

Schriften zur
Informations-
wissenschaft

74

Thomas Schmidt, Christian Wolff (Eds.)

Information between Data and Knowledge

**Information Science and its Neighbors
from Data Science to Digital Humanities**

Proceedings of the 16th International
Symposium of Information Science (ISI 2021)

Regensburg, Germany, 8th–10th March 2021



HOCHSCHULVERBAND
INFORMATIONSWISSENSCHAFT

vwh

Schmidt, Wolff (Eds.): Information between Data and Knowledge

Thomas Schmidt, Christian Wolff (Eds.)

Information between Data and Knowledge

Information Science and its Neighbors
from Data Science to Digital Humanities

Proceedings of the 16th International Symposium
of Information Science (ISI 2021)

Regensburg, Germany, 8th–10th March 2021

vwh

Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://www.d-nb.de> abrufbar.

Bibliographic information of the German National Library

The German National Library lists this publication in the German National Bibliography; detailed bibliographic data is available online at <http://www.d-nb.de>.

Der Tagungsband der ISI 2021 ist digital über den Publikationsserver der Universität Regensburg <https://epub.uni-regensburg.de/> via doi.org/10.5283/epub.44931 verfügbar, ferner über das Open-Access-Repository Zenodo, <https://zenodo.org/>.

The ISI 2021 Proceedings are available online via the University of Regensburg Publication Server, <https://epub.uni-regensburg.de/> by doi.org/10.5283/epub.44931, and, additionally, via the Zenodo open access repository, <https://zenodo.org/>.

© Verlag Werner Hülsbusch, Glückstadt, 2021

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

www.vwh-verlag.de

Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Glückstadt.

The Verlag Werner Hülsbusch, Glückstadt, owns rights of use for the printed version of this work.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

All trademarks used in this work are the property of their respective owners.

Satz/Typesetting: Werner Hülsbusch, Glückstadt

Druck und Bindung/Printing and binding: Schaltungsdienst Lange oHG, Berlin

Printed in Germany

ISSN: 0938-8710

ISBN: 978-3-86488-172-5

Inhaltsverzeichnis

Organisation und Sponsoren	11
Programmkomitee der ISI 2021	13
Vorwort	15
Keynotes	19
<i>Wolf Rauch</i>	
Was aus der Informationswissenschaft geworden ist	20
<i>Peter Bell</i>	
Digitale Kunstgeschichte und Bilddatenbanken als Synthese von Foto-, Dia- und Bibliothek	22
Session 1:	
Information Behavior and Information Literacy 1	23
<i>Joachim Griesbaum, Daphné Çetta, Thomas Mandl, Elke G. Montanari</i>	
What Is Information Literacy and How to Improve It?	24
<i>Athanasios Mazarakis, Isabella Peters, Jasmin Schmitz</i>	
#wenigerHashtagswirkenwissenschaftlicher	44
Der Zusammenhang von Tweet-Eigenschaften und wahrgenommener Wissenschaftlichkeit	
<i>Florian Meier</i>	
The Influence of Multilingualism and Mutual Intelligibility on Wikipedia Reading Behaviour	64
A Research Proposal	
<i>Vera Husfeldt, Sharon Alt</i>	
From Information Literacy to Data Literacy Education	73
A Case Study from Switzerland	

Session 2:	
Information Behavior and Information Literacy 2	79
<i>Jonas Ehrhardt, Timo Spinde, Ali Vardasbi, Felix Hamborg</i>	
Omission of Information: Identifying Political Slant via an Analysis of Co-occurring Entities	80
<i>Steffen Lemke, Paula Bräuer, Isabella Peters</i>	
Does the General Public Share Research on Twitter?	94
A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany	
<i>Tamara Heck, Ina Blümel</i>	
Open Practices of Early Career Researchers	115
A Qualitative Study on Research and Teaching Behavior	
<i>Thomas Schmidt, Elisabeth Salomon, David Elsweiler, Christian Wolff</i>	
Information Behavior towards False Information and “Fake News” on Facebook	125
The Influence of Gender, User Type and Trust in Social Media	
Session 3:	
Digital Humanities	155
<i>Thomas Schmidt, Johanna Dangel, Christian Wolff</i>	
<i>SentText</i> : A Tool for Lexicon-based Sentiment Analysis in Digital Humanities	156
<i>Manuel Burghardt, Jan Luhmann</i>	
<i>Same same, but different?</i>	
On the Relation of Information Science and the Digital Humanities	173
A Scientometric Comparison of Academic Journals Using LDA and Hierarchical Clustering	
<i>Michael Achmann, Florin Schwappach</i>	
Grundlagenermittlung für die digitale Werkbank qualitativ-hermeneutisch arbeitender Geisteswissenschaftlerinnen	200
Exploration geisteswissenschaftlicher Forschung mit Fokus auf Exzerpten und Literaturverwaltung	

Session 4:	
Information Retrieval	217
<i>Sebastian Schultheiß, Dirk Lewandowski</i>	
(Un)bekannte Akteure auf der Suchergebnisseite?	218
Ein Vergleich zwischen selbst eingeschätzter und tatsächlich vorhandener Suchmaschinenkompetenz deutscher InternetnutzerInnen	
<i>Carolin Schulz, Stefanie Elbeshausen, Christa Womser-Hacker</i>	
Kollaboratives Informationssuchverhalten von Personen unterschiedlicher Generationszugehörigkeit	247
<i>Sebastian Sünkler, Dirk Lewandowski</i>	
Den Einfluss der Suchmaschinenoptimierung messbar machen	273
Ein halb-automatisierter Ansatz zur Bestimmung von optimierten Ergebnissen auf Googles Suchergebnisseiten	
<i>Sebastian Sünkler, Dirk Lewandowski</i>	
Ist die Webseite suchmaschinenoptimiert?	299
Vorstellung eines Online-Tools zur Analyse der Wahrscheinlichkeit der Suchmaschinenoptimierung auf einer Webseite	
Session 5:	
Knowledge Representation	307
<i>Senan Kiryakos, Magnus Pfeffer</i>	
The Benefits of RDF and External Ontologies for Heterogeneous Data A Case Study Using the Japanese Visual Media Graph	308
<i>Luise Borek, Canan Hastik, Vera Khramova, Klaus Illmayer, Jonathan D. Geiger</i>	
Information Organization and Access in Digital Humanities TaDiRAH Revised, Formalized and FAIR	321
<i>Ingo Frank</i>	
Vom Datenkatalog zum Wissensgraph – Forschungsdaten im konzeptuellen Modell von FRBR	333

Session 6:	
Emerging Technologies	347
<i>Timo Breuer, Philipp Schaer</i>	
A Living Lab Architecture for Reproducible Shared Task Experimentation	348
<i>Fabian Deifuß, Cornelius Ihle, Moritz Schubotz, Bela Gipp</i>	
<i>procd</i> : A Privacy-preserving Robust Implementation to Discover Contacts in Social Networks	363
Session 7:	
Information Infrastructure	369
<i>Cornelia Veja, Julian Hocker, Christoph Schindler, Marc Rittberger</i>	
Educational Open Government Data in Germany	370
The Landscape, Status, and Quality	
<i>Sylvia Kullmann, Johannes Hiebl, Tamara Heck, Marc Rittberger</i>	
Potenziale von Open Educational Resources in der Informationswissenschaft?	397
<i>Heike Neuroth, Niklas Hartmann, Janine Straka, Ina Radtke, Ulrike Wuttke</i>	
Forschungsdatenmanagement in Brandenburg (FDM-BB)	405
Aufbau einer Landesinitiative	
Panel	413
<i>Vera Husfeldt, Sharon Alt</i>	
The Expanded Potential of Linking Different Types of Information and Data in the Digital Humanities	414
How Do Innovative Approaches Derived from Information Science and Data Science Contribute to a Knowledge Gain Across Disciplines?	
Doctoral Colloquium	419
List of Contributors / Teilnehmerliste	420

Gerhard Lustig Award Papers	421
<i>Mina Schütz</i>	
Detection and Identification of Fake News	422
Binary Content Classification with Pre-trained Language Models	
<i>Fabian Haak</i>	
Design and Development of an Emoji Sentiment Lexicon	432
<i>Neele Lange</i>	
Reflektive vs. intuitive Ästhetikwahrnehmung von Homepages	439
Eine Analyse kultureller Unterschiede zwischen Testpersonen aus Deutschland und Lateinamerika	
<i>Alexander Frummet</i>	
Towards the Identification of Information Needs in Conversational Search Dialogues	445
<i>Leyla Dewitz</i>	
Information Behavior von unbegleiteten, minderjährigen Geflüchteten unter Berücksichtigung der Rolle und Nutzung von Smartphones	452
<i>Sebastian Schultheiß</i>	
Der Einfluss des Verständnisses von Suchmaschinenwerbung auf das Rechercheverhalten am PC und am mobilen Endgerät	461
Eine Nutzerstudie	

Organisation und Sponsoren

Universität Regensburg, Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik

<http://mi.ur.de>



Universität Regensburg

Hochschulverband
Informationswissen-
schaft (HI) e.V.



HOCHSCHULVERBAND
INFORMATIONSWISSENSCHAFT

www.informationswissenschaft.org

Paper und Poster Chairs

Thomas Schmidt, Universität Regensburg

Raphael Wimmer, Universität Regensburg

Christian Wolff, Universität Regensburg

Lokale Organisation

Michael Achmann, Universität Regensburg

Alexander Bazo, Universität Regensburg

David Halbhuber, Universität Regensburg

Thomas Schmidt, Universität Regensburg

Christian Wolff, Universität Regensburg

Doktoranden-Kolloquium

Bela Gipp, Bergische Universität Wuppertal

Terry Ruas, Bergische Universität Wuppertal

Dr. Moritz Schubotzc, FIZ Karlsruhe

Gerhard-Lustig-Preis

Achim Oßwald, TH Köln

Studentisches Programm

Azra Eljezi-Bekiri, FH Graubünden, Chur

Marie Sautmann, Universität Regensburg

Programmkomitee der ISI 2021

Tatjana Aparac-Jelušić, Universität Zadar
Ingo Barkow, FH Graubünden, Chur
Bernard Bekavac, FH Graubünden, Chur
Christoph Bläsi, Johannes Gutenberg-Universität Mainz
Timo Breuer, TH Köln
Manuel Burghardt, Universität Leipzig
Maximilian Eibl, TU Chemnitz
David Elsweiler, Universität Regensburg
Rebecca Frank, Humboldt-Universität zu Berlin
Norbert Fuhr, Universität Duisburg-Essen
Maria Gaede, Humboldt-Universität zu Berlin
Ursula Georgy, TH Köln
Bela Gipp, Bergische Universität Wuppertal
Juan Gorraiz, Universität Wien
Michael Granitzer, Universität Passau
Joachim Griesbaum, Stiftung Universität Hildesheim
Vicente Pablo Guerrero-Bote, Universität Extremadura
Felix Hamburg, Universität Konstanz
Siegfried Handschuh, Universität St. Gallen
Gernot Heisenberg, TH Köln
Niels Henze, Universität Regensburg
Ulrich Herb, Universität des Saarlandes, Saarbrücken
Gerhard Heyer, Universität Leipzig
Hans-Christoph Hobohm, FH Potsdam
Isto Huvila, Universität Uppsala
Yvonne Kammerer, Leibniz-Institut für Wissensmedien, Tübingen
Markus Kattenbeck, TU Wien
Dowan Kim, Universität Pai Chai, Daejeon
Udo Kruschwitz, Universität Regensburg
Rainer Kuhlen, Universität Konstanz
Dirk Lewandowski, HAW Hamburg
Bernd Ludwig, Universität Regensburg
Thomas Mandl, Stiftung Universität Hildesheim

Zinaida Manžuch, Universität Vilnius
Philipp Mayr, GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim/
Köln
Florian Meier, Universität Aalborg, Kopenhagen
Norman Meuschke, Bergische Universität Wuppertal
Antje Michel, FH Potsdam
Heike Neuroth, FH Potsdam
Achim Osswald, TH Köln
Franjo Pehar, Universität Zadar
Isabella Peters, ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel/
Hamburg
Vivien Petras, Humboldt-Universität zu Berlin
Wolf Rauch, Karl-Franzens-Universität Graz
Gerhard Reichmann, Karl-Franzens-Universität Graz
Ulrich Reimer, OST – Ostschweizer Fachhochschule, Rapperswil
Jürgen Reischer, Universität Regensburg
Marc Rittberger, DIPF | Leibniz-Institut für Bildungsforschung und Bil-
dungsinformation
Philipp Schaer, TH Köln
Christian Schlögl, Karl-Franzens-Universität Graz
Thomas Schmidt, Universität Regensburg
Jasmin Schmitz, ZB MED – Informationszentrum Lebenswissenschaften,
Köln/Bonn
Stefan Schmunk, Hochschule Darmstadt
Wolfgang Semar, FH Graubünden, Chur
Melanie Siegel, Hochschule Darmstadt
Elisabeth Staudegger, Karl-Franzens-Universität Graz
Julia Maria Struß, FH Potsdam
Gabriela Tullius, Hochschule Reutlingen
Dirk Tunger, Forschungszentrum Jülich
Sirje Virkus, Universität Tallinn
Albert Weichselbraun, FH Graubünden, Chur
Raphael Wimmer, Universität Regensburg
Christian Wolff, Universität Regensburg
Christa Womser-Hacker, Stiftung Universität Hildesheim
Maja Žumer, Universität Ljubljana

Vorwort

Nach einer Pause im Jahr 2019, für das das 16. Internationale Symposium für Informationswissenschaft ursprünglich geplant gewesen war, findet das 16. Internationale Symposium für Informationswissenschaft im Jahr 2021 unter den Bedingungen der Covid-19-Pandemie erstmals virtuell statt. Das fröhliche Wiedersehen mit Kolleg*innen in der Kaffeepause oder der *social event* können daher nur mithilfe von Videokonferenzsystemen und anderen digitalen Plattformen stattfinden – *computer-mediated communication* zur Ermöglichung des Tagungsgeschehens, bereits das ein angemessen informationswissenschaftlicher Einstieg in die Tagung.

Das diesjährige Tagungsmotto – *Information zwischen Daten und Wissen. Die Informationswissenschaft und ihre Nachbarn von Data Science bis Digital Humanities / Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities* – thematisiert Entwicklungen der vergangenen Jahre, die die Informationswissenschaft in vielerlei Hinsicht berühren: Auf der einen Seite steht das Erstarren der digitalen Geisteswissenschaften in institutioneller, organisatorischer und personeller Hinsicht. Die digitalen Geisteswissenschaften zeigen Handlungsfelder traditioneller informationswissenschaftlicher Forschung, gleichzeitig erwächst durch die Etablierung eines weiteren, auch hilfswissenschaftlich zu interpretierenden, Lehr-, Forschungs- und Arbeitsfeldes auch Konkurrenz für die an Universitäten ausbaufähige Informationswissenschaft. Betrachtet man beispielsweise die aktuelle „Kleine Fächer“-Kartierung (Arbeitsstelle Kleine Fächer, 2020), so sind die digitalen Geisteswissenschaften unter den am stärksten wachsenden „kleinen Fächern“. Auf der anderen Seite, offenkundig bedingt durch die technischen Entwicklungen insbesondere im Bereich des maschinellen Lernens und des Deep Learning, kommt der Perspektive der Daten ein bisher unbekanntes Maß an Aufmerksamkeit zu. Aus informationswissenschaftlicher Sicht könnte man geneigt sein, das Zurückdrängen des Informationsbegriffs zugunsten der Betrachtung von Daten als Rückschritt zu betrachten. Einzuräumen ist, dass die neuartige Quantität und Qualität digitaler Daten diese Aufmerksamkeit rechtfertigt – und dabei neue informationswissenschaftliche Problemstellungen und Arbeitsfelder erzeugt (z.B. Metadaten, Forschungsdatenmanagement, digitale Repositorien). Im Ergebnis ist zu hoffen, dass über der neuen Datenseligkeit Probleme des Umgangs mit

Information nicht in den Hintergrund rücken. Jedenfalls ist das Spannungsverhältnis zwischen Data Science und Digital Humanities ein fruchtbares Feld für informationswissenschaftlichen Diskurs, wie dieser Tagungsband zeigen soll.

Der Tradition folgend, einen Tagungs- und Austauschort für die Informationswissenschaft im deutschsprachigen Raum zu bieten, bringt das Internationale Symposium für Informationswissenschaft auch in diesem Jahr Fachvertreter*innen insbesondere aus Deutschland, Österreich und der Schweiz (DACH) zusammen, um über aktuelle Forschungsvorhaben der Informationswissenschaft und benachbarter Fächer zu diskutieren.

Im Unterschied zu den beiden vorangegangenen Symposia, 2015 in Zadar und 2017 in Berlin, sind die Beiträge in diesem Jahr wieder in deutscher *und* englischer Sprache eingegangen. Dies ist der Überlegung geschuldet, dass ein regionales wissenschaftliches Treffen eine wichtige Aufgabe im Bereich der Nachwuchsförderung hat und gerade jüngeren Kolleg*innen und Doktorand*innen frühe Gelegenheit geben soll, ihre Arbeiten zu präsentieren. Die Möglichkeit, dies auch in der Muttersprache tun zu können, hat den Vorstand des Hochschulverbands Informationswissenschaft dazu bewogen, zu dieser Form der Multilingualität zurückzukehren. Die Stärkung des Deutschen als Wissenschaftssprache ist ein weiteres Argument, das hierfür spricht. Wie schnell sich pragmatische Aspekte der deutschen Wissenschaftssprache verändern, zeigen die vielfältigen Bemühungen um einen gendergerechten und inklusiven Sprachgebrauch. Die Herausgeber haben allerdings darauf verzichtet, den Autor*innen diesbezüglich konkrete sprachpolitische Vorgaben zu machen. Weitere Änderungen an den Einreichungsformaten betreffen insbesondere das Format bzw. den Umfang der Beiträge. In der Planung der Konferenz hat sich der Vorstand des Hochschulverbands Informationswissenschaft dem Vorbild anderer Fachtagungen angeschlossen und die Beitragsformate mit Blick auf die mögliche Beitragslänge an die Gepflogenheiten internationaler Fachtagungen angepasst.

Insgesamt wurden 2020 weniger Beiträge für die Tagung eingereicht, was unter anderem der Pandemiesituation, aber auch vergleichsweise kurzen Vorlaufzeiten geschuldet gewesen sein mag. In Verbindung mit den umfangreicheren Beitragsformaten ist daraus ein Tagungsband entstanden, der ein annähernd vergleichbares Volumen wie die Vorgängerausgaben aufweist.

Auch in diesem Jahr hat ein international besetztes Programmkomitee zur inhaltlichen Gestaltung und Qualitätssicherung der wissenschaftlichen Arbeiten beigetragen: Mehr als 60 Programmkomitee-Mitglieder aus elf Ländern

haben die Beiträge beurteilt und viele wertvolle Hinweise zu deren Verbesserung gegeben. Dabei wurden fast alle Beiträge von drei, manche sogar von vier Gutachter*innen bewertet. Den Mitgliedern des Programmkomitees sei für ihre exzellente Arbeit ein herzlicher Dank ausgesprochen.

Es ist gelungen, wesentliche Schwerpunkte aktueller informationswissenschaftlicher Forschungsarbeit angemessen zu repräsentieren. Dies zeigt ein Blick auf die für die Tagung geplanten Sessions:

Besonders erfreulich ist, dass der wichtige Aspekt der Informationsverhaltensforschung (bzw. der Forschung zu Informationsverhalten und Informationskompetenz als aufeinander bezogener komplementärer Perspektiven) einen klar erkennbaren Schwerpunkt der Tagung bildet – nur ihm sind zwei Sessions der Tagung gewidmet.

Weitere Sessions umfassen Beiträge zum Information Retrieval als dem wohl traditionsreichsten Arbeitsschwerpunkt der Informationswissenschaft, zu den digitalen Geisteswissenschaften als einem sich zuletzt sehr dynamisch entwickelnden Arbeits- und Forschungsgebiet, zur Wissensrepräsentation, zu Fragen sich neu entwickelnder Technologien (*emerging technologies*) und – last, not least – dem für Informationswissenschaftler*innen seit jeher besonders bedeutsamem Thema der Informationsinfrastruktur.

Neben den Fachbeiträgen ist es gute Tradition, auf dem Internationalen Symposium für Informationswissenschaft einen Wettbewerb um die beste Abschlussarbeit der vergangenen Tagungsperiode zu veranstalten. Die Siegerin / der Sieger wird mit dem Gerhard-Lustig-Preis ausgezeichnet, der nach dem Darmstädter Mathematiker, Informatiker und Informationswissenschaftler Gerhard Lustig benannt ist, der Pionierarbeiten auf dem Gebiet der Indexierung und des Information Retrieval geleistet hat und wegweisend für die deutschsprachige Informationswissenschaft gewesen ist. Auch 2020/2021 gab es hierfür zahlreiche Einreichungen, von denen sechs für die Präsentation auf der Tagung durch eine von Achim Oßwald, Köln, geleitete Jury ausgewählt wurden.

Ein weiteres, für die Nachwuchsförderung wichtiges Element ist das Doktoranden-Kolloquium, das Nachwuchswissenschaftler*innen einen geschützten Raum zur Präsentation und Diskussion ihrer Forschungsthemen jenseits der unmittelbaren Betreuung am jeweiligen Forschungs- und Arbeitsort ermöglichen soll und unter der Leitung von Bela Gipp, Wuppertal durchgeführt wird. Im Anschluss an die Tagung findet am 11. März 2021 der Workshop *eXpress Your Research* zur Wissenschaftskommunikation statt, ebenfalls ein Angebot, das sich vor allem an Nachwuchswissenschaftler*innen richtet.

Großer Dank gilt den Mitarbeiter*innen des Lehrstuhls für Medieninformatik an der Universität Regensburg, die in vielfältiger Weise die Vorbereitung und das Zustandekommen der Tagung unterstützt haben. Auch der Universität Regensburg und ihren zentralen Einrichtungen (Rechenzentrum, Universitätsbibliothek) ist für die Bereitstellung der technischen Infrastruktur sehr herzlich zu danken.

In bewährter Weise hat sich unser Verleger Werner Hülsbusch – selbst Diplom-Informationswissenschaftler Konstanzer Prägung und seit vielen Jahren eine wichtige Stütze des informationswissenschaftlichen Publikationswesens – nicht nur verlegerisch, sondern zudem als Lektor, Korrektor und Mentor für das informationswissenschaftliche Symposium engagiert – ganz herzlichen Dank dafür!

Schließlich ist dem Hochschulverband Informationswissenschaft und seinem Vorstand unter der Leitung von Vivien Petras für die großzügige finanzielle und organisatorische Unterstützung zu danken, die die Planung und Durchführung der Tagung erst möglich gemacht hat. Als *open conference* handelt es sich um eine wissenschaftliche Fachtagung, die allen Interessenten ohne Weiteres offensteht – auch das ein Novum!

Den Leser*innen des Tagungsbandes wünschen wir eine spannende und erkenntnisreiche Lektüre, den Teilnehmer*innen der Tagung anregende Diskussionen im vielfältigen Themenspektrum der Informationswissenschaft.

Regensburg, im Februar 2021

Thomas Schmidt und Christian Wolff

Literatur

Arbeitsstelle kleine Fächer (Hrsg.) (2020): Bericht zum Stand der Kartierung kleiner Fächer. Arbeitsstelle kleine Fächer bei dem Zentrum für Qualitätssicherung und -entwicklung, Johannes Gutenberg-Universität Mainz, August 2020. Online: https://www.kleinefaecher.de/fileadmin/user_upload/img/Kartierungsbericht_2020.pdf

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 15–18. DOI: doi.org/10.5283/epub.44932.

Keynotes

Was aus der Informationswissenschaft geworden ist

Wolf Rauch

Universität Graz, Österreich

wolf.rauch@uni-graz.at

Vor 30 Jahren, 1990, wurde das Symposium für Informationswissenschaft – ISI zum ersten Mal abgehalten. Vor 40 Jahren wurde der erste Lehrstuhl für Informationswissenschaft im deutschen Sprachraum eingerichtet (mit Rainer Kuhlen 1980 in Konstanz). Diese Jubiläen bieten eine Gelegenheit, der Frage nachzugehen, was aus der Informationswissenschaft bisher geworden ist.

In der ersten Generation von ca. 1950 bis 1980 (dem „Goldenen Zeitalter“ der Informationswissenschaft) wurde Information primär als öffentliches Gut verstanden. Staatliche Programme, Forschung an Universitäten, kostenfreier Zugang zum Wissen prägten diese Phase. An ihrem Ende, in den 1980er-Jahren, war die Informationswissenschaft mit eigenen Inhalten, Studienprogrammen und ersten Lehrstühlen an den Universitäten in bescheidenem Rahmen angekommen.

Die folgende Generation von 1980/1990 bis heute (das „Silberne Zeitalter“) brachte ein gradezu explosionsartiges Wachstum der Informationswissenschaft und hat dieses Fach zu einer der größten Wissenschaftsdisziplinen weltweit werden lassen – allerdings außerhalb der Universitäten. Die Ursache dafür war, dass Information in ihrem wirtschaftlichen Wert erkannt worden ist und ein riesiger Markt für Information, Kommunikation und Netzwerke entstanden ist. Eine Folge dieser Entwicklung war aber auch, dass ein Paradigmenwechsel in der Informationswissenschaft stattgefunden hat, weg von der geschriebenen Sprache und weg von Kausalität und Hypothesenbildung hin zu Künstlicher Intelligenz und Data Science.

Wenn die universitäre Informationswissenschaft (zumindest in Europa) wohl kaum eine Chance hat, im Bereich der Entwicklung von Systemen und Anwendungen wieder an die Spitze der Entwicklung vorzustoßen, bleiben ihr doch Gebiete, in denen ihr Beitrag in der kommenden Entwicklungsphase dringend erforderlich sein wird: Informationsethik, Informationskompetenz, Information Assessment. In den nächsten 30 Jahren, dem dritten Zeitalter der

Informationswissenschaft, wird sich entscheiden, in welche Richtung die Informationsgesellschaft unsere Welt verändern wird. Wir sollten uns daher gerade an den Universitäten nicht nur der Frage zuwenden, was wir mit dem Einsatz von Informationssystemen erreichen können, sondern was wir damit erreichen wollen bzw. welche Entwicklungen wir verhindern müssen.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 20–21. DOI: doi.org/10.5283/epub.44933.

Digitale Kunstgeschichte und Bilddatenbanken als Synthese von Foto-, Dia- und Bibliothek

Peter Bell

Department Medienwissenschaften und Kunstgeschichte,
Institut für Kunstgeschichte, Friedrich-Alexander Universität
Erlangen-Nürnberg, Deutschland

peter.bell@fau.de

Für Studierende der Kunstgeschichte ist das Dia ein fremdes Artefakt geworden. Auch die Fototheken werden meist nur noch digital genutzt. Ist also die Transformation ins Digitale konzeptionell abgeschlossen und es geht nur noch um das bloße Digitalisieren des noch Analogem? Inwieweit sind die Praktiken bzw. Produktionsweisen der Fotothek, Diathek und Bibliothek in die digitalen Bilddatenbanken eingegangen? Wo ergeben sich Defizite und Potenziale gegenüber den vorangegangenen Dispositiven – nicht nur durch das Digitale per se, sondern besonders bzgl. aktueller Entwicklungen (z.B. des maschinellen Lernens)? Entwickelt sich die digitale Kunstgeschichte dadurch zu einer bildgetriebenen Wissenschaft?

Die digitale Kunstgeschichte ist maßgeblich im Rahmen der Entwicklung von Bilddatenbanken entstanden und auch das mittlerweile ausdifferenzierte und ansatzweise etablierte Forschungsfeld kreist immer wieder um (zentrale) Bilddatenbanken. Dabei sieht sich die digitale Kunstgeschichte – etwa im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI) – geradezu als Anwalt der Bilder und unterstreicht deren Wichtigkeit als Daten und Bedeutungsträger. Um diese Überzeugung zu mehr zu machen als einem bloßen Selbstverständnis, empfiehlt sich der Blick in den Bild- und Mediengebrauch der Kunstgeschichte. In Etappen sollen bildzentrierte Arbeits- und Darstellungsweisen von Kunstgeschichte rekapituliert und ihr Fortbestehen und Fehlen in den Bilddatenbanken thematisiert werden. Es geht auch um das Verhältnis des Faches zu seinen Infrastrukturen. Zuletzt werfen alle Fragen nach dem Umgang mit dem Bild auch die Gegenfragen nach dem Text auf – ob als Metadaten oder Forschungsliteratur.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, p. 22.
DOI: doi.org/10.5283/epub.44934.

Session 1:

Information Behavior and Information Literacy 1

What Is Information Literacy and How to Improve It?

*Joachim Griesbaum, Daphné Çetta,
Thomas Mandl, Elke G. Montanari*

University of Hildesheim, Germany

{[griesbau](mailto:griesbau@uni-hildesheim.de), [cetta](mailto:cetta@uni-hildesheim.de), [mandl](mailto:mandl@uni-hildesheim.de), [montanar](mailto:montanar@uni-hildesheim.de)}@uni-hildesheim.de

Abstract

Information literacy is a controversially discussed topic with heterogeneous definitions. In the context of a project on “Information Literacy and Democracy” in Germany, 21 experts from diverse scientific backgrounds submitted position papers in which the authors provided their perspectives on the following questions: What is information literacy? How to foster information literacy? What are the central developments in the field of information literacy? What other aspects are relevant? Based on document analysis, this contribution unites the core messages of the 19 position papers for an international audience. Such a combined view of 21 experts from diverse scientific backgrounds is rather scarce. By merging the views of many experts and transcending disciplinary borders, this paper hopes to make a small contribution to stimulate multidisciplinary discussions on information literacy and corresponding approaches to improve it.

Keywords: information literacy; multidisciplinary; experts; document analysis

1 Introduction

Information literacy is a concept which has attracted much attention in Information Science. In the context of a project on “Information Literacy and Democracy” researchers from the University of Hildesheim aim to raise public awareness on the significance of information literacy, both for public and scientific discourse. By bringing together different stakeholders and actors

with diverse backgrounds, a broad discourse is initiated that aims to reveal a comprehensive picture of the subject information literacy, and possible approaches to foster information literacy. In the summer of 2019, a call for position papers was announced in the context of the first conference of the project¹. The goal was to collect the views of experts from different subject areas and professions on the following four questions: What is information literacy? How to foster information literacy? What are the central developments in the field of information literacy? What other aspects are relevant? Twenty-one authors answered the call and provided 19 position papers. The authors work within a wide range of subjects and professions. In addition to scientists from different subjects, librarians as well as journalists have contributed. Additionally, perspectives from political science, educational science, and teacher training expand the perspectives yielded by information science and library science. As a result, the position papers deliver a comprehensive view and diverse perspectives on information literacy. The position papers are available online in German language (Çetta et al., 2019). This contribution analyzes the 19 position papers for an international audience using document analysis that orients on inductive category formation as proposed by Mayring (2014). The authors believe this is worthwhile as such a perspective on information literacy is rather scarce. By unifying the views of many experts and transcending disciplinary borders, this paper hopes to make a small contribution to stimulate discussions on multidisciplinary views on information literacy. As indicated in the title of the paper, the research interest is on the question what constitutes information literacy and on approaches on how to improve it.

The paper is structured as follows. In the next chapter, we delineate information literacy as an evolving concept, portray multidisciplinary aspects of information literacy, discuss its relevancy and new approaches to improve it. In this way, the following explanations are contextually framed. In the next step, we describe the research aim and the analytical approach and procedure. Then, the papers are analyzed. Finally, the results are summarized and discussed. Please note that the authors of the position papers are the experts. The authors of this paper only build up upon their work.

1 The website of the conference is located at <https://informationskompetenz.blog.uni-hildesheim.de/>.

2 Information literacy as an evolving concept

According to literature, origins of information literacy can be traced back to a report written by Paul Zurkowski (1974) addressed to the *US National Commission on Libraries and Information Science* (Leaning, 2017). Zurkowski already stated at the beginning of the seventies that although most people today can read and write, only a small proportion of the population could actually be called information literate as “they have learned techniques and for utilizing the wide range of information tools” (p. 6). Another important milestone can be seen in the *Final Report of the Presidential Committee on Information Literacy* of the *American Library Association* (1989). Here, information literacy is understood as a behavioral and search process-oriented competency of the individual user. At the same time, the connection between information literacy and critical thinking, lifelong learning and social participation is already proclaimed here. Starting in the 1990s, further dissemination of information literacy concepts can be observed on an international level. For example, in 1999, the *Society of College, National and University Librarians* (SCONUL) proclaimed a model which consists of seven “pillars” of competence. The seventh pillar, *synthesis and create*, includes the generation of knowledge as a part of information literacy (SCONUL Advisory Committee on Information Literacy, 1999). In the year 2000, the *American Library Association* published the *Information Literacy Competency Standards for Higher Education*. Here, five behavioral standards with associated performance indicators and outcomes are prescribed. The model is behavioristic in principle, still, the inclusion of social and normative aspects of information behavior becomes clear. In the UNESCO *Global media and information literacy (MIL) assessment framework*, published in 2013, media and information literacy are connected and defined as “a set of competencies that empowers citizens to access, retrieve, understand, evaluate and use, create, as well as share information and media content in all formats, using various tools, in a critical, ethical and effective way, in order to participate and engage in personal, professional and societal activities” (p. 29). This framework has a wide focus on media and information literacy. It draws a normative connection between information literacy and individual and societal well-being. Finally, the *Framework for Information Literacy for Higher Education* provided by the *Association of College & Research Libraries* (ACRL, 2016), is based on six “frames”. Within each frame “knowledge practices” and “dispositions” are

described for learners developing information literate abilities. The ACRL framework is less deterministic and process-oriented than earlier concepts of information literacy. Rather, it aims to enhance users' comprehension of ontological and epistemological aspects of information (environments). In sum and seen from a historical perspective, it can be stated that the discourse on information literacy has become increasingly holistic. The original scope of information literacy, whose focus was on the individual user and behavior-based and also search process-related has expanded. Newer approaches additionally include a citizen and society related perspective. They take in account the context of the development and interests in knowledge generation as well as epistemological aspects.

Information literacy seems to be a rather broad concept. It is intertwined with other digital literacies. For example, the UNESCO bundles media and information literacy. The ACRL framework explicitly refers to the metaliteracy approach of Mackey and Jacobson (2014). In Germany, in the educational system, the term information literacy is known in the higher education area. In schools, teacher training, and in the curriculum, rather the term media literacy is used (e.g., Weisel, 2017; NLQ Hildesheim, 2015). In the ICIL 2018 study, Fraillon et al. (2019) speak of computer and information literacy. These "double"-competences explicitly show that the single competences are strongly interwoven and benefit from each other. E.g., the ability to act in an informational self-determined manner is important for the competent use of digital technologies and vice versa (ibid.). From the user's point of view, the respective context of action and his or her pragmatics are decisive for the adequate "competency bundle" that is required. Nevertheless, it is useful to define a core of information literacy which helps to identify the learning related areas that need to be addressed to foster information literacy. The historical perspective laid out above suggests that the ability to uncover and satisfy information needs in a self-determined manner is one central aspect and can be considered as the original core. Especially the ACRL framework emphasizes that information literacy nowadays also includes the ability to understand and reflect information on an ontological and epistemological level.

Concerning its significance, we can postulate information literacy touches on more stages and areas of life and is more important than ever. Von Loh and Henkel (2014) argue that information and media competence is already important in early childhood education. Furthermore, although often under the hood of media literacy, there seems to be an increasing integration of

information literacy into school education (cp. above). In higher education, information literacy expands to include aspects such as electronic publishing (e.g., open access) and research data management (Koltay, 2017). Beyond the academic sector, although still in the niche, the awareness of the relevance of information literacy also increases (Travis, 2017). Especially in everyday and societal life, information literacy becomes more important. Indeed, using search engines is easy. Still, information behavior in everyday life contexts is far from being trivial. In fact, information markets are complex and difficult to see through. In e-commerce, naive users are in danger of being manipulated or getting caught by fraudsters (Wang et al., 2014). In health-related areas, information literacy is directly related to personal well-being (Cano-Orón, 2019). With regard to political information, information literacy is seen as a critical basis for democracy (Khan, 2020). And we know, not all is well. Some researchers are seriously alarmed and are concerned about democracy (Breakstone et al., 2018; Morgan, 2018). With regard to information literate behavior of users, research shows there is much to be wished for. The ICIL study (Fraillon et al., 2019) indicates that many pupils only show a basic level of information literacy. Pursuant to Wineburg et al. (2016), many students struggle to assess the credibility of information. Users often rely on rather simple evaluation heuristics (such as familiarity or popularity) (Metzger & Flanagin, 2013). Typical patterns of search engine usage show preferences for top positions and specific result types (White, 2016, pp. 64–82). Queries are usually short and often users are satisfied with a single result or the information given directly on search result pages (Fishkin, 2017). In addition to these effort minimizing strategies, cognitive distortions are also a common phenomenon, such as confirmation errors, i.e., the preference for results that support the user's own point of view (Knobloch-Westerwick et al., 2015). The mentioned aspects make it apparent that the development of information literacy succeeds, and only then succeeds, when users really understand their information environment and are actually motivated to behave in an information literate manner. With regard to the latter, information literacy curricula that lead users to perform in-depth analysis of knowledge artefacts, e.g., websites to determine their credibility are increasingly criticized (Fielding, 2019). Such check-list based approaches like the CRAAP (Currency, Reliability, Authority, Accuracy, and Purpose) method originate from academic contexts. For everyday life information behavior, they were evaluated as unfit for two reasons. First, if such an analysis of knowledge artefacts is carried out carefully, it would be very costly, which seems unrealistic. As a result,

users typically diminish their effort. Secondly, if users take such a pragmatic approach and reduce their effort, they usually rely on very simple criteria such as the top domain type, the design of the website, self-declarations of expertise or the existence of graphics to assess quality. This is easily misleading (Warner, 2019). A lateral reading strategy is recommended – a procedure that professional fact-checkers also show (Wineburg & McGrew, 2017). Some newer initiatives that aim to build up information literacy in everyday life are based on such considerations (e.g., Caulfield, 2019).

So much to a literature-based background on information literacy as an evolving concept, its multidisciplinary aspects, its relevancy and new considerations on how to improve it. In the following, we will see into how far the experts correspond to or expand on the perspective drafted in this chapter.

3 Research aim, analytical approach, and procedure

The aim of our research is to get a comprehensive overview and multidisciplinary perspectives on information literacy and its perception. The pragmatic goal is to get insights on two questions: a) What is (understood of) information literacy? and b) How to improve information literacy? The hope is to come up with a current view of the field that is up-to-date and, somewhat representative. Of course, representativeness is difficult to achieve in qualitative research. Still, a merger of the views of many experts from different fields is more than the publicly available views and insights users, students, librarians, and researchers are encountering when researching the field.

Data collection and the resulting data set were as follows. The call for the position papers was disseminated in June 2019. Potential contributors were invited per email. In addition, the call was posted in mailing lists. A template was provided for the contributors that regulated copyright issues (in which the authors granted a non-exclusive right of publication) and specified the expected length of the position paper. Each of the four questions should be answered in 150–300 words. Thus, each position paper was targeted to be 2–4 pages long. Till October 2019, 19 position papers by 21 experts were submitted. Overall, the collection papers are 56 pages long and consist of roughly 17,000 words (Çetta et al., 2019).

How should the data be analyzed? The pre-defined questions resemble very generic perspectives on the topic information literacy (what it is, how to foster it, trends and other aspects). They do not structure information literacy in content-related or action-oriented pre-defined categories. In contrast, the analysis aims to reveal such perspectives of the experts. Thus, an inductive approach of analysis seems to be appropriate. The analytical procedure is oriented towards inductive category formation as proposed by Mayring (2014, pp. 79–87). The analysis was sequenced as follows:

1. The level of categories was pre-defined by the four questions that structured the position papers. The level of abstraction was determined by the topics found, not specific aspects of them. For example, with regard to the first question “What is information literacy”, e.g., *research expertise* is seen as a topic whereas the *ability to recognize an information need* is not seen as a topic of its own but as a part of the topic *research expertise*.
2. The first author of the paper worked through all of the 19 position papers and formulated categories. At the same time, the second author worked through five of the position papers and formulated categories.
3. Both authors developed an agreed collection of categories.
4. The first author again worked through the material and revised the categories, partly re-arranging the category system by building main and sub-categories.
5. The category system was checked and finalized in a discussion of the first author with the second author of the paper.

Microsoft Excel was used for the analysis. The analysis was carried out in March 2020.

4 Analysis

The analysis is structured as follows. First, we describe the profile of the experts. In the second step, we provide an overall consolidation of position papers. We first display the main categories across all four questions. Then, we get more specific and discuss the insights we gained about the single questions.

4.1 Profile of the experts

Overall, 21 experts contributed 19 position papers. Seventeen papers were written by one author and two papers by two authors. The majority of the experts were librarians (8) and information scientists (7). Furthermore, four educational scientists with a focus on teacher education, one communication scientist (a former journalist) with a focus on journalism and one teacher (also active as a media education expert) were involved. Two of the educational scientists co-authored a paper. Together, the communication scientist and the teacher also wrote one paper. Table 1 provides an overview.

Table 1:

Overview of the authors' profession and the number of position papers

Profession	Number of authors	Number of position papers
Librarian	8	8
Information scientist	7	7
Educational scientist with a focus on teacher education	4	3
Communication scientist with a focus on journalism (former journalist)	1	0,5
Teacher	1	0,5

Summarizing, we can see that information scientists and librarians make up the majority of the sample. In addition, educational scientists constitute a substantial part. Finally, individual views from the teaching and journalistic professions are also included.

4.2 Overview of categories

Table 2 provides an overview of all main categories across all four questions. This overview clarifies, first and foremost, information literacy is still seen as the ability to conduct research. This is the core of information literacy. In addition to that, most of the experts write that information literacy should not be seen in isolation but is overlapping and related to other competencies.

Table 2:

Overview of the main categories (fraction of papers mentioning the category)

Questions	Categories
What is information literacy?	<ul style="list-style-type: none"> research skills (inclusive evaluation competence) (79%) overlap with other competencies (64%) critical thinking (53%) societal participation (42%) competence in scientific work (32%) information processing and dissemination (26%) ethical use of information (26%) domain-specific (16%)
How to foster information literacy?	<ul style="list-style-type: none"> content (74%) places of learning (74%) didactic (58%) basic requirements (learning provision) (37%) intermediaries (26%) learning requirements (users) (16%)
What are the central developments in the field of information literacy?	<ul style="list-style-type: none"> new subject areas (extensions of the concept) (68%) evaluation of information becomes the central sub-competence (37%) information literacy becoming more important (32%) information literacy becoming more complex and comprehensive (32%) individual and adaptive mediation (21%) improved integration of learning provision (21%)
What other aspects are relevant?	<ul style="list-style-type: none"> necessity of a holistic epistemic perspective (26%) information literacy learning through gamification (11%) need to strengthen networking between institutions and intermediaries (11%) significance of information science (11%) certification (5%) cultural aspects (5%) fostering motivation to learn information literacy (5%) negative societal change (5%)

Roughly half of the authors argue that critical thinking is a part of information literacy. Large groups of the experts connect information literacy with the ability to participate in society and the competency for scientific

work. In addition to that, a fourth of the papers argue that information processing, dissemination and the ethical use of information also belong to information literacy. For some authors, it is important to mention that information literacy is domain-specific.

Referring to the question on how to foster information literacy, data shows that for most experts the learning content and the places of learning are a decisive factor. A majority also mentions didactical aspects as important for competency development. Furthermore, basic requirements on the supply side and requirements on part of the users are discussed. A quarter of the experts state a wide array of different kinds of intermediaries that are or should be involved in information literacy-related learning.

Concerning trends in the field of information literacy, two-thirds of the experts mention new subject areas to which information literacy should be related to. One-third of the experts argue that the ability to evaluate information has developed into the central part of information literacy. Respectively, also one-third assesses information literacy as becoming more important and also more complex and comprehensive. A fifth of papers request that learning supply should be more strongly focused on the needs of the individual learner and be adaptive. Two papers demand that learning provision should be more strongly integrated into the educational system. Also, two papers claim a greater personal responsibility for being information literate.

Finally, the analysis of the “other aspects” shows many issues already mentioned within the first three questions. Nevertheless, some hitherto unnamed facets are revealed, e.g., two authors contemplate on the role of information science in information literacy education. Besides, certification and culture are revealed as important aspects of information literacy.

4.3 Question 1: What is information literacy?

As mentioned, at the core, information literacy is seen as one’s individual ability “to paint a picture of the world around us that comes close to reality”. It encompasses research skills and critical thinking, scientific competencies and is a central foundation for lifelong learning, informal self-determination, and societal participation. Knowledge should also be used constructively and ethically. Some authors hint that information literacy is not a generic skill, fitting to any context, but highly domain-specific. Here, we can observe some tension between the generalizability of information literacy as basic

literacy and its affordances in specific informational contexts. One fourth of the papers also include information processing and further dissemination as a sub-competency of information literacy. Besides, information literacy is related to other competencies. Digital literacy is mentioned in one-third of the papers, data literacy and media literacy are also pointed out.

Table 3 contains the main and subcategories for questions one in detail.

Table 3: Question 1 – Overview of the main and sub categories (fraction of papers mentioning the category)

Question 1: main and sub categories
- research skills (inclusive evaluation competence) (79%)
- overlap with other competencies (64%)
▪ digital literacy (37%)
▪ data literacy (26%)
▪ media literacy (21%)
- critical thinking (53%)
- societal participation (42%)
- competence in scientific work (32%)
- information processing and dissemination (26%)
- ethical use of information (26%)
- domain-specific (16%)

4.4 How to foster information literacy

The experts provided manifold ideas to improve information literacy. Most of the answers were concerned with the content and places of learning. Also, didactical aspects, basic requirements of learning provision, intermediaries and learning requirements on part of the users were mentioned. Regarding the content, there is a field of tension concerning the delimitability of the rather generic research-oriented core area of information literacy and its relationship to individual subject-specific or pragmatic contexts. One expert wrote: “courses on search techniques or database applications are certainly justified, but they are ultimately only accessible to a limited group of society and can only cover a small part of the required skills.” That was kind of common sense, a generic basic understanding and research courses are helpful. Nevertheless, information literacy learning has to be subject-specific and should tie-up with learners’ world of experience.

There are many venues where information literacy can be learned. “Typical” places like schools, universities, libraries were mentioned most often. Several authors wrote that information literacy education should start at the time of entry into the school system. Furthermore, the experts made clear, that beyond the typical places, there are more areas out there where information literacy can be fostered, e.g., the working environment, private areas and even kindergartens, and museums. The answers grouped into the didactic category can be read like a table of contents on current trends in (e-)learning: gamification, self-directed learning, collaborative learning, flipped classroom and others. Two librarians mentioned information literacy frameworks and models. Two authors wrote information literacy education should orient on journalistic approaches.

In addition to the many “places” to learn, there are many ways seen to make information literacy education and learning effective and interesting. Still, information literacy education has its conditions for success. On the one hand, teaching staff must be trained, and information literacy has to be integrated into the curriculum. On the other hand, learners are confronted with a variety of cognitive and motivational demands. They also need to be motivated and able to learn information literacy, as the following comment illustrates: “The requirements for imparting any kind of ‘information literacy’ are, [...] the desire to increase knowledge (which is more important than the actual information), the will to seek information, and, above all, the scientific theoretical qualification and attitude.” To motivate learners, provocation, and pointing out the uncertainty and transparency of their information environment is seen as an anchor for access. In addition to teaching staff and librarians, the personal social environment, that is family, friends, and the users themselves are mentioned. In sum, the experts provide us with a collection of many touchpoints and opportunities to foster information literacy.

Table 4 contains the main and subcategories for questions two in detail.

*Table 4: Question 2 – Overview of the main and sub categories
(fraction of papers mentioning the category)*

Main categories	Sub categories
content (74%)	subject-related (53%) relation to learner's own world of experience (42%) basic understanding and research courses (16%)
places of learning (74%)	school (58%) university (32%) library (26%) adult education and citizens (16%) private area (16%) work environment (16%) further education (5%) kindergarten (5%) museum (5%)
didactic (58%)	e-learning (16%) collaborative learning (11%) applying frameworks and models (11%) gamification (11%) research based learning (11%) self-directed learning (11%) adapting to subject-specific epistemic cultures (5%) citizen science, real laboratories (5%) flipped classroom (5%) intermediaries setting an example (5%) orienting on the journalistic approaches (5%)
basic requirements (learning provision) (37%)	technical and curricular prerequisites (21%) staff prerequisites (21%) further training of teachers (16%)
intermediaries (26%)	teaching staff (16%) friends (14%) family (11%) libraries (5%) users themselves (5%)
learning requirements (users) (16%)	motivation (11%) critical thinking (11%)

4.5 Question 3: What are the central developments in the field of information literacy?

Concerning trends in the field of information literacy, the analysis shows the following main aspects. Two-thirds of the papers mentioned that information literacy is an expanding topical area. Respectively, one-third considers information literacy as getting more important and complex and wrote that the evaluation of information becomes the central sub-competence. Furthermore, a fifth of the papers argue an improved integration of learning provision, with regard to improved cooperation of intermediaries and the integration into the educational system. Finally, a fifth of the papers see a need for individual and adaptive mediation. Also, two papers argue a greater personal responsibility for information literacy.

With regard to the expanding topical area, the ‘old’ core of a primarily search-related information literacy is widely expanded. The most prominent single sub-topic mentioned is fake news. Furthermore, information literacy nowadays includes aspects of personal information management (data protection, IT security) and processes of scientific and commercial knowledge genesis. Moreover, there is a wide range of other topics, such as the structures and players in the information market (influencers, publishers, open access, ...), technical basics and implications of digitization (text and web mining, search and recommendation algorithms, ...). Depending on one’s point of view, information literacy is mostly unbounded in terms of content. Here again, we are confronted with the question where does information literacy start and where does it end? It seems information literacy requires a comprehension of ontological and epistemological aspects of many information environments. The scientific information market is prominently mentioned here. At the same time, everyday information behavior is very present. Besides, according to some experts, aspects of daily information management also are an inherent part of information literacy.

Table 5 contains the main and subcategories for questions three in detail.

Table 5: Question 3 – Overview of the main and sub categories (fraction of papers mentioning the category)

Question 3: main and sub categories
<ul style="list-style-type: none"> - new subject areas (extensions of the concept) (68%) <ul style="list-style-type: none"> ▪ fake News (37%) ▪ scientific work and scientific communities (26%) ▪ data protection (21%) ▪ (search) algorithms (21%) ▪ open educational resources and open access (16%) ▪ use of social media (16%) ▪ copyright (11%) ▪ information markets (11%) ▪ text mining and data mining (11%) ▪ use of collaborative work systems (11%) ▪ altmetrics (5%) ▪ linked open data (5%) ▪ use of cloud services (5%) - evaluation of information becomes central sub-competence (37%) - information literacy becoming more important (32%) - information literacy becoming more complex and comprehensive (32%) - individual and adaptive mediation (21%) - improved integration of learning provision (21%) <ul style="list-style-type: none"> ▪ improved integration into the educational system (16%) ▪ improved cooperation of intermediaries (11%)

4.6 Question 4: What other aspects are relevant?

Finally, experts were asked to address other aspects that they felt were important. As written, many points mentioned here were already presented before. Nonetheless, some new facets of and on information literacy were revealed. One expert wrote that information literacy also relates to cultural, social, geographical, political and temporal contexts. As far as we know, little research has been done on cultural and intercultural aspects of information literacy. Information literacy is widely treated as culture-agnostic which is surely a deficiency. In addition, two authors argue the relevance of information science research for information literacy education. One author formulates regret regarding a lack of visibility of information science. The

other author doubts the significance of the information science community: “German-language information behavior research is completely atrophied with regard to the interdisciplinary study of user behavior and the handling of information from an intentional perspective”. Concluding, some authors re-emphasize the necessity for a holistic perspective on the field and the significance of information literacy, not only for the individual but for society: “To realize effective social communication in the long term, it is necessary to strive for certain standards in information transfer and communication design in addition to critical reception behavior, and to commit the students (like all other participants in communication) to these standards.” That means information literacy is not only a skill but also a commitment.

5 Discussion and conclusion

What can we learn from this analysis and how should this research be categorized? With regard to the latter aspect, as written, the paper aims to get a comprehensive view and diverse perspectives on information literacy. First, it has to be remarked, research here is of exploratory value. For an expert sample, the sample is not that small, but it can in no way claim to be representative, not even for the German community of which its members consist of. In this way, the results here should not be seen as a kind of joint will of the involved communities. Still, the opinion papers resemble more than a solitary voice. At least in Germany, at current, it is the best representation of the topic that goes beyond individual disciplines. With the many viewpoints included, the collection could be used as an orientation and guideline of things to consider and ways to go, when the aim is to promote and foster information literacy.

With regard to the first question what is (understood of) information literacy we can conclude the following. As a first result, we can summarize for all their diversity in details the position papers show a relatively uniform view. First, information literacy is seen as an individual’s ability “to paint a picture of the world around us that comes close to reality”. It is seen as an essential aspect of modern life. The content and places of learning are decisive for fostering information literacy. Information literacy has expanded, covers many new areas and is intertwined with other literacies. The ability to evaluate information is one of the most decisive factors for being an infor-

mation literate person, which is also important seen from a societal perspective. Thus, if we relate the experts' insights with the literature-based overview in Chapter 2, we can state that both point in the same direction. Information literacy is a very broad and expanding field that shows significance for scientific contexts but also everyday information behavior in private or professional contexts. A holistic perspective and epistemic approach, most often named as critical thinking, is necessary.

The second question asked for insights on how to improve information literacy. Here, experts acknowledge the need to build up a basic understanding of information literacy, e.g., with the help of "search techniques" courses. But it is clear, information literacy education is so much more. It has to be context specific, that means first and foremost specialized and relate to the learners' world of experience. It should be integrated in the educational system, and we should also think beyond that. Information literacy can be experienced in work environments, adult and further education and even in the cultural leisure context (museums). The same is true for the intermediaries. According to the experts, we should again think beyond typical educators, and teaching staff, e.g., family, and friends are significant too. Concerning didactical arrangements of learning environments, the experts provide many ideas to make information literacy learning available, motivating and effective. If we compare the experts' recommendations with the initiatives presented in Chapter 2, we see some deviations. The recommendations of the experts are much more generic with regard to didactical aspects and the content of learning environments, than the mentioned initiatives in Chapter 2. Those rather focus on users' acquisition of worthwhile heuristics (lateral reading). Still, both perspectives, those reported on in the literature section and the views mentioned in the position papers, point in the same direction. They emphasize the importance of ontological and epistemic dimensions of information literacy. For that reason, it can be argued that the ACRL framework would be a good point of orientation for many of the ideas mentioned on how to improve information literacy.

As a concluding remark, we argue that the position papers show that information literacy is a basic literacy for the digital world. In many contexts there is an omnipresent need to be information literate. In addition to that, the position papers point out ways of leaving the niche in which information literacy is de facto still located, at least in the education system in Germany. According to the authors' assessment, interdisciplinary and trans-professional cooperation is especially worthwhile as the collection of the position

papers shows. Beyond stimulating the discussion, maybe the cooperation made visible here, serves as an anchor and starting point to which following joint initiatives for the promotion of information literacy will be linked to. Due to the importance of the topic, we hope for further initiatives and discussions. The project “Information Literacy and Democracy” ends in 2021, the cooperation hopefully continues and expands.

Acknowledgements

The authors of the paper want to thank the experts who provided such an impressive collection of perspectives and insights to the community.

References

- ACRL (2016). *Framework for Information Literacy for Higher Education* | Association of College & Research Libraries (ACRL). <http://www.ala.org/acrl/standards/iframeframework>
- American Library Association (1989). Presidential Committee on Information Literacy: Final Report. <http://www.ala.org/acrl/publications/whitepapers/presidential>
- American Library Association (2000). *Information Literacy Competency Standards for Higher Education*. <https://alair.ala.org/bitstream/handle/11213/7668/ACRL%20Information%20Literacy%20Competency%20Standards%20for%20Higher%20Education.pdf?sequence=1&isAllowed=y>
- Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Teaching students to navigate the online landscape. *Social Education*, 82(4), 219–221.
- Cano-Orón, L. (2019). Dr. Google, what can you tell me about homeopathy? Comparative study of the top10 websites in the United States, United Kingdom, France, Mexico and Spain. *El profesional de la información*, 28(2). <http://www.doi.org/10.3145/epi.2019.mar.13>
- Caulfield, M. (2019). Check, Please! Starter Course Released [Online Course]. *Hapgood*. <https://hapgood.us/2019/08/13/check-please-starter-course-released/>
- Çetta, D., Griesbaum, J., Mandl, T., Montanari, E. (Eds.) (2019). Positionspapier: Informationskompetenz und Informationskompetenzvermittlung: Aktueller Stand und Perspektiven. Projekt: Zukunftsdiskurse: Informationskompetenz und Demokratie (IDE): Bürger, Suchverfahren und Analyse-Algorithmen in der politischen Meinungsbildung, Universität Hildesheim, 12–13, <http://informationskom->

petenz.blog.uni-hildesheim.de/files/2019/10/Projekt_Informationskompetenz_und_DemokratieKompilation_aller_Positionspapiere.pdf

- Fielding, J. A. (2019). Rethinking CRAAP: Getting students thinking like fact-checkers in evaluating web sources. *College & Research Libraries News*, 80(11), 620–622. <https://doi.org/10.5860/crln.80.11.620>
- Fishkin, R. (2017). The State of Searcher Behavior Revealed Through 23 Remarkable Statistics. *Moz*. <https://moz.com/blog/state-of-searcher-behavior-revealed>
- Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). *IEA International Computer and Information Literacy Study 2018 Assessment Framework*. <https://doi.org/10.1007/978-3-030-19389-8>
- Khan, S. (2020). Negotiating (dis) Trust to Advance Democracy through Media and Information Literacy. *Postdigital Science and Education*, 2(1), 170–183. <https://doi.org/10.1007/s42438-019-00072-9>
- Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, 20(2), 171–187. <https://doi.org/10.1111/jcc4.12105>
- Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3–14. <https://doi.org/10.1177/0961000615616450>
- Leaning, M. (2017). *Media and information literacy: An integrated approach for the 21st century*. Cambridge, MA: Chandos Publishing.
- Loh, S. G. von, & Henkel, M. (2014). Information and media literacy in kindergarten. In *European Conference on Information Literacy*. Cham: Springer, 253–262. https://doi.org/10.1007/978-3-319-14136-7_27
- Mackey, T., & Jacobson, T. (2014). *Metaliteracy: reinventing information literacy to empower learners*. Chicago: ALA Neal-Schuman.
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt. URN: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210–220. <https://doi.org/10.1016/j.pragma.2013.07.012>
- Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, 3(1), 39–43. <https://doi.org/10.1080/23738871.2018.1462395>
- NLQ Hildesheim (2015). *Orientierungsrahmen „Medienbildung in der Schule“*. [Arbeitsfassung – Stand: 30.10.2015]. <https://www.nibis.de/uploads/rediedl/me->

[dienportal/OR/Orientierungsrahmen%20Medienbildung%20in%20der%20Schule%20-%20Arbeitsfassung.pdf](#)

- SCONUL Advisory Committee on Information Literacy (1999). Information skills in higher education: A SCONUL position paper. https://www.sconul.ac.uk/sites/default/files/documents/Seven_pillars2.pdf
- Travis, T. A. (2017). From the classroom to the boardroom: the impact of information literacy instruction on workplace research skills. *Education Libraries*, 34(2), 19–31. <http://dx.doi.org/10.26443/el.v34i2.308>
- UNESCO (2013). *Global Media and Information Literacy (MIL) Assessment Framework: Country Readiness and Competencies*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000224655>
- Wang, D. Y., Der, M., Karami, M., Saul, L., McCoy, D., Savage, S., & Voelker, G. M. (2014). Search + Seizure: The Effectiveness of Interventions on SEO Campaigns. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. ACM, 359–372. <https://doi.org/10.1145/2663716.2663738>
- Warner, J. (2019). Getting Beyond the CRAAP Test: A Conversation with Mike Caulfield. *Inside Higher Ed*. <https://www.insidehighered.com/blogs/just-visiting/getting-beyond-craap-test-conversation-mike-caulfield>
- Weisel, L. (2017). Ten years after – Stand und Perspektiven der DGI-Initiative für Informationskompetenz, Teil I – Sachstand. *Information – Wissenschaft & Praxis*, 68(4), 246–252. <https://doi.org/10.1515/iwp-2017-0058>
- White, R. W. (2016). *Interactions with Search Systems*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139525305>
- Wineburg, S., & McGrew, S. (2017). Lateral Reading: Reading Less and Learning More When Evaluating Digital Information. Stanford History Education Group-Working Paper No. 2017-AI. SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3048994
- Wineburg, S., McGrew, S., Breakstone, J., & Ortega, T. (2016). Evaluating Information: The Cornerstone of Civic Online Reasoning. Stanford Digital Repository. <http://purl.stanford.edu/fv751yt5934>
- Zurkowski, P. G. (1974). The Information Service Environment Relationships and Priorities. Related Paper No. 5; Report to the National Commission on Libraries and Information Science. <http://files.eric.ed.gov/fulltext/ED100391.pdf>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 24–43. DOI: doi.org/10.5283/epub.44935.

#wenigerHashtagswirkenwissenschaftlicher

Der Zusammenhang von Tweet-Eigenschaften und wahrgenommener Wissenschaftlichkeit

Athanasios Mazarakis

ZBW – Leibniz-Informationszentrum
Wirtschaft und Christian-Albrechts-
Universität zu Kiel
a.mazarakis@zbw.eu

Isabella Peters

ZBW – Leibniz-Informationszentrum
Wirtschaft und Christian-Albrechts-
Universität zu Kiel
i.peters@zbw.eu

Jasmin Schmitz

ZB MED –
Informationszentrum
Lebenswissenschaften
schmitz@zbmed.de

Abstract

Der Beitrag berichtet die Ergebnisse einer Proof-of-Concept-Studie zur wahrgenommenen Wissenschaftlichkeit von Tweets. Im Rahmen eines online durchgeführten Experiments wurden Proband_innen in einem Fragebogen manuell thematisch-kategorisierte Tweets aus einer Tweet-Sammlung zur Bewertung vorgelegt, die aus medizinischen Hashtags generiert wurden. Es wurde ermittelt, welche Eigenschaften und Themen von Tweets einen Einfluss auf ihre wahrgenommene Wissenschaftlichkeit haben. Die Ergebnisse beruhen auf einer Analyse von 162 klassifizierten Tweets in deutscher und englischer Sprache und ihrer Bewertung durch 109 deutschsprachige Proband_innen. Die Ergebnisse zeigen unter anderem, dass Tweets mit vielen Hashtags eher als „nicht wissenschaftlich“ wahrgenommen werden. Eindeutig zugeordnet wurden von Proband_innen solche Tweets, die sich mit den Kategorien „Methodik“, „Job“ und „Spaß“ beschäftigen, wobei nur zur ersteren Kategorie zugehörige Tweets überwiegend als „wissenschaftlich“ bezeichnet wurden.

Keywords: wahrgenommene Wissenschaftlichkeit; Social Media; Twitter; Informationsverhalten

1 Einleitung und Motivation

Soziale Medien, wie Twitter oder Facebook, werden mittlerweile sehr häufig in der Wissenschaft eingesetzt, um aktuelle Forschungsergebnisse zu kommunizieren (McKendrick/Cumming/Lee, 2012) oder auch mit verschiedenen Stakeholdern zu diskutieren (Salem/Borgman/Murphy, 2016; Sugimoto et al., 2017). Allerdings besteht aufseiten der Forschenden teilweise noch immer die Befürchtung, dass Wissenschaftler_innen, welche soziale Medien aktiv für ihre Arbeit nutzen, als unprofessionell angesehen werden (Lemke et al., 2019; Osterrieder, 2013). Trotzdem twittern viele Forschende, insbesondere während wissenschaftlicher Konferenzen (Letierce et al., 2010). Die Inhalte der Tweets sind auch persönlicher Natur, wie zum Beispiel Äußerungen zu aktuellen politischen Geschehnissen oder Nachrichten (McQuate, 2017). Auch Haustein et al. (2014) konnten zeigen, dass Astrophysiker_innen größtenteils nicht die gleichen Terme in Tweets nutzen wie in Abstracts zu publizierten Artikeln. Ob dies der anderen antizipierten Nutzerschaft von Twitter geschuldet ist oder der Trennung von wissenschaftlicher Kommunikation und eher informeller Privatkommunikation, ist offen geblieben.

Auf der anderen Seite zeigen Studien (u.a. Hölzig/Hasebrink, 2020), dass sich immer mehr Menschen auf Social-Media-Plattformen über das aktuelle Tagesgeschehen und z.B. Gesundheitsinformationen (Islam et al., 2020) informieren. Die Umfrage von Hölzig und Hasebrink (2020) kam auch zu dem Ergebnis, dass etwas mehr als ein Drittel der Befragten in Sorge sind, dass sie Fakten nicht von Falschinformationen unterscheiden können. Gleichzeitig sind Wissenschaftler_innen, Ärzt_innen und Gesundheitsorganisationen die am zweithäufigsten genutzten Nachrichten- und Informationsquellen über das Coronavirus SARS-CoV-2 nach „Nachrichtenorganisationen“.

Produzent_innen und Rezipient_innen von wissenschaftlichen Social-Media-Informationen haben dementsprechend gleichermaßen ein Interesse daran, eine gewisse „Wissenschaftlichkeit“ zu demonstrieren und als „wissenschaftlich“ wahrgenommen zu werden bzw. wahrhaftige „Wissenschaftlichkeit“ zu erkennen und den Konsum von Falschinformationen zu vermeiden. Doch wie kann diese „Wissenschaftlichkeit“ in Social-Media-Postings nachgewiesen werden – insbesondere im Hinblick darauf, dass nicht alle Rezipient_innen Fachexpert_innen sein können und damit in der Lage sind, die inhaltliche Qualität eines Beitrags zu beurteilen? Im Kontext von Tweets und Twitter haben Weller, Dröge und Puschmann (2011) folgende Signale

von Wissenschaftlichkeit diskutiert: wissenschaftlicher Inhalt oder Link zu einem wissenschaftlichen Beitrag (z.B. DOI), Tweet mit einem wissenschaftlichen Hashtag (z.B. Konferenz-Hashtag) oder wissenschaftliche_r Autor_in als Autor_in des Tweets. Obwohl Studien zu Konferenz-Hashtags populär sind (z.B. Letierce et al., 2010; Lemke/Mazarakis/Peters, 2016; Mazarakis/Peters, 2015), ist das Bestimmen von wissenschaftlichen Hashtags und Autor_innen der Flaschenhals dieser Ansätze. In etablierten Formen wissenschaftlicher Kommunikation, z.B. in Zeitschriftenartikeln, Konferenzpublikationen oder Fachbüchern, signalisieren weitere, leichter erhältliche Metadaten die Wissenschaftlichkeit eines Forschungsoutputs, z.B. Veröffentlichung bei einem bekannten Verlag, Impact Factor der Zeitschrift, durchgeführtes Peer Review. Auch die optische Erscheinung („Layout-Effekt“) des Beitrags trägt dazu bei, dass Texte als „wissenschaftlich“ wahrgenommen werden (Hahn et al., 2020).

Diese explorative Studie geht u.a. der Frage nach, ob sich ein derartiger Layout-Effekt auch für Tweets, d.h. Beiträge auf Twitter, nachweisen lässt. Denn auf Social-Media-Plattformen wie Twitter gibt es (noch) keine Konventionen oder gar Templates, die eine „Wissenschaftlichkeit“ optisch unterstreichen können oder in 280 Zeichen verdeutlichen, dass es sich um einen wissenschaftlichen Tweet handelt. Es fehlen zudem die Metadaten der etablierten wissenschaftlichen Publikationsformen, was die Beurteilung der Tweets schwierig macht. Zwei Fragestellungen wurden im Rahmen der Studie bearbeitet:

1. Wann wird ein Tweet als wissenschaftlich eingestuft?
2. Welche Eigenschaften bei Tweets wirken sich auf die wahrgenommene Wissenschaftlichkeit aus?

Für die Studie wurden typische Eigenschaften von Tweets, die sich auch optisch gut wahrnehmen lassen (z.B. Vorhandensein von Hashtags, Retweets, Gefällt-mir-Angaben; Bruns/Stieglitz, 2013), in einem online durchgeführten Experiment untersucht, um Hinweise darauf zu erhalten, welche Eigenschaften von Tweets einen Einfluss auf ihre wahrgenommene „Wissenschaftlichkeit“ haben. In die Auswertung der Online-Befragung wurden auch die Tweet-Inhalte eingebracht, um den Zusammenhang zwischen Thema, Tweet-Eigenschaften und „Wissenschaftlichkeit“ zu beleuchten. Für die Studie wurden Themenfelder aus der Medizin ausgewählt, da diese sowohl von Forschenden, aber auch von Patient_innen, interessierten Laien, Politiker_innen etc. aufgegriffen werden, sodass ein Datenpool zu erwarten ist, anhand dessen sich konkrete Unterschiede zwischen „Wissenschaftlichkeit“

und „Nicht-Wissenschaftlichkeit“ herausarbeiten lassen. In diesem Beitrag wird eine Proof-of-Concept-Studie präsentiert, welche die Grundlage für weiterführende Arbeiten z.B. im Bereich des Informationsverhaltens oder der Informationskompetenz bilden kann. Wir betrachten unsere Arbeiten zur wahrgenommenen „Wissenschaftlichkeit“ (Hahn et al., 2020) von Social-Media-Postings als Work-in-Progress.

2 Methoden und Daten

Im Folgenden werden die methodischen Ansätze zur Datenerhebung und -analyse sowie die erhobenen Daten beschrieben.

2.1 Datenbasis für die Online-Befragung

Die Datenbasis ist eine Tweet-Sammlung, welche zwischen April 2018 und Oktober 2018 erhoben und im Februar 2019 final bearbeitet wurde. Dafür wurden einschlägige medizinische Fachbezeichnungen in Deutsch und Englisch als Suchworte (d.h. Hashtags) auf Twitter verwendet und die gefundenen Tweets heruntergeladen. Die Hashtags entstammen einer explorativen Ad-hoc-Suchliste, bestehend aus medizinischen Hashtags. In Tabelle 1 sind die Hashtags verzeichnet, welche für die Suche verwendet wurden.

Tab. 1: Für die Suche verwendete Hashtags

Englisch	Deutsch
#ophthalmology	#augenheilkunde
#radiology	#radiologie
#midwifery	#geburtshilfe
#dermatology	#dermatologie
#orthopedics	#orthopädie
#microbiology	#mikrobiologie
#urology	#urologie
#epidemiology	#epidemiologie
#andrology	#andrologie
#oncology	#onkologie

Für die Auswahl der zu analysierenden Tweets wurden auch Antworten aufgenommen, welche auf einen vorherigen Tweet mit dem fachlich einschlägigen Hashtag geantwortet haben, selbst aber keinen Hashtag nachweisen konnten. Hierbei wird dennoch eine thematische Nähe zum Ursprungstweet angenommen.

Insgesamt wurden zum Erhebungszeitpunkt 186 Tweets ausgewählt, von denen final 162 Tweets in die anschließende Online-Befragung eingeflossen sind. Davon sind 22 Tweets in deutscher Sprache und 140 Tweets in englischer Sprache. Zum Erhebungszeitpunkt wurde der Tweet-Korpus in deutscher Sprache für die gesuchten Hashtags aus Tabelle 1 nahezu erschöpfend erfasst, da die Tweets nur in sehr geringer Anzahl auf Twitter vorhanden waren. Daher wurden zusätzlich überdurchschnittlich viele englischsprachige Tweets erhoben. Dies geschah, um ggf. relevante Sprachunterschiede zwischen den Tweets aufzuzeigen. Bei den aus dem Sample ausgeschlossenen Tweets handelt es sich primär um Antworten auf Tweets, die ohne die gleichzeitige Darstellung des Original-Tweets nicht verstanden werden. Auch Tweets, welche fachspezifische Abkürzungen und Fachtermini enthalten, sodass sie für Außenstehende nicht verständlich sind, wurden aus dem Sample entfernt.

Die Tweets wurden manuell beschrieben, um später den Einfluss spezifischer Eigenschaften bei der Einschätzung von „Wissenschaftlichkeit“ statistisch auswerten zu können. Zwei Kodierer haben dazu drei Arbeitsschritte vollzogen (s. Abb. 1). Zunächst wurden die 162 Tweets danach überprüft, ob in ihnen eine Frage gestellt wurde oder ob sie entweder ein Bild, ein Video oder einen Link enthalten haben. Es wurden außerdem die Anzahl der Retweets, der „Gefällt mir“-Angaben, der Hashtags und der @-mentions pro Tweet erfasst, welche gängige Metriken bei der Analyse von Tweets sind (Bruns/Stieglitz, 2013). Im Rahmen der Proof-of-Concept-Studie sollte untersucht werden, inwieweit diese Metriken im Hinblick auf die Bewertung von „Wissenschaftlichkeit“ eine Rolle spielen. Ein Tweet kann gleichzeitig mehrere Eigenschaften aufweisen, zum Beispiel kann er ein Bild und einen Link enthalten. Solche Tweets wurden dann in allen entsprechenden Eigenschaften annotiert, im gerade erwähnten Beispiel also sowohl in der Eigenschaft „Bild“ als auch in der Eigenschaft „Link“. Die Ergebnisse der Beschreibung sind in den Tabellen 2 und 3 dargestellt. Dreiviertel der Tweets mit den gesuchten Hashtags enthielten ein Bild, etwas weniger auch einen Link. Videos kamen praktisch nicht vor. Im Durchschnitt wurden die Tweets knapp zwei-

mal mit einer „Gefällt mir“-Angabe versehen und enthielten circa fünf Hash-tags.

Tab. 2: Vorhandensein bestimmter Eigenschaften
im analysierten Tweet-Sample, $n = 162$

	Ja	Nein
Tweet ist als Frage formuliert?	17 (10%)	145 (90%)
Tweet enthält ein Bild?	122 (75%)	40 (25%)
Tweet enthält ein Video?	4 (2%)	158 (98%)
Tweet enthält einen Link?	113 (70%)	49 (30%)

Tab. 3: Häufigkeit des Vorkommens weiterer Eigenschaften
in dem analysierten Tweet-Sample, $n = 162$

	Mittelw.	St.Abw.	Median	Min.	Max.
Anzahl der Retweets pro Tweet	0.88	2.96	0	0	27
Anzahl der „Gefällt mir“-Angaben pro Tweet	1.96	5.41	0	0	54
Anzahl der Hashtags pro Tweet	5.48	4.54	4	0	28
Anzahl der @-Erwähnungen pro Tweet	0.59	1.06	0	0	7

Um weitere Hinweise darauf zu erhalten, welche Tweet-Typen wissenschaftlich wirken, haben im zweiten Schritt die beiden Kodierer die Tweets im Rahmen einer Inhaltsanalyse verschiedenen Kategorien zugeordnet. Jeder Tweet wurde genau einer von sieben Kategorien zugeordnet (Kategorien und Beschreibungen s. u.). In acht Zweifelsfällen haben sich die Kodierer beraten und gemeinsam eine Zuweisung vorgenommen. Insgesamt wurde eine Übereinstimmung von ca. 95% erreicht (Cohens Kappa bei 0.91), was eine sehr hohe Übereinstimmung widerspiegelt.

Das Kodieren der Tweets resultierte in den folgenden sieben Kategorien:

- **Artikel** ($n = 35$) sind Tweets mit Informationen darüber, ob fachliche Artikel auf einem Fachportal oder in einer Fachzeitschrift veröffentlicht wurden, auf einer Universitätsseite, einem Blog, als Podcast usw., oder ob es sich um eine (wissenschaftliche) Nachricht auf einer Nachrichtenseite handelt. Die Studienergebnisse zeigen, dass Tweets, welche zu Artikeln auf wissenschaftlichen Nachrichtenseiten oder auf Fachportale leiten, eher als „wissenschaftlich“ betrachtet werden als Tweets, die auf Blogs, Podcasts oder Q&A-Seiten (Sonstiges) verlinken;

- **Job** ($n = 20$), bestehend aus Jobangeboten o.Ä.;
- **Meinung** ($n = 8$), was als Behauptung ohne Quellen zu verstehen ist;
- **Methodik** ($n = 18$), als Darstellung wissenschaftlicher Methodik sind Tweets kategorisiert, welche im Rahmen der wissenschaftlichen Abläufe entstanden sind, mit einer klaren Fokussierung auf wissenschaftliche Publikationen und Vorträge;
- **Publicity** ($n = 59$), bezüglich z.B. Buchvorstellungen, die Suche nach Personal, Teamvorstellungen, Werbung für einen Kongress, Kurse oder Organisationen. Die Ergebnisse zeigen, dass Tweets, die über Kongresse berichten, z.B. verstärkt als wissenschaftlich kategorisiert werden, auch wenn die meisten Tweets inhaltlich nur auf Kongresse aufmerksam machen und deshalb der Kategorie „Publicity“ zugeordnet wurden;
- **Spaß** ($n = 5$), wie z.B. sogenannte Funfacts;
- **Verständnis** ($n = 17$), mit einfachen Ergebnisdarstellungen oder allgemeinen Fragen.

Der überwiegende Teil der Tweets in dem Sample beschäftigte sich mit „Publicity“ ($n=59, 36\%$), zu denen auch Tweets zu Konferenzen gehören, gefolgt von Tweets über wissenschaftliche Artikel ($n=35, 22\%$). Die Kategorien sind nicht überlappend, jeder Tweet wurde einer Kategorie eindeutig zugeordnet. Beispiele finden sich in den Abbildungen 3, 4 und 5.

Im dritten Schritt wurden die Tweets durch die Kodierer bezüglich der wahrgenommenen „Wissenschaftlichkeit“ bewertet. Dies geschah, um eine Grundannahme des Experiments – nämlich, dass „wissenschaftliche“ und „nicht-wissenschaftliche“ Tweets von Laien und Expert_innen (d.h. Kodierern) ähnlich wahrgenommen werden – überprüfen zu können. Wie auch im tatsächlichen Experiment wurde keine Definition von „Wissenschaftlichkeit“ vorgegeben und nur die Frage gestellt: „Ist der Tweet aus Ihrer Sicht wissenschaftlich?“. Insgesamt wurde eine Übereinstimmung von ca. 94% zwischen den Kodierern erreicht, Cohens Kappa liegt bei 0.88.

Die Bewertung durch die Kodierer resultierte in folgender Verteilungshäufigkeit der Tweets im Sample: Ca. ein Viertel der Tweets wurden als „wissenschaftlich“ eingestuft und ca. drei Viertel als „nicht-wissenschaftlich“. Damit war die nötige Varianz in der wahrgenommenen „Wissenschaftlichkeit“ von Tweets für die nachfolgenden statistischen Auswertungen gegeben. Abbildung 1 gibt einen Überblick zur Sammlung und Kategorisierung des Tweet-Samples der Online-Befragung.

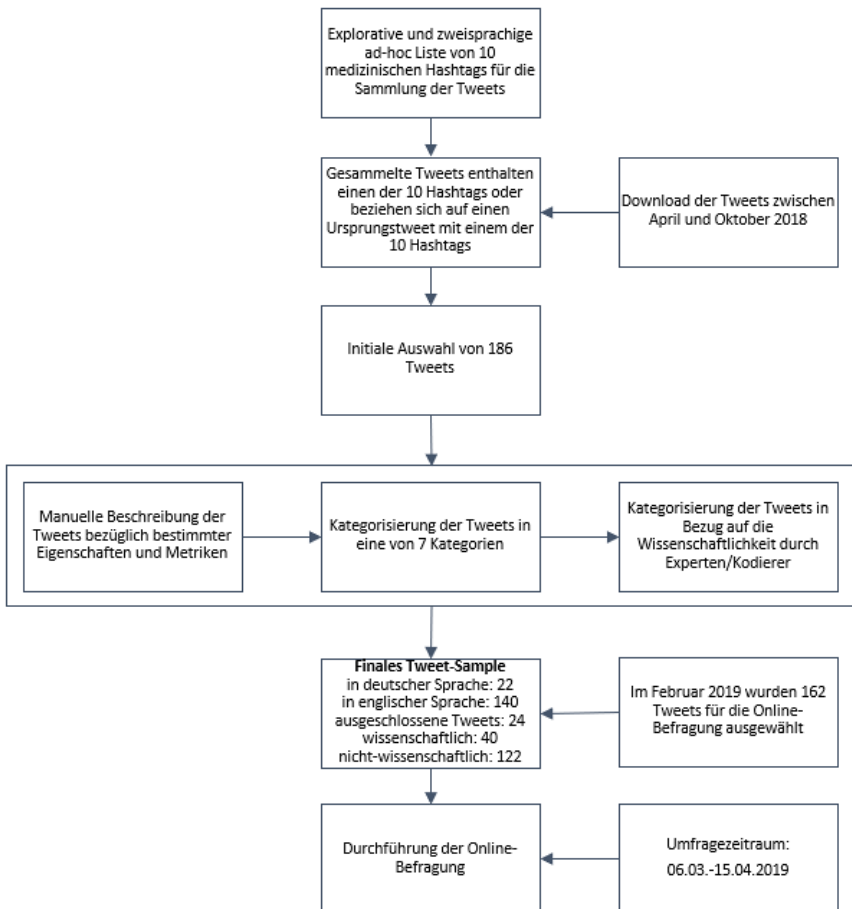


Abb. 1 Überblick zur Sammlung und Kategorisierung des Tweet-Samples der Online-Befragung

2.2 Online-Befragung

Die Online-Befragung wurde zwischen dem 6. März 2019 und dem 15. April 2019 durchgeführt, nachdem Teilnahmeaufrufe auf den Institutsseiten und auf Twitter und Facebook veröffentlicht wurden. Dazu wurde den Proband_innen jeweils eins von acht Sets mit je 20 unterschiedlichen Tweets aus dem Sample mit 162 Tweets angezeigt. Die Sets sollten dann von den Proband_innen bewertet werden. Die Proband_innen wurden gefragt, ob der angezeigte Tweet ihrer Meinung nach „wissenschaftlich“ ist (Antwortmög-

lichkeit: „Ja“ oder „Nein“). Im Anschluss an die Befragung erhielten die Proband_innen die Möglichkeit, ihre Votings noch einmal zu überprüfen. Mehrfachteilnahmen waren möglich, da die vorgegebenen Sets unterschiedliche Tweets beinhalteten und die Sets randomisiert verteilt wurden. Wir haben in unserer Studie keine Hinweise darauf, dass Mehrfachteilnahmen vorgekommen sind. In der Studie wurde mit Sets gearbeitet, damit eine möglichst gleichverteilte Anzeige aller Tweets stattfindet. Bei einer vollständig randomisierten Tweet-Anzeige und unter Beibehaltung der Abfrage von 20 unterschiedlichen Tweets je Proband_in wäre sonst eine mindestens vierstellige Proband_innenanzahl notwendig gewesen, um valide Ergebnisse zu erhalten. Die Proband_innen haben immer 20 Tweets bewertet, um den Zeitaufwand und die Abbruchquote gering zu halten.

Insgesamt haben 109 deutschsprachige Proband_innen an der Studie teilgenommen. Das durchschnittliche Alter der Proband_innen ist 43 Jahre, der Median liegt bei 41 Jahren, die Altersspanne reicht von 25 bis 65 Jahren. Der größte Teil der befragten Personen ist weiblich ($n = 47$, 43%), 32 Personen sind männlich (29%) und 30 (28%) Personen haben keine Angaben zum Geschlecht gemacht.

3 Ergebnisse

Zuerst werden die Ergebnisse zum Zusammenhang von Tweet-Eigenschaften und wahrgenommener „Wissenschaftlichkeit“ präsentiert, bevor auf die Tweets eingegangen wird, die als eindeutig „wissenschaftlich“ und eindeutig „nicht-wissenschaftlich“ kategorisiert wurden.

3.1 Tweet-Eigenschaften und wahrgenommene „Wissenschaftlichkeit“

Die Ergebnisse wurden über Visualisierungen wie Boxplots und deskriptive Tabellen ausgewertet. Außerdem wurden mehrere Spearman-Korrelationen durchgeführt, da die Ergebnisse nicht normalverteilt sind und durch Aggregation auf Ordinalskalenniveau vorliegen, womit die Voraussetzungen für die Durchführung von Spearman-Korrelationen erfüllt sind.

Insgesamt wurden im Durchschnitt die Tweets zu 41% (Akzeptanzwert = 0.41) als „wissenschaftlich“ von den Proband_innen eingestuft (Median = 0.40 und Standardabweichung = 0.29). Dies entspricht dem aggregierten Durchschnitt aller Bewertungen bezüglich der wahrgenommenen „Wissenschaftlichkeit“ je Tweet. Wurde also beispielsweise ein Tweet von 20 Proband_innen bewertet, wobei jedoch nur sechs Proband_innen den Tweet als „wissenschaftlich“ wahrgenommen haben, dann hat dieser Tweet einen Akzeptanzwert von 0.30, bzw. von 30%.

Um die Grundannahmen des Experiments mit den Ergebnissen zu vergleichen, wurde die wahrgenommene „Wissenschaftlichkeit“ der Tweets durch die Kodierer mit der wahrgenommenen „Wissenschaftlichkeit“ der Tweets durch die Proband_innen verglichen. Da die Voraussetzungen für die Verwendung eines parametrischen Verfahrens nicht gegeben sind (nicht vorhandene Normalverteilung), wurde ein Mann-Whitney-U-Test gerechnet. Der nicht-parametrische Mittelwertvergleich zeigt für die Gruppe mit „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 0.69 = 69%) gegenüber der Gruppe mit „nicht-wissenschaftlich“ wahrgenommenen Tweets (Mdn = 0.32 = 32%) einen statistisch signifikanten Unterschied mit $U = 799.00$, $z = -6.38$, $p < 0.001$. Dies bedeutet, dass die vermeintlich „nicht-wissenschaftlichen“ Tweets und die vermeintlich „wissenschaftlichen“ Tweets von den Proband_innen ebenfalls so bewertet wurden wie von den Kodierern. Die Grundannahme bezüglich der ausgewählten Tweets wurde bestätigt, „wissenschaftliche“ und „nicht-wissenschaftliche“ Tweets werden von Laien und Expert_innen (d.h. Kodierern) ähnlich wahrgenommen. Durch dieses Ergebnis sind nun auch die weiteren statistischen Analysen möglich.

Bezogen auf die ursprüngliche Forschungsfrage, nämlich die „wissenschaftliche“ Wahrnehmung von Tweets, kann ein statistisch signifikantes Ergebnis berichtet werden. Die Anzahl der Hashtags korreliert statistisch signifikant negativ mit $r_s = -0.23$, $p = 0.003$ mit der Klassifikation der Proband_innen, dass ein Tweet als „wissenschaftlich“ eingeschätzt wird. Um dieses Ergebnis nochmals zu überprüfen, wurde ein Mann-Whitney-U-Test gerechnet. Hierzu wurden alle Tweets, welche häufiger als 50% als „wissenschaftlich“ bewertet wurden, in eine Gruppe kategorisiert, Tweets mit weniger als 51%-Bewertung in eine zweite Gruppe hinzugefügt. In Bezug auf Hashtags zeigt der nicht-parametrische Mittelwertvergleich für die Gruppe mit mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 3.00) gegenüber der Gruppe mit nicht-mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 5.00) ebenfalls einen statistisch signifikanten Un-

terschied mit $U = 2070.50$, $z = -3.34$, $p = 0.001$. Je mehr Hashtags vorhanden sind, desto weniger „wissenschaftlich“ wirkt der Tweet. Andere Korrelationen mit Tweet-Eigenschaften – wie z.B., ob ein Tweet als Frage formuliert ist, die Anzahl der @-Erwähnungen oder ob ein Video vorhanden ist – führten zu Null-Korrelationen.

Da sich in der anfänglichen manuellen Kategorisierung der Tweets schon gezeigt hatte, dass Tweets häufig nicht nur einer Eigenschaft wie z.B. „enthält eine Abbildung“ zugeordnet werden konnten, wurde auch der Zusammenhang zwischen den verschiedenen Eigenschaften statistisch (über Spearman-Korrelationen) überprüft. Die Anzahl der Retweets korreliert mit einem statistisch signifikanten Wert von $r_s = 0.70$, $p < 0.001$ positiv mit der Anzahl der Favorisierungen („Gefällt mir“-Funktion bzw. Herz-Funktion). Je häufiger ein Tweet retweetet wird, desto mehr Gefallen findet er in der Community. Das Vorhandensein eines Links im Tweet korreliert mit $r_s = -0.21$, $p = 0.007$ negativ mit der Anzahl der Favorisierungen („Gefällt mir“-Funktion“ bzw. Herz-Funktion) und negativ mit der Anzahl an Retweets ($r_s = -0.21$, $p = 0.008$). Links sorgen also eher dafür, dass ein Tweet nicht favorisiert wird bzw. nicht retweetet wird. Das Alphaniveau wurde konservativ mit der Bonferroni-Methode korrigiert und beträgt für alle berichteten Ergebnisse 0.006. Derartige Auswertungen lassen Hinweise darauf zu, ob sich Tweet-Eigenschaften gegenseitig verstärken und damit als Bündel Einfluss auf die wahrgenommene „Wissenschaftlichkeit“ nehmen (was in zukünftigen Studien überprüft werden soll).

Die Ergebnisse der statistischen Auswertung zu den sieben Tweet-Kategorien sind als Boxplots in Abbildung 2 dargestellt. Es ist zu erkennen, dass Tweets, welche in der Kategorie „Methodik“ ($M = 0.76$, $SD = 0.19$ und ein Median von 0.80) enthalten sind, am ehesten als „wissenschaftlich“ anerkannt werden (hier benannt als „Akzeptanz“). Tweets in den Kategorien „Artikel“ ($M = 0.53$, $SD = 0.25$ und ein Median von 0.57) und „Verständnis“ ($M = 0.54$, $SD = 0.19$ und ein Median von 0.50) werden im Durchschnitt mindestens zur Hälfte als „wissenschaftlich“ bewertet (Akzeptanzwerte ab 50%). Solche Tweets, welche den Kategorien „Spaß“ ($M = 0.07$, $SD = 0.08$ und ein Median von 0.07), „Meinung“ ($M = 0.35$, $SD = 0.23$ und ein Median von 0.43), „Publicity“ ($M = 0.32$, $SD = 0.26$ und ein Median von 0.30) und „Job“ ($M = 0.15$, $SD = 0.16$ und ein Median von 0.10) zugeordnet wurden, werden mehrheitlich als „nicht-wissenschaftlich“ angesehen. Eine starke Streuung wie z.B. in den Kategorien „Artikel“ oder „Publicity“ deutet auf eine grundlegende Ambivalenz bei der Beurteilung hin.

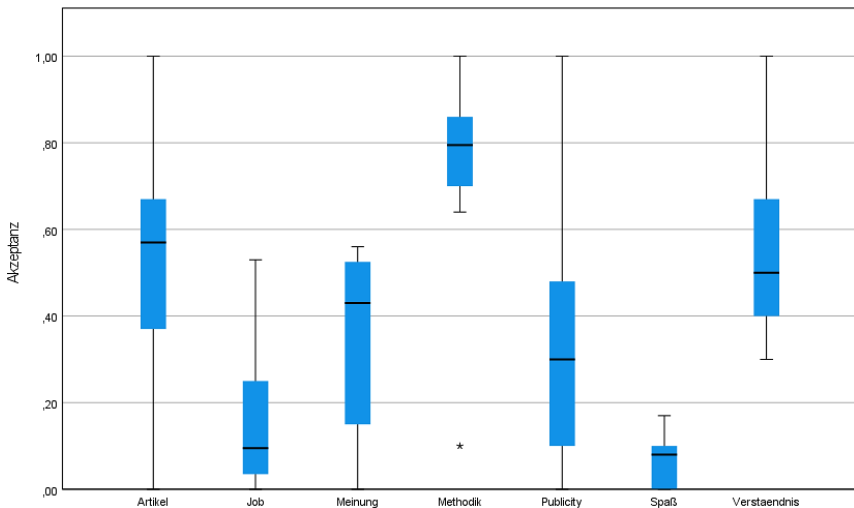


Abb. 2 Boxplots-Analyse der Tweets im Hinblick auf die Bewertung der „Wissenschaftlichkeit“ in Zusammenhang mit der Kategorisierung; x-Achse: Kategorien; y-Achse: Proband_inneneinschätzung: „wissenschaftlich“ ja/nein; * = Ausreißer

3.2 Vergleich von Tweets, die als eindeutig „wissenschaftlich“ und eindeutig „nicht-wissenschaftlich“ kategorisiert wurden

Im Folgenden werden zur Veranschaulichung bestimmte Tweets aus der Online-Befragung dargestellt, welche entweder sehr häufig (Abb. 3) oder nie (Abb. 4 u. 5) als „wissenschaftlich“ wahrgenommen wurden. Es werden nur Tweets dargestellt, die mindestens von zehn Proband_innen bewertet wurden.

Im direkten Vergleich beider Gruppen von Tweets fällt auf, dass Tweets, welche als „wissenschaftlich“ betrachtet werden, tatsächlich Aspekte aus der wissenschaftlichen Tätigkeit beschreiben. Das bedeutet, entweder wird auf eine wissenschaftliche Publikation, einen wissenschaftlichen Bericht oder medizinische und lebenswissenschaftliche Kongresse hingewiesen oder verlinkt.

Im Gegensatz dazu sind Tweets, welche als „nicht wissenschaftlich“ betrachtet werden, thematisch eher heterogen. Zwar wird der Themenbereich „Wissenschaft“ im weitesten Sinne angesprochen, z.B. in Form von wissenschaftlichen Jobangeboten sowie der Nennung oder Vorstellung von Institutionen oder Personal, jedoch fehlt in diesen Tweets der direkte Bezug zu wissenschaftlichen Ergebnissen bzw. zu den Aspekten aus der wissenschaftlichen Tätigkeit. Sobald aber dieser Bezug vorhanden ist, steigt die „Wissenschaftlichkeit“ der Tweets an. So werden Tweets, die auf Kongresse auf-

merksam machen, eher als „wissenschaftlich“ bewertet, da auf Kongressen wissenschaftliche Ergebnisse präsentiert werden.

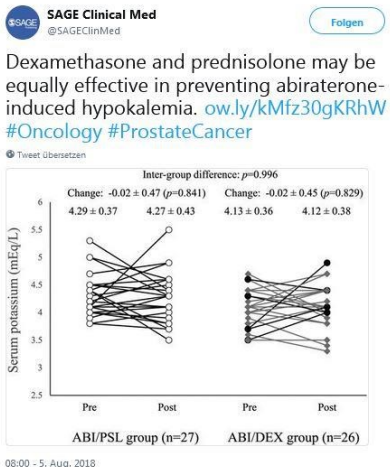
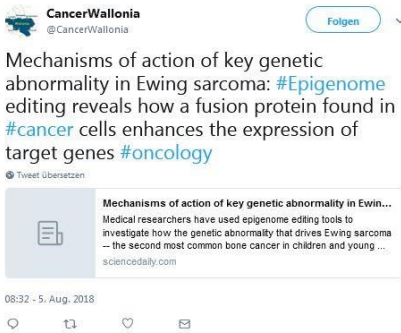

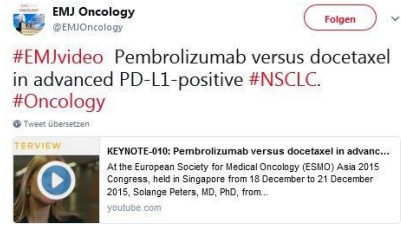
 <p>08:00 - 5. Aug. 2018</p>	 <p>08:32 - 5. Aug. 2018</p>
<p>100% Akzeptanz der ProbandInnen https://twitter.com/SAGEClinMed/status/1026120800175370241</p>	<p>93% Akzeptanz der ProbandInnen https://twitter.com/CancerWallonia/status/1026128705230385152</p>
<p>Dieser Tweet wurde von allen befragten Personen als „wissenschaftlich“ bewertet. Der Link im Tweet führt zu einem wissenschaftlichen Artikel. Der Text im Tweet gibt ein Resultat der Publikation wieder.</p>	<p>Der Text gibt ein fachliches Resultat wieder, der Link verlinkt auf einen Artikel in Sciedaily.com, eine wissenschaftliche Nachrichtenseite.</p>
 <p>00:47 - 5. Aug. 2018</p>	 <p>07:20 - 5. Aug. 2018</p>
<p>90% Akzeptanz der ProbandInnen https://twitter.com/IllingworthInfo/status/1026011901761925120</p>	<p>90% Akzeptanz der ProbandInnen https://twitter.com/EMJOnco/status/1036257481897455623</p>
<p>Im Tweet selbst steht ein Auszug des Titels einer Zeitschriftenpublikation. Der Link leitet zur Publikation weiter. Im Link-Erklärungstext ist ein Auszug aus dem Abstract zu lesen.</p>	<p>Der Tweet-Text ist ein Auszug des Titels einer Keynote. Der Link verweist auf ein Video, welches eine Vorschau auf die Keynote gibt.</p>

Abb. 3 Beispielhafte Tweets, die eindeutig als „wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: oben links: Methode; oben rechts: Artikel; beide unten: Methode)



<p>Lokmanya Hospitals @LIVHospitals Folgen</p> <p>Hydration during winter is especially essential as it fights inflammation and keeps your joints working fit & fine. If you too are looking for a comprehensive joint treatment, visit #LokmanyaHospitals: goo.gl/a4U8iu #Health #Healthy #Pune #Orthopedics #NoPainThisWinter</p>  <p>21:30 - 25. Okt. 2018</p> <p>1 „Gefällt mir“-Angabe</p>	<p>Q Laboratories, Inc. @QLaboratories Folgen</p> <p>@QLaboratories was honored to be named a Finalist for the Deal Maker Award by @ACGCincinnati @ACGGlobal bit.ly/2kiYxC #laboratory #microbiology #chemistry #analytical</p>  <p>08:13 - 23. Mai 2018</p> <p>1 „Gefällt mir“-Angabe</p>
<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/LMhospitals/status/1055677982961004547</p> <p>Der Text beschreibt den Zusammenhang zwischen Hydration und Entzündungen, um dann zur eigentlichen Werbung überzuleiten. Folgt man dem Link, kommt man auf die Internetseite der Institution, welche auch im Bild anhand des Logos erkennbar ist, bzw. welche diesen Tweet gepostet hat.</p>	<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/QLaboratories/status/999307365588619264</p> <p>Der Tweet weist auf eine Auszeichnung hin, welche die postende Institution bekommen hat. Im Bild ist der Award zu sehen. Der Link leitet auf die offizielle Pressemitteilung der Institution anlässlich einer Preisverleihung.</p>

Abb. 4 Beispielhafte Tweets, die eindeutig als „nicht wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: beide: Publicity)

<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/radiology_guide (Account ist derzeit gesperrt)</p>	<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/bonniejaddario/status/1026142076071759872</p>
<p>Der Tweet ist eine Jobanzeige. Hinter dem Link verbirgt sich eine Jobplattform.</p>	<p>Der Tweet ist ein Spendenaufruf und erklärt, wie dieser funktioniert, mit einem Bild, auf der die postende Person zu sehen ist (letzteres lässt sich nur durch Interpretation des Tweets erschließen).</p>
<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/INTREPYLLC/status/1027234266709745666</p>	<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/JenaHall/status/1055656894960975872</p>
<p>Der Link des Tweets leitet auf einen Artikel auf der Webseite der zu postenden Institution weiter. Ein Aufforderungssatz, gefolgt von Schlagwörtern in Form von Hashtags, leitet das Thema ein. Ein Bild, welches ein Zitat darstellen soll, beinhaltet einen Satz, der so nicht im Artikel hinter dem Link zu finden ist.</p>	<p>Der Tweet weist auf ein Team hin, welches auf dem Bild abgebildet ist.</p>

Abb. 5 Beispielhafte Tweets, die eindeutig als „nicht wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: oben links: Job; oben rechts: Publicity; beide unten: Publicity)

4 Diskussion der Ergebnisse

Die Antworten der Online-Befragung geben im Zusammenhang mit den Tweet-Eigenschaften Aufschluss darüber, welche Tweets als „wissenschaftlich“ wahrgenommen werden und welche nicht. In der Detailanalyse der Tweets, die seitens der Proband_innen weitestgehend eindeutig als „wissenschaftlich“ und „nicht-wissenschaftlich“ bewertet wurden, weisen diejenigen Tweets, die als „wissenschaftlich“ (primär aus der Kategorie „Methodik“) bewertet werden, eine stärkere Fokussierung auf Aspekte aus der wissenschaftlichen Tätigkeit auf und beinhalten Informationen zu Publikationen, Abbildungen aus Publikationen oder von Kongressen.

Als „nicht-wissenschaftlich“ eingestufte Tweets (primär aus den Kategorien „Job“ und „Spaß“) haben eine größere thematische Breite. Bei allen übrigen Kategorien gibt es eine große Streuung der Bewertungen. Eine mögliche Erklärung für dieses Ergebnis könnte sein, dass die Kategorien nicht ausreichend trennscharf sind. Dagegen spricht allerdings das quantitative Ergebnis, da keine bemerkenswerten Unterschiede in der Standardabweichung existieren. Es ist eher anzunehmen, dass Tweets aus den anderen Kategorien („Verständnis“, „Publicity“, „Meinung“ und „Artikel“) sehr ambivalent bewertet werden und hier eine sehr hohe Subjektivität vorhanden ist. Hierbei ist zu beachten, dass Tweets, welche fachspezifische Abkürzungen und Fachtermini enthielten, sodass sie für Laien nicht verständlich sind, aus dem Sample entfernt wurden. Es besteht aber die Möglichkeit, dass diese Tweets eventuell als „wissenschaftlich“ wahrgenommen werden, gerade weil man die Fachtermini nicht versteht (Weisberg et al., 2007).

Die durchgeführte Korrelationsanalyse und der nicht-parametrische Mittelwertvergleich zeigen zudem, dass Tweets mit vielen Hashtags als „nicht-wissenschaftlich“ wahrgenommen werden. Schaut man sich die beispielhafte Gegenüberstellung der Tweets in den Abbildungen 3, 4 und 5 an, dann zeigt sich, dass als „nicht-wissenschaftlich“ eingestufte Tweets häufig mehr als drei Hashtags verwenden, während als eindeutig „wissenschaftlich“ kategorisierte Tweets maximal drei Hashtags oder weniger verwenden.

Eine wichtige Einschränkung der Studie liegt in der Ausgestaltung der Online-Befragung begründet. In der Online-Befragung hatten die Proband_innen nur Zugriff auf ein Bild (Screenshot) des Tweets und mussten den Tweet ohne die Prüfung von z.B. weiterführenden Links im Hinblick auf seine „Wissenschaftlichkeit“ bewerten. Das Linkziel ist jedoch ein wichtiger

Bestandteil eines Tweets und wurde dementsprechend auch in die manuelle Klassifikation durch die Kodierer einbezogen. Für die Einschätzung eines Tweets spielt möglicherweise der potenzielle Zugriff auf die Ressource oder die Linkbezeichnung, auf die verwiesen wird, ebenfalls eine Rolle. Durch die fehlende Möglichkeit, wie auf Twitter, den Link anzuklicken, war es Proband_innen nicht direkt möglich, eine eventuell vorhandene subjektive Einschätzung bezüglich des Links zu überprüfen. Dies könnte sich auf die Ergebnisse ausgewirkt haben, sodass zukünftige Studien idealerweise diese Funktionalität gewährleisten sollten.

Eine weitere Einschränkung dieser Studie ist auch die relativ geringe Anzahl an Bewertungen je Tweet. Dies könnte teilweise die fehlende Eindeutigkeit bzw. die hohe Varianz bestimmter Ergebnisse erklären. Die quantitative Unausgeglichenheit zwischen den sieben Tweet-Kategorien in dieser Studie lässt sich mit der Wahl der Hashtags für die Suche nach Tweets erklären. Es ist zu vermuten, dass die ausgewählten Hashtags eher dazu genutzt werden, bereits sehr „wissenschaftliche“ Themen und Tweets zu beschreiben.

Während der größte Teil der präsentierten Tweets in englischer Sprache vorliegt (86%), die Proband_innen hingegen deutschsprachig sind und auch der Fragebogen in deutscher Sprache verfasst wurde, ist es denkbar, dass an dieser Stelle ein Bias vorliegen könnte. Zwar konnten wir in unserer Analyse keine ungewöhnlichen Ergebnisse bezüglich der Sprache entdecken, trotzdem ist es möglich, dass eine Studie mit einer einsprachigen Ausrichtung der Tweets, bzw. mit rein englischsprachigen Proband_innen, zu anderen Ergebnissen gekommen wäre.

Schließlich wäre noch zu bedenken, ob nicht schon allein die tweetende Person als ausschlaggebendster Faktor für Wissenschaftlichkeit wahrgenommen wird, also ein sogenannter „Guru-Effekt“ auftritt (Sperber, 2010), und diesbezüglich die Tweets auch entsprechend identifiziert werden. Dies würde allerdings ein anderes Forschungsdesign benötigen, um diese interessante Fragestellung experimentell zu untersuchen, z. B. die Verwendung von Tweets von bekannten und weniger bekannten Wissenschaftler_innen.

5 Zusammenfassung und Ausblick

Die vorgestellte Studie hatte sich zum Ziel gesetzt, die „Wissenschaftlichkeit“ von Tweets zu medizinischen Themen in Abhängigkeit von verschiede-

nen Tweet-Eigenschaften (z.B. Anzahl Retweets, Tweet-Typ) über eine Online-Befragung zu untersuchen. Die Ergebnisse beruhen auf einer Analyse von 162 klassifizierten Tweets und ihrer Bewertung durch 109 Proband_innen.

Die Ergebnisse lassen sich wie folgt zusammenfassen: Tweets mit vielen Hashtags werden eher als „nicht-wissenschaftlich“ wahrgenommen, wobei hier maximal drei Hashtags noch positiv sind, eine höhere Anzahl hingegen korreliert negativ mit der wahrgenommenen „Wissenschaftlichkeit“. Diese Erkenntnis beruht einerseits auf der Analyse der Abbildungen 3, 4 und 5, andererseits liegt der Median der Anzahl der Hashtags für die Gruppe mit mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets ebenfalls bei drei Hashtags. Da allerdings der Median der Anzahl der Hashtags für die Gruppe mit mehrheitlich „nicht-wissenschaftlich“ wahrgenommenen Tweets bei fünf Hashtags liegt, könnten auch vier Hashtags noch akzeptabel sein. Hierbei ist auch der zu Beginn angemerkte Layout-Effekt zu erwähnen, da die Anzahl der Hashtags durchaus einen starken Einfluss auf die optische Erscheinung der Tweets hat.

Eindeutig zugeordnet wurden von Proband_innen offenbar solche Tweets, die sich mit den Kategorien „Methodik“, „Job“ und „Spaß“ beschäftigen, wobei nur zur ersten Kategorie zugehörige Tweets überwiegend als „wissenschaftlich“ bezeichnet, die beiden anderen Kategorien hingegen als eindeutig „nicht-wissenschaftlich“ kategorisiert wurden. Tweets, die sich den Kategorien „Artikel“ und „Verständnis“ zuordnen lassen, liefern kein eindeutiges Ergebnis, enthalten aber ebenfalls mehrheitlich Tweets, die eindeutig als „wissenschaftlich“ bewertet wurden. Dies ist insbesondere dann der Fall, wenn die Tweets auch auf Aspekte aus der wissenschaftlichen Tätigkeit verweisen oder diese zum Gegenstand haben, wie zum Beispiel Tweets im Kontext von einem Kongress mit einer Fokussierung auf wissenschaftliche Publikationen und Vorträge.

Zukünftig sind weitere Studien geplant. Hierbei bieten sich mindestens zwei Anpassungen an. Einerseits können die medizinischen Fachbezeichnungen ersetzt oder ergänzt werden um Hashtags zu Krankheiten (z.B. Diabetes). Fachbezeichnungen wurden in dieser Studie deswegen verwendet, damit die thematische Breite der Hashtags und Tweets nicht zu groß wird. Weitere Untersuchungen sind notwendig, um die Validität dieses Ansatzes zu überprüfen. Andererseits kann auch ein nicht-medizinisches Setting in Betracht gezogen werden. Durch ein solches Setting besteht die Möglichkeit, die Ergebnisse dieser Studie zu generalisieren.

Literaturverzeichnis

- Bruns, A.; Stieglitz, S. (2013): Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology: Theory and Practice* 16 (2), 91–108.
- Haustein, S.; Bowman, T. D.; Holmberg, K.; Peters, I.; Larivière, V. (2014): Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. *Aslib Journal of Information Management* 66 (3), 279–296.
- Hölig, S.; Hasebrink, U. (2020): Reuters Institute Digital News Report 2020 – Ergebnisse für Deutschland. Arbeitspapiere des Hans-Bredow-Instituts | Projektergebnisse Nr. 50. Hamburg: Verlag Hans-Bredow-Institut. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/66q2yde_AP50_RIDNR20_Deutschland.pdf
- Islam, M. S.; Sarkar, T.; Khan, S. H.; Mostofa Kamal, A.-H.; Hasan, S. M. M.; Kabir, A. ... Seale, H. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *The American Journal of Tropical Medicine and Hygiene*. <http://www.ajtmh.org/content/journals/10.4269/ajtmh.20-0812>
- Lemke, S.; Mazarakis, A.; Peters, I. (2016): Characteristics of Twitter usage at scientific conferences. In: *Proceedings of the 18th General Online Research Conference* (pp. 34–35). Dresden: Deutsche Gesellschaft für Online-Forschung (DGOF).
- Lemke, S.; Mehrazar, M.; Mazarakis, A.; Peters, I. (2019): “When You Use Social Media You Are Not Working”: Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics* 3, 39. <https://doi.org/10.3389/frma.2018.00039>
- Letierce, J.; Passant, A.; Breslin, J.; Decker, S. (2010): Understanding how Twitter is used to spread scientific messages. In: *Proceedings of Web Science Conf. 2010*, April 26–27, 2010, Raleigh, NC, USA.
- Mazarakis, A.; Peters, I. (2015): Science 2.0 and Conference Tweets: What? Where? Why? When? *Electronic Journal of Knowledge Management* 13 (4), 269–282. <http://www.ejkm.com/issue/download.html?idArticle=730>
- McKendrick, D. R. A.; Cumming, G. P.; Lee, A. J. (2012): Increased use of Twitter at a medical conference: A report and a review of the educational opportunities. *Journal of Medical Internet Research* 14 (6), e176. <https://doi.org/10.2196/jmir.2144>
- McQuate, S. (2017): What all those scientists on Twitter are really doing. *Nature* 12, e0175368. <https://doi.org/10.1038/nature.2017.21873>

- Osterrieder, A. (2013): The value and use of social media as communication tool in the plant sciences. *Plant Methods* 9 (1), 26. <https://doi.org/10.1186/1746-4811-9-26>
- Salem, J.; Borgmann, H.; Murphy, D. G. (2016): Integrating Social Media into Urologic Health care: What Can We Learn from Other Disciplines? *Current Urology Reports* 17 (2), 13. <https://doi.org/10.1007/s11934-015-0570-2>
- Sperber, D. (2010): The Guru Effect. *Review of Philosophy and Psychology* 1 (4), 583–592. <https://doi.org/10.1007/s13164-010-0025-0>
- Sugimoto, C. R.; Work, S.; Larivière, V.; Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* 68 (9), 2037–2062.
- Weisberg, D. S.; Keil, F. C.; Goodstein, J.; Rawson, E.; Gray, J. R. (2007): The Seductive Allure of Neuroscience Explanations. *Journal of Cognitive Neuroscience* 20 (3), 470–477.
- Weller, K.; Dröge, E.; Puschmann, C. (2011): Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. In M. Rowe, M. Stankovic, A.-S. Dadzie, M. Hardey (Hrsg.), *Making Sense of Microposts (#MSM2011), Workshop at Extended Semantic Web Conference (ESWC 2011), Crete, Greece*. Heidelberg: Springer. http://ceur-ws.org/Vol-718/paper_04.pdf

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 44–63. DOI: doi.org/10.5283/epub.44936.

The Influence of Multilingualism and Mutual Intelligibility on Wikipedia Reading Behaviour

A Research Proposal

Florian Meier

Science, Policy and Information Studies
Department of Communication and Psychology
Aalborg University, Copenhagen, Denmark
fmeier@hum.aau.dk

Abstract

Given the important role of Wikipedia in our everyday lives, a better understanding of how language skills affect Wikipedia usage is needed. If content is not available in a reader's native language or a language that she can readily understand, access barriers and knowledge gaps are created, threatening Wikimedia's goal to create knowledge equity among all its projects and their consumers. This article argues for research on the effects of multilingualism and mutual intelligibility on Wikipedia reading behaviour, focusing on the Nordic countries, Denmark, Norway, and Sweden. Initial exploratory analysis shows that while residents of these countries use the native language editions quite frequently, they rely strongly on English Wikipedia, too. Research questions and methods for future work in this area are presented.

Keywords: multilingualism; mutual intelligibility; information behaviour; Wikipedia

1 Introduction and state of knowledge

With an average of around 1.4 billion page views per month,¹ Wikipedia is the most popular source of encyclopedic knowledge on the web. Although

1 <https://stats.wikimedia.org/#/all-projects>

Wikipedia is used in many countries and native language editions play an important role for local populations (Johnson, 2019), a considerable proportion of Wikipedia studies puts a strong focus on English Wikipedia. This is criticized by researchers, as recent studies found that English Wikipedia is an outlier regarding several aspects and not representative of the many language editions (Lemmerich et al., 2019). For example, why and how users read Wikipedia (i.e., their motivations, information needs prior knowledge and resulting usage behaviour), differs significantly between countries and language editions (Singer et al., 2017; Lemmerich et al., 2019). The anglophonic bias is also mirrored in the many ways and strategies that have been developed and used to improve and extend Wikipedia and other Wikimedia sister projects. However, this bias possesses a threat to one of the 2030 Wikimedia Foundations strategic directions² which aims to operationalize and quantify knowledge equity in Wikimedia projects thereby building the basis for bridging knowledge gaps, giving people equal access to unbiased knowledge around the globe (Zia et al., 2019; Redi et al., 2020).

Knowledge gaps are defined as “disparities in participation or coverage of a specific group of readers, contributors, or content” (Redi et al., 2020, p. 4). One of the most well-studied knowledge gaps is the gender gap, which evolved over the years because editors are mostly white and male (Lam et al., 2011). This resulted in gender-skewed content with, e.g., less than 20% of biographies being about women (Wagner et al., 2016; Konieczny & Klein, 2018). A recent study documented the gender gap even among Wikipedia readers, finding that two-thirds of all readers are men (Johnson et al., 2020).

A one-size-fits-all approach developed solely on the basis of research on English Wikipedia is likely not enough to bridge knowledge gaps like the aforementioned gender gap. Complete knowledge equity requires a deeper understanding of the differences between Wikipedia language editions and what role language plays in Wikipedia reading behaviour in general. Unlike other user-generated content platforms, where the effects of multilingualism on usage behaviour have been studied before (Eleta & Goldbeck, 2014; Lee & Chau, 2018), little is known about the dynamics that are at play when Wikipedia consumers speak or are able to readily understand multiple languages. Multilingual editing behavior has been studied before (Hale, 2014; Kim et al., 2016). Hale (2014), for example, studied multilingual Wikipedia editors and found that they play a key role in information diffusion. Editors editing the

2 <https://meta.wikimedia.org/wiki/Research:2030>

same page in multiple languages leads to the reduction of self-focus bias by 25%.

To the best of our knowledge, multilingualism among readers hasn't been studied yet. Anecdotal evidence from a recent study combining survey and log data in 14 countries shows that up to 95% of participants reported being monolingual and relying on native language editions (Johnson, 2019). Exceptions are the English and French Wikipedia, which are essential for non-native speakers in Africa (*ibid.*). Moreover, this data showed that although the survey was administered in only 14 languages, participants viewed articles in over 100 languages during their reading sessions (Johnson et al., 2020). Finally, preliminary analysis of the same survey data “shows that on average roughly 20% of reader sessions involve the reader switching from one Wikipedia language to another” (Lemmerich et al., 2019, p. 625).

To sum up, while there seems to be evidence for within-session language switching, this phenomenon hasn't been studied in detail before. Additionally, there is no study that reports on across-session language switching, i.e., users alternating between various language editions depending on context factors like the type of information need or knowledge coverage. Previous research shows that Wikipedia and search engines form a tightly-knit relationship, with Wikipedia pages being highly ranked on Search Engine Result Pages (SERP) for many common queries. This implies that the way people search and which languages they use etc. also has a significant impact on how they read Wikipedia (Vincent et al., 2019).

This short article argues for the need of research on how multilingualism, i.e., speaking multiple languages, and mutual intelligibility, defined as “the fact that some language pairs are so closely related that the speakers are able to communicate each using their own language without prior language instruction” (Gooskens et al., 2017, p. 170) on Wikipedia usage behaviour. As mutual intelligibility is a phenomenon that is very pronounced in the Nordic countries, i.e., Denmark, Norway and Sweden (Gooskens et al., 2017), initial analysis as well as ideas for future work will focus on these countries and Wikipedia language editions.

2 Initial analysis: Wikipedia usage in the nordic countries

The Nordic countries' Wikipedia editions are good examples of diverse types of Wikipedia projects. While Danish and Norwegian Wikipedia are what one could call medium-sized with around 260 K and 535 K articles, Swedish Wikipedia features around 3.7 M articles and is among the biggest encyclopedic projects.³ The reason for its size is that most articles are created by a bot called Lsjbot (Guldbrandsson, 2013). To study mutual intelligibility, we, in an initial exploratory analysis, looked at these language projects' page views and their origin meaning which countries the page views come from. We collected data via the Wikimedia REST API⁴, which allows retrieving access-data for the last five years.

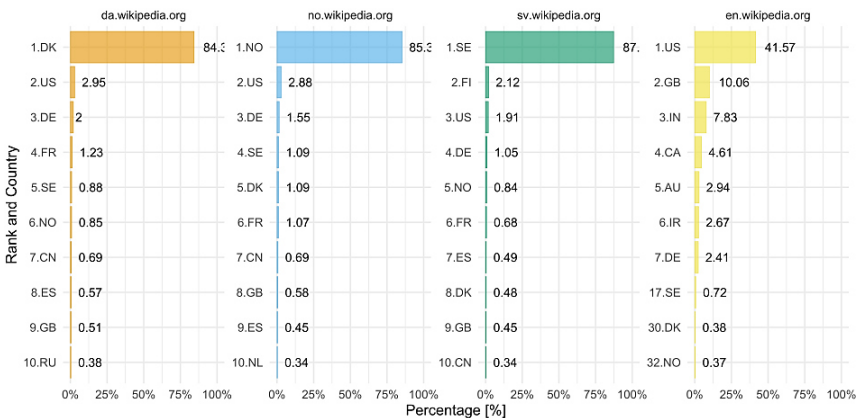


Fig. 1
Percentage of average page views per year coming from different countries for the three Nordic countries and English Wikipedia language editions (Data: 2016–2020)

Figure 1 visualizes the average percentage of page views for the Nordic countries' language projects and English Wikipedia. One can clearly see that most page views (84% or higher) come from within the country. Although, e.g., a Dane could easily read and understand a Norwegian or Swedish

3 https://wdo.wmcloud.org/cultural_context_content/

4 https://wikimedia.org/api/rest_v1/

Wikipedia article, signs that users are taking advantage of mutual intelligibility are very sparse. Only around 2% of access to the Danish Wikipedia comes from Sweden (SE: 0.88%) and Norway (NO: 0.85%). A similar ratio can be observed for the Norwegian and Swedish Wikipedia. A question that arises here is whether this traffic can be explained by, e.g., expatriates or tourists in these countries or whether there are other reasons that give an explanation of the, admittedly low, share of page views across countries and language editions. It is not surprising that English Wikipedia page views come from many different countries whereby most originate from the US, Great Britain and India. The statistic of most page views on en.wikipedia.org lists Sweden, Denmark, and Norway on rank 17, 30, and 32. Next, we tried to estimate the importance of English Wikipedia for the three Nordic countries. Earlier studies suggest that users mostly rely on the native language edition if one exists (Johnson, 2019), but what role does English Wikipedia play for languages/countries with quite large language editions? We estimated the importance of English Wikipedia by dividing the number of page views on en.wikipedia.org by the number of page views on the own language edition plus the number of page views on en.wikipedia.org. Figure 2 visualizes this share in a time-series analysis from January 2016 to October 2020.

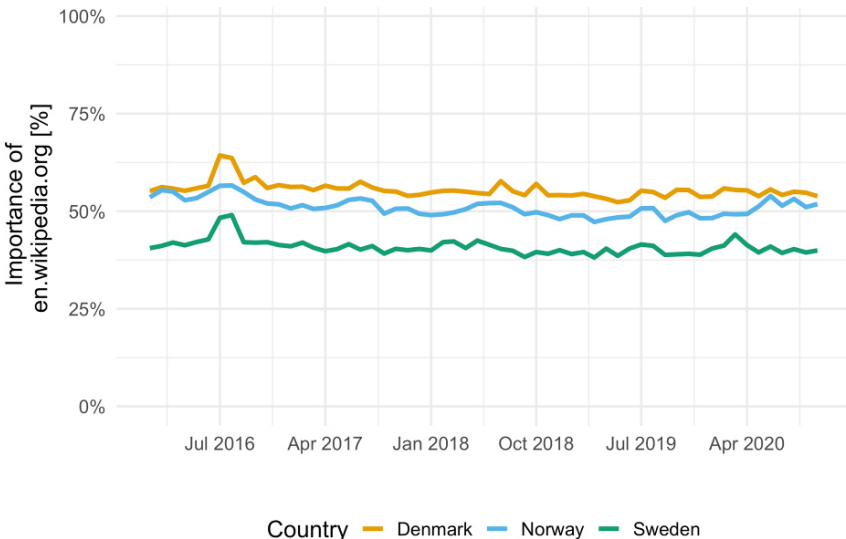


Fig. 2 Time series visualization of the importance of en.wikipedia.org for the three Nordic countries measured in percentage of page views (Data: 2016–2020)

The overall pattern is quite revealing. While the share of English Wikipedia accesses in Sweden lies constantly below 50% which means that the Swedes use sv.wikipedia.org more frequently than English Wikipedia, the Danes rely more strongly on the English Wikipedia. They use it more than their local language edition as is indicated by the line continuously lying above 50%. A possible explanation for this observation could be the smaller size of dk.wikipedia.org, and thus missing content that motivates users to look for information in other language projects. Observations like these are important as they help with prioritizing strategies for closing knowledge gaps.

3 Future work

Research on how language proficiency in multiple languages affects Wikipedia reading behaviour is anecdotal and hasn't been studied in depth. In the future, we want to investigate this phenomenon in detail, specifically focusing on the Nordic countries. The following questions will guide our research:

- How frequently do Wikipedia users make use of their language skills and switch between language editions?
- How proficient does a user need to be in a language so she would consider switching to it?
- What is the extent of within and across-session language switching?
- Which factors motivate this behaviour, i.e., what role do different types of information needs, prior knowledge or sociodemographic aspects play?
- How do the characteristics of Wikipedia language editions, i.e., size, quality of articles, the share of cultural context content, influence user choice (Miquel-Ribé & Laniado, 2020)?

To find answers to these questions, we will follow a mixed-methods approach combining a large-scale quantitative survey with a data science approach. While the survey will gain an understanding of the extent and characteristics of within and across-session language switching when reading Wikipedia, the data science approach will use the Wikipedia Cultural Diversity dataset (Miquel-Ribé & Laniado, 2019) and the Wikipedia Diversity Observatory⁵ to investigate the importance of local content for Wikipedia readers

5 <https://wdo.wmcloud.org/>

in the Nordic countries to help develop strategies for closing language proficiency-related knowledge gaps.

4 Summary and conclusion

This short paper presented a research proposal to bridge knowledge gaps in the Wikipedia, specifically to expand on the availability of qualitative and quantitative evidence on how the nature, scope and impact of reader's language proficiency affect Wikipedia usage. Which and how many languages Wikipedia consumers can understand impacts considerably what knowledge is available to them. However, hardly anything is known about how multilingualism and mutual intelligibility affect Wikipedia information behaviour. This is a threat to one of Wikimedia foundations 2030 strategic goals which aim to create knowledge equity, i.e., to support "the knowledge and communities that have been left out by structures of power and privilege" (Zia, 2019, p. 1). As multilingualism and mutual intelligibility are phenomena that are widespread in the Nordic countries (Denmark, Norway and Sweden), and their Wikipedia language editions are interesting projects regarding their different sizes. At first, this project will focus on Wikipedia reading behaviour in these countries. Initial exploratory analysis showed that although people could make use of their language skills and look up information on either the Danish, Swedish or Norwegian Wikipedia, most page views (~85%) of the three language editions in question come from within the own country. However, all three rely quite strongly on en.wikipedia.org. For Danes and Norwegians, English Wikipedia is even more important than their native language edition. In the future, a mixed-methods approach combining a large quantitative survey and data analysis from the Wikipedia Diversity dataset will study the role of language in Wikipedia usage in greater detail and help with bridging possible knowledge gaps.

References

- Eleta, I., & Golbeck, J. (2014). Multilingual use of Twitter: Social networks at the language frontier. *Computers in Human Behavior*, 41, 424–432. <https://doi.org/10.1016/j.chb.2014.05.005>
- Gooskens, C., van Heuven, V. J., Golubović, J., Schüppert, A., Swarte F., & Voigt, S. (2018). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, 15(2), 169–193, DOI: [10.1080/14790718.2017.1350185](https://doi.org/10.1080/14790718.2017.1350185)
- Guldbrandsson, Lennart (2013). Swedish Wikipedia surpasses 1 million articles with aid of article creation bot. <https://diff.wikimedia.org/2013/06/17/swedish-wikipedia-1-million-articles/>
- Hale, S. A. (2014). Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM conference on Web science (WebSci '14)*. New York, NY: ACM, 99–108. <https://doi.org/10.1145/26155692615684>
- Johnson, I. (2019). Research: Characterizing Wikipedia Reader Behaviour/Demographics and Wikipedia use cases. <https://meta.wikimedia.org/?curid=10664863>
- Johnson, I., Lemmerich, F., Sáez-Trumper, D., West, R., Strohmaier, M., & Zia, L. (2020). Global gender differences in Wikipedia readership. *ArXiv*. <https://arxiv.org/pdf/2007.10403.pdf>
- Kim, S., Park, S., Hale, S., Kim, S., Byun J., & Oh, A. (2016) Understanding Editing Behaviors in Multilingual Wikipedia. *PLOS ONE*, 11(5): <https://doi.org/10.1371/journal.pone.0155305>
- Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*, 20(12), 4608–4633. <https://doi.org/10.1177/1461444818779080>
- Lam, S. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011). WP:clubhouse?: an exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. New York, NY: ACM, pp. 1–10. <https://doi.org/10.1145/2038558.2038560>
- Lee, C., & Chau, D. (2018). Language as pride, love, and hate: Archiving emotions through multilingual Instagram hashtags. *Discourse, Context & Media*, 22, 21–29, <https://doi.org/10.1016/j.dcm.2017.06.002>
- Lemmerich, F., Sáez-Trumper, D., West, R., & Zia, L. (2019). Why the World Reads Wikipedia: Beyond English Speakers. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)*. New York, NY: ACM, pp. 618–626. <https://doi.org/10.1145/3289600.3291021>

- Miquel-Ribé, M., & Laniado D. (2019). Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. In *Proceedings of the 13th International AAAI Conference on Web and Social Media. ICWSM 2019*. [Palo Alto, CA]: Association for the Advancement of Artificial Intelligence.
- Miquel-Ribé, M., & Laniado D. (2020). The Wikipedia Diversity Observatory: A Project to Identify and Bridge Content Gaps in Wikipedia. In *Proceedings of the International Symposium on Open Collaboration (OpenSym 2020)*. New York, NY: ACM. 2 pages. <https://doi.org/10.1145/3412569.3412866>
- Redi, M., Gerlach, M., Johnson, I., Morgan, J., & Zia, L. (2020). A Taxonomy of Knowledge Gaps for Wikimedia Projects (First Draft). *ArXiv*. <https://arxiv.org/pdf/2008.12314v1.pdf>
- Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017). Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. Geneva: International World Wide Web Conferences Steering Committee, pp. 1591–1600. <https://doi.org/10.1145/3038912.3052716>
- Vincent, N., Johnson, I., Sheehan, P., & Hecht, B. (2019). Measuring the Importance of User-Generated Content to Search Engines. In *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01), 505–516. <https://ojs.aaai.org/index.php/ICWSM/article/view/3248>
- Wagner, C., Graells-Garrido, E., Garcia D., & Menczer, F. (2016). Women through the glass ceiling: gender asymmetries in Wikipedia. *EPJ Data Science*, 5(5). <https://doi.org/10.1140/epjds/s13688-016-0066-4>
- Zia, L., Johnson, I., Mansurov, B., Morgan, J., Redi, M., Saez-Trumper, D., & Taraborelli, D. (2019). Knowledge Gaps – Wikimedia Research 2030. doi.org/10.6084/m9.figshare.7698245

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 64–72.
DOI: doi.org/10.5283/epub.44937.

From Information Literacy to Data Literacy Education

A Case Study from Switzerland

Vera Husfeldt, Sharon Alt

University of Applied Sciences of the Grisons, Switzerland

{[vera.husfeldt](mailto:vera.husfeldt@fhgr.ch), [sharon.alt](mailto:sharon.alt@fhgr.ch)}@fhgr.ch

Abstract

Data literacy is considered a key competence for responsible, mature citizens of the 21st century. In this poster the interrelatedness of information literacy and data literacy is addressed. It then elaborates on how universities can make their students fit in dealing with data and information using the example of a new research project about the development of open educational resources to promote data literacy at Swiss universities. Thirteen universities work together in a unique cooperation to promote the key competence data literacy among university teachers and students. The project aims to ensure responsible, critical, ethical and sustainable handling of data during university education and builds upon existing initiatives to promote data literacy (e.g., by the Stifterverband in Germany).

Keywords: information literacy; data literacy; digital skills; future skills; OER; OEP

1 Introduction

Universities face the challenge of preparing their students for the digitalized world of work and research in teaching them so-called future skills. Data literacy, the responsible, critical, ethical and sustainable handling of data, is considered a key competence for responsible, mature citizens of the 21st century. In this poster we address the interrelatedness of information literacy and data literacy. We then elaborate on how universities can make their students

fit in dealing with data and information using the example of a research project about the promotion of data literacy at Swiss universities.

2 The interrelatedness of information and data literacy

The term information literacy was introduced by Paul Zurkowski in 1974 in a letter to the National Commission on Libraries and Information Science, USA (Kelly, 2013). At a time when the information industry changed, he recognized the need for techniques and skills to make the fullest possible use of the new information services. Information literacy is defined as “the ability to think critically and make balanced judgments about any information we find and use. It empowers us as citizens to develop informed views and to engage fully with society” (CILIP, 2018). It involves a set of skills and abilities to undertake information-related tasks, e.g., to discover, access, interpret, analyze, manage, create, communicate, store and share information. It incorporates also critical thinking and a sensibility and awareness of ethical and political issues regarding the use of information.

Today, in our data-driven world the concept of and need for data literacy is expressed (Hochschulforum Digitalisierung, 2018). It is defined as the ability to collect, manage, evaluate and apply data in critical ways (Ridsdale et al., 2015). Data literacy is considered a central competence for digitization and the global knowledge society in all sectors and disciplines, a central competence in the 21st century. When comparing the two concepts information and data literacy, it becomes clear, that they relate to each other and overlap. The concept information literacy can be understood as an overarching, receptive competence as it refers to the handling of all forms of information including data (Schüller et al., 2019). Data literacy describes the handling of data for the purpose of transforming data into knowledge. It requires an understanding of the possible reception as well as an understanding of the data-generating process, e.g., the origin of data, its limitations and the analysis tools used.

3 Promoting data literacy – an approach for Switzerland

The University of Applied Sciences of the Grisons has a strong track record regarding the education of information literacy in Switzerland (Blumer et al., 2013). With its focus on information science and data science, it is only natural to expand the teaching of information literacy to include data literacy. Therefore, a research project “develop data literacy” for the promotion of data literacy is to be developed, implemented and evaluated. The project is part of the large-scale project “Swiss Digital Skills Academy”, a cooperation of thirteen Swiss universities. The academy will target the development of digital skills, especially skills associated with the conception, the creation, and the implementation of open educational resources (OERs) and open educational platforms (OEPs), as well as the actual development of such open resources and their deployment in shared learning modules. “Develop data literacy” addresses the implementation and measurement of data literacy among lecturers and students. The project aims to ensure a responsible, critical, ethical and sustainable handling of data during higher education and beyond based on the data literacy framework developed by Schüller et al. (2019). The project builds upon existing initiatives to promote data literacy (e.g., by Stifterverband in Germany)¹. In order to meet students’ heterogeneous and individual needs, the didactical design of the modules will be organized based on a skills taxonomy of data literacy. Hence, students can acquire a basic, advanced or professional level of skills. The project is composed of five modules:

- a. develop OERs to foster basic data skills
- b. develop OERs to foster the application of data literacy
- c. develop OERs for critical data literacy and digital ethical sensibility
- d. develop OERs for responsible handling of data and data management
- e. develop tools to assess data skills.

As described above, the project focuses on knowledge about basic data skills (from understanding, analyzing to interpreting data), the application of data literacy, the responsible handling of data and data management as well as ethical aspects and critical thinking regarding the use and interpretation of

1 For more details about projects to promote data literacy in Germany funded by the Stifterverband see <https://www.stifterverband.org/data-literacy-education#netzwerk>.

data. Educational concepts may vary, depending on the students' previous knowledge and educational needs of the respective university. The project will develop modular training courses, lectures within study courses address students at the beginning of their studies, knowledge is deepened and applied in courses within a study program. Based on experience on teaching information literacy (Bättig, 2005) the courses are integrated in study programs where possible, the courses are developed in collaboration with other universities, the teachers will be trained in teaching data literacy and the courses will be developed methodically and pedagogically thought out with the help of teacher educators. All teaching and learning materials will be developed as OERs and will be published on an OEP for free re-use by other universities. The project finishes with the evaluation of the acquired competences by students and lecturers.

4 Conclusion

In this poster we describe a research program to promote data literacy in higher education in Switzerland. The modular approach aims for flexibility, expandability regarding the integration of sub-competencies and the strategic alignment to different universities programs. The four pillars strategic alignment, curricular integration, networks and provision of the materials as OERs on an already existing OEP secures a sustainable implementation and fosters re-use of teaching and learning materials regarding the promotion of data literacy.

Acknowledgements

The topic of the panel relates to the ongoing research project "Swiss Digital Skills Academy: Developing Data Literacy" and is supported by Swiss universities. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of Swiss universities.

References

- Bättig, E. (2005). Information Literacy an Hochschulen. Entwicklungen in den USA, in Deutschland und der Schweiz. Chur: Arbeitsbereich Informationswissenschaft.
- Blumer, E., Hügi, J., Bekavac, B., & Schneider, R. (2013). Information literacy competences of LIS-students in Switzerland – A Case Study. In S. Kurbanoğlu et al. (Eds.): *Worldwide Commonalities and Challenges in Information Literacy Research and Practice. European Conference on Information Literacy, ECIL 2013, Istanbul, Turkey, October 22–25, 2013, Revised Selected Papers* (pp. 596–602). Cham: Springer.
- CILIP (2018). CILIP definition of information literacy 2018. Information literacy group. Retrieved, Nov. 18, 2020. <https://infolit.org.uk/ILdefinitionCILIP2018.pdf>
- Hochschulforum Digitalisierung (2018). Strukturen und Kollaborationsformen zur Vermittlung von Data-Literacy-Kompetenzen – Stand der Forschung. Arbeitspapier Nr. 32. Berlin: Hochschulforum Digitalisierung.
- Kelly, J. (2013). Paul G. Zurkowski and information literacy: On his trip to the first European Conference on Information Literacy (ECIL). *Journal of information Literacy*, 7(2), pp. 163–167.
- Ridsdale, C., Rothwell, J., Smit, M., Hossam A.-H., Bliemel, M., Irvine, D., ... Wuetherick, B. (2015). Strategies and Best Practices for Data Literacy Education. Knowledge Synthesis Report. Halifax, Canada: Dalhousie University. DOI: [10.13140/RG.2.1.1922.5044](https://doi.org/10.13140/RG.2.1.1922.5044)
- Schüller, K., Busch, P., & Hindinger, C. (2019). Future Skills. Ein Framework für Data Literacy. Kompetenzrahmen und Forschungsbericht (Forschungsbericht Nr. 47/August 2019). Berlin: Hochschulforum Digitalisierung.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 73–77. DOI: doi.org/10.5283/epub.44938.

Session 2:

Information Behavior and Information Literacy 2

Omission of Information: Identifying Political Slant via an Analysis of Co-occurring Entities

Jonas Ehrhardt

University of Amsterdam
jonas.m.ehrhardt@gmail.com

Timo Spinde

University of Constance
timo.spinde@uni-konstanz.de

Ali Vardasbi

University of Amsterdam
a.vardasbi@uva.nl

Felix Hamburg

University of Constance
felix.hamburg@uni-konstanz.de

Abstract

Due to the strong impact the news has on society, the detection and analysis of bias within the media are important topics. Most approaches to bias detection focus on linguistic forms of bias or the evaluation and tracing of sources. In this paper, we present an approach that analyzes co-occurrences of entities across articles of different news outlets to indicate a strong but difficult to detect form of bias: bias by omission of information. Specifically, we present and evaluate different methods of identifying entity co-occurrences and then use the best performing method, reference entity detection, to analyze the coverage of nine major US news outlets over one year. We set a low performing but transparent baseline, which is able to identify a news outlet's affiliation towards a political orientation. Our approach employing reference entity selection, i.e., analyzing how often one entity co-occurs with others across a set of documents, yields an F1-score of $F1 = 0.51$ compared to $F1 = 0.20$ of the TF-IDF baseline.

Keywords: media bias; bias by omission; news articles; co-occurrences

1 Introduction

News articles play a significant role in informing the public about current events. Many individuals source their information about those events pre-

dominantly from news articles, making them a major influence in understanding and developing an opinion about the covered topics (Bernhardt, Krasa, & Polborn, 2008; Hamborg, Donnay, & Gipp, 2019; Spinde et al., 2021b)

The coverage by news outlets – such as newspapers, TV-channels, and online platforms – often inherits a particular political slant (Agirdas, 2015; Gentzkow & Shapiro 2010). The slant impacts the general way of reporting and manifests in formulations, source-selection, and perspectives (Dallmann et al., 2015; Gentzkow & Shapiro, 2010; Hamborg et al., 2019). While this provides readers with a constant perspective on covered topics, securing a loyal readership (Gentzkow & Shapiro, 2010), it carries the risk of reinforcing incomprehensive opinions and ideologies (Flaxman, Goel, & Rao, 2016; Garrett, 2009; Iyengar & Hahn, 2009; Munson & Resnick, 2010). With the increasing polarization of news outlets, media bias detection and highlighting have become an increasingly important topic (Spinde et al., 2020a, 2020b, 2020c).

Media bias appears in a variety of manifestations. It ranges from strategic decisions, like the selection of covered events and the selection of sources (Agirdas, 2015; Bourgeois, Rappaz, & Aberer, 2018; Gruenewald, Pizarro, & Chermak, 2009), over the operative placement of news articles and their size allocation within the news outlet (Stempel, 1969; Stovall, 1985), down to the structural level of information within a news article (Park et al., 2009) and the linguistic level of labeling and word-choices (Bhowmick, 2009; Papa-charissi & De Fatima Oliveira, 2008).

Whereas most manifestations of media bias already have been covered, research towards bias by omission of information is rare (Hamborg et al., 2019; Park et al., 2009). “Journalists need to select sources, e.g., press releases, other news articles, or studies, to be used when writing an article. Ultimately, the journalists must decide which information from the sources to be included and which to be excluded from the article to be written. This step is called commission or omission, and likewise affects which perspective is taken on the event” (Hamborg et al., 2019). Bias by omission of information therefore describes the *wanted or unwanted discrepancy of information* between news articles from different news outlets covering the same topic (ibid.). It manifests in the omission of perspectives or sources. Hence, it shows similarities with source or event-selection bias (ibid.).

In the social sciences, the detection of bias by omission of information heavily relies on human judgment for the selection of topics or search terms,

which are manually applied to the examined corpora (Smith et al., 2001; Stemler, 2001; Spinde et al., 2021c). In computer science, there have so far been no specific approaches for the detection of bias by omission of information (Hamborg et al., 2019).

In this paper, we derive a first basic approach for detecting political slant of news outlets, which sets the base for bias by omission of information, using co-occurrences of entities. We seek to identify whether mentions or commentions of specific entities in news outlets relate to the political ideology of the outlet. Therefore, we calculate co-occurrences over all detected entities in a news corpus. In the current version of the system, we excluded headlines, but we will integrate them in future work. By applying statistical analyses, we can depict relations in political slant between compared news outlets. As this approach heavily depends on the selection of co-occurring entity tuples, we apply and evaluate different entity selection methods, including standard procedures from *Natural Language Processing* (NLP) like *Term Frequency – Inverse Document Frequency* (TF-IDF) or *Latent Dirichlet Allocation* (LDA). We evaluate the differences in political slant between news outlet tuples by performing standard χ^2 -tests.

Our dataset¹ consists of 76,221 news articles from nine major news outlets in the US over one year. The average article length is 632 words. The results with a maximum accuracy of 0.6250 indicate the potential of future research, and our method² has the potential of being easily extended and improved or used as a simple baseline in further research.

This paper makes two main contributions. First, we provide a simple and comprehensible method for the estimation of political slant between news outlets based on omission of information. Second, we evaluate different approaches for the selection of co-occurring entities

The remainder is structured as follows. Section 2 gives an overview of the research on the detection of media bias by omission of information. Section 3 introduces the used datasets. Section 4 derives the methodology. Section 5 evaluates the outcomes and presents the findings. Section 6 discusses limitations and chances. Section 7 closes with the conclusion.

1 The dataset can be found in the GitHub repository.

2 <https://github.com/the-banandit/OmissionOrCommission>

2 Related work

Within this section, we summarize approaches for the detection of bias media which can be applied to detect bias by omission of information. Further, we derive the research gap and introduce our approach.

Although no distinctive research has been carried out on this topic, some approaches can be partially adapted for the detection of bias by omission of information. Bourgeois et al. (2018) developed a method to detect and depict selection bias regarding the selection of covered events within a news outlet, using an interaction matrix of news outlets and events recovered from the *Global Database of Events, Language, and Tone* (GDELT). This interaction matrix monitors a wide range of news outlets, annotates and records global events, and their coverage (ibid.). By modeling the decision process between a tuple of events, they can identify an inherent preference structure, which indicates a bias in event selection (ibid.). While this methodology applies to the event selection process, it could also be used in information selection by replacing events with sources or information fragments.

An approach by García-Uribe (2018) identified event selection bias on front pages through a pairwise comparison between different news outlets. By modeling a discrete choice model for the selection of an event tuple, she was able to predict the preference for the omission of events. Although this approach also predicts event selection bias, its application on the information selection process could predict bias by omission of information.

Due to their successful application in previous research, co-occurrences of entities are a suitable variable for detecting media bias. Since they can carry additional information about the author's perspective on a topic, they can outrun simple quantitative approaches that employ single entities.

Gentzkow and Shapiro (2010) detected the political slant of news outlets in the US by comparing co-occurrences of single entities or phrases that were initially published by members of Congress and later echoed by news outlets. As a result, they were able to calculate a slant ranking for each examined news outlet.

Following a similar approach, Groseclose and Milyo (2005) included phrases from policy groups and think tanks in their methodology, defining their own slant ranking of news outlets.

Dallmann et al. (2015) conducted a bias analysis of German online newspapers. They conducted a sentiment analysis, and analyzed co-occurring

ideological terms and party references in online news articles. Their results showed a significant difference in reporting between different political parties.

Within the social sciences, where researchers have studied various forms of media bias for decades, the detection of bias by omission of information has been approached with the manual construction and application of co-occurring search terms and phrases (Smith et al., 2001). This approach requires specific domain knowledge and is neither time efficient, nor can be easily automated. Hence, we see a potential in the automatic detection of bias by omission of information. However, no research on the detection of bias by omission of information has been carried out in the field of computer science (Hamborg et al., 2019).

Therefore, we propose an approach utilizing co-occurrences of entities between articles from different news outlets, to identify and detect political slant of news outlets as a first step towards detecting bias by omission of information. Since this approach heavily relies on the selected entities, we tested different compiling methods, utilizing standard procedures and techniques from NLP. With this approach, we aim to determine *whether co-occurrences are applicable to indicate a news outlets' political slant based on the omission of information*, and hence can be an indicator for bias by omission of information in news articles reporting on the same events.

3 Data

To test our approach, we employed a dataset scraped from the *Common Crawl Project*³. Since there is no comparable research regarding bias by omission of information, we had to rely on related research regarding other forms of media bias to create our dataset. We identified three leading publications that cluster US news outlets into slant groups (Budak, Goel, & Rao, 2016; Flaxman et al., 2016; Groseclose & Milyo, 2005; Spinde et al., 2021a). We interpolated the findings of the leading publications to identify the slant of a news outlet. As a simplification, we translated the numerical values into three distinct slant groups: *liberal*, *center* and *conservative*.

3 <https://commoncrawl.org/>

We scraped a total of 76,221 articles from nine different news outlets equally distributed over the three slant categories using the Common Crawl API (cf. Table 1). The dataset contains all articles from the “news” section of the year 2011. We consider articles from this section equivalent to front-page articles, containing only developments determined important by the editors. We chose 2011 since Common Crawl provides an uninterrupted dataset of articles over the whole year and all news outlets. Additionally, 2011 lies between the publication dates of the reference studies (Budak et al., 2016; Flaxman et al., 2016; Groseclose & Milyo, 2005), including the election campaigns of the 2012 US elections, which were especially ambivalently covered. The average word-count per article is 632 words.

Table 1: Composition of the main dataset, covering news articles from 2011-01-01 until 2011-12-31, in the category “news”

News outlet	Number of articles	Slant group
Cable News Network (CNN)	2,652	center
Chicago Tribune (CTB)	2,843	conservative
Fox News (FXN)	6,508	conservative
Huffington Post (HFP)	14,876	liberal
National Broadcasting Company (NBC)	3,958	center
New York Times (NYT)	11,281	liberal
Reuters (RET)	16,767	center
Washington Post (WPO)	14,814	liberal
Wall Street Journal (WSJ)	2,522	conservative

4 Methodology

After stating the research objective, we present the approaches used to extract co-occurrences of entities, select entities, and test for slant group affiliation. In this paper, we seek to answer to what extent co-occurrences of entities between articles of news outlets are beneficial towards identifying bias by omission of information. Specifically, we seek to define the most suitable method for the selection of entities.

4.1 Selecting entities

As baselines for the entity selection we chose *Term Frequency – Inverse Document Frequency* (TF-IDF) (Spärck-Jones, 1972) and *Latent Dirichlet Analysis* (LDA) (Blei, Ng, & Jordan, 2003), as they are simple standard measures in NLP. For the TF-IDF baseline we divided our dataset into corpora containing only news articles from one news outlet. We calculated a ranking for all occurring terms based on the TF-IDF statistic and extracted the top ranked $n = 9$ terms per corpus. For the LDA baseline we also divided the dataset into corpora only containing news articles from one news outlet. We performed the LDA on those corpora, setting the number of topics to $m = 5$. For each topic in each corpus, we selected the top ranked entity. We accumulated all entities over all topics and corpora in an entity list.

We tested the performance of two approaches on the baselines. The first approach is based on *one constant reference entity*. We derived the top $n = 9$ co-occurring entities within each news outlet's corpus for the reference entity. The reference entity has to carry the potential of being differently covered by news outlets from different political orientation. Therefore, the co-occurring entities would differ for each news outlet, but presumably be more similar for news outlets from the same slant group. We derived the reference entity from a test-dataset, engaging the mentioned requirements. All baselines and our reference entity approach are unsupervised approaches.

The second approach is based on *manual selection*. Therefore, we read 100 random articles from outlets of the outer spectrum of the slant ratings. We then manually selected $n = 9$ entities that carried the potential of being ambivalently covered by different news outlets from different slant groups (cf. Table 2). The selection criteria consisted of being mentioned in at least 30 of the 100 random articles. A personal emphasis was put on the eligibility of the considered entities. Of course, the emphasis was heavily dependent on our prior knowledge of linguistic media bias and the knowledge about the covered topics' outcomes.

4.2 Estimating affiliation towards slant groups

To estimate affiliation towards a slant group, we measured the dependence of the selected entities and a tuple of news outlets. Therefore, we constructed contingency tables over the two dimensions news outlets and quantified co-

occurrences. To measure the entities' dependence on the news outlets, we ran standard χ^2 -tests on the contingency tables.

We calculated the dependence of all news outlet combinations for each entity selection method. Thus, we were able to group the news outlets. We evaluated the performance of our methods by comparing the outcomes of all tuple tests with the slant group affiliation we derived from the literature (cf. Table 2), providing the metrics accuracy, precision, recall, and F1-score.

5 Results

Within this section, we present and evaluate our results. Our best-performing method excels the lowest-performing baseline's accuracy by 150% and F1-score by 154.5%. The results of the different methods for entity selection are summarized in Table 2. The reference entity method scored best with an accuracy of 0.6250 and an F1-score of 0.5090. The weakest performing method is the TF-IDF argmax with an accuracy of 0.2500 and an F1-score of 0.2000.

For each entity selection method, we ran the χ^2 -tests of independence for all combinations of news outlet tuples. As a threshold for rejecting the hypothesis, and hence a slant group affiliation, we chose $p = 0.05$.

Table 2: Results for the different methods of entity extraction; the metrics picture the performance of predicting slant group affiliation.

Entity selection method	Accuracy	Precision	Recall	F1-score
TF-IDF argmax	0.2500	0.1250	0.5000	0.2000
LDA argmax	0.5278	0.5278	0.5370	0.4963
Manual selection 1	0.4444	0.5151	0.5185	0.4375
Manual selection 2	0.5833	0.5000	0.5000	0.4958
Reference entity	0.6250	0.5089	0.5093	0.5090

The results in Table 2 show an increasing performance for more complex methods for entity selection. Simple methods, like *TF-IDF argmax*, utilize only high-level, general features of documents and corpora. Thus, the utilized features show similarities, particularly for large corpora; low performance was both expected and confirmed.

More sophisticated approaches, like the *LDA* or *reference entity* approaches, are of higher complexity and consider more abstract properties of the corpora, like the distribution of topics and their composition. Hence they are able to assess the omission of information between different corpora comparatively.

Especially the *reference entity* approach seems eligible for the detection of bias by omission of information, as it considers the top n co-occurring entities of a reference entity explicitly. It enables a direct comparison between the weighting of entities over different corpora and represents the omission of topics or information. Although the performance with an F1-score of 0.5090 is relatively low, it outruns the baselines. A variation and improvement of the parameters and settings could increase its performance.

The most sophisticated approach of *manual selection* involves human decision-making. The peculiarities of omitted information are perceived subconsciously and subjectively. Therefore, it is highly dependent on the reader's previous experience and sensitivity towards bias terms.

6 Discussion

Since this paper is a first approach for the detection of news outlets' political orientation based on bias by omission of information, it entails multiple constraints. This section discusses limitations in data, methods, and results.

6.1 Data

As there were no existing datasets for the detection of bias by omission of information, we had to compile a new dataset. Therefore, we interpolated the results of three reference publications that evaluated news outlets' political slant. The resulting dataset contains articles published only in 2011, since the reference publications are from 2005 and 2016, and Common Crawl could provide a complete, uninterrupted dataset for all examined news outlets only in 2011. The discrepancy in the publication dates of the reference publications and our dataset could challenge the internal validity of the interpolated biases. However, there is proof that news outlets that inherit a particular bias tend to retain it due to economic reasoning (Gentzkow & Shapiro, 2010). There-

fore, we could assume that there is no shift in the considered news outlets' inherent bias. Additionally, a period of one year of coverage could be considered too short to generate universally valid results. Since the examined type of media bias is based on the comparison of information, the period does not play a significant role, rather the integrity of the dataset. Providing complete coverage over multiple outlets is more important for evaluating and depicting bias by omission of information.

We simplified the numerical ratings of the slant group estimations into three distinct classes. This simplification can be seen as an oversimplification, which could distort the results. Certainly, a numerical representation of bias provides a more accurate indication of the political slant. As the publications we interpolated are measuring different forms of bias, a simplification into distinct abstract groups ensures the usability of the used classifications for our application. The duality of the United States' political landscape supports a division of outlets into the three used categories.

The article number discrepancy between the news outlets is significant for some outlets. As the outlets employ different guidelines for assigning the news articles to the domains, there are differences in the numbers and content of the articles. We tackle the discrepancy in numbers through our method (cf. Section 4).

6.2 Methodology and results

As this paper provides a first approach to detecting news outlets' political orientation based on bias by omission of information, we set our focus when devising methods on transparency and design simplicity. Therefore, we favored statistical tests over deep models.

For the automated approaches, the final number of co-occurrences might have been too little. We did not test the behavior of different numbers of co-occurrences. We applied the argmax criterion for the news outlets. A proper investigation of the number of co-occurrences, combined with improvements in the methodological approaches, could be a valuable extension of this research.

As our approach's performance heavily relies on selected entities for the co-occurrences, the improvement of the selection process would be a valuable extension. In our approach, we employed basic NLP-methods to select entities for further testing. Those approaches did not show high performance, as they did not consider the peculiarities of bias by omission of information.

A valuable extension to those methods would be the utilization of *ideology-* or *bias-terms* (cf. Dallmann et al., 2015). By combining those with reference entities, e.g., extracted with LDA, our approach's performance could be significantly improved.

Our method showed that co-occurrences are applicable for comparatively classifying the political orientation of news outlets. Though this is not a detection of media bias by omission of information, we consider it as a first step towards identifying potentially biased news outlets, and hence a first step towards identifying media bias by omission of information across news outlets.

7 Conclusion

In this paper, we presented a method that is able to depict the affiliation of news outlets towards a political orientation, based on the omission of co-occurring entities. We consider this method as a first step towards the detection of bias by omission of information in news articles reporting on the same event.

We tested our method on a dataset consisting of news coverage over a period of one year. Although the results show a low performance, we exceeded the baselines derived from standard NLP methods by far, i.e., $F1 = 0.20$ of the TF-IDF baseline compared to $F = 0.51$ yielded by our reference entity selection.

Further research can improve the performance by examining the number of co-occurrences. Also, we seek to focus on more comprehensive evaluations, e.g. testing the approach on datasets that strongly diverge in size and number of news outlets. Further, the entity selection method could be extended by an automatic matching with ideology- or bias-terms, automatic entity extraction, or sentiment analysis.

References

- Agirdas, C. (2015). What Drives Media Bias? New Evidence From Recent Newspaper Closures. *Journal of Media Economics*, 28(3), 123–41. [10.1080/08997764.2015.1063499](https://doi.org/10.1080/08997764.2015.1063499)
- Bernhardt, D., Krasa, S., & Polborn, M. (2008). Political Polarization and the Electoral Effects of Media Bias. *Journal of Public Economics*, 92(5–6), 1092–1104. doi: [10.1016/j.jpubeco.2008.01.006](https://doi.org/10.1016/j.jpubeco.2008.01.006)
- Bhowmick, P. K. (2009). Reader Perspective Emotion Analysis in Text through Ensemble Based Multi-Label Classification Framework. *Computer and Information Science*, 2(4). doi: [10.5539/cis.v2n4p64](https://doi.org/10.5539/cis.v2n4p64)
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bourgeois, D., Rappaz, J., & Aberer, K. (2018). Selection Bias in News Coverage. In *Companion of the The Web Conference 2018 – WWW '18* (pp. 535–543). New York, NY: ACM.
- Budak, C., Goel, S., & Rao, J. M. (2016). Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly*, 80, 250–271.
- Dallmann, A., Lemmerich, F., Zoller, D., & Hotho, A. (2015). Media Bias in German Online Newspapers. In R. Bergmann, S. Görg, & G. Müller (Eds.), *Proceedings of the 26th ACM Conference on Hypertext & Social Media – HT '15, {CEUR} Workshop Proceedings* (pp. 133–137). New York, NY: ACM Press.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80(S1): 298–320. doi: [10.1093/poq/nfw006](https://doi.org/10.1093/poq/nfw006)
- Garcca-Uribe, S. (2018). Multidimensional Media Slant: Complementarities in News Reporting by US Newspapers. *SSRN Electronic Journal*. doi: [10.2139/ssrn.3195751](https://doi.org/10.2139/ssrn.3195751)
- Garrett, R. K.. (2009). Echo Chambers Online? Politically Motivated Selective Exposure among Internet News Users. *Journal of Computer-Mediated Communication*, 14(2): 265–285. doi: [10.1111/j.1083-6101.2009.01440.x](https://doi.org/10.1111/j.1083-6101.2009.01440.x)
- Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1), 35–71. doi: [10.3982/ecta7195](https://doi.org/10.3982/ecta7195)
- Groseclose, T., & Milyo, J. (2005). A Measure of Media Bias. *The Quarterly Journal of Economics*, 120(4), 1191–1237. doi: [10.1162/003355305775097542](https://doi.org/10.1162/003355305775097542)

- Gruenewald, J., Pizarro, J., & Chermak, S. M. (2009). Race, Gender, and the Newsworthiness of Homicide Incidents. *Journal of Criminal Justice*, 37(3), 262–72. doi: [10.1016/j.jcrimjus.2009.04.006](https://doi.org/10.1016/j.jcrimjus.2009.04.006)
- Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated Identification of Media Bias in News Articles: An Interdisciplinary Literature Review. *International Journal on Digital Libraries*, 20(4), 391–415. doi: [10.1007/s00799-018-0261-y](https://doi.org/10.1007/s00799-018-0261-y)
- Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, 59(1), 19–39. doi: [10.1111/j.1460-2466.2008.01402.x](https://doi.org/10.1111/j.1460-2466.2008.01402.x)
- Munson, S. A., & Resnick, P. (2010). Presenting Diverse Political Opinions. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems – CHI '10* (Vol. 3, p. 1457). New York, NY: ACM Press.
- Papacharissi, Z., & De Fatima Oliveira, M. (2008). News Frames Terrorism: A Comparative Analysis of Frames Employed in Terrorism Coverage in U.S. and U.K. Newspapers. *International Journal of Press/Politics*, 13(1), 52–74. doi: [10.1177/1940161207312676](https://doi.org/10.1177/1940161207312676)
- Park, S., Kang, S., Chung, S., & Song, J. (2009). NewsCube. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems – CHI 09* (p. 443). New York, NY: ACM Press.
- Smith, J., McCarthy, J. D., McPhail, C., & Augustyn, B. (2001). From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C. *Social Forces*, 79(4), 1397–1423. doi: [10.1353/sof.2001.0053](https://doi.org/10.1353/sof.2001.0053)
- Spärck-Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, 28: 11–21.
- Spinde, T., Hamborg, F., Becerra, A., Donnay, K., & Gipp, B. (2020). Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2020.
- Spinde, T., Hamborg, F., & Gipp, B. (2020). An Integrated Approach to Detect Media Bias in German News Articles. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. Virtual Event China* (pp. 505–506). ACM Press. <https://doi.org/10.1145/3383583.3398585>
- Spinde, T., Lada, R., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., & Donnay, K. (2021). Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, 58(3). <https://doi.org/10.1016/j.ipm.2021.102505>
- Spinde, T., Rudnitskaia, L., Hamborg, F., & Bela, G. (2021). Identification of Biased Terms in News Articles by Comparison of Outlet-specific Word Embeddings. In

Proceedings of the 16th International Conference (iConference 2021). Virtual Event China. Springer Nature [Preprint].

Spinde, T., Rudnitckaia, L., Sinha, K., Hamborg, F., Bela, G., & Donnay, K. (2021). MBIC – A Media Bias Annotation Dataset Including Annotator Characteristics. In *Proceedings of the 16th International Conference (iConference 2021). Virtual Event China.* Springer Nature [Preprint].

Stemler, S (2001). An Overview of Content Analysis. *Practical Assessment, Research and Evaluation*, 7(17), 2000–2001. doi: [10.1362/146934703771910080](https://doi.org/10.1362/146934703771910080)

Stempel, G. H. (1969). The Prestige Press Meets the Third-Party Challenge. *Journalism Quarterly*, 46(4), 699–706. doi: [10.1177/107769906904600402](https://doi.org/10.1177/107769906904600402)

Stovall, J. G. (1985). The Third-Party Challenge of 1980: News Coverage of the Presidential Candidates. *Journalism Quarterly*, 62(2), 266–271. doi: [10.1177/107769908506200206](https://doi.org/10.1177/107769908506200206)

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 80–93. DOI: doi.org/10.5283/epub.44939.

Does the General Public Share Research on Twitter?

A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany

Steffen Lemke¹, Paula Bräuer^{1,2}, Isabella Peters^{1,2}

¹ ZBW – Leibniz Information Centre for Economics, Germany

² Kiel University, Germany

{[s.lemke](mailto:s.lemke@zbw.eu), [p.brauer](mailto:p.brauer@zbw.eu), [i.peters](mailto:i.peters@zbw.eu)}@zbw.eu

Abstract

The search for a final nuclear repository in Germany poses a societal and political issue of high national medial presence and controversy. The German Repository Site Selection Act demands the search to be a “participatory, science-based [...] process”. Also, the repository search combines numerous scientific aspects (e.g., geological analyses, technical requirements) with broad societal implications. For these reasons it constitutes a promising background to analyze the general public’s habits regarding referencing research on Twitter. We collected tweets associated with the conversation around the German nuclear repository search based on keywords. Subsamples of the resulting tweet set are coded regarding sending users’ professional roles and types of hyperlinked content. We found the most vocal group participating in the conversation to be activists and initiatives, while journalists constituted the follower-wise most influential accounts in the sample. Regarding references to scientific content, we found only very few cases of direct links to scholarly publications; however, several kinds of indirect references to academic findings could be identified, e.g., links to paraphrases of studies in news articles or blog posts. Our results indicate participation from a fairly diverse set of users in the observed communication around the German repository search; exchanges of research findings however appear to have happened rarely and been limited to very few particular studies. The findings also illustrate a central problem regarding the expressive power of social media-based altmetrics, namely that a large share of signals indicating a scholarly work’s influence will not be found by searching for explicit identifiers.

Keywords: Twitter; research references; scholarly use of social media; scholarly communication; nuclear repository search; altmetrics

1 Introduction

The rise of social media equipped researchers with highly promising new tools for scholarly communication. The newly created platforms are believed to enable immediate dissemination of research to a virtually unrestrained target audience, both from academic and from non-academic spheres. The anticipated result would be a faster, more responsive and more open landscape of science communication, as social media would not only allow for more immediate exchange within the research domain, but also to an extent move those exchanges to (semi-)public realms, therefore making scientific processes and outcomes easier accessible to the general public as well (Bartling & Friesike, 2014). While the concept of electronic publishing – titled the most recent of the four major revolutions in human kind’s production of knowledge by Harnad (1991) – has of course already been around for a significantly longer time than social media, the success of the latter led to the broad availability of an infrastructure to substantially increase and accelerate those exchanges of and about research, especially between academia and public.

Since those beginnings of social media, an ever-growing body of studies has made the effects of social media on scholarly and science communication their subject of examination (Sugimoto, Work, Larivière, & Haustein, 2017). Frequently, the focus of these studies was on the scientists, often addressing questions regarding which platforms researchers use in the context of their work, and which specific needs they have to meet (e.g., Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011; Tenopir, Volentine, & King, 2013; Van Noorden, 2014; Lemke & Peters, 2019). Presumably less attention so far has been paid to the question in which regard the general public participates in the communication of science on social media (see also Sugimoto et al., 2017). While there has been research on the backgrounds of the users behind interactions with research online (e.g., Haustein & Costas, 2015; Tsou, Bowman, Ghazinejad, & Sugimoto, 2015), there is comparatively little knowledge about whether members of the general public actively distribute academic research by referencing it when engaging in societal or political debates on

social media, e.g., to strengthen their own arguments. Whether this is the case is of particular interest for the field of scientometrics: altmetrics, which embody the comparatively young concept of capturing scientific publications' influence by measuring their prevalence on online domains, are frequently associated with the hope that they might reflect research's influence among the general public (Wouters & Costas, 2012) – in contrast to their traditional predecessors, bibliometric citations. Finding non-academic actors to follow habits like referencing scientific articles in online discussions would provide an empirical basis for this claim.

Of the various social media platforms that presented themselves to the scholarly community over the years, the microblogging service *Twitter* stands out as one of the most versatile. The literature has identified numerous aspects of academic work for which Twitter is used by researchers, such as the discovery of new research or collaborators, the identification of recent trends of public interest, or the external communication of science, to name a few (Lemke & Peters, 2019; Van Noorden, 2014). This versatility, alongside the fact that Twitter data can be obtained with comparative ease via its APIs, likely also contributed to its high popularity as a research subject (e.g., Java, Song, Finin, & Tseng, 2007; Priem & Costello, 2010; Hadgu & Jäschke, 2014; Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Holmberg, Bowman, Haustein, & Peters, 2014; Haustein & Costas, 2015; Syn & Oh, 2015; Robinson-Garcia, Costas, Isett, Melkers, & Hicks, 2017; Schmitt & Jäschke, 2017; Ke, Ahn, & Sugimoto, 2017; Vainio & Holmberg, 2017).

In this case study, we set out to characterize the role of academic research in the Twitter conversations around a political and societal controversy, the search for a nuclear repository in Germany. We choose this use case, as it has direct scientific aspects to it (e.g., analyses of geological conditions, legal foundations, or technical requirements for safe containers), while not being inherently academic in itself. The topic is also of particular interest because in Germany, the legislator in the description of the search procedure for a repository attaches great importance to the involvement of the population.¹ Twitter/social media could be a way to gain insight into ongoing discussions within the population and thus perhaps also offer options for participation in the search process. Since the first step of the search process is based on scientific evidence,¹ this is a promising context to observe whether non-academic users also refer to research on Twitter, e.g., to substantiate their

1 https://www.gesetze-im-internet.de/standag_2017/BJNR107410017.html

own claims. Also, the repository search has been a topic of recurring medial presence in Germany over the past years, particularly over the weeks around September 28, 2020, when the possible sites for a repository in Germany were announced.² We therefore expect a substantial amount of conversation on it to have happened on Twitter.

There have been several previous studies examining either the Twitter communication around critical societal controversies or the way research is referenced on Twitter and by whom. Pearce, Holmberg, Hellsten, and Nerlich (2014) analyzed Twitter conversations associated with the publishing of the Intergovernmental Panel on Climate Change's Working Group I report, which constituted a significant event in the public debate on climate change. They coded participants based on their stance towards climate change, finding participants to be most likely to converse with other participants of similar views. Related to the same real world event, Holmberg and Hellsten (2016) also analyzed the development of hashtag use over time. One of their findings is that hashtags are only to a limited extent capable of indicating shared communities of tweeters or completely shared issues online.

Moscrop, Wong, and Alperin (2020) analyzed the tweets sent by a small sample of Canadian political pundits for whether they use Twitter to share scholarly research and if so, for which motives. They found 78% of their sample to share scholarly research on Twitter, although most pundits would do so very infrequently. Tsou et al. (2015) coded 500 Twitter users that had referenced an article from one of four prestigious academic journals, finding more than a third of the coded users to possess a PhD – a much higher proportion than among the general population – suggesting that references to research articles in tweets are primarily made by academics. In line with this hypothesis are findings by Vainio and Holmberg (2017): in their extensive analysis of a sample of tweets mentioning academic articles by Finnish authors, they also coded subsamples of the senders of said tweets, again finding high shares of researchers and professors among them. Even more recently, Alperin, Gomez, and Hausteine (2019) analyzed the follower networks of users that had tweeted about a small sample of biology articles with overall high Twitter uptake. While they found diffusion patterns of scholarly articles on Twitter to take diverse forms, they also noticed their diffusion to the public to usually be low.

2 <https://www.bge.de/de/endlagersuche/meldungen-und-pressemitteilungen/meldung/news/2020/8/468-endlagersuche-2/>

Priem and Costello (2010) examined scholars' practices regarding citing on Twitter through a combination of semi-structured interviews and quantitative analyses of tweets. They found the scholars of their sample to frequently cite research publications on Twitter, although in about half of all measured cases in an indirect fashion ('second-order citations') by referencing intermediate webpages linking to the intended resource, instead of referencing that resource itself. Thelwall, Tsou, Weingart, and Holmberg (2013) coded tweets linking to a selection of prestigious journals or popular digital libraries, finding many of the coded tweets to be summaries of the linked research, often including the original publication's title but rarely any author attribution.

The political debate on social media platforms such as Twitter on nuclear energy has already been the subject of various studies as well. Several studies examine the discourse at the level of language (Kim & Kim, 2014; Liu & Na, 2018), actors (Arlt, Rauchfleisch, & Schäfer, 2018), and with regard to the delivery of narratives (Gupta, Ripberger, & Wehde, 2018). However, the extent to which scientific work can influence the discourse on these platforms on topics such as final disposal remains an open question. For our use case of the German repository search this question is of particular interest, as Twitter might provide us with an opportunity to observe hints as to whether its process is as participatory and science-based as the legislator meant it to be.

One way to tackle our research interest would be to look at mentions of scientific publications on Twitter that were captured by an altmetrics data provider, for instance Altmetric.com or PlumX, and then determine whether the users involved come from an (non-)academic background, e.g., by analyzing their Twitter profiles (see for instance Haustein & Costas, 2015). An advantage of such an approach is that it most certainly provides the analysts with a reasonably high number of publication mentions to analyze. However, it comes with downsides as well: first, it rigidly only factors in what the respective altmetrics data provider considers to be mentions of research publications. As said data providers rely on certain persistent identifiers and whitelisted domains to track such mentions,³ the data obtained will likely be an underestimation of the true amount of mentions and might miss relevant cases. This is especially problematic as previous research suggests that a

3 See also <https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track> for an example of this.

significant share of references to scholarly objects on Twitter happens indirectly, i.e., by linking to intermediate websites instead of by directly linking to the respective scholarly objects (Priem & Costello, 2010). Second, while query-based searches for altmetric mentions with said data providers can be very specific regarding properties of the scientific publications to consider, there is no easy solution to control for the topical context in which the tracked mentions took place, as long as full texts of the mentioning posts are not available (which is typically the case with Twitter data from altmetric providers). For example, it would be very laborious to track down a substantial amount of mentions of scientific research that were part of statements about a specific real-world event of interest, like the German nuclear repository search in our case. And third, only looking at the isolated examples of when an event occurred – in this case a mention of a research publication – can make it difficult to estimate how frequently said event really occurs in practice, depending on the comprehensiveness of data available.

Therefore, we choose a different approach by starting with data directly obtained from Twitter, which we then analyze with both automatic and manual approaches. Our analysis follows two main objectives:

1. to get an overview over the conversations surrounding our use case on Twitter, with a particular focus on the backgrounds of its most active participants;
2. to examine whether participants of said conversations actively reference research, either by posting DOIs of academic works, by including hyperlinks to scientific publications or academic websites, or by citing tweets by academic users, and if this happens dependent of their own professional role.

We start by fetching a corpus of tweets containing certain keywords indicating a relation to our use case. As a next step, we describe the corpus' properties statistically and identify the most influential participants contributing to the conversation. Finally, we code a sample of that corpus for references to research as well as senders' biographies for hints on their professional role, e.g., whether they are themselves part of academia.

2 Data and methods

For data collection we used TAGS⁴, a free Google Sheet-based tool for archiving tweets which utilizes the Twitter Search API⁵ to automatically perform keyword-based queries over extended periods of time. We programmed our TAGS instance to fetch tweets containing at least one of seven German keywords related to the nuclear repository search. The seven terms were intended to cover common German terminology on the subject of final disposal of radioactive waste and all related topics. Tweet collection ran from July 29 to October 13, 2020, so for a little more than twelve weeks around the date of the announcement of the final repository's potential sites on September 28, 2020. The keywords we used, their English translations, as well as the numbers of tweets retrieved per keyword are depicted in Table 1.

Table 1: Results of keyword-based tweet collection

Keyword used	English translations	# of retrieved tweets
Atomausstieg	nuclear (power) phase-out	501
Atomenergie	atomic/nuclear energy/power	373
Atomkraft	atomic/nuclear energy/power	2,541
Atommüll	atomic/nuclear/radioactive waste	5,361
Endlager	final/permanent disposal site/repository	4,807
Kernenergie	atomic/nuclear energy/power	7,612
Kernkraft	atomic/nuclear energy/power	325

During manual inspection of the tweet sets, we noticed a substantial number of Dutch tweets in the *Kernenergie*-set, due to the same word existing in the Dutch language. To keep a stronger focus on the German discussions surrounding the nuclear repository search and to avoid language-related complications during the content analysis, we therefore removed the *Kernenergie*-set from our data, leaving us with a total of 13,908 tweets. A removal of duplicates based on tweet ids further reduced our data to 10,884 unique tweets from a total of 5,616 individual users.

In the following, we examine the tweets' basic content properties and the most active users in our dataset to achieve an overview over the Twitter

4 <https://tags.hawksey.info/>

5 <https://developer.twitter.com/en/docs/api-reference-index>

communication on the German nuclear repository search. Afterward, we specifically look for traces of references to scientific publications and findings, to get an understanding of the role scientific results might play in the observed discussions. The coding of users and tweets is done by one author (SL), who manually inspects respective tweet- and profile pages; complementary Google searches are used to increase the coding's accuracy. The coding schemes are developed and continually adjusted during the coding process.

3 Results and discussion

Of the 10,884 unique tweets from our dataset, a total of 7,207 tweets (66.22%) could be identified as being retweets (meaning they are non-distinct copies of other, original tweets), indicated by their texts starting with the string “RT @” in TAGS' output. In the following, we use the remaining subset of 3,677 unique *and* original tweets to examine how different users participated in the conversations we tracked.

3.1 User analysis

The 3,677 unique and original tweets were sent by a total of 1,808 users, meaning that only 32.19% of the users involved in the communication covered by our dataset also participated actively by contributing at least one original tweet. Of those 1,808 users, 1,347 individuals appeared with exactly one tweet in our dataset. The distribution of tweets over the remaining 461 more active users is shown in Figure 1. As can be seen, few users are responsible for a large share of all original tweets from our sample, while most users are only represented by very few tweets each – a common pattern regarding communication both on Twitter as well as on social media in general (Haustein, 2019).

Next, we look at the characteristics of the users that contributed the most tweets to our sample. To do this, we manually examined the Twitter profile pages of the 50 most active accounts, who together were responsible for 1,161 (31.57%) of all unique original tweets. Based on the Twitter biographies provided by the users themselves, we determined whether said accounts be-

longed to (1) individuals or groups, (2) the primary profession or role of the account's owner, and (3) if said owner evidently has an academic background. Table 2 exemplarily shows the coding for the ten most active accounts alongside their counts of followers and tweets in the sample. Table 3 shows the shares of the groups resulting from the coding among all coded accounts, as well as the aggregated numbers of tweets (in the sample) and followers of all accounts from said groups. It should be noted that the same follower could of course be following multiple accounts from within the same group, in which case that follower would be counted multiple times in that group's aggregation. Table 4 contains additional short explanations of the most important criteria leading to an account being declared to belong to a specific category.

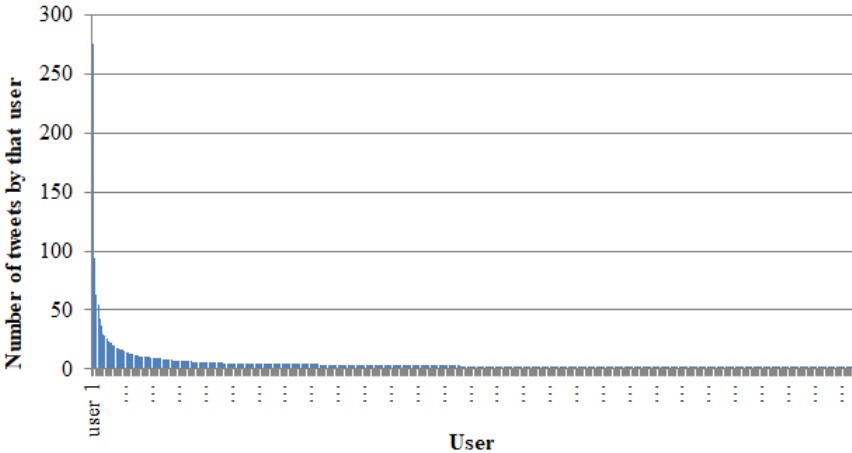


Fig. 1 Distribution of original tweets over users who each were responsible for at least two of the tweets in the sample

We can infer from Tables 2 and 3 that the accounts responsible for the most tweets belong to activists/initiatives, followed by accounts associated with journalism. If we look at follower counts, the category journalism appears to contain by far the most influential accounts in our sample. Out of the ten accounts with the most followers, nine are associated with journalism – particularly strong outliers to the top are @tagesschau, @zeitonline, and @Tagesspiegel, which as the top 3 together reach an aggregated number of 5,439,395 followers. Accounts belonging to scientists, engineers or research institutes seem to play a comparatively minor role, regarding numbers of accounts, tweets, or aggregated followers.

Table 2: Coding results of the ten most active accounts in the sample

#	Individual or group?	Role	Academic background?	# of tweets in sample	# of followers
1	individual	activism/initiative	no evidence	276	1686
2	group	activism/initiative	no evidence	94	10,759
3	group	activism/initiative	no evidence	63	31
4	group	company	no evidence	55	368
5	individual	science/engineering	yes	43	61
6	individual	journalism	no evidence	37	2,927
7	individual	journalism	no evidence	30	763
8	individual	activism/initiative	no evidence	28	17
9	group	website	yes	26	2,694
10	group	activism/initiative	no evidence	24	22

*Table 3:**Frequencies of coded variables among the 50 most active users in the sample*

Coding variable	Share	Aggregated # of tweets	Aggregated # of followers
individual	44%	631	36,109
group	54%	522	5,953,832
unidentifiable	2%	8	270
activism/initiative	26%	598	170,395
company	4%	69	1,268
science/engineering	12%	112	3,334
federal agency	2%	14	605
journalism	36%	230	5,802,531
website	4%	41	3,458
other/unidentifiable	16%	97	8,620
academic background	16%	157	13,309
no evidence for academic background	84%	1004	5,976,902

Table 4: Descriptions of coding categories for Twitter accounts

Coding variable	Description
individual	There is evidence to suggest that the account's posts reflect the thoughts and interests of exactly one user, e.g., a single real name.
group	There is evidence to suggest that the account's posts reflect the thoughts and interests of multiple people or users, e.g., a company.
unidentifiable	none of the above
activism/initiative	The account's primary purpose is to advertise a clear political or societal mission that its owner(s) aim to achieve; includes representatives of political parties.
company	The account represents a company or firm. Note: can also be a federally owned company – an example for this is @die_BGE.
science/engineering	The account represents an individual clearly identifying as an academic, researcher, or engineer, or an academic institution or group, e.g., a university.
federal agency	The account represents a governmental agency or institution.
journalism	The account represents an individual clearly identifying as a journalist, or a journalistic institution or group, e.g., a newspaper.
website	The account represents an online platform or service, e.g., a file sharing service or a non-journalistic and non-scientific blog.
other/unidentifiable	none of the above
academic background	There is evidence to suggest that the individual or group behind the account professionally identifies as a part of the academic enterprise, e.g., as an academic researcher or lecturer.
no evidence for academic background	There is no such evidence.

3.2 References to research or researchers

In our first attempt to obtain links to academic research, we performed a string search for the term “doi” over all collected tweet texts, to see if any scientific

publications were referenced via digital object identifiers⁶. This approach led to exactly one valid DOI⁷ belonging to a scientific journal article.

However, a significant amount of references to academic works could happen less directly (see also Priem & Costello, 2010), e.g., in the form of references to webpages who themselves paraphrase or link to scholarly publications. We therefore first determined all tweets from our sample containing at least one outgoing hyperlink by filtering for occurrences of the string “http”, leading to a subset of 2,463 tweets containing hyperlinks. Of these, we took a random subsample of 250 tweets, which we would then code regarding the types of resources they linked to by visiting the respective webpages. As our main interest was the identification of references to any kinds of research, we allowed for more specific subcategories in that area, even though this might lead to some very rare categories. Table 5 shows numbers of occurrences per category identified this way and provides descriptions of the individual categories’ meanings.

Regarding direct references to academic works, in addition to the one DOI found earlier, our manual coding retrieved only one further case, namely a link to a scholarly book on energy transition. Also, a link to a geographical educational resource might be considered as a direct reference to a specific academic work. However, the conversations we tracked are not as devoid of links to scientific content as this low number of direct references alone would suggest: the category “paraphrased study/report” refers to a number of tweets that link to pages or graphics summarizing or paraphrasing what appear to be results from academic studies or reports. Those paraphrasing websites – oftentimes blogs – vary substantially regarding their format, style, and attention to detail. Therefore, also the difficulty to unambiguously attribute studies or results they reference to certain academic publications varies strongly. On the pages linked to in our tweet sample, however, direct unambiguous links to scholarly articles seemed to be the exception rather than the rule. Similar observations could be made for the category “popular science/tidbits” – a group of online resources with the primary purpose of effectively conveying complex information, often from academic contexts.

While the aforementioned categories in many cases lacked explicit references to the scholarly works they were based on, they are quite easily distinguishable as resources based on some kind of research, although the latter’s

6 See <https://www.doi.org/> for more information on digital object identifiers.

7 <https://doi.org/10.1111/zygo.12268>

specific origins might often not be clear. A further, even less immediate but quite frequently occurring way for users to indirectly reference research can be found within the category “news/journalism”. As can be seen in Table 5, the most common type of links in our coded sample is to journalistic articles (which might be partially explained by the high number and reach of journalism accounts found during the user coding, see subsection 3.1). Journalistic media of course regularly report on findings from academic studies – a particular example from our coded sample was a study from the University of Sussex on nuclear power’s potential to lead to reduced carbon emissions (Sovacool, Schmid, Stirling, Walter, & MacKerron, 2020) that was published recently in *Nature Energy*. Said study’s findings were picked up by several news outlets, with seven observed references to such news articles in our coded sample alone. Again, it became apparent that the diligence with which original scientific sources are indicated varies substantially between different journalistic sources. Another type of indirect references to research(ers) included in the “news/journalism” category are links to interviews with researchers, for which we found four examples in our coded sample.

Table 5: Frequencies and descriptions of coding categories for linked resources among 250 tweets with outgoing links

Coding variable	Description	Occ.
educational content	a link to an educational resource aimed at usage by teachers or lecturers	1
embedded tweet	a direct embedding of another post on Twitter	27
image	a link to a picture, often a photograph	28
news/journalism	a link to a journalistic article reporting rather neutrally on past events, for instance from an online news platform or magazine	114
opinion piece	a link to a text deliberately expressing a non-neutral political or societal stance on a topic	35
paraphrased study/report	a link to a text or graphic with the purpose of summarizing or paraphrasing results from one or more (supposedly academic) studies or reports; often in the context of a blog	20
popular science/tidbits	a link to a text, video, or graphic explaining a complex and/or academic topic or fact, with a focus on being accessible	16
scientific publication	a link to a scholarly publication	1
other	something else, e.g., links meant for pure entertainment	8

One further category warranting more attention is “embedded tweets”. Figure 2 shows an example of such a tweet embedding from our sample. Depending on the tweet embedded, this category obviously provides another way of indirectly linking to research. In fact, among the 27 cases of embedded tweets in our sample we found three to be references to aforementioned *Nature Energy* article, three to be direct citations of statements by researchers, and two each referencing paraphrases of studies, reports, or popular science (according to the definitions from Table 5). Counting all these mentions as indirect references to research of some kind, we can conclude that about a third of the tweet embeddings in our coded sample are used to link to research content. Furthermore, five embedded tweets referenced news articles, which again could contain references to research, as we have seen before.



Fig. 2 Example of an embedded tweet (account information anonymized)

Summarizing our content analysis of outgoing links in our coded sample, we can conclude that direct references to scholarly publications, e.g., via DOI or by linking directly to a publication page, seem to happen extremely rarely. However, we found several examples for academic studies and actors being referenced indirectly in the conversations we observed. In these cases, outgoing links referenced mediating entities, which in turn (with varying degrees of traceability) referenced or paraphrased research. Such mediating entities include for instance news articles, blog posts, or other embedded tweets.

In our next and final step of analysis, we will combine our findings from the previous analyses by coding the accounts behind the 250 tweets for which we examined outgoing links regarding their roles, in line with our user analysis from subsection 3.1. This should allow us to see to which degree certain user groups are responsible for certain types of research mentions. Table 6 shows this data as a contingency table.

Table 6: Contingency table of tweet senders' roles and tweets' linked content

	Activism/ initiative	Com- pany	Federal agency	Journa- lism	Science/ enginee- ring	Other/ uniden- tified	Website	Sum
Educational content	0	1	0	0	0	0	0	1
Embedded tweet	12	1	0	2	4	8	0	27
Image	12	1	0	3	7	5	0	28
News/ journalism	67	3	1	32	7	2	2	114
Opinion piece	26	1	0	2	1	3	2	35
Other	3	0	1	1	1	1	1	8
Paraphrased study/ report	8	0	1	1	10	0	0	20
Popular science/ tidbits	2	7	3	0	0	3	1	16
Scientific publication	0	0	0	0	0	1	0	1
Sum	130	14	6	41	30	23	6	250

The sums in Table 6's bottom line confirm some observations we had made previously for the 50 most active accounts in Table 3, namely the very high share of Twitter activity contributed by accounts devoted to activism, with journalism and science/engineering being the next most vocal categories. Looking at the role-wise shares of link types reveals that some role categories indeed appear to have their specialties – for instance, most links to opinion pieces come from accounts of the category activism/initiative, the overwhelming majority of tweets from accounts identified as journalistic does link to news pages, and no other group links as frequently to para-

phrases of research studies as the category science/engineering. Although this analysis indicates an association between a Twitter user's professional role and its likelihood to disseminate research content, it also suggests that users with scientific background are far from being the only ones to share research in the conversation around the German repository search.

4 Conclusion and future work

We set out to characterize the Twitter conversation around the search for a German nuclear repository by analyzing a sample of tweets collected over approximately twelve weeks both quantitatively and qualitatively. In addition, we examined whether the different participants of said conversation would link to research and if so, in which form.

We found the conversation to be dominated by relatively few very active accounts, while most participating accounts only contributed few tweets, which is typical for most online conversations (Haustein, 2019). Most activity proceeded from accounts primarily devoted to activism or initiatives, while the follower-wise most influential accounts participating in the conversation were journalistic. Results from both our automatic as well as our manual search for research mentions suggest that direct links to scientific publications happened very rarely. We did however find evidence of several less direct ways of users referencing research, e.g., by linking to compositions based on that research in news articles, on blogs, or in other tweets. Most links to paraphrases of studies came from science accounts, followed by accounts of activists/initiatives. Summaries being a typical format for communicating about research on Twitter is an observation in line with findings by Thelwall et al. (2013), who made similar observations for a sample of tweets containing links to articles from high-profile journals and digital libraries. The prevalence of indirect references (or 'second-order citations') on the other hand was also found by Priem and Costello (2010) when analyzing researchers' reference behavior on Twitter. However, while Priem and Costello found roughly equal numbers of occurrences of first- and second-order citations in their sample, our observations suggest that the ratio would be leaning much stronger towards second-order citations in our case.

One intention of our study was to get an impression of Twitter's suitability as a platform to achieve insights about the participative and science-based

shaping of the selection process behind the German search for a nuclear repository. Regarding this question, our findings go both ways: First, the overall diversity of represented roles as well as the high degree of activity coming from activists and initiatives indicates that the Twitter conversation around the topic did not take place in an ‘ivory tower’ of academics and technocrats, but to an extensive degree also involved committed citizens. On the other hand, we could not observe much of an exchange around scientific findings and ideas – references to research were overall rather scarce, with most (indirect) references in our coded sample going to the same academic study. The validity of this interpretation should be checked by future studies, however. Such studies could for example apply methods of social network analysis (see for instance Alperin et al., 2019) to achieve more accurate descriptions of the pathways which research takes through the Twitter conversation examined in this study.

Our findings are also interesting from the perspective of altmetrics research: They indicate that the lion’s share of signals indicating the influence of scholarly articles on Twitter will most probably be missed with automated data collection approaches, as long as these approaches do not resolve links within tweets and process the contents behind them. In our case study, especially news articles proved to be popular as intermediaries, which were linked to more frequently as sources for scientific findings than the respective original studies. These findings raise the question whether news mentions, which also exist as an altmetric indicator, might possibly be better at capturing the attention scientific articles likely received on social media than the metrics derived directly from the social media platforms themselves, as in our case Twitter.

Our study has some limitations. First, while the manual coding of tweets and accounts enables high control over the exact parameters to analyze, the effort related to this approach limits it to comparatively small sample sizes, restricting the generalizability of our findings. In our case, we might just have been ‘unlucky’ with our random sample of 250 coded tweets and therefore underestimate the actual amount of direct references made in the conversations around the repository search (the opposite could be true of course as well). Also, the manual categorization of accounts relied to a large extent on the profile information provided by their owners, for which checks for correctness are often virtually impossible. Moreover, as our user profile analysis focused on the head of what appears to be a power law distribution of tweets across users, it might be interesting in the future to additionally compare our

respective findings with those of an analysis of a similarly sized subsample of users from the distribution's tail.

Furthermore, the debate on nuclear energy has shown to vary greatly from country to country (Arlt et al., 2018). In the USA, for example, unlike in Germany, nuclear energy is regarded as a climate-friendly form of energy generation and is being expanded. This in turn could also have an influence on whether and to what kind of scientific publications tweets refer. Since only German-language tweets were considered in this study, it can be assumed that the results cannot necessarily be transferred to the discourse in other countries.

Finally, the choice of the time of collection of tweets also influences the content of the collected tweets. By focusing on the date of announcement of possible sites for a repository, the discourse might have been much more focused on topics like geology or fairness. A more detailed analysis of the tweets' content and possibly also a comparison with a different time period could shed light on these issues. And lastly, it is of course possible that Twitter simply is not the platform on which the actual conversation about the German repository search takes place.

Acknowledgements

This study is a joint outcome of the research projects *MeWiKo*, funded by the German Federal Ministry of Education and Research (grant number 01PU17018A), and *TRANSENS*, funded by the Germany Federal Ministry of Economics and Energy, Volkswagen Foundation, Lower Saxony Ministry for Science and Culture (grant number 02E11849A-J).

References

- Alperin, J. P., Gomez, C. J., & Haustein, S. (2019). Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles. *Public Understanding of Science*, 28(1), 2–18. <https://doi.org/10.1177/0963662518761733>
- Arlt, D., Rauchfleisch, A., & Schäfer, M. S. (2018). Between Fragmentation and Dialogue. Twitter Communities and Political Debate About the Swiss “Nuclear

- Withdrawal Initiative.” *Environmental Communication*, 13(4), 440–456. <https://doi.org/10.1080/17524032.2018.1430600>
- Bartling, S., & Friesike, S. (Eds.) (2014). *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-00026-8>
- Gupta, K., Ripberger, J., & Wehde, W. (2018). Advocacy Group Messaging on Social Media: Using the Narrative Policy Framework to Study Twitter Messages about Nuclear Energy Policy in the United States. *Policy Studies Journal*, 46(1), 119–136. <https://doi.org/10.1111/psj.12176>
- Hadgu, A. T., & Jäschke, R. (2014). Identifying and analyzing researchers on twitter. In *WebSci '14: Proceedings of the 2014 ACM Conference on Web Science, June 2014* (pp. 23–32). New York, NY: ACM Press. <https://doi.org/10.1145/2615569.2615676>
- Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review*, 2(1), 39–53.
- Haustein, S. (2019). Scholarly Twitter Metrics. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 729–760). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_28
- Haustein, S., & Costas, R. (2015). Identifying Twitter audiences: Who is tweeting about scientific papers? Presented at *ASIS&T SIG/MET Metrics 2015 Workshop*. St. Louis, USA, November 7, 2015. <https://www.asist.org/sig/sigmet/events/past-sig-met-workshops/>
- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656–669. <https://doi.org/10.1002/asi.23101>
- Holmberg, K., Bowman, T. D., Haustein, S., & Peters, I. (2014). Astrophysicists’ Conversational Connections on Twitter. *PLOS ONE*, 9(8), e106086. <https://doi.org/10.1371/journal.pone.0106086>
- Holmberg, K., & Hellsten, I. (2016). Twitter Campaigns Around the Fifth IPCC Report: Campaign Spreading, Shared Hashtags, and Separate Communities. *SAGE Open*, 6(3), 2158244016659117. <https://doi.org/10.1177/2158244016659117>
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the Joint 9th WEB-KDD and 1st SNA-KDD Workshop 2007*. Berlin, Heidelberg: Springer. https://ebiquity.umbc.edu/file_directory/papers/369.pdf

- Ke, Q., Ahn, Y.-Y., & Sugimoto, C. R. (2017). A systematic identification and analysis of scientists on Twitter. *PLOS ONE*, 12(4), e0175368. <https://doi.org/10.1371/journal.pone.0175368>
- Kim, D. S., & Kim, J. (2014). Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter. *International Journal of Multimedia and Ubiquitous Engineering*, 9(11), 373–384. <https://doi.org/10.14257/ijmue.2014.9.11.36>
- LaPoe, V. L., Carter Olson, C., & Eckert, S. (2017). “Linkedin Is My Office; Facebook My Living Room, Twitter the Neighborhood Bar”: Media Scholars’ Liminal Use of Social Media for Peer and Public Communication. *Journal of Communication Inquiry*, 41(3). <https://doi.org/10.1177/0196859917707741>
- Lenke, S., & Peters, I. (2019). Coping with Altmetrics’ Heterogeneity—A Survey on Social Media Platforms’ Usage Purposes and Target Groups for Researchers. In *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics*, 2, 2320–2325. Rome, Italy: Edizioni Efesto.
- Liu, Z., & Na, J.-C. (2018). Aspect-Based Sentiment Analysis of Nuclear Energy Tweets with Attentive Deep Neural Network. In M. Dobрева, A. Hinze, & M. Žumer (Eds.), *Maturity and Innovation in Digital Libraries* (pp. 99–111). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-04257-8_9
- Moscrop, D., Wong, L., & Alperin, J. P. (2020). Have You Seen This? Why Political Pundits Share Scholarly Research on Social Media. *Scholarly and Research Communication*, 11(1). <https://doi.org/10.22230/src.2020v11n1a355>
- Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group I Report. *PLOS ONE*, 9(4), e94785. <https://doi.org/10.1371/journal.pone.0094785>
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1). <https://doi.org/10.1002/meet.14504701201>
- Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, 12(8), e0183551. <https://doi.org/10.1371/journal.pone.0183551>
- Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, 24(3), 183–195. <https://doi.org/10.1087/20110306>
- Schmitt, M., & Jäschke, R. (2017). What do computer scientists tweet? Analyzing the link-sharing practice on Twitter. *PLOS ONE*, 12(6), e0179630. <https://doi.org/10.1371/journal.pone.0179630>

- Sovacool, B. K., Schmid, P., Stirling, A., Walter, G., & MacKerron, G. (2020). Differences in carbon emissions reduction between countries pursuing renewable electricity versus nuclear power. *Nature Energy*, 5(11), 928–935. <https://doi.org/10.1038/s41560-020-00696-3>
- Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, 68(9), 2037–2062. <https://doi.org/10.1002/asi.23833>
- Syn, S. Y., & Oh, S. (2015). Why do social network site users share information on Facebook and Twitter? *Journal of Information Science*, 41(5), 553–569. <https://doi.org/10.1177/01655515155585717>
- Tenopir, C., Volentine, R., & King, D. W. (2013). Social media and scholarly reading. *Online Information Review*, 37(2), 193–216. <https://doi.org/10.1108/OIR-04-2012-0062>
- Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting Links to Academic Articles. *Cybermetrics*, 17, 1–8.
- Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? In *Proceedings of the 2015 International Conference on Scientometrics and Informetrics*. Presented at the ISSI2015, Istanbul, Turkey, June 29–July 04, 2015.
- Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 112(1), 345–366. <https://doi.org/10.1007/s11192-017-2368-0>
- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512(7513). <https://doi.org/10.1038/512126a>
- Wouters, P., & Costas, R. (2012). *Users, Narcissism and the Control: Tracking the Impact of Scholarly Publications in the 21st Century*. Utrecht: Stichting Surf. <http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 94–114. DOI: doi.org/10.5283/epub.44940.

Open Practices of Early Career Researchers

A Qualitative Study on Research and Teaching Behavior

Tamara Heck

Information Center for Education
DIPF | Leibniz Institute for Research
and Information in Education,
Germany
Heck@dipf.de

Ina Blümel

German National Library of Science
and Technology | Technische Infor-
mationsbibliothek (TIB) and
Hochschule Hannover | University of
Applied Sciences and Arts
Ina.Bluemel@tib.eu

Abstract

Many researchers have a positive attitude towards open science and are motivated to apply them. However, applying them requires a change in one's daily practices. Different factors might challenge a behavioral change. The introduced study wants to get deeper insights into the reasons and influences that lead early career researchers to apply open practices in their daily research and teaching work. The participatory design let ten participants choose open practices they wanted to learn and adapt in either research or teaching scenarios. The study accompanied them and collected their positive and challenging experiences via diverse methods like interviews, diary entries and workshops. This paper introduces the study design and preliminary results.

Keywords: open science practices; early career researchers; research behavior; user study

1 Problem and research question

Open science practices are the visible impact of the open science movement, which aims at opening up research and educational processes. As those practices are part of a researchers' work, many studies investigate researchers' attitudes towards open practices and their actual behavior. Topics investigated are data sharing and open data (Scherp, Siegfried, Biesenbender, & Breu-

er, 2020; Ünal, Chowdhury, Kurbanoglu, Boustany, & Walton, 2019), changes in open access publications (Bosman & Kramer, 2018; Piwowar et al., 2018), the influence of policies, e.g., in open education (Bossu & Stagg, 2018), and the influence of research communities and one's personality (Kim & Nah, 2018; Linek, Fecher, Friesike, & Hebing, 2017).

Research shows that many researchers have a positive attitude towards open science (Kramer & Bosman, 2016), although the understanding of openness differs in detail (Levin, Leonelli, Weckowska, Castle, & Dupré, 2016). However, it seems that open science practices are not yet adapted by many researchers. This discrepancy motivated the following study. Our assumption is that although researchers are motivated to apply open practices, they face challenges in adapting them to current research behavior.

This study investigates potentials and challenges of open science practices that early career researchers experience within their research and teaching in higher education. The participatory approach aimed at letting participants choose open practices they want to learn and apply, and afterward let them report on their experiences with those practices. Moreover, the study did not intend to set up an experimental design, but wanted open practices to be applied within the daily work of participants.

The following research questions will be answered in this contribution:

- Which open practices did early career researchers choose to test for which reasons?
- Which experiences did they make?

Section 2 introduces the method and study design. Section 3 discusses preliminary results and study limitations.

2 Method

2.1 Design

The study applied a participatory qualitative design and aimed to accompany participants for at least six months. The experience of open science practices was left to the participants. The participants chose practices related to aspects of openness in research and teaching, which they wanted to learn and explicitly apply within their current work. The design included an onboarding

workshop where participants chose their open practices. After this workshop, participants were prompted to write diary notes to tell about their progress of applying the new practices and personal experiences they gained with them. After this phase, participants joined a final workshop to discuss experiences among all participants. Due to Covid-19, this final workshop was held online.



Fig. 1 User study approach and time line of processes

2.2 Target group

The target group consisted of early career researchers from educational research, i. e., PhD candidates or Postdocs with a low academic age, professors

were excluded. Educational research is an interdisciplinary field, where dynamics and influences of open practices within diverse sub-fields like psychology, sociology and pedagogical research can be investigated and compared. The behavior of early career researchers is worth investigating as they are an important driver for establishing new research practices within a community. On the one hand they are motivated to become part of an established research community and gain scientific reputation. Thus, they might adopt traditional research practices that are not yet open practice. On the other hand, early career researchers are aware of digital technologies and possible open practices and might be willing to apply those (Tenopir et al., 2016). This group is in a challenging position to decide whether to apply accepted practices or change their behavior.

2.3 Adaptations made

In the first participant recruitment phase in spring 2019, we found five researchers who agreed to participate. As we aimed at having more participants, we started a second recruitment phase in autumn 2019, where we found five additional participants. Not all researchers stayed until the end of the study and engaged in diary notes. Participants were asked to engage during several months, and the experiences we got back were not satisfactory due to different reasons (participants said they had not time or forgot to make a diary entry). Thus, we slightly adapted data generation. We decided to do another interview with each participant and let them talk about their experiences, i.e., we did open interviews at the end of the study guided by the leading diary entry questions.

One participant did not answer any requests after the first interview and left the study in the second phase. Two other participants stepped out of the study when their jobs and work tasks changed and they were not able to engage in the chosen practices. Seven participants finished the study. The main data analysis is based on seven participants, although we included the interview and questionnaire data from all participants. The full process of the study design is shown in Figure 1.

2.4 Limitations

The qualitative study gives deeper insights, but is not representative for early career educational researchers in a discipline. Participants were researchers

with a positive attitude of and willingness to learn about open science. Thus, the study cannot contribute any reasons for not trying to apply open practices. Choosing diary entries seemed appropriate to directly let participants reflect on experiences. However, to get more entries and feedback, more explicit prompting would have been needed.

3 Results and discussion

Preliminary results of the first phase participants were summarized in an earlier paper (Heck & Brimiouille, 2019). The discussion of the second workshop, attended by five participants and two project team members ended in a summary report that discusses some relevant factors of open practices in research and teaching (Heck et al., 2020).

Table 1 shows all participants' demographics. All participants worked either at a German university or a research institution. Only one participant did her PhD as external candidate and worked part-time for a company. She opted out of the study when she changed jobs.

Table 1: Demographics of participants

#	Gender	Degree	Year of PhD degree/ start of PhD	Field of research
1	f	Postdoc	2019	education
2	m	Postdoc	2018	math didactics
3	m	Postdoc	2016	school education
4	f	PhD candidate	2019	education
5	f	Postdoc	2018	education
6	m	PhD candidate	2018	media education
7	f	Postdoc	2014	special needs education
8	f	PhD candidate	2013	education
9	f	Postdoc	2008	German philology
10	f	PhD candidate	2014	social sciences

Table 2 shows the answers from a short questionnaire on facets of open science and participant's awareness and knowledge of them. The answers

show a high interest in open educational practices, which is reflected in the scenarios the participants chose to apply.

Table 2: Participant answers on open science facets

Open science facets	Answers by participants ($n = 8$)		
	I heard of ...	I have experience with ...	I'd like to know about ...
open access	5	4	3
open data	4	2	5
open source	2	4	5
open peer review	2	0	7
open methodology	1	1	7
citizen science	5	1	2
open educational resources	3	5	4
open educational practices	4	0	8

The personal experiences of the participants are quite diverse – except one participant, who had not heard about open science before, the other participants heard about open science via colleagues or on workshops. The ideas of open practices differ slightly (compare Levin et al., 2016), e.g., one participant stated that open practices means not being predefined by theories but instead being open to different research approaches.

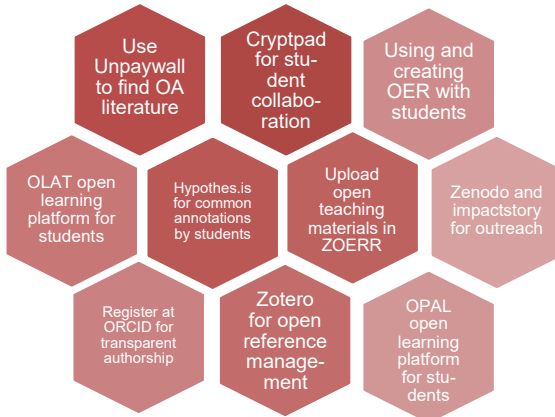


Fig. 2 Selection of open practices by participants

In most cases, participants heard of single aspects of open science like shown in Table 1. One reason is their different working tasks. For example, researchers, who are concerned with quantitative studies that generate research data reported to have heard about pre-registration and data sharing. Other researchers are in a phase where they focus on teaching and wanted to learn more about openness in education. Moreover, we could detect great differences in the participants' knowledge and use of digital tools to make resource openly available or allow collaborative editing. Two participants have more profound knowledge and engaged in open science projects, like creating open educational resources.

Open practices were discussed within the workshops and participants got shown examples on how to apply those practices with exemplary tools and services (see report on Wikiversity)¹. Each participant chose two to five open practices and Figure 2 shows a selection of those.

First results on the motivation and goals of adapting open tools and first experiences given in the diary entries during the study are summarized in Table 3. The reasons for testing open practices show two main aspects. Open driver is facilitating working and collaboration. The other driver is more ideological, e.g., participants want to apply more inclusive tools.

Table 3: Motivation and first experiences of participants for using open tools

Motivation and goals for participants to replace proprietary tools and use open online tools like pads and literature reference tool	Experiences
be more independent	strength: open tools allow location independency and are free of costs
allow for better outreach and transparency of work flows	constraint: technical difficulties with tools → lack of ease of use
improve student communication and train student to articulate themselves	challenge: students ask for more training on tools
improve collaboration among colleagues	opportunity: high acceptance among colleagues

Participants stated that the chosen digital tools were appropriate for their purpose. One participant introduced Zotero to his colleagues and reported that after a first reluctance, they accepted the new tool to share references

1 <https://de.wikiversity.org/wiki/OPER>

more easily. However, introducing online spaces for students to communicate and write collaboratively, like online pads or wikis, was more challenging in practice, as students appreciated the idea, but asked for more support to use the tools more effectively. Technical difficulties like unavailability of a service tool, are still common and narrowed the motivation to apply the open tools. Moreover, the lack of technical infrastructure was difficult: One participant gave a seminar and wanted her students to apply digital tools to commonly and synchronously write and edit texts. However, her lecture room only had one plug socket and “in the late afternoon the students’ notebooks went out of power”.

Besides those practical technical and infrastructural challenges, participant admitted that practices like applying new tools with students are more time consuming. Moreover, applying new practices need a commitment to test and implement them, like experiences with colleagues show. Convincing colleagues to rethink their practices is difficult as currently incentives are missing. However, one participant reports that he got positive feedback from his colleagues as open science expert and enables of new practices. He sees a main challenge in getting federates in one’s direct environment, who support each other in applying open practices.

Being asked about further adopting the new open practices tested after the project phase, participants answered differently with regard to the specific practices. It got obvious that practices they can adapt on their own like register one’s ORCID are easier to keep, whereas any practices where colleagues or students are involved are more complex. Applying open practices depends on the concrete context and situation and might change within different teams, research projects, and especially within education where lecturers teach diverse courses with different study levels like Bachelor and Master. Furthermore, participants stated that they are responsible for work tasks, for which they do not need open practices or open practices known are just not adaptable for those tasks. This reason was as well mentioned by the participants, who opted out of the study. Their work tasks changed and they just did not know, how they could apply any open practices in the new context. It stresses the relevant factor that open practices need to be valuable and bring a concrete benefit in a work task, like facilitate reference sharing or common text editing. The remaining participants stated they are now aware of options to apply open practices and will try to implement them as often as they can, internalizing the motto “as much openness as possible” (Heck et al., 2020).

4 Conclusion

The first conclusions of the project are:

- a) RQ1: Changing one's behavior comes with small steps. Participants chose to apply small practices like single tools in specific contexts. The reasons show that participants wanted to either facilitate specific tasks in research and education, or apply practices like using open source tools because they are free of costs and more inclusive.
- b) RQ2: Overall, open practices are being experienced as useful to improve tasks in research and foster student activities. The context and benefit for all people involved are crucial to apply them. There are still technical challenges and reluctance of other people that narrow the motivation to fully apply the new practices in the future.

In the future, we are planning a comprehensive evaluation of the results with a special focus on the potentials and challenges of open practices among early career researchers with regard to their disciplines and communities.

References

- Bosman, J., & Kramer, B. (2018). Open access levels: a quantitative exploration using Web of Science and oaDOI data. Advance online publication. *PeerJ Preprints* 6:e3520v1. <https://doi.org/10.7287/peerj.preprints.3520v1>
- Bossu, C., & Stagg, A. (2018). The potential role of Open Educational Practice policy in transforming Australian higher education. *Open Praxis*, 10(2), 145. <https://doi.org/10.5944/openpraxis.10.2.835>
- Heck, T., Blümel, I., Fahrner, S., Lohner, D., Schneider, J., & Visser, S. (2020). Open practice in science and education – a discussion with researchers and educators who tested to be open. *Zenodo*. <https://doi.org/10.5281/zenodo.3746018>
- Heck, T., & Brimiouille, P. (2019). Experiencing open practices – a qualitative long-term study among early career educational researchers. Presented at the Open Practices in Education (OPINE) – Research Symposium, Frankfurt a.M., Germany. *Zenodo*. <https://doi.org/10.5281/ZENODO.3532740>
- Kim, Y., & Nah, S. (2018). Internet researchers' data sharing behaviors. *Online Information Review*, 42(1), 124–142. <https://doi.org/10.1108/OIR-10-2016-0313>

- Kramer, B., & Bosman, J. (2016). Innovations in scholarly communication – global survey on research tool usage. *F1000Research*, 5, 692. <https://doi.org/10.12688/f1000research.8414.1>
- Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice. *Bulletin of Science, Technology & Society*, 36(2), 128–141. <https://doi.org/10.1177/02704676166668760>
- Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLOS ONE*, 12(8), e0183216. <https://doi.org/10.1371/journal.pone.0183216>
- Piwowar, H., Priem, J., Lariviere, V., Alperin, J. P., Matthias, L., Norlander, B., ... Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PEERJ*, 6. <https://doi.org/10.7717/peerj.4375>
- Scherp, G., Siegfried, D., Biesenbender, K., & Breuer, C. (2020). Die Bedeutung von Open Science in den Wirtschaftswissenschaften. Ergebnisbericht einer Online-Befragung unter Forschenden der Wirtschaftswissenschaften an deutschen Hochschulen 2019. BW – Leibniz Information Centre for Economics, Kiel/Hamburg. <http://hdl.handle.net/11108/448>
- Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R., ... Watkinson, A. (2016). Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire. *Journal of the Association for Information Science and Technology*, 67(10), 2344–2361. <https://doi.org/10.1002/asi.23598>
- Ünal, Y., Chowdhury, G., Kurbanoglu, S., Boustany, J., & Walton, G. (2019). Research data management and data sharing behaviour of university researchers. *Information Research: An International Electronic Journal*, 24(1).

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 115–124. DOI: doi.org/10.5283/epub.44941.

Information Behavior towards False Information and “Fake News” on Facebook

The Influence of Gender, User Type and Trust in Social Media

Thomas Schmidt

Media Informatics
Group, University of
Regensburg, Germany

thomas.schmidt@ur.de

Elisabeth Salomon

Information Science,
University of Regensburg,
Germany

elisabeth_salomon@gmx.de

David Elswailer

Information Science,
University of Regens-
burg, Germany

david.elsweiler@ur.de

Christian Wolff

Media Informatics Group,
University of Regensburg,
Germany

christian.wolff@ur.de

Abstract

In this paper, we present a survey study with 119 participants conducted in German, which investigates respondents' Facebook behavior. In particular, the survey provides insight into how the individual factors gender, user type and trust in social media influence information behavior with respect to false information on Facebook. Our participants' Facebook use is predominantly passive, the trust in social media is mediocre and most users claim to encounter false information on a weekly basis. If the truthfulness of information is verified it is mostly done by checking alternative sources and for the most part, users do not react actively to false information on Facebook. Of the different categories of Facebook users studied, more active and intensive users of Facebook (posters and heavy users) encounter false information the most. These users are the only user group to report posts with false information to Facebook or interact with the post. Participants with higher trust in social media tend to check the comments of a post to verify information.

Keywords: false information; fake news; social media; gender; user type; trust in social media

1 Introduction

Social media is becoming an ever more important part of the web. 65% of all American adults use social media (Perrin, 2015) and although the influence of social media is lower in Germany by comparison, it remains substantial with 48% of all web users in Germany reporting at least occasional social media use, with the figure increasing to 81% if only younger users are considered (Poushter, 2016). In light of these high usage numbers, substantial research literature exists dealing with how and why social media is used (e.g., Lee & Ma, 2012; Park, Kee, & Valenzuela, 2009). Social media has also been subject to frequent criticism. The World Economic Forum labelled the propagation of false information on social media as one of the top ten trends as early as 2014. Research in information science has examined this subject, for example, in the contexts of particular events, such as the 2010 earthquake in Chile (Mendoza, Poblette, & Castillo, 2010) and the Boston Marathon Bombing (Starbird et al., 2014). Recently, fake news behavior during the Covid-19 pandemic has been studied (Montesi, 2020) and the rise of this phenomena has led to various research in information science (cf. Chiluwu & Samoilenko, 2019). An active research topic in the area of false information is the prediction of the credibility of information using data mining methods (Fletcher, Schifferes, & Thurman, 2020). The source of false information – man or algorithm – has drawn some research interest, as well (Zimmer et al., 2019a, 2019b).

With the user-centered turn (Dervin & Nilan, 1986), personal and individual factors are regarded as very important to achieve a holistic view of a subject in information science. The individual factors personality and gender have been verified as important influencing factors in information seeking (Schmidt & Wolff, 2015, 2016; Schmidt, 2016), user interface aesthetics (Schmidt & Wolff, 2017, 2018), usability testing (Schmidt, Wittmann, & Wolff, 2019) and general social media behavior (Correa, Hinsley, & De Zuniga, 2010), however very little work has investigated how these factors interact with false information. Marret's and Joshi's (2009) work goes in this direction by analyzing correlations of the motivation and the normative influence on the propagation of false rumors in online forums. Chen and Sin (2013) examined the influence of gender and personality on the sharing of false information on the web. Although research on false information propagation on the web is limited, the results published so far hint that individual

factors are important and warrant further investigation. As individual factors, the study presented here examines gender, user type and the degree of trust in social media. Although gender could not be proven as relevant influencing factor in recent research, further analysis is justified since gender has been verified as important influencing factor in several research areas in information behavior (Venkatesh & Morris, 2000; Schmidt & Wolff, 2015). We referred to the distinction between lurkers and posters, as well as heavy and light users to analyze user type and information behavior towards false information in the context of social media. The final individual factor is the “trust in social media”. Johnson and Williams (2010) discovered that social media is the media channel that is trusted least. Jakob (2010) showed that the trust in social media influences how the social media platform is used. These results lead us to believe that trust in social media can also be an influencing factor on handling false information in social media. Further, the focus of current research relies heavily on the active act of false information propagation (Chen & Sin, 2013; Chen et al., 2015). The passive perception of false information is also an important part of information behavior towards false information and was examined closer in our study. Furthermore, post-perception behavior has not yet been analyzed. Our study tries to close this gap by also investigating the reaction and verification behavior after encountering false information. We also integrate the concept of mis- and disinformation in our research questions. Misinformation is false information that is shared without knowledge about the falsehood. Disinformation is shared with deceptive intention and knowledge about the falsehood (Karlova & Fisher, 2013). This taxonomy can lead to more differentiated results. We chose to study the social media platform Facebook and used a survey to learn about user behavior and perceptions. By analyzing known factors as well as unknown nuances, we gained new insights for the research area of false information.

2 Related work

Although research about individual differences in social media is still rare, information science has dealt with the general phenomenon of social media for some time. For example, theoretical works have formulated definitions and taxonomies (Ellison, 2007; Kaplan & Haenlein, 2012). Ellison (2007)

defines a social network as a web service and interconnected system one can create a profile and user lists as well as communicate and share information. The social network used in this study, *Facebook*, is a prototypical example. Empirical work deals with the question of who uses the social network (Lenhart, 2009), how much time is spent doing so (ibid.; Raacke & Bonds-Raacke, 2008) and in which way social media is used (Lenhart, 2009). These and similar factors have also been analyzed in the context of group-specific influences. Park, Kee, and Valenzuela (2009) examined the reasons for participation in Facebook-groups among college students. The primary reasons are socialization, entertainment, self-promotion and gathering information. Women living outside the city of the university, however, were found to use the groups significantly more for gathering information than men. Correa, Hinsley, and De Zúñiga (2010) showed that people who are open for new experiences, as well as more extraverted people use social media more frequently. Emotional stability, on the other hand, is a rather negative predictor for social media usage. The influence of personality on social media use interacts with gender and age, with the relationship being especially strong for young, extraverted users. However, men with lower emotional stability use more social media. Lee and Ma (2012) examined explicitly sharing information in social networks. They showed that persons that use social networks primarily for information seeking, socialization and self-promotion also tend to share news more than users with other motivations. Similar to our study the individual intensity of active participation in the social network was operationalized and it was proven that this intensity correlates with the intention to share news.

Regarding the concept of false information, theoretical as well as empirical research in information science can be found. Theoretical works deal with the definition and classification of false information (Fallis, 2009; Alexander & Smith, 2010; Karlova & Lee, 2011). Instead of referring to these sophisticated reflections, we refer to a simpler but gainful definition by Karlova and Lee (2013), who divide false information into misinformation (false information shared without knowledge about the falsehood) and disinformation (false information shared with deceptive intentions).

Empirical research on false information in social networks has typically focused on isolated catastrophic events. Mendoza, Poblete, and Castillo (2010) analyzed the information propagation on Twitter during the earthquake in Chile. They notice that mis- and disinformation is indeed shared but can be identified because it is less intensively shared and the truthfulness is

doubted in retweets. Starbird et al. (2014) performed a similar study concerning the Boston Marathon Bombing. Again, the creation and propagation of mis- and disinformation was shown. They identified crowd-based correction procedures; however, these corrections were less effective than the propagation intensity and speed of the false information propagation. Leeder (2019) presents a study of college students’ fake news handling behavior. They identified that students were unable to evaluate their own skills in detecting fake news but found correlations between the identification of fake news and specific critical evaluation strategies.

Research about the influence of individual and group-specific factors on the propagation of false information in the online-context is rare so far. Marret and Joshi (2009) examined the propagation of rumors and information in online-forums about sports. They analyzed several individual factors like the user type by differing between posters (active participants of the forum that post and answer frequently) and lurkers (passive user that mostly just read the online forum). Moreover, extrinsic and intrinsic motivations, as well as normative influence were studied. In a survey study with 471 participants, they were able to show that the motivation to share information and rumors differ between posters and lurkers. Posters are influenced by all three motivational factors while lurkers are mainly influenced by extrinsic motivation and normative influence. Marret and Joshi show that the user type can have an impact on handling false information and the analysis of this factor can lead to more sophisticated results in this research area. This idea is picked up in our study. Chen and Sin (2013) examined explicitly the influence of person-related factors on the motivation to share misinformation. As person-related factors they chose gender and personality, operationalized with the know Big-5-model. They conducted a survey study with 171 college students. More than two thirds of the participants affirm to having shared false information. Gender did not show any relevant results for these relationships. Regarding personality, extraverted participants tend to share misinformation more likely to socialize with others.

3 Research questions

In this section, we formulate our main research question. We focus our research solely on the platform Facebook. We motivate this decision by the

fact that Facebook remains by far the most popular social network in Germany where the study was conducted.¹ The research questions are structured by the independent variables: the individual factors

- gender,
- user type,
- and trust in social media.

We further refer to the user type by the more precise term *Facebook-usage*. Facebook-usage is on the one hand operationalized by the intensity and frequency of the active and passive Facebook-usage (“heavy user” and “light user”) – we will refer to this concept as *Facebook-intensity*. On the other hand, we distinguish between active participation (posters) and more passive reception (lurkers) – this concept will be named *Facebook-user-type*.

Trust in social media is operationalized by various statements. The higher the trust in social media the more informative, accurate, truthful, reliable and essential social media is regarded.

Our dependent variables are the passive and active handling of false information, the estimation of the proportion of mis- and disinformation and the verification- and reaction behavior when confronted with false information. Handling false information is operationalized analogue to Facebook-usage by differentiating between active and passive interactions as well as by regarding the frequency of consumption and interaction, this time with false information.

To integrate the taxonomy of mis- and disinformation we also collect data about the assessment of the proportion of both types to false information. In doing so, we want to gain first insights from the perspective of users concerning this taxonomy.

The verification behavior is divided in different possible verification actions possible on the web to verify the truthfulness of suspected false information. The reaction behavior is also divided in different actions one can perform on Facebook to react to false information. All factors are operationalized by single or multiple questionnaire items. More details on the operationalization follows in the next sections. To get a holistic view, interactions between the independent variables were analyzed as well. We structure the research questions by the primary independent variables into the categories G (gender), F (Facebook-usage) and T (trust in social media). Each research question consists of one or multiple hypotheses depending on the various

1 <https://gs.statcounter.com/social-media-stats/all/germany>

variables/questionnaire items that are examined via this research question. Please note that we also take a look at interactions between the independent variables.

Gender:

- G-RQ1: There is a relationship between gender and Facebook-usage.
- G-RQ2: There is a relationship between gender and trust in social media.
- G-RQ3: There is a relationship between gender and handling false information.
- G-RQ4: There is a relationship between gender and the estimation of the proportion of mis- and disinformation.
- G-RQ5: There is a relationship between gender and the verification behavior.
- G-RQ6: There is a relationship between gender and the reaction behavior towards false information.

Facebook-usage:

- F-RQ7: There is a relationship between Facebook-usage and trust in social media.
- F-RQ8: There is a relationship between Facebook-usage and handling false information.
- F-RQ9: There is a relationship between Facebook-usage and the estimation of the proportion of mis- and disinformation.
- F-RQ10: There is a relationship between Facebook-usage and the verification behavior.
- F-RQ11: There is a relationship between Facebook-usage and the reaction behavior towards false information.

Trust in social media:

- T-RQ12: There is a relationship between trust in social media and handling false information.
- T-RQ13: There is a relationship between trust in social media and the estimation of the proportion of mis- and disinformation.
- T-RQ14: There is a relationship between trust in social media and the verification behavior.
- T-RQ15: There is a relationship between trust in social media and the reaction behavior towards false information.

4 Methods

To answer the hypotheses of the research questions a survey study was conducted.

4.1 Questionnaire development and participant acquisition

The questionnaire is divided in several parts consisting of self-made questions as well as questions and statements oriented towards questionnaires of recent research. The survey was created with *Google Forms*. We performed a pre-test with five students of information science to test the structure and overall understanding of the questionnaire in two iterations of the survey development. Based on this feedback we improved upon the formulation of some questions and the structure of the questionnaire.

We gathered participants by posting the survey in multiple German Facebook-groups that are focused on the acquisition of participants for online questionnaires. Since these groups are mainly directed towards students performing similar studies, these make up the majority of our sample (see Chapter 5.3.1 for more information about the demographics). The questionnaire was online for one month.

4.2 Questionnaire structure

We describe the structure and content of our questionnaire. An anonymized version of the original German questionnaire as well as an English translation can be found online.² Overall, the questionnaire had 42 items including demographic questions. For an overview of the final important variables see Tables 1 and 2.

4.2.1 Demography

In the first part of the questionnaire, basic demographic data is gathered. We collected information on gender, age, education level and profession.

2 <https://www.dropbox.com/sh/oum3i6gvbyp4s6u/AAAvwyV6o7Nwbo-DyMpb413a>

4.2.2 *Facebook-usage*

Rosen et al. (2013) developed a questionnaire to capture daily media usage. By carrying out a factor analysis with 942 participants, they were able to identify eleven subscales. One of these subscales deals with the usage of social media. The subscales show a high level of reliability and validity. We chose the “General social media usage subscale” and modified it for our purposes. The items were translated into German. The questions assess frequency concerning several social media activities: calling the Facebook-page (in general, from the smartphone, from work), updating the status, posting photos, browsing through profiles, reading posts, commenting posts, liking posts. These nine questions were extended by a self-formulated question about the sharing of information (e.g., news) because this activity is important in relationship to information behavior towards false information. The frequency of the individual activities is answered by choosing one of eleven selections: never (1) / less than once a month (2) / once a month (3) / multiple times a month (4) / once a week (5) / multiple times a week (6) / once a day (7) / multiple times a day (11) / once per hour (9) / multiple times per hour (10) / all the time (1) (see Rosen et al., 2013). Like Rosen et al. we transform the answers to ordinal numerical values ranging from 1 (never) to 11 (all the time). The classification in lurkers and posters is explained in Chapter 5.1.

4.2.3 *Trust in social media*

To operationalize trust in social media we developed a questionnaire section that is oriented towards the “Scale to Measure Consumer Skepticism Toward Advertising” (SKEP; Obermiller & Spangenberg, 1998). Although the subject of the questionnaire has no direct relationship with the goal of the present study, it was possible to reformulate the items to the context of trust in social media. The questions were translated into German. Three questions were removed since they were not suitable for our use case. Overall, five statements are formulated in the form of: “social media is an informative / truthful / accurate / reliable / essential source of information”. On a 5-point Likert-scale participants could express their approval to the statement from 1 (disagree fully) to 5 (agree fully). By summing the numerical values of all items, we get a metric variable ranging from 5 to 30. The higher this value the higher the trust in social media.

4.2.4 *Handling false information*

To gather data concerning handling false information we designed a subsection of the survey that uses activities from the Facebook-usage section and reformulates them with false information. We asked how often participants believe to consume, share, “like” or comment false information. As explained in Chapter 4.2.2, items were answered on a scale from 1 (never) to 11 (all the time).

4.2.5 *Proportion of mis- and disinformation*

To integrate the concept of mis- and disinformation, participants were asked to estimate the proportion of mis- and disinformation to false information on Facebook. Since these terms can be very complex, we presented an explanation of the terms in this subsection of the questionnaire. The proportion of each type was assessed on a 5-point scale ranging from 1 (very low) to 5 (very high). We wanted to gather user-based data concerning these concepts but did not integrate other questions to avoid possible problems in understanding the definitions.

4.2.6 *Verification of information*

To obtain data about verification behavior when confronted with information assumed false on Facebook, we adopted and adjusted a questionnaire developed by Flanagin and Metzger (2000). They conducted a study about the credibility of information on the Internet and developed a questionnaire about the different ways of verification strategies of information on the internet. We reduced the number of questions and reformulated those remaining, such that they were suitable for a social media context. The chosen verification strategies are: “check the page / credentials / objectives of the poster / check the topicality / use other information sources / check the comments / check if other trusted persons or pages liked or shared the post / check if the information is an objective statement or an opinion / check if the information is complete”. Further, we added the verification strategy “examine the comments of the post”. Analogue to the original scale of Flanagin and Metzger, survey participants rated the usage of these activities on a scale from 1 (never) to 5 (always).

4.2.7 Reaction to false information

In the final section of the questionnaire, we gathered data about how people react once they are sure they are confronted with false information on Facebook. As no suitable questionnaire could be found, the authors determined four possible reactions regarding the functions of Facebook: “comment to the post and remark the falsehood / share the post and remark the falsehood / report the post to Facebook / unsubscribe the page or person”. Similar to the previous section participants had to rate how often they use the proposed reactions when confronted with false information on a scale from 1 (never) to 5 (always).

5 Results

The raw data as well as all results are also available online.³

5.1 Data preparation and -transformation

To define certain variables for statistical analysis, some gathered data had to be prepared or transformed. The items concerning Facebook-usage were summed to achieve an overall-variable *Facebook-intensity*. This variable can have a value between 11 and 110 and represents the intensity and activity of the Facebook-usage. It is, therefore, possible to distinguish between “heavy user” and “light user” on a metric scale.

To distinguish between lurkers and posters, a variable – *active Facebook-usage* – was defined by summing up only those activities of the Facebook-usage questionnaire that require active interaction and participation on Facebook (update status, post photos, comment posts, “like” something, share posts). By this, we get a value between 5 and 55. We further carried out a median split to classify participants in just one group: lurkers or posters. Median split is an established statistical technique in the social sciences and psychology in similar settings to create dichotomous variables (cf. Iacobucci et al., 2015). The median for the variable *active Facebook-usage* is 14. Every participant above this value is considered a poster; everyone below is classi-

3 <https://www.dropbox.com/sh/vnvdvb52ahmpa5g/AABzYuMInbJuR2DjQqOdIBLta>

fied as a lurker. By using this definition, posters are Facebook-users that update their status, post photos, comment posts and like something on Facebook in an above-average frequency. We decided upon this method since it divided the sample relatively equal (see Chapter 5.3.2). We refer to the variable differentiating between lurkers and posters as *Facebook-user-type*.

Tables 1 and 2 give an overview of all variables.

Table 1: Overview of the independent variables

Variable-group	Variable	Possible values
gender	gender	male/female
Facebook-usage	Facebook-intensity	11 (low usage rate) – 110 (high usage rate)
	Facebook-user-type	poster/lurker
trust in social media	trust in social media	5 (low trust) – 30 (high trust)

Table 2: Overview of the dependent variables

Variable-group	Variable	Possible values
handling false information	consume false information	1 (never) – 11 (all the time)
	share false information	
	like false information	
	comment false information	
proportion of mis- and disinformation	proportion of misinformation	1 (very low) – 5 (very high)
	proportion of disinformation	
verification of information	check the page	1 (never) – 5 (always)
	check the credentials of the poster	
	check the objectives of the poster	
	checking the topicality of the post	
	use other information sources	
	check the comments	
	check if other trusted persons/pages liked or shared the post	
	check if the information is an objective statement or an opinion	
check if the information is complete		
reaction to false information	comment the post	1 (never) – 5 (always)
	share the post	
	report the post	
	unsubscribe the poster	

5.2 Statistical procedure

Depending on the analyzed research questions and the used variables, different statistical tests were carried out: *t*-tests for pairwise group comparisons, Spearman’s Rho for correlations between ordinal or metric variables and one chi-square test for a relationship between two nominal variables (gender and Facebook-user-type). As level of significance, we chose 0.05. Since we do test multiple hypotheses on the same sample, we correct the level of significance via the Bonferroni-Holm-Method (Holm, 1979). The *p*-values we report are the corrected *p*-values according to this method. We do report non-significant results via descriptive analysis if we found interesting results. Please note however that these results have to be interpreted with caution. All statistical analysis was carried with the Software *IBM SPSS Statistics*.

5.3 Descriptive statistics

5.3.1 Sample

Overall, 119 persons participated in the study with 65 female and 54 male participants. Almost all had a high school degree (114). Most of the participants were students (89). 27 participants were employees and the rest pupils or other. The average age was 25.8 with the youngest person being 19 and the oldest being 62. Most participants were aged between 21 and 27 years ($n = 95$). Only five participants were older than 35 years. Thus, while we managed to acquire many participants from the user group of Facebook in Germany that represents most users according to current statistics⁴ (the age group between 19 and 34), we were not able to gather many participants of the age group above 35.

5.3.2 Facebook usage

Table 3 illustrates the descriptive statistics for all variables concerning Facebook usage.

4 <https://de.statista.com/statistik/daten/studie/512316/umfrage/anzahl-der-facebook-nutzer-in-deutschland-nach-alter-und-geschlecht/>

Table 3: Descriptive statistics for all variables/questionnaire items concerning Facebook usage (main variable is in bold)

Variable	Active/passive	Min	Med	Avg	Max	<i>Sd</i>
call Facebook	passive	2	8	7.65	11	1.71
call Facebook on mobile	passive	1	8	6.47	10	2.52
call Facebook from work	passive	1	7	6.39	10	2.45
update status	active	1	2	1.88	10	1.24
post photos	active	1	2	1.86	6	0.77
general browsing	passive	1	5	5.12	10	2.05
read posts	passive	2	8	6.85	10	1.96
comment posts	active	1	3	3.33	10	2.08
like posts	active	1	6	5.13	10	2.24
share posts	active	1	2	2.40	9	1.64
active Facebook-usage	–	5	14	14.61	41	6.35
Facebook-intensity	–	16	49	47.08	82	12.50

Facebook is visited on average multiple times a day ($M = 7.65$, $Sd = 1.71$). With a mean value of 6.47, it is visited less frequently with the smartphone. The value represents the statement “multiple times a week”. In a similar frequency Facebook is called while being on work ($M = 6.39$, $Sd = 2.45$). The most frequent activities on Facebook are passive, e. g. reading posts ($M = 6.84$, $Sd = 1.96$) or browsing through profiles and photos ($M = 5.12$, $Sd = 2.05$). Participants state to perform more active activities rather rarely, most of the time less than once a month, e.g., sharing posts ($M = 2.40$, $Sd = 1.64$), updating the status ($M = 1.88$, $Sd = 1.24$) and posting photos ($M = 1.86$, $Sd = 0.77$). Only the activity to like posts is performed more frequently, on average multiple times a week ($M = 5.13$, $Sd = 2.24$).

Using a median split as discussed in Chapter 5.3.2, the sample could be equally divided in lurkers and posters. 65 participants are lurkers and 54 are posters according to our definition using median split. With a value of 14, the median for active Facebook-usage is relatively low. The reason for this is that Facebook is rarely used active. Therefore, posters in our study are also persons that participate *in rather low levels* on Facebook in an active way by sharing or commenting occasionally.

5.3.3 Trust in social media

The results concerning the items about trust in social media point to an average degree of trust (see Table 4).

Table 4: Descriptive statistics for all variables/questionnaire items concerning trust in social media

Variable	Min	Med	Avg	Max	<i>Sd</i>
social media is informative	1	3	2.81	5	0.86
social media is truthful	1	2	2.03	5	0.91
social media is reliable	1	2	1.87	5	0.96
social media is accurate	1	2	2.11	4	0.96
social media is essential	1	3	2.76	5	1.13
overall: trust in social media	6	14	14.33	24	3.99

The highest agreement is found for the statement that social media are informative ($M = 2.81$, $Sd = 0.86$) and essential ($M = 2.76$, $Sd = 1.13$). The lowest agreement is found for the statement that social media are a reliable information source ($M = 1.87$, $Sd = 0.96$). The summed overall-variable points to an average trust and a moderate skepticism towards social media ($M = 14.33$, $Sd = 3.99$).

5.3.4 Handling false information

Descriptive statistics concerning dependent variables are summarized in Table 5.

A mean value of 4.87 shows that participants are confronted with false information on Facebook on average multiple times a month to multiple times a week ($Sd = 2.17$). However, the active interaction with false information is very low. More than 75% state that they would never share false information ($M = 1.36$, $Sd = 0.83$). Similar numbers can be found for liking ($M = 2.10$, $Sd = 1.32$) or commenting false information ($M = 1.66$, $Sd = 1.06$). The majority of participants assume to like or comment false information less than once a month or even never at all.

Table 5: Descriptive statistics for all dependent variables/questionnaire items

Variable-group	Variable	Min	Med	Avg	Max	<i>Sd</i>
handling false information	consume false information	1	4	4.87	11	2.17
	share false information	1	1	1.36	6	0.83
	comment false information	1	1	1.66	6	1.06
	like false information	1	2	2.10	8	1.32
proportion of mis- and dis-information	proportion of misinformation	1	4	3.56	5	0.98
	proportion of disinformation	1	3	3.06	5	1.00
verification of information	check the page	1	3	3.09	5	1.17
	check the credentials of the poster	1	3	2.56	5	1.18
	check the objectives of the poster	1	3	2.80	5	1.24
	check the topicality of the post	1	3	3.10	5	1.00
	use other information sources	1	4	3.49	5	1.11
	check the comments	1	3	3.31	5	1.14
	check if other trusted persons/ pages liked or shared the post	1	2	2.32	5	1.10
	check if the information is an objective statement or an opinion	1	4	3.40	5	1.13
	check if the information is complete	1	3	2.92	5	1.08
reaction to false information	comment the post	1	1	1.82	5	1.4
	share the post	1	1	1.30	4	0.70
	unsubscribe the poster	1	3	3.25	5	1.18
	report the post	1	2	2.03	5	1.18

5.3.5 Mis- and disinformation

Concerning the estimation of the proportion of mis- and disinformation to false information, the data shows that this proportion is estimated as equally distributed (see Table 5). The proportion of misinformation is estimated as slightly higher ($M = 3.56$, $Sd = 0.98$) than the proportion of disinformation ($M = 3.06$, $Sd = 1.00$). This means that overall, our participants believe that half of all false information on Facebook is misinformation (therefore mistakenly false) and the other half is disinformation (therefore shared with knowledge about the falsehood).

5.3.6 *Verification of information*

It is shown that the most frequent verification activities are using other information sources ($M = 3.49$, $Sd = 1.11$) as well as checking if the post is an objective statement or an opinion ($M = 3.40$, $Sd = 1.12$). Occasionally activities like checking the comments of the post ($M = 3.31$, $Sd = 1.14$) or reviewing the topicality are carried out ($M = 3.11$, $Sd = 1.01$). Very rarely participants examine if other known and trusted persons liked or shared the post ($M = 2.32$, $Sd = 1.10$).

5.3.7 *Reaction to false information*

The analysis of the questions about the reaction to false information confirms a known pattern (Table 5). Participants state to very rarely to never comment ($M = 1.82$, $Sd = 1.04$), share ($M = 1.30$, $Sd = 0.7$) or report a post ($M = 2.03$, $Sd = 1.18$). The only activity that is performed rather often is unsubscribing ($M = 3.25$, $Sd = 1.18$). However, in regard to the mean value, this activity is also only carried out occasionally. Once again, this activity is a more passive way of responding to false information.

5.4 **Inference statistics**

In the following section we present the results for the inference statistics structured according to research questions. All significant results concerning the main independent variables are also presented in Tables 6–8. All results and data online are also available online.⁵

5.4.1 *Gender and Facebook-usage (G-RQ1)*

We performed a *t*-test to identify significant differences between men and women concerning the way and intensity of Facebook-usage. Every single activity was analyzed in regard to the frequency of usage. No significant differences were found. On a descriptive level, women click more often the like-button on posts ($M = 5.66$, $SD = 1.97$) than men do ($M = 4.5$, $SD = 2.4$), $t(117) = 2.84$, $p = 0.095$. Rounded and transferred to the used scale, this means that men use the like-button on average once per week while women like

5 <https://www.dropbox.com/sh/vnvdv52ahmpa5g/AABZyuMInbJuR2DjQqOdIBLta>

posts once a day. Facebook-usage is very homogenous across the genders. We examined the relationship between gender and the groups of lurkers and posters with a chi-square test of independence, but this also shows no significant result according to the corrected p -values ($\chi^2(1, n=119) = 2.78, p = 0.81$). While the proportion of lurkers (31) and posters (34) is almost equal among women, lurkers are overrepresented (34) while there are fewer posters (20) among men.

5.4.2 *Gender and trust in social media (G-RQ2)*

An independent samples t -test could not prove any significant differences concerning trust in social media. This result is found for the overall variable as well as all the single statements. Women and men have a homogenous mediocre trust in social media (see Chapter 5.3.3).

5.4.3 *Gender and handling false information (G-RQ3)*

All questions of the questionnaire subsection of handling false information were used to perform a t -test to find significant differences between the genders. It can be stated, that women believe more so than men to consume (women: $M = 5.03, SD = 2.12$; men: $M = 4.68, SD = 2.24$), comment (women: $M = 1.74, SD = 1.28$; men: $M = 1.56, SD = 0.72$), share false information (women: $M = 1.43, SD = 0.99$; men: $M = 1.28, SD = 0.60$) and like false information more frequently (men: $M = 2.32, SD = 1.45$; women: $M = 3.35, SD = 1.07$). However, the differences were not significant. It was already shown that women hit the like-button on posts more often than men do. Therefore, this circumstance also occurs in relationship to false information.

5.4.4 *Gender and mis- and disinformation (G-RQ4)*

A t -test shows that there is no significant difference in the estimation of the proportion of misinformation to false information on Facebook. Women estimate the proportion of misinformation ($M = 3.74, SD = 0.87$) slightly higher than men ($M = 3.35, SD = 1.07$), $t(117) = 2.14, p = 1.0$. In regard to the estimation of the proportion of disinformation, the results are almost equal.

5.4.5 Gender and the verification of information (G-RQ5)

T-Tests on all frequency statements of verification activities prove that there are no gender-specific differences. Men and women behave almost the same in verifying information on Facebook.

5.4.6 Gender and the reaction to false information (G-RQ6)

Similar to the verification of information no significant results were found concerning the reaction to false information. The behavior is again homogeneous.

5.4.7 Facebook-usage and trust in social media (F-RQ7)

All significant correlations concerning Facebook-intensity are summarized in Table 6.

Table 6: Significant correlations of Facebook-intensity

Variable-group	Facebook-intensity correlation with ...	<i>r</i>	Corrected <i>p</i> -value
trust in social media	trust in social media	0.33	< 0.001
	social media is informative	0.32	< 0.001
	social media is reliable	0.26	< 0.001
	social media is accurate	0.29	< 0.001
handling false information	share false information	0.31	0.032
	comment false information	0.39	< 0.001
	like false information	0.33	< 0.001

A significant correlative relationship between the intensity of Facebook-usage and trust in social media is verified, $r(117) = 0.33$, $p < .001$. We also analyzed the isolated sub-items of trust in social media and were able to identify that this relationship is significant for the statements to consider social media informative, $r(117) = 0.32$, $p < .001$, reliable, $r(117) = 0.26$, $p < .001$, and accurate, $r(117) = 0.29$, $p < .001$. The other two statements (essential, truthful) show rather weak correlations.

The *t*-test to find differences between posters and lurkers in social media shows no significant result.

5.4.8 Facebook-usage and handling false information (F-RQ8)

All active interactions with false information show significant correlation with the intensity of Facebook-usage. The more Facebook is used the more likely and frequent it is that someone has shared, $r(117) = 0.31$, $p = 0.032$, liked, $r(117) = 0.33$, $p < .001$, or commented, $r(117) = 0.39$, $p < .001$, false information. For the general confrontation with false information this relationship is however not significant, $r(117) = 0.15$, $p = 0.81$.

By separating between lurkers and posters via the variable Facebook-user-type, we were able to specify these findings with t -tests. All significant results concerning these group-based comparisons and the dependent variables are illustrated in Table 7.

Table 7: Significant results for group-based comparisons between posters and lurkers for all dependent variables

Variable-group	Facebook-intensity differences with ...	Facebook-intensity: avg		t -value	Corrected p -value
		Poster	Lurker		
handling false information	share false information	1.67	1.10	-3.58	0.032
	comment false information	2.20	1.20	-4.23	< 0.001
	like false information	2.64	1.65	-5.37	< 0.001
reaction to false information	comment the post	2.20	1.50	-3.80	< 0.001
	report the post	2.44	1.70	-3.55	0.032

There are significant differences between posters and lurkers in sharing (posters: $M = 1.67$, $SD = 1.1$; lurkers: $M = 1.1$, $SD = 0.36$; $t(117) = -3.58$, $p = 0.032$), liking (posters: $M = 2.64$, $SD = 1.52$; lurkers: $M = 1.65$, $SD = 0.93$; $t(117) = -5.37$, $p < .001$) and commenting false information (posters: $M = 2.20$, $SD = 1.32$; lurkers: $M = 1.20$, $SD = 0.40$; $t(117) = -4.23$, $p < .001$). However, similar to Facebook-intensity, for this analysis no significant result concerning the general passive consumption of false information is found.

5.4.9 Facebook-usage and mis- and disinformation (F-RQ9)

No significant correlation between the estimation of the proportion of disinformation and the intensity of Facebook-usage is found, $r(117) = 0.2$, $p = 0.48$. Regarding lurkers and posters no significant differences in the estimation neither for misinformation nor for disinformation are shown as well.

5.4.10 Facebook-usage and the verification of information (F-RQ10)

We did not find significant results concerning the verification of information and differences between the Facebook-usage measured via intensity or differing between posters and lurkers.

5.4.11 Facebook-usage and the reaction to false information (F-RQ11)

For the reactions to false information, we found no significant correlations based on the degree of Facebook-intensity.

Nevertheless, we did find significant results for the reaction to false information when comparing lurkers to posters (Table 7). There is a significant difference for the statements to comment to posts with false information (posters: $M = 2.2$, $SD = 1.17$; lurkers: $M = 1.5$, $SD = 0.8$; $t(117) = -3.8$, $p < .001$) or report them to Facebook (posters: $M = 2.44$, $SD = 1.28$; lurkers: $M = 1.7$, $SD = 0.97$; $t(117) = -3.55$, $p = 0.032$). These activities are done significantly more often by posters than by lurkers.

5.4.12 Trust in social media and handling false information (T-RQ12)

Spearman's correlation coefficient proves that the overall trust in social media correlates negatively with the statement to consume false information, $r(117) = -0.19$, $p = 0.52$. However, this result is not significant but shows descriptively that participants with low degree of trust in social media claim to be confronted more often with false information than participants with a higher trust in social media. Regarding the single items of the questionnaire about trust in social media, we can identify precisely which statements lead to the assessment to be confronted with false information more frequently. Moderately strong and also significant correlations are found for the statements that posts on social media platforms are truthful, $r(117) = -0.38$, $p < 0.001$ and reliable, $r(117) = -0.28$, $p = 0.04$. If participants specifically had doubts concerning these attributes and social media, they affirm to be confronted with false information more often.

5.4.13 Trust in social media and mis- and disinformation (T-RQ13)

No significant results between the trust in social media and the estimation of mis- and disinformation were found.

5.4.14 *Trust in social media and verification of false information* (T-RQ14)

Correlations between the trust in social media and the frequency of performed verifications strategies for information assumed false were examined. For one verification activity a significant relationship is found (Table 8). It is a positive and moderately strong correlation with the activity to check the corresponding comments of a post to verify if the information is true or false, $r(117) = 0.32, p < .001$. Participants with higher trust in social media use this method significantly more often to verify information. For all the other verification strategies, the results show no significance.

Table 8: Significant correlations of trust in social media

Variable-group	Trust in social media correlation with ...	<i>r</i>	Corrected <i>p</i> -value
handling false information	check the comments	0.32	< 0.001

5.4.15 *Trust in social media and reaction to false information* (T-RQ15)

None of the reaction activities correlate significantly with the variable representing the overall trust in social media. In addition, when analyzing the sub-items there are also no significant results present. The strongest correlation is however between the statement to view social media as truthful and reporting posts to Facebook as a reaction to false information, $r(117) = -0.24, p = 1.0$.

6 Discussion

In the following section, the results are interpreted and put in context considering the research area. We also discuss implications for future research.

6.1 Overall results

The descriptive results, in addition to the analysis of the independent variables, provide insights for the research field of social media and false infor-

mation. The main activities considering Facebook-usage are passive activities like reading posts and browsing through profiles and photos. Regarding the average Facebook-usage, a user is considered a poster according to our definition using median split if she only interacts with Facebook actively *at all*, e.g., sharing posts and updating her profile. Trust in social media is mediocre. Indeed, participants doubt the accurateness of the information shared on social media. This verifies previous research about media usage that showed that in comparison social media platforms are less trusted than other media channels (Johnson & Williams, 2010). The reason for this low trust level might have to do with the finding that users believe themselves to be confronted with false information up to multiple times a week. However, our participants assume that they do not share or interact actively with this false information. It is uncertain how accurate these estimations are. Eventually these results are due to under-reporting of undesirable behavior, which is a known bias in self-report studies (Neeley & Cronley, 2004). The proportion of mis- and disinformation is considered even. Based on our results participants believe that false information on Facebook comprises overall of 50% misinformation and 50% disinformation. Future research can analyze this finding in more detail by examining if this subjective perception is actually true. Regarding the verification behavior, it is noticeable that the main verification strategy is to use other sources of information. This is very likely connected with the felt frequent consumption of false information on social media as well as the low degree of trust in social media. Considering the reaction to false information, we discovered that active participation and interaction is rather seldom. Users barely comment or share false information remarking it as false. Participants also barely report false information to Facebook, although it being an anonymous activity. Therefore, Facebook as well as research should strive to find methods to motivate users to react actively to false information. By this way, the propagation of false information can be better prevented. The only very frequent reaction is to unsubscribe the poster or Facebook-page. Knowing this, it might be possible to detect false information automatically. Further research about this topic is necessary to prevent false information propagation.

6.2 Gender

We did not find evidence that gender is a relevant influencing factor for the majority of research questions. Through descriptive analysis, we recognized

that women tend to be posters rather than lurkers and actively participate on Facebook. Women also use the like-button more frequent than men do. The increased usage of the like-button is also reflected when analyzing the way false information is handled. Here as well, women affirm to like false information more often than men do. Overall, we identified that women believe to be confronted with false information as well as to actively interact with false information slightly more often than men do. Future research should examine if this finding is based on a subjective estimation of women or if women are indeed confronted more often with objectively false information. Further, women estimate the proportion of misinformation higher than men do. However, none of these descriptive findings proved to be significant according to corrected p -values. We can state that the behavior between the genders regarding false information on Facebook is rather similar and homogenous according to our data. The missing influence of gender was already proven in previous studies (Chen & Sin, 2013) and is verified in the present study.

6.3 Facebook-usage

Facebook-usage was examined by differing between lurkers and posters (Marret & Joshi, 2009) and by using the intensity and frequency of Facebook-usage as individual variable. By this way, we were able to gain important insights. The intensity and frequency of Facebook-usage correlates with the trust in social media. It is not clear whether the frequent usage increases the degree of trust or whether persons with high trust in social media tend to use Facebook more intensively. The separation between posters and lurkers could not prove any significant results concerning Facebook-usage and trust in social media. The results about handling false information prove that the higher the Facebook-usage the more likely it is to be confronted with false information and to interact with false information. This finding verifies that, according to our participants, false information is indeed a part of Facebook. The general way of Facebook-usage is carried over to the way false information is handled. Participants that interact more actively with Facebook also affirm to interact more actively with false information than more passive users do. Depending on the general behavior (active or passive) participants believe to consume false information regularly. In combination with the results about trust in social media, this result is counterintuitive. Although more active and frequent Facebook-users assume to consume more false information, they also show higher levels of trust in social media. This para-

dox finding should be research object for future studies. The estimation of the proportion of mis- and disinformation shows no relevant relationships with Facebook-usage. Regardless of the user type, the proportion is estimated as even. Posters tend to react more actively to false information than lurkers and claim to comment posts with false information or even report them to Facebook. Overall, heavy users and posters tend to actively react to false information. At the same time this is also a problem for identifying false information in a social network because only a special and smaller user group reports and comments false information. In this context, it is especially alarming that even the anonymous activity of reporting is not popular with passive users when confronted with false information. Apart from unsubscribing, they rather tend to not react at all to false information. Future research has to find possibilities to better integrate the passive user group in detecting false information.

6.4 Trust in social media

We were able to find some punctual significant results concerning the variable trust in social media. While we did not find a significant correlation between this variable and the impression to consume false information, we did find that the more users doubt the truthfulness and reliability of Facebook posts the more they report to be confronted with false information. The relationship is rather obvious. The more one doubts the information on Facebook the more skeptical about social media she or he gets. An alternative explanation might be that this is a form of “self-fulfilling prophecy”. Someone is skeptical towards social media and tends to believe subjectively to be confronted more often by false information. Subsequent studies might gather qualitative data to answer abstract research questions like this more precisely. It is necessary for future research to discover if skeptical persons indeed are confronted by more false information or their doubt just leads to the subjective impression of consuming more false information. It is also possible that more skeptical persons pay more attention to false information and therefore are confronted more often while persons with high trust levels are more likely to overlook them. Regarding the verification and reaction behavior, trust in social media is not an important influencing factor. Solely for verification strategies we identified that trust in social media correlates with the verification activity to check the comments of a post. Apparently, higher trust in

social media is extended to higher trust in the participants of the social network and their statements.

7 Limitations

Several limitations of the study restrict the value and interpretation of the data, but also offer possibilities for subsequent studies. Compared to other studies the number of participants (119) is rather low (Chen & Sin, 2013; Marret & Joshi, 2009) and due to the nature of our recruitment process, we have an overrepresentation of younger students and way fewer regular employees in our sample. While the age group between 18 and 34 years represents the core group of Facebook in Germany, the sample is not fully representative since we were not able to gather enough participants for the age group above 35 who still represents an important chunk of Facebook users.⁶ Thus, our interpretations are mostly representative for younger students but not German Facebook users in general. Further, note that the study was conducted in Germany, so nationality-specific biases are possible. Future studies with bigger sample sizes and samples that are more international can gather insights that are more meaningful.

The fact that the majority of participants use Facebook rather passively might be problematic interpreting results considering the Facebook-usage. By using a median split and assigning participants with even low levels of activity as posters, the differentiation between these groups might be skewed. We rather recommend in future work to use metric scales similar to the variable Facebook-intensity in our work to deal with this problem.

Lastly, the focus of this study relies on the survey and analysis of quantitative data. The analysis of qualitative data, e.g., by using open questions or interview-techniques, can add value to the research area by identifying the underlying motivations and reasons behind several findings.

6 <https://de.statista.com/statistik/daten/studie/512316/umfrage/anzahl-der-facebook-nutzer-in-deutschland-nach-alter-und-geschlecht/>

8 Conclusion

The presented study contributes to the research field of false information on Facebook. Overall, we prove that users believe to be confronted with false information and fake news regularly and that the trust in social media platforms is mediocre. We also analyzed the actual verification and reaction behavior when confronted with false information. Another focus of this study is the analysis of the influence of individual factors. Important results were found for the factors user type (Facebook-usage) and trust in social media. Gender was not proven as relevant influencing factor. Future research can pick up several impulses of this study and gather data that is more detailed by using qualitative methods. False information in social networks is a problem that should concern society as a whole.

References

- Alexander, J. M., & Smith, J. M. (2010). Disinformation: A taxonomy. http://repository.upenn.edu/cgi/viewcontent.cgi?article=1966&context=cis_reports
- Chen, X., & Sin, S. C. J. (2013). ‘Misinformation? What of it?’ Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1–4.
- Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship*, 41(5), 583–592.
- Chiluwa, I. E., & Samoilenko, S. A. (2019). *Handbook of Research on Deception, Fake News, and Misinformation Online*. Hershey, PA: IGI Global.
- Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users’ personality and social media use. *Computers in Human Behavior*, 26(2), 247–253.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, 21, 3–33.
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Fallis, D. (2009). A conceptual analysis of disinformation. <http://cite-seerx.ist.psu.edu/viewdoc/download?doi=10.1.1.569.304&rep=rep1&type=pdf>

- Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, 77(3), 515–540.
- Fletcher, R., Schifferes, S., & Thurman, N. (2020). Building the ‘Truthmeter’: Training algorithms to help journalists assess the credibility of social media sources. *Convergence*, 26(1), 19–34. doi: [10.1177/1354856517714955](https://doi.org/10.1177/1354856517714955)
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). The median split: Robust, refined, and revived. *Journal of Consumer Psychology*, 25(4), 690–704.
- Jackob, N. G. E. (2010). No Alternatives? The Relationship between Perceived Media Dependency, Use of Alternative Information Sources, and General Trust in Mass Media. *International Journal of Communication*, 4, 589–606.
- Johnson, E. P. & Williams, D. (2010). Media Channel Effectiveness and Trust. <https://www.amstat.org/sections/srms/Proceedings/y2010/Files/400107.pdf>
- Kaplan, A. M., & Haenlein, M. (2012). Social media: back to the roots and back to the future. *Journal of Systems and Information Technology*, 14(2), 101–104.
- Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research*, 18 (1).
- Karlova, N. A., & Lee, J. H. (2011). Notes from the underground city of disinformation: A conceptual investigation. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–9.
- Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, 28(2), 331–339.
- Leeder, C. (2019). How college students evaluate and share “fake news” stories. *Library & Information Science Research*, 41(3), 100967. doi: <https://doi.org/10.1016/j.lisr.2019.100967>
- Lenhart, A. (2009). Adults and Social Network Websites. Pew Internet and American LifeProject. <https://www.pewresearch.org/internet/2009/01/14/adults-and-social-network-websites/>
- Marett, K., & Joshi, K. D. (2009). The decision to share information and rumors: Examining the role of motivation in an online discussion forum. *Communications of the Association for Information Systems*, 24(1), 47–68.
- Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter Under Crisis: Can we trust what we RT? In *SOMA '10 – Proceedings of the First Workshop on Social Media Analytics* (pp. 71–79). New York, NY: ACM Press.

- Montesi, M. (2020). Understanding fake news during the Covid-19 health crisis from the perspective of information behaviour: The case of Spain. *Journal of Librarianship and Information Science*, 1–12. doi: [10.1177/0961000620949653](https://doi.org/10.1177/0961000620949653)
- Neeley, S. M., & Cronley, M. L. (2004). When Research Participants Don't Tell It Like It Is: Pinpointing the Effects of Social Desirability Bias Using Self vs. Indirect-Questioning. *Advances in Consumer Research*, 31, 432–433.
- Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology*, 7(2), 159–186.
- Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsychology & Behavior*, 12(6), 729–733.
- Perrin, A. (2015). Social Media Usage: 2005–2015. <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>
- Poushter, J. (2016). Social networking very popular among adult internet users in emerging and developing countries. Pew Internet and American LifeProject. <http://www.pewglobal.org/2016/02/22/social-networking-very-popular-among-adult-internet-users-in-emerging-and-developing-nations/>
- Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & Behavior*, 11(2), 169–174.
- Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rökkum, J. (2013). The media and technology usage and attitudes scale: An empirical investigation. *Computers in Human Behavior*, 29(6), 2501–2511.
- Schmidt, T. (2016). *Der Einfluss der Persönlichkeit auf das Informationsverhalten: Eine empirische Studie zur Suche im Web*. Bachelor Thesis, University of Regensburg, April 2015. [urn:nbn:de:bvb:355-epub-338868](https://nbn-resolving.org/urn:nbn:de:bvb:355-epub-338868)
- Schmidt, T., Wittmann, V. & Wolff, C., (2019). The Influence of Participants' Personality on Quantitative and Qualitative Metrics in Usability Testing. In: Alt, F., Bulling, A. & Döring, T. (Eds.), *Mensch und Computer 2019 – Tagungsband*. New York, NY: ACM Press. DOI: [10.1145/3340764.3340787](https://doi.org/10.1145/3340764.3340787)
- Schmidt, T., & Wolff, C. (2015). Genderspezifische Unterschiede im Informationsverhalten am Beispiel E-Commerce. *Information – Wissenschaft & Praxis*, 66(1), 65–76.
- Schmidt, T., & Wolff, C. (2016). Personality and information behavior in web search. *Proceedings of the Association for Information Science and Technology*, 53, 1–6.
- Schmidt, T. & Wolff, C. (2017). Der Einfluss von User Interface-Attributen auf die Ästhetik. In: Burghardt, M., Wimmer, R., Wolff, C. & Womser-Hacker, C.

(Hrsg.), *Mensch und Computer 2017 – Tagungsband* (pp. 61–72). Regensburg: Gesellschaft für Informatik e.V.

- Schmidt, T., & Wolff, C. (2018). The Influence of User Interface Attributes on Aesthetics. *i-com*, 17(1), 41–55.
- Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston marathon bombing. In *iConference 2014 Proceedings*. (pp. 654–662). Berlin: Humboldt-Universität.
- Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, 24(1), 115–139.
- World Economic Forum. (2014). Top 10 trends of 2014: 10. The rapid spread of misinformation online. <http://bit.ly/1edZQQF>
- Zimmer, F., Scheibe, K., Stock, M., & Stock, W. (2019a). Echo chambers and filter bubbles of fake news in social media. Man-made or produced by algorithms. Paper presented at the *8th Annual Arts, Humanities, Social Sciences & Education Conference, Honolulu, Hawaii*.
- Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019b). Fake news in social media: Bad algorithms or biased users? *Journal of Information Science Theory and Practice*, 7(2), 40–53.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 125–154. DOI: doi.org/10.5283/epub.44942.

Session 3:
Digital Humanities

SentText: A Tool for Lexicon-based Sentiment Analysis in Digital Humanities

Thomas Schmidt

Johanna Dangel

Christian Wolff

Media Informatics
Group, University of
Regensburg, Germany

Media Informatics
Group, University of
Regensburg, Germany

Media Informatics
Group, University of
Regensburg, Germany

thomas.schmidt@ur.de

johannadangel97@web.de

christian.wolff@ur.de

Abstract

We present *SentText*, a web-based tool to perform and explore lexicon-based sentiment analysis on texts, specifically developed for the Digital Humanities (DH) community. The tool was developed integrating ideas of the user-centered design process and we gathered requirements via semi-structured interviews. The tool offers the functionality to perform sentiment analysis with predefined sentiment lexicons or self-adjusted lexicons. Users can explore results of sentiment analysis via various visualizations like bar or pie charts and word clouds. It is also possible to analyze and compare collections of documents. Furthermore, we have added a close reading function enabling researchers to examine the applicability of sentiment lexicons for specific text sorts. We report upon the first usability tests with positive results. We argue that the tool is beneficial to explore lexicon-based sentiment analysis in the DH but can also be integrated in DH-teaching.

Keywords: sentiment analysis; dictionary-based approaches; Digital Humanities; usability; tool; user-centered design; lexicon-based approaches

1 Introduction

Sentiment analysis (or opinion mining) is a term used to describe computational methods for predicting and analyzing sentiment, predominantly in written text (Liu, 2016, p. 1). Sentiment analysis is especially popular for social media content (Moßburger et al., 2020; Schmidt, Hartl, Ramsauer, Fischer, Hilzenthaler, & Wolff, 2020; Schmidt, Kaindl, & Wolff, 2020) and any

other form of user generated content (cf. Mäntylä et al., 2018). Sentiment analysis offers various benefits for industry and is used, for example, to investigate the popularity of enterprises or political parties on social media (cf. *ibid.*), predict the mood of users in usability testing (Schmidt, Schindwein, Lichtner, & Wolff, 2020), in health informatics (Hartl et al., 2019) and gaming (Halbhuber et al., 2019). Sentiment Analysis is focused on the prediction of the attitude towards an entity on a bipolar scale or nominal classes consisting of positive, neutral and negative (Vinodhini & Chandrasekaran, 2012). The neighboring research area of emotion analysis is similar in its approaches and main goals but focused on the prediction of more differentiated emotional classes like anger, sadness and happiness (cf. Binali et al., 2010).

Both research fields have gained increasing popularity in the field of Digital Humanities (DH), especially in the area of Computational Literary Studies (cf. Kim & Klinger, 2018). Sentiment and emotion analysis have been used to investigate various types of literary texts like novels (Jannidis et al., 2016; Kakkonen & Kakkonen, 2011; Reagan et al., 2016), fairy tales (Alm & Sproat, 2005; Mohammad, 2011), plays (Mohammad, 2011; Nalisnick & Baird, 2013; Schmidt & Burghardt, 2018a, 2018b; Schmidt, Burghardt, & Dennerlein, 2018b; Schmidt, 2019; Yavuz, 2021), fan fictions (Kim & Klinger, 2019a, 2019b), online writings (Pianzola et al., 2020), historical political texts (Sprugnoli et al., 2016) or pop song lyrics (Napier & Shamir, 2018; Schmidt, Bauer, Habler, Heuberger, Pils, & Wolff, 2020). Research goals vary for the application of sentiment and emotion analysis on these text sorts. Most research tries to explore general purpose applications on these texts to analyze descriptive results, e.g., with a focus on sophisticated visualizations of sentiment and emotion distributions and progression or comparisons of different works (Kakkonen & Kakkonen, 2011; Mohammad, 2011; Reagan et al., 2016; Napier & Shamir, 2018; Schmidt, Burghardt, & Dennerlein, 2018b; Schmidt, 2019). Others evaluate different methodological approaches for this challenging text sort comparing the performance on annotated text units (Schmidt & Burghardt, 2018a; Kim & Klinger, 2019a). Examples of other projects and research goals are the analysis and prediction of plot developments (Reagan et al., 2016), character relations (Nalisnick & Baird, 2013; Schmidt & Burghardt, 2018b; Yavuz, 2021) or “happy endings” (Jannidis et al., 2016) via sentiment and emotion analysis. One branch of research focuses on the annotation of texts with sentiment or emotion information to create well-curated corpora for evaluation and machine learning and to investigate annotation behavior and agreement statistics (Alm & Sproat, 2005; Sprugnoli et al.,

2016; Schmidt, Burghardt, & Dennerlein, 2018a; Schmidt, Burghardt, Dennerlein, & Wolff, 2019a; Schmidt, Jakob, & Wolff, 2019; Schmidt, Winterl, Maul, Schark, Vlad, & Wolff, 2019). Modern approaches explore multimodal methods to analyze sentiment in cultural artefacts (Schmidt, Burghardt, & Wolff, 2019; Ortloff et al., 2019).

From a methodological standpoint, the vast majority of the aforementioned research projects in DH often rely on the general purpose, rule-based method of lexicon- (also often called dictionary-) based sentiment analysis. The main idea of lexicon-based methods in Natural Language Processing (NLP) is purely descriptive and is to count words according to large lists of words of a specific category (e.g., positive or negative sentiment), which is called the dictionary or lexicon, to derive insights about the analyzed texts. In the context of sentiment and emotion analysis, some of the most famous English sentiment lexicons are the *Bing Lexicon* (Hu & Liu, 2004) and the *NRC Emotion Lexicon* (Mohammad & Turney, 2013). They consist of large lists of words annotated with their most likely sentiment connotation (e.g., positive, negative, neutral). These words are referred to as sentiment bearing words (SBWs). They are either assigned with a sentiment class or by values with a linear metric, e.g., between -3 (negative) and 3 (positive), to represent to what degree a word is connoted with a specific sentiment. In the case of sentiments consisting of nominal assignments for sentiment classes the value for a positive word is regarded as 1 and for a negative word as -1 . To gain an overall value of the expressed sentiment of a text, all values of the detected words are summed up. In the same way, one can perform emotion analysis (Mohammad & Turney, 2013) with lexicons that consist of list of words for specific emotion categories. For more information about the creation of sentiment or emotions lexicons and an overview of research cf. Mohammad (2020).

In general, the lexicon-based approach has been proven inferior compared to advanced machine learning approaches (cf. Kim & Klinger, 2018) for multiple use cases. Current state-of-the-art machine learning approaches in sentiment and emotion analysis involve large word embeddings and deep neural networks (cf. Zhang et al., 2018; Shmueli et al., 2019). Lexicon-based approaches are currently only recommended for areas in which not enough large-scale annotated corpora exist which would be necessary for modern methods. This is mostly the case for under-resourced languages or text sorts which are uncommon in the NLP-community which is the case for literary and historic texts. However, even outside of this use case, libraries and APIs

applying lexicon-based text analysis have regained popularity in the NLP-community to calculate benchmarks or perform first explorations (e.g., VADER; Hutto & Gilbert, 2014).

While rare compared to other text sorts, one can find first projects in DH exploring state-of-the-art deep learning techniques, e.g., on literary texts (Kim & Klinger, 2019a). Nevertheless, lexicon-based methods are still widely popular and the established sentiment analysis method in DH. The vast majority of recent and current projects use some sort of lexicon-based techniques (Alm & Sproat, 2005; Mohammad, 2011; Nalisnick & Baird, 2013; Reagan et al., 2016; Sprugnoli et al., 2016; Jannidis et al., 2016; Schmidt & Burghardt, 2018a, 2018b; Schmidt, 2019; Pianzola et al., 2020; Schmidt, Bauer, Habler, Heuberger, Pils, & Wolff, 2020; Yavuz, 2021). Reasons for this are certainly the lack of large-scale and well-annotated training corpora that would be necessary for state-of-the-art-methods but also that it is a rather transparent and easy method enabling researchers to perform fast and comprehensible explorations of textual sentiment. Schmidt, Burghardt, and Wolff (2018) discuss this dominance of lexicon-based methods in DH in more detail.

Nevertheless, to our knowledge, a usable and accessible tool to perform lexicon-based sentiment analysis specifically for DH researchers that is adaptable to the various use cases of DH and that allows the easy introspection of results does not exist so far. While this is certainly no problem for DH researchers familiar with coding and computational methods, humanities scholars with limited IT skills are quickly discouraged. As Burghardt and Wolff (2014) argue, it is especially important for tools in DH to be accessible and of high usability to lower the participation threshold and help establish digital methods in the humanities community. Furthermore, it is important to note that humanities scholars are a special user group with very specific needs and skills that have to be taken into account during the development phase. While one can identify a growing interest in projects developing such accessible tools (e.g., Schmidt, Burghardt, Dennerlein, & Wolff, 2019b; Schmidt, Jakob, & Wolff, 2019), there is still a lack of easy-to-use and accessible tools for various methods in DH and the tool *SentText* was developed to close this gap for lexicon-based sentiment analysis.

We present the tool *SentText* (see Fig. 1 for the logo), a web-based tool for performing lexicon-based sentiment analysis on texts via a user interface without the need of coding skills. We argue that the tool is beneficial for first explorations in DH research, e.g., to compare the applicability of various

sentiment lexicons for a specific text sort but also in DH teaching to introduce students into the possibilities but also challenges of lexicon-based sentiment analysis. While the tool is currently focused on sentiment analysis in its default settings and overall presentation it can easily be applied to any sort of lexicon-based analysis. To adapt to the critical user group of humanities scholars we apply the user-centered design process (cf. Vredenburg et al., 2002) and integrate the feedback of this user group early in the development process. Furthermore, we integrate usability tests at various steps of the development to evaluate and improve the usability of the tool.



Fig. 1 Logo of the tool *SentText*

2 Development and requirements analysis

The tool was developed with the Flask framework¹ in Python 3.7, JavaScript, CSS3 and HTML5. We use multiple libraries for natural language processing like the NLTK². Our development process uses ideas of the user-centered design process (cf. Vredenburg et al., 2002). We develop the tool in multiple iterations integrating the feedback of the final user group at various steps via methods of requirements engineering or usability testing.

We acquired first requirements via interviews at the beginning of development. We conducted semi-structured interviews with six researchers with backgrounds either in DH, literary studies or computational text analysis to gather an understanding of potential text sorts, their workflow and potential functionalities for a lexicon-based sentiment analysis tool. All researchers either have performed research in sentiment analysis and DH or plan to do so. The interviews were conducted either personally or via video call and the audio was transcribed afterward. The interview style was semi-structured

1 <https://flask.palletsprojects.com/en/1.1.x/>

2 <https://www.nltk.org/>

with loose guidelines consisting of questions and topics to talk about. We summarize some of the main requirements that had implications on the development of *SentText*. Some of the gathered requirements have also implications on the tool development in other research areas:

- Corpora come in different forms and shapes, the most important file formats that should be supported are XML, TEI and TXT.
- Preferred method to import text is a simple upload possibility.
- Important preprocessing steps before the sentiment analysis that should be integrated are: Lemmatization, stop words filtering, lower casing (However we noticed the higher the technical expertise of our participants the more likely they argued to perform these steps themselves and that they would eventually not trust a web tool.).
- Sentiment lexicons have to be adjusted (e.g., adding and removing words) and it should be possible to use own self-created lexicons; some standard sentiment lexicons should be offered by default.
- the most desired visualization types: a progression curve throughout the text, word clouds, visualizations enabling the comparison between texts or larger text groups
- Results should be traceable and transparent; it should be possible to follow to calculation process to the smallest unit.
- Web is the preferred platform since it is more accessible (no long installation process, no problems with different operating systems).
- All results should be downloadable in standard formats, e.g., CSV, JPEG.
- The tool should consist of a graphical user interface and should have a documentation.

Furthermore, we conducted a market analysis comparing various established online sentiment analysis tools to reflect upon useful functions, design elements and what we may offer with our own tool; to name a few: *SentiStrength*³, *Sentiment Analyzer*⁴ or the *Stanford Sentiment Analysis Tool*⁵. However, none of these tools is directed towards DH, works with transparent lexicon-based methods and oftentimes lacks important functionality like the upload of own material. Nevertheless, we systematically analyzed functional-

3 <http://sentistrength.wlv.ac.uk/>

4 <https://www.danielsooper.com/sentimentanalysis/default.aspx>

5 <http://nlp.stanford.edu:8080/sentiment/rmtnDemo.html>

ity, layout, advantages and disadvantages of these tools to derive potential features for our future tool and how to implement them.

3 Functionality

SentText is web-based and can therefore be used without any sort of installation. The tool is available online:

<https://thomasschmidtur.pythonanywhere.com/>.

We integrated a detailed documentation and “about” page but also offer information at various steps of the tool usage via text or tooltips.

Users can upload texts in UTF-8-encoded TXT- or XML-format and perform the sentiment analysis. Using advanced options users can choose to perform stop words removal, lemmatization (only for German: via *textblob*⁶) or the integration of negations into the sentiment calculation (negations before a sentiment bearing word reverse the sentiment value).

Please note that at the moment these functionalities are performed for German (as we currently focus on the support of the study of German literary texts), however, users can upload lexicons and stop word lists for other languages and we also plan to include modules for other languages in future work. Users can also choose the sentiment lexicon to be used for the sentiment analysis. We are currently offering per default *SentiWS* (Remus et al., 2010) and *BAWL-R* (Vo et al., 2009), two popular lexicons for German. However, users can also use self-created or adjusted sentiment lexicons if they follow a specific CSV-format, which is defined and explained in the documentation. We plan to add more free lexicons for German (e.g., Waltinger, 2010) as well as for other languages in the future (e.g., Mohammad & Turney, 2013). The adjustment of the lexicons enables users to change lexicons or use resources that are better suited for the specific text sort of interest. Once *SentText* has completed a sentiment analysis run, the results screen is shown (Fig. 2).

On the left side, users can investigate their documents overall but also create groups via “Create Folder” to compare collections of text. By double clicking on a document or a group, users can investigate the results in the

6 <https://pypi.org/project/textblob-de/>

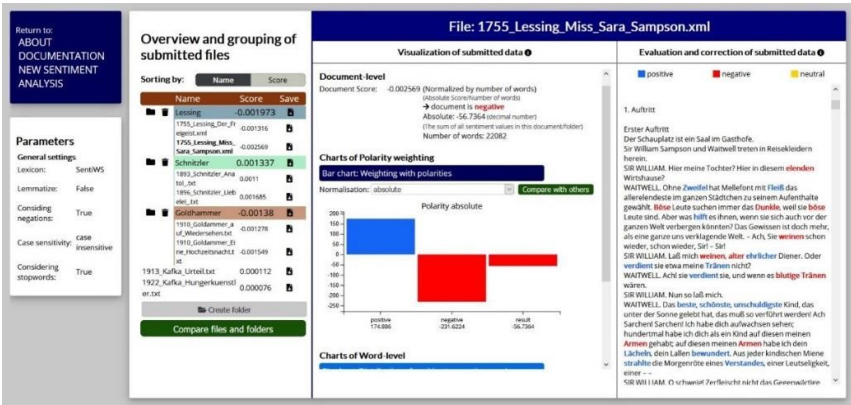


Fig. 2 Results-Screen – Three groups of documents of different authors are analyzed and compared to each other.

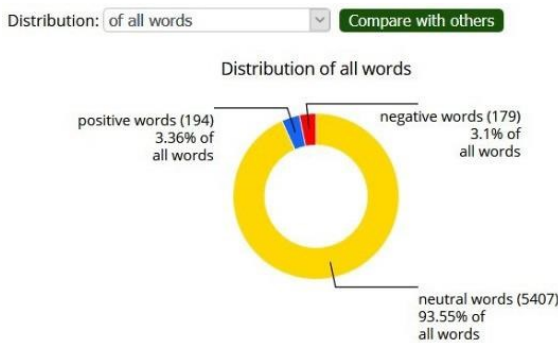


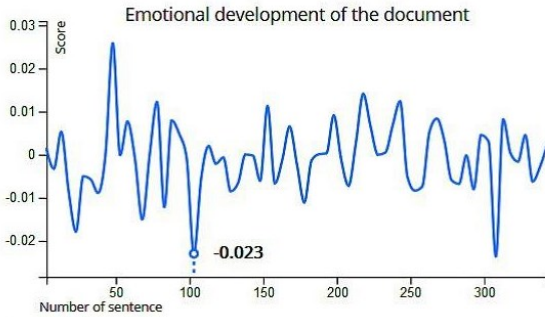
Fig. 3 Pie chart – Distribution of sentiment bearing words



Fig. 4 Word cloud – Strongest negative sentiment bearing words for a German text

Score:

Number of Sentence in each unit:



Current sentence

Number of Words: 68

Absolute score: -1.563

Normalized score: -0.022985

Text:

Zitternd lauscht Alles eine Minute. **Kampf** auf der Straße, **Mord** in den Häusern! eine **Furcht** war in den Allen — mit Ausnahme des Strohmers. Ueber das Haus her schallt das Traben der Reiter, das Klirren der Waffen, das Aechzen und Schreien der Getroffenen. Auf dem Hofe des Juden fangen die Weiber an zu **weinen**, und die Kinder richten neugierige Fragen an ihren stillbetenden Vater — der Strohmer blickt um sich.

Fig. 5 Sentiment progression throughout a text using normalized values and five sentences per data point unit (text: the novel “Auf Wiedersehen!” by Leo Goldhammer)

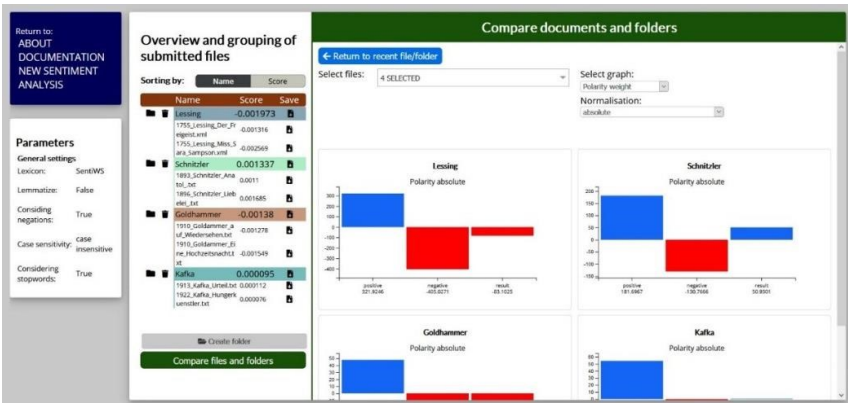


Fig. 6 Comparison of multiple document collections using bar charts

middle area of the page. The tool reports absolute but also normalized values (normalized by the number of tokens of a text) as well as overall values or

SBW-specific values. Users can select to analyze only sentiment bearing words or all words (Fig. 3). The results are shown via various forms of interactive visualization like bar charts, pie charts (Fig. 3) or word clouds (Fig. 4).

Results can be examined at the word as well as at the sentence level. Users can also visualize the sentiment progression throughout a text as a line graph (Fig. 5). If multiple documents are analyzed, visualizations have a “compare with others”-button (Fig. 6). On the right side of the screen (see Fig. 2 above), there is a “close reading” section (Fig. 7). Users can explore the selected text and what words are marked in what way by the lexicon in detail. This section enables users to investigate what words might be marked wrongly for a specific lexicon and how lexicons should be adjusted. Users can also adjust sentiment values of words during their analysis and download all data as CSV- or XML-files for further analysis as well as all visualizations as PNG-files.

Evaluation and correction of submitted data ●

■ positive ■ negative ■ neutral

ersten Becher Weins Ihr brachte, Das weißt du wohl noch selbst: sie fluchte uns Noch **grauenvoller**, als Kriemhild uns fluchte, Und loderte in Flammen auf, wie nie, Seit sie im **Kampf erlag**.

HAGEN.
Sie brauchte Zeit, Um sich hinein zu finden.

GUNTHER.
Als ich sie Nun mahnte, daß sie selbst es Ja geboten, Goß sie den Wein mir ins Gesicht und **lachte**. Wie ich die Menschheit noch nicht **lachen** hörte – Wars so? Sonst straf mich **Lügen!**

HAGEN.
Allerdings, Dann **tote** für immer aus.
GUNTHER.
Ja wohl! So völlig **kurzen** Augenblick. Durch ihren Feuerfluch voraus verzehrt, Denn nur als **Tote** stand sie wieder auf!

HAGEN. Als **Tote**?
GUNTHER.
Ja, obgleich sie ißt und trinkt Und in die Runen stiert. Du hattest **recht**. Nur Siegfried war im Weg.

HAGEN. Ich glaubte – – Nein!
GUNTHER.
Das mildste Wort entlockt ihr nie ein **Lächeln**, Und hätt ichs Volkers frischem Liedermund In einer goldnen Stunde abgefangen, Das **härteste** noch minder eine **Träne**, Sie kennt den **Schmerz** und auch die Lust nicht mehr.

UTE. So ists! Die **alte** Amme deckts nur zu!
GUNTHER.
Stumpf blickt sie drein, als wär ihr Blut vergraben Und

Key: tote
Sentiment of key: -0.1683
Is negated: False
Sentiment result: -0.1683

Fig. 7

Close reading section of the tool (text: “Kriemhilds Rache” by Friedrich Hebbel)

4 Evaluation

We have conducted preliminary small-scale usability tests for our tool after the first development iteration to gain some first insights in possible improvements and the overall quality of the tool. The first usability test we performed was focused on the detection of general user type independent usability issues. We conducted a task-based guerilla usability test (Nielsen, 1994) with ten participants (two female, eight male in the age group from 21 to 30 years). The sample consisted almost entirely of students of various degree programs and none of the participants had experience with sentiment analysis. The usability test was lab-based and conducted in a quiet room with one supervisor. The test consisted of nine tasks of which two were explorative ones. The test involved tasks like searching for specific information, analyzing text according to specific parameters, comparing and exporting results. Users were assigned to “think aloud” and give as much feedback as possible. The tests were recorded, and the supervisor made notes with important feedback during the test.

After analyzing the recordings of the test and the notes taken, we summarized the results by identifying all mentioned usability issues of which we found 21. We ordered these issues according to severity and formulated possible improvements. While we cannot discuss all issues in detail, we want to highlight some of the most important issue groups. A lot of problems can be summarized as missing explanations (e.g., what type of data can be uploaded, how to interpret this score, how to interpret colors). Thus, we integrated more help at fitting usage points and improved upon the documentation. Other issues dealt with an information overload which led us to hide a lot of results at the beginning and make them expandable. Other than that we also identified some basic missing functionality like saving image files by right-click. Despite the usability issues, the average task completion rate (ratio of successfully completed tasks to all tasks) was 89% which points to a tool that is effectively and efficiently to use. We improved upon all identified usability problems in a subsequent development iteration before proceeding to the next usability test.

We conducted the second usability test with four persons with DH or a humanities background (two female, two male) and between 25 and 31 years old. The focus of this test was to gain an overall impression of the tool from the viewpoint of the target group via quantitative parameters. The test

approach was similar to the previous test with the exception that it was performed as remote usability test via Skype. The task success rate was 96% and overall feedback was rather positive. After the test, participants were instructed to fill out a questionnaire: The System Usability Scale (SUS) by Brooke (1996). The SUS is a well-established questionnaire for measuring usability (Bangor et al., 2008) and the tool achieves a score of 92.5 (on a scale from 0 to 100) which is regarded as very good usability (ibid.). While we did not identify major usability issues via this test, we received some feedback for some minor new features.

5 Future work

We argue that *SentText*, in its current state, is a promising and accessible tool for integrating and exploring digital methods in the humanities and the results of the usability tests support this claim. While the sample size for the usability tests is rather small, the results are quite encouraging. We plan to continue our test with more researchers in literary studies to get a better understanding of (1) how and (2) at what steps the tool can best support sentiment analysis projects. Furthermore, we will integrate the usage of this tool in a course about computational literary studies to analyze how this tool can be beneficial for teaching purposes and get feedback from students. Following up, we will continue our next development cycle integrating more lexicons and support for other languages as well as the requirements and features we will acquire with the integration and collaboration of more researchers in literary studies. The application of the user-centered design process, the early integration of the user group as well as the usability tests were highly beneficial for the quality of the current tool and more so reasonable steps due to the special target group.

References

- Alm, C. O., & Sproat, R. (2005). Emotional sequencing and development in fairy tales. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 668–674). Berlin, Heidelberg: Springer.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
- Binali, H., Wu, C., & Potdar, V. (2010). Computational approaches for emotion detection in text. In *4th IEEE International Conference on Digital Ecosystems and Technologies* (pp. 172–177). Piscataway, NJ: IEEE.
- Brooke, J. (1996). SUS – A quick and dirty usability scale. In Jordan, Patrick W. (Ed.): *Usability Evaluation in Industry*. London: Taylor & Francis, pp. 189–194.
- Burghardt, M., & Wolff, C. (2015). Humanist-Computer Interaction: Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik. In *Book of Abstracts. Workshop "Informatik und die Digital Humanities"*, Leipzig.
- Halbhuber, D., Fehle, J., Kalus, A., Seitz, K., Kocur, M., Schmidt, T. & Wolff, C., (2019). The Mood Game – How to use the player’s affective state in a shoot’em up avoiding frustration and boredom. In: Alt, F., Bulling, A., & Döring, T. (Eds.), *Mensch und Computer 2019 – Tagungsband*. New York, NY: ACM Press. DOI: [10.1145/3340764.3345369](https://doi.org/10.1145/3340764.3345369)
- Hartl, P., Fischer, T., Hilzenthaller, A., Kocur, M. & Schmidt, T. (2019). AudienceAR – Utilising Augmented Reality and Emotion Tracking to Address Fear of Speech. In Alt, F., Bulling, A., & Döring, T. (Eds.), *Mensch und Computer 2019 – Tagungsband*. New York, NY: ACM Press. DOI: [10.1145/3340764.3345380](https://doi.org/10.1145/3340764.3345380)
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177), Berlin, Heidelberg: Springer.
- Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 216–225). Reading, UK: Academic Conferences Limited.
- Jannidis, F., Reger, I., Zehe, A., Becker, M., Hettinger, L. & Hotho, A. (2016). Analyzing Features for the Detection of Happy Endings in German Novels. *arXiv preprint [arXiv:1611.09028](https://arxiv.org/abs/1611.09028)*
- Kakkonen, T., & Kakkonen, G. G. (2011). SentiProfiler: creating comparable visual profiles of sentimental content in texts. In *Proceedings of Language Technolo-*

- gies for Digital Humanities and Cultural Heritage* (pp. 62–69). Cham: Springer Nature.
- Kim, E., & Klinger, R. (2018). A survey on sentiment and emotion analysis for computational literary studies. *arXiv preprint* [arXiv:1808.03137](https://arxiv.org/abs/1808.03137)
- Kim, E., & Klinger, R. (2019a). Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters. *arXiv preprint* [arXiv:1903.12453](https://arxiv.org/abs/1903.12453)
- Kim, E., & Klinger, R. (2019b). An Analysis of Emotion Communication Channels in Fan Fiction: Towards Emotional Storytelling. *arXiv preprint* [arXiv:1906.02402](https://arxiv.org/abs/1906.02402)
- Liu, B. (2016). *Sentiment Analysis. Mining Opinions, Sentiments and Emotions*. New York: Cambridge University Press.
- Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27(Febr.), 16–32.
- Mohammad, S. (2011). From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 105–114). Association for Computational Linguistics.
- Mohammad, S. M. (2020). Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons. *arXiv preprint* [arXiv:2011.03492](https://arxiv.org/abs/2011.03492)
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3), 436–465.
- Moßburger, L., Wende, F., Brinkmann, K., & Schmidt, T. (2020). Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum. In *Proceedings of the Fifth Social Media Mining for Health Applications (#SMM4H). Workshop & Shared Task* (pp. 70–81). Association for Computational Linguistics.
- Nalisnick, E. T. & Baird, H. S. (2013). Character-to-character sentiment analysis in shakespeare’s plays. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Vol. 2: Short Papers, pp. 479–483). Association for Computational Linguistics.
- Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4), 161–176.
- Nielsen, J. (1994). Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In Randolph G. Bias (Ed.): *Cost-Justifying Usability* (pp. 245–272). Boston, MA: Academic Press.

- Ortloff, A.-M., Güntner, L., Windl, M., Schmidt, T., Kocur, M., & Wolff, C. (2019). SentiBooks: Enhancing Audiobooks via Affective Computing and Smart Light Bulbs. In: Alt, F., Bulling, A., & Döring, T. (Eds.), *Mensch und Computer 2019 – Tagungsband*. New York, NY: ACM Press. DOI: [10.1145/3340764.3345368](https://doi.org/10.1145/3340764.3345368)
- Pianzola, F., Rebora, S., & Lauer, G. (2020). Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. *PLOS ONE*, 15(1), e0226708.
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1), 31.
- Remus, R., Quasthoff, U., & Heyer, G. (2010, May). SentiWS – A Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 1168–1171). European Language Resources Association.
- Schmidt, T. (2019). Distant Reading Sentiments and Emotions in Historic German Plays. In *Abstract Booklet, DH_Budapest_2019* (pp. 57–60). Budapest, Hungary.
- Schmidt, T., Bauer, M., Habler, F., Heuberger, H., Pilsl, F., & Wolff, C. (2020). Der Einsatz von Distant Reading auf einem Korpus deutschsprachiger Songtexte. In *Book of Abstracts, DHd 2020* (pp. 296–299). Paderborn, Deutschland.
- Schmidt, T., & Burghardt, M. (2018a). An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. In: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 139–149). Santa Fe, NM: Association for Computational Linguistics.
- Schmidt, T., & Burghardt, M. (2018b). Toward a Tool for Sentiment Analysis for German Historic Plays. In: Piotrowski, M. (Ed.), *COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018* (pp. 46–48). Lausanne, Switzerland: Laboratoire laussannois d'informatique et statistique textuelle.
- Schmidt, T., Burghardt, M., & Dennerlein, K. (2018a). Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior. In: S. Kübler, & H. Zinsmeister (Eds.), *Proceedings of the Workshop on Annotation in Digital Humanities (annDH 2018)* (pp. 47–52). Sofia, Bulgaria.
- Schmidt, T., Burghardt, M., & Dennerlein, K. (2018b). „Kann man denn auch nicht lachend sehr ernsthaft sein?“ – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In *Book of Abstracts, DHd 2018*.
- Schmidt, T., Burghardt, M., Dennerlein, K., & Wolff, C. (2019a). Sentiment Annotation in Lessing's Plays: Towards a Language Resource for Sentiment Analysis on

- German Literary Texts. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*. LDK Posters. Leipzig, Germany.
- Schmidt, T., Burghardt, M., Dennerlein, K., & Wolff, C. (2019b). Katharsis – A Tool for Computational Drametrics. In: *Book of Abstracts, Digital Humanities Conference 2019 (DH 2019)*. Utrecht, Netherlands.
- Schmidt, T., Burghardt, M., & Wolff, C. (2018). Herausforderungen für Sentiment Analysis-Verfahren bei literarischen Texten. In: Burghardt, M., & Müller-Birn, C. (Eds.), *INF-DH 2018*. Bonn: Gesellschaft für Informatik e.V.
- Schmidt, T., Burghardt, M. & Wolff, C. (2019). Towards Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing’s Emilia Galotti. In *Proceedings of the DHN (DH in the Nordic Countries) Conference* (pp. 405–414). Copenhagen, Denmark.
- Schmidt, T., Hartl, P., Ramsauer, D., Fischer, T., Hilzenthaller, A., & Wolff, C. (2020). Acquisition and Analysis of a Meme Corpus to Investigate Web Culture. In *Digital Humanities Conference 2020 (DH 2020)*. Virtual Conference.
- Schmidt, T., Jakob, M., & Wolff, C. (2019). Annotator-Centered Design: Towards a Tool for Sentiment and Emotion Annotation. In Draude, C., Lange, M., & Sick, B. (Eds.), *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)* (pp. 77–85). Bonn: Gesellschaft für Informatik. DOI: [10.18420/inf2019_wso8](https://doi.org/10.18420/inf2019_wso8)
- Schmidt, T., Kaindl, F., & Wolff, C. (2020). Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. Riga, Latvia.
- Schmidt, T., Schlindwein, M., Lichtner, K., & Wolff, C. (2020). Investigating the Relationship Between Emotion Recognition Software and Usability Metrics. *i-com*, 19(2), 139–151.
- Schmidt, T., Winterl, B., Maul, M., Scharf, A., Vlad, A., & Wolff, C. (2019). Inter-Rater Agreement and Usability: A Comparative Evaluation of Annotation Tools for Sentiment Annotation. In Draude, C., Lange, M. & Sick, B. (Eds.), *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)* (pp. 121–133). Bonn: Gesellschaft für Informatik. DOI: [10.18420/inf2019_wsi2](https://doi.org/10.18420/inf2019_wsi2)
- Shmueli, B., & Ku, L. W. (2019). SocialNLP EmotionX 2019 Challenge Overview: Predicting Emotions in Spoken Dialogues and Chats. *arXiv preprint* [arXiv:1909.07734](https://arxiv.org/abs/1909.07734)
- Sprugnoli, R., Tonelli, S., Marchetti, A., & Moretti, G. (2016). Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*, 31(4), 762–772.

- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267–307.
- Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(6), 282–292.
- Vo, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538.
- Vredenburg, K., Mao, J. Y., Smith, P. W., & Carey, T. (2002). A survey of user-centered design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 471–478). New York, NY: ACM Press.
- Waltinger, U. (2010). GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation – LREC 2010* (pp. 1638–1642). Dordrecht: Springer.
- Yavuz, M. C. (2021). Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams. In *Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it'20*. Bologna, Italy.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 156–172.
DOI: doi.org/10.5283/epub.44943.

Same same, but different? On the Relation of Information Science and the Digital Humanities

A Scientometric Comparison of Academic Journals Using LDA and Hierarchical Clustering

Manuel Burghardt

Jan Luhmann

Computational Humanities Group
Leipzig University, Germany
burghardt@informatik.uni.leipzig

Computational Humanities Group
Leipzig University, Germany
jan.luhmann@gmx.net

Abstract

In this paper we investigate the relationship of Information Science (IS) and the Digital Humanities (DH) by means of a scientometric comparison of academic journals from the respective disciplines. In order to identify scholarly practices for both disciplines, we apply a recent variant of LDA topic modeling that makes use of additional hierarchical clustering. The results reveal the existence of characteristic topic areas for both IS (*information retrieval, information seeking behavior, scientometrics*) and DH (*computational linguistics, distant reading and digital editions*) that can be used to distinguish them as disciplines in their own right. However, there is also a larger shared area of practices related to *information management* and also a few shared topic clusters that indicate a common ground for – mostly methodological – exchange between the two disciplines.

Keywords: scientometrics; Information Science; Digital Humanities; LDA topic modeling; hierarchical clustering

1 Introduction

In their ISI 2015 paper, Robinson, Priego, and Bawden (2015) discussed if *library and information science* (IS) and the *digital humanities* (DH) might

have a joint future, as there are obviously many connecting factors between both fields. A few years later, the relationship of IS and the DH is explicitly addressed by the conference theme of ISI 2021, as “information science and its neighbors, from data science to digital humanities”. Obvious connections between IS and the DH have been widely described in various related works in the past years. One of the earliest examples can be found in an article by Busa (2004), who summarizes the evolution of the DH and notes that there has been a “documentaristic current” and information-infrastructure focus in the DH from its very beginnings, e.g., with the *American Documentation Society* and the *Deutsche Gesellschaft für Dokumentation* in the 1950s. This documentaristic tradition of the DH is also reflected by a number of publications that highlight the relation of DH and libraries or library studies (Koltay, 2016; Millson-Martula & Gunn, 2017; Sula, 2013). In addition, Terras (2013a) stresses the important role of IS scholars as part of the early DH community, as about 15% of the authors of the 2005 DH conference at the University of Victoria (Canada) are actually coming from the field of library and information science. Along the same lines, Sula and Hill (2019) found that 5.4% of the authors from the journal “Computers and the Humanities” (1966–2004) and 7.4% of the authors from “Literary and Linguistic Computing” (1986–2004) are coming from information and library science. Burghardt et al. (2015) point out structural similarities of IS and the DH and make clear where traditional core topics of Information Science research – including information retrieval, information systems, tool science, user interface design and information behavior – can be transferred to novel research questions and applications in the Digital Humanities. Balck et al. (2015) also identify common, central themes of Digital Humanities and Information Science and argue that *digital* core competencies of Digital Humanities (e.g., indexing, databases, information retrieval) are largely covered in existing Information Science curricula. However, there are also more skeptical voices on the relation of IS and the DH. One of those voices is Gladney (2012, p. 203), who compares various definitions of IS and DH and comes to the conclusion that the DH actually are an “unneeded invention”.

Despite all the overlaps and synergies between IS and the DH, we believe it is crucial to distinguish the two disciplines from each other, as an overly inclusive approach to academic disciplines without explicit demarcations can have its downsides, as is noted by Terras (2013b) in her critique of the “big tent” metaphor in the DH. Terras describes a “crisis of inclusion” in the DH

that leads to a blurring of the boundaries of the discipline and therefore weakens its position as an emerging field in its own right.

“[...] if everyone is a Digital Humanist, then no-one is really a Digital Humanist. The field does not exist if it is all pervasive, too widely spread, or ill defined.”

This criticism can be easily expanded to the relation of IS and DH: If we just see IS as yet another visitor in the big tent DH, the legitimacy of both disciplines is strongly devalued, which in turn has practical, negative effects on the distribution of third-party funds and the development of study programs and professorships for both IS and the DH. The latter is reflected in a current study of German small subjects (“Kleine Fächer”). In this study, the number of IS professorships has more or less stagnated with 10.5 in 2009 and eleven in 2020, whereas the number in the DH has increased strongly from two in 2009 to 24 in 2020. While this increase is positive from the point of view of the DH, the question arises, whether this development might be at the expense of IS in some cases.

In this paper, we try to shed some more light on the scholarly practices of and the relation between IS and the DH. To this end, we present a scientometric study in which we compare a sample of scholarly articles from typical DH journals to articles that were taken from typical IS journals and compare differences and similarities by means of a topic modeling approach using LDA and hierarchical clustering. Using this approach, we hope to answer the following research questions:

- What are typical topics in a corpus of IS and DH research articles?
- How are these topics distributed over the two disciplines in question?

2 Related work

This study is mainly influenced by related work that uses scholarly publications for scientometric analyses of the DH. For example, Gao et al. (2017) visualize co-citation networks and identify citation patterns based on articles published in the journals “Computers and the Humanities / Language Resources and Evaluation”, “Literary and Linguistic Computing/Digital Scholarship in the Humanities” and “Digital Humanities Quarterly” from 1966 to 2016. Tang et al. (2017) present an extensive bibliometric study of a more diverse corpus of DH journal articles, which addresses citation patterns and

their correspondence with article keywords. Weingart (2015, 2016) periodically analyzes trends in the metadata of the DH conferences, focusing on co-authorship and author affiliation as well as critical reflection on diversity in DH. He also identifies trends in research topics, which in this case are not based on topic modeling, but on topic tags that the DH conference authors select for their contribution. A similar quantitative analysis is conducted for example by Sprugnoli et al. (2019) for the Italian conferences AIUCD and CLiC-it from 2014 to 2017. Puschmann and Bastos (2015) present a study of two academic networking platforms, which are related to DH, and analyze user posts based on co-word analysis and topic modeling. They reveal how the discourse on each platform is oriented towards certain sub-areas of DH and towards certain humanities disciplines. A study of Library and Information Science using topic modeling is offered by Figuerola et al. (2017). They collect research papers from the LISA database, show research trends among them and are able to measure the impact of the “movement of digital humanities” on information science, with a rapid increase of humanities-related research starting in 2008.

3 Methodology: LDA and hierarchical clustering

The method used in this study is inspired by related works that have successfully applied topic models to quantify academic journals, for instance Blei and Lafferty (2007), Goldstone and Underwood (2012), Griffiths and Steyvers (2004), Mimno (2012) and Wehrheim (2019). In this section we describe a recent method that combines LDA and hierarchical clustering (Vega-Carrasco et al., 2020), which we adapted from the area of marketing and transaction analysis to the analysis of academic journals. As the method can be considered rather novel for these purposes, we also provide insights from evaluations that accompanied the use of this approach.

3.1 Building the corpus

In order to investigate differences and similarities between IS and the DH, we created a corpus of peer-reviewed publications from both fields. Our corpus (see Tab. 1) comprises the full texts of 6,498 research articles published

between 1990 and 2019 in two journals associated with IS (28 M tokens total) and three journals associated with the DH (15 M tokens total). The selected journals are well-established in the respective disciplines and can be considered top tier.

Table 1: Corpus overview

Discipline	Journal	Time	Articles		
			Decade	Number of articles	Number of tokens
Digital Humanities	<i>Language Resources and Evaluation</i> (LRE; formerly known as <i>Computers and the Humanities</i>)	1990–2019	1990–1999	329	1,746,825
			2000–2009	236	1,283,827
			2010–2019	339	2,647,701
	<i>Digital Scholarship in the Humanities</i> (DSH; formerly known as <i>Literary and Linguistic Computing</i>)	1990–2019	1990–1999	283	1,334,671
			2000–2009	381	1,988,258
			2010–2019	558	3,267,314
	<i>Digital Humanities Quarterly</i> (DHQ)	2007–2019	1990–1999	–	–
			2000–2009	49	363,871
			2010–2019	338	2,534,583
Information Science	<i>Journal of the Association for Information Science and Technology</i> (JASIST)	1990–2019	1990–1999	252	1,600,561
			2000–2009	1588	10,537,002
			2010–2019	1222	9,298,150
	<i>Journal of Documentation</i> (J. Doc)	1990–2019	1990–1999	174	1,141,250
			2000–2009	327	2,260,000
			2010–2019	422	3,357,853

Since we will treat these articles as a sample of the research literature of both disciplines, we must bear in mind that these journals can represent only a fraction of the international discourse which is predominantly oriented towards North American and Western European research, and largely excludes scholarship across the Global South. Also, regarding the many flavors and sub-areas of Digital Humanities research (Huggett, 2012; Sahle, 2015; Burgardt, 2020), our sample of articles is probably biased toward literary computing and computational and corpus linguistics, given the traditions of LRE and DSH.

We obtain most of the above research articles in PDF format via URLs provided by the text mining services of CrossRef¹, except for articles of DSH, which we obtain from *Oxford University Press*², and articles of DHQ, which are available as XML straight away³. All PDFs are converted to XML using *Grobid*⁴. In addition to the fulltexts we also collect metadata (e.g., author, title, year) for all of the articles. All texts are tokenized, POS-tagged, lowercased and lemmatized using *spaCy*⁵. To improve lemmatization, especially of conjugated verbs, we use the *LemmInflect* dictionary⁶, but keep original *spaCy* lemmata for out-of-vocabulary terms. We applied some manual corrections for a couple of nouns (*humanities*, *linguistics*, *data*, *media*, etc.) that can be considered plural-only nouns or uncountable in the domain context and which *spaCy* would otherwise reduce to a singular form. To remove texts from the corpus which are not original research articles, but contain, e.g., reviews or organizational matters, we only keep texts which contain at least 1,000 tokens and whose authors are included in the metadata. We also removed any documents with generic titles such as “Editorial”, “News and notes”, “Book review”, etc.

3.2 Building a stabilized, aggregated topic model

Topic modeling is an unsupervised machine learning technique for inferring a set of latent semantic topics from a large collection of documents. For our study we use the popular Latent Dirichlet Allocation (LDA; Blei et al., 2003) with collapsed Gibbs sampling, implemented in the MALLET toolkit (McCallum, 2002). Since LDA is a probabilistic, non-deterministic algorithm that involves random initializations and random sampling, multiple LDA runs on the same data and the same parameters may lead to different results. Furthermore, results are heavily depending on parameter settings (number of topics, alpha and beta priors). To tackle this issue and obtain a stable and robust model of topics, we decided to aggregate topics from several LDA

1 <https://www.crossref.org/>

2 <https://academic.oup.com/>

3 <http://www.digitalhumanities.org/>

4 <https://github.com/kermitt2/grobid>

5 <https://spacy.io/>

6 <https://github.com/bjascob/LemmInflect/>

runs (Blair et al., 2016, 2020) by means of hierarchical clustering (with an agglomerative approach), as proposed by Vega-Carrasco et al. (2020). We justify this approach by evaluating the quality of the clustered topic model in comparison to individual LDA models. The metrics used for evaluation are:

1. **perplexity** on held-out documents to assess the generalizability of a model, estimated by the “iterated pseudo-count” method described by Wallach et al. (2009);
2. **topic coherence** (Aletas & Stevenson, 2013; Mimno et al., 2011), defined as the average value of pairwise collocation probabilities of the top 15 terms of a topic, based on Normalized Pointwise Mutual Information calculated on collocations in our corpus with a window size of 5L and 5R;
3. **topic distinctiveness** (Vega-Carrasco et al., 2020), defined as the minimum of the cosine distances a topic shows to all other topics within a model;
4. **topic stability** (Greene et al., 2014) between two models; utilizing the Hungarian algorithm to find an optimal alignment between the sets of topics of two models based on pairwise cosine distances, topic stability is then defined as the average value of cosine distances of all aligned topic pairs⁷.

The approach proposed by Vega-Carrasco et al. (2020), in a first step, involves the identification of a plausible number of topics for LDA models based on the afore-mentioned metrics. In a second step, a series of LDA runs with this number of topics are performed. We leave the optimization of alpha and beta priors to MALLETT. The resulting topics of the LDA model series are then merged by hierarchical clustering, based on cosine distance and average linkage. To obtain a final model of clustered topics, two parameters have to be determined: a cosine distance threshold below which topics are merged during the clustering process and a minimum cluster size at which we consider a topic cluster to be included in our aggregated model. Cluster

7 Please note that the cosine distance between topics refers to the cosine distance between the term probability distributions of topics. Our calculation of the topic stability differs from Greene et al. (2014) in that they calculate Average Jaccard distance of the sets of top terms of topics instead of cosine distance. However, the cosine distance outperformed other measures of topic distance when compared to human judgment (Aletas & Stevenson, 2014) and when used for topic matching (Niekler & Jähnichen, 2012).

size is also referred to as *recurrence* by Vega-Carrasco et al. (2020), reflecting the reappearance and therefore stability of a topic among several topic models.

3.3 Preprocessing and sampling

For reasons of efficiency (Martin & Johnson, 2015) we consider nouns and proper nouns as features, on which we perform LDA. We additionally concatenate multi-word units (2- to 5-grams) which are extracted using the NPMI association measure (Bouma, 2009), with ≥ 0.4 being reported to be a suitable threshold. Multi-word units are further filtered to contain at least one noun or proper noun, considering it a shallow approach to extracting noun phrases. These are also included as features. Finally, all terms occurring in less than 1% of texts are excluded, resulting in 6,721 terms.

Since the number of articles varies greatly between journals and decades, each of the following LDA runs is performed with a random subsample of 2,400 texts, stratified by discipline, decade and journal. For each decade, this subsample contains 800 texts, 400 for IS and 400 for DH. These 400 texts per discipline texts are being distributed as evenly as possible among the different journals of the discipline. We consider the resampling of texts on which individual topic models are fitted as an additional way to ensure the generalizability and robustness of our aggregated model. We additionally define a set of 240 test documents, which are stratified in the same way. These documents are never used for LDA training, but only for the calculation of perplexity.

3.4 Estimating the number of topics

In a first step we aim to determine a plausible number of topics. We run LDA using MALLET for a number of 50, 100, 150 and 200 topics, ten runs each. For all the resulting topic models we calculate the above-mentioned metrics (see Fig. 1). Although the perplexity values may suggest the use of a topic model with 200 or more topics, we chose to perform LDA using 100 topics. The 100-topic models show a significant decrease in perplexity, while they perform much better on topic coherence and stability than LDA models with a larger number of topics.

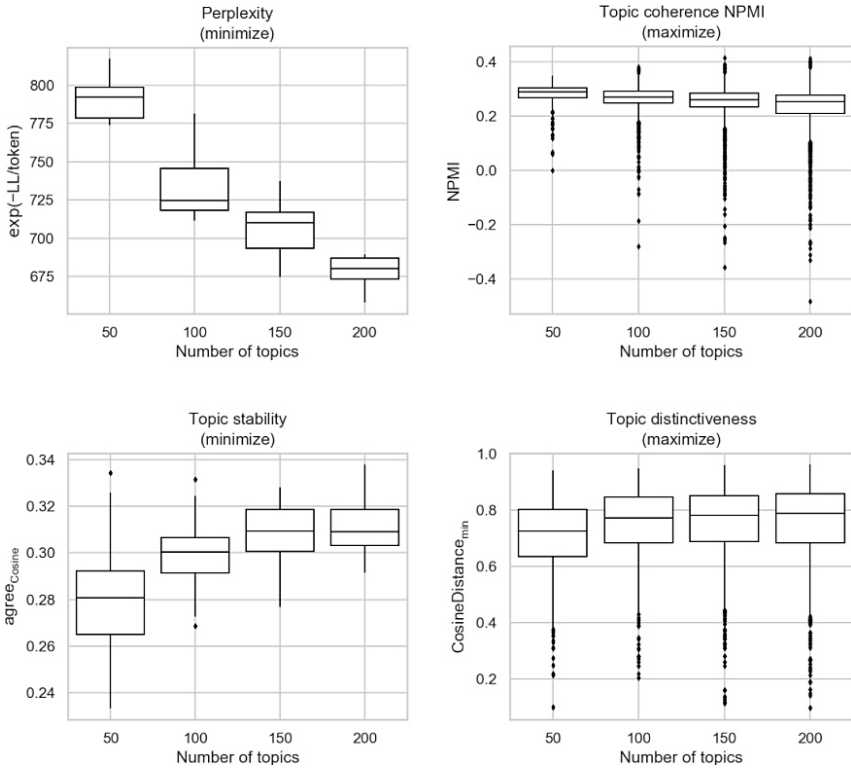


Fig. 1 Evaluation metrics of LDA models with 50, 100, 150 and 200 topics

3.5 Hierarchical clustering of topics

For topic clustering we apply a hierarchical agglomerative clustering algorithm with average linkage and cosine distance metric, providing a distance threshold for the merging of topics. To estimate in which range this threshold should be, we manually annotate 250 randomly selected topic pairs according to their estimated similarity, judging from the top 20 terms of each topic. Figure 2 shows the correlation of annotated topic similarity to cosine distances.

A cosine distance ≤ 0.2 is mostly associated with highly similar topic pairs, while a cosine distance ≥ 0.5 corresponds to dissimilar topic pairs. From this we conclude that a reasonable cosine distance threshold is in the range 0.2 to 0.5. An aggregated topic model is based on topics derived from multiple LDA topic models. In this case, we use 20 individual LDA topic models with 100 topics, meaning we perform a clustering of 2,000 topics. Each

of the 20 models is generated using MALLET, with a hyperparameter optimization of alpha and beta priors every 20th iteration, and a burn-in phase of 50 iterations. The resulting 2,000 topics are then clustered, based on the cosine distance of their term probabilities, once for each distance threshold from 0.25, 0.35 to 0.45. The term probabilities of a topic cluster (or clustered topic) are calculated as the average of the term probabilities of the topics which are part of the cluster. As mentioned above, the size of a topic cluster can be seen as the recurrence of highly similar topics among LDA models. A second parameter is the minimum cluster size at which we include a cluster in our model. To select these two parameters and to evaluate aggregated models, we repeat the entire procedure five times.

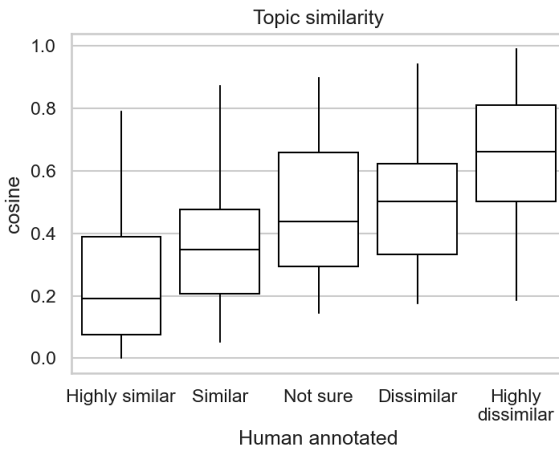


Fig. 2 Human-annotated topic similarity compared to cosine distance

Figure 3 shows the evaluation results of perplexity, topic coherence, topic stability and topic distinctiveness values of the aggregated models, for varying values of cosine distance threshold and minimum cluster size. The average values achieved by the original individual LDA models are displayed as black horizontal lines, one standard deviation as dashed lines. In terms of topic coherence and topic stability, the clustered models clearly outperform the basic topic models, especially if setting a minimum cluster size between 9 and 18.⁸ A distance threshold of 0.25 performs slightly better on topic

8 It should be noted at this point that resampling the articles for each LDA run most likely has a negative impact on topic stability among the LDA models, while the aggregated models may have an advantage. This should be evaluated in the future.

coherence, a threshold of 0.45 does slightly better on topic stability and a threshold of 0.35 seems to be a good middle ground for both measures. However, in terms of perplexity, the aggregated models using a distance threshold of 0.25 and a minimum cluster size > 8 perform far worse than original LDA models. This also applies to the models using a distance threshold of 0.35 and minimum cluster size > 10 , and those using a distance threshold of 0.45 and minimum cluster size > 12 . Regarding topic distinctiveness, the models with 0.25 distance threshold perform best and also better than original LDA models, but only if setting a minimum cluster size > 11 . For these parameter settings however, as we have already noted, the perplexity values are far from acceptable. We finally decide on using a distance threshold of 0.45 and a minimum cluster size of 12. The aggregated models using these parameter settings show a comparable perplexity and only marginally lower topic distinctiveness as the original LDA models, and outperform them by far in terms of topic coherence and topic stability. Among the five aggregated topic models we obtained, the one showing the best evaluation results is selected.

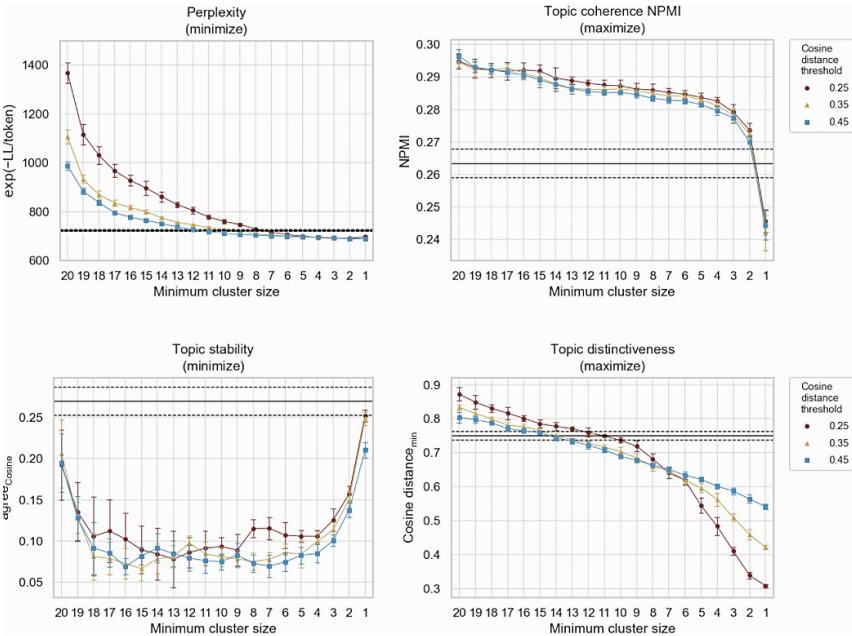


Fig. 3 Evaluation results of aggregated topic models. Horizontal lines show the mean value and standard deviations achieved by standard LDA models.

Figure 4 shows a significant difference in topic coherence values of this model compared to the LDA models it is aggregated from. In order to be able to post-hoc calculate average topic probabilities with respect to individual publication years, journals or disciplines, we require the topic distributions of all documents in our corpus, based on our newly aggregated model. We infer these using the collapsed Gibbs sampler, where we fix the already known topic term probabilities matrix and priors.

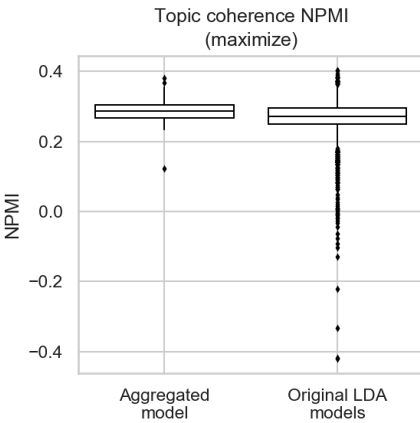


Fig. 4 Distribution of topic coherence values of the final aggregated model vs. the original LDA models

4 Results and discussion

As a result of our clustered topic model, we obtain a total number of 87 topic clusters.⁹ These are numbered by their initial cluster index. All clusters with a size < 12 were filtered out. To display the latent semantic content of a topic, we determine the most relevant terms of each topic cluster by applying the term relevance formula proposed by Sievert and Shirley (2014) with weight parameter $\lambda = 0.8$. They show that the top terms of the resulting ranking achieve a better interpretability to humans than simply displaying the most

9 We provide a list of these topic clusters together with their most relevant terms, their actual cluster size (i.e., the number of merged topics), the coherence score as well as average probability for both DH and IS in the Appendix of this article.

probable terms of a topic. The document topic distributions allow us to observe topic probabilities of higher-level groupings of documents. We define the topic probabilities for a given discipline as the weighted average of the mean values corresponding to decade and journal strata of the discipline (for details on stratification and sampling, see 3.3). The mean topic probabilities of a stratum are calculated as the mean values of the topic probabilities of the documents within this stratum.

Figure 5 shows an overview of the distribution of the mean probabilities among IS and DH for all topics. The topics are sorted by relative difference between their discipline-specific probability.

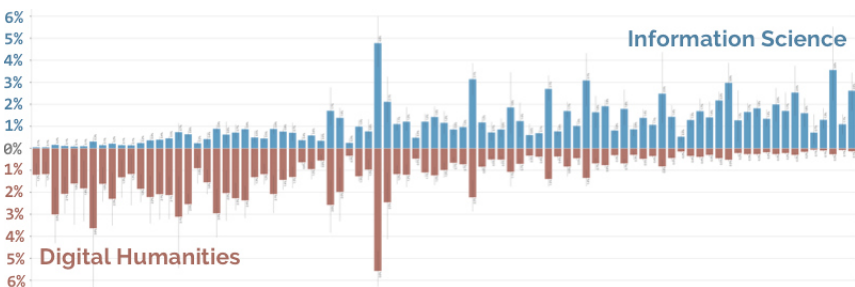


Fig. 5 Distribution of topic clusters among IS and the DH

On the left and right side, the plot clearly shows that there are a number of topic clusters that are rather characteristic for either IS (blue) or the DH (red). In the middle, there are some shared topic clusters that occur in both disciplines.

Figure 6 provides a spatial overview of all topic clusters. It shows a 2D projection of the topic distances based on the transposed document topic probability matrix. This matrix was reduced using singular value decomposition to a dimension of 100, then projected to two dimensions using UMAP (McInnes, 2018). The size of the topic clusters depends on the overall average topic probability. The color of the topic clusters indicates whether they show a statistically significant higher probability for a certain discipline. This was tested via a two-sided Mann-Whitney rank sum test on stratified document topic probability values. If a topic shows no significance for either discipline, it is displayed grey. If it is significant with $p < 0.01$, the intensity of the color depends on the value of the effect size R (rank biserial correlation). The position of topic clusters indicates co-occurrence with other topic clusters in the corpus, and thus allows us to identify superordinate cluster structures, which

we call *topic areas*. In the following we discuss how the topic clusters of Figure 6 can be interpreted to shed some light onto distinctive and overlapping scholarly practices in IS and the DH.

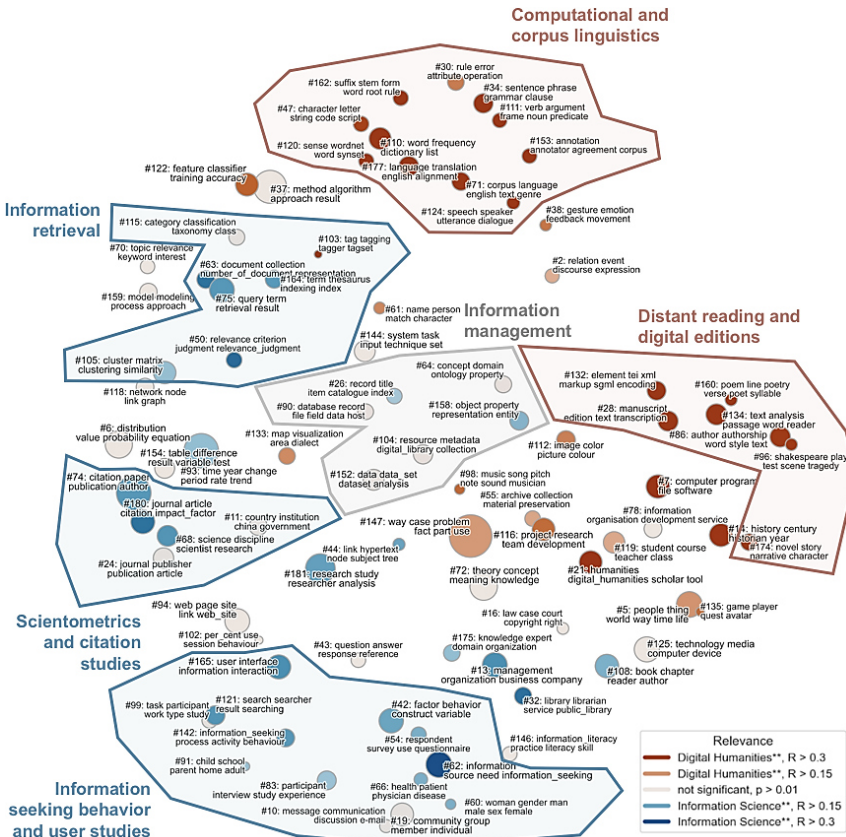


Fig. 6 Spatial overview of all topic clusters (colored bubbles). The size of the topic clusters depends on the overall average topic probability; color indicates whether a topic shows a statistically significant higher probability (Mann-Whitney rank sum test) for IS (blue) or DH (red). Topic clusters that form larger thematic areas are displayed as regions (e.g., “information retrieval”, in the top-left).

4.1 General observations on topic clusters

The previous topic clustering approach has resulted in mostly meaningful topics. Only two topic clusters (#5, #147) exclusively contain generic terms

that cannot be used to gain insight about disciplinary characteristics of IS and the DH. A few other topic clusters use more specific terms but eventually describes rather generic concepts, for example: *law* (#16), *time* (#93) or *publishing* (#24). Most of the topic clusters however are very revealing with regard to scholarly practices of IS and the DH, allowing us to show scholarly differences and similarities between the two disciplines.

The following bar charts (Figs. 7–12) show discipline-specific topic probabilities and the confidence interval based on decade-journal strata mean values as error bars. These charts also contain basic diachronic information in the form of a miniature line diagram to the left of the bar chart. These show the trend in topic probability for each discipline (red for DH, blue for IS) for the decades 1990s, 2000s and 2010s from left to right.¹⁰

4.2 IS topic clusters

Figure 6 shows three large topic areas that can be centrally assigned to IS, i.e., these topics play no significant role in DH publications. The largest area here is *information seeking behavior* and the closely related method of *user studies*. Figure 7 shows that this topic area clearly belongs to IS. We find a similar picture for the topic area of *scientometrics* and *citation studies*, which traditionally has been part of the IS research agenda and does not really seem to be contested by the DH at all (see Fig. 8).

Interestingly, things look similar for the topic area of *information retrieval* (IR; see Fig. 9), which – with its many techniques for text and data analysis – one might have expected to play an important role in the DH too. However, it seems that IR remains a central topic area for the IS and the DH rather seem to have developed their own IR-oriented practices for text analysis. We will see below that the DH indeed seem to borrow many of their text analytics methods – which often involve some form of machine learning – from corpus and computer linguistics. It is also interesting to see that the IR-related topic cluster on *question answering* (#43) is classified as a rather neutral topic that is neither explicitly addressed by IS nor by the DH (see Fig. 6).

10 We do provide these trends as an extra source of information but do not actually discuss diachronic developments in this paper, as we were not able to detect any meaningful trends in the data. For a more comprehensive diachronic perspective, the corpus probably should cover more than just three decades.

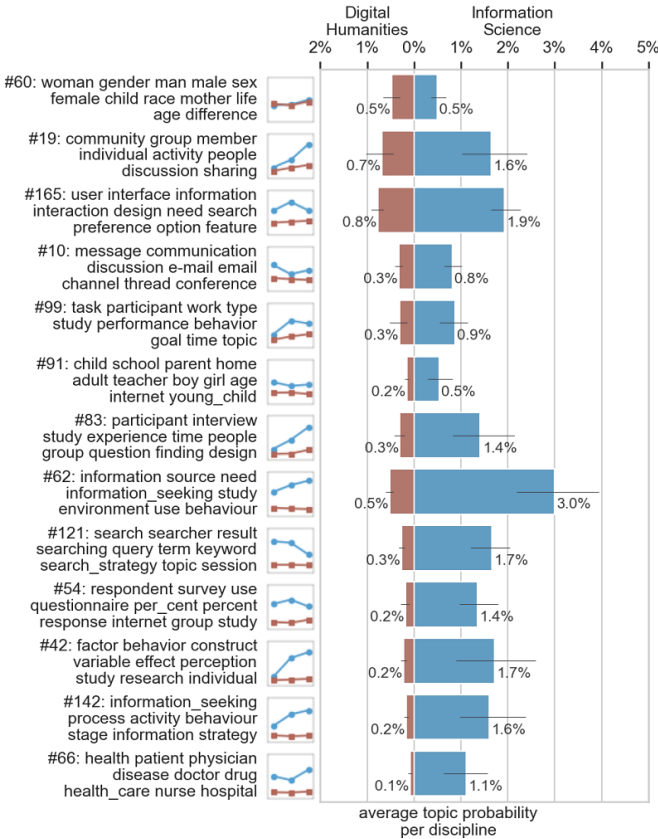


Fig. 7 Topic area: Information seeking behavior and user studies

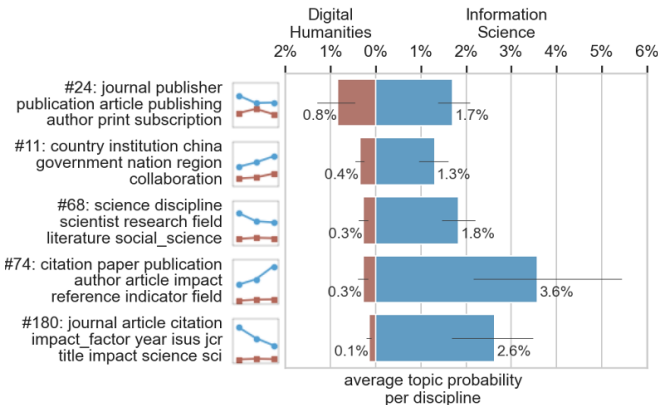


Fig. 8 Topic area: Scientometrics and citation studies

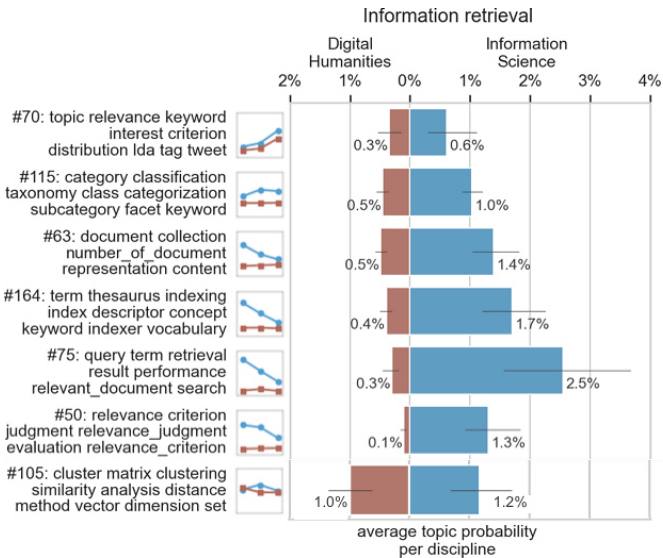


Fig. 9 Topic area: Information retrieval

Apart from these major areas of IS research practices, only a few topic clusters remain that can be clearly assigned to IS. These include a generic *hypertext* topic (#44) as well as *knowledge management* (#175), but also two topic clusters *library* (#32) and *book* (#108) that reflect the *library science* aspect of IS (#32, #108). The two topic clusters *research* (#181) and *business* (#13) illustrate that IS has both an academic, research-oriented perspective, but also a more applied, business-related perspective with industry co-operations etc.

4.3 DH topic clusters

As is the case for IS, the DH too have larger topic areas that are characteristic for the discipline. The DH seem to be strongly influenced by methods and practices from *computational* and *corpus linguistics* (see Fig. 10), which – then again – is not too surprising, as the DH traditionally have had a strong focus on *text*. This focus is also reflected in respective DH journals, which – apart from DHQ – often have a heavy bias toward *linguistics*.¹¹ The actually

11 We plan to include proceedings of DH-related conferences and workshops for future analyses, as they represent a much broader range of topics and practices in the DH.

interesting thing about this, is that IS has only a very small share of the linguistic clusters, although it has been claimed that IS has many connecting factors to linguistics (Montgomery, 1972, p. 195; Engerer, 2012).

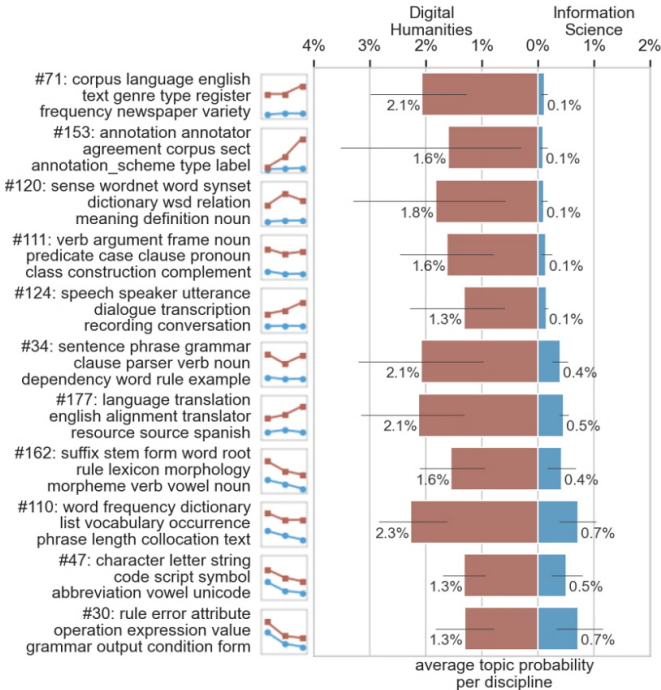


Fig. 10 Topic area: Computational and corpus linguistics

Another large topic area in the DH can be paraphrased as *distant reading* (Moretti, 2000) and *digital editions* (see Fig. 11), entailing the encoding and computational analysis of text documents with a focus on literary rather than on linguistic studies. Digital editions are represented by a topic cluster on manuscript / edition / transcript (#28) as well as by a cluster on TEI / XML / markup (#132). It is also worth mentioning that Shakespeare's plays seem to constitute a topic cluster of its own (#96), underpinning the status of the "Bard" as the *drosophila melanogaster* of the DH. Another topic cluster within this larger area of distant reading that should be highlighted is the method of *stylometry* and *authorship attribution* (#96), which is a frequently used method throughout the DH. The topic cluster on machine learning and

classification (#122) is an important methodological foundation for almost any computational analysis in the humanities.

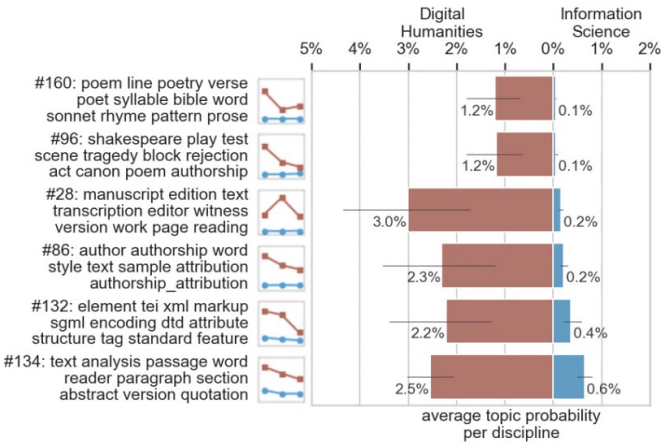


Fig. 11 Topic area: Distant reading and digital editions

Although having a strong focus on text, the DH are not restricted to text-based media, but also take into account *images* (#112) and *music* (#98) as well as *multimodal* genres (#38) such as *games* (#135). While being centered around linguistics and literary studies, the DH also address a wider spectrum of humanities disciplines, including *history* (#14), *geography* (#133) and *translation studies* (#177).

A number of clusters revolve around methodological topics such as named *entity recognition* (#61), *gesture recognition* (#38), and *event and discourse detection* (#2). A highly relevant topic cluster is about rather generic terms, such as *computer*, *program* and *software* (#7). While this may seem odd at first sight, it is actually very revealing for the DH, as they will often use these generic IT terms in order to highlight a fundamental epistemological shift in the humanities. Another fundamental characteristic of the DH is revealed by the *collaboration* (#116) topic cluster: Collaborative research will be the norm in many scientific disciplines, including IS, but it is not for many humanities disciplines. However, with the rise of digital tools and the general interdisciplinary nature of the DH, Digital Humanities have to be much more open to collaborative research practices than traditional humanists, implying another epistemological shift. Another topic cluster in the DH is dedicated to *teaching* (#119). Similar to the two previous topic clusters, this is very telling about the nature of the DH, who not only reflect upon the chances of digital

tools and methods for their research practices, but also for their teaching. One of the most prominent topic clusters for DH actually is dedicated to the reflection and discussion of the *digital humanities as a discipline* (#21), which illustrates the coming-of-age of a young discipline that is still trying to position itself in an interdisciplinary “big tent” setting (see also Terras et al., 2013).

4.4 Shared topic clusters

The previous sections highlighted characteristic topic clusters of IS and the DH. Here, we will highlight some of the topic clusters that are shared by both disciplines. There seems to be a larger topic area that we labeled *information management and library studies* (see Fig. 12). Although some of the topic clusters (#26, #158) here seem to be more contested by IS, more general activities such as *management of resources*, *data and meta data* as well as *object representation* also seem to be important for the DH.

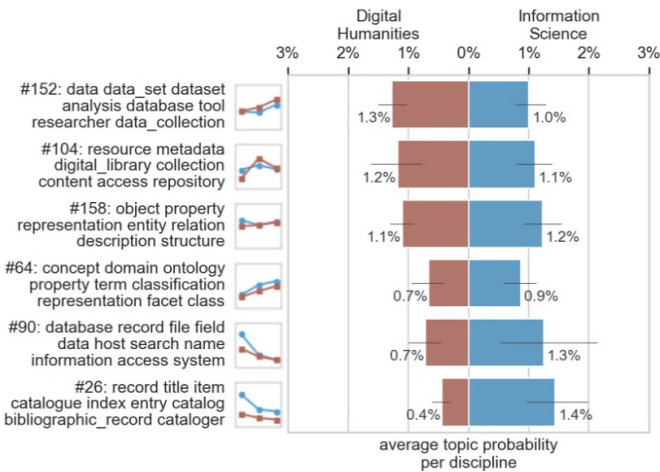


Fig. 12 Topic area: Information management and library studies

While we have seen that both IS and the DH have their specific methods, there are also some shared clusters on the methodological level. These include established *machine learning algorithms* for pattern recognition and information extraction (37) and *statistical tests* (#6, #154), but also techniques for *network analysis* (#118) and *topic modeling* (#70). Besides such methodological topic clusters, we also find two clusters that are rather on the concep-

tual level, as they represent *modeling practices* (#159) as well as *epistemological theory* (#72). Further shared topic clusters are rather general and focus for instance on the *web* as a resource (#94), *digitization and new media* (#125), *software systems and frameworks* (#144), *information institutions* (#78) and *information literacy* (#146) in general.

5 Conclusion

We performed a scientometric analysis of typical IS and DH journals from three decades, using a recent method that combines LDA with hierarchical clustering. The data on which Vega-Carrasco et al. (2020) originally developed and evaluated this novel method are transactions in the grocery retail market, which are treated as bags-of-products. In this paper, we have adapted the method back into the context of text analysis (bags-of-words) and were able to prove its success by means of various evaluation metrics.

The goal of this study was to shed some more light on the relation of IS and DH, which have been said to have a number of overlaps and similarities (see introduction). With the above method we were able to identify topic clusters that are characteristic for both disciplines: IS has three major research areas (*information behavior*, *information retrieval* and *scientometrics*) that are uncontested by the DH. While IS is dedicated to various aspects of libraries, the DH seem to address other cultural heritage institutions, such as archives and collections. The DH is still very heavily focused on text analysis and digital editions, borrowing many concepts and methods from linguistics and literary studies. However, there are also characteristic topic clusters in the DH that indicate that they also deal with other humanities disciplines and also with research objects that go beyond text.

Our results suggest that the generally expected overlap between IS and DH (see, e.g., Sula, 2013; Robinson et al., 2015; Burghardt et al., 2015) seems to be mostly in the broad area of *information management* and on the methodological level, whereby it must be stressed that both IS and DH also have clearly distinguishable research methods and practices. All in all, our results indicate that despite rather occasional overlaps, there is enough uncontested space for both IS and the DH to thrive as individual disciplines and to further develop unique research agendas and study programs.

References

- Aletras, N., & Stevenson, M. (2013). Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)* (pp. 13–22). Red Hook, NY: Curran.
- Aletras, N., & Stevenson, M. (2014). Measuring the similarity between automatically generated topics. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 2: Short Papers, pp. 22–27). Red Hook, NY: Curran.
- Balck, S., Büttner, S., Ducks, D., Lehfeld, A.-S., Schneider, E., & Vietze, E. (2015). Mit den Informationswissenschaften von Daten zu Erkenntnissen. In *DHd 2015*, Graz.
- Blair, S. J., Bi, Y., & Mulvenna, M. D. (2016). Increasing topic coherence by aggregating topic models. In *International Conference on Knowledge Science, Engineering and Management* (pp. 69–81). Cham: Springer Nature.
- Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, 50(1), 138–156.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17–35.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In C. Chiarcos, R. E. de Castilho, & M. Stede (Eds.): *Von der Form zur Bedeutung: Texte automatisch verarbeiten. Proceedings of the Biennial GSCL Conference 2009* (pp. 31–40). Tübingen: Narr.
- Burghardt, M. (2020). Theorie und Digital Humanities – Eine Bestandsaufnahme. AG Digital Humanities Theorie (Blog). <https://dhtheorien.hypotheses.org/680>
- Burghardt, M., Wolff, C. & Womser-Hacker, C. (2015). Informationswissenschaft und Digital Humanities. In: *Information – Wissenschaft & Praxis*, 66(5–6): 287–294.
- Busa, R. A. (2004). Foreword: Perspectives on the Digital Humanities. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.): *A Companion to Digital Humanities*. Oxford: Blackwell.
- Engerer, V. (2012). Informationswissenschaft und Linguistik: Kurze Geschichte eines fruchtbaren interdisziplinären Verhältnisses in drei Akten. *International Journal for Language Data Processing*, 36(2), 71–91.

- Figuerola, C. G., Marco, F. J. G., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics*, 112(3), 1507–1535.
- Gao, J., Duke-Williams, O., Mahony, S., Ramdarshan Bold, M., & Nyhan, J. (2017). The Intellectual Structure of Digital Humanities: An Author Co-Citation Analysis. In *Digital Humanities 2017. Alliance of Digital Humanities Organizations (ADHO)*.
- Gladney, H. M. (2012). Long-term digital preservation: A digital humanities topic? *Historical Social Research*, 37(3), 201–217.
- Goldstone, A., & Underwood, T. (2012). What can topic models of PMLA teach us about the history of literary scholarship. *Journal of Digital Humanities*, 2(1), 39–48.
- Greene, D., O’Callaghan, D., & Cunningham, P. (2014, September). How many topics? Stability analysis for topic models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 498–513). Berlin, Heidelberg: Springer.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. In *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228–5235.
- Huggett, J. (2012). Core or Periphery? Digital Humanities from an Archaeological Perspective. *Historical Social Research / Historische Sozialforschung*, 37(3), 86–105.
- Koltay, T. (2016). Library and information science and the digital humanities. *Journal of Documentation*, 72(4), 781–792.
- Martin, F., & Johnson, M. (2015). More efficient topic modelling through a noun only approach. In *Proceedings of the Australasian Language Technology Association Workshop 2015* (pp. 111–115). Red Hook, NY: Curran.
- McCallum, A. K. (2002). MALLETT: A machine learning for language toolkit. <http://www.cs.umass.edu/~mccallum/mallet>
- McInnes et al., (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861.
- Millson-Martula, C., & Gunn, K. (Eds.) (2017). *The Digital Humanities: Implications for Librarians, Libraries, and Librarianship*. London, New York: Routledge.
- Mimno, D. (2012). Computational historiography: Data mining in a century of classics journals. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(1), 1–19.
- Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Red Hook, NY: Curran.

- Montgomery, C. A. (1972). Linguistics and Information Science. *Journal of the American Society for Information Science*, 23(3), 195–219.
- Moretti, F. (2000). Conjectures on world literature. *New Left Review*, 1(4), 54–68.
- Niekler, A., & Jähnichen, P. (2012). Matching results of latent dirichlet allocation for text. In *Proceedings of ICCM 2012, 11th International Conference on Cognitive Modeling* (pp. 317–322). Berlin: Universitätsverlag der TU Berlin.
- Puschmann, C., & Bastos, M. (2015). How digital are the digital humanities? An analysis of two scholarly blogging platforms. *PLOS ONE*, 10(2), e0115035.
- Robinson, L., Priego, E., & Bawden, D. (2015). Library and Information Science and Digital Humanities: Two Disciplines, Joint Future? *Proceedings of the 14th International Symposium on Information Science, ISI 2015* (pp. 44–54). Glückstadt: Verlag Werner Hülsbusch.
- Sahle, P. (2015). Digital Humanities? Gibt's doch gar nicht! In Baum, C. & Stäcker, T. (Eds.), *Grenzen und Möglichkeiten der Digital Humanities* (Sonderband der ZfdG, 1).
- Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Association for Computational Linguistics.
- Sprugnoli, R., Pardelli, G., Boschetti, F., & Del Gratta, R. (2019). Un'Analisi Multi-dimensionale della Ricerca Italiana nel Campo delle Digital Humanities e della Linguistica Computazionale. *Umanistica Digitale*, 3(5).
- Sula, C. A. (2013). Digital humanities and libraries: A conceptual model. *Journal of Library Administration*, 53(1), 10–26.
- Sula, C. A., & Hill, H. V. (2019). The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). *Digital Scholarship in the Humanities*, 34(1), i190–i206.
- Tang, M. C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics*, 113(2), 985–1008.
- Terras, M. (2013a). Disciplined: Using Educational Studies to Analyse “Humanities Computing.” In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.), *Defining the Digital Humanities – A Reader* (pp. 67–96). Farnham, UK: Ashgate Publishing.
- Terras, M. (2013b). Peering Inside the Big Tent. In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.), *Defining the Digital Humanities – A Reader* (pp. 263–270). Farnham, UK: Ashgate Publishing.
- Terras, M., Nyhan J., & Vanhoutte, E. (2013). *Defining Digital Humanities: A Reader*. Farnham, UK: Ashgate Publishing.

- Vega-Carrasco, M., O’Sullivan, J., Prior, R., Manolopoulou, I., & Musolesi, M. (2020). Modelling Grocery Retail Topic Distributions: Evaluation, Interpretability and Stability. *arXiv:2005.10125*.
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 1105–1112). New York, NY: ACM Press.
- Wehrheim, L. (2019). Economic history goes digital: topic modeling the Journal of Economic History. *Cliometrica*, 13(1), 83–125.
- Weingart, S. B. (2015). Submissions to DH2016 (Pt. 1). <http://scottbot.net/submissions-to-dh2016pt-1/>
- Weingart, S. B. (2016). Submissions to DH2017 (Pt. 1). <http://scottbot.net/submissions-to-dh2017pt-1/>

Appendix

Table A1: All topic clusters in our model, their most relevant terms, cluster size, topic coherence (NPMI), and mean topic probability by discipline

No.	Most relevant terms by log-lift formula (Sievert & Shirley, 2014), $\lambda = 0.8$	Cluster size	Topic coherence	Mean prob. DH	Mean prob. IS
2	relation event discourse expression ontology entity	13	0.272	0.96%	0.59%
5	people thing world way time life idea sense day experience	12	0.266	2.58%	1.72%
6	distribution value probability equation function parameter	23	0.286	0.83%	2.50%
7	computer program file software machine computing version	24	0.26	3.11%	0.74%
10	message communication discussion e-mail channel	13	0.281	0.32%	0.81%
11	country institution china government nation region collaboration	17	0.31	0.35%	1.30%
13	management organization business company manager technology	16	0.295	0.46%	2.18%
14	history century historian year period war work time past	17	0.261	2.96%	0.90%
16	law case court copyright right report act authority	16	0.275	0.39%	0.70%
19	community group member individual activity people discussion	14	0.29	0.69%	1.65%
21	humanities digital_humanities scholar tool humanities_computing	21	0.329	3.63%	0.32%
24	journal publisher publication article publishing author	20	0.323	0.84%	1.71%
26	record title item catalog index entry catalog	16	0.275	0.45%	1.44%
28	manuscript edition text transcription editor witness	24	0.313	3.01%	0.16%
30	rule error attribute operation expression value grammar	14	0.269	1.31%	0.72%
32	library librarian service public_library book collection	21	0.316	0.27%	2.01%

No.	Most relevant terms by log-lift formula (Sievert & Shirley, 2014), $\lambda = 0.8$	Cluster size	Topic coherence	Mean prob. DH	Mean prob. IS
34	sentence phrase grammar clause parser verb noun dependency	25	0.308	2.09%	0.39%
37	method algorithm approach result experiment performance	16	0.288	2.46%	2.12%
38	gesture emotion feedback movement video hand body multimodal	15	0.25	0.91%	0.24%
42	factor behavior construct variable effect perception study	17	0.261	0.22%	1.71%
43	question answer response reference percent type wikipedia	17	0.261	0.38%	0.78%
44	link hypertext node subject tree hypermedia path marker	14	0.264	0.51%	0.73%
47	character letter string code script symbol abbreviation	14	0.287	1.32%	0.50%
50	relevance criterion judgment relevance_judgment evaluation	16	0.307	0.11%	1.30%
54	respondent survey use questionnaire per_cent percent	19	0.291	0.19%	1.36%
55	archive collection material preservation museum archivist	14	0.302	0.97%	0.77%
60	woman gender man male sex female child race mother life age	14	0.304	0.48%	0.49%
61	name person match character entity surname location place	18	0.241	0.64%	0.38%
62	information source need information_seeking study environment	21	0.299	0.52%	2.99%
63	document collection number_of_document representation	20	0.269	0.49%	1.40%
64	concept domain ontology property term classification	14	0.3	0.66%	0.87%
66	health patient physician disease doctor drug health_care	20	0.382	0.08%	1.11%
68	science discipline scientist research field literature	17	0.335	0.27%	1.83%
70	topic relevance keyword interest criterion distribution lda	14	0.239	0.34%	0.61%
71	corpus language english text genre type register frequency	20	0.284	2.07%	0.11%
72	theory concept meaning knowledge view information_science	20	0.295	1.35%	3.09%
74	citation paper publication author article impact reference	24	0.318	0.28%	3.57%
75	query term retrieval result performance relevant_document	25	0.323	0.31%	2.55%
78	information organization development service center	12	0.265	1.08%	1.87%
83	participant interview study experience time people group	14	0.279	0.30%	1.41%
86	author authorship word style text sample attribution	20	0.302	2.30%	0.22%
90	database record file field data host search name information	18	0.289	0.71%	1.25%
91	child school parent home adult teacher boy girl age	14	0.288	0.16%	0.53%
93	time year change period rate trend over_time figure	14	0.272	0.84%	1.19%
94	web page site link web_site web_page search_engine url	20	0.349	0.69%	1.80%
96	shakespeare play test scene tragedy block rejection act	19	0.26	1.18%	0.06%
98	music song pitch note sound musician composer mode listener	19	0.242	0.56%	0.35%
99	task participant work type study performance behavior goal	18	0.268	0.31%	0.87%
102	per_cent use session behavior log service usage number	14	0.263	0.07%	0.72%
103	tag tagging tagger tagset metadata flickr social_tagging	13	0.123	0.34%	0.25%
104	resource metadata digital_library collection content access	20	0.316	1.18%	1.11%
105	cluster matrix clustering similarity analysis distance	20	0.291	0.99%	1.17%
108	book chapter reader author reading section volume discussion	26	0.272	1.21%	1.22%
110	word frequency dictionary list vocabulary occurrence phrase	21	0.29	2.28%	0.72%

No.	Most relevant terms by log-lift formula (Sievert & Shirley, 2014), $\lambda = 0.8$	Cluster size	Topic coher- ence	Mean prob. DH	Mean prob. IS
111	verb argument frame noun predicate case clause pronoun	17	0.322	1.62%	0.15%
112	image color picture color painting photograph art pixel	22	0.266	1.44%	0.77%
115	category classification taxonomy class categorization	20	0.237	0.45%	1.03%
116	project research team development work resource collabora- tion	17	0.291	2.37%	0.88%
118	network node link graph connection edge relationship tree	20	0.268	0.52%	0.86%
119	student course teacher class teaching learning classroom	20	0.339	2.08%	0.89%
120	sense wordnet word synset dictionary wsd relation meaning	20	0.323	1.83%	0.10%
121	search searcher result searching query term keyword	19	0.295	0.27%	1.66%
122	feature classifier training accuracy classification set	20	0.321	2.04%	0.63%
124	speech speaker utterance dialogue transcription recording	19	0.295	1.32%	0.15%
125	technology media computer device internet environment	12	0.284	1.99%	1.40%
132	element tei xml markup sgml encoding dtd attribute struc- ture	20	0.355	2.22%	0.37%
133	map visualization area dialect distance location city	16	0.26	1.18%	0.45%
134	text analysis passage word reader paragraph section	20	0.27	2.54%	0.65%
135	_game player quest avatar adventure video _game narrative	20	0.274	1.17%	0.14%
142	information _seeking process activity behavior stage	13	0.307	0.17%	1.61%
144	system task input technique set processing evaluation	20	0.273	1.23%	1.44%
146	information _literacy practice literacy skill landscape	13	0.31	0.22%	1.29%
147	way case problem fact part use question example kind form	19	0.251	5.58%	4.79%
152	data data _set dataset analysis database tool researcher	20	0.276	1.28%	1.00%
153	annotation annotator agreement corpus sect annota- tion _scheme	20	0.295	1.60%	0.09%
154	table difference result variable test value number analysis	23	0.294	2.23%	3.15%
158	object property representation entity relation description	19	0.287	1.10%	1.22%
159	model modeling process approach figure simulation rela- tionship	20	0.243	0.73%	0.98%
160	poem line poetry verse poet syllable bible word sonnet	15	0.277	1.21%	0.06%
162	suffix stem form word root rule lexicon morphology mor- pheme	19	0.367	1.55%	0.42%
164	term thesaurus indexing index descriptor concept keyword	18	0.302	0.40%	1.70%
165	user interface information interaction design need search	20	0.289	0.77%	1.93%
174	novel story narrative character fiction genre literature	12	0.295	1.85%	0.25%
175	knowledge expert domain organization knowledge management	20	0.235	0.36%	1.08%
177	language translation english alignment translator resource	19	0.284	2.13%	0.46%
180	journal article citation impact _factor year isus jr title	20	0.341	0.14%	2.62%
181	research study researcher analysis field method area	17	0.301	1.40%	2.71%

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 173–199. DOI: doi.org/10.5283/epub.44944.

Grundlagenermittlung für die digitale Werkbank qualitativ-hermeneutisch arbeitender Geisteswissenschaftlerinnen

Exploration geisteswissenschaftlicher Forschung mit Fokus auf Exzerpten und Literaturverwaltung

Michael Achmann, Florin Schwappach

Universität Regensburg, Lehrstuhl für Medieninformatik

{[michael-markus.achmann](mailto:michael-markus.achmann@ur.de), [florin.schwappach](mailto:florin.schwappach@ur.de)}@ur.de

Abstract

Dieser Beitrag berichtet von einer Fallstudie zum wissenschaftlichen Arbeiten mit Nachwuchswissenschaftlerinnen. Der Fokus der Untersuchung lag auf den Aktivitäten Notieren, Exzerptproduktion und Literaturverwaltung bei qualitativ-hermeneutisch arbeitenden Geisteswissenschaftlerinnen. Das Ergebnis legt nahe, dass vorhandene Literaturverwaltungssoftware nicht zum mentalen Modell der Geisteswissenschaftlerinnen passt, da dieses Notizen und Exzerpte im Mittelpunkt eines bibliografischen Eintrags sehen. Entsprechend sollten die Digital Humanities bei der Entwicklung zukünftiger Werkzeuge das Notieren und die Literaturverwaltung sowie die Verknüpfung beider Aktivitäten stärker in den Fokus setzen. Zur Verifizierung der Erkenntnisse aus der Fallstudie sind weitere quantitative Studien notwendig.

Keywords: Digital Humanities; virtuelle Forschungsumgebungen; personal information management; wissenschaftliche Forschung; scholarly research; grounded theory; Informationsverhalten; information behavior

1 Zur Relevanz und Geschichte des Notierens beim wissenschaftlichen Arbeiten

Der Stellenwert von Notizen und der Erstellung von Exzerpten beim wissenschaftlichen Arbeiten stößt in der Forschung auf breite Anerkennung (Case,

1986, 1991; Cevolini, 2018; Palmer/Neumann, 2002), gerade beim Exzerpieren handelt es sich um eine jahrhundertealte Tradition. Zwischen dem 16. und 17. Jahrhundert lernten die Gelehrten zu vergessen und fingen an, Exzerpte als Speicherunterstützung für das vergängliche menschliche Gedächtnis zu nutzen. Es entwickelten sich Register und Wissenschaftler begannen damit, Notizen zu systematisieren (Cevolini, 2018). Die Geschichte der Exzerpte lässt sich über Zettelboxen bis zur Literaturverwaltungssoftware weiterzeichnen. Erstere wurden bereits vor der digitalen Revolution und damit vor der Entwicklung erster Spezialsoftware im viel zitierten Aufsatz „As We May Think“ von Vannevar Bush (1945) zu einem – hypothetischen – neuen System, dem Memex, weiterentwickelt. Dieses sollte die Verwaltung und Verknüpfung von Wissen und Informationen über Dokumentengrenzen hinweg ermöglichen. Die Forschung zum „Personal Information Management“ (PIM) wird bis heute von dieser visionären Idee beeinflusst und angetrieben (Jones, 2007; Gemell/Bell/Lueder, 2006). Während einzelne Technologien und Entwicklungen wie das World Wide Web (Berners-Lee, 1980) die Memex-Vision in greifbare Nähe rücken, ist fraglich, wie sich in der Forschungspraxis – genauer: im PIM der geisteswissenschaftlichen Forschung – widerspiegeln. Dowlings und Wilsons (2017) Arbeit legt nahe, dass die Praxis von Nachwuchswissenschaftlerinnen weniger durch technische Möglichkeiten, sondern durch den Aufwand, der für die Einarbeitung in neue Werkzeuge nötig ist, bestimmt wird. Derartiges Zögern bei der Nutzung von Software wurde auch im Rahmen der vorliegenden Studie beobachtet.

Aktuelle Entwicklungen im Bereich des Machine Learnings in Kombination mit moderner Eye-Tracking-Hardware oder Augmented-Reality-Geräten wie der Microsoft HoloLens könnten die Einarbeitung in Spezialsoftware zukünftig überflüssig machen: Was wäre, wenn die Software beim Arbeiten mit analoger Literatur automatisiert die Literaturverwaltung übernehmen könnte und die eigenen Notizen mit den gelesenen Werken verknüpft? Khan (2019) stellt ein System vor, das die Verknüpfung von Notizen mithilfe eines Eyetrackers in einem digitalen Text erlaubt, ein ähnliches Vorgehen wäre auch für analoge Quellen denkbar. Ein derartig intelligentes System könnte Nutzerinnen dabei helfen, zwei Hürden des PIM zu überwinden: die Fragmentierung von Informationen, insbesondere zwischen der analogen und der digitalen Welt, sowie hohe Kosten bei der Einführung in die Nutzung der Software, da sich in dieser Vision nicht die Nutzerin an den Computer anpassen muss, sondern vielmehr der Computer sich der Nutzerin anpasst. Der „Digital Desk“ (Wellner, 1993) und „Augmented Paper“ (z.B. Norrie/Signer/

Weibel, 2006; Margetis et al., 2019) sind lange formulierte Ideen, die die Verknüpfung der Affordanzen des Papiers (Sellen/Harper, 2002) in der analogen Welt mit den Vorteilen der digitalen Welt ermöglichen sollen. Nachfolgende Studie dient der Grundlagenermittlung für eine derartige augmentierte digitale Forschungsumgebung für Geisteswissenschaftlerinnen, für deren qualitative Forschung auch weiterhin Bücher und gedrucktes Papier eine zentrale Rolle spielen werden.

Zusammenfassend könnte die digitale Erfassung analoger Notizen in Kombination bei automatisierter Verknüpfung und Anreicherung mit vorhandenen digitalen Informationen (wie Metadaten der Literatur oder einer digitalen Version davon) „Refinding“-Aktivitäten vor allem im Bereich der Literaturverwaltung unterstützen. Andorfer (2015) formuliert in seiner Studie weitere Argumente für die Digitalisierung geisteswissenschaftlicher Forschungsdaten: Einerseits könnte es wünschenswert sein, alle Daten zu publizieren, andererseits ermöglichen strukturierte abgespeicherte Notizen eine sinnvolle Archivierung (Puhl et al., 2015). Im Folgenden wird der aktuelle Forschungsstand zur Untersuchung des wissenschaftlichen Arbeitens vorgestellt, anschließend wird die Interviewführung näher erläutert und schließlich werden die Ergebnisse der Studie mit Fokus auf Notieren und Literaturverwaltung berichtet.

2 Forschungsstand

Wissenschaftliches Arbeiten, Forschungsdaten in der Geisteswissenschaft und die Exzerptproduktion wurden in verschiedenen akademischen Disziplinen untersucht: Bei seinen Recherchen im Feld der Informationswissenschaft untersuchte Case (1986) primär Notizen im akademischen Bereich. Er identifiziert das Notieren als wichtigen Bestandteil der Forschungsarbeit: “In taking notes, the scholar screens large amounts of material, selecting only those elements that are relevant to his or her interests. More importantly, the scholar transforms what is read – condensing it, making inferences and drawing conclusions” (ebd., S. 99).

In einer weiteren Studie untersuchte Case (1991) die Arbeit von Historikerinnen. Er beobachtete einen Wandel in ihren Gewohnheiten: Manche Probandinnen begannen, ihre handschriftlichen Notizen digital zu transkribieren und in diesem Prozess mit weiteren Informationen anzureichern. Insgesamt

beschrieb er den Arbeitsprozess der Geisteswissenschaftlerinnen als nicht linear, zentriert um eine Gliederung, die sich ständig weiter entwickelte in einem Vor und Zurück zwischen Notieren, Sortieren, Schreiben und der Neuorganisation derselben.

Brockman et al. (2001) konnte zehn Jahre später feststellen, dass sich die wissenschaftliche Arbeit im Bereich der Geisteswissenschaften trotz Durchbruch der Computer nicht radikal gewandelt hat; weiterhin war der Texteditor die primär genutzte Software unter Geisteswissenschaftlerinnen. Damit liegt die Frage nahe, ob sie keine weiteren digitalen Werkzeuge neben dem Texteditor benötigen oder schlicht kein passendes Tool existiert. Toms (2008) evaluierte die Bedürfnisse von Wissenschaftlerinnen: Der geisteswissenschaftliche Werkzeugkasten benötigt Meta-Tags, um relevante Informationen abbilden zu können (abseits von z. B. Titel, Autor), passende Scanning- und Browsing-Werkzeuge, um Texte zu explorieren, sowie eine bessere Unterstützung, um digitale Texte herunterzuladen und zu speichern; außerdem “[a] note-taking feature that would allow researchers to compile, edit and save their reflections on the text; notes could be viewed either co-located on a single page or as annotations on the actual text, or as a composite” und ein standardisiertes, durchdachtes Textanalysewerkzeug. Während aber die Praxis des Annotierens in verschiedenen Disziplinen untersucht wurde, um neue Tools zu entwerfen oder vorhandene zu verbessern, wurden Notizwerkzeuge für Geisteswissenschaftlerinnen wenig beachtet (Bradley/Vetch, 2007; Brush et al., 2001; Marshall, 1997).¹ Zusammenfassend zeigt sich in der Forschung eine Lücke im Bereich der digitalen Werkzeuge für Geisteswissenschaftlerinnen oder aber zumindest die offene Frage, ob der Bedarf existiert, diese Lücke zu schließen.

Zur Analyse des Forschungsprozesses hat Unsworth (2000) eine Liste sogenannter „scholarly primitives“ vorgestellt. Diese wurden durch Blanke und Hedges (2013) weiterentwickelt und beispielsweise von Benardou et al. (2010) bei der Modellierung des wissenschaftlichen Arbeitens genutzt.

Zusammen mit Antonijević (2015) ethnografischer Studie dienen die genannten Arbeiten als Grundlage für die im Folgenden vorgestellte Fallstudie, insbesondere für die Analyse der durchgeführten Interviews. Die genannten

1 Beispielsweise für die Linguistik existieren diverse Annotationswerkzeuge. Diese Arbeit beschäftigt sich allerdings mit hermeneutisch arbeitenden Geisteswissenschaftlerinnen; eine Annotationssoftware für Linguistik erfüllt nicht unbedingt die Erwartungen einer Historikerin.

Studien basieren auf der Forschung aus dem Bereich des Personal Information Management, dass sich nach Barreaus Definition mit den Methoden beschäftigt, die Personen zur Erkundung, Speicherung und Organisation von Informationen in ihrer Arbeitsumgebung nutzen (Barreau, 1995). Jones schlägt auf dieser Definition aufbauend drei Hauptgruppen für PIM-Aktivitäten vor: “Finding & refinding activities, keeping activities, and meta-level activities” (Jones, 2007). Während die vorgestellte Studie sich insbesondere mit den Meta-Level-Aktivitäten beschäftigt, der Organisation und dem Management von Information im „Personal Information Space“, ergeben sich aus diesen Aktivitäten Konsequenzen für die Aktivitäten anderer Gruppen.

2 Vorstellung der Fallstudie

Inspiziert durch die eben vorgestellte Literatur wurde eine Interviewreihe durchgeführt, um typische Aktivitäten und Arbeiten während der Forschung in den Geisteswissenschaften und Nachbardisziplinen zu eruieren. Dabei wurde ein spezielles Augenmerk auf die Dateiverwaltung und das Notieren gelegt, da sie zu den wichtigsten Aktivitäten für Geisteswissenschaftlerinnen gezählt werden (Case, 1986, 1991; Cevolini, 2018; Palmer/Neumann, 2002).

Um den Arbeitsablauf von zwölf Nachwuchsforscherinnen im Master- oder Promotionsstudium zu erkunden, wurde eine Fallstudie durchgeführt. Die dazugehörigen Interviews wurden im Zeitraum eines Monats geführt, ihnen gingen zwei Testinterviews zu Eruierung des Fragebogens voraus. Die Gespräche wurden durch demografische Angaben zu akademischem Lebenslauf, Softwarenutzung und einer Selbsteinschätzung zur Computeraffinität ergänzt. Jede Befragung hat ca. eine Stunde in Anspruch genommen. Die Partizipierenden wurden via Sozialen Medien und Mundpropaganda akquiriert. Die daraus resultierende willkürliche Stichprobe wurde in Anbetracht des explorativen Charakters dieser Studie in Kauf genommen. Sie bestand aus fünf Doktorandinnen und sieben Masterstudierenden. Während Geisteswissenschaftlerinnen die primäre Untersuchungspopulation darstellen, wurden auch Interviews mit Studierenden aus Nachbardisziplinen durchgeführt: vier Geschichtsstudentinnen, davon zwei Doktorandinnen und zwei Masterstudierende, zwei Osteuropastudentinnen, eine Doktorandin der Slawistik, eine der Makroökonomie, eine der Sozialpsychologie, ein Masterstudent der Politikwissenschaft, eine Masterstudentin der Philosophie sowie eine Lehr-

amtsstudentin der Fächer Biologie, Chemie und Ethik. Letzteres Interview wurde bei der Auswertung verworfen, da sich offenbarte, dass die Studentin keinerlei Forschungsbezüge in ihrem Studium aufzeigen konnte. Zehn der Interviewpartnerinnen waren weiblich, zwei männlich; alle Interviewten wurden zwischen 1989 und 1996 geboren.

Alle Interviews wurden händisch transkribiert, dabei wurde alles ins Hochdeutsche übersetzt. Anschließend wurden die Transkripte im Sinne der Grounded Theory codiert (Corbin/Strauss, 2008; Strauss/Corbin, 1990). Zunächst wurde nach den Anweisungen von Corbin und Strauss eine offene Codierung durchgeführt, durch mehrere Iterationen wurden sie kondensiert. Schließlich wurden aus den konzeptionellen Codes in den letzten Durchläufen Aktivitäten klassifiziert und Konzepte, Attribute, Werte und Relationen ergründet (Milton, 2007). Dafür wurde die Software MAXQDA (VERBI Software, 2019) genutzt. Die Interviews konzentrierten sich vor allem auf folgende Aktivitäten des wissenschaftlichen Arbeitens: 1. Auswahl der Methode, 2. Notieren, 3. Schreiben, 4. Archivieren. Im Vergleich zur einschlägigen Forschungsliteratur kamen zwei Themen in den durchgeführten Interviews zu kurz: Das Publizieren wurde nicht im gewünschten Maße erfasst, Kollaboration wurde in keinem Gespräch thematisiert.

Bei den Interviews lag das Augenmerk darauf zu eruieren, wie jede der Aktivitäten durchgeführt wird, welche Werkzeuge die Forscherinnen nutzen, welche Probleme in den einzelnen Schritten auftreten und wie die Teilnehmerinnen wünschen, die erfragten Prozesse zu verbessern. Die Befragten wurden darum gebeten, sich zur Beantwortung der Fragen an ihre letzte größere Forschungsarbeit zurückzuerinnern.

3 Auswertung der Fallstudie

Durch den Grounded-Theory-Prozess wurden 23 Aktivitäten gefunden, die in fünf Kategorien gruppiert wurden: ‚universelle Aktivitäten‘, ‚Vorbereitung‘, ‚Wissensakquise‘, ‚Schreiben‘ und ‚Ende des Projekts‘. Mit Ausnahme der ‚Wissensakquise‘ wurden unter allen Interviewpartnerinnen die gleichen Aktivitäten beobachtet, unabhängig von der jeweiligen akademischen Disziplin. Innerhalb der Gruppe der Wissensakquise waren die beobachteten Aktivitäten heterogen und abhängig von der gewählten Methodologie. Gleiches gilt für die zeitliche Abfolge. Auch hier fielen starke Unterschiede zwischen

einzelnen Disziplinen, genauer aber einzelnen Methodologien auf: Während beispielsweise die Psychologin einen stark sequenziellen Arbeitsablauf beschrieb, begonnen mit der Definition einer Forschungsfrage, haben die wenigsten interviewten Geisteswissenschaftlerinnen diese Tätigkeit überhaupt erwähnt. Vielmehr beschrieben sie einen iterativen Prozess ihrer Forschung, der mit einem vage formulierten Forschungsinteresse beginnt und durch mehrere Zyklen des Recherchierens und Lesens von Literatur zu einer finalen Auswahl an Fragen führt. Damit unterscheidet sich der Forschungsprozess von qualitativ-hermeneutisch arbeitenden Geisteswissenschaftlerinnen von den Prozessen anderer Methodologien durch ihre nicht-linearen Abläufe und das Hin- und Herspringen zwischen verschiedenen Aktivitäten. Der Arbeitsablauf beim hermeneutischen Arbeiten ist somit durch synchron und rekursiv ablaufende Aktivitäten geprägt, wobei die Ergebnisse jedes Durchlaufs einer Aktivität jeweils das Fundament für die nächsten Aktivitäten bildet.

3.1 Literaturverwaltung

Weiterhin wurde ein Unterschied in der Literaturverwaltung der Befragten beobachtet. Die Verwaltungsstrategien konnten in fünf unterschiedliche Kategorien eingeteilt werden: Nutzung von Spezialsoftware (z. B. Zotero), Nutzung von Spezialsoftware-Excel-Hybrid, händische Verwaltung, Excel-händisch-Hybrid und Spezialsoftware-händisch-Hybrid. Aufgrund der qualitativen Natur der vorliegenden Studie wurden einzelne Strategien nur einmalig beobachtet, andere von mehreren Befragten berichtet. Um die vorgenannten Gruppen zu bestätigen, bedarf es einer quantitativen Anschlussstudie.

Diese Gruppierung stellt einen wichtigen Baustein für die Beobachtung zweier Hauptstrategien dar: Für diejenigen, die ihre Literatur händisch verwalten, standen die Exzerpte und Notizen zur Literatur im Mittelpunkt. Sie sind nach Referenzen zu den Texten geordnet. Damit stellen die Notizen einen Prototypen für das Zitieren während des Schreibens dar, während sie zeitgleich als Übersicht der gelesenen Schriften dienen (s. Type A in Abb. 1). Dem gegenüber stehen die Nutzerinnen der speziellen Literaturverwaltungsprogramme oder Excel. Sie organisieren ihre Literatur durch Verschlagwortung und das Hinzufügen von Metadaten, wodurch sie die Wiederauffindbarkeit von Texten verbessern und von der Möglichkeit Gebrauch machen, diese Listen in gelesene und zu lesende Literatur einzuteilen. Während es weiterhin persönliche Unterschiede innerhalb dieses Typs gibt – manche sortieren ihre Literatur in Bezug auf Projekte, andere pflegen eine fortwährende Sammlung

–, steht der bibliografische Eintrag im Mittelpunkt dieses Typs (vgl. Type B in Abb. 1). Während die Möglichkeit nur von einer Teilnehmerin explizit genannt wurde, erlaubt dieser Ansatz auch das Anhängen von (digitalen) Büchern und Papers an den Datenbankeintrag sowie das Verknüpfen von Notizen. Ob ein bibliografischer Eintrag während des Recherche- oder Leseprozesses erstellt wird, scheint individuell verschieden zu sein. Die Erstellung eines solchen kann beim Typ B wortwörtlich genommen werden, für Wissenschaftlerinnen, die ihre Quellen nach Typ A organisieren, handelt es sich um einen impliziten Vorgang, während sie ein neues Kapitel in ihren Notizen anlegen und darin die neue Literatur referenzieren. Die bibliografischen Informationen werden parallel zur Aktivität „Schreiben“ und „Gliederung entwerfen“ durchgeführt. Während der Abruf beim Schreiben klarer erscheint, haben einige Interviewpartnerinnen genannt, dass sie beim Entwurf einer Gliederung bereits spezifische Literatur mit einzelnen Überschriften und Abschnitten verknüpfen. Hier geschah die Verknüpfung teils implizit, teils explizit. Das Nachverfolgen von Zitaten, wofür die Notizen zu Literatureinträgen auch genutzt werden, wurde nur parallel zur Aktivität „Schreiben“ beobachtet.

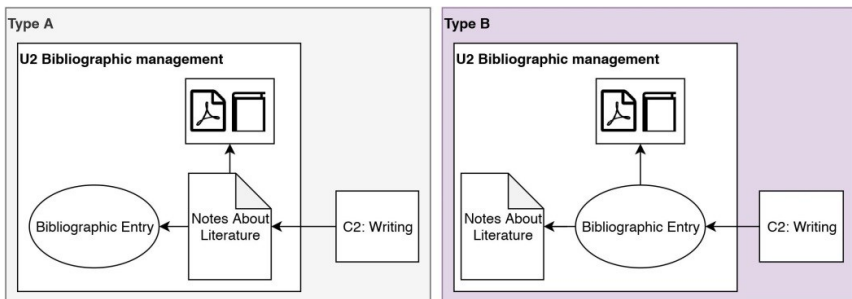


Abb. 1 Die zwei Typen der Literaturverwaltung

Das Bibliografiemanagement wird als universelle Aktivität kategorisiert, da von Studierenden berichtet wurde, dass diese Aktivität zu unterschiedlichen Zeitpunkten zu unterschiedlichen Zwecken im Verlauf eines Forschungsprojekts initiiert werden kann. Es dient der Speicherung von Informationen über Quellen und Literatur sowie dem Abruf der gespeicherten Informationen, insbesondere um Zitate nachzuverfolgen.

3.2 Notieren

Neben dem Literaturmanagement lag ein weiterer Fokus der Fallstudie auf dem Notieren beim wissenschaftlichen Arbeiten. Diese Aktivität wird, ähnlich wie das Bibliografiemanagement auch, quer durch alle in der Befragung vertretenen Disziplinen zu fast jedem Zeitpunkt der Forschung initiiert. Entsprechend wird auch sie als Universalaktivität bewertet. Durch das Notieren entstehen Zeitkapseln für Informationen, wobei verschiedene Arten von Notizen beobachtet wurden, die jeweils unterschiedlichen Zwecken dienen. Die Befragten haben von folgenden Arten der Notizen berichtet:

- Gesprächsnotizen,
- Übersetzungsnotizen,
- Vorlesungsnotizen,
- Notizen zum Denken,
- Notizen für sich selbst,
- Literaturnotizen (Exzerpte),
- To-do-Listen,
- Zeitstrahle.

Ein Teil der genannten Notizen entsprechen der Definition von „Information Scraps“ (Bernstein et al., 2008): To-do-Listen, die auf Schmierpapier notiert wurden, Notizen zum Denken oder Code-Kommentare tendieren stark dazu, von Standard-Informationsmanagement-Werkzeugen nicht erfasst zu werden.

Exzerpte jedoch stellen eine Ausnahme dar: Sie sind eine besondere Art der Notizen, die als digitale oder analoge Textdokumente existieren und sorgfältig abgeheftet oder eingeordnet werden. Dies geschieht entweder an einem zentralen Ort oder projektzentriert. Wenn sie aus der Ordnung der PIM-Werkzeuge herausfallen, stellt das ein großes Problem beim Erstellen von Arbeiten dar, da ein Großteil des wissenschaftlichen Arbeitens darauf beruht, das passende Exzerpt zu finden. Im Digitalen werden sie in einer kultivierten Ordner-Hierarchie abgelegt. Die Sinneinheiten der Dokumente werden nach persönlichem Geschmack gewählt: Einzelne Befragte haben ein einziges Dokument für verschiedene Quellen genutzt, andere wiederum für jede Quelle ein einzelnes Dokument angelegt. In der Papierwelt konnten ähnlich heterogene Strukturen beobachtet werden: angefangen bei College-Blöcken, die für alle Art der Notizen genutzt werden, bis hin zu Ordnern, die fein säuberlich sortiert Exzerpte zu allen gelesenen Quellen beinhalten.

Eine Person berichtet beispielsweise, ihre eigenen Gedanken beim Erstellen ihrer Literaturnotizen mit dem Inhalt der gelesenen Literatur zu vermengen. Diese Notizen werden damit zu ihrer eigenen Arbeit, die ihr den Schreibprozess vereinfacht, da Gedanken bereits mit einer passenden Quelle belegt sind. Dieser Bericht passt zu Ehlichs (1981) Beobachtung einer partiellen Paraphrasierung in Exzerpten. Zusammengefasst scheint die Motivation zur Erstellung von Literaturnotizen und Exzerpten derart ähnlich, dass den Studierenden, die nicht den Begriff „Exzerpt“ genutzt haben, unterstellt werden kann, dass sie ein Exzerpt anfertigen, auch wenn ihnen der Fachbegriff nicht geläufig ist.²

Mit dieser Unterstellung im Kopf können weitere Beispiele der Exzerptproduktion betrachtet werden: Eine Person berichtet davon, ihre eigenen Gedanken in ihren Notizen mit Textmarkern hervorzuheben, um den kopierten Inhalt von ihren eigenen Gedanken zu unterscheiden. Lesen und Notieren finden bei ihr zeitgleich statt. Die Notizen werden so zur Kombination aus einer Zusammenfassung der Quelle, vermischt mit eigenen Gedanken. Ähnliches berichtet eine Studentin der Philosophie: Sie sieht in Notizen den Zweck, Fragen zu einem Text beantworten zu können, ohne ebenjenes erneut komplett lesen zu müssen. Sie nutzt Randnotizen sowie Exzerpte für ihre Notizen während des Lesevorgangs. Die Notizen enthalten neben Argumentationsketten auch Referenzen zu anderen Dokumenten. Eine Geschichtsstudentin nutzt Post-its, um die ausgedruckten Kopien ihrer Quellen physisch zu erweitern. Eine weitere Historikerin benützt zwei unterschiedliche Begriffe, wenn sie von ihren Notizen spricht: einerseits „Transkripte“, andererseits „Exzerpte“. Der erste Begriff beschreibt selektive Kopien von Archiv-Quellen in Form eines Word-Dokuments. Den zweiten Terminus „Exzerpte“ benutzte sie vornehmlich bei der Beschreibung ihrer Arbeit mit Sekundärquellen. Sie nutzt außerdem die Citavi-Notizfunktion, um Verknüpfungen zwischen verschiedenen Werken herzustellen. Diese Funktion wird auch von einer Studentin der Slawistik verwandt, womit sie ihre Notizen zwischen der Literaturmanagementsoftware und der Verzeichnisstruktur ihres Disserta-

2 Während der Begriff „Exzerpt“ nur von Studierenden aus dem Bereich der Geschichtswissenschaft genutzt wurde, wird der Begriff im Folgenden eben auch für Literaturnotizen verwendet. Mögliche Erklärungen für die Nutzung unterschiedlicher Begrifflichkeiten: (1) die Grundlagenausbildung zum wissenschaftlichen Arbeiten an unterschiedlichen Lehrstühlen und (2) die Selbstreflexion des eigenen wissenschaftlichen Arbeitens und der genutzten Methodologie.

tionsprojekts fragmentiert. Eine weitere Person berichtet die Nutzung der Word-Kommentarfunktion, um innerhalb der Exzerpte weitere Notizen an sich selbst einzufügen – die Kommentarfunktion dient dabei dem Zweck, ihre eigenen Gedanken zu markieren; ihr Inhalt sind Meta-Informationen zu Verknüpfungen zwischen Notizen und Quelle. Insgesamt hat jede Interviewpartnerin davon berichtet, Zusammenfassungen der gelesenen Texte zu erstellen oder Notizen zu erarbeiten, die mit der Literatur verknüpft sind. Alle fertigen ihre Notizen während des Lesens der Quellen an. Abschließend muss noch eine besondere Beobachtung berichtet werden: Die Doktorandin der Psychologie berichtet davon, nach Abschluss des Leseprozesses und damit auch vollendetem Exzerpt dieses vom ursprünglichen Word-Dokument in ihre Excel-Literaturliste zu kopieren, um in Form der Excel-Liste an einem zentralen Ort sämtliche Informationen über alle Einträge hinweg durchsuchen zu können.

Eine weitere Art der Notizen, die besonders häufig genannt wurde, sind To-do-Listen. Sie tragen eine Erinnerungsfunktion und werden für eine begrenzte Lebensdauer angefertigt. Die anderen aufgeführten Typen von Notizen wurden in nur wenigen Fällen berichtet, entsprechend wird hier nicht weiter ausgeführt.

3.3 Digital vs. analog

Dieser Abschnitt beschäftigt sich mit der Modalität der Notizen: Werden Sie analog oder digital angefertigt? Eine Antwort kann durch die vorgestellte Fallstudie lediglich für die interviewte Gruppe gegeben werden, ohne von der geringen Zahl an Befragten auf Wissenschaftlerinnen im Allgemeinen schließen zu können. Die meisten der Interviewpartnerinnen nutzen für ihre Notizen irgendeine Art von Block, wenige den Seitenrand und eine Person Post-it-Zettel. Im Digitalen werden vor allem Word- und Plain-Text-Dateien verwendet, zwei Wissenschaftlerinnen nutzen Excel und jeweils eine Teilnehmerin Code-Kommentare und Citavi. Insgesamt betrachtet wird in den seltensten Fällen exklusiv digital oder analog notiert. Drei Personen verwenden nur Notizbücher oder Blöcke, demgegenüber eine Person nur Word. Vor dem Hintergrund bestehender Forschung erscheint dieses Ergebnis nicht besonders, da auch frühere Studien darauf hinweisen, dass es eine gleiche Aufteilung zwischen physischen und digitalen „information scraps“ gibt, mit nur wenigen, die komplett digital notieren (Bernstein et al., 2008; Boardman/Sasse, 2004; Jervis, 2014).

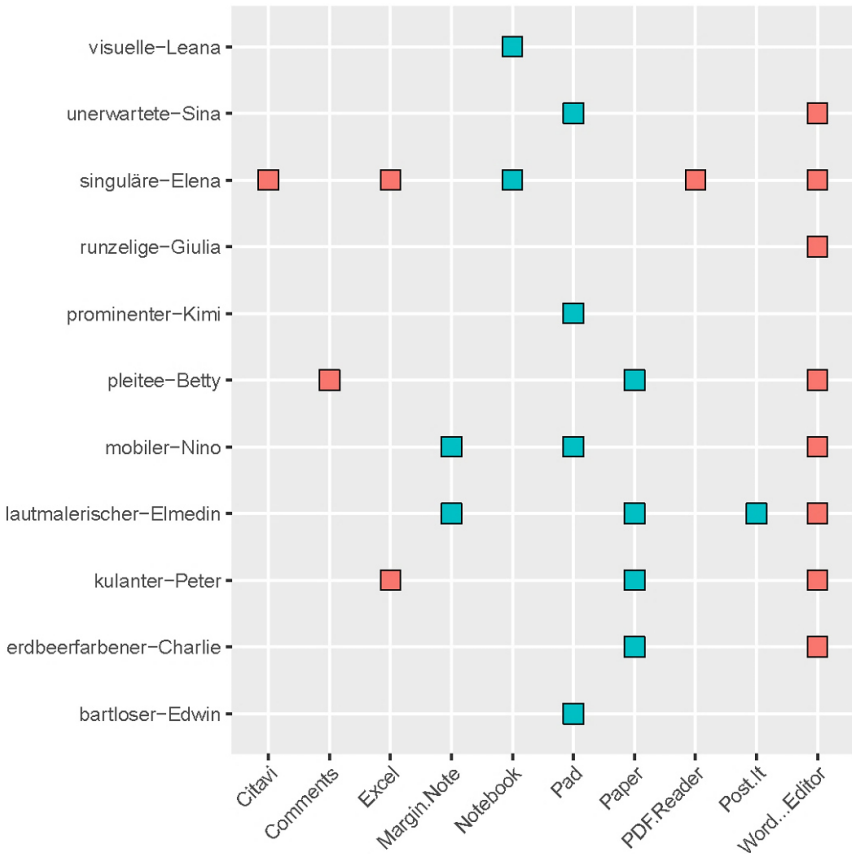


Abb. 2 Medien und Software, die für das Notieren genutzt werden: Rote Boxen markieren digitale Medien, blaue analoge Medien; links die Pseudonyme der Befragten.

Zusammenfassend haben die Befragten von zwei Arten der Notizen während des wissenschaftlichen Arbeitens berichtet: Manche Notizen haben den Charakter von Dokumenten, andere fallen eher in die Definition von „information scraps“. Exzerpte und Literaturnotizen sind die Art von Dokumenten, die am besten durch PIM-Werkzeuge erfasst werden können. Sie werden systematisch erstellt und organisiert. Normalerweise beinhalten die Exzerpte Referenzen zu den Seiten der Dokumente, auf die sie sich beziehen. Das Detaillevel dieser Zusammenfassungen hängt von den persönlichen Präferenzen ab, wobei manche der Befragten nur davon berichten, wichtige Schlagwörter zu notieren, während andere Paraphrasen von Textstellen in Verknüpfung mit Argumentationsketten und vermischt mit eigenen Gedanken und

Verknüpfungen niederschreiben. Spezielle Formatierung und Annotationen können genutzt werden, um eigene Ideen oder exakte Übernahmen aus den Quellen zu kennzeichnen. Unabhängig davon, ob ein Exzerpt handschriftlich erstellt oder getippt wird: Sie beinhalten Referenzen zu den originalen Dokumenten, beispielsweise bereits in Form von zitierfähigen Kennzeichnungen. Während eine Minderheit von der Nutzung stark strukturierter Werkzeuge wie Citavi oder PDF-Annotationswerkzeuge in Standardbetrachtern berichtet, entschieden sich die meisten Befragten für Textdokumente oder Blöcke zum Notieren. Die Art der Organisation variiert zwischen den Befragten: Während eine Person ihre Exzerpte sorgfältig in ein Notizbuch einträgt, das sie bereits nach Kauf durch die Einteilung in Kapitel strukturiert hat, fertigt eine andere Person ihre Literaturnotizen auf Schmierpapier an.

Nichtsdestotrotz berichtet auch die letztgenannte Studentin davon, die Notizblätter anschließend ungeordnet in einer Mappe zu sammeln, die für ein gesamtes Projekt genutzt wird. Somit organisieren alle Befragten ihre Exzerpte, auch wenn sich die Art der Organisation im Detail unterscheidet. Schlussendlich konnte ein Problem der Fragmentierung beobachtet werden: Manche der Befragten nutzen analoge und digitale Techniken beim Lesen und Notieren. Diese Beobachtung stimmt mit Ergebnissen bestehender Forschung überein (Boardman/Sasse, 2004; Jervis, 2014; Mizrachi/Bates, 2013).

3.4 Zusammenfassung

Es wurde die Analyse von elf Interviews mit Nachwuchswissenschaftlerinnen vorgestellt. Auf dieser Basis konnte festgestellt werden, dass ein Unterschied in der Praxis des wissenschaftlichen Arbeitens zwischen den Geisteswissenschaften und verwandten Fächern lediglich im Bereich der Wissensakquiseaktivität existiert. In dieser Aktivität wird die geisteswissenschaftliche-hermeneutische Praxis des Recherchierens und Lesens von Literatur in Disziplinen wie der Wirtschaftswissenschaft, Psychologie oder Linguistik durch zusätzliche Datenquellen wie die Durchführung von Experimenten oder Nutzung mathematischer Modelle ergänzt. Für die hermeneutischen Geisteswissenschaftlerinnen steht das Lesen im Mittelpunkt des wissenschaftlichen Arbeitens. Gelesen wird von einem Computermonitor oder gedruckten Texten, während das Lesen in allen Welten vom Notieren oder Annotieren begleitet wird.

Außerdem erlauben alle untersuchten Methodologien mit Ausnahme des qualitativ-hermeneutischen Arbeitens einen sukzessiven Arbeitsablauf. Die

Anfertigung einer Gliederung ist in diesen Fällen meistens durch typische Strukturen im Fach festgelegt und können nacheinander bearbeitet werden. Für hermeneutisch arbeitende Geisteswissenschaftlerinnen gestaltet es sich schwierig, derart strukturiert zu arbeiten: Ihre Gliederungen und Forschungsfragen entwickeln sich über die Zeit und mit laufender Recherche hinweg weiter. Zusammengefasst ist der hermeneutische Arbeitsablauf durch rekursive Aktivitäten gekennzeichnet, wobei der Abschluss jeder Aktivität jeweils das Fundament für den Beginn der nächsten Iteration darstellt.

4 Diskussion und Ausblick

Die vorgestellten Interviews unterstreichen den auch in der Literatur beschriebenen Stellenwert von Exzerpten für das wissenschaftliche Arbeiten. Bei näherer Betrachtung des Notierens und insbesondere des Literaturmanagements zeigt sich eine mögliche Erklärung für Probleme einiger Studierender bei der Nutzung von Literaturverwaltungssoftware: Das mentale Modell von bibliografischen Angaben eines Teils der Studierenden ist inkompatibel mit der Datenstruktur, wie sie von gängiger Software vorgegeben wird. Für Geisteswissenschaftlerinnen stellen Exzerpte den Einstiegspunkt in ihre Literatur dar. Außerdem wurden unterschiedliche Dateimanagement-Strategien beobachtet. Die Unterschiede in dieser Aktivität waren nicht derart auffällig wie bei der Literaturverwaltung.

Die Differenzierung zwischen Annotieren und Notieren könnte potenziell einen künstlichen Unterschied darstellen, da die beiden Aktivitäten möglicherweise in einer gegenseitigen Abhängigkeit stehen.

Es wurden die Ergebnisse einer explorativen Fallstudie vorgestellt; entsprechend gilt es nun, einen passenden Fragebogen zu entwickeln, um die vorgestellten Literaturverwaltungstypen und Erkenntnisse zur Exzerptproduktion sowie Aktivitäten und Arbeitsabläufe zu verifizieren. Bei einer positiven Überprüfung sollten die vorgestellten Ergebnisse bei der Entwicklung einer digitalen Forschungsumgebung für Geisteswissenschaftlerinnen Beachtung finden: Im Mittelpunkt einer solchen Software sollte zukünftig die Erstellung und Verwaltung von Exzerpten und ihrer Inhalte stehen. Weiterhin gilt es, die Ergebnisse zur Fragmentierung zwischen digitaler und analoger Welt im Bereich des geisteswissenschaftlichen Arbeitens weiter zu untersuchen, um bei der Entwicklung digitaler Tools potenzielle Affordanzen und Präferenzen

von bzw. für Papier zu beachten. Die digitale Erfassung analoger Vorgänge und das Stichwort „ubiquitous computing“ könnten hierfür interessante Lösungen darstellen. Für die Ausarbeitung generalisierbarer Anforderungen für solch eine Forschungsumgebung stellen die User Stories aus dem Text+-Projekt³ einen sinnvollen Anknüpfungspunkt für weitere Forschung dar.

Literatur

- Antonijević, Smiljana (2015): *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production*. New York, NY: Palgrave Macmillan.
- Barreau, D. K. 1995. Context as a factor in personal information management systems. *Journal of the American Society for Information Science* 46 (5): 327–339.
- Benardou, Agiatis; Constantopoulos, Panos; Dallas, Costis; Gavrilis, Dimitris (2010): Understanding the Information Requirements of Arts and Humanities Scholarship. *International Journal of Digital Curation* 5 (1): 18–33.
- Berners-Lee, Tim (1980): ENQUIRE V 1.1. Okt. 1980. <https://www.w3.org/History/1980/Enquire/manual/>
- Bernstein, Michael; Van Kleek, Max; Karger, David; Schraefel, M. C. (2008): Information Scraps: How and Why Information Eludes our Personal Information Management Tools. *ACM Transactions on Information and System Security* 26 (4): 24:1–24:46.
- Blanke, Tobias; Hedges, Mark (2013): Scholarly primitives: Building institutional infrastructure for humanities e-Science. *Future Generations Computer Systems: FGCS* 29 (2): 654–661.
- Boardman, Richard; Sasse, M. Angela (2004): ‘Stuff goes into the computer and doesn’t come out’: a cross-tool study of personal information management“. In: *Proceedings of the 2004 Conference on Human Factors in Computing Systems – CHI ’04*, S. 583–590. New York, NY: ACM Press.
- Bradley, John; Vetch, Paul (2007): Supporting Annotation as a Scholarly Tool—Experiences From the Online Chopin Variorum Edition. *Literary and Linguistic Computing* 22 (2): 225–241.
- Brockman, William S.; Neumann, Laura; Palmer, Carole L.; Tidline, Tonyia J. (2001): *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources.

3 <https://www.text-plus.org/en/research-data/user-stories-en/>

- Brush, A. J. Bernheim; Barger, David; Gupta, Anoop; Cadiz, J. J. (2001): Robust annotation positioning in digital documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '01* (S. 285–292). New York, NY: ACM Press.
- Bush, Vannevar (1945): As We May Think. *The Atlantic*, 1. Juli 1945. <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- Case, Donald Owen (1986): Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science and Engineering* 12 (3): 97–104.
- Case, Donald Owen (1991): The Collection and Use of Information by Some American Historians: A Study of Motives and Methods. *The Library Quarterly* 61 (1): 61–82.
- Cevolini, Alberto (2018): Exzerpieren. In: Heiko Christians, Nikolaus Wegmann, Matthias Bickenbach (Hrsg.): *Historisches Wörterbuch des Mediengebrauchs*, Bd. 2, S. 149–166. Köln: Böhlau Verlag.
- Corbin, Juliet M.; Strauss, Anselm L. (2008): *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 3. Aufl., Los Angeles, CA: Sage Publications.
- Dowling, Robyn; Wilson, Michael (2017): Digital doctorates? An exploratory study of PhD candidates' use of online tools. *Innovations in Education and Teaching International* 54 (1): 76–86.
- Ehlich, Konrad (1981): Zur Analyse der Textart Exzerpt. *Pragmatik. Theorie und Praxis (Amsterdamer Beiträge zur neueren Germanistik*; Bd. 13): 379–401.
- Gemmell, Jim; Bell, Gordon; Lueder, Roger (2006): MyLifeBits: a personal database for everything. *Communications of the ACM* 49 (1): 88–95.
- Jervis, Matthew (2014): How do people attempt to integrate the management of their paper and electronic documents? *Aslib Journal of Information Management* 66 (2): 134–155.
- Jones, William (2007): Personal Information Management. *Annual Review of Information Science and Technology* 41 (1): 453–504.
- Khan, Anam Ahmad (2019): Gaze assisted voice note taking system. In: *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers – UbiComp/ISWC '19*. New York, NY: ACM Press, S. 367 bis 371.
- Margetis, George; Ntoa, Stavroula; Antona, Margherita; Stephanidis, Constantine (2019): Augmenting natural interaction with physical paper in ambient intelligence environments. *Multimedia Tools and Applications* 78 (10): 13387–13433.

- Marshall, Catherine C. (1997): Annotation: from paper books to the digital library. In: *Proceedings of the Second ACM International Conference on Digital Libraries – DL '97*. New York, NY: ACM Press, S. 131–140.
- Milton, N. R. (2007): *Knowledge Acquisition in Practice: A Step-by-step Guide*. London: Springer,.
- Mizrachi, Diane; Bates, Marcia J. (2013): Undergraduates' personal academic information management and the consideration of time and task-urgency. *Journal of the American Society for Information Science and Technology* 64 (8): 1590 bis 1607.
- Norrie, Moira C.; Signer, Beat; Weibel, Nadir (2006): Print-n-link: weaving the paper web. In: *Proceedings of the 2006 ACM Symposium on Document Engineering, DocEng '06*. S. 34–43. New York, NY: ACM Press.
- Palmer, Carole L.; Neumann, Laura J. (2002): The Information Work of Interdisciplinary Humanities Scholars: Exploration and Translation. *The Library Quarterly: Information, Community, Policy* 72 (1): 85–117.
- Puhl, Johanna; Andorfer, Peter; Höckendorff, Mareike; Schmunk, Stefan; Stiller, Juliane; Thoden, Klaus (2015): *Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften* (DARIAH-DE Working Papers; 11). Göttingen: DARIAH-DE.
- Sellen, Abigail J.; Harper, Richard (2002): *The Myth of the Paperless Office*. Cambridge, MA: MIT Press.
- Strauss, Anselm L.; Corbin, Juliet M. (1990): *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage Publications.
- Toms, Elaine G. (2008): Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation* 64 (1): 102–130.
- Unsworth, John (2000): Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? Part of a symposium on “Humanities Computing: formal methods, experimental practice” sponsored by King’s College, London, May 13, 2000. <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>
- VERBI Software. 2019. MAXQDA 2020 (Version 2020). <https://www.maxqda.com/>
- Wellner, Pierre (1993): Interacting with paper on the DigitalDesk. *Communications of the ACM* 36 (7): 87–96.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 200–216. DOI: doi.org/10.5283/epub.44945.

Session 4:
Information Retrieval

(Un)bekannte Akteure auf der Suchergebnisseite?

Ein Vergleich zwischen selbst eingeschätzter und tatsächlich vorhandener Suchmaschinenkompetenz deutscher InternetnutzerInnen

Sebastian Schultheiß, Dirk Lewandowski

Hamburg University of Applied Sciences, Department of Information
Finkenau 35, 22081 Hamburg (Germany)

{[sebastian.schultheiss](mailto:sebastian.schultheiss@haw-hamburg.de), [dirk.lewandowski](mailto:dirk.lewandowski@haw-hamburg.de)}@haw-hamburg.de

Zusammenfassung

Kommerzielle Suchmaschinen sind ein wichtiges Rechercheinstrument für den Wissenserwerb in der Gesellschaft. Es ist wichtig, dass NutzerInnen verstehen, wie die Treffer in Suchmaschinen zustande kommen, damit sie fundierte Entscheidungen in ihrer Trefferauswahl treffen können. Die Darstellung und Reihenfolge der Treffer auf Suchergebnisseiten haben das Potenzial, Einfluss auf den Wissenserwerb der NutzerInnen zu nehmen. Dieser Einfluss erfolgt durch die Suchmaschinenbetreiber selbst sowie durch Akteursverbände aus Suchmaschinenoptimierung (SEO) und -werbung (SEA). Wie anhand der Daten aus einer repräsentativen Online-Befragung mit $n = 2.012$ deutschen InternetnutzerInnen gezeigt wird, verfügen diese über eine hohe Einschätzung der eigenen Suchmaschinenkompetenz. Die Einflüsse durch SEO und SEA sind jedoch dem Großteil der NutzerInnen nicht hinreichend bekannt. Dies gilt auch für diejenigen InternetnutzerInnen, die ihre Suchmaschinenkompetenz als hoch einschätzen. Dieser Widerspruch zwischen selbst eingeschätzter und tatsächlich vorhandener Suchmaschinenkompetenz ist problematisch. Fehlt die Kenntnis der Einflüsse durch SEO und SEA, fehlt damit auch die Grundlage für eine bewusste und kritische Ergebnisselektion sowie die Auseinandersetzung mit den Suchergebnissen. Aufgrund des hohen Stellenwerts von Suchmaschinen für den Wissenserwerb besteht somit der Bedarf, die Suchmaschinenkompetenz deutscher InternetnutzerInnen zu fördern und ein Bewusstsein externer Einflussgruppen zu schaffen.

Keywords: Suchmaschinen; Google; Suchmaschinenoptimierung; Suchmaschinenwerbung; Informationskompetenz

1 Einleitung

Suchmaschinen zählen zu den meistgenutzten Diensten des Internets und werden vom Großteil der Bevölkerung (76%) mindestens einmal wöchentlich genutzt. Damit finden sie sich an zweiter Stelle wieder, dicht hinter den Messenger-Diensten (80%) und noch vor dem Lesen oder Schreiben von E-Mails (65%). Auf den weiteren Rängen folgen mit großem Abstand die Nutzung von Online-Communitys (36%) sowie Blogs (5%) (Ergebnisse der ARD/ZDF-Onlinestudie 2020; Beisch/Schäfer, 2020). Neben den Suchmaschinenbetreibern und den NutzerInnen wirkt mit den Inhaltenanbietern ein weiterer Akteursverbund auf das Ranking ein. Die Inhaltenanbieter können die Sichtbarkeit ihrer Inhalte über Suchmaschinenoptimierung (*search engine optimization*; SEO) oder Suchmaschinenwerbung (*search engine advertising*; SEA) steigern (Röhle, 2010; s. Abb. 1).

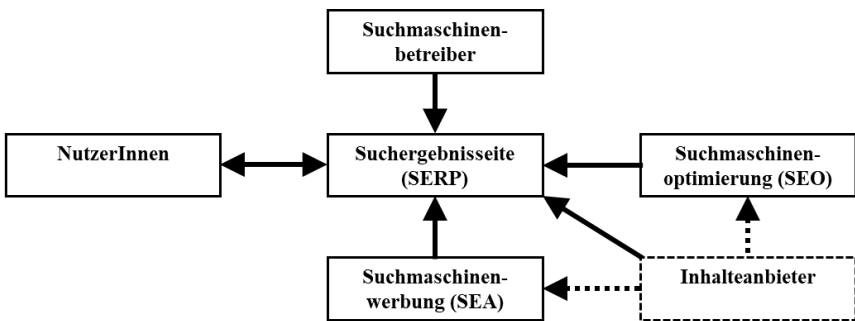


Abb. 1 Akteursverbände mit Einfluss auf die Suchergebnisseite

SEO und SEA sind Bestandteile des Online-Marketings (Griesbaum, 2013). SEO beschreibt Maßnahmen, die das Ziel haben, den Ranking-Algorithmen von Suchmaschinen möglichst gut zu entsprechen, um so eine verbesserte Platzierung innerhalb der organischen Suchergebnisse zu erreichen (ebd.). Bei Suchmaschinenwerbung (SEA) handelt es sich hingegen um zur Suchanfrage passende und gekennzeichnete Anzeigen, die ausgespielt werden, insofern ein Werbetreibender ein entsprechendes Gebot abgegeben hat (Google.com, 2020; Lewandowski, 2018). Die Bezahlung erfolgt auf Grundlage der Klicks (Google.com, 2020). Durch Anzeigen für Produkte und Dienstleistungen generiert Alphabet Inc., der Mutterkonzern Googles, den Großteil seiner Einnahmen. Von den Gesamteinnahmen in Höhe von 162

Milliarden US-Dollar wurden im Jahr 2019 83% (135 Milliarden US-Dollar) über Werbung erwirtschaftet (Alphabet Inc., 2020). Auf das, was die NutzerInnen auf den Suchergebnisseiten (*search engine result pages*; SERPs) zu sehen bekommen, wirken somit erhebliche externe Einflüsse.

Werden SuchmaschinennutzerInnen nach ihren eigenen Fähigkeiten im Kontext der Suchmaschinennutzung befragt, äußern sie meist ein hohes Vertrauen in die eigene Suchmaschinenkompetenz (s. Abschn. 3). Hinzu kommt, dass sie ein hohes Vertrauen in Suchmaschinen haben, ihnen die „besten“ Ergebnisse zuerst aufzuführen (u.a. Kammerer/Gerjets, 2014; Pan et al., 2007). Trotz der alltäglichen Nutzung und der selbstbewussten Einschätzung der eigenen Suchmaschinenkompetenz sind gering ausgeprägte Kenntnisse hinsichtlich der Funktionsweise (Stark et al., 2014) und des Geschäftsmodells Googles (Lewandowski et al., 2018) sowie der Lösung komplexer Rechercheaufgaben (u.a. Singer et al., 2012) zu verzeichnen. Es liegt somit eine Überschätzung der eigenen Informationskompetenz (IK) vor, wie sie in Studien häufig beobachtet und als Dunning-Kruger-Effekt bezeichnet wird (Mahmood, 2016). Daraus lässt sich ableiten, dass die NutzerInnen mit Suchmaschinen konfrontiert werden, deren vielfältige Einflussgruppen ihnen nur unzureichend bekannt sind. Als ursächlich hierfür kann gelten, dass Suchmaschinen trotz ihrer alltäglichen Präsenz eine Form der Infrastruktur darstellen, deren Komplexität durch die NutzerInnen kaum wahrnehmbar und durchdringbar ist (Sundin, 2020; Sundin et al., 2017).

Die fehlende Auseinandersetzung mit Suchmaschinen ist als problematisch zu erachten. So wirken zahlreiche Akteure mit unterschiedlichen Eigeninteressen auf die Suchergebnisseite ein, wodurch eine potenzielle Einflussnahme auf den Wissenserwerb der NutzerInnen entsteht. In einem Experiment mit $n = 297$ TeilnehmerInnen stellte Tremel (2010, S. 230) fest, dass knapp die Hälfte der Versuchspersonen Werbung als Faktenwissen nutzte. Bei unzureichenden Anzeigenkenntnissen kann zudem bereits der Primäreffekt (Primacy-Effekt; im Kontext des Klickverhaltens s. Murphy et al., 2006) dafür sorgen, dass kommerziell motivierte Botschaften bevorzugt aufgenommen werden und die Bewertung der organischen Ergebnisse dann in Form eines Vergleichs mit den bisher aufgenommenen Informationen erfolgt. Sucht ein Nutzer beispielsweise nach Behandlungsmöglichkeiten für Kopfschmerzen und erhält zuerst Anzeigen von Pharmaunternehmen, die eine medikamentöse Behandlung empfehlen, wird er folgende Ergebnisse (die vielleicht wirksame Hausmittel wie Entspannungsübungen und Spaziergänge an der frischen Luft empfehlen) *im Vergleich zu den Anzeigenergebnissen*

betrachten (White/Horvitz, 2014). Zusammen mit der Tendenz der NutzerInnen, dem Ranking der Suchmaschinen zu vertrauen (Purcell et al., 2012), wird dieser Nutzer die Relevanz/Vertrauenswürdigkeit der Anzeigenergebnisse höher einschätzen als die der ersten organischen Ergebnisse. Würde derselbe Nutzer hingegen den Unterschied zwischen Anzeigen und organischen Ergebnissen verstehen, würde er wahrscheinlich zuerst – oder gar nur – die organischen Ergebnisse in Betracht ziehen, zuerst also die Empfehlung der Hausmittel erhalten. Würde er später die Anzeigen ansehen, würde er diese Ergebnisse mit den bereits bekannten Aussagen aus den organischen Ergebnissen vergleichen.

Kenntnisse über SEO zu fördern, ist zum einen aufgrund der Dimension der SEO-Branche von Bedeutung. So wird allein in den USA für das Jahr 2020 ein Umsatz in Höhe von 80 Milliarden US-Dollar für die SEO-Branche erwartet (McCue, 2018). Zum anderen gilt der Einsatz von SEO-Maßnahmen aus Sicht von Inhabern als unabdingbar, um die Sichtbarkeit ihrer Inhalte zu gewährleisten. SEO ist somit im Großteil der Websuchen präsent, ohne für die NutzerInnen klar erkennbar zu sein (Schultheiß/Lewandowski, 2020).

Die Vermittlung grundlegender Kenntnisse über Suchmaschinen, zusammengefasst als Suchmaschinenkompetenz, tritt im Rahmen von Schulungen der Informationskompetenz (IK) nur unzureichend auf (Informationskompetenz.de, 2018).¹ Durch die alltägliche Verwendung von Suchmaschinen und ihren erheblichen Einfluss auf den Wissenserwerb stellt die Suchmaschinenkompetenz jedoch die Basis eines kompetenten Umgangs mit den Suchwerkzeugen dar und sollte den Ausgangspunkt von IK-Schulungen bilden (Lewandowski, 2016).

In diesem Beitrag werden Ergebnisse einer repräsentativen Online-Befragung deutscher InternetnutzerInnen berichtet. Darin wurde die Selbsteinschätzung der Suchmaschinenkompetenz erfragt sowie Kenntnisse zu SEO und SEA erhoben, welche die tatsächlich vorliegende Suchmaschinenkompetenz operationalisiert. Dadurch wird der zentralen Fragestellung nachgegangen, welche Zusammenhänge zwischen der selbst eingeschätzten und der

1 Suchmaschinen werden in der IK-Statistik 2017, für die 80 Bibliotheken befragt wurden, unter der Kategorie Internetrecherche aufgeführt. Die Gesamtzahl der Nennungen des inhaltlichen Schwerpunkts Internetrecherche ($n = 3.088$) bewegt sich weit hinter bibliothekarischen Themenschwerpunkten wie Bibliotheksnutzung ($n = 10.131$), Datenbanken ($n = 9.039$) und Fernleihe ($n = 3.837$).

erhobenen Suchmaschinenkompetenz deutscher InternetnutzerInnen bestehen. Die Ergebnisse zeigen, dass auch bei NutzerInnen mit hoher Selbsteinschätzung wesentliche Defizite hinsichtlich der Kenntnis über die Einflussnahmen durch SEO und SEA auszumachen sind. Diese sind als problematisch zu erachten, da Kenntnisse über die Akteure auf der Suchergebnisseite sowie deren Motivationen die Grundlage für eine kritische Auseinandersetzung mit den Suchergebnissen sind.

2 Nutzerverhalten auf Suchergebnisseiten von Web-Suchmaschinen

Google ist die am häufigsten verwendete Suchmaschine mit europaweiten Marktanteilen von über 90% (European Commission, 2017). Zunehmend werden die Suchanfragen, mittlerweile über 50%, dabei über mobile Endgeräte wie Smartphones abgegeben (Sterling, 2016).

Das Auswahlverhalten der SuchmaschinennutzerInnen wird stark durch die Position, Sichtbarkeit und Darstellung eines Suchergebnisses beeinflusst. So bevorzugen NutzerInnen die zuerst platzierten Ergebnisse (Granka et al., 2004; Petrescu, 2014), Treffer innerhalb des direkt (also ohne zu scrollen) sichtbaren Bereichs (Cutrell/Guan, 2007; Sachse, 2019) sowie Suchergebnisse mit ansprechender, beispielsweise bebildeter Gestaltung (Liu et al., 2015). Daneben verfügen die NutzerInnen über ein hohes Maß an Vertrauen in die Suchmaschinen. In einer repräsentativen Befragung US-amerikanischer InternetnutzerInnen antworteten 73% der Befragten, dass die meisten oder gar alle Informationen, die sie über Suchmaschinen finden, korrekt und vertrauenswürdig seien. Als faire und unvoreingenommene Informationsquelle wurden Suchmaschinen von 66% der Befragten bezeichnet (Purcell et al., 2012). Auch Eye-Tracking-Studien wiesen das hohe Vertrauen der NutzerInnen in Google, genauer gesagt in die Ergebnisreihenfolge, nach. Die TeilnehmerInnen der Studien wählten die ersten Ergebnisse auch dann aus, wenn diese durch experimentelle Manipulation von geringerer Relevanz oder Vertrauenswürdigkeit waren. Die Position hatte somit den stärksten Einfluss auf das Auswahlverhalten (Kammerer/Gerjets, 2014; Pan et al., 2007). In einer Replikationsstudie der Untersuchung von Pan et al. (2007) bestätigten Schultheiß, Sünkler und Lewandowski (2018) den großen Einfluss der Ergebnisrei-

henfolge auf das Blick- und Auswahlverhalten der ProbandInnen. Ausschlaggebend für die Auswahl eines Treffers war in dieser Studie jedoch die Relevanz des Ergebnisses. Ähnliche Ergebnisse wie die der Studie von Pan et al. (2007) erbrachte auch eine Logfile-Analyse von Keane, O'Brien und Smyth (2008). Trotz umgekehrter Ergebnisreihenfolge erfolgte die Mehrzahl der Klicks auf dem erstplatzierten Ergebnis. Nur wenige NutzerInnen schienen somit nach dem eigentlich erstplatzierten (nun zuletzt platzierten) Ergebnis gesucht zu haben.

Es lässt sich somit zusammenfassen, dass in der Nutzung kommerzieller Suchmaschinen überwiegend Google genutzt und stark darin vertraut wird, dass diese Suchmaschine die relevantesten Ergebnisse zuerst platziert. Dabei bleibt unklar, wie diese NutzerInnen zu erkennen meinen, dass Google ihnen die relevantesten Ergebnisse liefert, da ihnen zumindest für informationsorientierte Suchanfragen kein direkter Vergleichsmaßstab vorliegt. Die Sichtung der Suchergebnisse beschränkt sich in der Regel auf die vorderen Ränge mit nur geringer Bereitschaft, zu unteren Ergebn isrängen zu scrollen oder andere Ergebnisseiten als die erste in Betracht zu ziehen.

3 Informationskompetenz von Suchmaschinennutzern

Informationskompetenz (IK) kann als Fähigkeit definiert werden, zu erkennen, wann Informationen benötigt werden, sowie die benötigten Informationen zu finden, auszuwerten und effektiv zu nutzen (American Library Association, 1989). Leitlinien und Ratgeber in- und außerhalb des Bibliotheksbereichs zielen darauf ab, diese Fähigkeiten zu vermitteln (u. a. Lau, 2008) sowie eigenständig zu erwerben (in Bezug auf Suchmaschinen u. a. Machill, 2009; Machill et al., 2012). Aufgrund der Nutzungshäufigkeit und dem Einfluss von Suchmaschinen auf den Wissenserwerb ist die Suchmaschinenkompetenz als ein wesentlicher Baustein der Informationskompetenz zu betrachten. Neben Kompetenzen, um den eigenen Suchprozess zu optimieren, stehen hier insbesondere die Kenntnisse über den Aufbau und die Funktionsweise der Suchmaschinen im Fokus (Lewandowski, 2016).

Nutzerbefragungen zeigen, dass SuchmaschinennutzerInnen ihre eigenen Fähigkeiten im Umgang mit Suchmaschinen als hoch einschätzen. So gaben

91% der befragten US-amerikanischen InternetnutzerInnen an, üblicherweise das Gesuchte zu finden. Ein stark ausgeprägtes Zutrauen in die eigenen Recherchefähigkeiten äußerten 56% der Befragten (Purcell et al., 2012). Werden die tatsächlich vorhandenen Fähigkeiten und Kenntnisse hinsichtlich Suchmaschinen erhoben, zeigt sich jedoch ein anderes Bild. So zeigten Studien, dass ProbandInnen die notwendige Präzision in der Formulierung ihrer Suchanfragen vermissen lassen (Dragovic et al., 2016; Stark et al., 2014), unzureichende Kenntnisse über die Ranking-Kriterien von Suchmaschinen aufweisen (Stark et al., 2014) und trotz positiver Einschätzung des eigenen Rechercheerfolgs Schwierigkeiten beim Lösen komplexer Rechercheaufgaben haben (Aufenanger/Siebertz, 2014; Singer et al., 2012). Eine repräsentative Umfrage deutscher InternetnutzerInnen ergab, dass nur etwa 61% der Befragten die Werbung als einziges Geschäftsmodell Googles bekannt ist. Darüber hinaus ist der Großteil der InternetnutzerInnen nicht in der Lage, verlässlich zwischen Anzeigen und organischen Ergebnissen zu unterscheiden. Dieselben Befragten gaben jedoch meist (91%) an, über gute oder gar sehr gute Fähigkeiten im Umgang mit Google zu verfügen (Lewandowski et al., 2018). Auch außerhalb der Suchmaschinenkompetenz liegt eine Überschätzung der eigenen Fähigkeiten vor. Mahmood (2016) führte eine Literaturschau von 53 Studien zur Selbsteinschätzung der Informationskompetenz durch, wobei nicht näher auf die in den Studien eingesetzten Suchwerkzeuge eingegangen wurde. In den meisten Studien überstieg die Selbsteinschätzung die tatsächliche Informationskompetenz der TeilnehmerInnen („Dunning-Kruger-Effekt“). Zu ähnlichen Ergebnissen kamen Douglas et al. (2014), Ngo et al. (2019) sowie Rust et al. (2017) bei Studierenden sowie Cullen et al. (2011) bei sich in Ausbildung befindlichen Ärzten und der Recherche in medizinischen Datenbanken.

Es lässt sich somit zusammenfassen, dass bei kommerziellen Suchmaschinen wie auch anderen Informationssystemen die hohen Selbsteinschätzungen der NutzerInnen nicht mit deren tatsächlich vorhandenen Informations- bzw. Suchmaschinenkompetenz einhergehen.

4 Forschungsfragen

Die erste Forschungsfrage (FF1) zielt auf die in Befragungen häufig erhobene Selbsteinschätzung der eigenen Suchmaschinenkompetenz (u. a. Lewandows-

ki, 2017; Schweiger, 2003) ab. Kenntnisse über die Suchwerkzeuge und somit im Falle Googles über die Akteure auf der Suchergebnisseite sind Voraussetzung für einen kompetenten Umgang mit Suchmaschinen, sodass die Akteure der Bereiche SEO und SEA in der Erhebung der Suchmaschinenkompetenz (FF2) im Vordergrund stehen. Forschungsfrage FF3 dient schließlich dem Vergleich zwischen Selbsteinschätzung und tatsächlichem Kenntnisstand und weitergefasst auch der Frage, inwieweit derartige Selbstauskünfte zur Suchmaschinenkompetenz verlässliche Ergebnisse liefern.

FF1: *Welche Selbsteinschätzung haben deutsche InternetnutzerInnen hinsichtlich ihrer eigenen Suchmaschinenkompetenz?*

FF2: *Welche Kenntnisse haben deutsche InternetnutzerInnen von Suchmaschinenoptimierung (SEO) und Suchmaschinenwerbung (SEA) und wie zuverlässig können sie die Suchergebnisse mit beiden Bereichen in Verbindung bringen?*

FF3: *Welche Zusammenhänge bestehen zwischen der selbst eingeschätzten und der erhobenen Suchmaschinenkompetenz deutscher InternetnutzerInnen?*

5 Methode

Zur Beantwortung der Forschungsfragen wurde eine repräsentative Online-Befragung mit $n = 2.012$ deutschen InternetnutzerInnen zwischen 16 und 69 Jahren durchgeführt, wobei bei der Samplebildung den Kriterien der Arbeitsgemeinschaft Onlineforschung (AGOF) gefolgt wurde.²

Aufgrund der im Forschungsstand geschilderten zunehmenden Bedeutung der mobilen Suche nahmen die Befragten zur Hälfte am großen (Desktop-PC, Laptop, Tablet) oder am kleinen Bildschirm (Smartphone) teil. Die Befragung fand in Zusammenarbeit mit dem Marktforschungsunternehmen Fittkau & Maaß Consulting³ statt. Die Online-Umfrage wurde im Rahmen des Projekts „SEO-Effekt“⁴ durchgeführt. Die vollständige Online-Umfrage enthält Abschnitte, die über den Fokus dieses Beitrags hinausgehen. In diesem Bei-

2 <https://www.agof.de/>

3 <https://www.fittkaumaass.de/>

4 <https://searchstudies.org/seo-effekt/>

trag werden die Ergebnisse der Fragebogenabschnitte zur Selbsteinschätzung, zu SEA- und SEO-Kenntnissen sowie zur Identifikation von Suchergebnissen, die durch SEA sowie SEO beeinflussbar sind, behandelt. Andere Abschnitte, wie z. B. zur Personalisierung oder zu weiteren Nutzungsgewohnheiten, bleiben unberücksichtigt.

5.1 Fragebogen

Der Fragebogen wird in Tabelle 1 dargestellt und beinhaltet vier Abschnitte. Im ersten Abschnitt wurden die Befragten nach ihrer Selbsteinschätzung der eigenen Suchmaschinenkompetenz befragt. Daran schließen sich die Abschnitte II und III mit Fragen zum Kenntnisstand über SEA und SEO an. Im vierten und letzten Abschnitt folgen schließlich Screenshot-basierte Markierungsaufgaben, bei denen je nach Aufgabenstellung entweder durch SEO oder durch SEA beeinflussbare Ergebnisse zu identifizieren waren (s. dazu auch Abschn. 5.2). Vor dem vierten Abschnitt erhielten die Befragten einen Informationstext mit Beschreibungen zu SEO und SEA, um die Verständlichkeit der darauffolgenden Markierungsaufgaben zu erleichtern. Der Informationstext lautet:

Website-Betreiber haben verschiedene Möglichkeiten darauf hinzuwirken, dass ihre Webseiten bei bestimmten Google-Suchanfragen weiter oben auf der Seite erscheinen, und zwar:

- Bezahlung: Sie zahlen dafür Geld an Google.^[5]
- Suchmaschinenoptimierung: Sie gestalten/programmieren ihre Webseiten entsprechend, z.B. durch die Verwendung bestimmter Begriffe, schnelle Ladezeiten, sinnvolle Bildbeschriftungen.

Wir zeigen Ihnen gleich zwei verschiedene Google-Ergebnisseiten und würden gern von Ihnen erfahren, ob bzw. welche der dort gezeigten Suchergebnisse durch Bezahlung an Google und/oder Suchmaschinenoptimierung beeinflussbar sind.

Die Fragen der Online-Befragung wurden zum Teil an eine vorherige Untersuchung zum Verständnis deutscher InternetnutzerInnen von SEA (Lewandowski et al., 2018) angelehnt. Darüber hinaus wurden demografische Daten erhoben.

5 Hierbei handelt es sich um eine vereinfachte Erläuterung im Sinne der Verständlichkeit des Fragebogens. Wie in der Einleitung erläutert, erfolgt die Bezahlung nicht auf Grundlage der Platzierung, sondern anhand der auf den Anzeigen erfolgten Klicks.

Tab. 1: Fragebogen der Online-Befragung

Abschnitt	Nr. Fragestellung	Antwortmöglichkeiten
I) Selbst- einschätzung	1.1* Wenn es darum geht, mithilfe von Suchmaschinen etwas im Internet zu finden: Wie schätzen Sie Ihre eigenen Fähigkeiten diesbezüglich ein?	Meine Fähigkeiten bei der Suchmaschinenutzung sind ... sehr gut gut befriedigend eher schlecht sehr schlecht. Weiß ich nicht.
	II) Kenntnisse: SEA	2.1* Was denken Sie: Wodurch generiert Google den Großteil seiner Einnahmen? 2.2* Haben Website-Betreiber bzw. Unternehmen Ihrer Kenntnis nach die Möglichkeit, dafür zu bezahlen, dass sie bzw. ihre Produkte auf der Suchergebnisseite von Google weit oben erscheinen? 2.3* Sind die eben erwähnten bezahlten Suchergebnisse Ihrer Meinung nach von den übrigen Suchergebnissen zu unterscheiden? 2.4* Und wodurch unterscheiden sich die bezahlten Suchergebnisse auf Google von den übrigen Ergebnissen, für die nicht bezahlt worden ist?
III) Kenntnisse: SEO	3.1 Haben Website-Betreiber bzw. Unternehmen Ihrer Meinung nach Möglichkeiten bzw. Einfluss darauf, bei bestimmten Suchanfragen in der Google-Ergebnisliste weiter oben zu erscheinen, ohne dafür an Google Geld zu bezahlen?	Ja, solche Möglichkeiten gibt es. Nein, solche Möglichkeiten gibt es nicht. Weiß ich nicht.
	3.2 Wissen Sie, mit welchem Begriff diese Maßnahmen zur Verbesserung der Platzierung in der Google-Suchergebnisliste (ohne Bezahlung an Google) bezeichnet werden?	Ja, man nennt das: ... offene Eingabe: z. B. SEO, Suchmaschinenoptimierung, Search Engine Optimization Weiß ich nicht.

Abschnitt	Nr. Fragestellung	Antwortmöglichkeiten
	3.3 Und mit welchen Maßnahmen kann eine Webseite so gestaltet bzw. programmiert werden, dass sie in den Google-Suchergebnislisten weiter oben aufgeführt wird?	Bitte tragen Sie hier alle Möglichkeiten/Maßnahmen ein, die Sie kennen. Mithilfe folgender Maßnahmen: offene Eingabe: z. B. passende Keywords, optimierte Ladezeiten, ... Weiß ich nicht.
<i>Infoblock „SEO/SEA“</i>		
IV) Kenntnisse: Identifikation SEO/ SEA	4.1* Kommen wir nun zu der ersten Google-Ergebnisseite. Existieren auf dieser Seite Suchergebnisse, auf die Einfluss genommen werden kann, indem Google dafür vom Website-Betreiber bezahlt wird?	Nein, es gibt auf dieser Seite keine Suchergebnisse, die durch Zahlungen an Google beeinflusst werden können. Ja, auf folgende Suchergebnisse kann Einfluss genommen werden, indem dafür Geld an Google gezahlt wird. <i>Klicken Sie bitte auf die entsprechenden Suchergebnisse.</i>
	4.2 Noch eine weitere Frage zu dieser Suchergebnisseite: Gibt es hierauf auch Suchergebnisse, auf die mithilfe von Suchmaschinenoptimierung Einfluss genommen werden kann?	Nein, es gibt auf dieser Seite keine Suchergebnisse, die durch Suchmaschinenoptimierung beeinflusst werden können. Ja, auf folgende Suchergebnisse kann durch Suchmaschinenoptimierung Einfluss genommen werden. <i>Klicken Sie bitte auf die entsprechenden Suchergebnisse.</i>
	4.3* Kommen wir nun zu unseren Fragen zur zweiten (und letzten) Google-Ergebnisseite. Existieren auf dieser Seite Suchergebnisse, auf die Einfluss genommen werden kann, indem Google dafür vom Website-Betreiber bezahlt wird?	Nein, es gibt auf dieser Seite keine Suchergebnisse, die durch Zahlungen an Google beeinflusst werden können. Ja, auf folgende Suchergebnisse kann Einfluss genommen werden, indem dafür Geld an Google gezahlt wird. <i>Klicken Sie bitte auf die entsprechenden Suchergebnisse.</i>
	4.4 Noch eine weitere Frage zu dieser Suchergebnisseite: Gibt es hierauf auch Suchergebnisse, auf die mithilfe von Suchmaschinenoptimierung Einfluss genommen werden kann?	Nein, es gibt auf dieser Seite keine Suchergebnisse, die durch Suchmaschinenoptimierung beeinflusst werden können. Ja, auf folgende Suchergebnisse kann durch Suchmaschinenoptimierung Einfluss genommen werden. <i>Klicken Sie bitte auf die entsprechenden Suchergebnisse.</i>

* Fragestellung angelehnt an Lewandowski et al. (2018)

5.2 Markierungsaufgaben

Für die Markierungsaufgaben wurden zwei Blöcke mit insgesamt vier Aufgabenstellungen erstellt. Aufgaben A und B wurden dem Block I (einfach), Aufgaben C und D dem Block II (komplex) zugeordnet. Die zwei Blöcke dienten dazu, SERPs mit unterschiedlich komplexen Zusammensetzungen zu testen. Die Struktur der beiden SERPs ist je Block hinsichtlich der sich darauf befindlichen Elemente identisch. Tabelle 2 zeigt die Blöcke und Aufgaben inklusive der zugehörigen Suchanfragen sowie SERP-Elemente.

Tab. 2: Markierungsaufgaben: Suchanfragen und SERP-Elemente

Block	Aufgabe	Suchanfrage	SERP-Elemente
I (einfach)	A	steuererklärung hilfe	organische Ergebnisse,
	B	rechtsberatung	Textanzeigen (am oberen und unteren Ende der SERP)
II (komplex)	C	apple iphone	organische Ergebnisse,
	D	samsung galaxy	Textanzeigen (am oberen Ende der SERP), Shoppinganzeigen, News, Knowledge Graph

Da alle Aufgaben auf zwei Endgeräten (großer und kleiner Bildschirm) getestet wurden, mussten insgesamt acht SERP-Screenshots erstellt werden, also Screenshots der Aufgaben A–D in den Formaten „großer Bildschirm“ und „kleiner Bildschirm“. Die Teilnehmenden erhielten jeweils zwei Aufgaben, eine aus Block I und eine aus Block II, jeweils zufällig zugewiesen (z.B. Aufgaben A und D im Format „kleiner Bildschirm“, wenn der Zugriff auf die Befragung per Smartphone stattfand). Die SERP jeder Aufgabe wurde zweimal angezeigt. Zunächst sollten darauf alle Ergebnisse des Einflussbereichs SEA markiert werden (d.h. Anzeigen) und im Anschluss alle Ergebnisse, die durch SEO beeinflussbar sind (d.h. organische Suchergebnisse), wie in Tabelle 3 dargestellt wird. Insgesamt nahm jede an der Umfrage teilnehmende Person somit Markierungen auf vier SERPs vor (z.B. Aufgaben A und D, separate Markierung von Anzeigen und organischen Ergebnissen).

Die Screenshots wurden mit der Desktopversion des Chrome-Browsers und darin integrierten Erweiterungen erstellt. Die Erweiterungen dienten der Anpassung der Fenstergröße⁶ und des User-agents⁷ zur Simulation des An-

6 <https://chrome.google.com/webstore/detail/window-resizer/kkelicaakdanhinjdeammmlgefongfh>

droid-Betriebssysteme sowie der Erstellung der Screenshots⁸. Um die Screenshots nach der Erstellung auf die zu untersuchenden Elemente (s. Tab. 3) zu reduzieren, nutzten wir die Software GIMP⁹ in der Version 2.10.14. Hierbei wurden die beiden zu einer Aufgabe gehörenden Screenshots (großer und kleiner Bildschirm) aufeinander angeglichen, sodass dieselben Ergebnisse an derselben Position der jeweiligen SERP platziert waren. Andernfalls wäre ein Vergleich der Markierungsaufgaben erschwert und Unterschiede im Markierungserfolg nicht mehr auf die unterschiedliche Darstellung, sondern auf gegebenenfalls vorhandene inhaltliche Unterschiede der Treffer und deren Reihenfolgen zurückzuführen.

Tab. 3: Markierungsaufgaben: zu markierende Ergebnisse

Aufgabe	Endgerät	Bereich	Zu markierende Suchergebnisse
A	großer und kleiner Bildschirm	SEO	organische Ergebnisse (10×)
A	großer und kleiner Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Textanzeigen, unteres SERP-Ende (2×)
B	großer und kleiner Bildschirm	SEO	organische Ergebnisse (10×)
B	großer und kleiner Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Textanzeigen, unteres SERP-Ende (2×)
C	großer Bildschirm	SEO	organische Ergebnisse (6×) News (3×)
C	großer Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Shoppinganzeigen (8×)
C	kleiner Bildschirm	SEO	organische Ergebnisse (6×) News (2×)
C	kleiner Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Shoppinganzeigen (2×)
D	großer Bildschirm	SEO	organische Ergebnisse (6×) News (3×)
D	großer Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Shoppinganzeigen (8×)
D	kleiner Bildschirm	SEO	organische Ergebnisse (6×) News (2×)
D	kleiner Bildschirm	SEA	Textanzeigen, oberes SERP-Ende (2×) Shoppinganzeigen (2×)

7 <https://chrome.google.com/webstore/detail/user-agent-switcher-for-c/djflhoibgkdhkhcedjklpkjnoahfmg>

8 <https://chrome.google.com/webstore/detail/full-page-screen-capture/fdpohaocaechifmbbbkbknoalclal>

9 <https://www.gimp.org/>

In Abbildung 2 wird ein Ausschnitt der SERP zu Aufgabe B im großen Darstellungsformat abgebildet. Die hier zu markierenden SEA-Ergebnisse sind durch eine gestrichelte Linie gekennzeichnet, die SEO-Ergebnisse durch eine durchgezogene Linie. Für die weiteren SERP-Screenshots siehe Abschnitt *Forschungsdaten*.

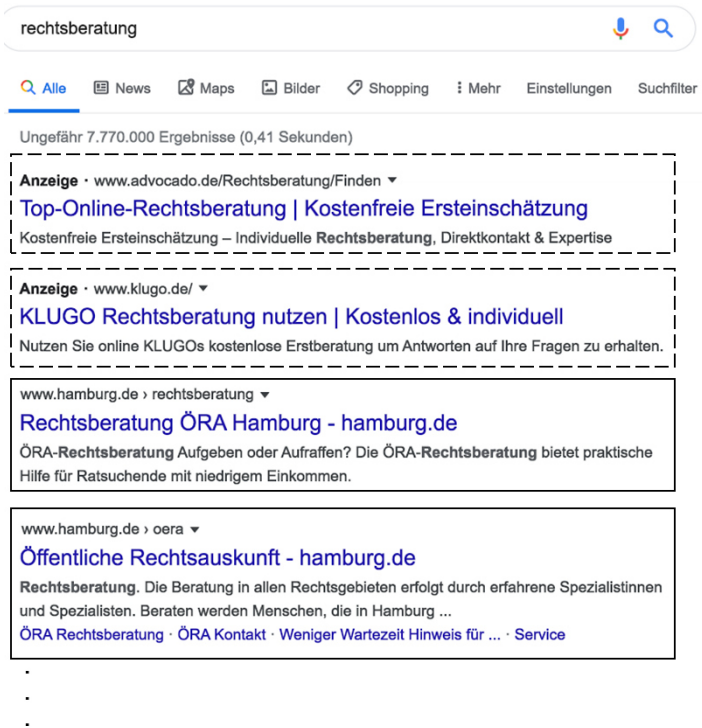


Abb. 2 SERP für Aufgabe B, großer Bildschirm
(SEA-Ergebnisse: gestrichelte Linie, SEO-Ergebnisse: durchgezogene Linie)

Auf Grundlage der jeweils markierten Ergebnisse wurde für alle ProbandInnen je Aufgabe (A–D), Bildschirmgröße (groß, klein) sowie Bereich (SEO, SEA) eine Identifikationsquote berechnet. Diese Quote berücksichtigt die richtig positiv sowie die falsch positiv markierten Suchergebnisse anhand der Formel

$$\frac{n(\text{richtig}) - n(\text{falsch})}{n(\text{zu markieren})}$$

Es folgen zwei Beispiele. Das erste Beispiel beschreibt den Fall, in dem eine positive Identifikationsquote für Aufgabe A, großer Bildschirm, SEO-

Ergebnisse erzielt wird. In diesem Fall sind zehn organische Ergebnisse zu markieren, von denen die beispielhafte Testperson acht Ergebnisse korrekt markiert (acht richtig). Darüber hinaus markiert die Testperson fälschlich zwei Anzeigen als organische Ergebnisse (zwei falsch). Daraus ergibt sich eine Identifikationsquote von 0,6. Das zweite Beispiel betrifft einen Fall, in dem eine negative Identifikationsquote erreicht wird. Negative Quoten entstehen dann, wenn ein Proband mehr falsche als richtige Markierungen vornimmt, z.B. für Aufgabe B, kleiner Bildschirm, SEA-Ergebnisse. In diesem Fall sind insgesamt vier Anzeigen zu markieren. Wenn eine Testperson alle vier Anzeigen identifiziert (vier richtig), aber zusätzlich sechs organische Ergebnisse fälschlicherweise markiert (sechs falsch), erreicht die Person eine Identifikationsquote von $-0,5$ bei dieser Markierungsaufgabe.

5.3 Codierung offener Fragen

Tabelle 4 stellt die Codiervorgaben der offenen Wissensfragen dar, deren Antworten in „richtig“ und „falsch“ unterschieden wurden. Die Codierung wurde von einer Person vorgenommen. Dies wurde als ausreichend erachtet, da kein nennenswerter Interpretationsspielraum bei der Codierung bestand.

Tab. 4: Codierung offener Fragen

Nr.	Fragestellung	Codierung
2.1	Was denken Sie: Wodurch generiert Google den Großteil seiner Einnahmen?	richtig: Werbung o.Ä. (z.B. advertisement, sponsored results, search engine advertising, SEA, paid search marketing) falsch: vollständig falsche Eingabe (z.B. Datenverkauf, Spenden)
2.4	Und wodurch unterscheiden sich die bezahlten Suchergebnisse auf Google von den übrigen Ergebnissen, für die nicht bezahlt worden ist?	richtig: Anzeigenkennzeichnung o.Ä. (z.B. Anzeigenbegriff, Kennzeichnung, Markierung) falsch: vollständig falsche Eingabe (z.B. unterschiedliche Schriftart)
3.2	Wissen Sie, mit welchem Begriff diese Maßnahmen zur Verbesserung der Platzierung in der Google-Suchergebnisliste (ohne Bezahlung an Google) bezeichnet werden?	richtig: Suchmaschinenoptimierung o.Ä. (z.B. search engine optimization, SEO) falsch: vollständig falsche Eingabe (z.B. Anzeigen, Bots)
3.3	Und mit welchen Maßnahmen kann eine Webseite so gestaltet bzw. programmiert werden, dass sie in den Google-Suchergebnislisten weiter oben aufgeführt wird?	richtig: Keyword-Platzierung oder andere SEO-Techniken falsch: vollständig falsche Eingabe (z.B. Bezahlung, Anzeigen)

6 Ergebnisse

Die Ergebnisse werden zunächst anhand der einzelnen Fragebogenabschnitte, dann im Vergleich der selbst eingeschätzten mit der gemessenen Suchmaschinenkompetenz dargestellt.

6.1 Selbsteinschätzung der Internetnutzer

Die Online-Umfrage wurde von $n = 2.012$ Personen absolviert, 51,6% davon männlich und 48,4% weiblich. Wie Abbildung 3 zeigt, schätzten mehr als die Hälfte der deutschen InternetnutzerInnen (52,4%) ihre eigenen Fähigkeiten in der Suchmaschinennutzung als „gut“, ein knappes Drittel (31,9%) als „sehr gut“ ein.

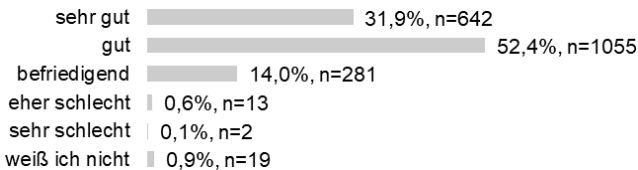


Abb. 3 Selbsteinschätzung der Suchmaschinenkompetenz

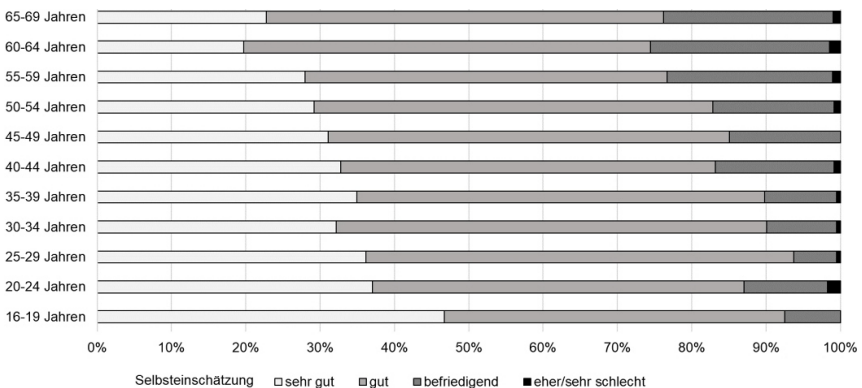


Abb. 4 Selbsteinschätzung nach Altersgruppen

Wie Abbildung 4 zeigt, ist bei steigendem Alter eine niedrigere Selbsteinschätzung zu beobachten, jedoch bleibt diese auf hohem Niveau. So schätzten in der jüngsten Altersgruppe der 16- bis 19-Jährigen 93% ihre Suchmaschinenkompetenz als „sehr gut“ oder „gut“ ein. In der Altersgruppe der

65- bis 69-Jährigen liegt dieser Anteil bei 76%. Angaben mit „eher schlecht“ und „sehr schlecht“ wurden aufgrund geringer Fallzahlen in der Abbildung zusammengefasst, „weiß ich nicht“ nicht berücksichtigt.

6.2 Kenntnisstand zu Suchmaschinenwerbung und -optimierung

Abbildung 5 zeigt die Kenntnisse deutscher InternetnutzerInnen zu Suchmaschinenwerbung (SEA; Fragen 2.1–2.4). Die offene Frage 2.1 nach der Haupteinnahmequelle Googles zeigt, dass 59,6% der befragten InternetnutzerInnen wissen, dass es sich dabei um das Anzeigengeschäft handelt. Ein wesentlich höherer Anteil (79,1%) hält die Einflussmöglichkeit auf die Sichtbarkeit einzelner Ergebnisse gegen Bezahlung an Google prinzipiell für möglich (Frage 2.2). Weniger als die Hälfte aller InternetnutzerInnen (42,3%) ist bewusst, dass sich bezahlte von organischen Ergebnissen unterscheiden (Frage 2.3) und ein gutes Viertel (27,7%) weiß, dass beide Ergebnistypen durch die Anzeigenkennzeichnung zu unterscheiden sind (Frage 2.4).

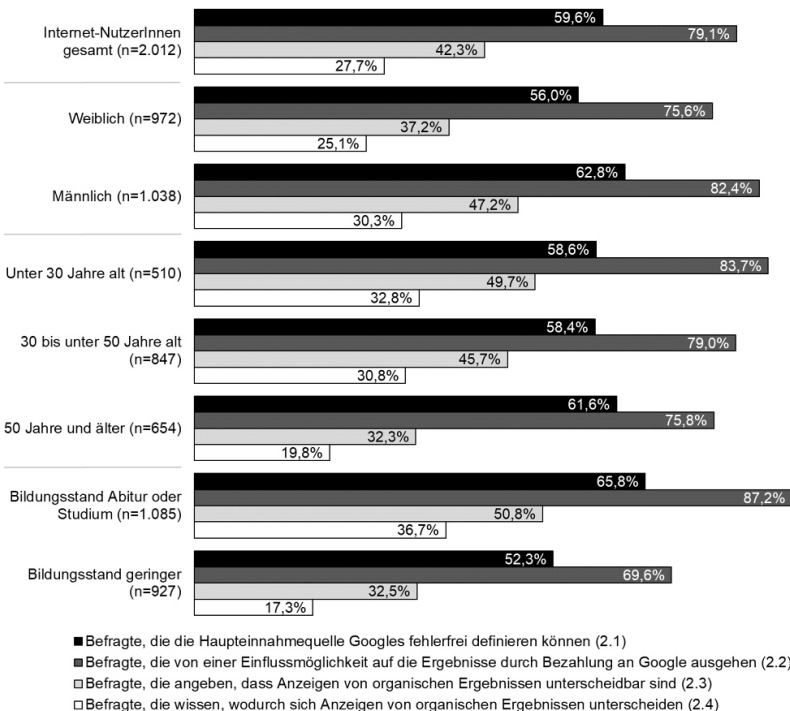


Abb. 5 Kenntnisse deutscher InternetnutzerInnen zu SEA

Darüber hinaus werden Unterschiede mit Blick auf die Altersgruppen und Bildungsstände deutlich. Jüngere InternetnutzerInnen sowie NutzerInnen mit Abitur oder Studienabschluss verfügen über alle Fragen hinweg über höhere SEA-Kenntnisse, mit Ausnahme der Frage zu Googles Haupteinnahmequelle (2.1). Hier sind keine wesentlichen Unterschiede zwischen den Altersgruppen erkennbar.

In Abbildung 6 werden die Kenntnisse deutscher InternetnutzerInnen zu SEO (Fragen 3.1–3.3) aufgeführt. Während die Einflussmöglichkeit auf die Suchergebnisse ohne Bezahlung an Google (Frage 3.1) noch 43,4% aller InternetnutzerInnen bekannt ist, wird damit lediglich von 8,1% der NutzerInnen der Begriff „Suchmaschinenoptimierung“ bzw. „SEO“ in Verbindung gebracht (3.2). SEO-Maßnahmen, wie beispielsweise Backlinks oder den gezielten Einsatz bestimmter Keywords, können 12,6% der InternetnutzerInnen nennen (3.3). Auch im Fragebogenabschnitt zu SEO sind die zuvor geschilderten Unterschiede hinsichtlich Alter und Bildungsgrad erkennbar.

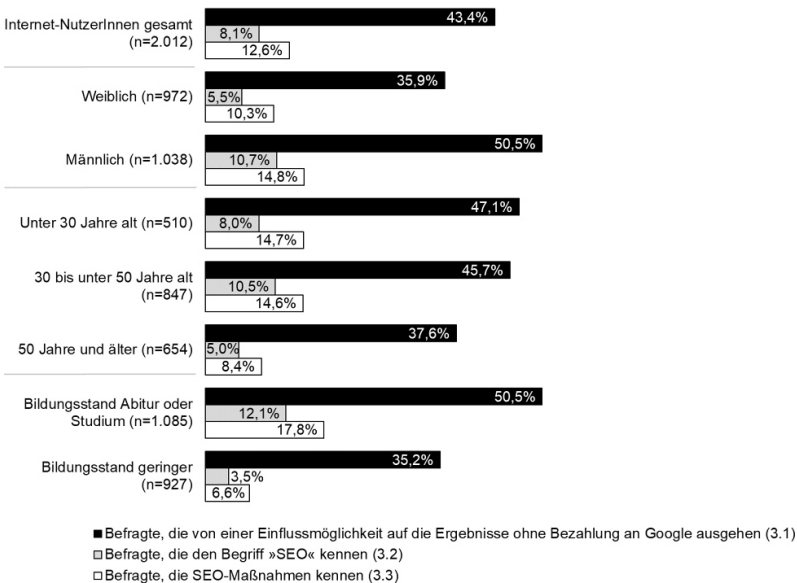


Abb. 6 Kenntnisse deutscher InternetnutzerInnen zu SEO

Im Vergleich von SEA und SEO zeigt sich somit, dass den deutschen InternetnutzerInnen SEA wesentlich bekannter ist als SEO.

6.3 Fähigkeit zur Identifikation organischer und bezahlter Suchergebnisse

In diesem Abschnitt stellen wir die Fähigkeit der ProbandInnen dar, Suchergebnisse, die durch SEO oder SEA beeinflussbar sind, korrekt zu identifizieren. Die Identifikationsquoten wurden wie in Abschnitt 0 beschrieben berechnet.

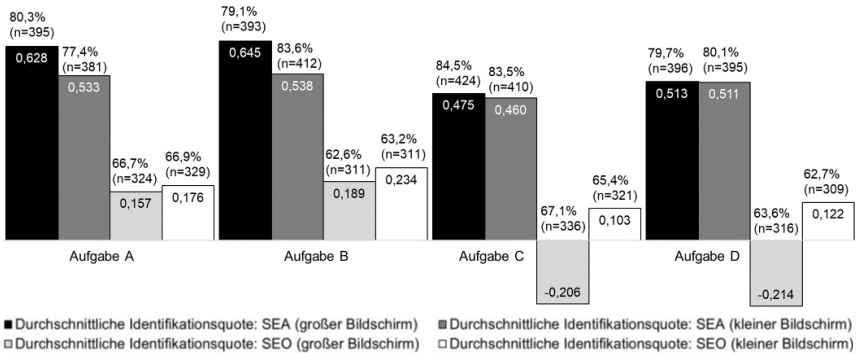


Abb. 7 Fähigkeit zur Identifizierung organischer und bezahlter Suchergebnisse

Die durchschnittliche Identifikationsquote je Aufgabe wird innerhalb der Balken dargestellt, z.B. 0,628 für Aufgabe A am großen Bildschirm und der Kennzeichnung der Anzeigen (SEA). Oberhalb der Balken ist die Anzahl der TeilnehmerInnen angegeben, auf die sich die jeweilige Identifikationsquote bezieht. Im Fall von Aufgabe A, großer Bildschirm, SEA, bedeutet dies, dass 80,3% der Personen, die diese Aufgabe zu bearbeiten hatten, auch tatsächlich Markierungen für Anzeigen vorgenommen haben, während die übrigen 19,7% fälschlicherweise angaben, dass es keine Anzeigen auf der SERP gab.

Beim Blick auf die SEA- und SEO-Identifikationsquoten wird deutlich, dass Anzeigen auf allen SERPs wesentlich zuverlässiger identifiziert worden sind als durch SEO zu beeinflussende (organische) Ergebnisse. Darüber hinaus wurden beide Ergebnistypen bei den einfach strukturierten SERPs (Aufgaben A und B) besser identifiziert als bei den komplexeren SERPs (Aufgaben C und D). Außerdem haben die ProbandInnen bei den SEA-Ergebnissen wesentlich häufiger Markierungen vorgenommen (bis zu 84,5%) als bei den SEO-Ergebnissen (bis zu 67,1%).

Beim Vergleich zwischen den Bildschirmgrößen fällt auf, dass die SEA-Ergebnisse bei den einfach strukturierten SERPs (Aufgaben A und B) auf

dem großen Bildschirm besser identifiziert wurden als auf dem kleinen Bildschirm. Auf dem großen Bildschirm zeigen die negativen Identifikationsquoten für die SEO-Ergebnisse auf den komplexen SERPs (Aufgaben C und D), dass im Durchschnitt mehr falsche als richtige Markierungen vorgenommen wurden. Bei den komplexen SERPs wurden die SEO-Ergebnisse auf dem kleinen Bildschirm zudem erfolgreicher identifiziert als auf dem großen Bildschirm.

Dabei ist zu berücksichtigen, dass in die Berechnung der Identifikationsquoten die falsch-positiven Markierungen (in diesem Fall die Anzeigenmarkierungen) mit einfließen. Auf dem großen Bildschirm der komplexen SERPs wurden acht Shoppinganzeigen dargestellt, auf dem kleinen Bildschirm nur zwei. Die Identifikationsquote sinkt also wesentlich stärker, wenn ein Befragter alle acht Shoppinganzeigen auf dem großen Bildschirm fälschlicherweise als SEO-Ergebnisse einstuft, als wenn der Proband die lediglich zwei Shoppinganzeigen auf dem kleinen Bildschirm fälschlicherweise als SEO-Ergebnisse markiert. Wir haben diese Unterschiede in der Gestaltung der Befragung beibehalten, da uns eine realistische Darstellung der SERPs wichtiger erschien als eine exakte Übereinstimmung der Anzeigenzahl zwischen beiden Bildschirmformaten.

6.4 Vergleich der selbst eingeschätzten mit der gemessenen Suchmaschinenkompetenz

In Tabelle 5 wird dargestellt, wie erfolgreich die Befragten mit unterschiedlicher Selbsteinschätzung die einzelnen Wissens- sowie Markierungsaufgaben bearbeitet haben. Im Falle der Wissensfragen werden die Anteile korrekter Antworten in Prozent, im Falle der Markierungsaufgaben die Identifikationsquoten (zur Berechnung s. Abschn. 5.2) in Mittelwerten angegeben.

Zunächst wird deutlich, dass InternetnutzerInnen mit hoher Selbsteinschätzung („sehr gut“ und „gut“) die einzelnen Wissensfragen zu SEA und SEO häufiger korrekt beantwortet haben als auch die SEA-Ergebnisse zuverlässiger identifizieren konnten als NutzerInnen mit niedrigerer Selbsteinschätzung. Dennoch bewegt sich auch ihr Kenntnisstand auf einem niedrigen Niveau. So ist die Einflussmöglichkeit außerhalb der Bezahlung an Google nur weniger als der Hälfte (48%), der SEO-Begriff 12% und SEO-Maßnahmen 19% der Befragten trotz Selbsteinschätzung „sehr gut“ bekannt (Fragen 3.1–3.3). In der Markierung der SEO-Ergebnisse erreichten die ProbandInnen mit hoher Selbsteinschätzung zudem keine höheren Identifikationsquoten als

ProbandInnen mit niedrigerer Selbsteinschätzung. Somit lässt sich zum einen festhalten, dass InternetnutzerInnen mit hoher Selbsteinschätzung häufiger die Wissensfragen zu SEA und SEO korrekt beantworten und stärker darin sind, Anzeigen zu identifizieren. Zum anderen verfügen jedoch auch diese ProbandInnen über erhebliche Defizite hinsichtlich der SEO-Wissensfragen sowie über dieselben Schwächen in der Identifizierung organischer Ergebnisse wie Personen mit niedrigerer Selbsteinschätzung.

Tab. 5: Anteile korrekter Antworten sowie Identifikationsquoten nach Selbsteinschätzung

	<i>n</i> =	Korrekte Antworten auf Wissensfragen (in Prozent)							Identifikationsquoten (Mittelwerte)			
		Wissensfragen SEA			Wissensfragen SEO				Identifikation SEA		Identifikation SEO	
Selbsteinschätzung		2.1	2.2	2.3	2.4	3.1	3.2	3.3	ein-fach*	kom-plex**	ein-fach*	kom-plex**
sehr gut	642	63,2	81,5	49,2	33,5	48,3	11,7	18,8	0,60	0,50	0,20	0,04
gut	1.055	62,4	81,9	42,6	30,5	43,4	8,2	13,6	0,61	0,50	0,19	0,04
befriedigend	281	44,5	66,9	29,2	17,1	33,1	6,4	10,3	0,44	0,38	0,20	0,03
eher schlecht	13	38,5	76,9	23,1	7,7	38,5	7,7	0,0	0,42	0,52	0,26	0,08
sehr schlecht	2	0,0	50,0	0,0	0,0	50,0	0,0	0,0	0,50	0,50	-0,20	-0,11

* einfach strukturierte SERPs: Aufgaben A und B
 ** komplex strukturierte SERPs: Aufgaben C und D

Die weitreichende Unkenntnis über alle Selbsteinschätzungsgruppen hinweg wird auch in einem weiteren Vergleich deutlich, welcher die Fragestellungen zu den Themenkomplexen SEA und SEO jeweils gruppiert betrachtet.¹⁰ Für den Themenkomplex SEA wurden die ProbandInnen hinsichtlich ihrer Selbsteinschätzung (zeilenweise)¹¹ und der Beantwortung des Fragenkomplexes zu SEA (spaltenweise) aufgeteilt. In den Spalten finden sich auf der linken Seite diejenigen ProbandInnen wieder, die sowohl den Großteil (mindestens drei von vier) der Wissensfragen zu SEA korrekt beantwortet haben als auch den Großteil der Anzeigen korrekt identifiziert haben (durch-

10 In der Stichprobe finden sich lediglich zwei ProbandInnen, die sämtliche Wissens- als auch Markieraufgaben beider Themenkomplexe (SEO, SEA) korrekt gelöst haben.

11 Aufgrund geringer Fallzahlen wurden die TeilnehmerInnen mit den Selbsteinschätzungen „befriedigend“, „eher schlecht“ sowie „sehr schlecht“ zusammen betrachtet.

schnittliche Identifikationsquote für SEA > 0,5). In der Spalte rechts daneben finden sich somit jene ProbandInnen wieder, denen dies nicht gelungen ist.

Tab. 6: Themenkomplex SEA nach Selbsteinschätzung

Selbsteinschätzung	Großteil der Wissensfragen zu SEA korrekt* <u>sowie</u> Großteil der SEA-Markierungen** erfolgreich	Großteil der Wissensfragen zu SEA inkorrekt und/oder Großteil der SEA-Markierungen nicht erfolgreich
sehr gut	32%	68%
gut	27%	73%
befriedigend oder schlechter	15%	85%

* mind. drei von vier Wissensfragen zu SEA korrekt

** SEA-Identifikationsquoten von durchschnittlich > 0,5

Wie Tabelle 6 zeigt, haben InternetnutzerInnen, die ihre Suchmaschinenkompetenz als „sehr gut“ einordnen, zu 32% den Großteil der Wissens- und Markierungsaufgaben zu SEA erfolgreich absolviert. Die weiteren InternetnutzerInnen mit sehr hoher Selbsteinschätzung (68%) weisen somit mindestens in der Bearbeitung der Wissens- oder der Markierungsaufgaben Mängel auf. ProbandInnen mit Selbsteinschätzung „befriedigend“ oder schlechter absolvierten die SEA-Fragestellungen lediglich zu 15% erfolgreich. Zwischen der Selbsteinschätzung und der Beantwortung der SEA-Fragestellungen besteht ein signifikanter Zusammenhang $\chi^2(2) = 27.952, p < .001$.

Für den Themenkomplex zu SEO wurde identisch vorgegangen. Bei den Wissensfragen bezieht sich der Großteil nun auf mindestens zwei von drei Wissensfragen, bei den Markierungen beziehen sich die Identifikationsquoten nun auf die SEO-Ergebnisse (durchschnittliche Identifikationsquote für SEO > 0,5).

Tab. 7: Themenkomplex SEO nach Selbsteinschätzung

Selbsteinschätzung	Großteil der Wissensfragen zu SEO korrekt* <u>sowie</u> Großteil der SEO-Markierungen** erfolgreich	Großteil der Wissensfragen zu SEO inkorrekt und/oder Großteil der SEO-Markierungen nicht erfolgreich
sehr gut	8%	92%
gut	6%	94%
befriedigend oder schlechter	5%	95%

* mind. zwei von drei Wissensfragen zu SEO korrekt

** SEO-Identifikationsquoten von durchschnittlich > 0,5

Beim SEO-Themenkomplex wird deutlich, dass nur ein Bruchteil (8%) der InternetnutzerInnen mit Selbsteinschätzung „sehr gut“ die zugehörigen Wissens- und Markieraufgaben erfolgreich bearbeiten konnte. Zwischen der Selbsteinschätzung und der Beantwortung der SEO-Fragestellungen besteht kein signifikanter Zusammenhang $\chi^2(2) = 4.885, p = .087$. Anschließend an die zuvor gezeigten Ergebnisse, die Vorteile in der Beantwortung einzelner SEO-Wissensfragen für die NutzerInnen mit hoher Selbsteinschätzung zeigten, bedeutet dies, dass über alle SEO-Fragestellungen hinweg (Wissens- und Markieraufgaben) die NutzerInnen über eine vergleichbare Unkenntnis bezüglich SEO verfügen, unabhängig davon, wie sie ihre eigene Suchmaschinenkompetenz einschätzen.

7 Diskussion und Fazit

Die repräsentative Online-Befragung zeigt, dass deutsche InternetnutzerInnen größtenteils eine hohe Einschätzung ihrer eigenen Suchmaschinenkompetenz haben. Anhand von Frage- bzw. Aufgabenstellungen zu organischen und bezahlten Suchergebnissen konnte jedoch gezeigt werden, dass die hohe Selbsteinschätzung nicht der tatsächlichen Suchmaschinenkompetenz entspricht. Dies gilt im Besonderen für den Bereich der Suchmaschinenoptimierung, der nur einem Bruchteil deutscher InternetnutzerInnen bekannt ist und mit den organischen Ergebnissen in Verbindung gebracht wird. Die Ergebnisse legen somit nahe, dass deutsche InternetnutzerInnen nicht hinreichend verstehen, wie die Ergebnisse auf Suchergebnisseiten kommerzieller Suchmaschinen entstehen und in welcher Weise sie beeinflusst werden können. Verglichen mit der Studie von Lewandowski et al. (2018), die als Vorlagenstudie für die SEA-bezogenen Fragestellungen diente, wissen mehr NutzerInnen, dass es eine Möglichkeit der bezahlten Platzierung gibt (79% vs. 73%), genauso viele NutzerInnen, dass sich organische Ergebnisse von Anzeigen unterscheiden (43%), und weniger NutzerInnen, wodurch Google seine Einnahmen generiert (68% vs. 81%) und *wodurch* sich Anzeigen von organischen Ergebnissen unterscheiden (28% vs. 34%).

Der Vergleich der Selbsteinschätzung mit der tatsächlich vorhandenen Suchmaschinenkompetenz zeigt, dass InternetnutzerInnen mit hoher Selbsteinschätzung („sehr gut“ oder „gut“) über ausgeprägtere Kenntnisse im Bereich SEA verfügen als InternetnutzerInnen mit niedrigerer Selbsteinschät-

zung. Dies gilt für die Wissens- sowie Markieraufgaben. Auch im SEO-Bereich absolvieren die NutzerInnen mit hoher Selbsteinschätzung einzelne Wissensfragen erfolgreicher, weisen insgesamt betrachtet (Wissens- und Markieraufgaben) jedoch vergleichbare Defizite wie NutzerInnen mit niedrigerer Selbsteinschätzung auf. Zudem bewegen sich die Kenntnisse hinsichtlich SEO und SEA auch bei den NutzerInnen mit hoher Selbsteinschätzung auf niedrigem Niveau. Es kann somit von einer Überschätzung der eigenen Suchmaschinenkompetenz gesprochen werden. An dieser Stelle ist auch die generelle Problematik bei Selbsteinschätzungsfragen zu nennen. Bei diesen liegt ein hohes Maß an Subjektivität vor, sodass sich für die Erhebung tatsächlich vorhandener Leistungen bzw. Kenntnisse Test-Items wie Wissens- oder Markieraufgaben anbieten (Döring/Bortz, 2016, S. 408; Womser-Hacker/Mandl, 2012).

Die Überschätzung – also die Fehlannahme, über eine hohe Suchmaschinenkompetenz zu verfügen – ist als problematisch zu erachten. Sie hat das Potenzial, in einem unreflektierten Umgang mit den Suchergebnissen kommerzieller Suchmaschinen zu münden. Da den NutzerInnen die Einflüsse durch Akteursverbände aus SEO und SEA nicht hinreichend bekannt sind, fehlt die Basis für eine kritische Auseinandersetzung mit den erheblichen (finanziellen) Eigeninteressen der Suchmaschinenbetreiber und den Einflüssen durch externe Akteure. Diese Problematik wird zum einen dadurch verstärkt, dass SuchmaschinennutzerInnen mit hoher Selbsteinschätzung den Großteil der Internetnutzerschaft stellen. Zum anderen erfolgt die Verstärkung durch die Verhaltensweisen der SuchmaschinennutzerInnen, den Suchmaschinen und insbesondere Google stark zu vertrauen und bei der Recherche nur wenige, prominent platzierte Suchergebnisse in ihrer Auswahl zu berücksichtigen.

Eine Förderung der Suchmaschinenkompetenz ist somit entscheidend dafür, den über Suchmaschinen erfolgenden Wissenserwerb der InternetnutzerInnen auf eine fundierte Basis zu stellen. Im Kontext der Informationskompetenz wird Suchmaschinenkompetenz allerdings kaum berücksichtigt (Informationskompetenz.de, 2018). Auf der Basis unserer Ergebnisse argumentieren wir, dass die Schulung von Informationskompetenz gerade dort ansetzen muss, wo NutzerInnen in ihrem Alltag suchen, nämlich bei den kommerziellen Suchmaschinen. Wenn sie anhand dieser Systeme auch ihre Kompetenzdefizite erkennen, dürften sie auch offener für weitergehende Schulungen, beispielsweise zur Auswahl und Verwendung spezieller Datenbanken, sein.

Einschränkungen der Studie sind im Wesentlichen in der Fragebogenkonstruktion zu vermuten. So ist nicht vollends sicherzustellen, dass sämtliche Befragte die Erläuterungstexte zu SEO und SEA gelesen und verstanden haben. Bezüglich der Markierungsaufgaben ist anzumerken, dass zum einen aufgrund des großen Gesamtumfangs des Fragebogens lediglich zwei Suchergebnisseiten getestet werden konnten. Zum anderen stellen die vorgefertigten Suchergebnisseiten keine natürliche Nutzungssituation dar und könnten die Identifikation der jeweils zu markierenden Ergebnisse erschwert haben. Eine Stärke der Studie stellt die repräsentative Stichprobe dar, die Aussagen über die deutsche Online-Bevölkerung nach den Kriterien der AGOF zulässt. Eine regelmäßige Wiederholung der repräsentativen Befragung wäre fruchtbar, um die Entwicklung der Suchmaschinenkompetenz der Internetnutzerschaft nachvollziehen und so gezielte Fördermaßnahmen entwickeln zu können.

Forschungsdaten

Die Daten der Online-Befragung können über die OSF-Plattform (dx.doi.org/10.17605/OSF.IO/JYV9R) abgerufen werden.

Förderung

Das Projekt „SEO-Effekt“, aus dem diese Studie hervorgeht, wird von der Deutschen Forschungsgemeinschaft (DFG) unter der Projektnummer 417552432 gefördert.

Literatur

- Alphabet Inc. (2020): Alphabet Announces Fourth Quarter and Fiscal Year 2019 Results. https://abc.xyz/investor/static/pdf/2019Q4_alphabet_earnings_release.pdf
- American Library Association (1989): Presidential Committee on Information Literacy: Final Report. <http://www.ala.org/acrl/publications/whitepapers/presidential>
- Aufenanger, S.; Siebertz, T. (2014): Informationskompetenz als notwendige Voraussetzung zur Nutzung von Suchmaschinen: Eine empirische Studie zu Such-

- strategien bei komplexen Aufgaben. In B. Stark, D. Doerr, S. Aufenanger (Hrsg.): *Die Googleisierung der Informationssuche*. de Gruyter, S. 160–180. <https://doi.org/10.1515/9783110338218.160>
- Beisch, N.; Schäfer, C. (2020): Internetnutzung mit großer Dynamik: Medien, Kommunikation, Social Media. *Media Perspektiven*, 9, 462–481. https://www.ard-zdf-onlinestudie.de/files/2020/0920_Beisch_Schaefer.pdf
- Cullen, R.; Clark, M.; Esson, R. (2011): Evidence-based information-seeking skills of junior doctors entering the workforce: an evaluation of the impact of information literacy training during pre-clinical years. *Health Information & Libraries Journal* 28 (2), 119–129. <https://doi.org/10.1111/j.1471-1842.2011.00933.x>
- Cutrell, E.; Guan, Z. (2007): What are you looking for? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '07*, S. 407–416. <https://doi.org/10.1145/1240624.1240690>
- Döring, N.; Bortz, J. (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Aufl., Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Douglas, K.; Wertz, R.; Fosmire, M.; Purzer, S.; Van Epps, A. (2014): First-Year and Junior Engineering Students' Self-Assessment of Information Literacy Skills. *2014 ASEE Annual Conference & Exposition Proceedings*, S. 24.607.1–24.607.10. <https://doi.org/10.18260/1-2--20498>
- Dragovic, N.; Madrazo Azpiazu, I.; Pera, M. S. (2016): “Is Sven Seven?”. A Search Intent Module for Children. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval – SIGIR '16*, 1, S. 885–888. <https://doi.org/10.1145/2911451.2914738>
- European Commission (2017): Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service—Factsheet. http://g8fiprpkplyr33r3krz5b97d1.wp-engine.netdna-cdn.com/wp-content/uploads/2017/06/Google-MEMO-17-1785_EN.pdf
- Google.com (2020): Werbekosten bei Google: Pay-per-Click – Google Ads. https://ads.google.com/intl/nl_de/home/pricing/
- Granka, L. A.; Joachims, T.; Gay, G. (2004): Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval – SIGIR '04*, S. 478–479. <https://doi.org/10.1145/1008992.1009079>
- Griesbaum, J. (2013): Online-Marketing. In: R. Kuhlen, W. Semar, D. Strauch (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. 6. Aufl., Berlin, Boston, MA: de Gruyter Saur, S. 411–423. <https://doi.org/10.1515/9783110258264.411>

- Informationskompetenz.de (2018): Informationskompetenz – Bundesstatistik 2017. <http://www.informationskompetenz.de/index.php/veranstaltungsstatistik/>
- Kammerer, Y.; Gerjets, P. (2014): The Role of Search Result Position and Source Trustworthiness in the Selection of Web Search Results When Using a List or a Grid Interface. *International Journal of Human-Computer Interaction* 30 (3), 177–191. <https://doi.org/10.1080/10447318.2013.846790>
- Keane, M. T.; O'Brien, M.; Smyth, B. (2008): Are people biased in their use of search engines? *Communications of the ACM* 51 (2), 49–52. <https://doi.org/10.1145/1314215.1314224>
- Lau, J. (2008): Richtlinien zur Informationskompetenz für Lebenslanges Lernen. <https://www.ifla.org/files/assets/information-literacy/publications/ifla-guidelines-de.pdf>
- Lewandowski, D. (2016): Suchmaschinenkompetenz als Baustein der Informationskompetenz. In: Wilfried Sühl-Strohmeier (Hrsg.): *Handbuch Informationskompetenz*. Berlin: de Gruyter Saur, S. 115–126. <https://doi.org/10.1515/9783110403367-013>
- Lewandowski, D. (2017): Users' Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice* 5 (4), 6–25. <https://doi.org/10.1633/JISTaP.2017.5.4.1>
- Lewandowski, D. (2018). *Suchmaschinen verstehen*. 2. Aufl., Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-56411-0>
- Lewandowski, D.; Kerkmann, F.; Rümmele, S.; Sünkler, S. (2018): An empirical investigation on search engine ad disclosure. *Journal of the Association for Information Science and Technology* 69 (3), 420–437. <https://doi.org/10.1002/asi.23963>
- Liu, Z.; Liu, Y.; Zhou, K.; Zhang, M.; Ma, S. (2015): Influence of Vertical Result in Web Search Examination. In: R. A. Baeza-Yates, M. Lalmas, A. Moffat, B. A. Ribeiro-Neto (Hrsg.): *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval – SIGIR '15* (S. 193–202). New York, NY: ACM Press. <https://doi.org/10.1145/2766462.2767714>
- Machill, M. (2009). 12 goldene Suchmaschinen-Regeln. <https://www.medienanstalt-nrw.de/fileadmin/lfm-nrw/Medienkompetenz/ratgeber-suchmaschinen-farbe.pdf>
- Machill, M.; Beiler, M.; Gerstner, J. R. (2012): Der Info-Kompass: Orientierung für den kompetenten Umgang mit Informationen. https://www.klicksafe.de/fileadmin/media/documents/pdf/Broschren_Ratgeber/Info_Kompass.pdf
- Mahmood, K. (2016): Do People Overestimate Their Information Literacy Skills? A Systematic Review of Empirical Evidence on the Dunning-Kruger Effect. *Communications in Information Literacy* 10 (2), 199–213. <https://doi.org/10.15760/comminfolit.2016.10.2.24>

- McCue, T. (2018): SEO Industry Approaching \$80 Billion But All You Want Is More Web Traffic. *Forbes*. <https://www.forbes.com/sites/tjmccue/2018/07/30/seo-industry-approaching-80-billion-but-all-you-want-is-more-web-traffic/>
- Murphy, J.; Hofacker, C.; Mizerski, R. (2006): Primacy and Recency Effects on Clicking Behavior. *Journal of Computer-Mediated Communication* 11 (2), 522 bis 535. <https://doi.org/10.1111/j.1083-6101.2006.00025.x>
- Ngo, H. T.; Pickard, A. J.; Walton, G. (2019): Information literacy capabilities of upper secondary students: the case of Vietnam. *Global Knowledge, Memory and Communication* 68 (6/7), 453–470. <https://doi.org/10.1108/GKMC-03-2019-0037>
- Pan, B.; Hembrooke, H.; Joachims, T.; Lorigo, L.; Gay, G.; Granka, L. (2007): In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication* 12 (3), 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>
- Petrescu, P. (2014): Google organic click-through rates in 2014 [Weblog post]. <https://moz.com/blog/google-organic-click-through-rates-in-2014>
- Purcell, K.; Brenner, J.; Rainie, L. (2012): Search Engine Use 2012. Pew Internet and American Life Project. https://www.eff.org/files/pew_2012_o.pdf
- Röhle, T. (2010): *Der Google-Komplex. Über Macht im Zeitalter des Internets*. Bielefeld: transcript. <https://doi.org/10.14361/transcript.9783839414781>
- Rust, K.; Schlögl, C.; DongBack, S. (2017): Comparing Information Literacy of Students from University of Graz (Austria) and Chungbuk National University (Republic of Korea). In: M. Gäde, V. Trkulja, V. Petras (Hrsg.): *Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin* (S. 24–36). Glückstadt: Verlag Werner Hülsbusch. <https://dx.doi.org/10.18452/1454>
- Sachse, J. (2019): The influence of snippet length on user behavior in mobile web search. *Aslib Journal of Information Management* 71 (3), 325–343. <https://doi.org/10.1108/AJIM-07-2018-0182>
- Schultheiß, S.; Lewandowski, D. (2020): “Outside the industry, nobody knows what we do”. SEO as seen by search engine optimizers and content providers. *Journal of Documentation* (Preprint). <https://doi.org/10.1108/JD-07-2020-0127>
- Schultheiß, S.; Sünkler, S.; Lewandowski, D. (2018): We still trust in Google, but less than 10 years ago: an eye-tracking study. *Information Research* 23 (3), paper 799. <http://www.informationr.net/ir/23-3/paper799.html>
- Schweiger, W. (2003): Suchmaschinen aus Nutzersicht. In: M. Machill, C. Welp (Hrsg.): *Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen*. Gütersloh: Verlag Bertelsmann Stiftung, S. 133–208. <https://sfb.tu-dortmund.de/handle/sfbs/141>

- Singer, G.; Norbistrath, U.; Lewandowski, D. (2012): Ordinary search engine users assessing difficulty, effort, and outcome for simple and complex search tasks. In: *Proceedings of the 4th Information Interaction in Context Symposium on – IIIX '12*. New York, NY: ACM Press, S. 110–119. <https://doi.org/10.1145/2362724.2362746>
- Stark, B.; Magin, M.; Jürgens, P. (2014): Navigieren im Netz: Befunde einer qualitativen und quantitativen Nutzerbefragung. In: B. Stark, D. Dörr, S. Aufenanger (Hrsg.): *Die Googleisierung der Informationssuche*. Berlin, Boston, MA.: de Gruyter, S. 20–74. <https://doi.org/10.1515/9783110338218.20>
- Sterling, G. (2016): Report: Nearly 60 percent of searches now from mobile devices. <https://searchengineland.com/report-nearly-60-percent-searches-now-mobile-devices-255025>
- Sundin, O. (2020): Where is Search in Information Literacy? A Theoretical Note on Infrastructure and Community of Practice. In: A. Sundqvist, G. Berget, J. Nolin (Hrsg.): *Sustainable Digital Communities. 15th International Conference, iConference 2020*. Cham: Springer Nature, S. 373–379. https://doi.org/10.1007/978-3-030-43687-2_29
- Sundin, O.; Haider, J.; Andersson, C.; Carlsson, H.; Kjellberg, S. (2017): The searchification of everyday life and the mundane-ification of search. *Journal of Documentation* 73 (2), 224–243. <https://doi.org/10.1108/JD-06-2016-0081>
- Tremel, A. (2010): *Suchen, finden – glauben? Die Rolle der Glaubwürdigkeit von Suchergebnissen bei der Nutzung von Suchmaschinen*. Diss., LMU München. https://edoc.ub.uni-muenchen.de/12418/1/Tremel_Andreas.pdf
- White, R. W.; Horvitz, E. (2014): Belief Dynamics and Biases in Web Search. *ACM Transactions on Information Systems* 33 (4). <https://doi.org/10.1002/asi.23128>
- Womser-Hacker, C.; Mandl, T. (2012): Information Seeking Behaviour (ISB). In: R. Kuhlen, W. Semar, D. Strauch (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. 4. Aufl., Berlin: de Gruyter Saur, S. 97–108. <https://doi.org/10.1515/9783110258264.97>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 218–246. DOI: doi.org/10.5283/epub.44946.

Kollaboratives Informationssuchverhalten von Personen unterschiedlicher Generationszugehörigkeit

Carolin Schulz, Stefanie Elbeshausen, Christa Womser-Hacker

Stiftung Universität Hildesheim, Deutschland
{[schulzca](mailto:schulzca@uni-hildesheim.de), [elbesh](mailto:elbesh@uni-hildesheim.de), [womser](mailto:womser@uni-hildesheim.de)}@uni-hildesheim.de

Zusammenfassung

Im vorliegenden Paper wird eine Studie dargestellt, in der das kollaborative Informationssucherverhalten (*Collaborative Information Seeking*, kurz: CIS) zwischen Generationen untersucht wurde. In Teams von zwei Personen sollten die Suchgruppen eine Reise planen und dafür im Internet nach Reiseinformation recherchieren. Anhand der Forschungsfrage: „*Wie wird die kollaborative Informationssuche zwischen unterschiedlichen Generationen durchgeführt?*“ fand eine qualitative Studie statt, in welcher die Organisation und das generelle Vorgehen bei der Suche beobachtet wurden. Im Fokus standen dabei das Rollenverhalten der Beteiligten und die Frage danach, ob eine Person eine leitende Position während der Suche einnimmt. Die Suche und Zusammenarbeit wurden zu diesem Zweck mittels Videos aufgenommen und nachfolgend analysiert. Im Anschluss partizipierten die Testpersonen in leitfadengestützten Gruppeninterviews. Die Ergebnisse deuten in allen Teams auf einen ähnlichen Ablauf der gemeinsamen Suche hin. Eine Leitungsfunktion war nur bedingt zu beobachten, wobei die Testpersonen jedoch spezifische Rollen während der Suche eingenommen haben.

Keywords: Collaborative Information Seeking; Generationen; qualitative Forschung; videobasierte Inhaltsanalyse

1 Einleitung

Die Nutzung des Internets ist mittlerweile in allen Generationen angekommen. Die ARD/ZDF-Onlinestudie verzeichnet für 2020, dass 94% der

deutschsprachigen Bevölkerung ab 14 Jahren das Internet täglich nutzen. In der Bevölkerungsgruppe von 50 bis 69 Jahren beträgt die durchschnittliche Nutzungsdauer 68 Minuten täglich, in der Gruppe ab 70 Jahren sind es noch 23 Minuten am Tag (vgl. ARD/ZDF-Onlinestudie 2020¹). Gerade durch die aktuelle Corona-Pandemie ist das Internet auch für die älteren Generationen immer wichtiger geworden – sei es, um Kontakt zur Familie halten zu können, gefahrlos einzukaufen oder Unterhaltungsmedien zu nutzen.

Neben der Nutzung in der Freizeit ist es auch im Arbeitskontext in vielen Branchen unabdingbar geworden, online zu sein. In der Wissenschaft ist die Nutzung von im Netz verfügbaren Literaturquellen alltäglich, die Kommunikation und Zusammenarbeit mit örtlich verteilten Projektpartnern verläuft meist online. Im Einzelhandel werden Bestellungen über das Internet durchgeführt, Kliniken und Ärzte nutzen es, um Krankheitsbilder zu recherchieren, Experten zu kontaktieren oder Bestellungen durchzuführen. Dies sind nur wenige Beispiele, die andeuten, wie relevant die Online-Vernetzung in unterschiedlichsten Branchen mittlerweile ist. Dem Arbeits- sowie dem Freizeitkontext ist darüber hinaus gemein, dass Menschen verschiedener Generationen zusammenarbeiten, auch wenn in vielen Organisationen ein baldiger Generationenwechsel bevorsteht. Bei dieser Zusammenarbeit werden selbstverständlich auch Computer und Internet genutzt. Im Kontext von komplexeren Projekten – bspw. in der Wissenschaft oder auch in der Freizeit, etwa bei der Reiseplanung – führen die älteren und jüngeren Generationen auch gemeinsame Informationssuchen durch. Im Zusammenspiel der immer stärker anwachsenden Internetnutzung, gerade auch bei den älteren Generationen, dem anstehenden Generationenwechsel in vielen Unternehmen und Organisationen sowie der oftmals notwendigen Online-Kommunikation und Zusammenarbeit stellt sich die Frage, wie sich die Arbeit im Team zwischen Menschen verschiedener Generationszugehörigkeit gestaltet. Durch ein besseres Verständnis der Abläufe und des Vorgehens können entsprechende Unterstützungsmaßnahmen bereitgestellt werden, um die genannten Abläufe so effizient wie möglich auszuführen.

In der vorliegenden Untersuchung wird ein Aspekt der Kollaboration – die gemeinsame Informationssuche in der Freizeit – herausgegriffen, um diesen näher zu beleuchten. Anhand der Fragestellung „Wie wird die kollaborative Informationssuche zwischen unterschiedlichen Generationen durch-

1 ARD/ZDF-Onlinestudie 2020. Online verfügbar unter: <https://www.ard-zdf-online-studie.de/>

geführt?“ wurde eine qualitative Studie mit Zwei-Personen-Teams vollzogen, welche hinsichtlich ihrer Generationszugehörigkeit heterogen zusammengesetzt sind. Neben der generellen Gestaltung der gemeinsamen Suche standen insbesondere mögliche Rollenverteilungen sowie die Übernahme der Leitung während der Zusammenarbeit im Fokus. Nachfolgend werden zunächst die theoretischen Grundlagen und angrenzende Studien dargestellt (Abschn. 2). Der Fokus liegt hierfür zum einen auf der kollaborativen Informationssuche (*Collaborative Information Seeking*, kurz: CIS), zum anderen auf dem Generationenbegriff und Generationenmodellen. Anschließend erfolgt eine Darstellung des methodischen Vorgehens (Abschn. 3) sowie der Untersuchungsdurchführung (Abschn. 4). Im Abschnitt 5 wird die Vorgehensweise der Auswertung erläutert und anschließend werden die Ergebnisse präsentiert (Abschn. 6). Den Abschluss bilden Fazit und Ausblick (Abschn. 7), wobei neben Anschlussmöglichkeiten für nachfolgende Studien auch Implikationen für die Praxis gezo-gen werden.

2 Theoretische Grundlagen

Die vorliegende Untersuchung basiert vornehmlich auf zwei zentralen Themenbereichen: Collaborative Information Seeking und dem Generationenbegriff. Die nachfolgenden Ausführungen geben eine knappe Übersicht zu beiden Bereichen, wobei der Schwerpunkt auf angrenzenden Studien liegt.

2.1 Collaborative Information Seeking

CIS umfasst, im Gegensatz zum singulären Information Seeking, die gemeinsame Suche nach Information. Poltrock et al. (2003, S. 239) definieren CIS als “[...] activities that a group or team of people undertakes to identify and resolve a shared information need”. CIS ist insbesondere dann gewinnbringend, wenn Suchaufgaben zu komplex sind, um von einer einzelnen Person bearbeitet zu werden (vgl. Shah, 2010). Es kann Shah (2014) zufolge als ein vorsätzlicher, interaktiver Informationssuchprozess verstanden werden, der im Kontext eines kollaborativen Projekts in einer kleinen Gruppe stattfindet. Um diesen besonders gewinnbringend zu gestalten, sollten die Mit-

gliedert unterschiedliches Wissen und unterschiedliche Fähigkeiten mitbringen (ebd., S. 219).

Kollaboration wird als die engste Form der Zusammenarbeit verstanden (vgl. Denning/Yaholkovsky, 2008, S. 20). Begriffe wie Kooperation oder Kontribution stehen zwar in engem Zusammenhang, stellen allerdings abgeschwächte Formen der Kollaboration dar.

CIS wurde bereits in diversen Kontexten und Domänen untersucht. Die gemeinsame Suche im wissenschaftlichen Kontext erforschten bspw. Hyldegård (2006), Haseki, Shah und Gonzales-Ibanez (2012), Vick, Nagano und Popadiuk (2015), Leeder und Shah (2016) oder Elbeshausen (2019). Im medizinischen Bereich erforschten Reddy und Jansen (2008) das CIS von Krankenpflegeteams, Reddy und Spence (2008) fokussierten sich auf CIS in der Notaufnahme. Sonnenwald (1995) und Sonnenwald und Pierce (2000) setzten sich mit kollaborativem Suchverhalten im militärischen Führungsbereich (Command & Control) auseinander. Hansen und Järvelin (2005) untersuchten CIS im Patentbereich. Für in der Freizeit ausgeführte gemeinsame Suchen findet sich bisher wenig Forschung. Elbeshausen et al. (2015) widmeten sich in ihrer Studie dem mobilen CIS in der Freizeit und stellten einen Vergleich zu Work-/Task-basiertem, kollaborativem Suchverhalten auf.

Generell dürfen die Vor- und Nachteile der kollaborativ ausgeführten Suche nicht außer Acht gelassen werden. Für Menschen als soziale Wesen ist es natürlich, kollaborativ zusammenzuarbeiten. Müssen komplexe Aufgaben bearbeitet werden, kann das Resultat bei CIS besser sein als beim individuellen Information Seeking (IS), da Beteiligte mit unterschiedlicher Expertise zusammenkommen und gemeinsam eine Lösung generieren (vgl. Shah 2015, S. 3 ff.). Gerade bei längerfristigen Suchen im Kontext einer komplexen Aufgabe kann es schnell zu einem Information Overload kommen. Um die große Menge an Information effizient bewältigen zu können, ist es hilfreich, wenn mehrere Personen zusammenarbeiten (vgl. Reddy et al. 2009). Insbesondere die Expertise der Mitglieder ist ein wichtiger Faktor, der für eine Zusammenarbeit spricht, denn von dem Wissen und den Fähigkeiten der einzelnen Mitglieder kann die gesamte Gruppe profitieren (vgl. Shah 2015, S. 7).

Nachteile liegen insbesondere im hohen Aufwand, den eine kollaborativ ausgeführte Zusammenarbeit mit sich bringt. Problematisch ist diese Form dann, wenn der Aufwand, auch hinsichtlich der Koordination, zu hoch ist, um dem Nutzen gerecht zu werden. Auch Interessenskonflikte der Gruppenmitglieder, Vertrauensprobleme oder unausgeglichene Machtverhältnisse können zu Problemen führen (vgl. ebd., S. 4). Zu bedenken ist auch, dass eine

kollaborative Zusammenarbeit einen zeitintensiven Prozess darstellt und somit für Aufgaben, die schnell gelöst oder entschieden werden können, nicht sinnvoll ist. Weiterhin zeigt sich, dass Kollaboration und entsprechend auch CIS am besten in kleinen Gruppen funktioniert. Ist ein Team zu groß, kann ein Projekt schnell scheitern (vgl. London, 1995).

2.2 Rollen bei der kollaborativen Suche

Rollen und ihre Auswirkungen auf das Suchverhalten wurden bereits im Kontext des individuellen IS untersucht. So setzte sich bspw. Heinström (2002) mit dem Einfluss der Persönlichkeit und daraus resultierenden Rollen bei der Informationssuche auseinander. Als Ergebnis ihrer Forschungsarbeit identifizierte sie drei Rollenmuster: *Fast Surfers*, *Broad Scanners* und *Deep Divers*, die ihre Suchen rollenspezifisch divers ausführen.

Auch im Zusammenhang mit CIS finden sich Forschungsprojekte, welche sich mit der Übernahme von Rollen auseinandersetzen. Pickens et al. (2008) identifizierten Rollen, die entweder den Fokus auf die Ausführung der Suche oder auf die Expertise legen. Besonders gewinnbringend für die gemeinsame Suche scheint dabei ein heterogen zusammengesetztes Team mit Such- und Domänenexperten zu sein. Soulier et al. (2014) setzten auf die Ergebnisse von Pickens et al. (2008) auf und entwickelten einen sogenannten *user-driven system-mediated approach*, um diese Rollen systembasiert zu identifizieren und die Suchperformanz zu verbessern. In einer weiteren Studie untersuchten Tamine und Soulier (2015) den Einfluss von vor Beginn der Suche zugeordneten und festen Rollen im Gegensatz zu solchen, welche sich spontan während der Suche ergaben. Aus ihrer Studie konnten sie weiterhin schließen, dass diese sich während der Suche ändern können, d.h. die Nutzenden nicht unbedingt durchgehend eine gleichbleibende Rolle besetzen. Die aufgeführten Rollenmuster sind primär im Kontext von kurzen Suchsessions bzw. beim *Collaborative Information Retrieval* (CIR) anwendbar. Elbeshausen (2019) und Elbeshausen et al. (2019) untersuchten, welche Rollen sich im Kontext von längeren oder wiederholten Suchsessions, die in einen komplexeren Task eingebunden sind (vgl. Hansen/Widén, 2016), identifizieren lassen. Solche Situationen, in denen ein Team über einen längeren Zeitraum zusammenarbeitet, bedürfen Rollen, welche sowohl das Such- als auch das Teamverhalten und generell die Zusammenarbeit adressieren.

Um die Rollenübernahme bei der gemeinsamen Suche verstehen zu können, ist das Verhältnis der Zusammenarbeitenden relevant. Dieses kann ent-

weder symmetrisch oder asymmetrisch sein – je nachdem, wie das Informationsbedürfnis aufgeteilt wird und welche Rolle jedes Gruppenmitglied einnimmt, um dieses Bedürfnis zu erfüllen. Bei einer symmetrischen Kollaboration haben die Gruppenmitglieder das gleiche Informationsbedürfnis und nehmen während der Suche einheitliche Rollen ein. Bei der asymmetrischen besetzen die Gruppenmitglieder unterschiedliche Rollen. Die Gründe dafür können sich unterscheiden; bspw. kann eine Arbeitsteilung der Suchaufgabe in Hinblick auf die Vertrautheit mit der Technik, der Jobhierarchie oder auf besonderer Expertise beruhen (vgl. Morris/Teevan 2010, S. 9).

Sofern die gemeinsame Suche an einem geteilten Endgerät stattfindet, können hier zwei spezifische Rollen entstehen. Die Person, welche die Kontrolle über das technische Gerät hat, übernimmt die Rolle des *driver*, die Rolle des *observer* beobachtet und gibt Vorschläge zur Ausführung der Suche (vgl. Amershi/Morris, 2008, S. 1651).

Studien, welche die kollaborative Suche von Personen unterschiedlicher Generationenzugehörigkeit untersuchen, sind bisher nicht in Erscheinung getreten. Allerdings finden sich in unterschiedlichen Arbeiten Hinweise darauf, dass die gemeinsame Suche von Angehörigen einer Familie und somit Mitgliedern verschiedener Generationen, die bspw. eine Urlaubsreise planen, ein typisches Szenario ist (vgl. Shah, 2015, S. 5 f.; Morris/Teevan, 2010, S. 7).

2.3 Generationenbegriff und Generationenmodelle

Der Generationenbegriff ist Gegenstand unterschiedlicher Forschungsdisziplinen und wird entsprechend heterogen definiert. Im sozialwissenschaftlichen Sinne ist laut Liebau (1977) eine Unterscheidung in drei Grundkonzepte notwendig, wobei Höpflinger (1999) zufolge im gesellschafts- und sozialpolitischen Bereich meist eine historisch-gesellschaftliche Definition Anwendung findet. Laut Liebau dient der Generationenbegriff nach diesem Konzept „zur Unterscheidung kollektiver historisch-sozialer Gruppierungen, die sich durch ihre gemeinsame Lage im historischen Raum, durch gemeinsame prägende Erlebnisse und durch gemeinsame Verarbeitungs- und Handlungsforderungen auszeichnen“ (Liebau, 1977, S. 20). Höpflinger bezeichnet als Generation eine soziale Kategorie, in der Menschen gleichzeitig aufwachsen und somit gemeinsame Erfahrungen machen. Entsprechend werden soziale, kulturelle und historische Gemeinsamkeiten erweckt. Als Beispiel hierfür lässt sich die Kriegsgeneration heranziehen (Höpflinger, 1999, S. 10). Kupperschmidt definiert Generationen als soziale Kohorten, in denen Men-

schen von verschiedenen Faktoren beeinflusst werden, diese jedoch umgekehrt genauso beeinflussen können. Dazu zählen etwa soziale, wirtschaftliche und politische Veränderungen und Geschehnisse. Als Ergebnis dieser geteilten Erfahrungen entwickeln die Kohorten eine gemeinsame Persönlichkeit beziehungsweise generationsbedingte Charakteristika wie Werte, Einstellungen, Prioritäten und das generelle Verhalten (Kupperschmidt, 2000, S. 66).

Wie erwähnt können Menschen, die gleichzeitig aufwachsen und gemeinsame Erfahrungen teilen, Generationen und somit sozialen Kategorien zugeordnet werden. Diese lassen sich nach Bruch et al. (2010) in verschiedene Gruppen einteilen: Nachkriegsgeneration (1935–1945), Wirtschaftswundergeneration (1946–1955), Baby-Boomer-Generation (1956–1965), Generation Golf (1966–1980) und Internet-Generation (ab 1981) (Bruch et al., 2010, S. 95). Aktuellere Forschung zeigt jedoch auf, dass diese Kategorien modifiziert werden sollten. Klaffke schlägt hierzu vor, die junge Generation (Generation Internet) nochmals in Generation X, Y und Z zu unterteilen. Die Begründung hierfür liegt darin, dass sich mit der fortschreitenden Mediatisierung auch die Sozialisationsbedingungen der letzten Jahre geändert haben (vgl. Klaffke, 2014, S. 11). Die Einteilung nach Klaffke ist in Tabelle 1 dargestellt.

Tab. 1: Einteilung der Generationen nach Klaffke (2014)

Generation	Geburtszeitraum
Nachkriegsgeneration	zwischen 1945 und 1955
Baby Boomer	zwischen 1956 und 1965
Generation X	zwischen 1966 und 1980
Generation Golf	zwischen 1965 und 1980
Generation Y	zwischen 1981 und 1995
Generation Z	ab 1996 bis heute

Die *Baby Boomer* machen zahlenmäßig den Großteil aller Generationen aus. Die Mitglieder repräsentieren die Mehrheit der Führungskräfte in Unternehmen und werden, mit Sicht aus 2020, in etwa fünf Jahren in den Ruhestand gehen (vgl. Klaffke, 2014, S. 12 f.). Die Kinder der *Baby Boomer* sind erwachsen, teilweise haben sie bereits selbst Kinder (vgl. Bruch et al., 2010, S. 104).

Die Mitglieder der *Generation X / Generation Golf* sind zurzeit die „Eltern“ und übernehmen erste Führungspositionen. Sie wurden durch die Kritik am Wohlstand ihrer Eltern und die Abgrenzung zu ihnen geprägt (vgl. Klaffke, 2014: 12 f.). Ab Mitte der 80er-Jahre fand für die *Generation X* und die

Generation Golf insbesondere im Berufsleben eine Konfrontation mit der Mediatisierung statt. Sie besitzen daher eine hohe Bereitschaft, neue Technologien zu erproben.

Die *Generation Y* wird aufgrund ihrer Charakteristik, das Leben und etliche Dinge zu hinterfragen, auch *Generation „why“* genannt. Die Mitglieder sind hauptsächlich die Kinder der Baby Boomer und stehen am Anfang ihrer beruflichen Karriere oder studieren. Sie befinden sich in einem Lebensabschnitt, in dem private und berufliche Entscheidungen zu treffen sind, und werden durch das globalisierte Spektrum an Möglichkeiten durch Unsicherheiten und Entscheidungsschwierigkeiten geplagt. Die Mitglieder dieser Generation verbrachten bereits einen Großteil ihrer Jugend mit der Existenz des Internets und sind daher besser ausgebildet als ihre Vorgängergenerationen (vgl. ebd., S. 13).

Die *Generation Z* ist als erste vollständig im digitalen Zeitalter aufgewachsen und wurde bereits in jungen Jahren mit dem Internet konfrontiert. Sie sind die Kinder der *Generationen X* sowie *Baby Boomer* und gehen, abhängig von ihrem Geburtsjahr, noch zur Schule oder haben mit einem Studium oder einer Ausbildung begonnen. Im Gegensatz zu ihren Vorgängergenerationen werden sie selbstständiger und individueller erzogen (vgl. ebd., S. 14)

3 Methodisches Vorgehen

Wie im vorhergehenden Kapitel aufgeführt, finden sich bisher keine Studien, welche das CIS in Gruppen, die hinsichtlich ihres Alters heterogen zusammengesetzt sind, untersuchen. Daraus ergibt sich für die vorliegende Untersuchung die Fragstellung „*Wie wird die kollaborative Informationssuche zwischen unterschiedlichen Generationen durchgeführt?*“. Um diese beantworten zu können, wurde ein qualitativer Forschungsansatz herangezogen, um das noch offene Feld anzugehen. Qualitative Forschung folgt dem Prinzip der Offenheit und lässt so neue und unbekannte Ergebnisse (vgl. Flick et al., 2012, S. 17) sowie Überraschungen und Einblicke in die Tiefgründigkeit von sozialen Phänomenen (vgl. Strübing 2018, S. 23) zu. Durch die offene Beobachtung können entsprechend neue Erkenntnisse entstehen und Hypothesen aufgestellt werden.

Untersuchungsplanung

Die Erhebung ist an die Vorgehensweise bei Huang et al. (2012) angelehnt. In dieser Studie untersuchten die Forschenden, wie ältere Personen nach Gesundheitsinformation suchen. Bildschirmaktivitäten und externalisierte Gedanken der Probanden wurden bei der Aufgabenbearbeitung mittels zweier Softwaretools aufgezeichnet. Im Anschluss fanden Interviews zu den Sucherfahrungen der Teilnehmenden statt, die aufgenommen und transkribiert wurden. Den Abschluss bildete ein Fragebogen zu demografischen Angaben und bisherigen Sucherfahrungen im Bereich der Gesundheitsinformation (vgl. ebd.). Für die vorliegende Untersuchung wurde dies insoweit abgewandelt, als dass nicht nach Gesundheits-, sondern auch Reiseinformation gesucht werden sollte, da dieses Szenario für den Kontext von CIS mit generationsgemischten Gruppen realitätsnäher erscheint. Die Aufzeichnung der Interaktion zwischen den Probanden erfolgte mittels einer Kamera.

Für die Untersuchung wurde folgendes Suchszenario erstellt:

Ihr möchtet gerne im Sommer zusammen verreisen und plant eure Reise. Diese soll etwa zwei Wochen lang und außerhalb Europas sein. Ihr möchtet euch vorab schon einmal grundlegend über Unterkünfte, Flüge und weitere wichtige Aspekte in dem Land eurer Wahl informieren und grob eure Reise gestalten.

1. Überlegt euch gemeinsam ein Land, in das ihr für zwei Wochen verreisen wollt.
2. Überlegt euch gemeinsam die Art der Reise (z. B. Städtetrip, Strandurlaub, Wanderung etc. Eine Pauschalreise ist nicht möglich.).
3. Überlegt euch gemeinsam, wie ihr arbeiten wollt, um die benötigte Information zu beschaffen. Nutzt jeder ein eigenes Gerät oder teilt ihr euch eins?
4. Überlegt euch gemeinsam, wie ihr vorgeht, um die benötigten Daten zu beschaffen. Teilt ihr die Aufgabe oder nicht?
5. Überlegt euch gemeinsam, wie ihr während der Suche kommuniziert und eure Ergebnisse miteinander teilt. Ihr sollt im engen Austausch stehen und miteinander interagieren.
6. Überlegt euch gemeinsam, wie und wo ihr die relevanten Ergebnisse speichert. Am Ende der Suche sollen alle Informationen an einem Ort gespeichert werden, der für euch beide zugänglich ist.

Die recht detaillierte Arbeitsanweisung sollte sicherstellen, dass alle Faktoren, welche für eine Kollaboration relevant sind, berücksichtigt werden. Die Planung des Vorgehens sollte gemeinsam erfolgen, insbesondere auch im Hinblick darauf, ob die Probanden ein Endgerät teilen oder jeweils ein eigenes nutzen. Das vorliegende Szenario sah keine längerfristig angelegte Suche über mehrere Sessions vor. Entsprechend fand die Suche an einem Ort

(co-located) und synchron statt (vgl. Morris/Teevan, 2010, S. 45 ff.). Die Aufgabenbeschreibung sah vor, dass die Probanden auch die Form der Kommunikation untereinander festlegen sollten, sowie die Art und Weise, wie Ergebnisse gesammelt und miteinander geteilt werden. Durch die Untersuchungsleitung wurde explizit darauf hingewiesen, dass die Kommunikation untereinander und der fortlaufende Austausch hinsichtlich des generellen Vorgehens, erfolgter Zwischenschritte usw. essenziell für die Bearbeitung der Aufgabe sind.

Damit die Reiseplanung möglichst individuell und unabhängig von Vorgaben, d.h. möglichst realitätsnah, verlaufen konnte, wurde das Szenario nach dem Prinzip der Offenheit gestaltet. Finanzielle oder zeitliche Vorgaben wurden nicht gemacht. Dennoch sollte eine Reise geplant werden, welche die Teilnehmenden auch in der Realität antreten würden. Um einen zu schnellen Abschluss der Suche zu vermeiden, wurde jedoch das Buchen einer Pauschalreise ausgeschlossen. Somit sollte auch sichergestellt werden, dass die Teams ausreichend interagieren, um so eine Beobachtung des Rollenverhaltens zu ermöglichen. Die Wahl der offenen Gestaltung und des freien Interagierens innerhalb des Suchszenarios erfolgte auch deshalb, um die Testatmosphäre zu reduzieren, eine möglichst natürliche Suche und Interaktion zwischen den Beteiligten zu begünstigen und so dem Sinn des qualitativen Forschungsansatzes gerecht zu werden (vgl. Flick et al., 2012, S. 17).

Anschließend an die Suche wurde angelehnt an Huang et al. (2012) ein Interview geplant. Dieses sollte für die vorliegende Untersuchung in der Gruppe durchgeführt werden, um dem Charakter der Untersuchung gerecht zu werden. Misoch zufolge liegen Vorteile des Gruppeninterviews auch darin, „[...] dass sich die Befragten gegenseitig ergänzen, korrigieren, zum Erzählen und Detaillieren anregen [...]“ (Misoch, 2015, S. 160). Um Meinungen und Erfahrungen der Kollaborateure zu untersuchen, handelt es sich hierbei demnach um eine effektive Methode. Der Leitfaden für das Interview sah folgendermaßen aus:

Leitfadeninterview Fragenkatalog

1. *Zu Beginn interessiert mich, ob ihr beide schon mal eine Reise zusammen organisiert und euch dazu gemeinsam im Internet informiert habt?*
2. *Wie sind eure Erfahrungen mit der Online-Reiseplanung generell?*
3. *Habt ihr vorher schon einmal gemeinsam eine kollaborative Informationssuche (in einem anderen Kontext) durchgeführt, z. B. nach Gesundheitsinformationen gesucht, weil eure Oma krank war oder etwas ähnliches?*
4. *Ihr habt jetzt die Aufgabe abgeschlossen. Erzählt doch mal detailliert, wie ihr dabei vorgegangen seid.*
5. *Wieso habt ihr verschiedene/ein Gerät(e) genutzt?*
6. *Für welche Kommunikationsmöglichkeit habt ihr euch entschieden und warum?*
7. *Wie empfanDET ihr eure Kommunikation während der Suche?*
8. *Wo habt ihr eure gefundenen Informationen abgespeichert und füreinander zugänglich gemacht?*
9. *Welche Alternativen fallen euch ein?*
10. *Würdet ihr sagen, dass eine/r während der Durchführung mehr die Leitung übernommen hat?*
11. *Wie habt ihr Entscheidungen getroffen?*
12. *Sind Differenzen jeglicher Art aufgetreten?*
13. *Was waren die größten Schwierigkeiten während eurer Aufgabe?*
14. *Würdet ihr eure Vorerfahrungen als hilfreich für die Durchführung ansehen?*
15. *Würdet ihr in Erwägung ziehen, auch in der Realität eine Reise wie eben durchgeführt zu planen?*
16. *Wie erfolgreich schätzt ihr eure kollaborative Suche ein?*

4 Untersuchungsdurchführung

Zur Vorbereitung der eigentlichen Studie wurde ein Pretest durchgeführt. Dieser fand im Januar 2019 im Haus einer Testgruppe statt, um eine natürliche Testumgebung zu generieren. Die Gruppe bestand aus einer weiblichen Person aus der Generation *Baby Boomer* und einer ebenfalls weiblichen Per-

son aus der *Generation Z*. Bei den Probandinnen handelt es sich um ein Mutter-Tochter-Verhältnis. Die Durchführung des Pretests ergab einige kleinere Modifikationen für die Hauptuntersuchung. Diese bezogen sich primär auf Anpassungen am Szenario und Änderungen am Interviewleitfaden. Dieser wurde insgesamt stärker spezifiziert, um so für die Untersuchung weniger relevante Ausführungen zu reduzieren. Weiterhin erfolgte eine noch präzisere Abfrage hinsichtlich der kollaborativen Zusammenarbeit der Beteiligten.

Die Hauptuntersuchung fand im Zeitraum von Februar bis März 2019 statt. Bei der älteren Generation handelte es sich um Personen aus der *Generation X* und den *Baby Boomern*, der Altersdurchschnitt hier betrug 52 Jahre. Die jüngeren Testteilnehmenden sind alle der *Generation Z* zuzuordnen, der Altersdurchschnitt lag bei 22,5 Jahren. Sechs der sieben Suchteams setzten sich aus einer Mutter und einer Tochter bzw. einem Sohn zusammen, eines aus einem Vater und seiner Tochter (vgl. Tab. 2). Die zwei Generationenausprägungen wurden bewusst ausgewählt, um konträre generationenspezifische Aspekte zu vereinen. Im Zuge der Untersuchung wurden die sieben Gruppen gebeten, eine kollaborative Informationssuche in Form eines Nutzertests durchzuführen. Für jede Gruppe fand eine individuelle Sitzung statt, wobei der gesamte Testdurchlauf je etwa 1,5 Stunden in Anspruch nahm, davon 45 bis 60 Minuten für die kollaborative Suche.

Tab. 2:

Übersicht über die CIS-Gruppen, die an der Hauptuntersuchung teilnahmen

Gruppe	Ältere Generation	Jüngere Generation	Verwandschaft
1	Generation X (1970)	Generation Z (1997)	Mutter/Tochter
2	Baby Boomer (1964)	Generation Z (1996)	Mutter/Tochter
3	Baby Boomer (1964)	Generation Z (1997)	Vater/Tochter
4	Generation X (1970)	Generation Z (1996)	Mutter/Sohn
5	Generation X (1968)	Generation Z (1996)	Mutter/Sohn
6	Generation X (1966)	Generation Z (1996)	Mutter/Sohn
7	Baby Boomer (1965)	Generation Z (1997)	Mutter/Tochter

Die Datenerhebung der einzelnen Gruppen erfolgte bei den Probanden zu Hause, um eine gewohnte Umgebung sicherzustellen und eine Laborsituation weitestgehend zu vermeiden. Nach der Begrüßung und informellem Austausch zur Auflockerung der Situation fand eine kurze Einführung in das Forschungsinteresse und die Durchführung der Untersuchung statt. Nach Einwilligung in die Studie und Unterschreiben der Datenschutzerklärung

füllten die Teilnehmenden einen kurzen demografischen Fragebogen aus, welcher auch das Verwandtschaftsverhältnis der Beteiligten erfasste. Daraufhin wurde das Suchszenario ausgeteilt, gelesen und mit der Testleitung besprochen. Anschließend begannen die ProbandInnen mit der Durchführung des Szenarios (s. Abschn. 3.1). Das Vorgehen wurde mithilfe einer Kamera, welche den Bildschirm und die Suchenden filmte, aufgenommen. Jeder Gruppe stand eine Stunde für das Durchlaufen und den Abschluss des Szenarios zur Verfügung. Abschließend wurden kurz die Ergebnisse der Suche reflektiert und daraufhin das leitfadengestützte Gruppeninterview durchgeführt.

5 Auswertung

Aus den Aufzeichnungen der Team-Interaktion und Bildschirmaktivität resultierten Videos, die alle Vorgänge und die Kommunikation der Gruppen enthielten. Um eine einheitliche Analyse des Datenmaterials zu gewährleisten, wurde ein Template entwickelt, welches die wesentlichen Vorgänge in komprimierter Form zusammenfasst. So ließ sich eine effiziente und einheitliche Auswertung ermöglichen. Das Template ist in folgende Kategorien eingeteilt:

TEMPLATE GRUPPE 2							
Generation Baby Boomer, Generation Z: Zusammenarbeit an einem Laptop (Person 1 gibt ein)							
Sequenz	Zeit	Vorgehen	Kommunikation	Person 1 (Gen. Z)	Person 2 (Gen. Baby Boomer)	Entscheidungsverhalten	Anmerkungen
1	00:00:14– 00:01:20	<i>Website:</i> google.de <i>Suchbegriff:</i> flug deutschland bali <i>Aktion:</i> Auswahl Website aus Ergebnissen	Reden über Reisedaten und Zeit	Person 1 sucht nach Flug	Person 2 schlägt Reisezeit vor	Person 2 entscheidet über Reisezeit	

Abb. 8 Beispielhafter Ausschnitt aus einer Tabelle zu den Suchaktivitäten (hier Gruppe 2). Die Tabelle basiert auf dem für die Untersuchung erstellten Template.

- Sequenz: Einteilung der Vorgänge und Zeiten in Abschnitte
- Zeit: Abschnitte, in denen ein bestimmtes Vorgehen stattfindet
- Vorgehen: besuchte/aufgerufene Websites, verwendete Suchbegriffe, Aktion innerhalb der Sequenz
- Kommunikation: Themen der Unterhaltung während der Durchführung

- Person 1: besonderes Verhalten der jüngeren Person
- Person 2: besonderes Verhalten der älteren Person
- Entscheidungsverhalten: Wie wurden Entscheidungen getroffen und wer veranlasste diese?
- Anmerkungen: besondere Vorkommnisse, welche zum Verständnis aufgeführt werden.

Für alle Gruppen erfolgte auf Basis des Templates die Zusammenfassung der Vorgehensweise während der Suche in tabellarischer Form. Anschließend wurde dieses in Kategorien unterteilt und in einer gemeinsamen Tabelle für alle Gruppen komprimiert. Folgende Kategorien ließen sich als Vorgänge während des CIS zwischen Generationen extrahieren:

- Suche nach Reiseinformation: Rundreisen; Flug; Unterkunft; Aktivitäten, Transfer, Mietwagen und Wohnmobil; Information über Orte; Information über das Land (bspw. Einreisebedingungen, Währung, notwendige Impfungen etc.)
- Datensicherung
- Sichtung Land auf Karte
- Planung der Route
- generelle Information: umfasst jegliche Information, welche nicht auf das zugrundeliegende Szenario bezogen ist.

Dieses Vorgehen erlaubte es, die relevanten Aspekte nebeneinander zu stellen und zu vergleichen. Die Klassifizierung der Vorgänge in Kategorien zeigt, dass jede Gruppe ungefähr die gleichen Schritte während der kollaborativen Suche durchlief, welche im nachfolgenden Abschnitt 6 dargestellt sind.

Die Auswertung der Interviews erfolgte in Anlehnung an die qualitative Inhaltsanalyse nach Mayring (2015). Es fand eine zusammenfassende Analyse statt, wobei aus den Interviews die relevanten Passagen ausgewählt wurden. Um diese für die Beantwortung der Forschungsfrage heranziehen zu können, wurden verschiedene Kategorien gebildet. Es entstanden die Kategorien „Gerätenutzung“, „Visualisierung“, „Alternativen“ und „Leitung und Rollen“. Die Passagen der einzelnen Interviewteilnehmenden wurden sinngemäß transkribiert und in die jeweiligen Kategorien eingeordnet. Die Zielsetzung hierbei war primär, ergänzende Daten zu den Videos zu sammeln, um die Hintergründe besser nachvollziehen zu können. Auch sollte den Beteiligten so die Möglichkeit geboten werden, ihr Vorgehen zu begründen und dieses für die Versuchsleitung besser nachvollziehen zu können. Der Leit-

faden des Interviews (s. Abschn. 3.1) verdeutlicht die gestellten Fragen und die Zielsetzung.

6 Ergebnisse

Nachfolgend sind die Ergebnisse dargestellt. Diese beziehen sich auf die zusammengeführten Daten aus Videoanalyse und Interview.

6.1 Eingennomene Rollen

Wie vorhergehend dargestellt, haben sechs der sieben Gruppen gemeinsam an einem Computer gearbeitet. Entsprechend konnten auch die Bedienelemente wie Maus und Tastatur nur von einer Person aktiv genutzt werden. Interessant ist, dass diese Rolle in allen Gruppen von der jüngeren Person (P1) übernommen wurde. Die Ergebnisse deuten also darauf hin, dass in einem synchronen, ortsgebundenen CIS-Szenario die Kontrolle über die Hardware durch die jüngere Generation vorgenommen wird. Diese Beobachtung bestätigt sich auch in den Aussagen der ProbandInnen in den Interviews. Dies lässt sich beispielhaft an den nachfolgenden Belegstellen aus den Transkripten zeigen:

„[P1] hat das alles gemacht, das können die Jungen generationsbedingt besser und sie saß auch direkt vor dem Computer.“ (P2, Gruppe 3)

„[...] derjenige, der die Maus bedient, macht halt mehr.“ (P1, Gruppe 4)

„Dadurch, dass ich das Laptop bedient habe, habe ich mehr gemacht.“ (P1, Gruppe 6)

Auch die Erfahrung im Umgang mit der generellen Technologie und der Internetnutzung scheint ein Grund dafür zu sein, warum die jüngere Generation die Leitung, zumindest in Bezug auf die Gerätenutzung, übernommen hat, wie die folgenden Ausschnitte aus den Transkripten zeigen:

„[P1 hat übernommen], weil sie erfahrener ist und das schon öfter in der Art und Weise gemacht hat.“ (P2, Gruppe 1)

„[P1] hat das am PC ausgeführt, das ist für sie auch leichter, ich brauche da viel länger für.“ (P2, Gruppe 2)

Aus den Antworten der älteren Generation (P2) lässt sich ableiten, dass diese die Erwartung hat, dass die junge Generation im Umgang mit dem Computer kompetenter ist. Diese Form der Organisation des CIS zwischen zwei Gene-

rationen scheint demnach ein fast schon automatischer, erwartungskonformer Prozess zu sein. Diese Ergebnisse gehen mit den Befunden von Morris einher, wonach sich junge Personen eher an kollaborativen Suchen beteiligen, wofür die Nutzung und Akzeptanz von neuen Technologien ursächlich zu sein scheint (Morris, 2013, S. 1189). Allerdings sollte die Übernahme der Führung der Technik nicht als generelle Leitung der Suche verstanden werden: Auch die ältere Generation gibt Handlungsanweisungen zur Informationssuche und fällt Entscheidungen zum Vorgehen. Diese Befunde spiegeln sich in den Interviews wider:

„Also, wir haben da die Aufgaben ein bisschen verteilt.“ (P1, Gruppe 2)

„[...] aber von der Reiseplanungsführung war das ausgeglichen.“ (P1, Gruppe 4)

„Ideenmäßig würde ich das auch ausgewogen bezeichnen.“ (P2, Gruppe 4)

Gruppe 2 gibt auch an, dass die ältere Person mehr Wünsche bezüglich der Reise geäußert hat, die jüngere Person nur die Ausführung übernommen hat. Interessant ist, ob sich in Gruppe 5, in der beide Personen ein eigenes Endgerät benutzt haben, eine Leitung zeigte. Laut der Beteiligten hat hier die jüngere Person die Leitung über die Organisation der Zusammenarbeit übernommen:

„Weil ich den Zettel bei mir hatte, wie das Szenario aussehen soll und ich gerne Aufgaben verteile.“ (P1, Gruppe 5)

„Genau, [P1] hat eigentlich durch den Prozess geführt.“ (P2, Gruppe 5)

Das Treffen von Entscheidungen wurde hingegen als ausgewogener Prozess wahrgenommen:

„Die Entscheidungen haben wir zusammen getroffen.“ (P1, Gruppe 5)

Tab. 3: Anzahl der Entscheidungen in den Gruppen

	G1	G2	G3	G4	G5	G6	G7
gemeinsam	12	6	8	7	7	2	7
P1	3	2	1	0	4	3	0
P2	2	2	6	3	1	0	1

G = Gruppe, P1 = Person der jüngeren Generation, P2 = Person der älteren Generation

Tabelle 3 zeigt eine Übersicht über die quantitative Verteilung der Entscheidungen in den Gruppen. Daran wird deutlich, dass der überwiegende Teil gemeinsam getroffen wurde. Eine Ausnahme stellt Gruppe 6 dar, in der die jüngere Person eine Entscheidung mehr gefällt hat als das Team gemeinsam. Allerdings ist hier anzumerken, dass in dieser Gruppe insgesamt deutlich weniger Entscheidungsprozesse stattgefunden haben als im Durchschnitt

über alle Teams hinweg (arithmetisches Mittel aller Entscheidungen je Gruppe = 11; arithmetisches Mittel gemeinsame Entscheidungen = 7). Auffällig ist auch Gruppe 3, in der die ältere Person deutlich mehr individuelle Entscheidungen getroffen hat als die jüngere und insgesamt mehr Entscheidungen als Einzelpersonen in den anderen Gruppen. Interessant ist, dass in Gruppe 5, im Gegensatz zu den übrigen sechs Gruppen, die ältere Person männlich (Vater, s. Tab. 2) ist. Ob dies allerdings tatsächlich ausschlaggebend ist oder nicht, lässt sich anhand der kleinen Stichprobe nicht belegen.

Die quantitative Betrachtung der Entscheidungen in Verbindung mit den Aussagen aus den Interviews und der Beobachtung des Verhaltens während der Suche spricht insgesamt für ein symmetrisches Entscheidungs- und damit auch Kollaborationsverhältnis. Auch wenn die individuell gefällten Entscheidungen unterschiedlich ausfallen, wie bspw. in Gruppe 4 oder Gruppe 6, spricht das nicht für eine Übernahme der Leitung. So sollte bedacht werden, dass auch individuelle Persönlichkeitsmerkmale der Beteiligten einfließen, etwa in Hinblick auf Extrovertiertheit oder Dominanz, welche das Entscheidungsverhalten beeinflussen.

Insgesamt lassen sich den Angehörigen der beiden Generationen nach Amershi und Morris (2008) folgende Rollen zuteilen: Die junge Generation spiegelt sich in der Rolle des *drivers* wider, da sie die Kontrolle über das technische Gerät hat. Die ältere Generation lässt sich in die Rolle des *observers* einordnen, da sie zuschaut und Vorschläge gibt. Demnach hat das CIS zwischen Angehörigen unterschiedlicher Generationen unterschiedliche Kollaborationsausprägungen: Im Hinblick auf die Vertrautheit mit Technik und die Expertise ist das Verhältnis zwar asymmetrisch, in Bezug auf die Autorität der Kollaborateure ist jedoch eine symmetrische Ausprägung vorhanden.

6.2 Facetten des CIS zwischen unterschiedlichen Generationen

Wie in Abschnitt 5 erwähnt, kamen in allen Teams ähnliche Schritte während der Suche vor. Die Gruppen 2, 3, 4 und 6 suchten zunächst nach Flügen, während sich die Gruppen 1, 5 und 7 zuerst über Rundreisen informierten. Die Suche nach Flügen stellte in mehr als der Hälfte der Gruppen den Startpunkt des Szenarios dar und wurde demnach als eine grundlegende Information für die weitere Planung der Reise aufgefasst. Auch die Suche nach Aktivitäten und Unterkünften war Bestandteil jeder Suche und scheint demnach ein grundlegend relevanter Aspekt der Reiseplanung zu sein. Die Mehrzahl der Beteiligten beendete das Szenario nach 45 Minuten. Eine Ausnahme dazu

stellt Gruppe 5 dar. Diese wählte eine andere Form der Zusammenarbeit und führte die kollaborative Suche mit zwei Rechnern durch. Dadurch ist eine umfangreichere Suche entstanden, welche einen höheren Kommunikationsbedarf erforderte.

Während in allen Gruppen zwar ähnliche Schritte erfolgten, ist dennoch anzumerken, dass ein direkter Vergleich des konkreten Vorgehens Abweichungen im Prozessablauf zeigt. Dafür lassen sich diverse Gründe annehmen, bspw. die Art der gewählten Reise oder Unterschiede hinsichtlich der Expertise der einzelnen Gruppen.

Interessant ist, dass sich sechs der sieben Gruppen für die Nutzung eines gemeinsamen Endgerätes für die Suche entschieden haben. In Gruppe 4 war dies ein Desktop-Computer, Gruppe 5 entschied sich für die Nutzung von zwei Laptops, die übrigen Gruppen führten die Suche gemeinsam an einem Laptop durch. Die Auswertung der Interviews ergab sechs kategorisierbare Gründe, die für die gemeinsame Nutzung eines Gerätes zur Durchführung der kollaborativen Informationssuche sprechen. Ein Aspekt, der genannt wurde, ist die Geschwindigkeit:

„Irgendwo ging es dann auch relativ schnell.“ (P2, Gruppe 1)

„[...] es geht bei uns auch um die Geschwindigkeit.“ (P2, Gruppe 6)

Weiterhin wurde die anfängliche Unsicherheit über die Reisegestaltung als Grund genannt:

„Die Route stand noch nicht genau fest und wir guckten erst einmal gemeinsam, wo der Weg uns hinführt und dann ist man dabei geblieben.“ (P2, Gruppe 1)

Eine Gruppe erwähnte den „*Information Overload*“, welcher bei der Nutzung von zwei Geräten entstehen würde, als Nachteil:

„Es sind eh schon immer so viele Informationen und auf zwei Geräten dann noch einmal mehr, das verwirrt und man verliert den Überblick.“ (P1, Gruppe 2)

Als Grund wurde auch genannt, dass beide Personen eine gemeinsame Zielsetzung verfolgen und gemeinsame Entscheidungen treffen müssen:

„Man sucht eh das Gleiche, also kann man es auch auf einem Rechner machen.“ (P1, Gruppe 4)

„[...] es ist angenehmer, weil man über das Gleiche spricht und das Gleiche sieht und dann ist es nicht so anstrengend.“ (P1, Gruppe 7)

„[...] beide Augenpaare können darauf gucken und eine Meinung dazu geben.“ (P1, Gruppe 4)

In Gruppe 6 betonte das Mitglied der jüngeren Generation zudem, dass eine „einfachere Kommunikation“ möglich sei.

Als weiteres Argument wurde die Computeraffinität der jüngeren Generation angeführt:

„Ich kann das schneller bedienen.“ (P1, Gruppe 6)

„Ich finde das deutlich angenehmer so, weil ich noch nie gerne am Computer gearbeitet habe und meine Tochter viel flinker ist. Ich muss viel länger gucken.“ (P2, Gruppe 7)

Gruppe 5 gab mehrere Begründungen für die Nutzung von zwei Endgeräten an, unter anderem die bessere Möglichkeit des Vergleichs der Information:

„Jeder geht auf eine andere Seite und dann kann man vergleichen“ (P2, Gruppe 5)

Auch die größere Vielfalt wurde genannt:

„Ich finde, man hat auch unterschiedliche Auswahl [...]“ (P2, Gruppe 5)

„[...] man hat ein viel größeres Spektrum an Informationen gleichzeitig bekommen.“ (P1, Gruppe 5).

Weiterhin gab die Gruppe an:

„Wenn ich etwas machen muss, dann möchte ich lieber selbst schnell eingeben, bevor ich es vergesse.“ (P1, Gruppe 5)

Das Team hob auch hervor, dass so eine parallel durchgeführte Suche möglich sei:

„Wir konnten uns auf einem Rechner die Karte angucken und auf einem anderen Rechner die Orte nochmal extra, so musste man nicht zwischen den Tabs springen.“ (P1, Gruppe 5)

Eine weitere Anforderung aus dem Szenario war die Visualisierung der Ergebnisse. Die Gruppenmitglieder sollten diese stets für einander zugänglich machen und Optionen für Überarbeitungen und Ergänzungen ermöglichen. Die Vorgabe war, dass die Daten so abgelegt werden, dass das Gruppenmitglied zu jeder Zeit darauf zugreifen kann. Da die Teams kein CIS-System für die Suche verwendeten, werden die genutzten Alternativen nachfolgend kurz aufgeführt. Die Idee dazu, wie und wo die Information abgespeichert werden soll, entstand meist durch das jüngere Gruppenmitglied. Aus den Interviews ergaben sich hierzu unterschiedliche Kategorien.

Als eine Methode wurde das Einfügen der relevanten Weblinks in ein Dokument genannt. Die Teilung mit dem Teammitglied fand auf unterschiedliche Art und Weise statt:

„Ich würde das dann an meine Mutter weiterleiten, vermutlich per WhatsApp, dann kann sie auch nochmal darauf zugreifen.“ (P1, Gruppe 2)

Gruppe 1 fügte die Websitelinks in eine Mail ein und verschickte diese, Gruppe 3 hat die Ergebnisse in eine Excel-Tabelle eingefügt und ebenfalls

via E-Mail versendet. Als weitere Methode wurden Screenshots der relevanten Websites erstellt und in ein Textdokument eingefügt. Nur eine Gruppe (Gruppe 7) hat Lesezeichen genutzt und dies wie folgt begründet:

„Weil wir dann beide sofort darauf Zugriff haben [...]“ und „Mama weiß dann auch, wo sie hin muss und die Seite ist gleich auf.“ (P1, Gruppe 7).

Zwei Gruppen (Gruppen 5 und 6) entschieden sich dazu, Google Drive zum Teilen der Ergebnisse zu nutzen. Gruppe 5, die an zwei Rechnern arbeitete, begründete das folgendermaßen:

„Das war eine gute Lösung, weil beide die Ablage bearbeiten können“ und „man kann auch im Nachhinein noch Information hinzufügen und muss dafür nicht zusammensitzen und man kann darüber kommunizieren.“ (P1, Gruppe 5)

Diese Ergebnisse decken sich mit der Studie von Morris (2013), welche aufzeigt, dass traditionelle, d.h. bekannte und (häufig) genutzte, Kommunikationstechnologien bevorzugt für das kollaborative Arbeiten genutzt werden, auch wenn spezifische Kollaborationswerkzeuge herangezogen werden könnten.

Eine Übersicht der relevanten Aspekte, gegliedert nach dem Vorgehen bei der kollaborativen Suche, technischen und sozialen Aspekten, ist in Abbildung 9 dargestellt.

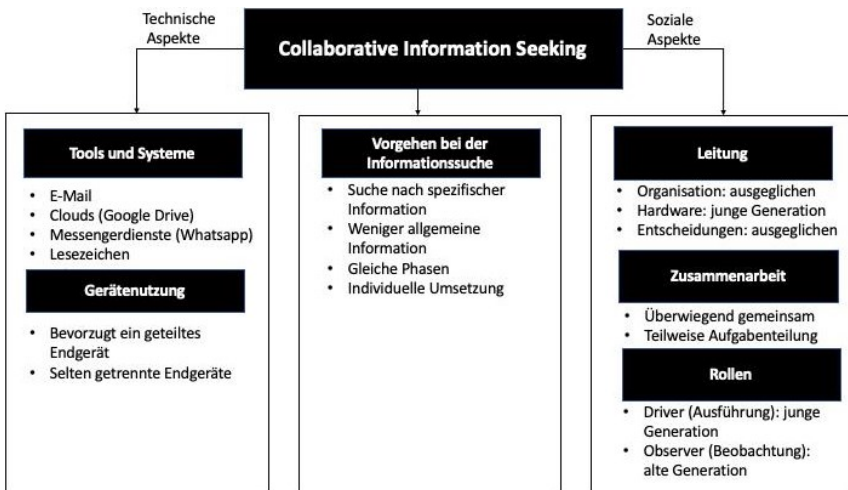


Abb. 9

Beobachtete relevante Aspekte des CIS in generationenspezifisch heterogenen Teams

7 Fazit und Ausblick

Anhand der Fragestellung „*Wie wird die kollaborative Informationssuche zwischen unterschiedlichen Generationen durchgeführt?*“ war es in der vorliegenden Untersuchung möglich, einige interessante Aspekte und Prozessschritte des CIS mit Mitgliedern unterschiedlicher Generationen aufzuzeigen. Das methodische Vorgehen – basierend auf Suchszenario, Videoanalyse und Interviews – hat sich für die vorliegende Zielsetzung als effektiv erwiesen und konnte dazu beitragen, die kollaborative Suche mit heterogener Generationenzusammensetzung näher zu beleuchten.

Die Untersuchung konnte aufzeigen, dass in allen Gruppen ähnliche Prozessschritte durchlaufen werden, diese jedoch nicht unbedingt in der gleichen Reihenfolge auftreten.

Den ProbandInnen wurde die freie Wahl gelassen, ob sie an einem oder zwei Geräten arbeiten und ob und inwiefern sie eine Aufgabenteilung vornehmen. Interessant ist hier, dass sechs von sieben Gruppen die Arbeit an einem Endgerät und mit wenig Aufgabenteilung bevorzugten, wofür sich mehrere Gründe annehmen lassen. Ein Grund könnte in der Form des gewählten Szenarios liegen, welches auf einen Zeitraum von einer Stunde ausgelegt war. Dieser eher kurze Zeitraum für die Suche hat möglicherweise dazu geführt, dass den Teilnehmenden eine aufwendige Koordination und Organisation der Aufgabenteilung nicht sinnvoll erschien und schon von daher die gemeinsame Suche vor einem Bildschirm gewählt wurde. Eine weitere Begründung könnte in der Teamzusammensetzung liegen. Die Gruppen waren alle aus einander vertrauten Personen zusammengesetzt, die aus einer Familie entstammen und sich entsprechend lange und vermutlich gut kennen. Die Vertrautheit unter den Gruppenmitgliedern lässt sich als weiterer Grund für die gewählte Art der Zusammenarbeit annehmen. Die Befragten selbst gaben an, dass sie diese Form primär als effizienter einschätzten. Eine Gruppe entschied sich für die Arbeit an zwei Endgeräten und begründete dies mit einer größeren Vielfalt an Informationen und die schnellere oder einfachere Umsetzung der Suche.

Hinsichtlich möglicher Rollen, welche während der Suche eingenommen wurden, konnte analog zu Amershi und Morris (2008) die Übernahme der Rollen *driver* und *observer* beobachtet werden. Die Rolle des *driver* wurde dabei in den sechs Gruppen, welche die Aufgabe teilten, von der jungen Generation, jene des *observer* von der älteren Generation übernommen. Die

Teilnehmenden begründeten dies vornehmlich mit der höheren Technikaffinität der jungen Generation und der Vertrautheit mit dem Internet. Eine generelle Leitung der Suche oder Zusammenarbeit ergab sich daraus nach Aussagen der ProbandInnen nicht. Auch eine Auswertung der Entscheidungen in den Gruppen (s. Tab. 3) weist darauf hin, dass diese in erster Linie gemeinsam getroffen wurden und sich kein Muster in Hinblick auf die Generationenzugehörigkeit zeigt.

Für nachfolgende Untersuchungen ist geplant, die Studie unter veränderten Bedingungen, etwa im Hinblick auf die Zeit, durchzuführen. Dadurch, dass den Gruppen mehr Zeit für die Suche zur Verfügung steht, steigt die Wahrscheinlichkeit dafür, dass die Suche nicht durchgehend synchron und an einem Ort ausgeführt wird. Ferner ist anzunehmen, dass CIS in einem längerfristig angesiedelten Szenario auch stärker strukturiert und koordiniert wird. Weiterhin wäre es interessant, die Gruppengröße zu erweitern, was ebenfalls Auswirkungen auf die Koordination erwarten lässt. So ist denkbar, dass ab drei oder vier Personen tatsächlich eine Leitung notwendig wird, welche die Aufgaben verteilt und den Überblick über die Durchführung der Suche behält (vgl. Elbeshausen, 2019).

Im vorliegenden Fall setzten sich die Gruppen aus miteinander vertrauten Personen zusammen. Geplant ist, die Studie noch einmal mit einander unbekanntem Personen unterschiedlicher Generationenzugehörigkeit durchzuführen und zu prüfen, ob sich hierdurch geänderte Abläufe und Rollenübernahmen ergeben. Ein solches Vorgehen wäre auch in einem professionellen Kontext interessant. Denkbar wäre, die Untersuchung noch einmal in einem Unternehmen durchzuführen, in dem ältere und jüngere Generationen an komplexen Projekten zusammenarbeiten, und hier zu prüfen, inwiefern die veränderte Situation Einfluss auf die Ergebnisse hat.

Generell muss bedacht werden, dass es sich vorliegend um eine kleine Stichprobe mit sieben Kleingruppen, d.h. 14 Personen handelt. Für die durchgeführte qualitative Untersuchung ist das zwar ein ausreichender Rahmen, um erste Befunde in Hinblick auf die gewählte Fragestellung zu erhalten. Um belastbare Ergebnisse zu erhalten, welche die hier erfolgten Befunde bestätigen oder widerlegen, müssen aufsetzende quantitative Untersuchungen folgen.

Die genutzten Tools deuten darauf hin, dass die Feststellung von Morris (2013) noch immer zutreffend ist und vertraute Kommunikationstools für CIS bevorzugt werden. Spezifische Kollaborationswerkzeuge wurden nur in Form von cloudbasierten Diensten genutzt. Die herangezogenen Work-

arounds sind im Kontext von kurzen kollaborativen Suchen als ausreichend nützlich einzuschätzen, was sich auch in der Bearbeitungszeit (vorliegend i.d.R. 45 min.) widerspiegelt. Gerade bei längerfristig angelegtem CIS im Kontext einer größeren gemeinsamen Aufgabe ist der dadurch entstehende zusätzliche kognitive Workload problematisch. Zum einen sinkt so die Effizienz bei der Aufgabebearbeitung, sodass mehr Zeit investiert werden muss. Zum anderen werden die Suchenden unnötig belastet und verschwenden wertvolle kognitive Ressourcen, welche dadurch nicht für die Bearbeitung der eigentlichen Aufgabe zur Verfügung stehen können. Gerade im professionellen Bereich sollte dies vermieden werden, sodass die Beteiligten konzentriert an ihren Kernaufgaben arbeiten und Synergien, welche durch die unterschiedliche Expertise bei der Kollaboration entstehen können, befördert werden können.

In den vorliegenden Ergebnissen zeigt sich, dass die ältere Generation eine gewisse Unsicherheit in Bezug auf die Nutzung neuerer Technologien hat. Auf der anderen Seite treffen die Gruppenmitglieder jedoch viele Entscheidungen gemeinsam und es lässt sich nicht erkennen, dass eine Generation die Leitung übernimmt. Um die Zusammenarbeit gerade auch im professionellen Kontext zu befördern und generell ein effizienteres Arbeiten zu ermöglichen, könnte es sinnvoll sein, die Sicherheit der älteren Generation im Umgang mit den zugrundeliegenden Technologien zu erhöhen. Denkbar wären hierfür bspw. Schulungen oder Hilfestellungen, welche bei der Nutzung von Hard- und Software herangezogen werden können.

Generell zeigt gerade die aktuelle Corona-Pandemie, wie wichtig die Onlinenutzung auch für die älteren Generationen geworden ist. Das betrifft nicht nur den vorliegend untersuchten kollaborativen Rahmen, sondern auch die individuelle Nutzung, was die eingangs erwähnte Studie zur Online-Nutzung bestätigt. Geeignete Unterstützungsmaßnahmen, welche die Kompetenz der Nutzenden und damit das Selbstvertrauen bei der Anwendung stärken, könnten hier Abhilfe schaffen und so die Vernetzung untereinander befördern. So ließe sich auch auf längere Sicht ermöglichen, dass unterschiedliche Generationen pandemieresistent zusammenarbeiten und voneinander profitieren können.

Quellen

- Amershi, Saleema; Morris, Meredith Ringel (2008): CoSearch: A System for Co-located Collaborative Web Search. In: *Proceedings of CHI 2008*, S. 1647–1656. doi: [10.1145/1357054.1357311](https://doi.org/10.1145/1357054.1357311).
- Bruch, Heike; Kunze, Florian; Böhm, Stephan (2010): *Generationen erfolgreich führen. Konzepte und Praxiserfahrungen zum Management des demographischen Wandels*. Wiesbaden: Gabler.
- Denning, Peter J.; Yaholkovsky, Peter (2008): Getting to “We”. *Communications of the ACM* 51 (4), 19–24.
- Elbeshausen, Stefanie (2019): *Collaborative Information Seeking. Integrierte Prozessmodellierung für die Ableitung systembasierter Unterstützungsmaßnahmen*. Hildesheim: Universitätsverlag Hildesheim. doi: <http://dx.doi.org/10.18442/020>
- Elbeshausen, Stefanie; Mandl, Thomas; Womser-Hacker, Christa (2015): Collaborative Information Seeking in the Context of Leisure and Work Task Situations: A Comparison of Three Empirical Studies. In: Hansen, P.; Shah, C.; Klas, C. P. (Hrsg.): *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Cham: Springer International Publishing, S. 73–98.
- Flick, Uwe; Kardorff, Ernst von; Steinke, Ines (2012): Was ist qualitative Forschung? Einleitung und Überblick. In: Flick, Uwe (Hrsg.): *Qualitative Forschung. Ein Handbuch*. 9. Aufl., Reinbek: Rowohlt, S. 13–30.
- González-Ibáñez, Roberto; Haseki, Müge; Shah, Chirag (2012): Understanding effects of time and proximity on collaboration: implication for technologies to support collaborative information seeking. In: *CHI EA '12: CHI '12 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: ACM Press, S. 1805–1810.
- Hansen, Preben; Järvelin, Kalervo (2005): Collaborative Information Retrieval in an information-intensive domain. *Information Processing & Management* 41 (5), 1101–1119. doi: [10.1016/j.ipm.2004.04.016](https://doi.org/10.1016/j.ipm.2004.04.016)
- Hansen, Preben; Widen, Gunilla (2016): The embeddedness of collaborative information seeking in information culture. *Journal of Information Science* 43 (4). doi: [10.1177/0165551516651544](https://doi.org/10.1177/0165551516651544)
- Heinström, Jannica (2002): *Fast surfers, broad scanners, and deep divers. Personality and information seeking behaviour*. Åbo, Finland: Åbo Akademis förlag.
- Höpflinger, Francois (1999): *Generationenfrage – Konzepte, theoretische Ansätze und Beobachtungen zu Generationenbeziehungen in späteren Lebensphasen*. Lausanne: Réalités Sociales.

- Huang, Man; Hansen, Derek; Xie, Bo (2012): Older Adults' Online Health Information Seeking Behavior. In: *iConference '12 Proceedings of the 2012 iConference*. New York, NY: ACM Press, S. 338–345. doi: [10.1145/2132176.2132220](https://doi.org/10.1145/2132176.2132220)
- Hyldegård, Jette (2006). *Between individual and group – exploring group members' information behavior in context*. Diss., Department of Information Studies, Royal School of Library and Information Science, Copenhagen.
- Klaffke, Martin (2014): Erfolgsfaktor Generationen-Management – Handlungsansätze für das Personalmanagement. In: Klaffke, Martin (Hrsg.): *Generationen-Management. Konzepte, Instrumente, Good-Practice-Ansätze*. Wiesbaden: Springer Gabler, S. 3–26.
- Kupperschmidt, Betty (2000): Multigenerational Employees: Strategies for Effective Management. *The Health Care Manager* 19 (1), 65–76.
- Liebau, Eckart (1997): Generation – ein aktuelles Problem? In: Liebau, Eckart (Hrsg.): *Das Generationenverhältnis. Über das Zusammenleben in Familie und Gesellschaft*. Weinheim, München: Juventa Verlag, S. 15–38.
- London, Scott (1995): Building collaborative communities. <http://www.scottlondon.com/articles/oncollaboration.html>
- Mayring, Phillipp (2015): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Misoch, Sabina (2015): *Qualitative Interviews*. Berlin: de Gruyter Oldenbourg.
- Morris, Meredith Ringel; Teevan, Jaime (2010): Collaborative Web Search – Who, What, Where, When and Why. *Synthesis Lectures on Information Concepts Retrieval and Services* 1 (1.). doi: [10.2200/S00230ED1V01Y200912ICR014](https://doi.org/10.2200/S00230ED1V01Y200912ICR014)
- Pickens, Jeremy; Golovchinsky, Gene; Shah, Chirag; Qvarfordt, Pernilla; Back, Maribeth (2008): Algorithmic mediation for collaborative exploratory search. In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08)*. New York, NY: ACM Press, S. 315–322. doi: [10.1145/1390334.1390389](https://doi.org/10.1145/1390334.1390389)
- Poltrock, Steven; Grudin, Jonathan; Dumais, Susan T.; Fidel, Raya; Bruce Harry; Pejtersen, Annelise Mark (2003): Information seeking and sharing in design teams. In: *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '03)*. New York, NY: ACM Press, S. 239 bis 247.
- Reddy, Madhu; Jansen, Bernard; Krishnappa, Rashmi (2009): The role of communication in collaborative information searching. *Proceedings of the American Society for Information Science and Technology* 45 (1), 1–10.
- Reddy, Madhu; Spence, Patricia Ruma (2008): Collaborative Information Seeking: A Field Study of a Multidisciplinary Patient Care Team. *Information Processing & Management* 44 (1), 242–255.

- Shah, Chirag (2010): Collaborative Information Seeking: A Literature Review. In: *Advances in Librarianship* 32, 3–33.
- Shah, Chirag (2015): Collaborative Information Seeking: From “What” and “Why?” to “How?” and “So What?”. In: Hansen, Preben; Shah, Chirag; Klas, Claus-Peter (Hrsg.): *Collaborative Information Seeking. Best Practices, New Domains and New Thoughts*. Cham: Springer International Publishing, S. 3–17.
- Shah, Chirag; Leeder, Chris (2016): Exploring collaborative work among graduate students through the C5 model of collaboration: A diary study. *Journal of Information Science* 42 (5), 609–629.
- Sonnenwald, Diane (1995): Contested collaboration: A descriptive model of intergroup communication in information system design. *Information Processing & Management* 31 (6), 859–877.
- Sonnenwald, Diane; Pierce, Linda (2000): Information Behavior in Dynamic Group Work Contexts: Interwoven Situational Awareness, Dense Social Networks and Contested Collaboration in Command and Control. *Information Processing & Management* 36 (3), 461–479. doi: [10.1016/S0306-4573\(99\)00039-4](https://doi.org/10.1016/S0306-4573(99)00039-4)
- Soulier, Laure; Shah, Chirag; Tamine, Lynda (2014): User-driven system-mediated collaborative information retrieval. In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14), Gold Coast, Australia*. New York, NY: ACM Press, S. 485–494.
- Strübing, Jörg (2018): *Qualitative Sozialforschung. Eine komprimierte Einführung*. München: de Gruyter Oldenbourg.
- Tamine, Lynda; Soulier, Laure (2015): Understanding the impact of the role factor in collaborative information retrieval. In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM '15), Melbourne, Australia (S. 43–52)*. New York, NY: ACM Press.
- Vick, Thais; Nagano, Marcelo Seido; Popadiuk, Silvio (2015): Information culture and its influences in knowledge creation: Evidence from university teams engaged in collaborative innovation projects. *International Journal of Information Management* 35 (3), 292–298.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 247–272. DOI: doi.org/10.5283/epub.44947.

Den Einfluss der Suchmaschinenoptimierung messbar machen

Ein halb-automatisierter Ansatz zur Bestimmung von optimierten Ergebnissen auf Googles Suchergebnisseiten

Sebastian Sünkler, Dirk Lewandowski

Hamburg University of Applied Sciences, Department of Information,
Finkenau 35, 22081 Hamburg (Germany)
{[sebastian.suenkler](mailto:sebastian.suenkler@haw-hamburg.de), [dirk.lewandowski](mailto:dirk.lewandowski@haw-hamburg.de)}@haw-hamburg.de

Abstract

Suchmaschinenoptimierung (SEO) kann einen erheblichen Einfluss auf die Suchergebnisseiten in kommerziellen Suchmaschinen haben. Bisher ist jedoch unklar, welcher Anteil der (Top-)Ergebnisse tatsächlich optimiert wurde und wie dies die Auswahl und das Ranking der angezeigten Treffer bestimmt. Für die Untersuchung über den Einfluss von SEO auf die Suchergebnisse haben wir halb-automatische Prozesse und ein Software-Framework entwickelt, das mit einem regelbasierten Klassifikator die Wahrscheinlichkeit von Optimierungsmaßnahmen auf Suchergebnissen zu ermittelt. Der Ansatz basiert auf 20 Merkmalen, zu denen die Analyse über genutzte SEO-Plug-ins und Analytics Tools, die Auswertung technischer Indikatoren wie der Ladezeit und der Verwendung bestimmter Tags wie zur Description sowie eine manuelle Klassifikation auf Basis von vorab zusammengestellten Listen mit optimierten und nicht optimierten Webseiten gehören. Dieser Ansatz wurde auf drei Datensätzen mit insgesamt 2.043 Suchanfragen und 263.790 Ergebnissen angewendet. Die Ergebnisse zeigen, dass ein großer Teil der in Google gefundenen Seiten zumindest wahrscheinlich optimiert ist, was im Einklang mit Aussagen von SEO-Experten steht, die sagen, dass es sehr schwierig ist, ohne die Anwendung von SEO-Techniken in Suchmaschinen sichtbar zu werden.

Keywords: Suchmaschinen; Suchmaschinenoptimierung; SEO; Datenanalyse

1 Einleitung

Mit Suchmaschinenoptimierung (SEO) werden Maßnahmen bezeichnet, die das Ziel haben, Ranking-Algorithmen von Suchmaschinen zu entsprechen, um dadurch eine verbesserte Platzierung in den organischen Suchergebnissen zu erreichen (Griesbaum, 2013). Sie kann dadurch zwischen der Unterstützung von Suchmaschinen beim Auffinden und Indexieren relevanter Inhalte und der Manipulation ihrer Ergebnisse eingeordnet werden. Inhaltsanbieter wie Online-Shops und Medienangebote sind auf gute Platzierungen in den Suchergebnissen in kommerziellen Suchmaschinen angewiesen, um Traffic für ihre Angebote zu generieren, da Suchmaschinen den wichtigsten Zugang zu Inhalten im Web darstellen (Van Couvering, 2009). Dadurch entsteht eine oft stark ausgeprägte Abhängigkeit insbesondere von Google, dem Marktführer in der Websuche. Google hat einen Marktanteil von 87% in den Vereinigten Staaten (StatCounter, 2020a) und 93% in Europa (StatCounter, 2020b) über alle Plattformen hinweg (Stand: August 2020). Durch die hohe Relevanz von Suchmaschinenoptimierung für Webseiten hat sich eine starke Branche für SEO ausgebildet, die beispielsweise in den USA im Jahr 2020 einen Umsatz von 80 Milliarden Dollar erreichen wird (McCue, 2018). Eine zentrale Frage, die sich daraus ergibt, ist, inwieweit Ergebnislisten in Suchmaschinen durch Suchmaschinenoptimierung (SEO) extern beeinflusst werden. Nach unserem Wissen gibt es keine Untersuchungen darüber, wie man technisch messen kann, ob eine Webseite (oder Website) durch SEO-Maßnahmen „manipuliert“ wurde.

Die Untersuchung dieses Themas ist für die Informationswissenschaft auf verschiedenen Ebenen relevant. Für den Bereich des Information Retrieval (IR) ist sie von Bedeutung, da es externe Einflüsse auf die Ergebnisse von Information-Retrieval-Systemen berücksichtigt und somit unsere Sichtweise auf IR-Systeme erweitert wird. Abgesehen von der Konzentration auf Maßnahmen, die Anbieter von IR-Systemen ergreifen können, um ihre Ranking-Funktionen zu verbessern, hat sich die bisherige Forschung darauf konzentriert, die Benutzer durch die Analyse von Klicks und weiteren Interaktionen mit den Ergebnissen in die Ranking-Modelle einzubeziehen (z.B. Wang et al., 2016; Zamani et al., 2017). Darüber hinaus gibt es einige Forschungsarbeiten über die Eigeninteressen von Suchmaschinenunternehmen und darüber, wie diese das, was auf den Ergebnisseiten angezeigt wird, beeinflussen können (z.B. Lewandowski & Sünkler, 2013). Einige Forschungs-

arbeiten haben sich auch auf die Mischung aus bezahlten und organischen Ergebnissen auf den Ergebnisseiten von Suchmaschinen (SERPs) konzentriert und darauf, wie Nutzer, die zwischen den beiden Ergebnistypen unterscheiden können bzw. nicht unterscheiden können, ein unterschiedliches Auswahlverhalten zeigen (z.B. Lewandowski, 2017).

Neben den genannten Aspekten sind auch die Fragen über den Einfluss von suchmaschinenoptimierten Inhalten auf den Wissenserwerb von Suchmaschinennutzenden und auch im Bereich der Informationskompetenz relevant. Suchmaschinenoptimierung kann als Teil von Suchmaschinenwerbung gesehen werden, da diese dazu führen soll, dass Inhalte prominenter in Suchergebnissen auftauchen. Dazu werden Maßnahmen durchgeführt, die zum Teil auch als Manipulation des Rankings eingesetzt werden. Zur Suchmaschinenwerbung gehört auch das Suchmaschinenmarketing (SEM) durch die Schaltung von Textanzeigen. Diese unterscheiden sich von den organischen Ergebnissen vor allem dadurch, dass Suchmaschinen sie als Werbung kennzeichnen; für suchmaschinenoptimierte Inhalte gibt es allerdings keine entsprechende Kennzeichnung. Während bereits Studien zu der Wahrnehmung von Werbeanzeigen in Suchmaschinen durchgeführt wurden und zeigen, dass Benutzer Schwierigkeiten bei der Unterscheidung zwischen Werbung und organischen Suchergebnissen haben (Lewandowski et al., 2018), gibt es keine Forschung zur Wahrnehmung von SEO – auch nicht dazu, wie sich diese Inhalte auf den Wissenserwerb auswirken können.

Für die Durchführung von Studien zu den genannten Bereichen ist es wichtig, dass Möglichkeiten geschaffen werden, um Aussagen darüber zu treffen, ob Dokumente suchmaschinenoptimiert sind oder nicht. In diesem Beitrag stellen wir einen Ansatz zur Identifizierung von suchmaschinenoptimierten Inhalten vor. Dabei wird aus einer Kombination von Analyseverfahren eine Aussage über die Wahrscheinlichkeit von SEO auf einer Webseite getroffen. In diesem Beitrag beschreiben wir zunächst die Vorgehensweise zur Identifizierung der SEO-Merkmale als Basis für unseren regelbasierten Klassifikator zur Einordnung der URLs. Darauf folgend stellen wir erste Ergebnisse über den Einsatz unseres Ansatzes auf drei Datensätzen mit insgesamt 263.790 Suchergebnissen vor. Anschließend folgen ein Fazit und eine Darstellung des weiteren Vorgehens zur Weiterentwicklung unserer Methoden.

2 Identifizierung von SEO-Faktoren

Unsere Diskussion der SEO-Faktoren und ihrer Umsetzung in unserem System basiert auf einer ausführlichen Durchsicht der Fachliteratur (z.B. Enge, 2015; Erlhofer, 2019) und Interviews mit SEO-Experten (Schultheiß/Lewandowski, 2020). Insgesamt besteht unser Modell aus 47 Faktoren, die wir für die Umsetzung in unserem System priorisiert haben. Es sei darauf hingewiesen, dass das aktuelle Modell nur 20 Faktoren berücksichtigt.

Da dies jedoch diejenigen sind, die von den Experten und Forschern als die fruchtbarsten angesehen wurden, sind wir davon überzeugt, dass wir bereits in diesem Stadium zuverlässig optimierte Inhalte identifizieren können. Bei der Auswahl der Kriterien haben wir uns ferner auch nicht daran orientiert, ob es sich um Merkmale handelt, die auf gute oder erfolgreiche SEO-Praktiken hinweisen. Auf den ersten Blick könnte man annehmen, dass Suchmaschinenoptimierer versuchen, mit ihren Optimierungsmaßnahmen die Ranking-Faktoren von Google zu bedienen und so zu beeinflussen, dass ihre Inhalte sichtbarer werden. Dadurch wären SEO-Faktoren und Ranking-Faktoren identisch. Allerdings sind zum einen die genauen Ranking-Faktoren unbekannt und zum anderen sind auch nicht alle SEO-Bemühungen erfolgreich. Einige Methoden werden auch von den Suchmaschinen bestraft oder einfach ignoriert. Zum Beispiel könnte jemand, der versucht, Sichtbarkeit in Google zu erlangen, Keyword-Stuffing verwenden, d.h. ein Keyword auf einer Webseite sehr oft wiederholen, um der Suchmaschine zu suggerieren, dass die Seite für dieses Keyword relevant ist. Offensichtlich wird dieser Ansatz nicht funktionieren, da Suchmaschinen solche einfachen Manipulationsversuche erkennen können. Da wir jedoch feststellen wollen, ob Inhaltsanbieter versuchen, ihre Seiten zu optimieren, kann das Keyword-Stuffing immer noch ein Faktor zur Erkennung optimierter Seiten sein. Ob Inhaltsanbieter ihre Seiten erfolgreich optimiert haben, ist kein Kriterium, das für unsere Klassifizierung relevant ist. Wir nutzen verschiedene halb-automatisierte Methoden, um die Wahrscheinlichkeit von Suchmaschinenoptimierung zu bewerten.

2.1 Halbautomatische Analysemethoden zur Bewertung der SEO-Wahrscheinlichkeit auf einer Webseite

Für die abschließende Analyse der Suchmaschinenoptimierung sind drei Prozesse notwendig, die zum einen zur Berechnung der Input-Variablen für die regelbasierte Klassifikation notwendig sind und zum anderen die eigentliche Klassifikation durchführen. Abbildung 1 zeigt die Prozesse, die in folgende Schritte eingeteilt sind:

1. Abrufen der URL, um den HTML-Code und die Metadaten zu extrahieren
2. Generierung des Inputs für den regelbasierten Klassifikator auf drei Stufen und
3. Bestimmen der Wahrscheinlichkeit von SEO.

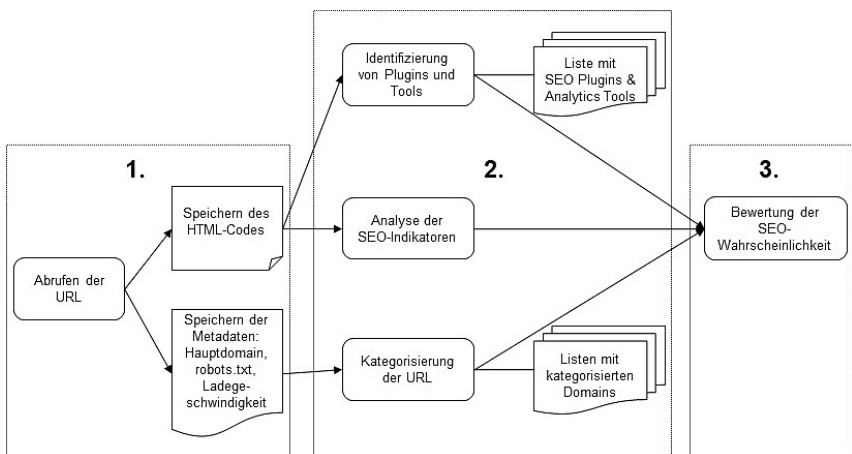


Abb. 1 Halb-automatisierte Methoden zur Bewertung der SEO-Wahrscheinlichkeit auf einer Webseite

Im ersten Schritt findet ein Scraping der Ergebnissen von Suchmaschinen statt. Scraping bezeichnet allgemein den Vorgang, durch technische Hilfsmittel automatisiert auf Texte und Webseiten zuzugreifen, um Informationen daraus zu extrahieren und für weitere Zwecke, z.B. Analysen, verfügbar zu machen. In diesem Fall werden die URLs von den Suchergebnissen von Google extrahiert und in einer Datenbank gespeichert. Dabei werden sowohl der HTML-Quelltext als auch einige zusätzliche Daten wie die Hauptdomain, der Inhalt der robots.txt und die Ladegeschwindigkeit der

URL erfasst. Anschließend werden SEO-Indikatoren sowie SEO-Plug-ins and Analytics Tools aus dem Quelltext extrahiert und die URL wird mit Listen abgeglichen, die optimierte und nicht optimierte Dokumente enthalten. Die Ergebnisse aus diesem Schritt sind die Grundlage für die abschließende Bewertung. Im Folgenden werden die einzelnen Prozesse aus dem zweiten und dritten Schritt näher erläutert.

2.2 Identifizierung von Plug-ins und Tools

Für die Identifizierung von SEO-Plug-ins und Tools wurden manuelle Listen erstellt, die sowohl die Bezeichnung als auch ein Suchmuster enthalten, um Hinweise auf die Nutzung im Quelltext der URL zu finden. Eine Nutzung von SEO-Plug-ins ist ein eindeutiger Hinweis darauf, dass Optimierungsmaßnahmen eingesetzt werden, aber auch Analytics Tools lassen zumindest darauf schließen, dass Webanalysen durchgeführt werden. Webanalysen sind ein sehr guter Indikator für ein kommerzielles Interesse des Anbieters und da insbesondere kommerzielle Anbieter Maßnahmen für gute Positionen im Suchmaschinenranking durchführen, sind Analytics Tools ebenfalls ein guter Indikator.

Bei den SEO-Plug-ins umfasst die Liste momentan 58 Plug-ins und bei den Analytics Tools wurden bisher 54 Tools manuell in einer Datenmenge von 30.000 gespeicherten Quelltexten identifiziert. Die folgenden Ausschnitte aus dem HTML-Quelltext zeigen jeweils ein Beispiel für ein SEO-Plug-in und ein Analytics Tool:

SEO-Plug-in: Yoast SEO Plugin¹

HTML-Code: <!--This site is optimized with the Yoast SEO plugin v12.4 -
https://yoast.com/wordpress/plugins/seo/-->

Suchmuster: "*yoast seo*"

Analytics Tool: Google Analytics²

HTML-Code: <!-- Google Analytics -->
<script>(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;...</script>

Suchmuster: "*google analytics*"

1 <https://yoast.com/wordpress/plugins/seo/>

2 <https://analytics.google.com/analytics/web/>

2.3 Kategorisierung der Seite anhand bekannter Domains

Für unseren Ansatz kombinieren wir die automatische Identifizierung von SEO-Indikatoren aus Webseiten und Websites mit der manuellen Generierung von Listen mit optimierten und nicht optimierten Websites. Wir verfolgen diesen Ansatz, da die Ergebnisse in den Suchmaschinen nicht gleichmäßig verteilt sind, d.h. es gibt nur eine relativ kleine Anzahl von Webseiten, die einen großen Teil der gezeigten URLs ausmachen (Petrescu, 2014) und in den Top-Ergebnissen angeklickt werden (Goel et al., 2010). Dies bedeutet, dass wir durch die manuelle Klassifizierung einer begrenzten Anzahl von Websites bereits einen relativ großen Teil der optimierten Seiten erkennen können (Petrescu, 2014). Für die Klassifizierung haben wir Listen mit Websites nach Kategorien zusammengestellt, die wir anhand der Wahrscheinlichkeit von Suchmaschinenoptimierung in den Kategorien unterteilt haben. Insgesamt wurden sechs Kategorien definiert. Die erste Kategorie ist eine Liste mit Kunden von SEO-Agenturen, da davon ausgegangen werden kann, dass Kunden von solchen Agenturen Suchmaschinenoptimierung betreiben.

Die Agenturen wurden anhand von Recherchen in Fachportalen identifiziert und die Websites für die Liste anhand der Kundenreferenzen auf den Seiten der Agenturen ermittelt. Durch diese Vorgehensweise konnten insgesamt 1.004 Websites identifiziert werden. Eine weitere Kategorie bezieht sich auf Websites, die ganz sicher keine Suchmaschinenoptimierung betreiben. Da unser Ansatz noch in der Entwicklung ist, sind Aussagen zu solchen Seiten schwer zu treffen. Daher befindet sich momentan nur Wikipedia in dieser Liste, da von Wikipedia bekannt ist, dass dort keine aktiven SEO-Maßnahmen genutzt werden. Für die weiteren Kategorien wurde ein Datensatz mit 13.403 Dokumenten manuell ausgewertet. Die URLs wurden dabei klassifiziert und in die restlichen Kategorien eingeteilt. Wir sind dabei von den möglichen Interessen der Seitenbetreiber ausgegangen. So ist definiert, dass klare kommerzielle Absichten die Wahrscheinlichkeit von SEO erhöhen. Die Kategorie, die dabei die meisten Dokumente enthält, sind Nachrichten-Websites wie beispielsweise spiegel.de. Insgesamt konnten 1.203 Nachrichtenangebote im Datensatz gefunden werden. Die weiteren Kategorien sind Online-Shops (178 Seiten), Websites, die Werbeanzeigen schalten (325 Seiten), und Unternehmens-Webseiten (72 Seiten). Alle Dokumente in den Listen sind dabei allerdings nicht exklusiv eingeordnet. Es sind Überschneidungen möglich. Die Listen werden fortlaufend erweitert und dabei ist auch insbesondere die Erweiterung der nicht optimierten Webseiten hochrelevant, da bisher nur eine

Domain in diese Liste aufgenommen wurde. Zum jetzigen Zeitpunkt ist eine ganz konkrete Einordnung solcher Seiten allerdings schwierig, da unser Klassifikator noch nicht abschließend und genau genug die SEO-Wahrscheinlichkeit bestimmen kann, um konkret auszusagen, dass eine Seite nicht optimiert ist.

2.4 Analyse der SEO-Indikatoren

Bei den SEO-Indikatoren extrahieren wir Informationen direkt aus dem HTML-Quelltext der Seiten und speichern dazu auch weitere Merkmale wie die Hauptdomain, die Inhalte der robots.txt und die Geschwindigkeit, mit der die Seite vollständig geladen wird. Bei den aus dem Quelltext gewonnenen Daten handelt es sich beispielsweise um den Seitentitel (Title Tag), die Seitenbeschreibung (Description Tag) und Nofollow-Links. Ferner wird noch geprüft, ob eine Sitemap vorhanden ist. In Bezug auf die robots.txt werden typische Inhalte gesucht, die auf SEO hinweisen. Eine robots.txt wird genutzt, um Anweisungen an Suchmaschinen-Crawler zu geben. Wir messen, ob diese vorhanden ist und SEO-relevante Anweisungen enthält (z.B. explizite Ausschlüsse von Inhalten auf der Domain für die Crawler). Schließlich messen wir die Geschwindigkeit der Seiten, da einer der grundlegenden technischen Faktoren bei der Suchmaschinenoptimierung die Optimierung der Seiten für ein schnelles Laden ist. Der Vorteil bei dieser Vorgehensweise liegt darin, dass wir unabhängig von externen Indikatoren agieren können, da wir nur auf direkt verfügbare technische Informationen zugreifen.

Durch die Kombination mit der Auswertung der Plug-ins und Tools sowie der Kategorisierung der URL durch die Abgleiche mit bekannten Domains wird eine Basis für die Einschätzung der Suchmaschinenoptimierung geschaffen. Eine vollständige Übersicht der genutzten Merkmale findet sich in Tabelle 1.

Die halb automatische Erfassung und Analyse dieser Merkmale bildet die Grundlage für die regelbasierte Klassifikation, um die Wahrscheinlichkeit von Suchmaschinenoptimierung auf der Seite zu bewerten.

Tab. 1: SEO-Merkmale

Merkmal	Beschreibung	Beispiel
SEO Plugins und Analytics Tools		
SEO Plug-ins	Tools, die direkt zur Suchmaschinenoptimierung eingesetzt werden	Yoast SEO Plugin
Analytics Tools	Tools, die für das Tracking und für die Webseitenanalyse verwendet werden	Google Analytics
URL-Listen		
Kunden von SEO-Agenturen	In dieser Liste sind Domains hinterlegt, die zu Kunden von SEO-Agenturen gehören (1.004 Seiten).	faz.net
nicht optimierte Seiten	Liste mit Domains, die garantiert keine Suchmaschinenoptimierung betreiben (1 Seite)	wikipedia.de
Nachrichtendienste	Domains von Nachrichtendiensten (1.203 Seiten).	spiegel.de
Online-Shops	Liste mit Domains von Online-Shops (178 Seiten)	amazon.de
Unternehmens-Webseiten	Liste mit Domains von Unternehmen (72 Seiten)	dhl.de
Webseiten mit Werbung	Liste mit Domains, die Werbeanzeigen schalten (325 Seiten)	facebook.com
Technische Indikatoren		
Microdata-Formate	Nutzung von Microdata-Formaten auf einer Webseite strukturiert	JSON-LD
Textanzeigen auf der Seite	Nutzung von Diensten zur automatischen Schaltung von kontextualisierten Werbeanzeigen	Google Ads
HTTPS	Verwendung des Hypertext Transfer Protocol Secure	–
SEO in der robots.txt	Hinweise zur Konfiguration von Suchmaschinen-Crawler in der robots.txt, die zu der Hauptdomain der Seite gehört	crawl-delay
Sitemap	Verwendung einer Sitemap auf der Seite für eine bessere Navigation	–
Viewport	Definition eines Viewports für eine responsive Darstellung der Webseite.	<meta name="viewport" content="width=device-width, initial-scale=1.0">

Merkmal	Beschreibung	Beispiel
Nofollow-Links	Nofollow-Links werden gesetzt, um Suchmaschinen anzuweisen, diese Links nicht für das Ranking zu berücksichtigen. Die Anweisung wird über das rel-Tag gesetzt.	<code></code>
Canonical-Links	Canonical-Links werden gesetzt, um bei mehrfach verwendeten Inhalten auf die Originalressource zu verweisen. Die Anweisung wird über das rel-Tag gesetzt.	<code></code>
Lade-geschwindigkeit	Messen der Ladegeschwindigkeit. Wenn der Wert unter drei Sekunden liegt, ist das ein Indikator für optimierte Inhalte.	<code>driver.execute_script("""" var loadTime = ((win- dow.performance.timing. domComplete - win- dow.performance.timing. navigationStart)/1000); return loadTime;) """"</code>
Description	Überprüfung, ob eine Description auf der Seite vorhanden ist. Das Tag für die Beschreibung der Seite; die Beschreibung kann dabei sowohl in den meta-Tags einer Seite und in Open Graph Tags gesetzt werden.	<code><meta name="description" content="example" /></code>
Title	Überprüfung, ob das Tag für den Titel der Seite gesetzt ist. Der Titel wird normalerweise in dem dafür vorgesehenen Tag gesetzt. Es ist aber auch möglich, dafür Open Graph Tags zu verwenden.	<code><title>example title</title></code>
Open Graph Tags	Überprüfung, ob Open Graph Tags auf der Seite verwendet werden.	<code><meta property="og:title" content="example" /></code>

2.5 Regelbasierte Klassifikation der SEO-Wahrscheinlichkeit

Durch die regelbasierte Klassifikation wird eine Aussage darüber getroffen, ob eine Dokument höchstwahrscheinlich optimiert, wahrscheinlich optimiert, wahrscheinlich nicht optimiert oder höchstwahrscheinlich nicht optimiert ist. Die Regeln werden im Folgenden näher erläutert.

2.5.1 *Höchstwahrscheinlich optimiert*

Ein Dokument ist höchstwahrscheinlich optimiert, wenn entweder mindestens ein SEO-Plug-in im Quelltext identifiziert wurde oder die Seite Kunde einer SEO-Agentur ist oder die Seite ein Nachrichtenangebot ist oder Werbeanzeigen auf der Seite geschaltet sind oder Microdata-Formate gefunden wurden.

Bei dieser Regel wurden die eindeutigsten Merkmale für SEO genutzt. So sind die Verwendung von SEO-Plug-ins und die Zusammenarbeit mit einer SEO-Agentur ganz klare Hinweise für Suchmaschinenoptimierung. Zusätzlich sind Nachrichtenangebote als Informationsangebote stark an guten Positionen in den Suchergebnissen interessiert und daher ist auch hier von einer ausgeprägten SEO-Intention auszugehen. Diese Annahme wurde durch unseren Beirat von Suchmaschinenexperten bestätigt. Zusätzlich sind eindeutige kommerzielle Absichten mit der Schaltung von Werbeanzeigen ein klarer Hinweis auf SEO und auch Microdata-Formate werden häufig dafür genutzt, sogenannte Rich Snippets, also angereicherte Trefferbeschreibungen für Suchmaschinen, zu generieren (Ronallo, 2012).

2.5.2 *Wahrscheinlich optimiert*

Die Seite ist wahrscheinlich optimiert, wenn sie nicht höchstwahrscheinlich optimiert ist und mindestens eines der folgenden Kriterien erfüllt: (1) Die Seite ist ein Online-Shop oder eine Unternehmensseite, (2) auf der Seite wurde mindestens ein Analytics Tools identifiziert, (3) als Übertragungsprotokoll wird HTTPS eingesetzt, (4) es wurden Hinweise zu SEO in der robots.txt gefunden, (5) die Seite hat eine Sitemap, (6) ein Viewport ist definiert, (7) es wurde mindestens ein Nofollow- oder ein Canonical-Link gefunden, (8) die Ladegeschwindigkeit der Seite liegt unter drei Sekunden.

Bei dieser Regel wird eine Vielzahl von Merkmalen einbezogen, die Hinweise auf eine wahrscheinliche Suchmaschinenoptimierung geben. Hier wird davon ausgegangen, dass kommerzielle Anbieter wie Online-Shops und Unternehmensseiten sowie Seiten, die Programme für kontextualisierte Werbeanzeigen wie Google Ads verwenden, ein großes Interesse an guten Positionen im Ranking von Suchmaschinen haben. Die weiteren Indikatoren, die geprüft werden, ergeben sich aus der Analyse des Quelltexts, der robots.txt und der gemessenen Ladegeschwindigkeit. Hier werden die Nutzung von HTTPS als Übertragungsprotokoll, das Angebot einer Sitemap für eine klare Navigation und als expliziter Pfad für Suchmaschinen-Crawler (Erlhofer,

2019), die Verwendung von Viewports zur Sicherstellung einer responsiven Darstellung der Inhalte sowie die Ladegeschwindigkeit berücksichtigt. Noch eindeutiger sind allerdings explizite Anweisungen für Suchmaschinencrawler in der robots.txt sowie die Definition von Nofollow- und Canonical-Links, da diese auch als direkte Anweisungen für Suchmaschinen und Crawler definiert werden (ebd.).

2.5.3 *Wahrscheinlich nicht optimiert*

Die Seite ist wahrscheinlich nicht optimiert, wenn sie nicht höchstwahrscheinlich nicht optimiert und nicht höchstwahrscheinlich optimiert ist und wenn mindestens eines der folgenden Kriterien erfüllt ist: (1) kein Description Tag, (2) kein Title Tag oder, wenn identische Title Tags auf Unterseiten sind, (3) keine Open Graph Tags definiert.

Wie bereits bei den anderen Regeln gezeigt, sind viele Merkmale als Hinweise für SEO zu deuten, auch wenn diese zum Teil als gute Praktiken für die Erstellung von Websites gelten können. Daher wird bei der Einschätzung, ob ein Dokument nicht optimiert ist, auf eine Prüfung der Mindestanforderungen für Suchmaschinenoptimierung zurückgegriffen, d.h., wenn diese nicht erfüllt sind, kann davon ausgegangen werden, dass wahrscheinlich keine Optimierungsmaßnahmen unternommen wurden. Zu diesen Mindestanforderungen zählt, dass eine Seitenbeschreibung in Form eines Description Tags gesetzt wurde. Dies gilt ebenfalls für den Seitentitel, der als Title Tag definiert wird. Bei dem Seitentitel wird allerdings ebenfalls geprüft, ob sich die Seitentitel auf Unterseiten einer Webseite unterscheiden, denn die Nutzung eines beispielsweise automatisch generierten identischen Titels zeigt, dass keine Anstrengungen unternommen wurden, diese entsprechend den Seiteninhalten anzupassen. Da Description und Title Tags auch in den sogenannten Open Graph Tags definiert werden können, wird bei dieser Regel auch geprüft, ob solche Tags gefunden wurden. Open Graph Tags werden von Social-Media-Angeboten wie Facebook und Twitter unterstützt und dienen einer strukturierten Vorschau darstellung der Seiteninhalte in diesen sozialen Medien (Krrabaj et al., 2017).

2.5.4 *Höchstwahrscheinlich nicht optimiert*

Die Domain der Seite ist auf der Liste mit nicht optimierten URLs.

Bei dieser Regel wird bisher nur geprüft, ob das Suchergebnis ein Wikipedia-Artikel ist. Wenn unsere weiteren Auswertungen eindeutiger Hinweise auf nicht optimierte Dokumente liefern, wird diese Liste um die Domains dieser Dokumente erweitert.

2.6 Technische Umsetzung der Methoden in einem Software-Tool

Für die Anwendung der Ansätze und Methoden wurde ein Software-Tool in Python programmiert, das sich aus verschiedenen Programmbibliotheken zusammensetzt, die für die Datenerfassung und Datenanalyse gut geeignet sind.

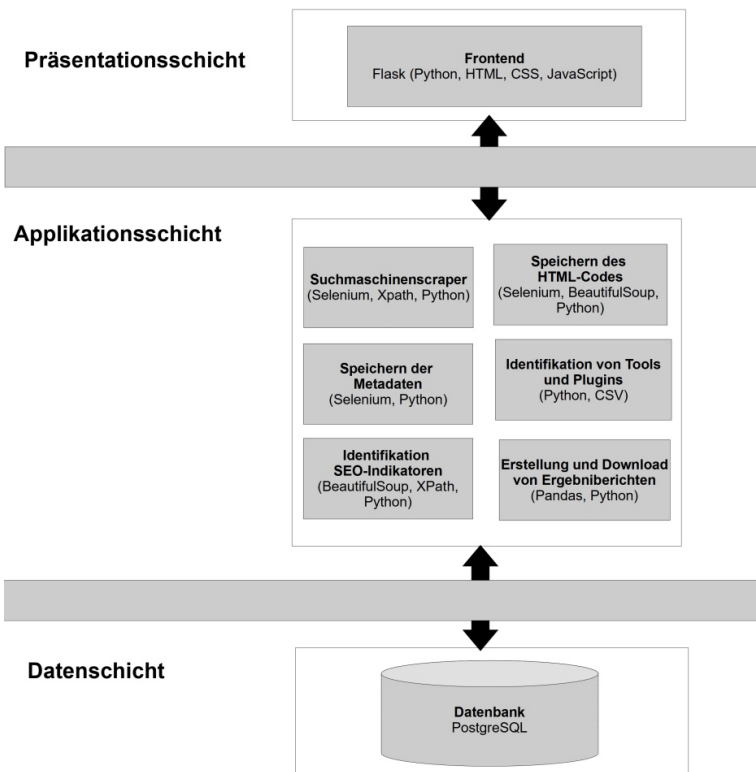


Abb. 2 MVC-Architektur der Software

Sämtliche Ergebnisse des Tools werden in einer PostgreSQL-Datenbank³ gespeichert. Für eine Strukturierung der Software wurde auf die MVC-Architektur zurückgegriffen, die durch die Aufteilung der Präsentations-, Applikations- und Datenschicht eine bessere Wartbarkeit der Software ermöglicht. Die Datenbank ist dabei als zentraler Speicher auf einem Datenbankserver installiert, während die Applikationen verteilt arbeiten. Damit ist es möglich, sämtliche Module für die Speicherung und Analyse verteilt auf mehreren Servern zu installieren, um damit die Performance zu erhöhen. Abbildung 2 zeigt die Software-Architektur des Tools mit den Schichten und den jeweiligen Modulen. Im Folgenden werden die Module näher erläutert.

2.6.1 Präsentationsschicht

Das Frontend der Software basiert auf dem Web-Framework Flask⁴, da dieses sämtliche Funktionen anbietet, die für die Webentwicklung notwendig sind. So lassen sich mit geringem Aufwand Formulare erstellen und über HTML-Templates abbilden. Das Frontend in der Anwendung dient dazu, neue Datensätze zu generieren und die Datenanalysen zu steuern. Der Nut-zende kann in dem Frontend Namen für Studien anlegen und dazu Suchanfragen definieren. Sämtliche weitere Prozesse werden dann anschließend automatisiert auf der Applikationsschicht durchgeführt.

2.6.2 Applikationsschicht

Auf dieser Schicht werden sämtliche Prozesse durchgeführt, die zu Beginn dieses Abschnitts näher erläutert wurden und in Abbildung 1 dargestellt sind. Eine weitere Hauptanwendung in dieser Schicht ist der Suchmaschinen-scrapers, der auf Selenium⁵ (portables Framework zum Testen von Webanwendungen) und BeautifulSoup⁶ (Paket zum Parsen von HTML und XML) basiert. Mit dem Scraper werden vorab definierte Suchanfragen automatisiert an Google gesendet und alle Suchergebnisse von allen Suchergebnisseiten (von der ersten Seite bis zur letzten Seite) gespeichert. Die URLs der Such-

3 <https://www.postgresql.org/>

4 <https://flask.palletsprojects.com/>

5 <https://www.selenium.dev/>

6 <https://www.crummy.com/software/BeautifulSoup/>

ergebnisse werden mithilfe von XPath extrahiert und anschließend werden der Quelltext, die Trefferposition, die Domain, die robots.txt und die Ladegeschwindigkeit jedes einzelnen Suchtreffers erfasst und in Relation mit der Suchanfrage in der Datenbank gespeichert. Alternativ zum Suchmaschinen-scrapers lassen sich auch Listen mit URLs für eine Analyse importieren. Die URLs werden über eine eindeutige ID in der Datenbank gespeichert.

Die Software prüft automatisch, ob neue URLs in der Datenbank erfasst wurden, und falls noch keine Auswertung für die SEO-Indikatoren zu der URL durchgeführt wurde, werden diese ausgeführt. Dabei werden in dem Modul für die Analyse der SEO-Indikatoren die technischen Merkmale über verschiedene Prüfalgorithmen aus dem Quelltext extrahiert und gespeichert. Für die Erfassung der SEO-Plug-ins und Analytics Tools werden die jeweiligen Listen mit den Suchmustern für jedes Plug-in geladen und abgeglichen. Für die Kategorisierung des Dokuments wird die Hauptdomain mit den hinterlegten URL-Listen verglichen und, falls ein Treffer vorliegt, in der Datenbank gespeichert. Sind alle Informationen zu einer URL gespeichert, wird die Wahrscheinlichkeit von SEO in diesem Dokument über den regelbasierten Klassifikator ermittelt und ebenfalls gespeichert. Mit dem Berichtsmodul können zu jeder Zeit Ergebnisberichte zu den gespeicherten Datensätzen erzeugt und als CSV-Datei heruntergeladen werden. Dabei werden sowohl die Rohdaten zur Verfügung gestellt als auch Zusammenfassungen zu den jeweiligen Datensätzen mit deskriptiven Kennzahlen zur SEO-Wahrscheinlichkeit.

Alle Module in dieser Schicht kommunizieren sowohl mit der Datenschicht, die bei dieser Anwendung aus einer großen PostgreSQL-Datenbank besteht, als auch mit der Applikationsschicht, auf der die Anwendung gestartet und gesteuert wird.

2.6.3 Datenschicht

Die Datenschicht der Anwendung ist eine PostgreSQL-Datenbank, die insgesamt aus sieben Tabellen besteht. Dort werden die Studien, die Suchanfragen, eine Jobverwaltung für die Scraper, die Suchergebnisse, die Quelltexte und die SEO-Merkmale gespeichert.

2.6.4 Demo-Tool

Ein Teil der Software ist ein Demo-Tool, das genutzt werden kann, um eine Analyse gezielt für eine gewünschte URL durchzuführen. Dadurch lassen sich schnelle Tests für den regelbasierten Klassifikator durchführen, um die Ansätze weiterzuentwickeln. Das Tool kann direkt im Web aufgerufen und für eigene Tests genutzt werden. Abbildung 3 zeigt einen Ergebnisbericht, der direkt in der Webanwendung generiert wird.

	Most probably optimized	Probably optimized	Most probably not optimized	Probably not optimized	Uncertain
SEO Assessment					
Tools & Plugins					
SEO Tools		✗			
Analytics Tools		✓		Google Tag Manager, GoogleTagManager tracker	
URL Category					
Not optimized		✓		Website is definetly not optimized	
Customer of a SEO Agency		✗		Website is a customer of a SEO agency	
News Service		✓		Website is a news service	
Website with ads		✗		Website has online advertisement	
Business Website		✗		Website is a business website	
Online Shop		✗		Website is an online shop	
Indicators for SEO					
Description		✓		auf stern.de finden sie news spannende hintergründe sowie bildstarke reportagen aus allen bereichen von politik und wirtschaft bis kultur und wissenschaft.	
Title		✓		nachrichten hintergründe & reportagen	
Identical Title tags		✗		No identical title tags on subpages	
Loading speed		✗		Loading speed is 4.397s > 3s	
Hypertext Transfer Secure (https)		✓		Page uses https	
SEO in robots.txt		✓		SEO in robots.txt found	
Viewport		✓		Viewport defined	
Microdata		✓		Microdata definitions found	
nofollow-Links		✗		0 nofollow-links found	
canonical-Links		✗		0 canonical-links found	

Abb. 3 Demo-Tool

In dem Bericht werden die Ergebnisse aus allen Analysen in dazugehörigen Kategorien ausgewiesen. So wird am Anfang die Einschätzung zur SEO-Wahrscheinlichkeit angezeigt. Darauf folgt die Auswertung zu den SEO-Plug-ins und Analytics Tools, die Kategorisierung der URL und darauffolgend werden die technischen SEO-Indikatoren erläutert. Der Bericht kann ebenfalls als CSV-Datei für weitere Verarbeitungen heruntergeladen werden.

3 Studien zur Überprüfung von Suchmaschinenoptimierung in kommerziellen Suchergebnissen

Für die praktische Anwendung unseres Ansatzes haben wir bisher drei Studien zu verschiedenen Themenbereichen durchgeführt, um zum einen unsere Vorgehensweise zu überprüfen, aber auch um erste Ergebnisse zu der Anzahl von Dokumenten in Suchergebnissen zu erhalten, die durch Suchmaschinenoptimierung beeinflusst sind. Insgesamt wurden 263.790 Suchergebnisse aus drei Datensätzen analysiert. Bei der Erhebung wurden dafür jeweils alle Suchtreffer berücksichtigt, die von Google zu einer Suchanfrage zurückgegeben wurden. Dabei sind dies in der Regel nicht mehr als 300 Ergebnisse je Anfrage. Die Datensätze werden im Folgenden näher erläutert und anschließend die Ergebnisse der Studien präsentiert.

3.1 Datensätze

Für die Anwendung unserer Methoden wurden drei sehr verschiedene Datensätze definiert, die sich thematisch stark voneinander unterscheiden. So wurden beispielsweise die populärsten Suchanfragen bei Google durch die Ausweisung in Google Trends über einen definierten Zeitraum gesammelt, um eine vielfältige Sammlung von Suchergebnissen zu generieren, die thematisch nicht festgelegt sind. Die weiteren Datensätze sind dagegen themenspezifisch und weisen eine stärkere Heterogenität auf. So konnten durch die Zusammenarbeit mit der Medienanstalt Hamburg Schleswig-Holstein Suchanfragen generiert werden, die potenziell strafrechtlich relevante Dokumente in Bezug auf rechtsradikale Inhalte auffinden. Der letzte Datensatz wurde

mithilfe von deutschen Suchanfragen generiert, die sich auf die Corona-Pandemie beziehen.

3.1.1 *Google Trends*

Bei diesem Datensatz wurden 244.985 Suchergebnisse durch 1.478 Suchanfragen erfasst. Die Suchanfragen wurden dabei in dem Zeitraum von März bis Juni 2020 direkt von der Google-Trends-Seite⁷ entnommen.

Dabei wurde deutlich, dass sich die Suchanfragen in der Regel immer wieder auf ähnlich populäre Themen über alle Monate hinweg beziehen. Dazu zählen durch die Corona-Pandemie Suchanfragen zu Corona, zu Prominenten, zu Fernsehsendungen und Anfragen zu Sport, z. B. zur Fußball-Bundesliga. Dieser Datensatz wurde erstellt, um ein möglichst großes und vielfältiges Set von Suchergebnissen zu generieren.

3.1.2 *Rechtsradikale Inhalte*

Dieser Datensatz ist durch eine Kooperation mit den Landesmedienanstalten Hamburg Schleswig Holstein entstanden. Eine Aufgabe einer Medienanstalt ist die Prüfung strafrechtlich relevanter Inhalte im Web. Daher werden dort spezialisierte und ständige angepasste Suchanfragen genutzt, um solche Inhalte zu finden. Durch den Zugriff auf diese Anfragen wurden insgesamt 82 Suchanfragen im März 2020 automatisiert an Google geschickt und insgesamt 13.403 Suchergebnisse mit potenziell rechtsradikalen Inhalten gespeichert und durch unsere Software analysiert. Die Auswahl dieses Themas für einen Datensatz folgte dabei der Frage, ob Betreiber solcher Angebote Suchmaschinenoptimierung betreiben. Die Annahme war, dass zu solchen Themen häufiger Privatangebote zu finden sind, die weniger professionalisiert im Online-Marketing agieren und daher auch kein SEO einsetzen.

3.1.3 *Corona*

Microsoft Bing stellt Suchanfragen zur Verfügung, die als Suchintention zur Corona-Pandemie passen.⁸ Dafür haben wir 483 Top-Anfragen für Deutschland aus dem Datensatz von Microsoft Bing im September 2020 extrahiert

7 <https://trends.google.de/trends/?geo=DE>

8 <https://github.com/microsoft/BingCoronavirusQuerySet>

und 5.402 Suchergebnisse gespeichert. Dieser Datensatz wurde generiert, da die Corona-Pandemie ein hochaktuelles Thema ist und wir auf tatsächlich genutzten Suchanfragen von Nutzer*innen zugreifen konnten.

3.2 Auswertung der Datensätze

Im Folgenden werden die Ergebnisse der Analyse der SEO-Wahrscheinlichkeit in den Dokumenten aus den Datensätzen näher erläutert und anschließend diskutiert.

3.2.1 Ergebnisse

Eine Analyse der Datensätze zeigt, dass die überwiegende Mehrheit der Ergebnisse entweder höchstwahrscheinlich oder zumindest wahrscheinlich optimiert ist. In Abbildung 3 sind die jeweiligen Ergebnisse zu sehen. Je nach Datensatz können wir sehen, dass zwischen 41,9% und 62% der gefundenen Ergebnisse als höchstwahrscheinlich optimiert eingestuft werden.

Tab. 2: Verteilung der Domains (Top-10)

Domain	Anzahl der Dokumente mit dieser Domain in allen Datensätzen	Anteil der Dokumente in allen Datensätzen in %
books.google.de	5.724	12,13
de.wikipedia.org	2.442	5,18
youtube.com	2.122	4,50
t-online.de	2.075	4,40
focus.de	1.915	4,06
welt.de	1.814	3,85
spiegel.de	1.713	3,63
sueddeutsche.de	1.699	3,60
stern.de	1.622	3,44
rtl.de	1.535	3,25

Die Unterschiede zwischen den Datensätzen lassen sich auf den höheren Anteil von Nachrichteninhalten im Trends- und Corona-Datensatz im Vergleich zum rechtsradikalen Datensatz zurückführen. So waren im Google-Trends-Datensatz 56% der Ergebnisse Nachrichtenangebote (136.806 Dokumente), im Corona-Datensatz 45% (2.454 Dokumente) und im Datensatz mit

den potenziell rechtsradikalen Inhalten nur 34%. Insgesamt konnten 47.177 eindeutige Domains über alle Datensätzen hinweg bei 263.790 Dokumenten festgestellt werden. Eine Auswertung der Top 10 zeigt, dass insbesondere Angebote von Google sehr häufig in den Dokumenten zu finden sind (books.google.de und youtube.com). Wikipedia ist auf dem zweiten Platz. Alle weiteren Angebote sind Nachrichtenangebote.

Eine weitere Analyse in Bezug auf genutzte SEO Plug-ins und Analytics Tools ergibt eine ähnliche Verteilung in allen Datensätzen. Tabelle 3 zeigt die Verteilung.

Tab. 3: Verteilung der SEO-Plug-ins und Analytics Tools

Datensatz	Anzahl an Dokumenten mit SEO-Plug-ins	Anzahl an Dokumenten mit Analytics Tools
Google Trends (244.985 Dokumente)	8% (19.515 Dokumente)	36% (87.175 Dokumente)
Potenziell rechtsradikale Inhalte (13.403 Dokumente)	8% (1.206 Dokumente)	36% (4.765 Dokumente)
Corona (5.402 Dokumente)	6% (332 Dokumente)	33% (1.793 Dokumente)

Es ist deutlich, dass die Anzahl an SEO-Plug-ins über alle Datensätze hinweg nicht sehr stark ausgeprägt ist. Das ist darauf zurückzuführen, dass diese in der Regel bei Webangeboten eingesetzt werden, die auf dem CMS Wordpress basieren und dies in der Regel nicht bei professionellen Nachrichtenangeboten verwendet wird. Eine Auswertung der Analytics Tools zeigt allerdings, dass die meisten Angebote Webanalysen für ihre Angebote durchführen, was auf kommerzielle Absichten hindeutet.

Die Auswertung der technischen Merkmale zeigt, dass 96% aller Dokumente HTTPS einsetzt (252.519 Dokumente) und Open Graph Tags sogar bei 98% (257.396 Dokumente) gesetzt werden. Das zeigt beispielsweise, wie hoch die Relevanz von Social Media für die Webseitenbetreiber ist, da fast alle Anbieter diese speziellen Tags für eine optimierte Darstellung einer Vorschau ihrer Inhalte definieren.

Abbildung 4 zeigt die Auswertung der SEO-Wahrscheinlichkeit in den Datensätzen. Dabei zeigt sich, dass nur ein kleiner Teil der Ergebnisse als höchstwahrscheinlich nicht optimiert eingestuft wurde (1,6% über alle Datensätze). Dies sind alles Ergebnisse von Wikipedia, da dies die einzige

Website auf unserer Liste der definitiv nicht optimierten Seiten ist. Ein relativ kleiner Teil der Seiten (zwischen 3,5 und 7%) ist wahrscheinlich nicht optimiert. Für eine bessere Darstellung wurden die Ergebnisse aus höchstwahrscheinlich nicht optimiert und wahrscheinlich nicht optimiert zusammengefasst. Wir fanden auch eine leichte Überlappung von bis zu 3% über alle Datensätze hinweg. Die Überlappung lässt sich dadurch erklären, dass die Ergebnisse von unserem Klassifikator sowohl als wahrscheinlich optimiert als auch als wahrscheinlich nicht optimiert eingestuft wurden. Zusammenfassend haben wir festgestellt, dass ein großer Teil der in Google gefundenen Ergebnisse entweder definitiv oder wahrscheinlich optimiert ist. Über 90% der gefundenen Ergebnisse gehören zu diesen Kategorien.

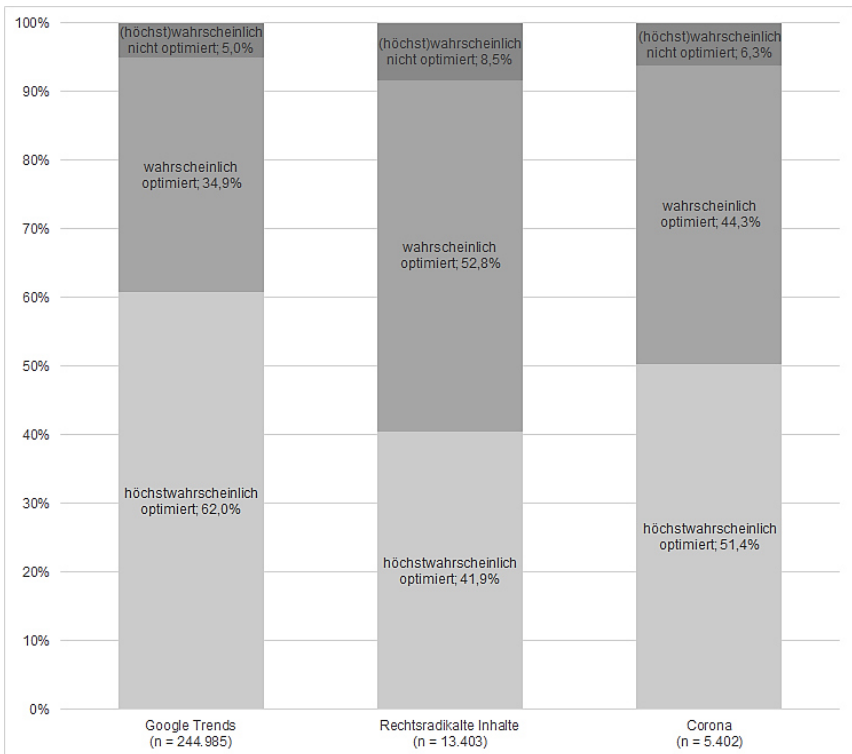


Abb. 4 Verteilung der SEO-Wahrscheinlichkeit

3.2.2 Diskussion

Die ersten Studien zur Ermittlung der Wahrscheinlichkeit der Suchmaschinenoptimierung zeigen, dass davon auszugehen ist, dass optimierte Inhalte die Suchergebnisse stark dominieren. So sind nur bis zu 10% nicht oder wahrscheinlich nicht optimiert. Dies ist keine Überraschung, da wir wissen, dass SEO ein Multimilliarden-Dollar-Geschäft ist und Unternehmen und andere Akteure oft von der Sichtbarkeit abhängig sind, die ihre Websites durch Traffic von Websuchmaschinen erhalten. Damit kann die Annahme bestätigt werden, dass Suchmaschinenoptimierung einen starken Einfluss auf die Suchergebnisse hat – auch unabhängig von den Themenbereichen, in denen recherchiert wird.

So unterscheiden sich die Datensätze inhaltlich stark voneinander. Während in dem Ergebnis-Set zu den Google Trends ein breites Themenspektrum nach der Popularität von Suchanfragen abgedeckt wird, sind die Ergebnisse zu den potenziell rechtsradikalen Inhalten und zu dem Coronavirus sehr spezifisch und weniger kommerziell ausgerichtet. Allerdings lässt sich eine Aussage darüber, ob der Einfluss der Suchmaschinenoptimierung sich positiv oder negativ auf die Ergebnisqualität auswirkt, auch in Bezug auf einzelne Suchanfragen oder die Gesamtheit der Suchanfragen mit den durchgeführten Analysen nicht beantworten. Für solche Analysen müsste die Methodik beispielsweise durch Retrievalstudien, in denen Juroren die Qualität der Suchergebnisse bewerten, ergänzt werden.

Ferner ist es auch notwendig, den bisherigen regelbasierten Ansatz weiter auszudifferenzieren, da bisher nur eine beschränkte Menge von Merkmalen in die Analyse einfließt. Dabei sind insbesondere die Prüfung auf den Einsatz von SEO-Plug-ins, was ein sehr starker Indikator für Suchmaschinenoptimierung ist, und die Überprüfung auf fehlende Trefferbeschreibungen (Description-Tag), was als starker Indikator für fehlende Suchmaschinenoptimierung ist, besonders aussagekräftig Indikatoren. Die weiteren Merkmale, die wir bisher betrachten, sind zum Teil ebenfalls gut für die Einschätzung von SEO geeignet, doch fehlt es an einer klaren Gewichtung – auch, um die Aussagekraft besser einschätzen zu können. Weitere Merkmale, die sich beispielsweise auf die Keywords in den Suchanfragen beziehen, oder auch sogenannte Offsite-Faktoren wie die Anzahl von Verlinkungen auf eine Webseite, fließen bisher nicht in das Modell ein.

In Bezug auf die Kategorisierung der Webseiten nach eher optimierten und nicht optimierten Sites kann die Identifikation eines Seitenbetreibers als

Kunde einer SEO-Agentur als sicherer Indikator angenommen werden. Zusätzlich ist eine hohe Suchmaschinenoptimierung bei einem Nachrichtenangebot anzunehmen. Dies zeigt auch die nähere Auswertung der Häufigkeiten in Bezug auf die Domains in den Datensätzen. Es ist davon auszugehen, dass besonders häufige aufgetretene Domains auch optimiert sind, und da insbesondere Nachrichtenangebote in der Quellenverteilung auftauchen, kann angenommen werden, dass diese Art von Websites auch SEO-Maßnahmen nutzt.

In der Auswertung der Verteilung der häufigsten Quellen in den Datensätzen waren in sieben von zehn Fällen prominente Nachrichtenanbieter mit einem Anteil von ca. 3% an der Gesamtheit aller Suchergebnisse zu finden. Dies lässt sich zum einen auf die überwiegend informationsorientierten Suchanfragen für alle Datensätze zurückführen, zum anderen erklären wir uns diese Dominanz von Nachrichtendiensten aber auch damit, dass die Suchmaschinen verlässliche Quellen in den Suchergebnissen bevorzugen, um beispielsweise die Gefahr von Falschinformationen und Fake News zu minimieren (Wingfield et al., 2016). Durch diese Vorgehensweise können allerdings kleinere und von großen Medienunternehmen unabhängige Informationsangebote benachteiligt werden. Eine weitere Benachteiligung für Inhaltsanbieter besteht auch dadurch, dass Eigenangebote von Suchmaschinentreibern sehr prominent in den Suchergebnissen aufgeführt werden. So zeigt unsere Analyse bei der Verteilung der Domains, dass eigene Angebote von Google (Google Books und YouTube) stark vertreten sind. Diese Art der Benachteiligung führt bereits seit mehreren Jahren zu rechtlichen Auseinandersetzungen (Edelman, 2015).

Die Auswertung aller genannten Aspekte zeigt, wie komplex der Einfluss verschiedener Faktoren auf die Anzeige von Dokumenten in den Suchergebnissen ist. So zeigen unsere Auswertungen bereits, dass neben dem Einfluss der Suchmaschinenoptimierung auf die Ergebnisse auch Aspekte wie die Diversität von Suchergebnissen und die Eigenangebote der Suchmaschinenbetreiber relevant sind.

4 Fazit und weiteres Vorgehen

Der Einfluss von Suchmaschinenoptimierung auf die Suchergebnisse in kommerziellen Suchmaschinen ist bisher kaum erforscht. Mit unserem halb-

automatisierten Verfahren und einer Umsetzung in einem Software-Tool konnten wir erste Auswertungen dazu vornehmen, wie groß der Anteil von optimierten Inhalten in Suchergebnissen unabhängig von Themenschwerpunkten ist. Die Vorgehensweise enthält dabei bewährte Ansätze, um Suchergebnisse automatisiert zu speichern, relevante Merkmale zu extrahieren und eine Kategorisierung der URLs vorzunehmen, um den Seitentyp nach Kategorien, aufgeteilt nach wahrscheinlich und nicht wahrscheinlich optimierten Seiten, aufzuteilen. Für diese Analysen haben wir Merkmale identifiziert, die verlässliche Aussagen über die Wahrscheinlichkeit von SEO auf Webseiten ermöglichen. Allerdings ist dieses Modell noch nicht vollständig und wird durch weitere Merkmale, Faktorenanalysen, Gewichtungen von Merkmalen und Merkmalsgruppen sowie Analysen mit Machine-Learning-Methoden aus dem unüberwachten Lernen (z. B. durch Clustering) erweitert. Dabei wird beispielsweise auch mit semi-überwachten Methoden aus der automatischen Klassifikation gearbeitet und mit Regressionsanalysen soll ein konkreter Wert für die SEO-Wahrscheinlichkeit berechnet werden. Ferner wird auch die Trefferpositionen im Ranking der Suchmaschinen näher betrachtet, um die Zusammenhänge von SEO und den Positionen zu untersuchen.

Mit der Durchführung weiterer Studien werden wir unsere Methodik weiterentwickeln und den Effekt der Suchmaschinenoptimierung weiter untersuchen. Dabei werden wir auch noch stärker auf die Diversität von Quellen in den Suchergebnissen und auf die Verteilung von Eigenangeboten der Suchmaschinenbetreiber eingehen, da sich bereits jetzt gezeigt hat, dass diese Faktoren ebenfalls einen Einfluss auf die Dokumentenauswahl für die Suchergebnisse und auf die Darstellung der Suchergebnisse haben.

Forschungsdaten

Der Quelltext des Tools kann über die OSF-Plattform (dx.doi.org/10.17605/OSF.IO/ETZHD) abgerufen werden. Die Suchergebnisse für die Auswertung sind ebenfalls auf der OSF-Plattform unter dx.doi.org/10.17605/OSF.IO/RVX54 abrufbar.

Förderung

Das Projekt „SEO-Effekt“, aus dem dieses Tool hervorgeht, wird von der Deutschen Forschungsgemeinschaft (DFG) unter der Projektnummer 417552432 gefördert.

Literaturverzeichnis

- Edelman, B. (2015): Does Google Leverage Market Power Through Tying and Bundling? *Competition Law & Economics* 11 (2), 365–400. <https://doi.org/10.1093/joclec/nhv016>
- Enge, E.; Spencer, S.; Stricchiola, J. (2015): *The Art of SEO: Mastering Search Engine Optimization*. Sebastopol, CA: O'Reilly.
- Erlhofer, S. (2019): *Suchmaschinen-Optimierung: Das umfassende Handbuch*. Bonn: Rheinwerk Verlag.
- Goel, S.; Broder, A.; Gabrilovich, E.; Pang, B.(2010): Anatomy of the long tail. In: Davison, B. D.; Suel, T.; Craswell, N.; Liu, B. (Hrsg.): *Proceedings of the Third ACM International Conference on Web Search and Data Mining - WSDM '10*. New York, NY: ACM Press.
- Griesbaum, J. (2013): Online-Marketing. In: R. Kuhlen, W. Semar, D. Strauch (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. Berlin, Boston, MA: de Gruyter Saur. https://doi.org/10.1515/9783110258264_411
- Krrabaj, S.; Baxhaku, F.; Sadrijaj, D. (2017): Investigating search engine optimization techniques for effective ranking. A case study of an educational site. In: *6th Mediterranean Conference on Embedded Computing (MECO), Bar, 2017*. IEEE. DOI: <https://doi.org/10.1109/meco.2017.7977137>
- Lewandowski, D. (2017): Users' Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice* 5 (4), 6–25. <https://doi.org/10.1633/JISTaP.2017.5.4.1>
- Lewandowski, D.; Kerkmann, F.; Rümmele, S.; Sünkler, S. (2018): An empirical investigation on search engine ad disclosure. *Journal of the Association for Information Science and Technology* 69 (3), 420–437. DOI: <https://doi.org/10.1002/asi.23963>
- Lewandowski, D.; Sünkler, S (2013): Representative online study to evaluate the revised commitments proposed by Google on 21 October 2013 as part of EU competition investigation AT.39740-Google: Country comparison report. http://searchstudies.org/wp-content/uploads/2015/10/Google_Country_Comparison_Report.pdf
- Li, K.; Lin, M.; Lin, Z.; Xing, B. (2014): Running and Chasing – The Competition between Paid Search Marketing and Search Engine Optimization. In: *47th Hawaii International Conference on System Sciences, Waikoloa, HI*. IEEE, S. 3110–3119. <https://doi.org/10.1109/HICSS.2014.640>

- McCue, T. (2018): SEO Industry Approaching \$80 Billion But All You Want Is More Web Traffic. *Forbes*. <https://www.forbes.com/sites/tjmccue/2018/07/30/seo-industry-approaching-80-billion-but-all-you-want-is-more-web-traffic/>
- Petrescu, P. (2014): Google Organic Click-Through Rates in 2014. <https://moz.com/blog/google-organic-click-through-rates-in-2014>
- Ronallo, J. (2012): HTML5 Microdata and Schema.org. *Code4Lib Journal* (16). <https://journal.code4lib.org/articles/6400>
- Schultheiß, S.; Lewandowski, D. (2020): “Outside the industry, nobody knows what we do”. SEO as seen by search engine optimizers and content providers. *Journal of Documentation* 77 (2), 542–557. <https://doi.org/10.1108/JD-07-2020-0127>
- StatCounter (2020a): StatCounter: Search Engine Market Share United States Of America | StatCounter Global Stats.
- StatCounter (2020b): StatCounter: Search Engine Market Share Europe | StatCounter Global Stats.
- Van Couvering, E. (2009). *Search Engine Bias. The Structuration of Traffic on the World-Wide Web*. Diss., London School of Economics and Political Science. <http://theses.lse.ac.uk/41/>
- Wang, X.; Bendersky, M.; Metzler, D.; Najork, M. (2016): Learning to rank with selection bias in personal search. In: *SIGIR '16: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: ACM, S. 115–124. <https://doi.org/10.1145/2911451.2911537>
- Wingfiel, N.; Isaas, M.; Benner, K. (2016): Google and Facebook take aim at fake news sites. *The New York Times*. <https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html>
- Zamani, H.; Bendersky, M.; Wang, X.; Zhang, M. (2017): Situational context for ranking in personal search. In: *WWW '17: Proceedings of the 26th International Conference on World Wide Web*. Geneva: International World Wide Web Conferences Steering Committee, S. 1531–1540. <https://doi.org/10.1145/3038912.3052648>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 273–298. DOI: doi.org/10.5283/epub.44948.

Ist die Webseite suchmaschinenoptimiert?

Vorstellung eines Online-Tools zur Analyse der Wahrscheinlichkeit der Suchmaschinenoptimierung auf einer Webseite

Sebastian Sünkler, Dirk Lewandowski

Hamburg University of Applied Sciences, Department of Information,
Finkenau 35, 22081 Hamburg (Germany)

{[sebastian.suenkler](mailto:sebastian.suenkler@haw-hamburg.de), [dirk.lewandowski](mailto:dirk.lewandowski@haw-hamburg.de)}@haw-hamburg.de

Abstract

Das SEO-Tool ist eine Webanwendung, die die Wahrscheinlichkeit von Suchmaschinenoptimierung (SEO) auf einer Webseite ermittelt. Für die Berechnung der Wahrscheinlichkeit werden insgesamt 20 Merkmale halb-automatisch erhoben und in drei Prozessen ausgewertet. Dafür analysiert das Tool zuerst den Quelltext der URL auf Informationen über die Verwendung von SEO-Plugins und Analytics Tools. Zweitens bestimmt es die Kategorie der gegebenen URL anhand manuell klassifizierter Websites und schließlich berechnet es verschiedene technische und inhaltliche SEO-Indikatoren. Die Ergebnisse aus diesen Prozessen bilden anschließend die Basis für die Einordnung der URL anhand eines regelbasierten Klassifikators. Die Demo des Tools ist unter <http://5.189.155.20:5000/> verfügbar.

Keywords: Software demonstration; Suchmaschinen; Suchmaschinenoptimierung; SEO; Datenanalyse

1 Einleitung

Inhaltsanbieter im Web sind auf Top-Positionen im Ranking in kommerziellen Suchmaschinen wie Google angewiesen, da sie dadurch einen großen Anteil ihres Website-Traffics generieren.

Eine wichtige Möglichkeit, um gute Platzierungen zu erhalten, ist die Suchmaschinenoptimierung (*search engine optimization*, SEO) (Li et al., 2014). Durch die hohe Relevanz von SEO für die Sichtbarkeit der Anbieter, die erheblichen Mittel, die in SEO-Maßnahmen fließen (McCue, 2018), sowie nach Aussagen von Expert/innen (Schultheiß/Lewandowski, 2020) kann davon ausgegangen werden, dass suchmaschinenoptimierte Inhalte einen erheblichen Einfluss auf die Suchergebnisseiten haben. Dieser Einfluss ist bisher kaum erforscht. Mit halb-automatisierten Prozessen und einem regelbasierten Ansatz in der Umsetzung in einem Software-Framework soll dieser Effekt erforscht werden. Ein Teil dieses Frameworks ist ein Online-Tool zur Klassifizierung der Wahrscheinlichkeit von SEO auf einer bestimmten Webseite, um schnelle Tests durchzuführen. Die Demo zu dem Tool ist unter <http://5.189.155.20:5000/> verfügbar.

2 Prozesse zur Identifikation von SEO auf einer Webseite

Die Identifikation von SEO auf einer Webseite basiert zum einen auf relevanten Indikatoren für Suchmaschinenoptimierung und auf einer Klassifikation der URL anhand eines Abgleichs mit Websites, die in Kategorien eingeteilt sind, die auf optimierte und nicht optimierte Websites hinweisen. Insgesamt werden 20 Merkmale geprüft. Der Prozessablauf ist in Abbildung 1 dargestellt und kann in drei Schritte unterteilt werden:

1. Abrufen der URL, um den HTML-Code und die Metadaten zu extrahieren,
2. Generierung des Inputs für den regelbasierten Klassifikator auf drei Stufen,
3. Bestimmen der Wahrscheinlichkeit von SEO.

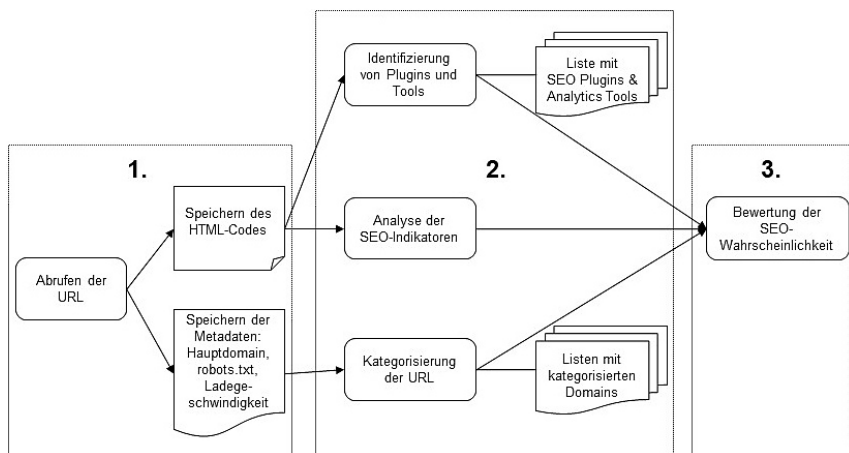


Abb. 1 Prozesse für die Bestimmung von SEO auf einer Webseite

2.1 Abrufen der URL und Speichern der Metadaten und des HTML-Codes

Im ersten Schritt wird eine eingegebene URL heruntergeladen und ihr Quelltext gespeichert. Aus dem Quelltext werden relevante Merkmale extrahiert, die Hinweise auf eine mögliche Suchmaschinenoptimierung geben. Zu diesen Informationen werden noch die Ladegeschwindigkeit der Seite und eine Kopie der robots.txt der Website erfasst.

2.2 Identifizierung von SEO-Plug-ins und Analytics Tools

In diesem Prozess wird der HTML-Code auf Hinweise für SEO-Plug-ins und Analytics Tools überprüft. Dafür werden manuell zusammengestellte Listen genutzt, die den Namen der Tools und ein Suchmuster enthalten.

2.3 Kategorisierung der URL

Die Kategorisierung der URL erfolgt anhand von vorab definierten Listen, die sich durch ihren Ergebnistyp und der Wahrscheinlichkeit von SEO auf diesen Typen voneinander abgrenzen. Diese Listen werden fortlaufend erweitert. Die Festlegung der Kategorien folgt der Annahme, dass insbesondere Seiten mit kommerziellen Absichten und große informationelle Angebote

wie Nachrichtendienste Suchmaschinenoptimierung einsetzen. Die Auswahl und Zusammenstellung der Seiten für die jeweiligen Kategorien basiert dabei auf den Top-Quellen in Suchergebnissen und auf der Identifizierung von SEO-Agenturkunden, die anhand von Kundenlisten dieser Agenturen erfolgte. Insgesamt erfolgte dieser Ansatz anhand der Annahme, dass die Diversität an Quellen in Top-Suchergebnissen relativ gering ist (Goel et al., 2010) und dazu die meisten Klicks nur auf eine geringe Anzahl von Domains erfolgen (Petrescu, 2014). Durch diese Vorgehensweise lässt sich die Vielzahl an Domains gut abdecken. Die Domains sind in folgende Kategorien eingeteilt und haben dabei den jeweils genannten Umfang an Seiten:

- nicht optimiert (1 Seite)
- Kunde von SEO-Agenturen (1.004 Seiten)
- Nachrichtendienst (1.203 Seiten)
- Websites mit Werbung (325 Seiten)
- Unternehmens-Webseite (72 Seiten)
- Online-Shops (178 Seiten)

2.4 Analyse der SEO-Indikatoren

Die ausgewählten Indikatoren für SEO basieren auf einer Sichtung der Fachliteratur (u. a. Enge, 2015; Erlhofer, 2019) und aus einer Befragung von SEO-Experten (Schultheiß/Lewandowski, 2020). Die Analyse erfolgt anhand folgender Kriterien, die auf ihr Vorkommen im Quelltext geprüft werden:

- Microdata-Formate
- Online-Werbung
- HTTPS
- SEO in robots.txt
- Sitemap
- Viewport
- Nofollow Tags
- Canonical Links
- Description Tags
- Title Tags.

2.5 Bewertung der SEO-Wahrscheinlichkeit

Für den regelbasierten Klassifikator sind folgende Kategorien und Regeln definiert:

- *Höchstwahrscheinlich optimiert*: Die Webseite ist höchstwahrscheinlich optimiert, wenn entweder ein SEO-Plug-in im Quelltext gefunden wurde, der Seitenbetreiber Kunde einer SEO-Agentur ist, die Seite ein Nachrichtenangebot ist, Werbeanzeigen auf der Seite sind oder mindestens ein Microdata-Format vorhanden ist.
- *Wahrscheinlich optimiert*: Die Seite ist wahrscheinlich optimiert, wenn sie nicht höchstwahrscheinlich optimiert ist und zumindest eines der folgenden Kriterien erfüllt: (1) Die Seite ist ein Online-Shop oder eine Unternehmensseite, (2) auf der Seite wurden Analytics-Tools identifiziert, (3) als Übertragungsprotokoll wird HTTPS eingesetzt, (4) SEO-spezifische Hinweise wurden in der robots.txt gefunden, (5) die Website hat eine Sitemap, (6) ein Viewport ist definiert, (7) es wurde ein Nofollow oder ein Canonical Link gefunden, (8) die Ladezeit liegt unter drei Sekunden.
- *Höchstwahrscheinlich nicht optimiert*: Die Domain der Seite ist auf der Liste mit nicht optimierten URLs.
- *Wahrscheinlich nicht optimiert*: Die Seite ist wahrscheinlich nicht optimiert, wenn sie nicht höchstwahrscheinlich nicht optimiert ist und wenn mindestens eines der folgenden Kriterien erfüllt ist: (1) kein Description Tag, (2) kein Title Tag, (3) wenn identische Title Tags auf Unterseiten sind und (4) keine Open Graph Tags definiert sind.

3 SEO-Tool

Das SEO-Tool ist eine Webanwendung, die die Durchführung der Prozesse demonstriert. Nach Eingabe einer URL wird die Analyse durchgeführt. Die Anwendung ist in Python entwickelt und nutzt Flask (in Python geschriebenes Web-Framework), Selenium (portables Framework zum Testen von Webanwendungen), BeautifulSoup (Paket zum Parsen von HTML und XML) und Pandas (Programmbibliothek zur Datenmanipulation und -analyse). Abbildung 2 zeigt ein Beispiel für einen generierten Ergebnisbericht. Der Bericht ist nach den Ergebnissen aus den Prozessen eingeteilt:

1. SEO-Bewertung: Wahrscheinlichkeit von SEO auf der gegebenen Webseite
2. Tools und Plug-ins: Verwendung von SEO-Tools und Analyse-Tools
3. URL Category: Ergebnis der Kategorisierung der URL
4. Indicators for SEO: Übersicht der Indikatoren für SEO.



SEO Assessment



Tools & Plugins



URL Category



Indicators for SEO

Most probably optimized	Probably optimized	Most probably not optimized	Probably not optimized	Uncertain
-------------------------	--------------------	-----------------------------	------------------------	-----------

SEO Tools	✗	
Analytics Tools	✓	Google Tag Manager, GoogleTagManager tracker

Not optimized	✓	Website is definetly not optimized
Customer of a SEO Agency	✗	Website is a customer of a SEO agency
News Service	✓	Website is a news service
Website with ads	✗	Website has online advertisement
Business Website	✗	Website is a business website
Online Shop	✗	Website is an online shop

Description	✓	auf stern.de finden sie news spannende hintergründe sowie bildstarke reportagen aus allen bereichen von politik und wirtschaft bis kultur und wissenschaft.
Title	✓	nachrichten hintergründe & reportagen
Identical Title tags	✗	No identical title tags on subpages
Loading speed	✗	Loading speed is 4.397s > 3s
Hypertext Transfer Secure (https)	✓	Page uses https
SEO in robots.txt	✓	SEO in robots.txt found
Viewport	✓	Viewport defined
Microdata	✓	Microdata definitions found
nofollow-Links	✗	0 nofollow-links found
canonical-Links	✗	0 canonical-links found

Abb. 2 Ergebnisbericht im SEO-Tool

In den Spalten wird das Merkmal mit einer Checkbox und weiteren Erklärung angezeigt. Das Tool zeigt auch tiefer gehende Erklärungen zu allen Prozessen an, die durch Anklicken einer Kategorie geöffnet werden können. Schließlich kann der Bericht als CSV-Datei heruntergeladen werden.

4 Zusammenfassung und weiteres Vorgehen

Das SEO-Tool bietet die Möglichkeit einer Echtzeitanalyse zur Identifizierung von Suchmaschinenoptimierung auf einer Seite. Dabei basiert die Analyse bisher auf technischen Indikatoren und einer Kategorisierung der URL durch Abgleiche mit manuell zusammengestellten Listen, in denen wahrscheinlich optimierte und nicht optimierte Websites gespeichert sind. Diese Listen werden kontinuierlich aktualisiert. Die definierten halb-automatischen Prozesse werden ebenfalls stetig getestet und weiterentwickelt. Mithilfe des Tools können die entwickelten Prozesse und Regeln schnell überprüft werden. In dem weiteren Vorgehen werden die Indikatoren durch externe SEO-Signale wie die Anzahl der Backlinks auf einer Website erweitert. Zusätzlich werden Klassifikationen mit Machine-Learning-Algorithmen getestet, um zusätzliche Indikatoren zu ermitteln.

Forschungsdaten

Der Quelltext des Tools kann über die OSF-Plattform (dx.doi.org/10.17605/OSF.IO/ETZHD) abgerufen werden.

Förderung

Das Projekt „SEO-Effekt“, aus dem dieses Tool hervorgeht, wird von der Deutschen Forschungsgemeinschaft (DFG) unter der Projektnummer 417552432 gefördert.

Literaturverzeichnis

- Enge, E.; Spencer, S.; Stricchiola, J. (2015): *The Art of SEO: Mastering Search Engine Optimization*. Sebastopol, CA: O'Reilly.
- Erlhofer, S. (2019): *Suchmaschinen-Optimierung: Das umfassende Handbuch*. Bonn: Rheinwerk Verlag.
- Goel, S.; Broder, A.; Gabrilovich, E.; Pang, B.(2010): Anatomy of the long tail. In: Davison, B. D.; Suel, T.; Craswell, N.; Liu, B. (Hrsg.): *Proceedings of the Third*

ACM International Conference on Web Search and Data Mining - WSDM '10. New York, NY: ACM Press.

- Li, K.; Lin, M.; Lin, Z.; Xing, B. (2014): Running and Chasing – The Competition between Paid Search Marketing and Search Engine Optimization. In: *47th Hawaii International Conference on System Sciences, Waikoloa, HI*. IEEE, S. 3110–3119. <https://doi.org/10.1109/HICSS.2014.640>
- McCue, T. (2018): SEO Industry Approaching \$80 Billion But All You Want Is More Web Traffic. *Forbes*. <https://www.forbes.com/sites/tjmccue/2018/07/30/seo-industry-approaching-80-billion-but-all-you-want-is-more-web-traffic/>
- Petrescu, P. (2014): Google Organic Click-Through Rates in 2014. <https://moz.com/blog/google-organic-click-through-rates-in-2014>
- Schultheiß, S.; Lewandowski, D. (2020): “Outside the industry, nobody knows what we do”. SEO as seen by search engine optimizers and content providers. *Journal of Documentation* 77 (2), 542–557. <https://doi.org/10.1108/JD-07-2020-0127>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 299–306. DOI: doi.org/10.5283/epub.44949.

Session 5:
Knowledge Representation

The Benefits of RDF and External Ontologies for Heterogeneous Data

A Case Study Using the Japanese Visual Media Graph

Senan Kiryakos, Magnus Pfeffer

Stuttgart Media University, Germany
{[kiryakos](mailto:kiryakos@hdm-stuttgart.de), [pfeffer](mailto:pfeffer@hdm-stuttgart.de)}@hdm-stuttgart.de

Abstract

Across numerous fan created and curated websites, there exists a wealth of semantically rich descriptive data for a variety of Japanese visual media, such as anime, manga, and video games. The amount and granularity of these data makes it valuable for domain researchers, but the semantic heterogeneity and lack of interconnectedness makes analysis cumbersome. Seeking to address this issue, the Japanese Visual Media Graph (JVMG) project aims to create a type of global database built using RDF and external ontologies to better enable data-driven research in the domain. We discuss the benefits this approach has when compared to the local relational databases used at the fan-sites, such as enabling the easy creation of aggregate resources using data from multiple providers, and the ability to constantly update and alter a schema over time. This encourages the incorporation of additional data in the future, while still maintaining rich data semantics and provenance. While many of the benefits are discussed in the context of the JVMG project or the Japanese visual media domain, we discuss how this database approach may be similarly advantageous to other projects that seek to create aggregate resources, or collect heterogeneous data from diverse sources.

Keywords: ontology-based data access; semantic heterogeneity; knowledge graph; fan-curated data; digital humanities

1 Introduction

For researchers interested in Japanese visual media (e.g., anime, manga, video games), the largest collection of semantically rich descriptive data exists not at existing memory institutions, but across a multitude of fan-created and curated websites. These data are abundant, highly granular, multilingual, and provide a rich source for which data-driven research can be conducted. The nature of these data make it valuable for researchers studying various aspects of the Japanese visual media domain, such as themes, genres, and characters, as well as trends, receptions, and influences. In attempting to better meet the needs of researchers, the Japanese Visual Media Graph (JVMG) project (Pfeffer & Roth, 2019) seeks to utilize this community-created data to develop a graph-based, highly interconnected database (i.e., a type of knowledge graph), combined with flexible and powerful querying and analytical tools.¹ It should be noted that while the database may contain some links to visual media found at the original data providers, the primary contents of the database are descriptive metadata and not the visual materials themselves.

The desire to utilize this fan-created data in an interconnected database poses several issues, but also unique opportunities. Relying on multiple data providers means the data is fairly heterogeneous, available in different formats, covering diverse entity levels, described according to different data models, and using different vocabularies. Thus, the integration of this disparate data into a single database involves significant data transformation, property mapping, the identification of related or identical resources, and the creation of a unified global ontology. Though a challenge, this integration presents us with unique opportunities, such as merging related resources with data sourced from multiple providers, developing a more formal data model and ontology for the domain of Japanese visual media, and exploring different ways of making these data accessible to, and able to be queried by, researchers.

In detailing this database approach, this paper outlines the issues and opportunities encountered by the JVMG project thus far, and highlights the advantages and disadvantages presented when using a database that relies on

¹ The database is currently for internal use and will be made more publicly available later this year. Up-to-date information on the state of the project and means of access will be published on the project homepage <https://jvmg.iuk.hdm-stuttgart.de/>.

linked data formats and external ontologies. We discuss the benefits that this approach has had for primary aspects of our project specifically, such as the integration of heterogeneous data, and the creation of aggregated resources, while also presenting this approach as one that may be of interest to projects with similar goals regarding data integration.

Following this introduction, Section 2 describes the data providers, the type of data being collected and integrated, and data processing. Section 3 outlines the differences between databases currently used with these data and the approach we use, along with discussing two specific advantages. Lastly, we conclude with thoughts and future plans in Section 4.

2 Data providers

In order to aid those seeking to conduct data-driven research on the Japanese visual media domain, a large amount of descriptive data is required. The best source of these data can change depending on the domain. Memory institutions, such as libraries or archives, for example, contain a wealth of data for more traditional literary materials, though this is often of a limited granularity due to factors such as cataloguing rules. Past research (e.g., Kiryakos et al., 2017) has shown that for media types like anime, manga, and video games, the best source of data is instead found across a variety of hobbyist webpages. The amount of data is not only vast, but extremely granular, as communities with various interests and perspectives describe the same materials in distinct ways.

As such, we rely on these communities, several of whom we have made formal collaboration agreements with, to utilize their data. These partners include the Anime Characters Database, AnimeClick, and the Visual Novel Database (VNDB). Other sources that allow for open data use, such as Wikidata, are also utilized. Though the coverage and contents of these sites differ, they generally contain traditional descriptive data, such as titles, creators, and dates, along with more granular data such as genres and tags, character lists and traits, relationships between related entities (e.g., characters or creative works), and organizational provenance. The amount of some of these data is shown in Table 1. This is the type of information that we seek to leverage to make data-driven research questions regarding Japanese visual media more easily investigated. Before doing so, however, work on the data must be per-

formed, as the communities provide us with data in different ways, using different formats, and according to their own individual data models and vocabularies. Though a detailed description of the data transformation pipeline is outside the scope of this paper, an outline of this transformation will help better illustrate how the data comes to be integrated and connected in our database.

Table 1: Approximate numbers of various data collected from providers

Visual Novel Database					
Visual Novel works	Unique releases	Characters	Descriptive work tags	Descriptive character tags	Agents
28,000	73,000	91,000	2600	2800	31,000
AnimeClick					
Anime	Manga	Characters	Related works	Authors	Other staff
9400	11,000	102,000	15,000	28,000	39,000
Anime Characters Database					
Characters	Works	Character relations	Descriptive character tags	Descriptive work tags	Voice actors
101,000	11,000	17,000	3800	1100	4900

Data transformation and ontology creation

Our data is sourced in a variety of ways, e.g., as data dumps or through APIs, depending on the provider, and is typically either in the form of SQL tables or JSON. As our database uses Resource Description Framework (RDF) serializations, data transformation is the first technical step, though this is preceded by an analysis of each provider to better understand their data model. Conversion is done using Python, and is then processed using the RDFLib package. The resulting output is a series of RDF files able to be integrated and queried using the query language SPARQL. While other existing methods of creating RDF output based on relational data exist (see, e.g., Arenas et al., 2012; Das et al., 2012), our method differs, as we do not create a mapping layer over the relational data, but instead ingest and transform it directly.

This processing is a central part of the data integration process, as RDFLib allows us to define and output RDF triples using non-RDF input data we receive from our data providers. Through an analysis of the data, an ontology is created for each data provider, which includes a class structure

based on how resources are presented by a provider, and an accompanying RDF vocabulary file. At its most basic, an example of this vocabulary conversion would be the ‘title’ column header in an SQL table from the VNDB becoming `<http://mediagraph.link/vndb/ont/title>` – a property in our namespace that represents that same title column header. An excerpt of this original tabular data from VNDB with limited properties is shown in Table 2, with its RDF transformation shown in Figure 1. While this RDF data is able to be queried using SPARQL, a separate RDF vocabulary file combined with our web frontend allows for these data to be displayed and browsed in a more human-readable way via a web browser; a sample page is shown in Figure 2.

Table 2: A sample of tabular data for a visual novel from VNDB

Visual Novel table					
id	title	original	pid	l_wp	l_wikidata
7014	Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei	ダンガンロンパ 希望の学園と絶望の高校生	1761	Danganronpa:_Trigger_Happy_Havoc	Q1035512
Producer table					
id	name	original	type	l_wp	l_wikidata
1761	Spike	スパイク	co	Spike_%28company%29	Q2541040

Data transformation and ontology creation for each provider allows the loading and querying of each dataset separately, providing users with the ability to access datasets from individual or multiple providers. Creating ‘local’ schemas in this way also allows for a more straightforward relation between the providers’ data model and a unified ‘global’ schema, which must occur at some point during data integration (Cruz & Xiao, 2005; Doan et al., 2012). This relation mapping between the ontology of each data provider and our own unified ontology is performed primarily using the Web Ontology Language (OWL) and Simple Knowledge Organization System (SKOS), allowing us to map properties and concepts between datasets.

```

@prefix vndb: <http://mediagraph.link/vndb/ont/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://mediagraph.link/vndb/vns/7014> a vndb:VisualNovel ;
  rdfs:label "Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei"@en,
    "ダンガンロンパ 希望の学園と絶望の高校生"@ja ;
  vndb:l_wp "<https://en.wikipedia.org/wiki/Danganronpa:_Trigger_Happy_Havoc>" ;
  vndb:l_wikidata <https://www.wikidata.org/wiki/Q10335512> ;
  vndb:original "ダンガンロンパ 希望の学園と絶望の高校生" ;
  vndb:title "Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei" ;
  vndb:producedBy <http://mediagraph.link/vndb/producers/1761> ;
  vndb:vndbLink <http://vndb.org/v7014> .

<http://mediagraph.link/vndb/producers/1761> a vndb:Producer ;
  rdfs:label "Spike" ;
  vndb:l_wikidata <https://www.wikidata.org/wiki/Q2541040> ;
  vndb:l_wp <https://en.wikipedia.org/wiki/Spike_%28company%29> ;
  vndb:name "Spike" ;
  vndb:original "スパイク" ;
  vndb:type "co" ;
  vndb:vndbLink <http://vndb.org/p1761> .

```

Fig. 1 Table 2 after RDFLib data processing and transformation

[en](#) [ja](#) [all](#) [simple](#) [XML](#) [Turtle](#) [Dark Mode](#)

Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei

Property	Value
label	<ul style="list-style-type: none"> • Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei en vndb • ダンガンロンパ 希望の学園と絶望の高校生 ja vndb
Original en ja	ダンガンロンパ 希望の学園と絶望の高校生 vndb
Produced by en ja	Spike vndb
Title en ja	Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei vndb
type	Visual Novel en ja vndb
VNDB Link en ja	http://vndb.org/v7014 vndb
Wikidata Link en ja	https://www.wikidata.org/wiki/Q10335512 vndb
Wikipedia Link en ja	https://en.wikipedia.org/wiki/Danganronpa:_Trigger_Happy_Havoc vndb

Fig. 2 Web frontend screenshot for RDF data shown in Figure 1. Note the ‘Producer’ data is a separate page linked to via the ‘Produced by – Spike’ value URL.

In addition to the benefits discussed in Section 4, an intrinsic advantage resulting from this transformation is that much of the formerly tabular data is turned into web-resolvable URIs, enabling their inclusion in other web applications, and as a browser-based entry point for the database. This “things, not strings” (Singhal, 2012) approach is fundamental to providing the benefits of knowledge graphs, and our data transformation brings these benefits to a large amount of data in the Japanese visual media domain. This applies not only to properties, such as the ‘title’ example mentioned previously, but to their values, resulting in unique web identifiers for single creative works, franchises, their creators, and their contents.

3 Benefits and drawbacks of an external ontology database approach

The data sourced from our providers belongs to schema-bound, local databases (i.e., tabular relational databases). This approach requires the schema to be developed first, and results in a data model that is difficult to later alter. This also means the data is viewed from a single perspective, and can discourage the integration of new data not considered during initial development. This also makes it difficult to incorporate data from additional sources, as the schemas will inevitably differ, requiring adjustment based on these new sources.

The database approach we instead use is that of a schema-less, global database combined with external ontologies. As they are separate from the data itself, the ontologies can easily be altered after their initial creation, which more readily supports the incorporation of new heterogeneous data. The external ontologies here refer to individual RDF ontologies, mentioned in Section 2. The varying granularity of the data we use also encourages the use of RDF-based ontologies, as they are better able to specify domain knowledge and semantic richness when compared to relational databases (Buron et al., 2019; Munir & Sheraz, 2018). This is not to say that the approach we use is preferred over relational databases in all cases, but rather this approach has provided significant benefits for the JVMG project specifically. We feel these benefits are worth highlighting so that other similar projects (e.g., those working with heterogeneous data or those seeking to create aggregate re-

sources) may be informed on the advantages and disadvantages of this approach. The following section discusses some of these benefits in further detail.

3.1 Creating aggregate resources

Each data provider has a separate dataset and receives its own unique ontology for use in our database, allowing for the browsing and querying of individual datasets when required. As much of the data across providers describes the same or related resources, researchers would also benefit from the ability to query combined datasets which contain this related, merged data. This is particularly beneficial to our project due to the high granularity and differing perspectives of the community data providers. If our main source of bibliographic data were libraries, for example, we would expect a significant amount of duplicate data due to restrictive cataloguing rules, and the benefits of aggregating data would be limited. The advantages of aggregation become more clear when understanding the nature of diverse fan-created source data, some of which focuses on particular aspects of Japanese visual media rather than the creative work itself, such as the characters (i.e. the Anime Characters Database).

Our RDF-based, external ontology approach allows for the easy creation of these aggregate resources. During the merging process, all of the matching data from all datasets related to a single related resource is aggregated, including duplicate or conflicting data. In a next step, a curated subset of attributes is created for each entity type, and existing data is mapped to these attributes. Duplicate and conflicting information is resolved in this step by different means: A preferred piece of information can be selected (e.g., the canonical title) and the remaining pieces are listed as variations (e.g., title variants). As ontologies are external to the data, we can choose a unique part of the namespace to be used with these aggregate resources. Any researcher working with the domain data can then solely focus on this consolidated ontology, as it will cover the whole domain and make the heterogeneous data accessible in a uniform way. As the original provider data as well as the curated data is available using the same entity URL, but with different ontologies, both provenance and all original values from the different sources are preserved and are immediately available. Though the creation of a separate ontology for the merged dataset requires additional labor, it makes the data more convenient to browse, and allows for the modelling of the domain as a

whole, rather than the modelling of single datasets from individual providers. As this step takes place after the creation and analysis of both individual provider and aggregate datasets, the resulting ontology reflects a more thorough understanding of the complete domain and its data; this is a key advantage of the schema-less approach presented here, as in more traditional databases, this step would have come prior to merging, and unexpected changes to the schema may then need to take place, requiring difficult ontology or data pipeline adjustments.

The creation of aggregated resources also allows for an easier managing of different data quality levels, as data is additive and conflicting statements are more visible. As we do not edit source data, and researchers may be interested specifically in conflicting or contentious data across communities, these conflicts and other quality issues are generally left as is. That said, in order to assess data quality and possible types of mistakes and to ease matching across providers, a random selection of statements is currently manually checked against ground truth where available (e.g., the visual media itself, credit rolls) as a part of a study on fan community data quality.

One drawback inherent to this data integration step is that when merging data from multiple providers, care must be taken to ensure they are describing not only the same resource, but the same entity level of that resource. Entity levels here refer to a separation of ‘levels’ of a resource that may be described by a given provider. For example, one provider may describe an entire franchise or intellectual property as a single resource, another may describe its manifestation in a single medium such as anime. A prominent example of this entity partitioning can be seen on Wikipedia, where the franchise *Pokémon* has separate articles for the umbrella franchise, the series of video games, single games within that series, etc. One popular model defining such entity levels is the FRBR entity-relationship model (IFLA, 2009), which defines Work, Expression, Manifestation, and Item entities. Though we do not currently define unique entities in the data we collect, we do analyze the data structure and properties used by each provider so that we can determine whether or not descriptions are about identical or related entities.

The significance of this issue is dependent on the domain and heterogeneity of the data being merged. For our project, particular attention must be paid here, as entity levels described for Japanese visual media amongst enthusiast communities frequently vary (Kiryakos & Sugimoto, 2019), and because relationships across different entity levels is quite common in Japanese visual media (Lee et al., 2018). If granular data from each provider is main-

tained (e.g., if a merged resource contains information on both manga volumes and anime episodes), it becomes difficult to define what exact entity or concept that resource is describing. As each resource in our database is given a unique URI, merged resources can be referred to globally in Linked Data or other web contexts. If an aggregate resource contains data from various entity levels, we cannot claim that its URI represents a common entity type, such as a franchise or single manifestation, as is possible if data is more strictly partitioned and merged. This problem can arise with any project integrating heterogeneous data, but it should be noted that our approach does not provide a solution for this, and that entity identification and separation must be performed if desired.

3.2 External ontologies

The ontology being separate to the data itself provides a level of flexibility that is not present with traditional relational databases. The class structure and vocabulary is able to be changed whenever desired, either to reflect an updated understanding of the data, or to incorporate new data requiring the addition of new properties. Through the use of multiple external ontologies mapped to one another, a single dataset can be viewed and presented in distinct ways. Datatype constraints are also not rigid, allowing for different value types, e.g., URI and string, to exist for a single property. This becomes important when merging data, as one provider may describe an author using a plain text string, whereas another may use a URI linking to something such as the Virtual Internet Authority File.

As our approach utilizes RDF, the use of existing, widely used vocabularies, such as those provided by Dublin Core and Schema.org, is also possible. Depending on the domain, such vocabularies may be sufficient for modeling single or aggregate datasets. As the Japanese visual media domain is fairly niche and the source data is highly granular, the creation of a new ontology was preferred (though some related niche media controlled vocabularies exist, such as the Comic Book Ontology). Though this requires significantly more labor, it allows for a given dataset to be modeled and described in an exact manner, and results in a more thorough understanding of the domain. The creation of a new RDF ontology, particularly for a previously uncovered domain, also enables its inclusion in other Linked Data applications.

A drawback with this approach is that despite the flexibility and conveniences that come with using external ontologies, the resulting schema can be

difficult to enforce, unlike relational database tables, particularly once resource merging takes place. For example, one may define what value types are valid for a given set of attributes based on an analysis of existing data. If new data is integrated and merged that provides alternative values for these attributes, the schema definition will be invalid (though the data will remain retrievable). One remedy to this is to regularly update the ontology to reflect the newly integrated data, which external ontologies easily support. If one wishes to define a more strict data model, one method is to use RDF validation and constraint methods, such as Shape Expressions (Prud'hommeaux et al., 2014) and Shapes Constraint Language (Corman et al., 2018), which allows for the definition of conditions that must be present in an RDF graph to be considered conformant.

4 Conclusion

Through the development of a database which utilizes RDF and external ontologies, the JVMG project is poised to meet its goal of serving researchers studying Japanese visual media. This database approach allows for the easy and ongoing integration of heterogeneous data from multiple providers, while maintaining semantic richness, and modelling the domain in the process. Additionally, this approach allows for the seamless creation of aggregate resources, which merge related data from multiple providers. The use of external ontologies encourages the integration of new data sources, further improving the coverage and granularity of the database, as ontologies can be easily adapted to incorporate additional heterogeneous data over time. The resulting database provides researchers with a single source with which to browse or query a wealth of granular, interconnected data on the Japanese visual media domain.

Future work seeks to build on this development in multiple ways. First, we plan to obtain more data from various web resources, particular those that offer existing query or collection methods, such as Wikipedia, Wikidata, and various fandom pages. While a 'coverage ceiling' will eventually be reached, this extra data can still be useful, e.g., for finding additional connections between entities. Also, we will focus on making the data more accessible to researchers who do not have the knowledge to create large SPARQL queries on the fly. In order to learn more about data-driven research methods,

we employ “tiny use cases”², i.e., small research questions that require data from the domain. Mixed groups of researchers from media/Japanese studies and information/computer science respectively work together on these questions and document their approaches and findings. These can be used to familiarize external researchers with the database and its ontologies as well as introduce typical data analysis workflows and techniques. Once more external researchers have started doing active research using the project’s data, we will have a better understanding of what features and functions could be added that go beyond predefined queries or query wizards.

References

- Arenas, M., Bertails, A., Prud’hommeaux, E., & Sequeda, J. (2012). A Direct Mapping of Relational Data to RDF, W3C Recommendation.
- Buron, M., Goasdoué, F., Manolescu, I., & Mugnier, M. L. (2019). *Ontology-Based RDF Integration of Heterogeneous Data* [Technical Report]. LIX, Ecole polytechnique; Inria Saclay. <https://hal.inria.fr/hal-02266517/document>
- Corman, J., Reutter, J. L., & Savković, O. (2018). Semantics and Validation of Recursive SHACL. In Vrandečić, D. et al. (Eds.), *The Semantic Web – ISWC 2018, Proceedings, Part I* (318–336). Cham: Springer Nature.
- Cruz, I. F., & Xiao, H. (2005). The role of ontologies in data integration. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 13(4), 245–252.
- Das, S., Sundara, S., & Cyganiak, R. (2012). R2RML: RDB to RDF Mapping Language. <https://www.w3.org/TR/r2rml/>
- Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of Data Integration*. Amsterdam: Elsevier.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (2009). *Functional Requirements for Bibliographic Records*. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- Kiryakos, S., & Sugimoto, S. (2019). Building a bibliographic hierarchy for manga through the aggregation of institutional and hobbyist descriptions. *Journal of Documentation*, 75(2), 287–313.

2 Several writeups about these can be found at <https://jvmg.iuk.hdm-stuttgart.de/category/tiny-use-case/>.

- Kiryakos, S., Sugimoto, S., Nagamori, M., & Mihara, T. (2017). Aggregating Metadata from Heterogeneous Pop Culture Resources on the Web. *International Conference on Dublin Core and Metadata Applications 2016* (pp. 65–74). Silver Spring, MD: Dublin Core Metadata Initiative.
- Lee, J. H., Jett J., Cho, H., Windleharth, T., Disher, T., Kiryakos, S., & Sugimoto, S. (2018). Reconceptualizing superwork for improved access to popular cultural objects. *Proceedings of the Association for Information Science and Technology* 55, 274–281.
- Munir, K., & Anjum, M. S. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, 14(2), 116–126.
- Pfeffer, M., & Roth, M. (2019). Japanese Visual Media Graph: Providing researchers with data from enthusiast communities. *International Conference on Dublin Core and Metadata Applications* (pp. 136–141). Silver Spring, MD: Dublin Core Metadata Initiative.
- Prud'hommeaux, E., Gayo, J. E. L., & Solbrig, H. (2014). Shape expressions: an RDF validation and transformation language. *Proceedings of the 10th International Conference on Semantic Systems* (pp. 32–40). New York, NY: ACM Press.
- Singhal, A. (2012): Introducing the Knowledge Graph: things, not strings. In Google Official Blog. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 308–320. DOI: doi.org/10.5283/epub.44950.

Information Organization and Access in Digital Humanities

TaDiRAH Revised, Formalized and FAIR

Luise Borek

Technical University of
Darmstadt, Germany

luise.borek@tu-darmstadt.de

Canan Hastik

DIPF | Leibniz Institute
for Research and Infor-
mation in Education,
Frankfurt/M., Germany

hastik@dipf.de

Vera Khramova

Darmstadt University
of Applied Sciences,
Germany

vera.khramova@stud.h-da.de

Klaus Illmayer

Austrian Centre for Digital Humanities
and Cultural Heritage (ACDH-CH),
Vienna, Austria

klaus.illmayer@oeaw.ac.at

Jonathan D. Geiger

Academy of Sciences and
Literature, Mainz, Germany

jonathan.geiger@adwmainz.de

Abstract

Classifying and categorizing the activities that comprise the digital humanities (DH) has been a longstanding area of interest for many practitioners in this field, fueled by ongoing attempts to define the field both within the academic and public sphere. Several European initiatives are currently shaping advanced research infrastructures that would benefit from an implementation of a suiting taxonomy. Therefore, new humanities and information science collaborations have been formed to provide a service that meets their needs. This working paper presents the transformation of the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) in order to make it machine-readable and become a formalized taxonomy. This includes the methodology and realization containing a complete revision of the original version, decisions in modelling, the implementation as well as organization of ongoing and future tasks. TaDiRAH addresses a wide range of humanities disciplines and integrates application areas from philologies as well as epigraphy, and musicology to name just a few. For this reason, the decision in favor of SKOS was made purely pragmatically in terms of technology, concept and domains. New language versions can now be easily integrated

and low-threshold term extensions can be carried out via Wikidata. The new TaDiRAH not only represents a knowledge organization system (KOS) which has recently been released as version 2.0. According to the FAIR principles this new version improves the Findability, Accessibility, Interoperability, and Reuse of research data and digital assets in the digital humanities.

Keywords: taxonomy; digital humanities; linked open data; knowledge organization

1 Motivation

The knowledge organization within the digital humanities is located at the intersection of scientific practice and state of the art. The investigation of the genesis of this domain shows that knowledge modelling has become established as a standard procedure (Unsworth, 2000b; McCarty, 2003; Schnapp, 2011) not only in text-oriented sciences. This practice integrates both subject-specific and information science methods that are generally known as ‘scholarly primitives’ within the digital humanities domain (Unsworth, 2000a). While the objects of research and their materiality vary greatly in the participating disciplines, ‘methodological commons’ may be identified, under which generic and more specific research activities may be subsumed.

For a better understanding of the general digital scientific process and for the shaping of knowledge, new knowledge models are being continuously developed and expanded. The Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) is such a model and was designed in close collaboration with the community to organize and categorize digital humanities content and generate academic credibility as well as greater visibility to the field (Borek et al., 2016). At the same time, it has become an instrument to reflect upon digital humanities as a discipline as its broader implementation provides insights on which activities are being used in which contexts. In this sense, the taxonomy is never complete and continuously depends on the community to help shape and adapt it according to their needs.

Since 2015 this vocabulary is used in a variety of contexts to categorize bibliographic data (DARIAH Zotero bibliography “Doing Digital Human-

ties¹), research projects (e.g., in the registries DHCommons² and AGATE European Science Academies Gateway for the Humanities and Social Sciences³), study programmes (Digital Humanities Course Registry⁴), conference abstracts, and research tools (TAPoR Text Analysis Portal for Research⁵, CLARIAH-DE⁶, and the SSH Open Marketplace⁷).

From the beginning, TaDiRAH has been a community-driven initiative that pursues a practice-oriented approach, in which it benefits from further contributions, collaborations, development and reuse. The interdisciplinary community of digital humanities scholars have not only helped shape the relevant terminology and the associated definitions, they have also provided translations on concept level to several languages such as German, French, Spanish, Portuguese, Serbian, and quite up-to-date Italian.

While TaDiRAH has been in wide use over the years, its application had not gone beyond keywording as the non-funding project did not have the necessary resources to provide a formalized, persistent and machine-readable version. With several European initiatives currently shaping advanced research infrastructures that could benefit from an implementation of the taxonomy, new humanities and information science collaborations have been formed to bring TaDiRAH to its next level.

In this paper, we describe the steps that were undertaken in order to transfer TaDiRAH to a machine readable taxonomy as a knowledge organization system (KOS) currently released as version 2.0 within the Vocab service at the Austrian Academy of Sciences.⁸ This working paper covers the methodology and realization that includes a complete revision of the original version, decisions in modelling, the implementation as well as the organization of ongoing and future tasks.

-
- 1 Doing Digital Humanities – A DARIAH Bibliography, https://www.zotero.org/groups/113737/doing_digital_humanities_-_a_dariah_bibliography (Nov. 18, 2020).
 - 2 now within <https://dhcenter.net.org/> (Nov. 18, 2020)
 - 3 AGATE. A European Gateway for the Academies of Sciences and Humanities, <https://agate.academy/de.html> (Nov. 18, 2020)
 - 4 <https://dhcr.clarin-dariah.eu/> (Nov. 18, 2020)
 - 5 TAPoR 3. Discover research tools for studying texts, <http://tapor.ca> (Nov. 18, 2020)
 - 6 <https://www.clariah.de/> (Nov. 18, 2020)
 - 7 SSH Open Marketplace, <https://www.sshopencloud.eu/ssh-open-marketplace> (Nov. 18, 2020)
 - 8 TaDiRAH: Taxonomy of Digital Research Activities for the Humanities, Version 2.0.0, <https://vocabs.dariah.eu/tadirah/> (Nov. 18, 2020)

2 Methodology

In information science, a term represents an abstract concept for a set of objects and items with common properties. This term is linguistically represented by a name. These designations can be natural language words but also a notation of a classification or order of knowledge. Various methods can be used to introduce these concepts into taxonomy. An empirical procedure for the concept analysis would be the semi-automatic calculation of word similarities in texts by means of cluster analysis. Therefore, a representative reference corpus must already be available. The method of formal concept analysis is used to determine connections and hierarchical relationships between concepts. A tree-like order structure is designed using superordinate and subordinate conceptual relations, with which common properties are combined into concepts. Following the hermeneutic approach, terms are subject to change and this change must always be viewed in its context. In pragmatics, the meaning, purpose and aim of terms are in the foreground while language always influences conceptualization (Hjørland, 2009).

The first TaDiRAH version 0.5.1⁹ with its research activities, techniques and objects was created through a case study in pragmatic classification based on a hermeneutical and iterative approach intensively discussed with the community (Borek et al., 2016). The resulting concept order comprises three independent concept trees, each with one superordinate and various subordinate levels. The term definitions contain information about synonyms and their semantic context but also minimal deviations from adjacent terms.

In order to eliminate the vagueness and to optimize the usefulness of the existing vocabulary for the growing knowledge domain of digital humanities, the need arose to revise and harmonize the structure and semantics of the model, as well as the concept terms and definitions, the so called scope notes, with respect to research activities and research techniques. This also includes expanding the model on the basis of community usage. As a result, a focus has been set on the research activities and the relation of techniques associated to them. To achieve this, existing implementations such as TAPoR, SSH und CLARIAH-DE were evaluated and aggregated in the model. At the same time, this led to the decision to exclude related research objects from

9 TaDiRAH: Taxonomy of Digital Research Activities for the Humanities, <http://tadirah.dariah.eu/vocab/index.php> (Nov. 18, 2020)

the model as no reliable systematic data has been available for modelling. However, research objects may be associated through suiting vocabularies in the future (within TaDiRAH or from additional sources).

In addition, criteria for the descriptions have been established to develop consistent definitions of terms with a focus on their usefulness for the domain. Scope notes should in future not be too narrow nor too broad, and should not contain any or explicit demarcations to other terms. Moreover, scope notes should not include contradictions or negative definitions. In principle, multiple definitions within a scope note, redundancies with other definitions and unnecessary references should be avoided. Finally, TaDiRAH terms should be consistently written in small letters and gerundiva to represent the performativity of the activities. This is to be largely continued in the subclasses. In addition to the scope notes that are assigned to the TaDiRAH core model on top and sub-concept level, so-called ‘aggregated concepts’ – mostly represented by research techniques – are extended by external Wiki-data definitions.

Only the linking of the totality of all concepts with each other allows for a knowledge order to emerge from a concept definition. These quasi hard-wired terms and their semantic relations provide an overview of the hierarchical, associative relationship between terms and their corresponding expressiveness of content in the resulting knowledge model and enable an improved search and retrieval. While it is required for a vocabulary to be consistent and coherent, it does not necessarily have to be complete, in order to perform search and retrieval. A taxonomy must always be open for extensions without the need to revise existing definitions (Gruber, 1993). This modelling process integrates analysis and organization activities while the next process step indicates a strong relation to knowledge engineering methods by formalizing and implementing the TaDiRAH conceptualization in standardized Simple Knowledge Organization System (SKOS)¹⁰. As a specification and standard according to the FAIR principles¹¹ SKOS and the digital infrastructure where it is published makes TaDiRAH findable, accessible, interoperable and reusable as a knowledge organization system.

10 SKOS Simple Knowledge Organization System, <https://www.w3.org/2004/02/skos/> (Nov. 18, 2020)

11 FAIR Principles, <https://www.go-fair.org/fair-principles/> (Nov. 18, 2020)

3 Modelling and implementation

Besides the redesign of TaDiRAH and the transformation of a prior nomenclature into a standardized taxonomy, its SKOSification was also implemented in several iterations. The goal was to make the simple keyword list of different research activities, research techniques and research objects, used in the digital humanities until then, interoperable available in a machine-readable form and to further improve the search for relevant information by semantically linking these data. This structured representation of the data and the relationships between the data is an important prerequisite to demonstrate the applicability and range of the developed model, but also for its use and reuse.

The machine-readable form will be achieved with the help of SKOS. As a formal language based on the Resource Description Framework (RDF)¹², SKOS with its clear structure and low level of detail allows easy to carry out adjustments and extensions of the model. SKOS was developed specifically for the representation of data of controlled vocabularies and taxonomies. It can be integrated into the Semantic Web and the modelled data can be stored as Linked Open Data (LOD) and further be linked to external resources. In this way, not only the availability and visibility of knowledge can be improved, but also interoperability with related disciplines can be achieved. For the creation of the SKOSified TaDiRAH model, Protégé¹³ was used, an ontology editor that facilitates the modelling process and also provides a visual representation of the model.

3.1 SKOSification: Concept hierarchy and semantification

Currently the revised TaDiRAH metadata profile includes 168 terms. The task included the creation of a model that would be broad enough to cover the most important research activities in the digital humanities but would be specific enough to help structure various digital humanities tools, projects, websites and bibliographies.

12 W3C Recommendation 25 February 2014, <https://www.w3.org/TR/rdf11-concepts> (Nov. 18, 2020)

13 Protégé v.5.5.0, a free, open-source ontology editor and framework for building intelligent systems, from <https://protege.stanford.edu/> (Nov. 18, 2020)

All terms in the scheme are represented as concepts by **skos:Concept** which is the central structural element in SKOS. Each concept is unique and identified by a Uniform Resource Identifier (URI), thus ensuring that each resource is uniquely identifiable on the web. All individuals must belong to a specific concept scheme, which is expressed through the class **skos:ConceptScheme**. Moreover, they are represented with a scope note or an external Wikidata definition, some broader and narrower terms, and also assigned to multilingual labels, currently a total of seven languages. Since the translations are based on the ‘old’ TaDiRAH version, they are neither complete nor consistent, and should be revised and adapted to the new model version 2 as soon as possible. All properties are expressed in the form of relations to represent the individual concept as well as its interrelation with other concepts. An example of the SKOS structure of TaDiRAH version 2 is shown in the next figure (Fig. 1).

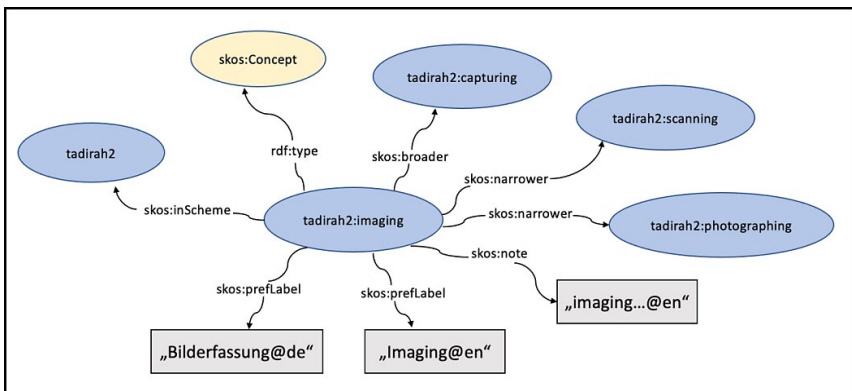



Fig. 1 SKOS structure of the concept *Imaging*

Hierarchical conceptual order between the scheme and the top concepts is achieved using the properties **skos:hasTopConcept** and **skos:topConceptOf**. From a total of 168 terms only seven are top concepts, namely *Analyzing*, *Capturing*, *Creating*, *Disseminating*, *Enriching*, *Interpreting* and *Storing*. The affiliation of all concepts to the schema is made recognizable with the property **skos:inScheme**. To represent the hierarchical relationships between the various concepts the properties **skos:broader** and **skos:narrower** are used. The properties **skos:prefLabel** and **skos:altLabel** are required to indicate preferred and alternative labels of the resources. To express the multilingual labels the property **skos:prefLabel** is used. Information about

the meaning of a resource is held by the properties **skos:scopeNote** or **skos:definition**. These two properties serve to distinguish the self-written scope notes from external Wikidata definitions, which are identified with the property **skos:definition**. In addition, the property **skos:closeMatch** is needed to create the mapping between the first and the new TaDiRAH version, but also through the terms taken from Tapor, SSH and CLARIAH-DE (e.g., *Optical Character Recognition, Searching, Archiving*) which are successfully aggregated in the model.

The new features outlined here simultaneously mean formalization and a semantic enrichment that improve the taxonomy and facilitate its application. The SKOSification resolved former ambiguity and vagueness, transformed implicit information into explicit representation, and led to more flexibility.

Prior only hierarchically organized concepts have been transformed into a formal classification, thus enabling semantic interpretation of concepts, their relations and represented entities.

Capturing > Imaging													
PREFERRED TERM	Imaging 												


TYPE	owl:NamedIndividual												
BROADER CONCEPT	Capturing												
NARROWER CONCEPTS	Photographing Scanning												
NOTE	imaging refers to capturing texts, images, artifacts or spatial formations in 2D or 3D using optical or acoustic means (light, laser, infrared, ultrasound). Imaging usually does not lead to the identification of discrete semantic or structural units in the data, such as words or musical notes.												
IN OTHER LANGUAGES	<table border="0"> <tr> <td>Imagerie</td> <td>French</td> </tr> <tr> <td>Bilderfassung</td> <td>German</td> </tr> <tr> <td>Gestione delle Immagini</td> <td>Italian</td> </tr> <tr> <td>Captura em formato de imagem</td> <td>Portuguese</td> </tr> <tr> <td>Усликавање</td> <td>Serbian</td> </tr> <tr> <td>Captura en formato imagen</td> <td>Spanish</td> </tr> </table>	Imagerie	French	Bilderfassung	German	Gestione delle Immagini	Italian	Captura em formato de imagem	Portuguese	Усликавање	Serbian	Captura en formato imagen	Spanish
Imagerie	French												
Bilderfassung	German												
Gestione delle Immagini	Italian												
Captura em formato de imagem	Portuguese												
Усликавање	Serbian												
Captura en formato imagen	Spanish												
URI	https://vocabs.dariah.eu/tadirah/imaging 												
Download this concept:	RDF/XML TURTLE JSON-LD												
CLOSELY MATCHING CONCEPTS	http://tadirah.dariah.eu/vocab/index.php?tema=12&/imaging tadirah.dariah.eu												

Fig. 2 Concept – *Imaging*, from <https://vocabs.dariah.eu/tadirah/imaging> (Nov. 18, 2020)

3.2 Implementation and publication with Vocabs

The TaDiRAH vocabulary is published on the vocabulary server of DARIAH-EU.¹⁴ This service is run by the Austrian Centre for Digital Humanities and Cultural Heritage at the Austrian Academy of Sciences, one of the partners in the CLARIAH-AT consortium.¹⁵ For showing and browsing the vocabularies the open source software Skosmos¹⁶ is used. This software works natively on a triple store database where it expects a SKOS compatible vocabulary to then show in the browser. The triple store that is used is an Apache Jena Fuseki setup.¹⁷ Usually, the workflow for inserting and updating a vocabulary expects an RDF import file, e.g., in a Turtle format. Such a file is manually imported into Apache Jena Fuseki in a dedicated namespace. This namespace is also declared in the configuration file of Skosmos and after a successful import the changes are immediately taken over by Skosmos. There is also an interface for SPARQL if there is a need for complex queries.¹⁸ This service is used by a growing number of vocabularies, focusing on such vocabularies that are used in the DARIAH community (mainly research disciplines out of the humanities and arts). This governance model guarantees sustainability and maintainability of the service and the data of the vocabularies.

Skosmos allows easy browsing in a vocabulary, it enables a search option and it does in general present the concepts of a vocabulary in a structured, human-readable way. It enables an easy way to interact with a SKOS based vocabulary and therefore supports dissemination and usage of the TaDiRAH vocabulary. The web server is set up in a way that the concept URIs dissolves as expected. Relations between concepts of the new and the first version of TaDiRAH allow easy mapping and give a good overview on the differences between the versions. Changes at the new version are communicated by a

14 Vocabulary services (Vocabs), <https://vocabs.dariah.eu> and general information on DARIAH-EU, <https://www.dariah.eu> (Nov. 18, 2020)

15 Information on CLARIAH-AT, <https://clariah.at> (Nov. 18, 2020)

16 Download Skosmos, <http://skosmos.org> (Nov. 18, 2020)

17 Apache Jena Fuseki, <https://jena.apache.org/documentation/fuseki2/index.html> (Nov. 18, 2020)

18 For more information on the SPARQL interface, <https://vocabs.dariah.eu/en/about#sparql> (Nov. 18, 2020).

versioning system.¹⁹ The approach is to be as open and as transparent as possible. Therefore the RDF export of the vocabulary can be downloaded in a turtle format and due to the licensing of TaDiRAH vocabulary under the Creative Commons public domain license (CCo 1.0) reuse is highly appreciated.²⁰ All in all, the publication settings take care that the TaDiRAH vocabulary conforms to the FAIR data principles, such as the use of URIs, to uniquely identify the concepts, the providing of machine readable form of the model, using W3C Standard SKOS and the usage of CCo 1.0 licence.

4 Future ongoing tasks

After the revision and publication of the new TaDiRAH version, a workflow for further maintenance, development, supervision, community engagement, monitoring, and quality management has been designed and implemented.

A TaDiRAH board will be responsible for process management in the future.²¹ The board consists of the original core team, new developers and other contributors. The future tasks include the maintenance of the published model, for example by correcting errors and editing GitHub-issues²², and the further development of the model according to the needs and commitment of the community. In addition, the board will coordinate other translations, such as the Norwegian language, which is currently being developed, and will ensure technical and professional mentoring for the use and reuse of the model. Another important task of the board is the continuous documentation and dissemination of current information via the website²³ providing visualization of the model, but also in publications and via social media²⁴. Project

19 Using the relation **owl:versionInfo**, the version numbers are based on Semantic Versioning 2.0.0, see: <https://semver.org/> – additionally the relation **dct:modified** informs about the date of the last change.

20 CCo 1.0 Universal (CCo 1.0) Public Domain Dedication, <https://creativecommons.org/publicdomain/zero/1.0/> (Nov. 18, 2020)

21 See TaDiRAH website for further activities: <https://tadirah.info/> (Jan. 15, 2021).

22 <https://github.com/dhtaxonomy/TaDiRAH> (Jan. 15, 2021)

23 TaDiRAH as Linked Open Data, <https://tadirah.info/> (Nov. 18, 2020)

24 https://twitter.com/tadirah_dh (Nov. 18, 2020)

and publication lists will also be constantly maintained. Finally, the quality of the new TaDiRAH model depends on the commitment and involvement of the community, which is invited to participate in the revision of the multilingual terminology and scope notes, to submit new GitHub issues, and to continue the list of applications, users, and publications.

The community is also invited to further develop the model in terms of content and structure. For this purpose, an approach has been created which, by interlinking with Wikidata, enables a workflow in which missing definitions are iteratively supplemented, corrected and ideally transformed into high-quality scope notes. Current aggregated concepts link to Wikidata items, which partly do not yet contain any or only rudimentary definitions. At this point, the model can be further extended.

Another task for the future of TaDiRAH is to deepen the relationship between TaDiRAH and Wikidata. That means on the one hand to add definitions of terms or concepts provided by the taxonomy but not by Wikidata or to improve existing but deficient definitions on Wikidata. This will improve Wikidata as a provider of linked open data and expand its scope towards terms highly relevant within the domain of the digital humanities. On the other hand, that means to make good use of precise definitions given in Wikidata for TaDiRAH especially for its scope notes. This two-folded contrastive method will not only improve collections of controlled terms but also inspire further development of TaDiRAH as ideas for new concepts, terms, relations, or their reshaping, may be provoked. This will be a good step to link TaDiRAH and Wikidata data-wise and methodically and might serve as a best practice example for similar approaches.

References

- Borek, L., Dombrowski, Q., Perkins, J., & Schöch, C. (2016). TaDiRAH – A Case Study in Pragmatic Classification. *Digital Humanities Quarterly (DHQ)*, 10(1). <http://digitalhumanities.org/dhq/vol/10/1/000235/000235.html>
- Gruber, Thomas R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Hjørland, Birger (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8), 1519–1536.

- McCarty, W. (2003). Humanities Computing. In Drake, M. (Ed.), *Encyclopedia of Library and Information Science* (2nd ed., pp. 1224–1235). New York: Marcel Dekker.
- Schnapp, Jeffrey T. (2011). Emerging Disciplines \Rightarrow Knowledge Design. <http://jeff-reyschnapp.com/2011/08/26/finding-new-field-descriptions/>
- Suominen, O., Ylikotila, H., Pessala, S., Lappalainen, M., Frosterus, M., Tuominen, J. ... Retterath, A (2015). Publishing SKOS vocabularies with Skosmos. <http://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf>
- Unsworth, J. (2000a). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>
- Unsworth, J. (2000b). What is Humanities Computing and What is Not? <http://www.people.virginia.edu/~jmu2m/mith.00.html>
- Zaytseva, K., & Ďurčo, M. (2020). Controlled Vocabularies and SKOS. Version 1.0.0. DARIAH-Campus [Training module]. <https://campus.dariah.eu/resource/controlled-vocabularies-and-skos>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 321–332. DOI: doi.org/10.5283/epub.44951.

Vom Datenkatalog zum Wissensgraph – Forschungsdaten im konzeptuellen Modell von FRBR

Ingo Frank

Leibniz-Institut für Ost- und Südosteuropaforschung,
Regensburg, Deutschland
frank@ios-regensburg.de

5

Abstract

Die Beschreibung von Forschungsdaten im Bereich des Forschungsdatenmanagements ist oft ungenau, unvollständig oder nicht konsistent bzw. leidet unter einer nicht konsequent durchgeführten Metadatenkuration. Der Beitrag stellt einen bibliotheks- und informationswissenschaftlich motivierten Ansatz vor, wie Metadaten über Forschungsdaten mithilfe des konzeptuellen Modells von FRBR (Functional Requirements for Bibliographic Records) verbessert werden können. Das konkrete Ziel dabei ist die Konstruktion eines Wissensgraphen, der FRBRisierte Metadaten aus einem Datenkatalog mit Metadaten aus einem Bibliothekskatalog sowie mit Forschungsinformation integriert. Die Methode baut auf einem Datenkatalog mit einem auf DCAT (Data Catalog Vocabulary) und Disco (DDI-RDF Discovery Vocabulary) basierenden Anwendungsprofil als Metadaten-schema auf. Die Metadaten im Datenkatalog werden mit SHACL (Shapes Constraint Language) validiert und dienen somit als Grundlage für die FRBRisierung zum Aufbau des Wissensgraphen mit FaBiO (FRBR-aligned Bibliographic Ontology) als FRBR-basiertem Datenmodell. Die FRBRisierten und integrierten Metadaten im Wissensgraphen unterstützen schließlich aufgrund der besseren Metadatenqualität und der Vorgehensweise zur Verlinkung von Entitäten aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem insbesondere die Versionierung und Provenienzinformation von Forschungsdaten und nicht zuletzt auch die Datenzitation. Der FRBRisierungsansatz trägt dadurch zur Verbesserung des Information Retrieval in Discovery-Systemen für Forschungsdaten bei.

Keywords: Forschungsdatenmanagement; FRBR; Informationsintegration; Wissensgraphen; Metadatenqualität; Digital Humanities; Forschungsdateninfrastruktur

1 Einleitung

Daten bzw. Forschungsdaten werden in der Forschungsdatenmanagement-Community oft unzureichend definiert (s. Hjørland, 2018). Problematisch ist auch die verbreitete Auffassung, dass Daten die Grundlage für Information und Wissen bilden, wovon im Modell der Wissenspyramide ausgegangen wird (s. Frické, 2019). Dazu kommt die häufig vorkommende Verwechslung von Daten und (Daten-)Dokument (s. Schöpfel et al., 2020) – was sich alles negativ auf die Qualität der Metadaten zur Beschreibung von Forschungsdaten auswirken kann. Wir wollen mit unserem kurzen Arbeitsbericht (work in progress) zeigen, welchen bibliotheks- und informationswissenschaftlichen Beitrag man im konzeptuellen Rahmenwerk von FRBR (Functional Requirements for Bibliographic Records) zur Verbesserung der Metadatenqualität und des Information Retrieval im Forschungsdatenmanagement leisten kann.

Datenkataloge, die DCAT (Data Catalog Vocabulary) verwenden, stellen Metadaten über Datensätze und deren Distribution in verschiedenen Formaten oder über verschiedene Dienste bereit. Trotz Best Practices zur Forschungsdatenkuratation gelingt diese Trennung nicht immer präzise. Bibliothekskataloge stellen Metadaten über Publikationen in der Regel auf der Ebene von FRBRs *Manifestation*-Begriff bereit. Datensätze in DCAT-Datenkatalogen sind allerdings mehr oder weniger auf *FRBR-Expression*-Ebene zu verorten, was ein weiteres Problem für die Integration von Information aus Datenkatalogen und Bibliothekskatalogen darstellt. Mit der FRBRisierung der Metadaten wollen wir diese Probleme angehen und die implizit vorhandenen Unterscheidungen von Entitäten und deren Zusammenhänge explizit machen und exakt erfassen, um die Qualität der in einem Wissensgraphen integrierten Metadaten zu verbessern.

Im Rahmen des institutionellen Forschungsdatenmanagements und des Aufbaus eines institutionellen Repositoriums, das sowohl Digitalisate, Publikationen, aber eben auch Forschungsdaten einschließt, besteht unsere konkrete Anforderung darin, den Datenkatalog mit dem Bibliothekskatalog und auch mit Forschungsinformation zu integrieren. Im Prinzip geht es darum, die für das institutionelle Repositorium relevante Information in einem Wissensgraphen zu integrieren.¹ Unsere Vorgehensweise zur Integration von

¹ Hintergrund bzw. Projektkontext ist die Integration des institutionellen Forschungsdatenrepositoriums LaMBDA mit einem im Aufbau befindlichen Informationsintegra-

Metadaten über Forschungsdaten aus den verschiedenen Quellen in einem ‚Institutional Knowledge Graph‘ wird im nächsten Abschnitt dargestellt.

2 Vorgehensweise

*FRBR is not a data model. FRBR is not a metadata scheme.
FRBR is not a system design structure.
It is a conceptual model of the bibliographic universe.*
– Tillett (2005)

Ausgehend von einem mit SHACL (Shapes Constraint Language) validierten Datenkatalog wird dieser FRBRisiert und schließlich mit zusätzlicher Information aus dem Bibliothekskatalog und dem Forschungsinformationssystem angereichert und in einem Wissensgraphen integriert. FRBR verwenden wir dabei als konzeptuelles Modell des bibliografischen Bereichs. Als konkretes Datenmodell zur Modellierung der FRBRisierten Metadaten verwenden wir die FRBR-aligned Bibliographic Ontology (FaBiO) (Peroni/Shotton, 2012).

2.1 DCAT- und Disco-basiertes Anwendungsprofil als Metadatenschema für erweiterte Datenkataloge

Zur Beschreibung von Forschungsdaten im institutionellen Forschungsdaten-repositorium haben wir ein im Wesentlichen auf DCAT basierendes Anwendungsprofil entworfen. Wesentliche Grundlage des Anwendungsprofils sind die Klassen `dc:Dataset` und `dc:Distribution` und deren für uns relevante Eigenschaften, die aber keine Information über Erhebungsmethoden etc. vorsehen. Heery und Patel (2000) definieren Anwendungsprofile als Metadatenschemas, “which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application”. Damit wir demnach methodologische Information über die Erhebung bzw. Erstellung der Forschungsdaten erfassen können, haben wir das Metadatenschema um Eigenschaften aus Disco (DDI-RDF Discovery Vocabulary) (Bosch et al., 2013) erweitert. Disco legt den Schwer-

punkt auf die Unterstützung der für das Information Retrieval in Discovery-Systemen relevanten Elemente aus dem DDI-Metadatenstandard.

Um beim Aufbau unseres Metadatenschemas keine ‚Frankenstein-Ontologie‘ (vgl. Corcho et al., 2015) zu entwerfen, bei der – womöglich lediglich von der Bezeichnung her passende – Klassen und Eigenschaften aus bereits vorhandenen Ontologien zusammengesucht werden, verwenden wir SHACL zur Vermeidung von Inkonsistenzen und zur Validierung der Metadaten. Der folgende Auszug aus den SHACL-Constraints zeigt als Beispiel die Festlegung der gemäß Forschungsdatenmanagement-Policy im Forschungsdaten-Repositorium möglichen Zugriffsrechte auf Forschungsdaten in einem *property shape*.

```
sh:property [
  sh:name "Availability of Data" ;
  sh:description "Access rights or restrictions to data
    distributions" ;
  sh:path dct:accessRights ;
  sh:class skos:Concept ;
  sh:nodeKind sh:IRI ;
  sh:in (
    <http://purl.org/coar/access_right/c_abf2>
    <http://purl.org/coar/access_right/c_flcf>
    <http://purl.org/coar/access_right/c_16ec>
    <http://purl.org/coar/access_right/c_14cb>
  ) ;
  sh:maxCount 1 ;
] ;
```

Mit dieser *property shape* wird eine bestimmte Auswahl an Zugriffsrechten aus dem *Access Rights*-Vokabular der COAR Controlled Vocabularies festgelegt. Auf ähnliche Weise können mit SHACL auch Einschränkungen formuliert werden, die festlegen, welche Wissensorganisationssysteme zum Einsatz kommen können. So haben wir etwa festgelegt, dass die Erhebungs- bzw. Erstellungsmethoden anhand der Eigenschaft `disco:collection-Mode` nicht nur mit dem auf die Sozial- und Wirtschaftswissenschaften zugeschnittenen *Mode of Collection*-Vokabular der DDI Controlled Vocabularies (Jaaskelainen et al., 2010), sondern auch mit Begriffen aus der TaDiRAH (Taxonomy of Digital Research Activities in the Humanities)² beschrieben werden können, um den (digitalen) Geisteswissenschaften entgegenzukommen.

2 TaDiRAH im Skosmos-Browser des ACDH-CH: <https://vocabs.dariah.eu/tadirah2/en/>

Wir gelangen so zu einer interdisziplinären Wissensorganisation – im Prinzip einer Facettenklassifikation – zur Erfassung und Beschreibung der Untersuchungsgegenstände, Erhebungs-/Erstellungsmethoden sowie außerdem dem theoretischen und disziplinären Hintergrund. Diese methodologische Information geht über die üblichen bibliografischen Angaben hinaus, kann aber anhand entsprechender SHACL-Festlegungen auch auf die – im nächsten Schritt – FRBRisierten Metadaten, d.h. `fabio:DataFile` auf FRBR-Expression-Ebene, eingebunden werden.

2.2 FRBRisierung des Datenkatalogs mit FaBiO

Zur FRBRisierung der DCAT-Entitäten `dcat:Dataset` und `dcat:Distribution` wenden wir FaBiO an.³ FaBiO ist gerade bei der Integration in einen Wissensgraphen deutlich einfacher zu handhaben (s. Abb. 1) als die Ontologie FRBRoo (Functional Requirements for Bibliographic Records object-oriented)⁴, die auf ereignisbasierter Modellierung mit CIDOC CRM (Doerr, 2003) basiert.

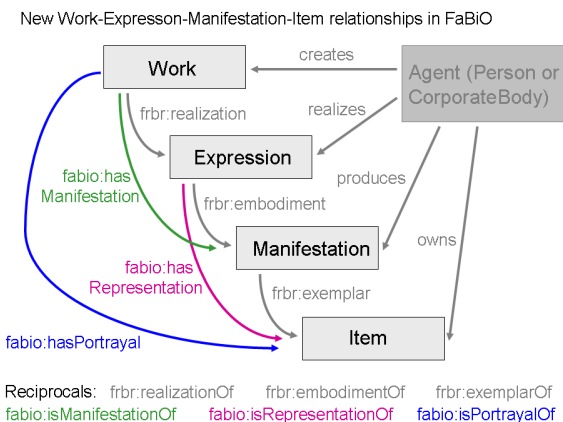


Abb. 1 Die FRBR-Konzepte *Work*, *Expression*, *Manifestation* und *Item* (WEMI-Entitäten) in FaBiO (Diagramm aus der Online-Dokumentation unter <http://purl.org/spar/fabio/2019-02-19>)

- 3 Das Metadatenschema von DataCite hat diese Unterscheidung übrigens nicht, was die Indexierung von DCAT-basierten Datenkatalogen in DataCite auf die Datensatz-Ebene (bzw. FRBR-Expression-Ebene) einschränkt. Dieses Problem soll hier aber nicht weiter behandelt werden.
- 4 Siehe OWL-Implementierung von FRBRoo unter <http://erlangen-crm.org/efrbroo>.

Genau genommen sollte eine `dcat:Distribution` die Daten eines Datensatzes repräsentieren. Bei mehreren Datendistributionen kann sich somit das Datenformat unterscheiden, aber nicht der Inhalt, d.h. die Daten selbst.⁵ Distributionen werden über die Eigenschaft `dcat:distribution` mit dem Datensatz, den sie in verschiedenen Formaten repräsentieren, verbunden. Verbindungen zu anderen Publikationen oder auch zur Datendokumentation sollten über die Eigenschaft `dcat:qualifiedRelation` jeweils als `dcat:Relationship` explizit gemacht werden.

Eine `dcat:Distribution` kann daher als `fabio:DigitalManifestation` (Unterklasse von `fabio:Manifestation`) aufgefasst werden. Wie bereits im vorigen Abschnitt angedeutet, betrachten wir ein `dcat:Dataset` mitsamt den zugehörigen bibliografischen, aber auch methodologischen Metadaten als `fabio:DataFile`. Der Klassenname ‚DataFile‘ ist etwas irreführend: Ein `fabio:DataFile` ist jedoch eine Unterklasse von `fabio:Expression` und als “realisation of a `fabio:Dataset` (a `frbr:Work`) containing a defined collection of data with specific content and possibly with a specific version number, that can be embodied as a `fabio:Digital Manifestation` (a `frbr:Manifestation` with a specific format)”⁶ definiert.

Ein typischer Anwendungsfall für den Ansatz ist die Versionierung von Datensätzen. Wie in Abbildung 2 zu sehen ist, ergibt sich durch den Einsatz von FRBR-Klassen – z.B. durch die Einbindung von `dcat:Dataset` und `dcat:Distribution` als Unterklassen von `fabio:DataFile` und `fabio:DigitalManifestation` – der Vorteil, dass der Zusammenhang von Versionen eines Datensatzes (als Werk) explizit gemacht werden kann. Das ist insbesondere für das Information Retrieval in einem Discovery-System relevant, um dem Benutzer in den Suchergebnissen oder der Navigation den Zugriff auf alle Versionen eines Datensatz zu ermöglichen. Bei der FRBRisierung des Datenkatalogs können die *Work*-Entitäten (`fabio:Dataset`) anhand der typisierten Relationen zwischen den versionierten Datensätze (`dcat:Dataset`) abgeleitet und generiert werden.

5 Vgl. dazu die DCAT-Dokumentation unter <https://www.w3.org/TR/vocab-dcat-2/#Class:Distribution>.

6 Siehe FaBiO-Dokumentation unter <https://sparantologies.github.io/fabio/current/fabio.html#d4e2752>.

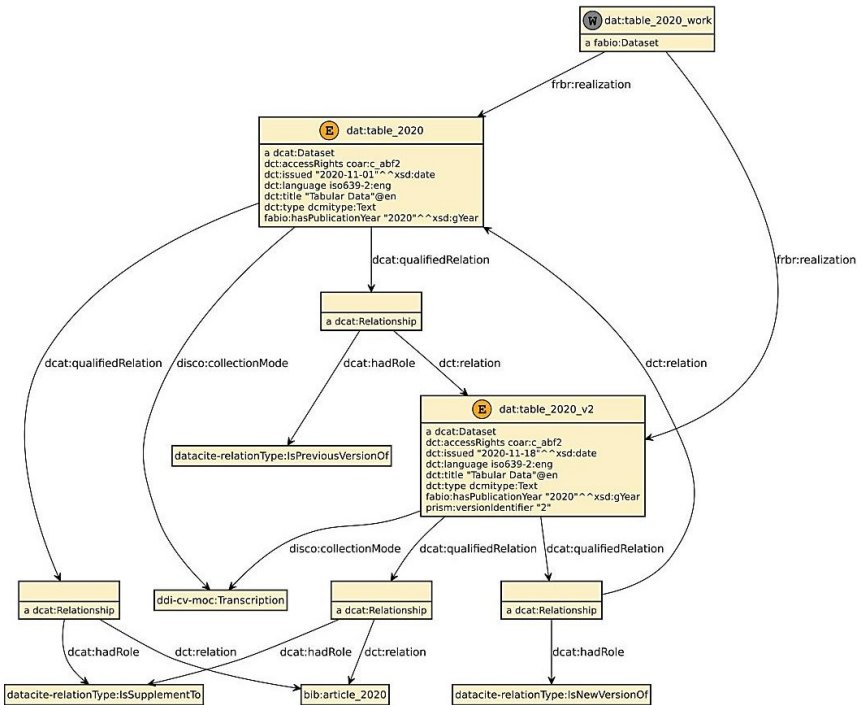


Abb. 2 DCAT hat keine Klasse zur Repräsentation eines Werkes im Sinne von FRBR. Deshalb kann in DCAT ein Datensatz nicht als Entität auf FRBR-Work-Ebene identifiziert werden, der z. B. in zwei Expressions realisiert ist, wie etwa zwei Versionen des Datensatzes. (Die Metadaten im Diagramm sind aus Platzgründen und zur übersichtlicheren Visualisierung nicht vollständig dargestellt. Es fehlen z. B. eigentlich notwendige Angaben zu Autoren, Sacherschließung, DOIs, etc.)

Einen vergleichbaren Ansatz zur FRBRisierung von Datenkatalogen hat bereits McCusker et al. (2012) vorgeschlagen. Die Autoren wollten damit die Provenienzinformation für integrierte offene Verwaltungsdaten verbessern.⁷ Sie haben dazu die FRBR-Konzepte angewendet, um zwischen der Transformation des Inhalts und der Transformation des Formats der Daten unter-

7 Hourlé (2008) verwendet das konzeptuelle Model von FRBR, um die Verwechslung zwischen Daten und digitaler Repräsentation der Daten bei wissenschaftlichen Daten – also Forschungsdaten – aufzuklären. Klump et al. (2020) schlagen einen aktuellen Ansatz vor, um mit FRBR ebenfalls die Metadaten über die Versionierung von Forschungsdaten zu verbessern.

scheiden zu können (s. Abb. 3). McCusker et al. (ebd., S. 57) beschreiben und begründen ihren FRBRisierungsansatz wie folgt:

FRBR's four levels of abstraction also naturally apply to electronic information sources. Copies of files (Items) are exemplars of the same Manifestation (byte sequence). Similarly, an Excel file created from a CSV will have a different Manifestation, but maintain the same Expression because they both store the same data. Finally, if the data is modified, the original and resulting spreadsheets have different Expressions (visual content) of the same abstract Work.

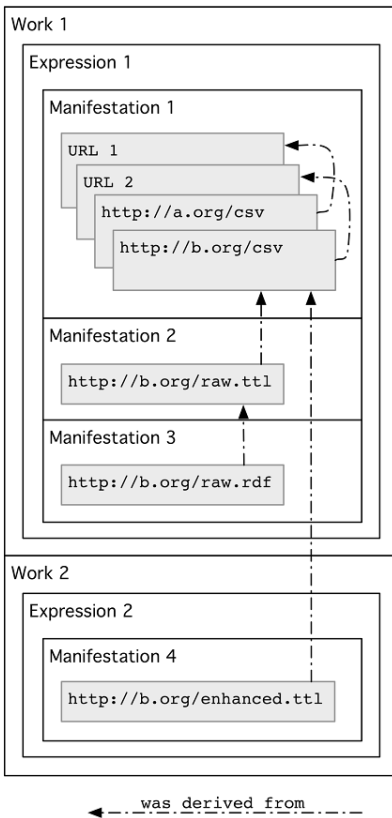


Abb. 3 Die WEMI-Entitäten von FRBR ermöglichen z.B. die Repräsentation eines Datensatzes (Work 1) und seiner (digitalen) Manifestationen sowie eines weiteren Datensatzes (Work 2) mit neuem *Inhalt* (Expression 2 von Work 2), der durch die erweiterte Manifestation 4 verkörpert wird. Diese liegt im Turtle-Format vor und wurde von der CSV-Repräsentation des ersten Datensatzes abgeleitet, was im Beispiel durch die *was derived from*-Relation ausgedrückt wird (Diagramm aus McCusker et al. 2012).

2.3 Integration von Datenkatalog, Bibliothekskatalog und Forschungsinformation in einem Wissensgraph

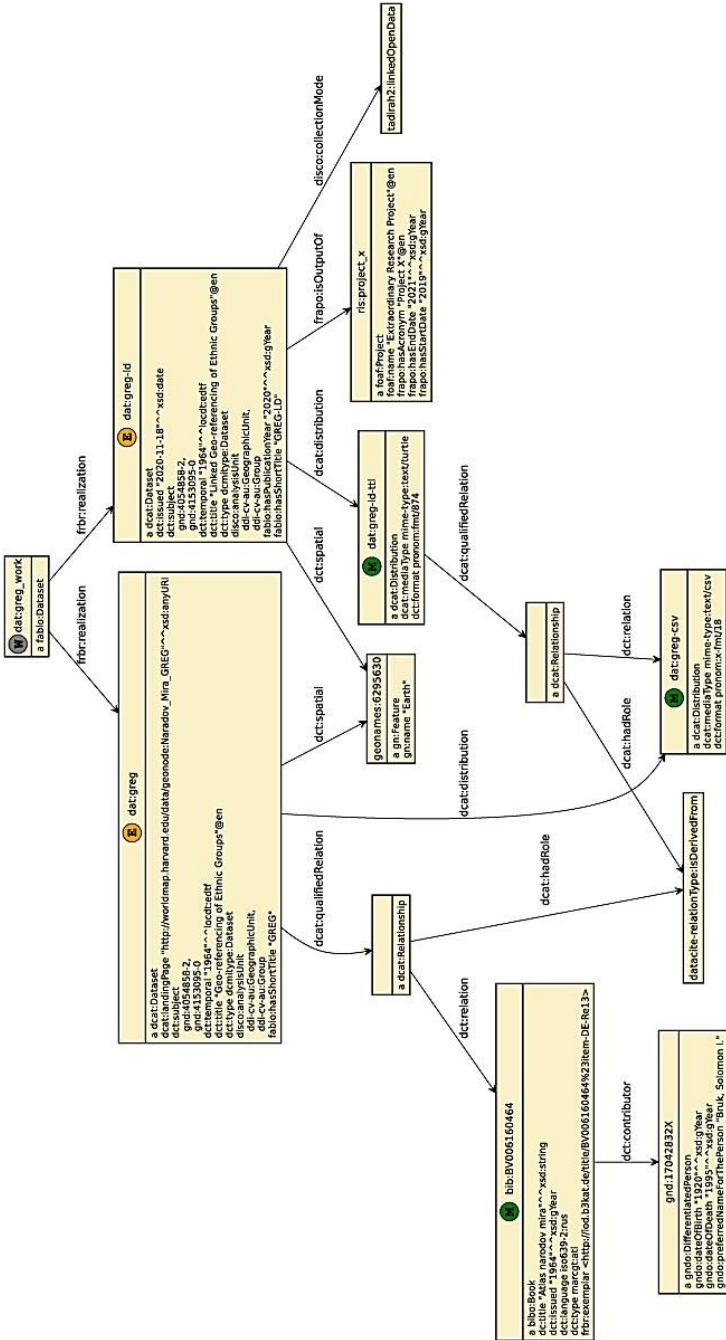
Nach der Definition von Ontotext (2020) repräsentiert ein Wissensgraph “a collection of interlinked descriptions of entities – real-world objects and events, or abstract concepts (e.g., documents)”. Mit Wallis (2013) kann man daher unsere Vorgehensweise auch sehr treffend mit “From Cataloging to Catalinking” bezeichnen. Unser Ansatz zur Konstruktion des Wissensgraphen ist somit die Verlinkung der relevanten Entitäten, die mit verschiedenen Ontologien repräsentiert werden:

- DCAT u. Disco: Datensätze und deren Datendistributionen im Datenkatalog
- BIBO u. FRBR: Publikationen im Bibliothekskatalog
- FRAPO u. FOAF: Forschungsinformation über Projekte, Drittmittelförderung, Forschungoutput, etc.
- FaBiO: FRBRisierte Entitäten im Wissensgraph.

Forschungsaktivitäten im Forschungsdatenlebenszyklus, wie z.B. die für Digital-Humanities-Projekte typischen Digitalisierungsprozesse, können auf einfache Weise mit der PROvenance Interchange Ontology (PROV-O) (Lebo et al., 2013) als `prov:Activity` dokumentiert werden. Die Forschungsinformation kann im Kontext des Forschungsdatenmanagements noch granularer aufbereitet werden, indem etwa die Scholarly Ontology (SO) (Pertsas/Constantopoulos 2017) zur ereignisbasierten Modellierung zum Einsatz kommt. Der Life Cycle of Historical Information (Meroño-Peñuela et al., 2014) aus dem Bereich der Historical Information Science (s. ebd.) ist auf typische Digital-History-Projekte zugeschnitten und eignet sich dabei als fachspezifischer Forschungsdatenlebenszyklus zusammen mit der TaDiRAH zur Klassifikation der Forschungsaktivitäten und -methoden besser für das Forschungsdatenmanagement entsprechender Projekte als der allgemeine Forschungsdatenlebenszyklus.

Abschließend betrachtet kann aufgrund der FRBRisierten Metadaten zu den Datensätzen die Datenzitation wesentlich genauer erfolgen, indem auf die im Wissensgraph integrierte Information aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem zurückgegriffen werden kann (s. Abb. 4).

Abb. 4 Beispiel zu FRBRisierten und integrierten Metadaten aus Datenkatalog (Präfix `dat:`), Bibliothekskatalog (Präfix `bib:`) und Forschungsinformationssystem (Präfix `ris:`) mit verbesserter Provenienzinformation als Grundlage für genauere Datenzitation (Im Beispiel wird davon ausgegangen, dass die Aufbereitung des GREG-Datensatzes als Linked Data noch kein neues Werk ausmacht, sondern lediglich eine neue FRBR-Expression bildet. Die Frage nach der Schöpfungshöhe, die ein neues Werk ausmacht, ist ggf. individuell zu diskutieren.) →



3 Zusammenfassung und Ausblick

Wir haben gezeigt, wie mithilfe des konzeptuellen Modells von FRBR und FaBiO als FRBR-basiertem Datenmodell ein ‚Institutional Knowledge Graph‘ aus Metadaten aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem konstruiert werden kann. FRBR dient dabei beim Übergang von Metadaten zu Wissensgraphen nicht nur als konzeptuelles Rahmenwerk, sondern quasi auch als Werkzeug zur Begriffsklärung – um auf den Beginn der Einleitung zurückzukommen: Was sind ‚Forschungsdaten‘? In welchen ‚Dokumenten‘ werden Forschungsdaten beschrieben und repräsentiert?

Der Mehrwert der bibliotheks- und informationswissenschaftlich informierten FRBRisierung und Integration von Metadaten in einem Wissensgraphen besteht zusammengefasst in

- der Möglichkeit zur Erfassung von Versionen (d. h. *FRBR-Expressions*) eines Datensatzes (als *Werk*) für das Information Retrieval,
- der erweiterten Information über den Entstehungskontext und die Quellen von Forschungsdaten anhand der Modellierung verschiedener – z. B. in verschiedenen Digitalisierungsprozessen erstellter – digitaler *Manifestationen* eines Datensatzes und
- nicht zuletzt der Verbesserung der Datenzitation aufgrund von detaillierterer Information über Datensätze und deren verschiedene *Datendistributionen* bzw. *-manifestationen*.

Als Implikation für die Entwicklung von Forschungsdatenmanagementsystemen ist allerdings festzuhalten, dass erst auf Semantic-Web-Basis eine direkte Verwendung der Metadatenstandards DCAT, Disco, FaBiO, etc. und die konsequente Nutzung von einheitlichen SKOS-Vokabularen zur interdisziplinären Wissensorganisation möglich ist.⁸ Dadurch wird die automatische Evaluierung der Konsistenz und Vollständigkeit der Metadaten mit SHACL ermöglicht.⁹ Ebenso kann die aus verschiedenen Quellen zusammengetragene

8 Eine Datenmanagementplattform, die direkt Semantic-Web- und Linked-Data-Technologie unterstützt, wird derzeit von Kirstein et al. (2020) entwickelt.

9 Die Wichtigkeit der Erfüllung der klassischen Qualitätskriterien für Metadaten – Korrektheit, Vollständigkeit und Konsistenz (s. Park 2009) – zeigt sich insbesondere auch in der gegenwärtigen Pandemie in der Schwierigkeit, aufgrund von fehlenden, inkonsistenten oder gar falschen Metadaten, gezielt nach relevanten klinischen Stu-

ne Information im Wissensgraphen mit SHACL evaluiert werden. Wissensgraphen eröffnen schließlich auch Möglichkeiten zur Verbesserung der Usability. Sehr naheliegend ist dazu die Verknüpfung von Forschungsdatensätzen und in Beziehung stehender Publikationen mittels eines interaktiven Visualisierungswerkzeugs zur Forschungsdatenkuration.¹⁰

Literatur

- Bosch, T.; Cyganiak, R.; Gregory, A.; Wackerow, J. (2013): DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data. In: *Proceedings of the WWW2013 Workshop on Linked Data on the Web. CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-996/papers/ldow2013-paper-12.pdf>
- Corcho, O.; Poveda-Villalón, M.; Gómez-Pérez, A. (2015): Ontology Engineering in the Era of Linked Data. *Bulletin of the Association for Information Science and Technology* 41 (4), 13–17. doi: [10.1002/bult.2015.1720410407](https://doi.org/10.1002/bult.2015.1720410407)
- Doerr, M. (2003): The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24 (3), 75–92.
- Frické, M. (2019): The knowledge pyramid: the DIKW hierarchy. *Knowledge Organization* 49 (1), 33–36.
- Heery, R.; Patel, M. (2000): Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne* (25). <http://www.ariadne.ac.uk/issue/25/appprofiles/>
- Hjørland, B. (2018): Data (with big data and database semantics). *Knowledge Organization* 45 (8), 685–708.
- Hourclé, J. A. (2008): FRBR applied to scientific data. *Proceedings of the American Society for Information Science and Technology* 45 (1), 1–4. doi: [10.1002/meet.2008.14504503102](https://doi.org/10.1002/meet.2008.14504503102)

dien oder Daten (z. B. für die Durchführung von Metastudien) zu suchen (vgl. Schriml et al., 2020). In Anlehnung an den Artikel „Medical Usability: How to Kill Patients Through Bad Design“ von Nielsen (2005) kann man das Problem mangelhafter Metadatenqualität hier treffend mit ‘How to Kill Patients Through Bad Metadata’ beschreiben.

¹⁰ Eine entsprechende Implementierung scheint es bisher ansatzweise nur für das Datenmanagementsystem CKAN zu geben: Milić et al. (2015) schlagen die Architektur Linked Relations (LIRE) zur visuellen Erstellung von Verknüpfungen von Datensätzen vor.

- Jaaskelainen, T.; Moschner, M.; Wackerow, J. (2010): Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability. *IASSIST Quarterly* 33 (1–2). <https://iassistquarterly.com/index.php/iassist/article/view/649>
- Kirstein, F.; Stefanidis, K.; Dittwald, B.; Dutkowski, S.; Urbanek, S.; Hauswirth, M. (2020): Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies. In: A. Harth et al. (Hrsg.): *The Semantic Web – 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings* (S. 648–664). Cham: Springer International Publishing. doi: [10.1007/978-3-030-49461-2_38](https://doi.org/10.1007/978-3-030-49461-2_38)
- Klump, J.; Wyborn, L.; Downs, R.; Asmi, A.; Wu, M.; Ryder, G.; Martin, J. (2020): Principles and best practices in data versioning for all data sets big and small. Revised Report of the Research Data Alliance Data Versioning Working Group. doi: [10.15497/RDA00042](https://doi.org/10.15497/RDA00042)
- Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D. ... Zhao, J. (2013): PROV-O: The PROV Ontology. W3C Recommendation. World Wide Web Consortium. <http://www.w3.org/TR/2013/REC-prov-o-20130430/>
- McCusker, J.; Lebo, T.; Chang, C.; McGuinness, D. L.; McGuinness, D. (2012): Parallel Identities for Managing Open Government Data. *IEEE Intelligent Systems* 27 (3), 55–62. doi: [10.1109/MIS.2012.5](https://doi.org/10.1109/MIS.2012.5)
- Meroño-Peñuela, A.; Ashkpour, A.; van Erp, M.; Mandemakers, K.; Breure, L.; Scharnhorst, A. ... van Harmelen, F. (2014): Semantic Technologies for Historical Research: A Survey. *Semantic Web* 6, 539–564.
- Milić, P.; Veljković, N.; Stoimenov, L. (2015): Linked Relations Architecture for Production and Consumption of Linksets in Open Government Data. In: *Open and Big Data Management and Innovation. 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015* (S. 212–222). Cham: Springer International Publishing. doi: [10.1007/978-3-319-25013-7_17](https://doi.org/10.1007/978-3-319-25013-7_17)
- Nielsen, J. (2005): Medical Usability: How to Kill Patients Through Bad Design. <https://www.nngroup.com/articles/medical-usability/>
- Ontotext (2020): What is a Knowledge Graph? Ontotext Knowledge Hub. Ontotext Fundamentals. <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>
- Park, J.-R. (2009): Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging and Classification Quarterly* 74 (3–4), 213–228. doi: [10.1080/01639370902737240](https://doi.org/10.1080/01639370902737240)
- Peroni, S.; Shotton, D. (2012): FaBIO and CiTO: Ontologies for Describing Bibliographic Resources and Citations. *Journal of Web Semantics First Look* 17, 33–43. doi: [10.1016/j.websem.2012.08.001](https://doi.org/10.1016/j.websem.2012.08.001)

- Pertsas, V.; Constantopoulos, P. (2017): Scholarly Ontology: modelling scholarly practices. *International Journal on Digital Libraries* 18, 173–190. doi: [10.1007/s00799-016-0169-3](https://doi.org/10.1007/s00799-016-0169-3)
- Schöpfel, J.; Farace, D.; Prost, H.; Zane, A.; Hjørland, B. (2020): Data documents. In: B. Hjørland, C. Gnoli (Hrsg.): *Encyclopedia of Knowledge Organization*. http://www.isko.org/cyclo/data_documents
- Schriml, L. M.; Davies, M. C. N.; Eloë-Fadrosh, E. A.; Finn, R. D.; Hugenholtz, P.; Hunter, C. I. ... Walls, R. (2020): COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data* 7. doi: [10.1038/s41597-020-0524-5](https://doi.org/10.1038/s41597-020-0524-5)
- Tillett, B. (2005): FRBR and Cataloging for the Future. *Cataloging and Classification Quarterly* 39 (3), 197–205.
- Wallis, R. (2013): Linked Data for Libraries: Great Progress, but What Is the Benefit? In: *SWIB13 – Semantic Web in Libraries Conference*.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 333–346. DOI: doi.org/10.5283/epub.44952.

Session 6:
Emerging Technologies

A Living Lab Architecture for Reproducible Shared Task Experimentation

Timo Breuer, Philipp Schaer

TH Köln – University of Applied Sciences, Germany
{[timo.breuer](mailto:timo.breuer@th-koeln.de), [philipp.schaer](mailto:philipp.schaer@th-koeln.de)}@th-koeln.de

Abstract

No existing evaluation infrastructure for shared tasks currently supports both reproducible on- and offline experiments. In this work, we present an architecture that ties together both types of experiments with a focus on reproducibility. The readers are provided with a technical description of the infrastructure and details of how to contribute their own experiments to upcoming evaluation tasks.

Keywords: reproducibility; evaluation infrastructure; shared tasks

1 Introduction

Experimental evaluation has a long history in Information Retrieval (IR) and Recommender System (RecSys) research and is the main driver of innovations and scientific progress in these fields. Shared task initiatives bring together community efforts and offer a platform for sharing and comparing different ideas and approaches. Especially in the field of IR, most experiments are based on test collections (Sanderson, 2010) with topical relevance judgments. Since pertinent relevance decisions are not covered, we can consider these in vitro approaches as *offline* experiments.

Very often, evaluations are solely based on system-oriented measures, and *online* experiments with real users are neglected. Opposed to this, Interactive Information Retrieval (IIR) experiments have a user-oriented focus (Kelly, 2007). Even though user-related aspects can be investigated in a very controlled manner, these experiments are costly, and thus are usually conducted on a smaller scale. As a compromise, session logs from online field experi-

ments allow us to include user interactions (as part of the evaluations) on a larger scale at the price of control over user characteristics.

A recent survey on shared task platforms shows, none of existing infrastructures ties together both worlds of on- and offline evaluations and additionally guarantees reproducible experiments (Schaible et al., 2020). In this sense, we introduce a novel architecture that complements the portfolio of the existing evaluation infrastructures by making online as well as offline evaluations fully reproducible. In the following, we broadly outline this infrastructure for which an overview is provided in Figure 1.

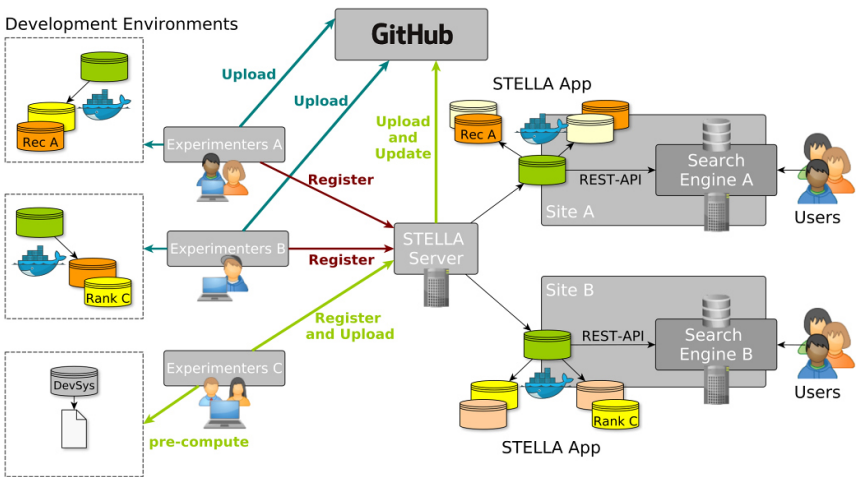


Fig. 1 Overview of the architecture

Our framework mainly relies on Docker and its containerization technology. An additional component is Git (or GitHub) as an attempt to make the whole infrastructure as transparent as possible. This is within the spirit of OpenScience and lets us focus on versioning, openness and reproducibility of runs and systems.

Experimenters provide systems with retrieval and recommendation algorithms and package them in Docker containers that are integrated into a multi container application (MCA). Multiple systems from possibly different experimenters are incorporated into the MCA. Search engine providers are referred to as sites. They provide access to data collections and, most importantly, user interaction data. Each site will deploy one instance of the MCA on their backend servers. Via a REST-API the users’ queries can be conducted to the experimental systems. In return, the rankings and recom-

mendations are sent back. Sites will then log user interactions and send this feedback to the MCA. Eventually, the MCA sends the user interaction data to the central server where it is stored and can be used for further analysis, training, and optimization of the experimental systems.

The remainder is organized as follows. At first, we introduce each main component of our infrastructure and its functionalities, starting with the container, and continuing with the MCA and the central server. Afterward, we revise existing work and previous attempts to provide newcomers with the current state-of-the-art in the field of living lab experimentation. Finally, we conclude and end with future work.

2 The micro-services

As mentioned before, experimenters contribute their IR and RecSys algorithms as micro-services (or containers) that are integrated into the MCA. At the current state, the infrastructure provides two ways of contributing the corresponding systems or their outputs. In the following, we outline both to give the reader an impression of which one fits best to their evaluation practices.

2.1 Pre-computed results

Instead of submitting the entire IR or RecSys application, only its results are contributed for experimentation as illustrated in Figure 2. Like in previous living lab attempts (Jagerman et al., 2018), the pre-computed results are restricted to a specific set of queries that is extracted from the top- k results of query logs. For these head queries, rankings and recommendations must be derived from a given document collection (that can be retrieved after registration). In order to make the transition from offline to online evaluations as smooth as possible, these results are contributed with TREC run file syntax as illustrated below.

```
<qid> <Q0> <docid> <rank> <score> <identifier>
```

Run file syntax used for precomputed results

Each line contains the numeric query identifier (<qid>), a string identifying the document (<docid>), an increasing rank number (<rank>), the corresponding score (<score>), and the tag chosen by the experimenters (<identifier>). The files have to be uploaded to the central server either with the help of HTTP requests (via the REST-API) or via the user interface. The infrastructure service will automatically prepare the uploaded files for integration into the MCA after they have been checked for validity and consistency.

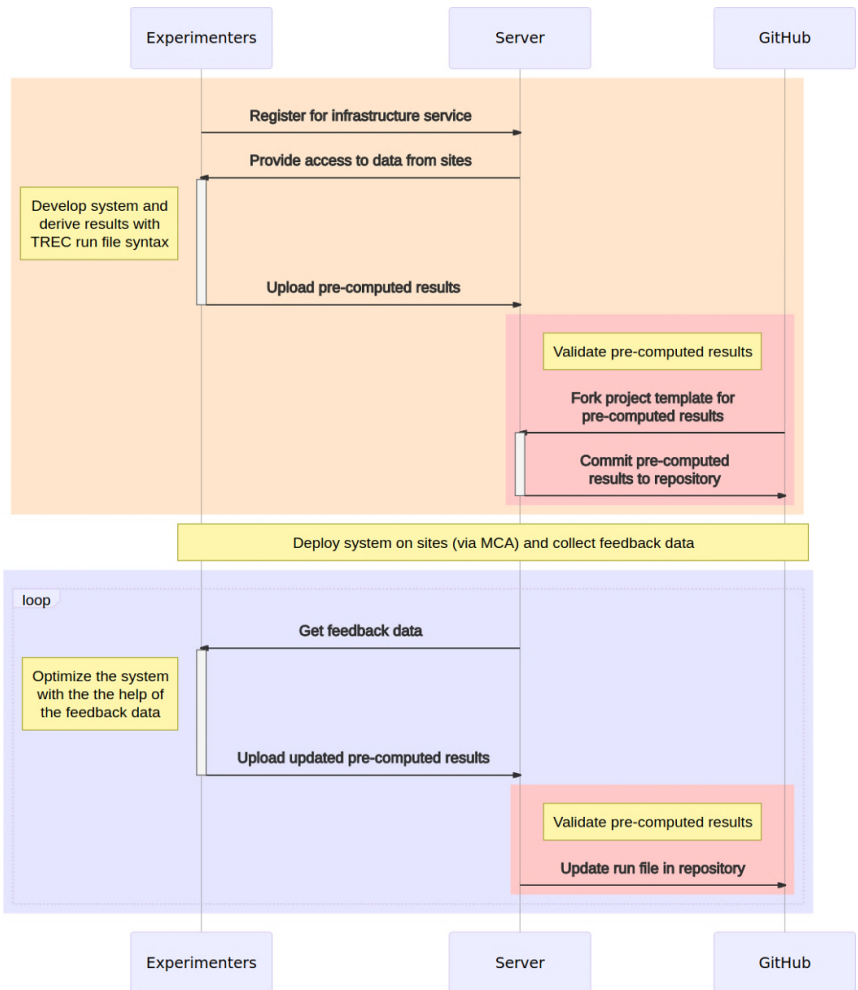


Fig. 2 Sequence diagram for contributing precomputed results

More specifically, a new micro-service is automatically set up and integrated into the MCA (cf. to the red boxes in Fig. 2). Furthermore, this submission type makes it possible to directly transfer over results from purely offline to living lab environments, facilitating the comparison of system-oriented evaluation measures to those inferred from, e.g., click feedback.

2.2 Encapsulated systems

Alternatively, experimenters can choose to submit the entire IR or RecSys application instead of systems' outputs only. In this case, the applications are contributed as individual micro-services, as illustrated in Figure 3.

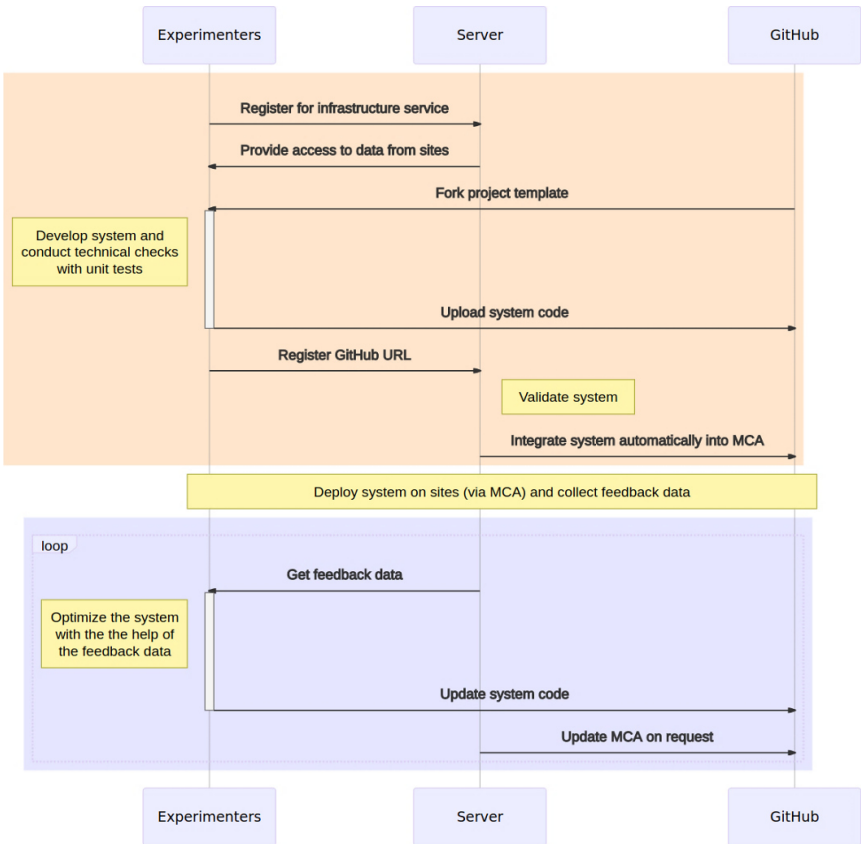


Fig. 3 Sequence diagram for contributing encapsulated systems

A project template in the form of a minimal Docker application is provided for the sake of compatibility between experimental systems and the MCA. Experimenters adapt these templates when integrating their applications. As a starting point, the project templates provide a minimal Python application based on the Flask web framework. But, of course, experimenters are not restricted to use Python at all and can modify the template as they see fit with the only requirement that resulting Docker containers respond to the defined REST endpoints correctly. These endpoints are then used by the MCA to index data or retrieve rankings and recommendations. An example of how to retrieve a ranking containing the first five results on the first page for the query *vaccine* is given below.

```
GET container/ranking?query=vaccine&page=1&rpp=5
```

REST endpoint for retrieving rankings from a single container

In comparison to pre-computed results, these micro-services are more comprehensive in the sense that their responses are not limited to pre-selected queries (or seed documents) only when rankings or recommendations are dynamically derived *on-the-fly*. The corresponding project template can be found in a public GitHub repository.¹

3 The multi container application

Each IR or RecSys application is a micro-service integrated into the MCA. By using Docker as a packaging tool, it is assured that all systems are reproducible and set up as intended, and likewise, the deployment effort is reduced to a minimum. Having a reproducible system evaluated across different domains may help to gain insight in terms of generalizability.

The MCA is the entrance for site providers to the infrastructure with a broker-like role. Sites communicate with the MCA (by a REST-API), mainly for retrieving ranking and recommendations and sending user feedback. When a user enters a query into the user interface, the site conducts this query to the MCA and receives a JSON-formatted response like it is exemplified below.

¹ <https://github.com/stella-project/stella-micro-template>

```

{
  "body":{
    "1":{ "docid":"M27622217", "type":"BASE" },
    "2":{ "docid":"M27251231", "type":"EXP" },
    "3":{ "docid":"M27692969", "type":"BASE" },
    "4":{ "docid":"M26350569", "type":"EXP" },
    "5":{ "docid":"M26715777", "type":"EXP" }
  },
  "header":{
    "container":{
      "base":"rank_base",
      "exp":"rank_elastic"
    },
    "page":0,
    "q":"vaccine",
    "rid":3,
    "rpp":5,
    "sid":1
  }
}

```

JSON-formatted ranking returned by the MCA

Especially for small- to mid-scale experiments, traditional A/B tests are not feasible due to limited amounts of feedback data (Hofman et al., 2016). As a solution, we rely on relative user preferences by comparing systems side-by-side. Radlinski et al. (2008) introduced the Team Draft Interleaving (TDI), which interlaces two result lists. In contrast to A/B tests, less feedback data is required to infer which system performs better.

In the given example, the baseline (BASE) and an experimental (EXP) system compose the ranking. If multiple experimental systems can deliver results for a query, the system having the lowest number of impressions contributes its results. At the current state, sites should log if and when there was a click on a document and send this information to the MCA. From there on, the MCA takes over and sends the logged feedback to the central server at regular intervals. While the MCA also offers the possibility to conduct traditional A/B tests, two things must be considered. First, it requires tremendously more feedback data to gain reliable (and significant) insights between two test groups (ibid.). Second, interleaved rankings can be used implicitly to control the overall retrieval and recommendation quality. When interlacing the result lists with outputs of established IR or RecSys algorithm, it can be guaranteed that at least half of the rankings or recommendations deliver reasonable results. In contrast, exposing site users with entirely experimental results (by subpar systems) may cause frustration during search and, at the worst, damage the site's reputation.

Technically, the MCA is built with docker-compose² – a build automation tool for multi container Docker applications. With the help of a YAML file, it is possible to define multiple micro-services and combine them into one application. Each micro-service is defined by a separate entry in the YAML file. These entries will be automatically added by the infrastructure service once the systems passed the validation. A central micro-service administrates all other micro-services and provides the RESTful interface. It distributes incoming queries to the experimental systems, receives the logged feedback data, and posts them to the central server of the infrastructure. The implementation of the MCA is available in a public GitHub repository.³

4 The server

The central server of the infrastructure serves four basic functionalities, including 1) the administration of the infrastructure users, 2) the dashboard service, 3) automated updates of the MCA, and 4) hosting the user feedback data. The underlying technology stack is given in Figure 4.

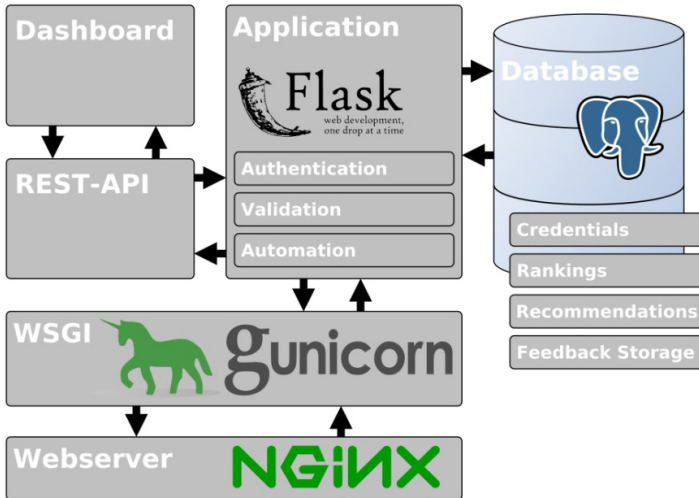


Fig. 4 Technology stack of the server

2 <https://docs.docker.com/compose>

3 <https://github.com/stella-project/stella-app>

Like the MCA and the template of the micro-services, the server application is implemented with Flask. It handles the authentication of users, validates submissions by experimenters and automates the build process of the MCA. Furthermore, it offers a RESTful API that is used by the MCA to post feedback data. Likewise, experimenters can use this API to retrieve feedback data which is stored in a PostgreSQL database.

After registration, new systems can be submitted by adding pointers to the corresponding GitHub-URL or by uploading pre-computed system results. Once systems' outputs have been exposed to site users, the dashboard service provides plots with the number of impressions and total clicks. Since we evaluate the systems' outputs with TDI, we can derive measures like wins, losses, and ties (Schuth et al., 2015). Figure 5 provides two examples of analytic tools that are provided by the dashboard service. These visualizations give first impressions of how the system performs in comparison to the baseline. Once enough data is available, experimenters can start the feedback loops that are illustrated in Figures 2 and 3. Clicks and other interactions can be used as points of reference for improving the ranking and recommender systems and the changes made can be compared to those of the previous evaluation phase. The implementation of the server is available in a public GitHub repository.⁴

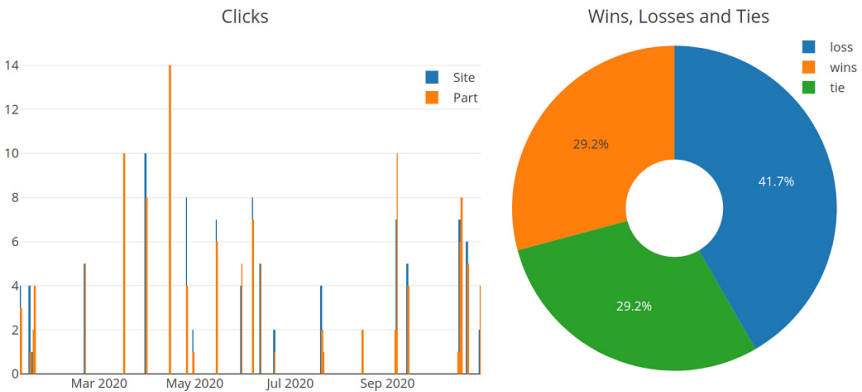


Fig. 5 Examples from the dashboard

4 <https://github.com/stella-project/stella-server>

5 Related work

Primarily, IR and RecSys methods should return results that are relevant to target users. While it is out of the scope of this work to elaborate on what constitutes relevance, researchers commonly agree that different types of relevance exist. Topical relevance judgments complement test collections (Sanderson, 2010) and simplify experimental evaluation but do not reflect the pertinent needs of specific searchers with different contexts.

Kelly's spectrum of IR experimentation contrasts experiments based on test collections with human-focused behavior studies at opposing ends (Kelly, 2007). IIR experiments can grasp very specific aspects of human behavior when interacting with search systems. However, these types of experiments are usually conducted on a small scale, and the recruitment of test subjects is costly. In the middle of her spectrum, Kelly locates log-based studies. They serve as a compromise between these two extremes of experiment designs, and one possible implementation of log-based studies are living labs (Kelly et al., 2009).

Living labs became part of IR conferences and workshops in 2014. Balog et al. (2014) introduced an API that can be used to retrieve experimental rankings for pre-selected queries, whereas Hopfgartner et al. (2014) introduced an infrastructure for real-time news recommendations. Both approaches share the idea of evaluating system performance with the help of user feedback that is provided in the form of clicks or other types of interaction with the presented search results. The living lab API for rankings was implemented in the LL4IR workshops (Schuth et al., 2015) as well as in the OpenSearch track (Jagerman et al., 2018), whereas the recommendation platform was used for the NEWSREEL campaign (Brodt & Hopfgartner, 2014).

While these attempts demonstrate the feasibility of living labs in IR and RecSys research, there remain many open questions and issues, e.g., regarding the reusability of such test collections that are augmented by user feedback data (Tan et al., 2017). More recent examples of living lab implementations are APONE (Marrero & Hauff, 2018), Macaw (Zamani & Crawell, 2020) and arXivDigest⁵. APONE is a living lab platform that is tailored to A/B tests as they are usually conducted in IIR studies. It is based on the PlanOut language developed by Facebook, which facilitates designing online

5 <https://arxivdigest.org>

field experiments by scripting them (Bakshy et al., 2014). Macaw has a focus on conversation information seeking, while its backend can either be fully algorithmic or set up as part of *wizard of oz* experiments. arXivDigest is a service that provides recommendations for research articles with the help of personalized email updates on recent publications based on interest profiles.

Recently, Schaible et al. (2020) provided a state-of-the-art overview on shared task evaluation platforms and found that there is no platform that combines both reproducible and online field experiments (as implemented by living labs). The presented infrastructure's underlying idea combines these two orthogonal research branches and can be framed with respect to larger reproducibility efforts in information retrieval (Breuer et al., 2019). Conceptually, the infrastructure takes account of the PRIMAD model (Ferro et al., 2016) that defines essential components of a reproducible experiment (including **Platform**, **Research goal**, **Implementation**, **Method**, **Actor**, **Data**).

6 Conclusions and future work

In this work, we introduced a novel architecture for living lab experiments in IR and RecSys research. Living labs offer a new perspective when evaluating ranking and recommender systems. Conventional test collections merely support topical relevance, but not pertinent information needs of specific users. As a solution, feedback data, inferred from living lab experiments, complements system-oriented evaluations with new insights.

How does the proposed architecture differ from previous living labs? First, the way how systems are integrated into the infrastructure focuses on making the experiments reproducible. Furthermore, there is a frictionless transition from offline batch-style experiments to online environments. With the help of the introduced submission mechanism for pre-computed runs, system-oriented measures can be easily complemented by user feedback. Experimenters feed the datasets into their available ranking or recommender systems and simply upload the outputs for selected queries or target items to the central server of the infrastructure.

Previous living labs made use of pre-computed results for top- k queries that were available from a web-based service (Jagerman et al., 2018). When contributing full-fledged systems in Docker containers, there is no limitation to the top- k queries and, likewise, latencies should be reduced to a minimum,

since results do not need to be transferred over the web but will be retrieved from local servers at the sites. Furthermore, systems should be able to deliver rankings or recommendations *on-the-fly* when implementing them as micro-services. In comparison to pre-computed results, the outputs of these dockerized systems are more comprehensive since they should be able to deliver results for more than just pre-selected queries or target items. To lower the entrance barrier even further, we plan to provide project templates with dockerized versions of popular open-source search engines like Elasticsearch or Solr. Likewise, it is worth investigating the feasibility of a shared index for multiple experimental systems, as exemplified by the CIFF format (Lin et al., 2020).

At the current state, the provided datasets of the upcoming evaluation campaigns do not have topical relevance judgments. While it is possible to annotate specific rankings or recommendations after feedback data has been collected, it is a major challenge to infer the query logs' underlying information need and this may be a topic for future research. Queries are the users' verbalization and understanding of their information needs. In the outlined setup, we do not have any information about the users, thus no context knowledge. Different information needs may translate into the same query. Even if we are provided with the query, the resulting ranking, and the corresponding feedback data, it is not enough to reach any definitive conclusions about the information need. Similar problems occur for RecSys evaluations. In contrast to rankings, recommendations are not part of an active search process, but likewise we do not have any context knowledge and thus do not know what may have caused the users to click on a recommended item.

As part of our future evaluations, it is worth evaluating the infrastructure by comparing results from encapsulated systems to pre-computed results. Since experiments with pre-computations resemble older living lab attempts, we can treat them as a baseline when evaluating the new implementation of the living lab paradigm with the help of Docker.

In 2021, there is the Living Labs for Academic Search (LiLAS⁶) evaluation campaign at CLEF 2021 (Schaer et al., 2020). Interested readers are invited to contribute their IR and RecSys experiments and have them evaluated in the living lab environment that is introduced in this work. As part of the STELLA project, TH Köln, GESIS, and LIVIVO team up to offer lab partici-

6 <https://clef-lilas.github.io>

pants the possibility to expose their systems' results to users from the social and life sciences. A demo setup of the entire infrastructure can be found in a public GitHub repository.⁷

References

- Bakshy, E., Eckles, D., & Bernstein, M. S. (2014). Designing and deploying online field experiments. In *Proceedings of the 23rd International Conference on World Wide Web – WWW '14*, pp. 283–292. <https://doi.org/10.1145/2566486.2567967>
- Balog, K., Kelly, L., & Schuth, A. (2014). Head First: Living Labs for Ad-hoc Search Evaluation. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management – CIKM '14*, pp. 1815–1818. <https://doi.org/10.1145/2661829.2661962>
- Breuer, T., Schaer, P., Tavakolpoursaleh, N., Schaible, J., Wolff, B., & Müller, B. (2019). STELLA: Towards a Framework for the Reproducibility of Online Search Experiments. In *OSIRRC 2019, the Open-Source IR Replicability Challenge* (pp. 8–11). Aachen: RWTH Aachen. <http://ceur-ws.org/Vol-2409/position01.pdf>
- Brod, T., & Hopfgartner, F. (2014). Shedding light on a living lab: The CLEF NEWSREEL open recommendation platform. In *Proceedings of the 5th Information Interaction in Context Symposium – IliX '14* (pp. 223–226). New York, NY: ACM. <https://doi.org/10.1145/2637002.2637028>
- Ferro, N., Fuhr, N., Järvelin, K., Kando, N., Lippold, M., & Zobel, J. (2016). Increasing Reproducibility in IR: Findings from the Dagstuhl Seminar on “Reproducibility of Data-Oriented Experiments in e-Science”. *SIGIR Forum*, 50(1), 68–82. <https://doi.org/10.1145/2964797.2964808>
- Hofmann, K., Li, L., & Radlinski, F. (2016). Online Evaluation for Information Retrieval. *Foundations and Trends® in Information Retrieval*, 10(1), 1–117. <https://doi.org/10.1561/15000000051>
- Hopfgartner, F., Kille, B., Lommatzsch, A., Plumbaum, T., Brodt, T., & Heintz, T. (2014). Benchmarking News Recommendations in a Living Lab. In E. Kanoulas, M. Lupu, P. Clough, M. Sanderson, M. Hall, A. Hanbury, & E. Toms (Eds.), *Information Access Evaluation. Multilinguality, Multimodality, and Interaction* (pp. 250–267). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-11382-1_21

⁷ <https://github.com/stella-project/stella-search>

- Jagerman, R., Balog, K., & Rijke, M. D. (2018). OpenSearch: Lessons Learned from an Online Evaluation Campaign. *Journal of Data and Information Quality*, 10(3), 1–15. <https://doi.org/10.1145/3239575>
- Kelly, Diane (2007). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, 3(1–2), 1–224. <https://doi.org/10.1561/1500000012>
- Kelly, D., Dumais, S., & Pedersen, J. O. (2009). Evaluation Challenges and Directions for Information-Seeking Support Systems. *Computer*, 42(3), 60–66. <https://doi.org/10.1109/MC.2009.82>
- Lin, J., Mackenzie, J., Kamphuis, C., Macdonald, C., Mallia, A., Siedlaczek, M., Trotman, A., & de Vries, A. (2020). Supporting Interoperability Between Open-Source Search Engines with the Common Index File Format. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2149–2152). New York, NY: ACM. <https://doi.org/10.1145/3397271.3401404>
- Marrero, M., & Hauff, C. (2018). A/B Testing with APONE. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 1269–1272). New York, NY: ACM. <https://doi.org/10.1145/3209978.3210164>
- Radlinski, F., Kurup, M., & Joachims, T. (2008). How does clickthrough data reflect retrieval quality? In *Proceeding of the 17th ACM Conference on Information and Knowledge Mining – CIKM '08* (pp. 43–52). New York, NY: ACM. <https://doi.org/10.1145/1458082.1458092>
- Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends® in Information Retrieval*, 4(4), 247–375. <https://doi.org/10.1561/1500000009>
- Schaer, P., Schaible, J., & Garcia Castro, L. J. (2020). Overview of LiLAS 2020 – Living Labs for Academic Search. In A. Arampatzis et al. (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (pp. 364–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58219-7_24
- Schaible, J., Breuer, T., Tavakolpoursaleh, N., Müller, B., Wolff, B., & Schaer, P. (2020). Evaluation Infrastructures for Academic Shared Tasks: Requirements and Concept Design for Search and Recommendation Scenarios. *Datenbank-Spektrum*, 20(1), 29–36. <https://doi.org/10.1007/s13222-020-00335-x>
- Scuth, A., Balog, K., & Kelly, L. (2015). Overview of the Living Labs for Information Retrieval Evaluation (LL4IR) CLEF Lab 2015. In J. Mothe et al. (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (pp. 484–496). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24027-5_47

- Tan, L., Baruah, G., & Lin, J. (2017). On the Reusability of “Living Labs” Test Collections: A Case Study of Real-Time Summarization. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 793–796). New York, NY: ACM. <https://doi.org/10.1145/3077136.3080644>
- Zamani, H., & Craswell, N. (2020). Macaw: An Extensible Conversational Information Seeking Platform. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2193–2196). New York, NY: ACM. <https://doi.org/10.1145/3397271.3401415>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 348–362. DOI: doi.org/10.5283/epub.44953.

procd: A Privacy-preserving Robust Implementation to Discover Contacts in Social Networks

Fabian Deifuß

University of Wuppertal, Germany
fabian.deifuss@uni-wuppertal.de

Cornelius Ihle

University of Wuppertal, Germany
ihle@gipplab.org

Moritz Schubotz

FIZ Karlsruhe, Germany
moritz.schubotz@fiz-karlsruhe.de

Bela Gipp

University of Wuppertal, Germany
gipp@uni-wuppertal.de

Abstract

Current instant messengers store the users' phone book contacts typically unencrypted or hashed on a central server. In case of a server's corruption, all contacts are either directly available in plaintext or can be unmasked using a simple dictionary attack. To solve this problem, we present *procd* [pʁo:st], a python implementation for privacy preserving contact discovery. *procd* is a trustless solution that requires neither plaintext numbers nor hashes of single phone numbers to retrieve contacts. Instead, we transfer hashed combinations of multiple phone numbers, which increases the effort for dictionary attacks to an unfeasible level using today's hardware.

Keywords: private contact discovery; private set intersection; secure multi-party computation; private information retrieval

1 Introduction

State-of-the-art social networks and messaging services store a social graph of its users to suggest communication options. Having a service provider storing and sharing a social graph is not privacy-preserving and should be avoided whenever possible.

procd [pɹɔ:st] – **Private RObust Contact Discovery** is our approach to private contact discovery with increased robustness against brute force attacks and without the need to store a social graph. We instead use a minimal social graph each user already has on its phone, the address book. Thus, the contact data remains distributed and owned by the user

Our goal is to find a way to increase the complexity of a brute force attack to a point where it is computationally infeasible to find an input that hashes to the processed values. For a dictionary attack on contacts, an attacker systematically tries each possible phone number as an input to match and unmask a hashed value to reveal numbers and connections.

2 Related work

The Signal¹ messenger includes one of the most promising implementations. Signal clients hash their phone number locally before uploading (Marlinspike, 2017) it to the Signal servers. However, even though this is better than uploading and storing everything in plaintext, a typical phone number only consists of about ten digits. Hence, these hashes are vulnerable to dictionary attacks (Bošnjak et al., 2018). Signal is aware of this issue and therefore has to rely on a hardware solution called Software Guard Extension (SGX) from Intel. This, however, moves the trust issue to another party – the hardware manufacturer.

3 Method

For our approach, we aim to meet three criteria:

1. no exchange of plaintext contact information
2. robustness against brute-force attacks
3. no dependency on single proprietary hardware solutions.

We hence, propose an unbalanced private set intersection with increased input complexity.

1 <https://signal.org/blog/private-contact-discovery/>

3.1 Pairwise hashing

Instead of hashing a single phone number, we form a hash of a pair of numbers. Each hashed pair consists of a user's phone number and one of her contacts.

In pairwise hashing, the server only ever sees the published hashes and does not gather any registered client's information. Even though the input complexity increases drastically, most benefits vanish if an attacker already has information like the (1) relation between the targets or (2) the individual's phone number. Therefore, it is necessary to salt the hash with a privately disclosed secret known only to the two parties trying to communicate.

Once a client knows about already registered contacts, a Diffie-Hellman key exchange (Li, 2010) is used to enable authentication and initiate private communication. The public keys can be extracted from our public database. However, we cannot just post each party's public key linked to their phone number. Hence, we apply a way to get a hold of each other's public key without exposing the corresponding phone number, neither in plaintext nor hashed. As mentioned earlier, we suppress false positives by hashing our combinations in two different orders, starting with Bob or starting with Alice. This way it is possible to store one's public key alongside the hash known by both parties.

3.2 Experiment

A REST API serves as an interface to GET an intersection of already registered contacts and POST a user registration to our exemplary service. Figure 1 shows all used system entities and their interfaces.

The REST API processes the user requests and inserts or retrieves information from the public PostgreSQL database, which holds all hashed phone number combinations together with their public keys. The client constructs two dictionaries of number combinations before interacting with the API. These two dictionaries differ in the order of phone numbers, but both can contain a common secret (salt), unique to each contact. The first dictionary (Dict1) is used to publish all contacts to the contact discovery service. The second dictionary (Dict2) is not published but is used to verify any retrieved hash from the service and filter false positives. Additionally to the dictionary, a user's public key is appended to the hash combinations before publishing so that the desired public key can be retrieved and used for an encrypted communication initiation through the messaging service.

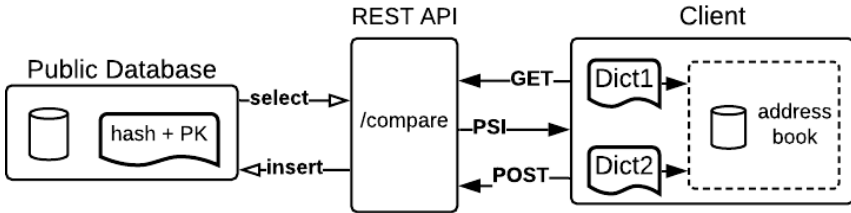


Fig. 1 System overview

4 Evaluation

In the following paragraphs, we analyze and discuss our architecture with regard to each of the three criteria we aimed for.

4.1 Robustness against dictionary attacks

The workload to compute hashes is horizontally scalable. Thus, the critical metric to evaluate the feasibility of computing a specific dictionary is the price to pay for the necessary computing resources.

On a modern computer (six cores, 2.8 GHz), it takes 0.00063 milliseconds ($6.3e-7$ seconds) to compute a SHA1 hash. This translates to $1.5e+6$ hashes per second. Assuming the desired output hash is computed after half of the possible combinations (input complexity of $8e+20$ without salt), it would take approximately 8 million years of computing to get the desired hash. A comparable VM rental on Azure is about 0.30 \$ per hour. Hence an attacker would need 21 billion USD for a successful dictionary attack.

Table 1: Dictionary attack cost estimation

Input complexity	Estimated time	Estimated cost of computation
German number hash ($1e+6$)	1 second	< 0.01 USD
Number hash ($4e+11$)	1.5 days	10 USD
Pairwise German number hash ($1e+12$)	7.3 days	52 USD
<i>procd</i> German number hash ($5e+18$)	10,000 years	262 million USD
Pairwise WhatsApp User hash ($8e+20$)	8 million years	21 billion USD
<i>procd</i> WhatsApp User hash ($4e+27$)	40 trillion years	105 quadrillion USD

Table 1 shows the estimated time required to compute the desired hashes for a contact list of 200 entries on modern hardware (single machine) with their estimated costs alongside the different hashing complexities.

Using our methodology, the upfront cost of resources necessary to compute a specific dictionary in question is incredibly high. Further, increasing the hash’s complexity through a salt is reasonable, as unmasking a single hash would otherwise lead to the exploitation of the corresponding dictionary. Hence, not only a registered phone number but its entire address book would be exposed.

4.2 Comparison

Compared with the contact discovery methods of other state-of-the-art mobile messaging applications, none of the popular applications meets all our criteria.

Table 2: Privacy protection overview (Kales et al., 2019)

	Whats-App	Tele-gram	Signal	Threema	paired	<i>procd</i>
Processes phone numbers in plaintext	✘	✘	–	–	–	–
Processes hashes of contacts	–	–	✘	✘	✘	✘
Uses salted hashes to prevent dictionary attacks	–	–	–	–	–	✘
Relies on trusted hardware	–	–	✘	–	–	–
Cost to unmask a single phone number (self-discovery)	\$ 0	\$ 0	\$ 10	\$ 10	\$10	\$ 276
Cost to unmask German numbers (10^6)	\$ 0	\$ 0	\$ 0.01	\$ 0.01	\$52	\$ 262M

5 Conclusion

We introduced a new unique method (pairwise phone number hashing) for private contact discovery and increased our robustness against dictionary attacks effectively using a common shared secret. Additionally, we implemented a mechanism to retrieve a public key to initiate communication. With

our *procd* approach, we successfully improved unbalanced private set intersections for the specific use case of contact discovery.

Our experiment shows that a trustless private contact discovery design is possible, and no exchange of plaintext data is needed.

References

- Bošnjak, L., Sres, J., & Brumen, B. (2018). Brute-force and dictionary attack on hashed real-world passwords. <https://doi.org/10.23919/MIPRO.2018.8400211>
- Cimpanu, C. (2020). New Platypus attack can steal data from Intel CPUs. *ZDNet*. <https://www.zdnet.com/article/new-platypus-attack-can-steal-data-from-intel-cpus/>
- Evans, D., Kolesnikov, V., Rosulek, M. (2018). *Pragmatic Introduction to Secure Multi-Party Computation*. Norwell, MA: Now Publishers.
- Ihle, C., Schubotz, M., Meuschke, N., & Gipp, B. (2020). A First Step towards Content Protecting Plagiarism Detection. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (pp. 341–344). New York, NY: ACM. <https://doi.org/10/ghg7rw>
- Kales, D., Rechberger, C., Schneider, T., Senker, M., & Weinert, C. (2019). Mobile Private Contact Discovery at Scale. In *28. USENIX Security Symposium (USENIX Security'19)* (pp. 1447–1464). Berkeley, CA: USENIX Association.
- Lee, K., Kaiser, B., Mayer, J., & Narayanan, A. (2020). An empirical study of wireless carrier authentication for SIM swaps. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)* (pp. 61–79). [Berkeley, CA]: USENIX Association. <https://www.usenix.org/conference/soups2020/presentation/lee>
- Li, N. (2010). Research on Diffie-Hellman key exchange protocol. *2010 2nd International Conference on Computer Engineering and Technology* (Vol. 4, pp. 634–637). Piscataway, NJ: IEEE. <https://doi.org/10/bdtfv3>
- Marlinspike, M. (2017). Technology preview: Private contact discovery for Signal. *Signal Messenger*. <https://signal.org/blog/private-contact-discovery/>
- Yanai, A. (2020). Private Set Intersection. <https://decentralizedthoughts.github.io/2020-03-29-private-set-intersection-a-soft-introduction/>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 363–368. DOI: doi.org/10.5283/epub.44954.

Session 7:
Information Infrastructure

Educational Open Government Data in Germany

The Landscape, Status, and Quality

*Cornelia Veja¹, Julian Hocker¹,
Christoph Schindler¹, Marc Rittberger^{1,2}*

¹ DIPF | Leibniz Institute for Research and Information in Education

² Darmstadt University of Applied Science, Germany

{[veja](mailto:veja@dipf.de), [julian.hocker](mailto:julian.hocker@dipf.de), [schindler](mailto:schindler@dipf.de), [rittberger](mailto:rittberger@dipf.de)}@dipf.de

Abstract

This paper examines the educational datasets of Open Government Portals (OGPs) in Germany. While OGPs has become an important player in making public data available, the quality and coverage are increasingly problematized. This study analyzes the quality of 28 OGPs, 18 on the city-level, seven on the state-level (the city-states of Berlin, Hamburg, and Bremen were counted as states), and three at the national level. The focus of this study is on identifying the gap in open educational data. The results show that the total number of datasets and the educational data varies very much between the portals. While the portals contain 94% ‘open’ datasets, most of the portals do not provide comprehensive usage metrics like the data download, evidence of usage in research or applications. Finally, the paper highlights the main shortcomings of the existing open data portals regarding the quality of datasets and the lack of findability and granularity.

Keywords: open data; OGD – Open Government Data; OGP – Open Government Portals; open education; German open data portals; local government; municipality; usability

1 Introduction

For several years open data, as part of the open movement, calls for transparency and collaboration for the benefit of the society at large.

A range of public administration followed this so-called open government movement and established Open Government Portals (OGP). While these developments resulted in various policies, concepts of use, and different underlying technologies with a diversity of content, functionalities, and appropriateness, the metadata quality, content relevance, and format proved crucial in the endeavor to re-use the data.

This paper's motivation lies in these challenges by following the question: What is the current status and quality of open government data portals in Germany? Therefore, this paper presents an analysis of currently available portals on the national and local level in Germany and aims, in particular, to understand whether these portals provide open data in a way that, in fact, facilitates their re-use and public accountability. Thereby, we especially focus on the educational content of these Open Government Data Portals (OGDPs). While domain-specific perspectives to the quality of open data are still the main desideratum, the most unrepresented domain in this respect is education (Atenas et al., 2019). Whether the role of education in producing and consuming open data is emphasized in the literature, the focus is mostly on Open Educational Resources (OER) rather than on the educational sectors' administrative, economic and social context. Especially in Germany, with its heterogeneous federated educational system, the openness of related educational government data is the main basis and need for accountability and improvement.

This study analyzed 28 OGPs, from which 21 of the largest cities (three of them are city/states), four portals at the state level and three on the national level. Following Charalabidis et al. (2018), the research uses an approach that refines the information system success model and systematizes and adjusts the indicators to the educational domain. We collected the data via the API-interface and enriched it by intellectual research and by analyzing the portals.

The novelty of this study is twofold:

1. The deep, thematic study of OGPs in Germany, with a focus on education. There are not many papers in literature, except that of Wang, Chen, and Richards (2018), which assesses the Open Government Data in education. There is no paper assessing the German local Open Government Data in deep, most of the studies address only the national portal.
2. The analysis of the OGPs put emphasis on the re-usability of datasets, and provides a model and indicators to measure it.

The paper's organization is as follows: Section 2 introduces the literature review and background of the research context on open data, Open Govern-

ment Portals, and their quality requirements. In Section 3, the research methodology is formulated, and the research method is described and justified. Section 4 presents the results of open data portals quality concerning. Section 5 discusses the results concerning other studies. Finally, concluding remarks are provided in Section 6. Supplementary materials are provided at the end of this paper.

2 Research context

2.1 Open data movement

The last two decades represent important steps in establishing the definition, principles, and policies of open data, open knowledge, and open content. In 2005, the Open Knowledge Foundation published the open definition¹ to establish under which conditions data and content can be seen as open. Since then, organizations, governments, and standardization bodies have been in place to establish legal and technical frameworks concerning the open definition. According to the *Open Data Handbook*² and Gerunov (2016), open data are “freely accessed, used, modified and shared by anyone for any purpose – subject only, at most, to requirements to provide attribution and/or share-alike”. This definition leads to two aspects of open data: the technical aspect and the legal aspect, going hand in hand. These aspects influence how open data are published and used.

Open data initiatives aim to open all non-personal and noncommercial data, especially (but not exclusively) all data collected and processed by government organizations. In 2007, the Open Government Working Group³ established the eight fundamental pillars that support the concept of open data. The data should be: complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and under a free license. These are superseded in 2013 when the International Open Data Charter launched the foundation for access to data and the release and use of admin-

1 <https://opendatacharter.net/the-open-definition/>

2 <https://opendatahandbook.org>

3 https://public.resource.org/open_government_meeting.html

istrative data in six principles: open by default, timely and comprehensive, accessible and usable, comparable and interoperable, for improved governance and citizen engagement, for inclusive development and innovation (Open Data Charter).

The six principles aim at a globally-agreed set of aspirational norms for publishing open data as a pre-requisite of all activities related to open data. To implement these, it is recommended to take into account the respective national political and legal framework. While some aspects of open data quality align with the ones of web portals, domain-specific quality perspectives in the context of open data (e.g., data management system, the openness of provided data based on the license or format, metadata) need to be identified and evaluated.

2.2 Open data and education

Whether the role of education in producing and consuming open data is emphasized in the literature, the focus is mostly on Open Educational Resources (OER), rather than on the administrative, economic and social context of educational sectors. Education is seen as one of the most under-represented domains in open data (Atenas et al., 2019).

The results of several international open data surveys pose questions about open data and its challenges, especially in education: The United Nations E-Government Survey, an assessment containing specific questions about open data since 2014, covering 194 countries in the eight editions of the survey, remarks, as a common approach for data gathering, data are often in non-machine-readable format, for example, in PDF. While non-machine-readable data has doubled in the past two years across various sectors, machine-readable datasets increase incrementally (United Nations, 2018). In Education, the number of countries providing machine-readable formats increases from 39 (2016) to 69 (2018), as the non-machine-readable formats decreases from 91 in 2016 (United Nations, 2016) to 88 in 2018 (United Nations, 2018).

Further on, the *Open Data Barometer*⁴ peer-reviewed expert survey covered 155 countries in 2017, and the Global Open Data Index covered 94 countries in 2016/2017. In the last edition, *Open Data Barometer* assesses primary and secondary education performance data.

4 <https://opendatabarometer.org/>

2.3 Open data in Germany

Germany's engagement in open data started in 2013 when the G8 countries adopted an Open Data Charter. The German action plan was adopted in 2014 and contained four commitments (Wiebe, 2020):

1. a clear direction signal for open data in Germany
2. the publication of records
3. the publication of the data on a national portal
4. measures for consultation, engagement, and exchange of experience.

The third commitment became a fact in 2015, when the national portal GovData.de was created to implement the National Action Plan. GovData.de aggregates datasets from local OGPs and various public institutions.

On May 18, 2017, the German Bundestag passed the draft first law to amend the E-Government Act presented by the Federal Minister of the Interior. The draft law specifies central criteria for open data. This includes, in particular, free provision, free access to the data, and machine readability (Klein, 2017). On March 27, 2019, the federal government started the consultation process on the second national action plan for the open government partnership in response to the increasing public interest in open data. Correlated with the EU Directive 2019/1024 of June 20 (European Commission, 2019), Germany established in September 2019 the second national action plan for 2019–2021.

As part of the local contribution to the second national action plan of the open government partnership, a primer initiative offers an overview of which data sets are being made available by Germany's cities and towns for the first time. As a first step, the catalog contains the generally accessible data sets published by communities in the state of North-Rhine-Westphalia.⁵

The *Open Data Barometer*⁶ scores Germany at 60% openness because of a lack of machine-readable format, open license or dataset identifiers. The European Data Portal yearly report (EU28+ Open Data Maturity of 2019) describes Germany's open data quality level as below average, at 63% (65% EU level). The usage of open data portals and open data awareness are reported on an ascendant trend. Overall, Germany scores slightly above the EU average (68%, and 66% respectively) (Blank et al., 2019).

5 <https://okfn.org/>

6 https://opendatabarometer.org/?_year=2017&indicator=ODB

2.4 Evaluating open government data

At a broader level, an early study from 2012, driven by the Technical University of Dresden, surveyed the top 50 open data platforms retrieved from the Open Knowledge Foundation catalog. The study benchmarked those repositories regarding the re-usability of open data (Braunschweig et al., 2012). The reports of the Open Data Barometer are more up-to-date. In the same way, Open Data Monitor⁷ provides information about dataset consistency automatically obtained from open data portals across Europe. The general approach of this kind of report is rounded-off with an assessment of sectors, like education.

Other studies focus mostly on the non-technical aspect of open data, discussing policies and legal coverage, social impact, and future possible development (Wiebe, 2020). In the same category, the authors investigate the relationship between the objectives of open government data initiatives and the benefits delivered (Zuiderwijk et al., 2019). They pointed out the contrast between the objectives and their attainment in open government data initiatives (OGDI). They proved that the benefits are often in areas other than those of the open government data initiative's objectives, and it is not clear whether the intended benefits have been delivered.

In Corrêa and Corrêa da Silva (2019), Neumaier et al. (2016), and Kubler et al. (2018), the authors have also automatically explored the quality of metadata from OGPs around the world. They assessed data portals and discussed general quality issues like retrievability of the data. The papers also define a set of quality metrics for the Data Catalog Vocabulary (DCAT) metadata standard. In the same category, the recent study from Chapman et al. (2020) highlights the different mechanisms used to successfully search datasets, and concludes that “dataset search itself is in its infancy” (ibid.).

The study by Wiczorkowski (2019) focuses on the quality of access methods and proper publishing of OGD and the economic benefits. The author has identified data publication's main problems, based on Central Repositories for Public Information (CRPI) in Poland, the USA, the UK, and Germany. At the European Union level, the study by de Juana-Espinosa Luján-Mora (2019) monitors and clusters the data collected from OGD portals in the 28 countries to showcase their similar involvement in open data.

7 <https://www.opendatamonitor.eu/>

Some studies focused on the province level at open government data portals and datasets. Noteworthy is the Chinese OGP study driven by Wang, Chen, and Richards (2018). At the time of their study, they note the lack of a national level open government data portal in China. Similar studies have been carried out in Italy (Molinari et al., 2017), Brazil (dos Santos Brito, 2015), Columbia and Spain (Benitez-Paez et al., 2018), India (Buteau et al., 2018), and Bulgaria (Gerunov, 2016). All these papers are not domain specific, except the Chinese paper (Wang, Chen, & Richards, 2018), which assesses the impact in various domains, among them education.

A comprehensive study of open data usability was carried out by Alexopoulos (2017). The study highlighted education, health, and finances as public sector areas with major open data publication incidence. Other studies like Berends et al. (2017), Benitez-Paez et al. (2017), Corrêa et al. (2017), Máchová and Lnénicka (2017), Kubler et al. (2018), Wiczorkowski (2019), Wang, Button, and Shepherd (2018), Braunschweig et al. (2012), Schmidt et al. (2016) and Zuiderwijk et al. (2019) identified the following main barriers preventing the re-usability of open datasets: public engagement, culture, economic, political and technical factors.

In recent years, several projects focused on open data in Germany and took steps forward to establish a clear roadmap of what open data of Germany stands for, and intending to clarify the classification of open datasets and their providers. One of these projects is *Open Data Map*.⁸ The project aimed to provide a complete and up-to-date overview of all offers of open data from as many public bodies as possible in the public sector and make this information accessible, among other visualizations, via Germany's map. The outcome of the project mirrors the status of open data of 2014, in an intuitive visual map, regardless of the content and theme of the datasets.

3 Methodology for OGPs study

This section presents the methodology for assessing German OGPs quality, grounded in the existing literature and adapted for the current research purpose. The methodological approach focuses on the objective study of OGPs, considering the potential usage in educational research. The approach refines

8 <https://www.open-data-map.de>

the information system success model (Charalabidis et al., 2018) with the metrics from the usability model (Osagie et al., 2017) appropriate to the educational field.

The Information System (IS) success theoretical model was first developed by William H. DeLone and Ephraim R. McLean in 1992. The most widely used system success model is the one by DeLone and McLean: Model of IS Success, developed in 2003. It proposes seven IS success measures, which are structured in three layers:

1. first layer: ‘information quality’, ‘system quality’ and ‘service quality’,
2. second layer: affecting ‘user satisfaction’, and
3. third layer: ‘actual use’ of the IS.

The model proposed by Charalabidis et al. in 2014, for the evaluation of the advanced second generation of OGDs, was primarily based on the IS success model. The new model adopts a layered evaluation approach and includes measures of both information and system quality.

The usability model proposed by Osagie et al. in 2017 refines the second and third layers of the IS success model, envisaging the third generation of OGD. These models emphasize capabilities like linked data, open collaboration (interagency, and with the public), the possibility of co-creating value-added services, learnability, accessibility, and feedback.

Starting from the aforementioned models of Open Government Data indexes resented in Charalabidis et al. (2018), Osagie et al. (2017), and the specificity of educational subject, the analysis considers the following perspectives and dimensions to signal the portal quality, presented in Table 1.

Table 1: The indicators for the study

Dimensions	Perspectives	Description	Literature reference correspondence
1. System quality	1.1 Type of portal	direct provider or aggregator	collaboration spaces in Charalabidis et al. (2018)
	1.2 Data provision	providing an API, web interface for search	Charalabidis et al. (2018), accuracy in Osagie et al. (2017)
2. Information quality	2.1 Thematic perspective	Portal provides a categorization of the datasets upon subjects.	availability and access dimension in Máchová et al. (2018)
	2.2 Number of education datasets	group of education datasets	Charalabidis et al. (2018)
	2.3 Number of organizations	number of organizations that collaborate in the portal for education datasets	de Juana-Espinosa et al. (2019)

Dimensions	Perspectives	Description	Literature reference correspondence
	2.4 Content format	number of machine-readable formats against non-machine-readable formats (per resources)	Charalabidis et al. (2018), availability and access dimension in Máchová et al. (2018)
	2.5 Metadata format	the metadata openness and metadata capabilities	Charalabidis et al. (2018), Mons et al. (2017)
	2.6 Metadata semantic	Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.	metadata interoperability in Mons et al. (2017)
	2.7 PID, ID	persistent identifier or at least permalink	Charalabidis et al. (2018), Mons et al. (2017)
	2.8 Information granularity ⁹	raw or aggregated content	Atenas and Haverman (2019)
	2.9 Keywords (tags) presence and representation	using keywords or a tag system, other than thematic categorization, helping users in searching datasets	learnability, accessibility in Osagie et al. (2017), availability and access dimension in Máchová et al. (2018), Chapman (2020)
3. Service quality	3.1 Licenses	number of open licenses against close, or no licenses	Máchová et al. (2017), Charalabidis et al. (2018), de Juana-Espinosa et al. (2019)
	3.2 Quality rating and feedback mechanisms	users interaction with the portal, in the form of blogs, user's activity, users rating	co-creating and value-added services in Charalabidis et al. (2018), communication and participation dimension in Máchová et al. (2018)
	3.3 Number of applications	number of reported open data applications (re-use)	de Juana-Espinosa et al. (2019), co-creating and value-added services in Charalabidis et al. (2018)

For the operationalization, several metrics are defined in the form of quantifiers or descriptive variables. The taxonomy of OGP's metrics is presented in Table 2.

The analysis methodology adopted in this section comprises four main stages: raw data collection, data aggregation, data processing and data analysis. These stages are described in detail below.

⁹ The current research does not evaluate the information granularity.

3.1 Data collection

For the data we selected the portals on three different levels: on national level, state level, and the level of the largest cities (March, 2020). Further criteria for selection have been:

1. Educational datasets: The portals were considered when either the group name containing the word ‘Education’ [German: Bildung] is explicitly set up at the inception of the OGP, or if this group not exists, the search interface and specific keywords were used to identify the ‘Education’ datasets.
2. The underlying software offers an API needed for automatic processing.

The final sample consists of 28 OGPs, 18 on the city-level, seven on the state-level (the city-states of Berlin, Hamburg, and Bremen were counted as states), and three at the national level.

This analysis considers two categories of OGPs: One category contains OGPs direct providers, whereas the other category considers the OGPs that are aggregators.

Collecting and normalizing the features of an OGP is quite challenging because a standard that all mentioned administrations accept and apply does not exist. All data were collected semi-automatically, employing an online search of the OGP by country, state, district, and city.

System quality

The survey used the search web interface of each OGP for collecting the type of portal and the provision capabilities (Table 1: 1. System quality).

Information quality

Where the open data portal is powered by CKAN¹⁰ or DKAN¹¹, as specialized tools for a data management system that makes open data accessible, they always provide an API. This facility was used to collect data concerning the number of organizations, keywords, metadata format, PID, content format, and licenses. For the rest of the perspectives, the manual method was used (Table 1: 2. Information quality).

Service quality

Data were collected manually by searching each of the OGPs web interfaces (Table 1: 3. Service quality). After completing the data acquisition and collec-

10 <https://ckan.org/about/>

11 <https://getdkan.org/>

tion, we performed data cleaning on the resulting dataset to identify and correct possible errors such as missing values, outlier values, or different data formats (Neumaier, 2016). This guarantees the highest degree of data reliability.

3.2 Data aggregation and processing

Considering the literature and the model presented in Table 1, the following numeric variables were calculated in the first step:

Table 2: Metrics and Measures

Dimensions	Perspectives	Metrics	Measure
1. System quality	1.1 Type of portal	direct provider or aggregator	[P/A]
	1.2 Data provision	API	[CKAN, DKAN, SMW, Web service]
		web interface for search	[YES/NO]
2. Information quality	2.1 Thematic perspective	Portal provides a categorization of the datasets upon subjects.	[YES/NO]
	2.2 Number of education datasets	group of education datasets	numeric, aggregated
	2.3 Number of organizations	number of organizations that collaborate in the portal for education datasets	numeric, aggregated
	2.4 Content format	number of machine-readable formats against non-machine-readable formats (per resources)	numeric, aggregated
	2.5 Metadata format	the metadata openness and metadata capabilities	[RDF, JSON, XML]
	2.6 Metadata semantic	metadata standard	[YES/NO]
	2.7 PID, ID	persistent identifier or at least permalink	[YES/NO]
	2.8 Information granularity*	raw or aggregated content	<i>not evaluated</i>
	2.9 Keywords (tags) presence and representation	extended tag system	[YES/NO]
3. Service quality	3.1 Licenses	number of open licenses against close, or no licenses	numeric, aggregated
	3.2 Quality rating and feedback mechanisms	user interaction/rating	[YES/NO]
		Blog	[YES/NO]
		statistics	[YES/NO]
		dataset content rating	[YES/NO]
3.3 Number of applications	number of reported open data applications (re-use)	numeric, aggregated	

Following Charalabidis et al. (2018), Open Data Platforms capabilities also include descriptive variables about datasets and sources, functionalities provided by the Open Data Portals in terms of dataset discovery, data provision capabilities, and data visualization. These qualitative metrics were aggregated in a second step (see Table 2).

4 Survey results

Table 3 shows all 28 analyzed Open Government Data Portals (OGDPs) with some main metrics. 18 portals are from the largest cities. Seven further relevant portals have been identified at the level of states, whereby Berlin, Hamburg, and Bremen are city-states. Three have been found on the national level. While the 25 portals of the city and state levels create data by themselves (data providers), the three national portals aggregate data.

Table 3: Several metrics of 28 Open Government Data Portals

Portal	Level	Datasets	Machine-readable	Datasets Education	Open license
Open Data Aachen	city	83	0	2	2
Open Data Bielefeld	city	97	5	7	7
Open Data Bonn	city	545	19	24	21
Open Data Chemnitz	city	75	9	9	9
Open Data Dortmund	city	335	66	16	16
Open Data Dresden	city	936	12	12	12
Open Data Duisburg	city	66	17	8	8
Open Data Düsseldorf	city	296	55	23	23
Open Data Frankfurt	city	88	2	1	1
Open Data Karlsruhe	city	100	2	4	4
Open Data Kiel	city	128	7	7	7
Open Data Köln	city	382	13	15	15
Open Data Leipzig	city	710	6	6	6
Open Data Moers	city	356	27	31	31
Open Data Nordrhein-Westfalen	state	3856	341	396	396
Open Data Potsdam	city	142	12	2	2
Open Data Rheinland-Pfalz	state	4511	0	56	56

Portal	Level	Datasets	Machine-readable	Datasets Education	Open license
Open Data Rostock	city	221	84	14	14
Open Data Wuppertal	city	178	2	3	3
Open Data-Portal München	city	153	2	2	2
Berlin Open Data	state/city	1655	86	14	12
Open Data Bavaria	state	901	110	55	55
Open Data Schleswig-Holstein	state	8272	254	372	367
Transparenzportal Bremen	state/city	153	1	3	3
Transparenzportal Hamburg	state/city	112,588	14	2426	2426
Federal Ministry of Education and Research	nation	264	180	264	264
GovData.de	nation	35,660	4061	2882	2521
Municipal Education Database	nation	816	2448	816	816

4.1 System quality

The analysis discovered 25 OGPs, which, as direct providers, publish open datasets in the field of education. They are at the city/state level. This survey counts 21 portals (three of them are also states) from cities having more than 100,000 inhabitants. Other OGPs (three) are aggregators at the national level. With eight exceptions, all of them use CKAN¹² as underlying software. Also, with two exceptions, all OGPs use either CKAN-API or DKAN-API¹³. One of the portals has a web service with similar functionality to CKAN-API. The web interface for searching provides advanced search capabilities for all the portals analyzed. All of them provide the capability to filter the information using several degrees of flexibility and dimensions. Most of them use CKAN as underlying software and CKAN-API (19 on their own, five via the national portal CKAN-API, and four via other kinds of web service) for automatic access.

The aggregators (the Federal Ministry of Education and Research and the Municipal Education Database) use a custom web service for automatic accessibility. The national portal uses CKAN-API.

12 <https://docs.ckan.org/en/2.9/api/>

13 <https://dkan.readthedocs.io/en/latest/introduction/index.html>

4.2 Information quality

We used several metrics to analyze information quality:

Thematic perspective. All of the OGP's group open datasets in themes. The themes differ slightly, depending on the various approaches adopted at the inception of each portal and/or administrative and stewardship reasons.

The number of datasets in education. This survey counts 7473 datasets in the education group for the whole of Germany. Most of the datasets are aggregated by the national portal, GovData.de, and OGP Hamburg.

Following de Juana-Espinosa (2019) and Yang and Wu (2016), the number of datasets concerning the population could be a measure of the possible usefulness of the portal. Therefore, Figure 1 depicts the number of datasets in education per 100,000 inhabitants of direct providers:

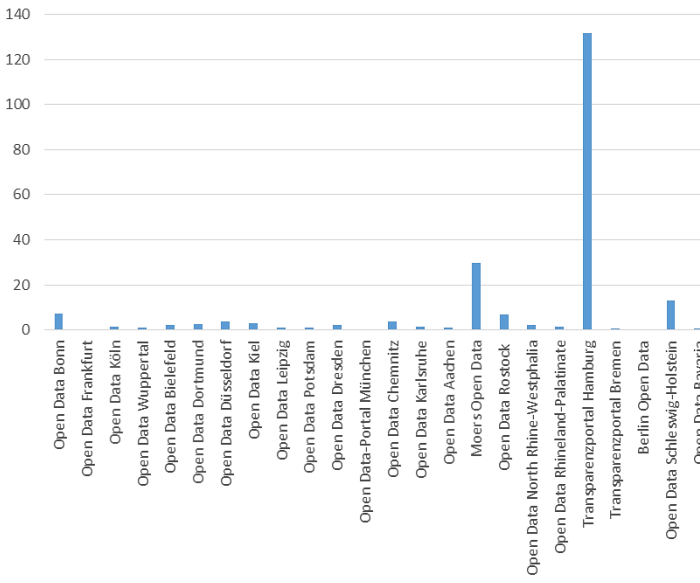


Fig. 1 Number of educational datasets/100,000 inhabitants of direct providers

See Figure 6 (a/b) for aggregator's numeric evaluation.

The number of organizations. The number of organizations that contribute to the OGDs varies from one to 19. This is an indicator that demonstrates the organizations' engagement in providing content to the portal and co-creating public knowledge as part of the ecosystem (Yang & Wu, 2016). Fig-

ure 2 gives a glimpse of this indicator for direct providers, whereas Figure 6 (a/b) represent the numeric indicators of aggregators at the national level.

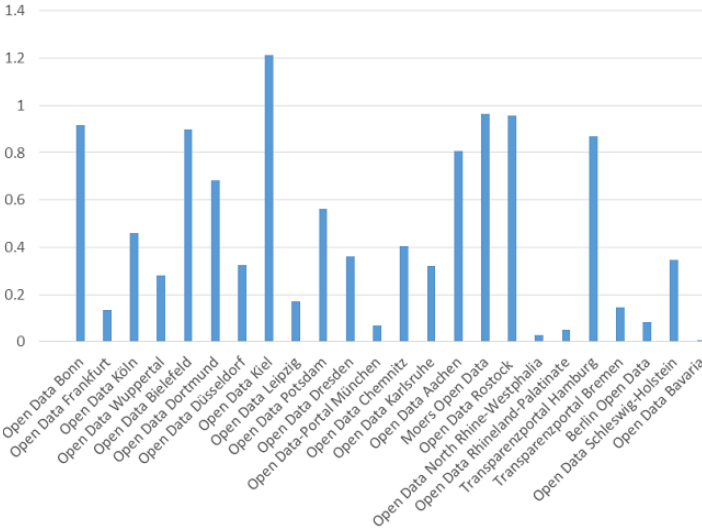


Fig. 2 Number of organizations/100.000 inhabitants of direct providers

The number of machine-readable formats. This number represents how many of the resources might be re-usable by automatic applications.

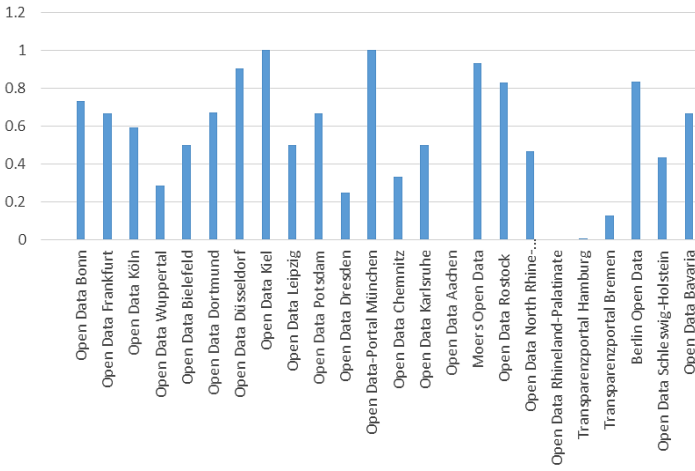


Fig. 3 The number of machine-readable formats/total number of resources formats of direct providers

The number of machine-readable formats per total number of resources content formats is represented in the Figure 3. This indicator is calculated at the resource level.

In the same way, Figure 6 (a/b) represents the number of machine-readable formats per total number of resources of aggregators at the national level.

Metadata format. In terms of metadata semantics, the most important initiative that a data portal should accommodate to facilitate interoperability, is an RDF vocabulary named Data Catalog Vocabulary (DCAT) by the World Wide Web Consortium (W3C).¹⁴ Using DCAT to describe datasets, publishers increase discoverability and enable applications to easily consume metadata from multiple catalogs (Máchová, 2017). Given the heterogeneity of the implementation at the state level, it is very desirable to consider only portals that implemented the DCAT-AP catalog standard. Of 28 OGPs, 22 adhere to the metadata format DCAT-AP¹⁵, which uses RDF.

Keywords or tags. In OGPs, tagging refers to the metadata assigned to a piece of content by the data provider. Tagging digital content enables data providers to structure, group, and order content around themes or topics, complementing the general categorization priority established by the portal administrators. The tags might be part of a vocabulary or might be termed freely, at the disposal of their creators. Also, they can be single words or compound words, as in the case of the GovData.de portal. At the cities/states level, seven OGPs are not using a system of keywords, and they rely on the coarse categorization of the thematic perspective. At the national level, only the national GovData.de portal uses an extensive system of tagging, the other two aggregators use a simple categorization of the datasets.

PID. The majority of the OGPs uses the permalinks. Only three portals use local identifiers for the PID.

4.3 Service quality

License. The number of open licenses measures the accessibility of datasets. The 'open license' is 94% of the total educational datasets, as this survey assessed the open licensed as opposed to closed or non-licensed datasets. It is

14 <https://www.w3.org/TR/vocab-dcat-2/>

15 <https://www.dcat-ap.de/>

interesting to point out the type of the license is attribution ('BY'), like 'Data license Germany – attribution – version 2.0', or attribution share-alike ('BY-SA'). The zero-type license, in the German version so-called 'Data license Germany – Zero – Version 2.0' or equivalent, is not so popular. Figure 4 represents this numeric indicator of direct providers.

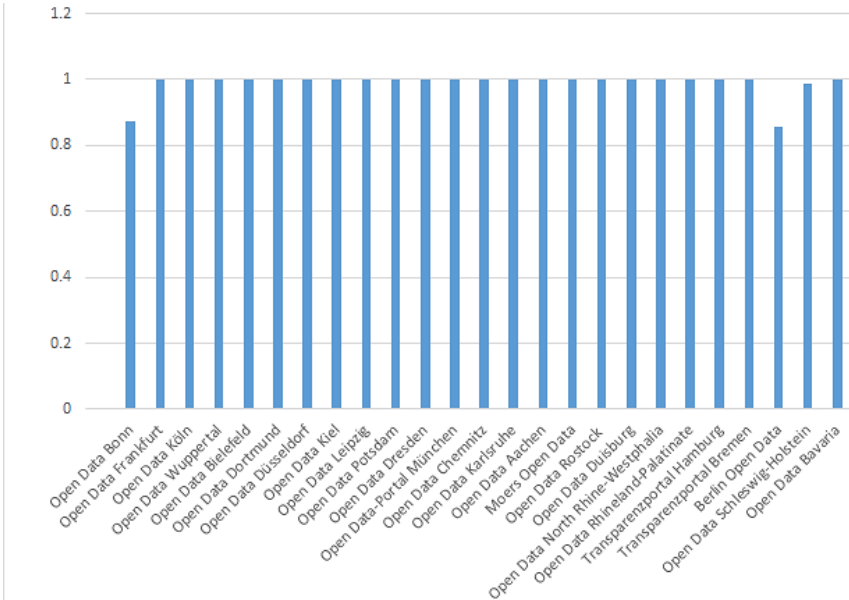


Fig. 4 Number of open licenses/total number of datasets of direct providers

At the aggregators' level, the national portal is 95% 'open licensed' of the total educational datasets (Fig. 6 b), whereas the rest of the aggregators are 100% 'open licensed', as is represented in Figure 6a.

Quality rating and user feedback measure the accuracy of the OGP. The accuracy is the measure of meeting the information need by the user. For OGP Hamburg and OGP Berlin, it was possible to assess the number of users' views (via API or web interface) for 2019, as they provided the statistical data as a dataset. Unfortunately, Hamburg has stopped counting the number of views in September 2019.

Several other portals present data ranking (Bonn, Potsdam, Bavaria) or blogs (Bonn, Cologne, Bielefeld). The adoption of blogs is mostly related to the DKAN software, or in the case of Bonn, with Semantic MediaWiki soft-

ware¹⁶. Both of them are content management systems based software, which allows for genuine user interaction.

It is worth to mention Open Data Schleswig-Holstein, as the portal presents a resources content classification upon 5 stars scale of Berners-Lee model¹⁷ on their web interface. This aspect gives a better chance for datasets re-usability (Farrow, 2014).

The number of applications reported by each of the OGP's is unknown in most portals; only eleven from 28 present this information in their web interfaces.

Re-use of data is considered a crucial indicator of OGP success since there is a symbiotic relationship between users and producers of Open Government Data (de Juana-Espinosa, 2019; Osagie, 2017). Figure 5 represents the number of application reported for direct providers.

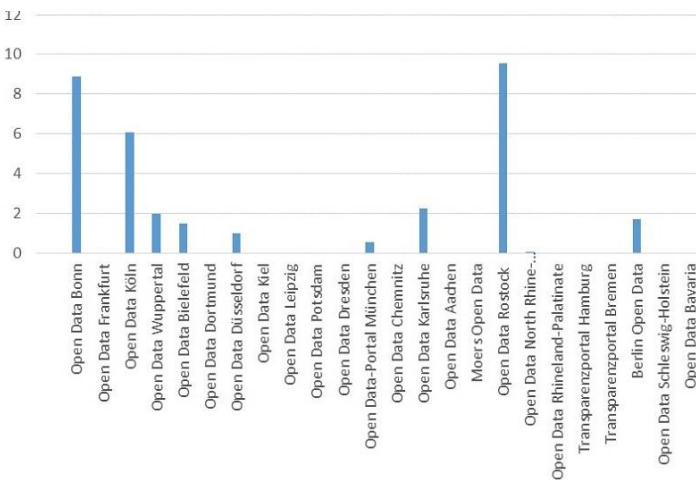


Fig. 5 Number of applications/100.000 inhabitants of direct providers

For the national portal, GovData.de, we use external information to assess the number of applications built on the datasets.

For the other aggregators at the ministries level, this study cannot find any references to the applications using directly their datasets.

16 https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

17 <https://5stardata.info/en/>

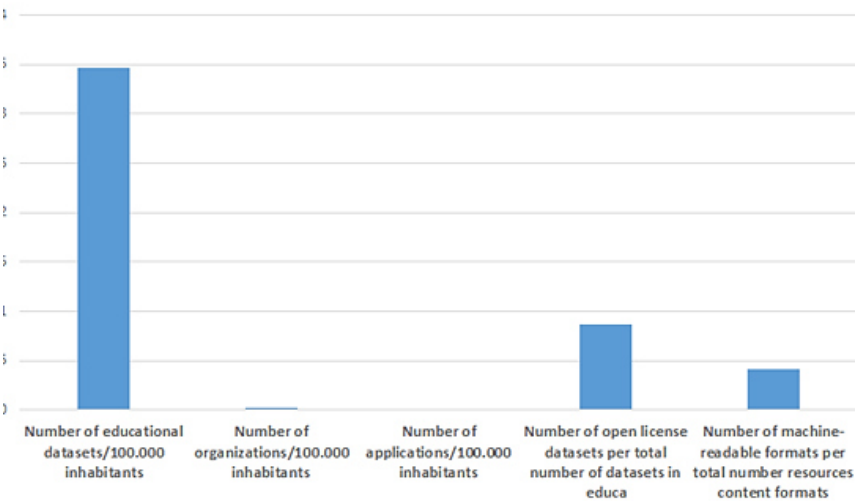
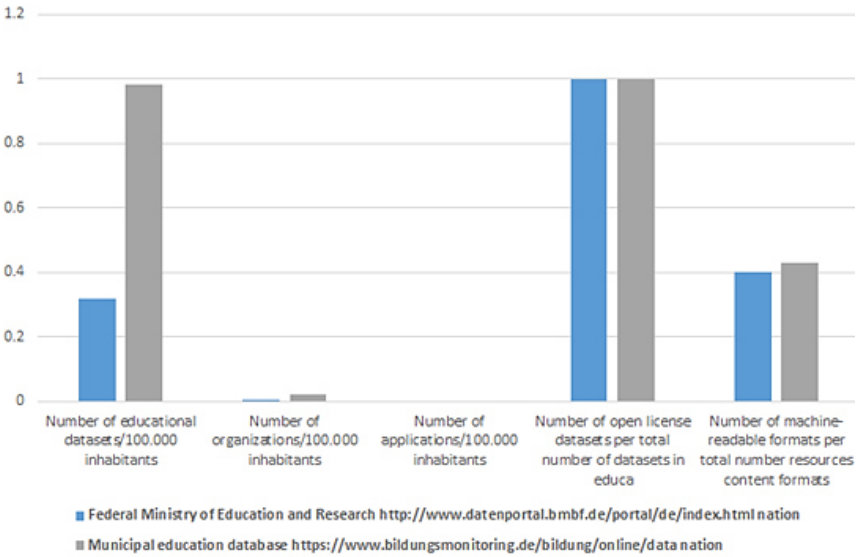


Fig. 6 National aggregators’ numeric indicators; a: on the top figure, the Federal Ministry of Education and Municipal Educational Database; b: on the bottom, GovData.de

5 Discussion

There are not many papers in literature, except that of Wang, Chen, and Richards (2018a) for China, which assesses the Open Government Data in education, at the local level. Other studies identified the main barriers preventing the re-usability of open datasets. A comprehensive study of open data usability was carried out by Alexopoulos (2017). The study highlighted education, health, and finances as public sector areas with major open data publication incidence, without a deep assessment of the quality of the datasets. Studies like Open Maturity Report, Open Data Barometer consider for evaluation only the datasets of GovData.de. The deep, thematic study of OGPs in Germany, with a focus on education, is still missing.

In the category of direct providers (cities/states), the OGP of the city of Rostock performs the best. This high score is given by the number of reported applications and a high number of machine-readable formats of datasets. It also adheres to CKAN-API and DCAT catalog format and presents statistics about the usage of the datasets. The city/state of Hamburg portal, even as they collect the highest number of educational datasets, needs improvements on the machine-readable format perspective, publishing most datasets in PDF format.

Open Data Bonn is one of the most interesting OGPs, performing well on the descriptive metrics but not very high on the quantitative metrics. Also, using a non-standard metadata format is a serious minus for this portal.

Open Data Bavaria portal was highly neglected in the last years; the last update of datasets dates back to 2015 at the time of this writing.

At the national level, the Federal Ministry of Education and Research and the Municipal education database aggregate datasets from the Federal Statistical Office¹⁸. There are several drawbacks, as they do not adhere to DCAT-AP standards, and access via API for third party applications is not easy. These portals lack reported application, interoperability, and metadata standardization.

On the second category of aggregators (national level), the national portal GovData.de performs best on all dimensions. They do not report the number of applications on the website, and user feedback capability was only recently considered.

18 <https://www.destatis.de/>

Overall, the service quality analysis signals an important drawback: The service's effectiveness cannot be measured due to the lack of important parameters: users' feedback, blogs, statistics of usage, and data ranking.

The 2020 Open Maturity Report¹⁹ of EU27+ situates Germany in the 8th place, up from 14th place in 2019. This report considers only the national portals at the EU level, not very relevant for federative countries. The thematic perspective is missing in the EU report; the open data in education is considered within the large mass of other open data. The report also mentioned that "Germany is one of the few countries where the national portal does not have a designated area for open data use cases", which is also reported in this study.

6 Conclusion

Our paper provides a preliminary overview of the quality of Germany's OGPs, including the perspective of usage in educational research. The study evaluates the OGPs at the city/state and national levels.

The study refines the information system success model from the indexed literature, put the emphasis to the re-usability, and attached evaluation metrics. The objective analysis considers descriptors and quantitative metrics for OGPs, with an emphasis on educational research. The analysis of the main 28 OGPs in Germany draws the following conclusions:

1. There is a solid German open data policy at the national level, which adheres to The European Directive 2019/1024 of June 20 (European Commission, 2019; Wiebe, 2020). This was also pointed out to the European Commission in other studies, like the EU28+ Open Data Maturity Report, in Blank (2019).
2. This study collected and analyzed around 7400 educational datasets and over 16,000 resources. The largest provider of educational data is Hamburg, and the largest aggregator is GovData.de.
3. Very few portals provide direct metadata or indirect (via the programmatic API) access to the usage metrics (i.e., views and download) in each portal. The national OGP GovData.de published only statistics of the

19 <https://www.europeandataportal.eu/en/dashboard/2020#table>

search queries. This practice makes it less immediate for researchers to evaluate a dataset's reception that might be of interest. This fact also leads to a disruption in feedback information for dataset providers, as they cannot assess their work's final impact.

4. There is also a lack of machine-readable formats. One of the most impressive portals in terms of the number of educational datasets, OGP Hamburg, publishes most of the datasets in PDF format. This is a serious drawback of the re-usability of the OGPs datasets, and automatic tools cannot find and use the appropriate content they are searching for.
5. The German version of metadata standard DCAT is used by 22 of the 28 portals, indicating that standardization is being used, but all portals should use it to enable better search in all portals. This standard, introduced by GovData.de, is aligned with high metadata standards (RDF), enabling better metadata interoperability, and this is a good start to a much more usage-oriented approach towards Linked Open Data. However, since not all OGPs use the metadata standards of GovData.de, it is not possible to find data from all cities/states in this portal. However, the usage of keywords is also not standardized, which made our data collection harder.
6. One of the best perspectives of the service quality dimension is licensing. 94% of the datasets are using an open license. The national portal GovData.de performs slightly better, 95%.

We have found the open educational data also have other possibilities of usage: The people involved in schools and education could profit from this data: teachers, parents, and people working in school administration. Since Germany's educational system is very heterogeneous, this study would also help compare the situation in the different states. Another potential of this study results might help guide parents who plan to move to another state with a different educational system.

In Germany, one large study about the educational system is the so-called "Bildungsbericht"²⁰ [English: Education Report]. This report gives an overview of the educational system in Germany every two years. In recent years, the trend is to have these reports at the state-level. These studies may benefit from this paper's results by making them more comprehensible to the public.

As part of our study, we encountered school data published in non-machine-readable formats on other local platforms with low accessibility

20 <https://www.bildungsbericht.de/de>

during this research. Bringing all this data together at the OGPs level would yield a great benefit and enable comparison among cities and states respectively. An easy presentation of this data also allows the general public to get deeper insights into the educational system and make educational research and educational policy more transparent.

7 Limitations of this work

For a comprehensive analysis of the topic, the information system success model allows for additional perspectives (Charalabidis, 2018). At the current stage of our research, the limitation of access to the relevant data prevented us from exploring them. For example, an important drawback is that no OGP shows how often a dataset has been downloaded. This information would give great feedback on the usage of the datasets. Only the OGPs of Hamburg and Berlin present a statistic of datasets views monthly. Both of the portals present search terms frequency statistics in two different formats. Only one OGP presents a statistic of datasets download and links to applications that use datasets from that portal (Open Data Rostock). Providing this information would primarily help the OGDs see what kind of data is used most, and it would also allow researchers to see what the most used data are across several OGDs, encouraging other OGDs to provide the same kind of data.

This study might be framed to a larger extent, for example, in a comparative study of educational Open Governmental Data of other federative countries, like Switzerland or the USA.

Acknowledgements

This research was funded in the context of the eHumanities Center CEDIFOR by The German Federal Ministry of Education and Research (BMBF) no. 01UG1416C.

Supplementary materials

Supplementary material associated with this article can be found at DOI: doi.org/10.5281/zenodo.4277847.

References

- Alexopoulos, C., Diamantopoulou, V., & Charalabidis, Y. (2017). Tracking the Evolution of OGD Portals: A Maturity Model. In Janssen M. et al. (Eds.), *Electronic Government. EGOV 2017*. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-64677-0_24
- Alves Neto, A. J., Neves, D. F., Santos, L. C., Rodrigues Jun., M. C., & do Nascimento, R. P. C. (2018). Open Government Data Usage Overview: A Systematic Literature Mapping. In *Euro American Conference on Telematics and Information Systems (EATIS '18)*, Fortaleza, Brazil. New York, NY: ACM. DOI: <https://doi.org/10.1145/3293614.3293619>
- Atenas, J., & Havemann, L. (2019). Open data and education. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The State of Open Data: Histories and Horizons* (pp. 91–102). Cape Town, Ottawa: African Minds and International Development Research Centre.
- Atenas, J., Havemann, L., Nascimbeni, F., Villar-Onrubia, D., & Orlic, D. (2019). Fostering Openness in Education: Considerations for Sustainable Policy-Making. *Open Praxis*, 11(2), 167–183.
- Benitez-Paez, F., Degbelo, A., Trilles, S., Huerta, J. (2018). Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain: A Data User's Perspective. *ISPRS International Journal of Geo-Information*, 7(1), 6.
- Berends, J., Carrara, W., Engbers, W., & Vollers, H. (2017). *Re-using Open Data*. Technical Report. Luxembourg: Publications Office of the European Union.
- Blank, M. (2019). *The Open Data Maturity Report 2019*. European Commission and Cagemini Invent. doi: [10.2830/073835](https://doi.org/10.2830/073835)
- Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The State of Open Data – Limits of Current Open Data Platforms. In *Proceedings of the 21st World Wide Web Conference 2012*, Web Science Track at WWW'12, Lyon, France, April 16–20, 2012. New York, NY: ACM.
- Buteau, S., Rao, P., Mehta, A. K., & Kadirvell, V. (2018). Developing a Framework to Assess Socio-Economic Value of Open Data in India. In *Proceedings of The 14th International Symposium on Open Collaboration – OpenSym '18* (Article No. 8). New York, NY: ACM, DOI: <https://doi.org/10.1145/3233391.3233532>
- Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L.-D., Kacprzak, E., & Groth, P. (2020). Dataset search: a survey. *The VLDB Journal*, 29, 251–272. DOI: <https://doi.org/10.1007/s00778-019-00564-x>
- Charalabidis, Y., Zuidervijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018). Open Data Evaluation Models: Theory and Practice.

- In *The World of Open Data. Concepts, Methods, Tools and Experiences* (pp. 137–172). Cham: Springer International Publishing.
- Corrêa, A. S., & da Silva, F. S. C. (2019). Laying the foundations for benchmarking open data automatically: a method for surveying data portals from the whole web. In *dg. o 2019: 20th Annual International Conference on Digital Government Research, 2019, Dubai, United Arab Emirates* (pp. 287–296). New York, NY: ACM. DOI: <https://doi.org/10.1145/3325112.3325257>
- Corrêa, A. S., Couto de Paula, E., Corrêa, P. L.P., & da Silva, F. S. C. (2017). Transparency and open government data: A wide national assessment of data openness in Brazilian local governments, *Transforming Government: People, Process and Policy*, 11(1), 58–78. DOI: <https://doi.org/10.1108/TG-12-2015-0052>
- de Juana-Espinosa, S., & Luján-Mora, S. (2019). Open government data portals in the European Union: Considerations, development, and expectations. *Technological Forecasting and Social Change*. 149(Dec.), Art. No. 119769. DOI: [10.1016/j.techfore.2019.119769](https://doi.org/10.1016/j.techfore.2019.119769)
- dos Santos Brito, K., da Silva Costa, M. A., Cardoso Garcia, V., & de Lemos Meira, S. R. (2015). Is Brazilian Open Government Data Actually Open Data? An Analysis of the Current Scenario. *International Journal of E-Planning Research*, 4(2), 57–73.
- European Commission. 2019. Directive 2019/1024 of June 20, 2019 on open data and re-use of public sector information. *Official Journal of the European Union*, L 172/56, June 26, 2019.
- Fane, B. (2019). What is the State of Open Data in 2019? In *The State of Open Data 2019 – A selection of analyses and articles about open data*. Figshare, Digital Science. <https://knowledge.figshare.com/articles/item/state-of-open-data-2019>
- Farrow, R. (Ed.) (2014). *Open Education Handbook*. <http://education.okfn.org/handbooks/handbook/>
- Gerunov, A. (2016). Understanding Open Data Policy: Evidence from Bulgaria. *International Journal of Public Administration*. 40(8), 1–9. DOI: [10.1080/01900692.2016.1186178](https://doi.org/10.1080/01900692.2016.1186178)
- Katte, A. (2017). Open Data Movement and its impact on the world. <https://your-story.com/2017/10/open-data-movement-and-its-impact-on-the-world>
- Klein, M. (2017). Was ist Open Data? *e-Government Computing*. 27.2.2017. <https://www.egovernment-computing.de/was-ist-open-data-a-693134/>
- Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13–29. DOI: <https://doi.org/10.1016/j.giq.2017.11.003>

- Máchová, R., Hub, M., & Lněnička, M. (2018). Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and re-usability from a stakeholders' perspective. *Aslib Journal of Information Management*, 70(3), 252–268. DOI: <https://doi.org/10.1108/AJIM-02-2018-0026>
- Máchová, R., & Lněnička, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12(1), 21–41. DOI: <https://dx.doi.org/10.4067/S0718-18762017000100003>
- Marjanovic, O., & Cecez-Kecmanovic, D. (2017). Exploring the tension between transparency and datification effects of open government IS through the lens of Complex Adaptive Systems. *The Journal of Strategic Information Systems*, 26(3), 210–232. DOI: <https://doi.org/10.1016/j.jsis.2017.07.001>
- Molinari, A., Maltese, V., Vaccari, L., Almi, A., & Bassi, E. (2014). Big Data and Open Data for a Smart City, IEEE-TN Smart Cities White Papers, Trento, Italy. DOI: [10.13140/RG.2.2.33170.56001](https://doi.org/10.13140/RG.2.2.33170.56001)
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49–56. DOI: [10.3233/ISU-170824](https://doi.org/10.3233/ISU-170824)
- Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated Quality Assessment of Metadata across Open Data Portals. *Journal of Data and Information Quality*, 8(1), 2:1–2:29. DOI: <https://doi.org/10.1145/2964909>
- Open Data Charter, <https://opendatacharter.net/principles/>
- Osagie, Edobor, Mohammad Waqar, Samuel Adebayo, Arkadiusz Stasiewicz, Lukasz Porwol, and Adegboyega Ojo. (2017). Usability Evaluation of an Open Data Platform. In *Proceedings of the 18th Annual International Conference on Digital Government Research (dg.o '17)* (pp. 495–504). New York, NY: ACM,. DOI: <https://doi.org/10.1145/3085228.3085315>
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. *PLOS ONE*, 11(1), e0146695. DOI: <https://doi.org/10.1371/journal.pone.0146695>
- SPARC (the Scholarly Publishing and Academic Resources Coalition), <https://sparc-open.org/open-education/>
- United Nations (2016). *United Nations E-Government Survey 2016: E-Government in Support of Sustainable Development*. New York, NY: United Nations Publications. https://drive.google.com/file/d/1VpiL5islLigFMKnmlM5_Q_upwS7Ey6RD/view
- United Nations (2018). *United Nations E-Government Survey 2018: Gearing E-Government to Support transformation towards Sustainable and Resilient Societies*. New York, NY: United Nations Publications. <https://publicadministra->

[tion.un.org/Portals/1/Images/E-Government%20Survey%202018_FINAL%20for%20web.pdf](https://www.un.org/Portals/1/Images/E-Government%20Survey%202018_FINAL%20for%20web.pdf)

- Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, 35(4), 644–656. DOI: <https://doi.org/10.1016/j.giq.2018.10.006>
- Wang, V., Button, M., & Shepherd, D. (2018). The Barriers to the Opening of Government Data in the UK – A View From the Bottom. *Information Polity*, 24(1), 59–74. DOI: <http://dx.doi.org/10.2139/ssrn.3363165>
- Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: an empirical view from the Citizen’s perspective. *Information Systems Frontiers*, 19, 285–300. DOI: <https://doi.org/10.1007/s10796-016-9679-1>
- Wiebe, A. (2020). *Open Data in Deutschland und Europa – Vorschlag zur Weiterentwicklung des rechtlichen Rahmens einer Informationsordnung (Open Data – Public-Service-Information Richtlinie)*. Berlin: Konrad-Adenauer-Stiftung.
- Wieczorkowski, J. (2019). Barriers to Using Open Government Data. In *ICEEG 2019: Proceedings of the 2019 3rd International Conference on E-Commerce, E-Business and E-Government, June 2019* (pp. 15–20). New York, NY: ACM. DOI: <https://doi.org/10.1145/3340017.3340022>
- Yang, T.-M., & Wu, Y.-J. (2016). Examining the socio-technical determinants influencing government agencies’ open data publication: a study in Taiwan. *Government Information Quarterly*, 33(3), 378–392. DOI: <https://doi.org/10.1016/j.giq.2016.05.003>
- Zuiderwijk, A., Shinde, R., & Janssen, M. (2019). Investigating the attainment of open government data objectives: is there a mismatch between objectives and results? *International Review of Administrative Sciences*, 85(4), 645–672. DOI: <https://doi.org/10.1177/0020852317739115>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 370–396. DOI: doi.org/10.5283/epub.44955.

Potenziale von Open Educational Resources in der Informationswissenschaft?

*Sylvia Kullmann¹, Johannes Hiebl¹,
Tamara Heck¹, Marc Rittberger^{1,2}*

¹ Information Center for Education, DIPF | Leibniz Institute for Research
and Information in Education, Frankfurt/M., Germany

² Darmstadt University of Applied Science
{[Kullmann](#), [Hiebl](#), [Heck](#), [Rittberger](#)}@dipf.de

Abstract

Open Educational Resources (OER) werden durch ihre offene Lizenzierung und den damit verbundenen Möglichkeiten zur Nachnutzung Potenziale zur Verbesserung von Lern-/Lehrsituationen zugeschrieben. In unserer Mixed-Methods-Studie erheben wir die Einschätzung zum Nutzen von OER für die Lehrpraxis von Lehrenden der deutschsprachigen Informationswissenschaft. Außerdem fragen wir nach Umsetzungsmöglichkeiten eines gemeinsamen OER-Pools und den Einstellungen zu gemeinschaftlichen Erstellungsprozessen von Lern-/Lehrmaterial. Im folgenden Posterbeitrag werden erste Ergebnisse einer qualitativen Befragung ausgewählter Lehrender vorgestellt, die Basis für eine darauf aufbauende quantitative Befragung unter Lehrenden der deutschsprachigen Informationswissenschaft ist.

Keywords: Open Educational Resources; OER; Informationswissenschaft; Hochschullehre

1 Hintergrund und Forschungsfragen

Open Educational Resources (OER) sind offene Lern-/Lehrressourcen, die von Akteuren erstellt, geteilt und in gleicher oder bearbeiteter Form wiederverwendet werden können. Der Unterschied zu anderen Lern-/Lehrmateri-

alien ist eine offene Lizenz (Creative Commons¹), die diese Nutzung erlaubt. OER weisen unterschiedliche Grade an Offenheit auf und liegen in verschiedenen medialen Formen und Granularitäten vor (Kerres et al., 2015). Ihnen wird insbesondere in Verbindung mit offenen Lehrpraktiken (Ehlers, 2011) und offener Pädagogik (Wiley et al., 2018) eine nachhaltige Verbesserung der Qualität von Lern- und Lehrprozessen zugeschrieben (Orr et al., 2015; Farrow et al., 2020).

Der Zusammenarbeit zwischen Lehrenden wird mit Blick auf die Erstellung von OER besondere Bedeutung zugemessen (Kerres, 2002; Orr et al., 2015). Ein Beispiel für eine solche Zusammenarbeit bildet das Repositorium für mathematische Selbstlernmaterialien für Studierende, optes². Optes stellt ein gemeinschaftlich entwickeltes OER-Lernangebot zu einem disziplinübergreifenden Thema mit dem Ziel dar, mathematisches Grundlagenwissen von Studierenden verschiedener Fächer zu optimieren.

In der Informationswissenschaft existieren bereits verschiedene Beispiele für eine Zusammenarbeit im Bereich der Lehre. Zu nennen sind hier u.a. gemeinsame Kursangebote mit mehreren beteiligten Institutionen (Griesbaum et al., 2005; Semar et al., 2004). Zudem gab es Bestrebungen, einen Ressourcenpool an Lern-/Lehrmaterialien für diverse Fachthemen aufzubauen (Weisbrod et al., 2004; Leichtweiß et al., 2004). Im Projekt nestor wurden Qualifizierungsangebote zur Langzeitarchivierung entwickelt (Nestor Handbuch, 2010). Ein neueres Beispiel ist der offen lizenzierte und mehrsprachige Kurs zur Vermittlung von Informationskompetenz, Information Literacy Online (Libbrecht et al., 2019).³

Im Zuge der Forderungen nach offenen Lern-/Lehrmaterialien stellt sich die Frage, wie eine wissenschaftliche Fachgemeinschaft mit der Entwicklung von offenen Lernangeboten umgeht. Von besonderem Interesse ist dabei, ob eine fachliche Community OER als Möglichkeit zur Verbesserung ihrer Lehre einstuft.

Die folgende Studie hat zum Ziel, Potenziale von OER für die Lehre in der deutschsprachigen Informationswissenschaft zu untersuchen. Sie ordnet sich in die Reihe von Untersuchungen zum Umgang mit OER und digitalen Infrastrukturen in der Hochschullehre ein (Perez-Paredes et al., 2018; Weller et al., 2015; Beaven, 2013) und fokussiert hierbei die Perspektive der Hoch-

1 <https://creativecommons.org/>, 13.01.2021

2 <https://optes.de/>, 13.01.2021

3 <https://informationliteracy.eu/de>, 13.01.2021

schullehrenden. Darüber hinaus wird das Potenzial von OER im Hinblick auf eine Fachgemeinschaft und deren Praktiken der Zusammenarbeit betrachtet, was den Blick auf die bisher untersuchten Potenziale von OER erweitert.

Die zentrale Forschungsfrage lautet: *Welche Potenziale und Hürden von OER sehen Lehrende für die Lehre in der Informationswissenschaft?*

2 Methode und Prozess

Die Studie ist als Mixed-Methods-Ansatz in Form einer sequenziellen Verbindung qualitativer und quantitativer Methoden (Flick, 2017, S. 43 f. u. 2018, S. 21 f.) geplant. In ca. einstündigen halb-strukturierten Experten-Interviews (Flick, 2017, S. 214–219 u. 2018, S. 236–239) werden ausgewählte Lehrende der deutschsprachigen Informationswissenschaft zunächst zu ihrer Nutzung und Erstellung von (offenen) Lern-/Lehrmaterial befragt. Die Interviews finden über ein Online-Konferenztool statt. Zur Auswahl der Interviewpartner wird die Mitgliedsliste des Hochschulverbands Informationswissenschaft (HI) herangezogen. Teilnehmende sollen die Idee von OER bereits kennen. Die Merkmale werden bei der Einladung zur Studie abgefragt und auch über persönliche Kontakte und Erfahrung ermittelt.

Die Fragen zielen neben der Erhebung von bereits gemachten Erfahrungen mit OER insbesondere auf die Einschätzung der Potenziale und Hürden dieser ab. Dabei wird nach dem Prinzip der theoretischen Sättigung (Flick, 2017, S. 158–166 u. 2018, S. 176–181) vorgegangen und nach den ersten fünf Interviews und der Analyse der Daten entschieden, ob weitere Teilnehmende befragt werden. Ziel dieser explorativen Phase ist, einen Erkenntnisstand zum Thema zu erreichen, auf Basis dessen eine repräsentative Umfrage entwickelt werden kann. Die Interviews werden nach den zu betrachtenden Merkmalen angemessen transkribiert (ebd., S. 379–384 u. 2018, S. 438–442) und mithilfe der qualitativen Inhaltsanalyse (ebd., S. 409–419 u. 2018, S. 481 bis 489) deduktiv nach den zentralen Leitfragen ausgewertet.

Die Ergebnisse, zusammen mit Erkenntnissen aus weiteren Untersuchungen (Heck et al., 2020; Deutscher Bildungsserver, 2016), dienen als Basis für die repräsentative Umfrage in der deutschsprachigen Informationswissenschaft.

3 Erhebung von Potenzialen und Hürden von OER

Die Leitfragen schließen an aktuelle Diskussionen aus der Forschung zu OER und OER-Infrastrukturen an. Im Folgenden werden die relevanten Aspekte kurz dargelegt.

3.1 Optimierung von Lern-/Lehrmöglichkeiten

OER wird allgemein das Potenzial zur Verbesserung von Lern-/Lehrmöglichkeiten zugeschrieben. Allerdings fehlen hinreichende Studien, die OER mit Blick auf den Lernerfolg untersuchen, bspw. im Vergleich zu Lernen ohne OER (Wiley et al., 2018). Weiterhin sehen Hochschullehrende den Bedarf nach fachspezifischen Einstiegen in die OER-Suche (Baas et al., 2019). Dies wird oft als Grund dafür angegeben, dass Lehrende OER nicht nutzen und auch nicht erstellen (Heck et al., 2020). Das Potenzial von OER in der Hochschulpraxis muss also noch belegt werden, für Lehrende als auch für Lernende. Im Rahmen unserer Studie wird untersucht, inwiefern Lehrende OER als mögliche Verbesserungen für die Lehre in der Informationswissenschaft einstufen und wie diese realisiert werden können.

3.2 Wertschätzung von Lehre

Für Lehrende wird durch OER die Möglichkeit gesehen, Anerkennung für ihre Arbeit zu erhalten, da ihre Aktivitäten in der Lehre durch OER und offene Praktiken sichtbar werden (D'Antoni, 2009). Ähnlich ist die Idee des Open Scholarship (Tennant et al., 2019), die Forschung und Lehre zugänglich und inklusiv gestalten will. Offene Fragen sind hierbei, welchen Impact Hochschullehrende auf Lernende und die Fachgemeinschaft haben, wenn sie OER erstellen und teilen, und inwiefern sie damit zu einer inklusiveren und qualitativ hochwertigeren Lehre beitragen. Zudem ist eine reine offene Lizenzierung nicht ausschlaggebend für die Verbesserung von Lernmöglichkeiten (Wiley et al., 2018). Die Studie untersucht, wie Lehrende das Potenzial zur Erhöhung der Anerkennung der eigenen Lehrtätigkeit einschätzen und ob diese eine Motivation zum Engagement im OER-Kontext darstellt.

3.3 OER als Objekt und als Prozess

Ein Ziel ist, dass OER gemäß den ‚5 R‘ nach Wiley (2014) von möglichst vielen Lehrenden verbessert, wieder genutzt und verbreitet werden. Die Nachnutzung bereits bestehender OER kann dazu beitragen, den Aufwand des Erstellens von qualitativ hochwertigem Lern-/Lehrmaterial für den einzelnen Lehrenden zu verringern und gleichzeitig den Lebenszyklus von Lern-/Lehrmaterial zu erhöhen (Orr et al., 2015). Konkret bedeutet dies, dass sich Lehrende aus einem Pool von OER ihres Faches „das Beste“ mit Blick auf ihre Lehrinhalte und -kontexte herausnehmen und veränderte, aktualisierte oder neu erstellte Materialien in diesen zurückführen können. Diese ideologische Idee von OER gilt es in der Praxis der Zielgruppe, den Hochschullehrenden, genauer zu untersuchen. Hochschullehrende sind meistens auch Wissenschaftler. Zum einen bestehen innerhalb eines Faches unterschiedliche Lehrmeinungen. Zum anderen gibt es eine gewisse Konkurrenz, auch im Hinblick auf Studiengangsangebote.

Diese Überlegungen schließen an die Forschungsfragen an, ob es einen gemeinsamen OER-Pool in der Informationswissenschaft geben – und wenn ja, wie dieser mit Blick auf die Materialien (mediale Formen, Granularitäten) und Funktionen (infrastrukturelle Merkmale) konkret ausgestaltet werden sollte.

3.4 Intensivierung der Zusammenarbeit

Für eine Fachgemeinschaft kann nicht nur das individuelle Erstellen von Ressourcen aus geteiltem Material interessant sein, sondern auch die gemeinsame Entwicklung offener Materialien (Kerres, 2002). Die oben genannten Beispiele zeigen, dass es dies in der Informationswissenschaft schon gibt. Das Konzept OER könnte mitunter die Sichtbarkeit und inklusivere Zugänglichkeit dieser Aktivitäten erhöhen. Allerdings gibt es hierbei kritische Fragen. Die Studie eruiert daher auch, inwiefern eine Zusammenarbeit bei der Erstellung von Materialien von Lehrenden in der Informationswissenschaft überhaupt favorisiert wird.

Das Poster wird erste Ergebnisse aus den Interviews vorstellen.

Referenzen

- Baas, Marjon; Admiraal, Wilfried; van den Berg, Ellen (2019): Teachers' Adoption of Open Educational Resources in Higher Education. *Journal of Interactive Media in Education* (1). <https://doi.org/10.5334/jime.510>
- Beaven, Tina (2013): Use and Reuse of OER: professional conversations with language teachers, *Journal of e-Learning and Knowledge Society* 9 (1), 59–71.
- Deutscher Bildungsserver (2016): Machbarkeitsstudie zum Aufbau und Betrieb von OER-Infrastrukturen in der Bildung. <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-117154>
- Ehlers, Ulf-Daniel (2011): Extending the Territory: From Open Educational Resources to Open Educational Practices. *Journal of Open, Flexible, and Distance Learning* 15 (2), 1–10. <https://www.learntechlib.org/p/147891>
- Farrow, Robert; Pitt, Rebecca; Weller, Martin (2020): Open Textbooks as an Innovation Route for Open Science Pedagogy. *Education for Information* 36 (3), 227 bis 245. <https://doi.org/10.3233/EFI-190260>
- Flick, Uwe (2017): *Qualitative Sozialforschung: Eine Einführung*. 8. Aufl., Reinbek: Rowohlt-Taschenbuch-Verlag.
- Flick, Uwe (2018): *An Introduction to Qualitative Research*. 6. Aufl., Los Angeles, London: Sage.
- Griesbaum, Joachim; Rittberger, Marc (2005): A collaborative lecture in information retrieval for students at universities in Germany and Switzerland. In: *World Library and Information Congress: 71th IFLA General Conference and Council "Libraries – a Voyage of Discovery"*, Oslo, Norway. https://archive.ifla.org/IV/ifla71/papers/o68e-Griesbaum_Ritterberg.pdf
- Heck, Tamara; Kovalenko, Valentyna; Rittberger, Marc (2020): User Experience to Inform the Design of a Search Infrastructure for Open Educational Resources. In: Sundqvist, A.; Berget, G.; Nolin, J.; Skjerdingsstad K. (Hrsg.): *Sustainable Digital Communities. iConference 2020*. Cham: Springer International Publishing, S. 419–427. https://doi.org/10.1007/978-3-030-43687-2_33
- Heck, Tamara; Peters, Isabella; Mazarakis, Athanasios; Scherp Ansgar; Blümel, Ina (2020): Open science practices in higher education: Discussion of survey results from research and teaching staff in Germany. *Education for Information* 36 (3), 301–323. <https://doi.org/10.3233/EFI-190272>
- Kerres, Michael; Heinen, Richard (2015): Open informational ecosystems: The missing link for sharing resources for education. *The International Review of Research in Open and Distributed Learning* 16 (1). <https://doi.org/10.19173/irrodl.v16i1.2008>

- Kerres, Michael (2002): Medien und Hochschule. Strategien zur Erneuerung der Hochschullehre. In: Issing, Ludwig J.; Stärk, Gerhard (Hrsg.): *Studieren mit Multimedia und Internet Ende der traditionellen Hochschule oder Innovationsschub?* Münster: Waxmann. http://mediendidaktik.uni-due.de/sites/default/files/kerres-tu-darmstadt_o.pdf
- Leichtweiß, Angela; Oßwald, Achim; Weisbrod, Dirk; Knorz, Gerhard (2004): 2 MN – Evaluation digitaler Wissensvermittlung. In: *26. Online-Tagung der DGI, Frankfurt a. M., Germany, 15.–17.06.2004*. <https://fizi.fh-potsdam.de/volltext/dgio4/05462.pdf>
- Libbrecht, Paul; Dreisiebner, Stefan; Buchal, Björn; Polzer, Anna (2019): Creating Multilingual MOOC Content for Information Literacy: A Workflow. In: *Conference on Learning Information Literacy across the Globe. Frankfurt am Main, Germany*. https://informationliteracy.eu/conference/assets/papers/LILG-2019_Libbrecht-et-al_Creating_ILO_MOOC.pdf
- Nestor Handbuch (2020): *Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3*, hrsg. v. Neuroth, H.; Oßwald, A.; Scheffel, R.; Strathmann, S.; Huth, K. Projekt: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland. <http://nbn-resolving.de/urn:nbn:de:0008-2010071949>
- Orr, Dominic; Rimini, Michele; Van Damme, Dirk (2015): Open Educational Resources: A Catalyst for Innovation. *Educational Research and Innovation*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264247543-en>
- Perez-Paredes, Pascual; Ordoñana Guillamón, Carlos; Aguado, Pilar (2018): Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning* 31 (5–6), 522–545. <https://doi.org/10.1080/09588221.2017.1418754>
- Semar, Wolfgang; Griesbaum, Joachim; König-Mistic, Jagoda; Lenich Andreas; Schütz, Thomas (2004): K3 – Wissensmanagement über kooperative verteilte Formen der Produktion und der Aneignung von Wissen zur Bildung von konzeptueller Informationskompetenz durch Nutzung heterogener Informationsressourcen – Stand und Aussichten. In: Hammwöhner, R.; Rittberger, M.; Semar, W. (Hrsg.): *Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen*. Konstanz: UVK Verlagsgesellschaft, S. 329–347. <https://doi.org/10.5281/zenodo.4137154>
- Tennant, Jonathan; Beamer, Jennifer; Bosman, Jeroen; Brembs, Björn; Chung, Neo Christopher; Clement, Gail ... Turner, Andy (2019): Foundations for Open Scholarship Strategy Development. Preprint DOI: <https://doi.org/10.31222/osf.io/b4v8p>
- Weisbrod, Dirk; Oßwald, Achim (2004): Module für die multimediale netzbasierte Hochschullehre: Abschlussbericht des Projektes „2 MN – Evaluation“ am Institut

für Informationswissenschaft der FH-Köln (= Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; Bd. 44). <https://epb.bibl.th-koeln.de/front-door/deliver/index/docId/88/file/bando44.pdf>

Weller, Martin; de los Arcos, Bea; Farrow, Robert; Pitt, Beck; McAndrew, Patrick (2015): The impact of OER on teaching and learning practice. *Open Praxis* 7 (4), 351–361. <http://dx.doi.org/10.5944/openpraxis.7.4.227>

Wiley, David (2014): The access compromise and the 5th R. <https://opencontent.org/blog/archives/3221>

Wiley, David; Hilton, John (2018): Defining OER-Enabled Pedagogy. The International Review of Research. *Open and Distributed Learning* 19 (4), 133–147. <https://doi.org/10.19173/irrodl.v19i4.3601>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 397–404. DOI: doi.org/10.5283/epub.44956.

Forschungsdatenmanagement in Brandenburg (FDM-BB)

Aufbau einer Landesinitiative

Heike Neuroth

Fachhochschule Potsdam,
Deutschland

heike.neuroth@fh-potsdam.de

Niklas Hartmann

Universität Potsdam,
Deutschland

niklas.hartmann@uni-potsdam.de

Janine Straka

Fachhochschule Potsdam,
Deutschland

janine.straka@fh-potsdam.de

Ina Radtke

Universität Oldenburg, Deutschland

ina.radtke@uni-oldenburg.de

Ulrike Wuttke

Fachhochschule Potsdam,
Deutschland

ulrike.wuttke@fh-potsdam.de

Abstract

Bundeslandinitiativen mit dem Schwerpunkt Forschungsdatenmanagement (FDM) spielen neben der Nationalen Forschungsdateninfrastruktur (NFDI) eine wichtige Rolle, um Kompetenzen, Beratungsangebote, technologische Dienstleistungen und dergleichen „in die Fläche zu bringen“. Der Beitrag stellt am Beispiel FDM-BB vor, wie in wissenschaftsgeleiteter Weise Handlungsempfehlungen erarbeitet und die Entwicklung einer Landesstrategie zu Forschungsdatenmanagement unterstützt werden kann. Mit einer solchen Strategie sollen alle brandenburgischen Hochschulen – unter Berücksichtigung des Diskussions- und Entwicklungsstandes der außeruniversitären Forschungseinrichtungen – in zwei Bereichen befähigt werden: (1) dem Aufbau von FDM-Aktivitäten vor Ort und (2) der kooperativen (Weiter-)Entwicklung gemeinsamer FDM-Dienste. Damit wird die Anschlussfähigkeit der brandenburgischen Hochschulen zu den nationalen Entwicklungen (NFDI) und der European Open Science Cloud (EOSC) hergestellt und das koordinierte Vorgehen der Einrichtungen unterstützt.

Keywords: Forschungsdatenmanagement; FDM; Bundeslandinitiative; FDM-BB

1 Ziele

Ziel des Projekts Forschungsdatenmanagement in Brandenburg (FDM-BB) ist die Erarbeitung von Handlungs- und Implementierungsempfehlungen für die Entwicklung einer bedarfsorientierten Forschungsdatenstrategie in Brandenburg, um Forschungsdaten den FAIR-Prinzipien¹ folgend intellektuell und maschinenlesbar nachnutzen zu können (Wilkinson et al., 2016). Damit wird der Forderung nach zugänglichen Daten und (inter)nationalen Entwicklungen Rechnung getragen, wie z.B. die Erarbeitung umfangreicher Datenstrategien durch die Europäische Kommission und die Regierung der BRD (DFG, 2019).² Im Gegensatz zu Open Access steht nicht nur das Publizieren von Texten im Mittelpunkt, sondern die ganzheitliche Unterstützung, Begleitung und Verbesserung des Forschungsprozess mit digitalen Methoden und Diensten durch forschungsnahe Beratungsangebote, nachhaltige (technologische) Infrastrukturen, wissenschaftliche Dienste und Qualifizierung. Die brandenburgische Forschungsdatenstrategie wird in enger Zusammenarbeit aller Hochschulen und außeruniversitären Forschungseinrichtungen Brandenburgs erarbeitet.³

2 Einordnung der Bundeslandinitiative

Auf Bundeslandebene gibt es in Deutschland zurzeit sechs geförderte FDM-Initiativen.⁴ Von diesen von einem Landesministerium finanzierten Bundes-

1 <https://www.go-fair.org/fair-principles/>

2 Vgl. European Commission: Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee und The Committee of the Regions. 2020. A European strategy for data. Brüssel. COM (2020) 66 final sowie Eckpunkte der Datenstrategie der Bundesregierung, s. <https://www.bundesregierung.de/breg-de/themen/digitalisierung/datenstrategie-1693546>.

3 Auch wenn die außeruniversitären Forschungseinrichtungen durch das Projekt FDM-BB nicht direkt adressiert werden konnten, beteiligen sie sich über eine Arbeitsgruppe an den Diskussionen und Ergebnissen.

4 Baden-Württembergisches Begleit- und Weiterentwicklungsprojekt für Forschungsdatenmanagement (bw2FDM), Forschungsdatenmanagement in Brandenburg (FDM-BB), Hamburg Open Science (HOS), Hessische Forschungsdateninfrastrukturen

landinitiativen ist FDM-BB die jüngste Initiative. Auch wenn FDM-BB im Vergleich zu den anderen deutschen FDM-Bundeslandinitiativen qua Laufzeit und Fördersumme eher als klein zu bezeichnen ist, ist das Identifizieren der wesentlichen Bausteine für die Weiterentwicklung eines systematischen institutionellen und bundeslandweit kooperativen Ansatzes von Relevanz. Eine Besonderheit in Brandenburg ist dabei, dass Universitäten und Fachhochschulen von Beginn an gleichermaßen beteiligt waren.

3 Vorgehensweise und Ergebnisse

In Brandenburg wird seit November 2019 für die Dauer von zunächst 14 Monaten eine Landesinitiative FDM analog zu den anderen Bundeslandinitiativen aufgebaut und vom Ministerium für Wissenschaft, Forschung und Kultur (MWFK) gefördert. Zu Beginn des Projekts wurde eine Arbeitsgruppe mit Vertreter*innen der operativen Arbeitsebene der acht Hochschulen und außeruniversitären Forschungseinrichtungen sowie auf der Leitungsebene eine Runde der Vize-Präsident*innen für Forschung (Entwicklung, Transfer etc.) etabliert.

Parallel wurden zwei Analysen erarbeitet, um die dringenden thematischen Schwerpunkte für die Landesinitiative FDM zu identifizieren und um, in Abhängigkeit der jeweiligen Entwicklungsstände an den Hochschulen sowie finanziellen Rahmenbedingungen, kooperative Strukturen herauszuarbeiten.

3.1 Umfeldanalyse

In der Umfeldanalyse wurden die aktuellen und sehr dynamischen Entwicklungen im Bereich FDM auf Bundesland-, nationaler (z.B. NFDI⁵, RfII⁶) sowie internationaler Ebene (z.B. EOSC⁷, RDA⁸) betrachtet und analysiert

(HeFDI), Landesinitiative für Forschungsdatenmanagement (fdm.nrw), Kompetenznetzwerk Forschungsdatenmanagement der Thüringer Hochschulen (TKFDM)

5 <https://www.nfdi.de/>

6 <http://www.rfii.de>

7 <http://www.eosc.de/>

(Wuttke et al., 2021). Der vertikal-horizontale Ansatz für diese Analyse ist in Abbildung 1 dargestellt. Der richtige Umgang mit Forschungsdaten wird seit über zehn Jahren durch (inter)nationale förderpolitische und institutionelle Richtlinien und Handlungsempfehlungen vorangetrieben. Forschungsdaten selbst als wissenschaftliches Gut, z.B. als Teil einer Publikation, gewinnen zunehmend an Bedeutung. Diese Entwicklung ist auch durch die zunehmende Digitalisierung von Forschungsprozessen in den Fachdisziplinen begründet. Das Teilen und Nachnutzen von Forschungsdaten bleibt jedoch weiterhin eher die Ausnahme.

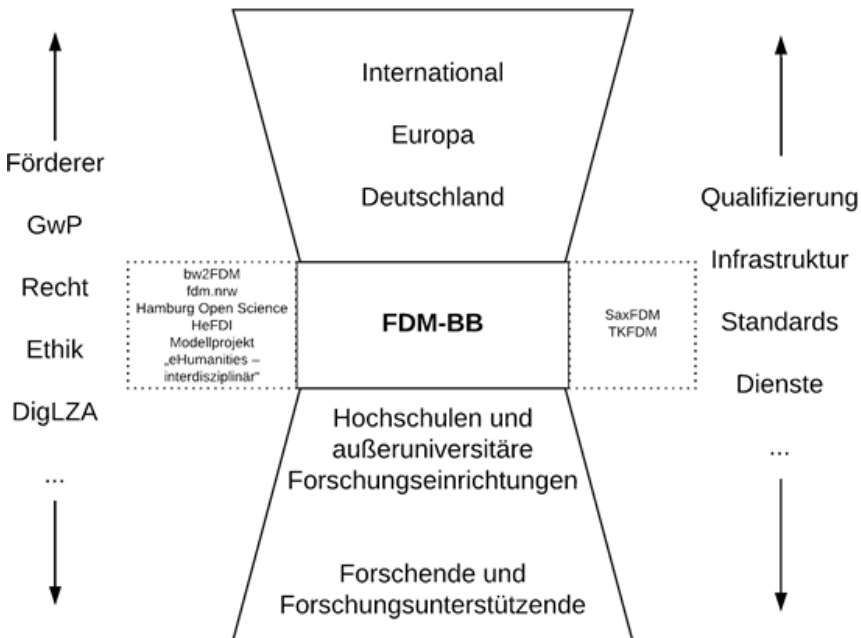


Abb. 1 Vertikal-horizontaler Ansatz der Umfeldanalyse zur inhaltlichen Verortung (GwP – Gute wissenschaftliche Praxis, DigLZA – Digitale Langzeitarchivierung; Bildnachweis: FDM-BB, 2020, CC BY 4.0).

Durch die Umfeldanalyse wurden folgende Aktivitätsfelder für Brandenburg erarbeitet:

- Bewusstseinsbildung auf Leitungs- und operativer Ebene,
- Aufbau eines bundeslandweiten FDM-Netzwerks,

- gemeinsame Entwicklung einer bundeslandweiten FDM-Strategie,
- Entwicklung von institutionellen FDM-Policies,
- Unterstützung der rechtlich verbindlichen Umsetzung des DFG-Kodex,
- Unterstützung bei der Institutionalisierung von FDM in den Hochschulen,
- Abschätzung von Kosten z. B. für kooperativ genutzte FDM-Dienste,
- Unterstützung und Berücksichtigung der unterschiedlichen Geschwindigkeiten und FDM-„Reifegrade“ insbesondere kleinerer und regional abgelegener Hochschulen.

3.2 Anforderungserhebung

Um den Status quo an den acht Hochschulen und deren Ziele in den Bereichen Relevanz von FDM und Institutionalisierung an der Hochschule, Technologien und Infrastrukturen, Personal und Qualifizierung und Organisation, Kooperation und Finanzierung zu identifizieren, wurde eine Organisationsuntersuchung durchgeführt (Radtke et al., 2020). Ein Auftakt-Workshop im Februar 2020 mit FDM-Expert*innen war die Grundlage für den konzeptionellen und inhaltlichen Zugang. Zur Datenerhebung dienten 16 leitfadengestützte Interviews auf Arbeits- und Leitungsebene sowie die Auswertung von Primärdokumenten (z. B. FDM-Policies, Organigramme).

Weil die Relevanz des Themas FDM einem dynamischen Änderungsprozess unterliegt, konnten nur Momentaufnahmen erzielt und bei fehlender Abstimmung zwischen Arbeits- und Leitungsebene innerhalb einer Einrichtung nur Einzelmeinungen wiedergegeben werden. Da das Ziel der Anforderungserhebung in der Durchführung praxisrelevanter Analysen bestand, wurde eine induktive Herangehensweise gewählt. Dazu wurden zentrale Handlungsfelder identifiziert, für eine differenzierte Analyse operationalisiert und an den jeweiligen praktischen Bedarfen ausgerichtet.

Die Untersuchung zeigte sehr unterschiedliche Grade der Institutionalisierung von FDM und damit große Divergenzen bezüglich Relevanz und Entwicklungsstand an den staatlichen Hochschulen Brandenburgs. Es gibt sowohl wenige Hochschulen, die auf langjährige FDM-Erfahrung und die Entwicklung entsprechender Dienste, Policies etc. zurückblicken können, als auch Hochschulen, die erst am Anfang stehen. Daraus ergeben sich unterschiedliche Präferenzen für die Weiterentwicklung von FDM-Diensten vor Ort und kooperativer Landesdienste.

3.3 Handlungs- und Implementierungsempfehlungen

Im nächsten Schritt wurden leitfadengestützte Interviews mit Koordinator*innen anderer FDM-Bundeslandinitiativen durchgeführt, um ein Lernen aus deren Erfahrungen (*lessons learned*) zu ermöglichen. Ferner wurden in sogenannten *Fact Sheets* FDM-Rahmendaten der Hochschulen erhoben (z. B. Anzahl unbefristeter FDM-Expert*innen, Anzahl sowie Art der Forschungsprojekte, Forschungsschwerpunkte etc.). Im Rahmen einer sogenannten *FDM-Palette* wurden darüber hinaus der Status quo, die geplanten FDM-Aktivitäten sowie deren Priorisierung entlang von 28 Handlungsfeldern in den Kategorien *Eigenqualifikation und Sensibilisierung hochschulweit, Koordination und Strukturbildung, Information, Beratung und Schulung, IT-Dienste für FDM und Forschung* und *Einbindung Verwaltungsprozesse, -strukturen* erhoben. Die FDM-Verantwortlichen der einzelnen Einrichtungen sind nicht immer klar festgelegt, sodass die Einordnung der Relevanz zum Teil auf Einzelmeinungen beruht. Zudem wurde die Relevanz der Kategorien, je nach Perspektive und Entwicklungsstand der Hochschule, unterschiedlich bewertet. Insgesamt konnten aber innerhalb der 28 Handlungsfelder deutliche Priorisierungen identifiziert werden, die als gute Diskussionsgrundlage für eine landesweite FDM-Strategie dienen können.

Diese Analysen bilden zusammen die Basis für Handlungs- und Implementierungsempfehlungen für den nachhaltigen und perspektivisch institutionell zu verankernden Aufbau und Betrieb von Forschungsdatenmanagement in Brandenburg (Neuroth et al., 2021), welche in einer zweiten Förderphase des Projekts im Jahr 2021 implementiert werden sollen. Das Ziel ist es, die Institutionalisierung von FDM an den brandenburgischen Hochschulen unter Einbeziehung nationaler und internationaler Entwicklungen zu unterstützen und die brandenburgischen Forschenden und Forschungsunterstützenden zu befähigen, FDM dauerhaft in Forschungsprozesse zu integrieren.

Acknowledgements

FDM-BB dankt dem Ministerium für Wissenschaft, Forschung und Kultur (MWFK) des Landes Brandenburg für die Förderung. Besonderer Dank gilt den studentischen Hilfskräften Mirjam Zeunert und Carsten Schneemann (FH Potsdam), die das Projekt von Beginn an eng begleitet und jederzeit unterstützt haben, sowie der vorherigen wissenschaftlichen Mitarbeiterin Laura Rothfritz (jetzt HU Berlin).

Quellen

- DFG – Deutsche Forschungsgemeinschaft (2019): Leitlinien zur Sicherung guter wissenschaftlicher Praxis: Kodex. https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf
- Neuroth, H.; Straka, J.; Zeunert, M.; Schneemann, C.; Hartmann, N.; Radtke, I. (2021): *Forschungsdatenmanagement in Brandenburg (FDM-BB) – Technologien, Kompetenzen, Rahmenbedingungen: Handlungs- und Implementierungsempfehlungen*. Potsdam: Universitätsverlag. Im Erscheinen.
- Radtke, I.; Hartmann, N.; Neuroth, H.; Rothfritz, L.; Wuttke, U.; Straka, J.; Zeunert, M.; Schneemann, C. (2020): *Forschungsdatenmanagement in Brandenburg (FDM-BB) – Technologien, Kompetenzen, Rahmenbedingungen: Anforderungserhebung bei den brandenburgischen Hochschulen*. Potsdam: Universitätsverlag. <https://doi.org/10.25932/publishup-48091>
- Wilkinson, M. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, Art. No. 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wuttke, U.; Neuroth, H.; Rothfritz, L.; Straka, J.; Zeunert, M.; Schneemann, C.; Hartmann, N.; Radtke, I. (2021): *Forschungsdatenmanagement in Brandenburg (FDM-BB) – Technologien, Kompetenzen, Rahmenbedingungen: Umfeldanalyse zum Aufbau einer neuen Datenkultur in Brandenburg*. Potsdam: Universitätsverlag. <https://doi.org/10.25932/publishup-48090>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 405–411. DOI: doi.org/10.5283/epub.44957.

Panel

The Expanded Potential of Linking Different Types of Information and Data in the Digital Humanities

How Do Innovative Approaches Derived from Information Science and Data Science Contribute to a Knowledge Gain Across Disciplines?

Vera Husfeldt, Sharon Alt

University of Applied Sciences of the Grisons, Switzerland
{[vera.husfeldt](mailto:vera.husfeldt@fhgr.ch), [sharon.alt](mailto:sharon.alt@fhgr.ch)}@fhgr.ch

Abstract

This panel will provide insights into current initiatives and projects linking information objects and data from different sources to generate knowledge. It will also shed light on gaps and give an outlook on what is possible with innovative approaches such as artificial intelligence to gain new findings from disparate data sources.

Keywords: information science; data science; data analysis; machine learning; artificial intelligence; data linkage; education research; data regulations

Introduction

Information science has a long tradition in the collection, selection, organization, processing, management and dissemination of data, information and knowledge (Borko, 1968; Saracevic, 2009). What often falls short is the linkage of different types of information as well as combining data from different sources to generate knowledge. However, where information and data remain isolated without a link, their potential can only be used to a limited extent (Husfeldt et al., 2020). Those data collections would become even more valuable and informative when individual characteristics can be linked (FORS,

2020). Linked data from different sources increases accuracy, helps save on costs, and reduces the burden on respondents.

So, the advantages are evident. However, there are many challenges to be overcome: Comprehensive metadata and documentation do often not exist, access to data is complicated or denied, or the use of data and the linking of it can only be done to a very limited extent, just to name a few (*ibid.*).

Which innovative approaches derived from information science and data science enable or improve the linking process? Which legal, political and institutional aspects and frameworks should be considered to improve access and data protection? What can we learn from other fields? Which processes and initiatives are under way to foster data linkage?

In the panel recognized practitioners and researchers will discuss the potential of linking existing data collections for scientific analysis. They will also talk about the opportunities, limits and risks of frameworks and approaches such as artificial intelligence for bringing relevant data sets together using the example of educational data collections.

Format

The event will take place in a 90 minute panel session slot. The program will start with an introduction to the potential of data linkage approaches for a knowledge gain in the digital humanities. Following this brief overview each panelist will give a three minute presentation about their data linkage projects. The second part will be a question and answer session, moderated by Prof. Dr. Vera Husfeldt.

Benefits

The panel will assemble data linkage efforts from the information science and digital humanities perspective. It also provides a good venue to introduce innovative artificial intelligence approaches to the audience.

Panel Chair

Prof. Dr. Vera Husfeldt, Education Informatics, University of Applied Sciences of the Grisons

Participants

Prof. Dr. Christof Wolf, Präsident von GESIS-Leibniz-Institut für Sozialwissenschaften und Inhaber des Lehrstuhls für Sozialstrukturanalyse an der sozialwissenschaftlichen Fakultät der Universität Mannheim

Dr. Manfred Antoni, Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, Forschungsdatenzentrum

Dr. Jacques Babel, Leiter „Bildungsperspektiven und Längsschnittanalysen“ Eidgenössisches Departement des Innern EDI, Bundesamt für Statistik BFS

Prof. Dr. Georg Lutz, Director FORS (Swiss Centre of Expertise in the Social Sciences), professor of political science at the University of Lausanne

Prof. Dr. Heiko Rölke (DAViS), Leiter Kompetenzzentrum für Datenanalyse, Visualisierung und Simulation, Professor für Data Science, Fachhochschule Graubünden

Andreas Klausling, Mitglied der Geschäftsleitung educa.ch, Fachagentur für ICT und Bildung, Entwicklung einer Datennutzungspolitik für den Bildungsraum Schweiz

Prof. Dr. Ben Jann, Professur für Sozialstrukturanalyse, Universität Bern

Dr. Marieke Heers, FORS (Swiss Centre of Expertise in the Social Sciences)

Dr. Andrea Diem, SKBF (Schweizerische Koordinationsstelle für Bildungsforschung), Longitudinal-Studien

Acknowledgements

The topic of the panel relates to the ongoing research project “Virtual Educational Observatory”. It is supported by the Swiss National Science Foundation within the framework of the National Research Program 77 “Digital Transformation”. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the Swiss National Science Foundation. Details of the project are available via: <https://www.fhgr.ch/fhgr/angewandte-zukunftstechnologien/schweizerisches-institut-fuer-informationswissenschaft-sii/projekte/das-erweiterte-potenzial-von-bildungsdaten/#c12947>.

References

- Borko, Harold (1968). Information science. What is it? In: *American Documentation*, 19(3).
- FORS (2020). Accessing and linking data for research in Switzerland. <https://link-hub.ch/report/>
- Husfeldt, V., Schiller, D., & Alt, S. (2020). Das erweiterte Potenzial von Bildungsdaten. In: *Information – Wissenschaft & Praxis*, 71(5–6), 298–302.
- Saracevic, Tefko (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 2570–2585). New York: Taylor and Francis.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 414–417. DOI: doi.org/10.5283/epub.44958.

Doctoral Colloquium

List of Contributors / Teilnehmerliste

Alexander Schönhals

Provide Computer-Aided Innovation by Intellectual Property Protection
Using Distributed Ledger Technology
Bergische Universität Wuppertal

Corsin Capol

Untersuchung des Nutzer-Anfrageverhaltens in Frage-Antwort-Systemen:
Prototypentwicklung und Evaluation
Universität Hildesheim/Fachhochschule Graubünden

Dennis Trautwein

Towards a Decentralized Plagiarism Detection System
Bergische Universität Wuppertal

Stephanie van de Sandt

The Tracking of Research Data (Re-)Use in Scholarly Works
CERN

Timo Spinde

Detecting and Visualizing Polarizing Language in News Articles
Bergische Universität Wuppertal

Verena Weimer

Scientometric-based Monitoring Instruments in Educational Research –
Conditions for Success, Possibilities and Limits
German Institute for International Educational Research (DIPF)

Chair: Prof. Dr. Bela Gipp, Bergische Universität Wuppertal

Co-Chair: Dr. Terry Ruas, Bergische Universität Wuppertal

Co-Chair: Dr. Moritz Schubotzc, FIZ Karlsruhe

**Gerhard Lustig Award
Papers**

Detection and Identification of Fake News

Binary Content Classification with Pre-trained Language Models

Mina Schütz

Darmstadt University of Applied Sciences

Haardtring 100, 64295 Darmstadt, Germany and

Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna, Austria

ORCID-ID: [0000-0002-1102-306X](https://orcid.org/0000-0002-1102-306X), mina.schuetz@ait.ac.at

Abstract

Fake news has emerged as a critical problem for society and professional journalism. Many individuals consume their news via online media, such as social networks and news websites. Therefore, the demand for automatic fake news detection is increasing. There is still no agreed upon definition for fake news, since it can include various concepts, such as clickbait, propaganda, satire, hoaxes, and rumors. This results in a broad landscape of machine learning approaches, which have a varying accuracy in detecting fake news. This masterthesis focused on a binary content-based classification approach, with a bidirectional Transformer (BERT), to detect fake news in online articles. BERT creates a pre-trained language model during training and is fine-tuned on a labeled dataset. The FakeNewsNet dataset is used to test two variants of the model (cased/uncased) with articles, using only the body text, the title, and a concatenation of both. Additionally, both models were tested with different preprocessing steps. The models gain in all 29 carried out experiments high accuracy results, without overfitting. Using the body text and the concatenation resulted in five models with an accuracy of 87% after testing, whereas using only titles resulted in 84%. This shows that short statements could be already enough for fake news detection using language models. Also, the preprocessing steps seem to have no major impact on the predictions. It is concluded that transformer models, such as BERT, are a promising approach to detect fake news, since it achieves notable results, even without using a large dataset.

Keywords: fake news; fake news detection; BERT; transformer; pre-trained language model; binary classification

1 Introduction

Fake news has recently gained a lot of attention in the media and research community. Through the information overload in the Internet and an increased usage of social media for news consumption the propagation of disinformation has risen (Figueira & Oliveira, 2017). Zhou and Zafarani (2018) even state that fake news is “[...] one of the greatest threats to democracy, journalism, and freedom of expression” (p. 1). This is also based on several psychological reasons, such as echo chambers (Rana et al., 2018). Each individual has a social circle, which typically has the same beliefs as them. Those views are propagated with other people, who share their ideology and therefore the facts seem more credible (Shu et al., 2017a). Individuals are more likely to believe facts the more often they have read it (validity effect) and tend to believe facts that confirm their views. A manual method to solve this problem is expert-based fact-checking, which is expensive and slow (Graves, 2018). Another approach is the automatic detection of fake news with machine learning methods and natural language processing (NLP), which could help users to identify a possible deception in a claim (Mahid et al., 2018).

2 State-of-the-art

Since there is no common definition for fake news to-date, we propose a new definition, which includes all concepts (disinformation, misinformation, hoax, propaganda, rumor, clickbait, satire) that could be categorized as fake news, regardless of the intention or factuality of the article: *Fake news is an article which propagates a distorted view of the real world regardless of the intention behind it.* The content can be examined through the extraction of linguistic features in the body text and titles with NLP methods. This can be on character, word, sentence or document level. Usually Bag-of-Words, N-Gram models, and Part-of-Speech Tagging play an important role. However, many researchers used a variant of models, such as SVM, logistic regression, decision trees or neural networks (convolutional, recurrent, LSTM) to compare the significance of features and their impact on predictions. Since 2018 several surveys were published (Rana et al., 2018; Oshikawa et al., 2018; Zhou & Zafarani, 2018; Sharma et al., 2019), which give an overview of the

various papers and methodologies. Besides content-based, there are also knowledge-based, social-context-based, and hybrid approaches, which can include user-information, network-analysis and stance detection (Shu et al., 2017a; Zhou & Zafarani 2018).

The newly invented transformers have shown to outperform many NLP state-of-the-art results and are usually pre-trained on a large corpus and then fine-tuned on a dataset for the specific task, which makes them very usable for fake news detection, since available corpora are often small. During pre-training transformers create language models, which can be used for sentence- or token-level tasks (Devlin et al., 2019).

BERT (Bidirectional Encoder Representations from Transformers) is one of the latest innovations in machine learning techniques for NLP and was developed by Google in 2019 (ibid.). It is based on the original transformer structure by Vaswani et al. (2017). This transformer “[...] is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers” (ibid., p. 1). Additionally, BERT is the first bidirectional transformer, which means that the model can read a sentence in both directions at the same time, unlike unidirectional transformers (e.g., left-to-right, right-to-left) (Devlin et al., 2019).

3 Main contribution

Text-based fake news detection is still an open research challenge, because the currently proposed solutions are not yet satisfying. These studies use various datasets with no common ground truth and are stand-alone or a mix of the different approaches. Their results are varying to a high degree and often focus on multiple fake news concepts. The purpose of this work is to set out the current state-of-the-art methods in fake news detection and use a pre-trained language model for a binary content classification to gain further knowledge about the textual representation of fake news, especially the usage of a Transformer model for long body text and short titles. Furthermore, the main research question of this work is:

To which extent is a pre-trained language model useful for content-based fake news detection and does it gain promising results in predicting the classification of body texts and titles of news articles?

Therefore, the practical part of the conducted thesis proposes an approach for fake news detection with BERT and the FakeNewsNet (Shu et al., 2017a, 2017b, 2018) dataset. This leads to the following sub-questions:

- How do the results change, if the body text or the title of a news article are used separately, or both are used together?
- Do the results change based on different preprocessing steps, as removing non-ASCII characters, transcripts or short/long articles?
- Does it make a difference, in respect to the preprocessed files, if the cased or uncased version of BERT is used?

4 Related work

Several studies focused on fake news and Transformers but with different tasks: sentence comparison/fact classification (Yang et al., 2019; Mao & Liu, 2019), propaganda detection (Gupta et al., 2019), relationship of body text/title (Jwa et al., 2019), stance detection (Slovikovskaya, 2019), evidence retrieval (Soleimani et al., 2019). Directly comparable are the works of Liu et al. (2019) and Rodríguez and Iglesias (2019). For a multi-classification for short statements and extra metadata, Liu et al. (2019) gained a result of 40.58% accuracy. However, Rodríguez and Iglesias (2019) also did a binary classification of fake news using three neural networks. They had an accuracy of 98% and used a similar approach as this work, but tested BERT only on one experiment and combined titles and text as one input but did not differentiate between using body text or title and the cased/uncased version.

5 Methodology

In this work we used the BERT implementation by HuggingFace¹, which is based on PyTorch. The FakeNewsNet contains more than 20,000 already labeled (fake/real) articles. It is notable, that the datasets contained more real than fake articles, depending on the pre-processing steps. The content was

1 <https://github.com/huggingface/transformers>

pre-processed in various ways, which lead to a total of 13 different files (Table 1) with the proposed pre-processing steps and size: deleting articles based on length (removing outliers), deleting transcripts (spoken language in mostly real articles), deleting HTML-tag [edit] (mostly fake articles), deleting non-ASCII symbols and digits. Those were carried out to prevent the model from learning false features based on the classes.

Table 1: Pre-processed dataset files

File no.	Type	Length	Transcript	Edit	ASCII/Digits	Dataset size
1	Text	Yes	Yes	Yes	Yes	11,944
2	Text	Yes	Yes	Yes	No	11,944
3	Text	Yes	Yes	No	No	11,944
4	Text	Yes	No	No	No	12,172
5	Text	No	No	No	No	21,041
6	Title	No	No	No	No	21,041
7	Title	Yes	No	No	No	12,172
8	Title	Yes	No	No	Yes	12,172
9	Both	No	No	No	No	21,041
10	Both	Yes	No	No	No	15,355
11	Both	Yes	Yes	No	No	15,103
12	Both	Yes	Yes	Yes	No	15,103
13	Both	Yes	Yes	Yes	Yes	15,103

Those files were tested in a total of 29 experiments. The hyper parameters for the model were tuned with the first file (completely pre-processed, only body text). The best hyper parameters were then used for the following experiments (text, title, both) with the BERT-base-cased/uncased versions (Fig. 1).

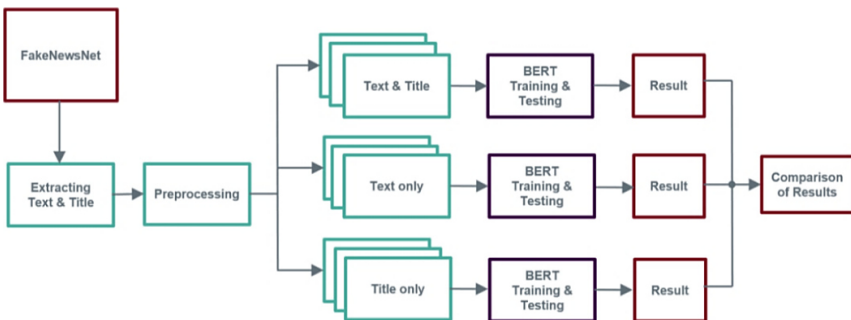


Fig. 1 Methodology overview

Afterward the most promising combinations were tested with the BERT-base-uncased model (Fig. 2). All experiments were evaluated with precision,

recall, accuracy and F1. For each experiment the loss after validation was calculated and compared to the various preprocessed files and both models.

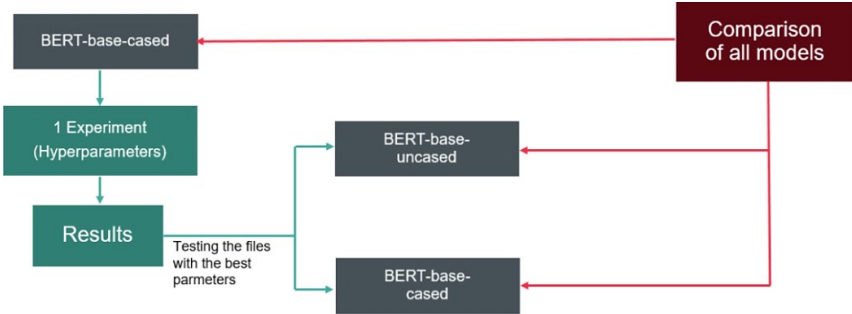


Fig. 2 Workflow for both BERT models

6 Results

The results in Table 2 show that, with the right parameters, both models gain very good detection results. Almost all experiments had results over 80% accuracy in the validation and testing set without overfitting the data. As shown, the learning rate has the highest influence on the results in comparison to other hyper parameters (see experiment 1). Furthermore, even if there is only a small difference in the evaluation metrics, the datasets with a combination of titles and text have the best results (files 9–13). The datasets with the most promising metrics were additionally carried out with the BERT-base-uncased-version (last block). Therefore, we conclude that lowercasing the text has no influence with the used data. Also, in all experiments it is shown that the pre-processing steps have no major impact on the outcome. Although for titles, a larger dataset might be helpful, since file 6 gained the best results with twice the size as other pre-processed data.

Table 2: Results

Model	File	Epochs	Batch	LR	MSL	W-U	ACE	VL	ACT	F1	P	R
B-B-C	1	5	6	5e-5	512	100	0.75	0.60	0.72	0.48	0.43	0.57
B-B-C	1	3	6	2e-5	512	100	0.86	0.30	0.86	0.79	0.81	0.80
B-B-C	1	3	6	1e-5	512	100	0.87	0.30	0.87	0.80	0.82	0.82
B-B-C	1	3	16	2e-5	256	100	0.86	0.23	0.86	0.81	0.83	0.82
B-B-C	1	5	16	5e-5	256	100	0.84	0.07	0.85	0.82	0.81	0.80
B-B-C	1	5	16	2e-5	256	100	0.85	0.09	0.85	0.80	0.83	0.81
B-B-C	1	15	16	2e-5	256	100	0.85	0.01	0.86	0.81	0.84	0.81
B-B-C	1	15	6	2e-5	512	100	0.86	0.02	0.87	0.79	0.81	0.81
B-B-C	1	10	6	2e-5	512	100	0.85	0.02	0.85	0.77	0.79	0.79
B-B-C	1	5	16	2e-5	256	0	0.86	0.09	0.85	0.81	0.83	0.81
B-B-C	1	10	16	2e-5	256	0	0.86	0.02	0.86	0.81	0.84	0.81
B-B-C	1	100	16	2e-5	256	100	0.85	0.01	0.86	0.80	0.84	0.80
B-B-C	2	10	16	2e-5	256	0	0.84	0.02	0.85	0.80	0.81	0.81
B-B-C	3	10	16	2e-5	256	0	0.85	0.02	0.86	0.80	0.82	0.81
B-B-C	4	10	16	2e-5	256	0	0.84	0.02	0.86	0.81	0.83	0.82
B-B-C	5	10	16	2e-5	256	0	0.87	0.08	0.86	0.78	0.82	0.78
B-B-C	6	10	16	2e-5	256	0	0.85	0.08	0.84	0.75	0.78	0.76
B-B-C	6	5	32	3e-5	128	0	0.84	0.11	0.84	0.76	0.79	0.76
B-B-C	7	5	32	2e-5	128	0	0.84	0.11	0.84	0.80	0.80	0.80
B-B-C	8	5	32	2e-5	128	0	0.83	0.10	0.83	0.79	0.80	0.79
B-B-C	9	10	16	2e-5	256	0	0.87	0.07	0.87	0.78	0.82	0.79
B-B-C	10	10	16	2e-5	256	0	0.87	0.08	0.86	0.78	0.82	0.78
B-B-C	11	10	16	2e-5	256	0	0.87	0.09	0.86	0.77	0.83	0.77
B-B-C	12	10	6	2e-5	512	0	0.87	0.10	0.87	0.78	0.80	0.80
B-B-C	13	10	16	2e-5	256	0	0.87	0.08	0.86	0.78	0.86	0.83
B-B-U	1	10	6	2e-5	512	0	0.86	0.02	0.86	0.82	0.84	0.83
B-B-U	9	10	16	2e-5	256	0	0.87	0.07	0.87	0.79	0.84	0.80
B-B-U	6	10	32	2e-5	128	0	0.85	0.08	0.85	0.77	0.85	0.80
B-B-U	6	30	32	2e-5	128	50	0.85	0.06	0.85	0.75	0.85	0.79

(B-B-C: BERT-base-cased, B-B-U: BERT-base-uncased, LR: Learning Rate, MSL: Maximum Sequence Length, W-U: Warm Up, ACE: Accuracy Evaluation, VL = Validation Loss, ACT = Accuracy Testing, R = Recall, P = Precision)

7 Conclusion

The results of this work show that a content-based approach can gain promising results for detecting fake news, even without setting hand-engineered features. We conclude that BERT can be used for short statements as well as complete articles. Out of 29 experiments, five gained an accuracy of 87% on the test set. For only testing the titles of the articles the best accuracy was

84%. This shows that there is only a minimal difference in accuracy using the body text or the title (or both) with pre-trained language models. In future work, especially social-context-based features should be looked upon to gain better results. Another important part is to explore methods of explainable artificial intelligence, to help understanding the difference in fake news concepts, which could help to gain further knowledge about linguistic features. Lastly, this approach should be tested with another dataset to evaluate the cross-domain prediction results.

Acknowledgements

This thesis was conducted at the Darmstadt University of Applied Sciences at the Faculty of Media within the course of studies in Information Science in collaboration with the Austrian Institute of Technology GmbH at the Center for Digital Safety and Security under the supervision of Kawa Nazemi (h_da), Melanie Siegel (h_da), and Alexander Schindler (AIT). Additionally, this work was supported by the Research in Information Science (<https://sis.h-da.de/>) and the Research Group on Human-Computer Interaction and Visual Analytics (<https://vis.h-da.de>).

Bibliography

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171–4186). Stroudsburg, PA: Association for Computational Linguistics.
- Figueira, A., & Oliveira, L. (2017). The Current State of Fake News: Challenges and Opportunities. *Procedia Computer Science*, 121, 817–825.
- Graves, L. (2018). Understanding the promise and limits of automated fact-checking. <https://reutersinstitute.politics.ox.ac.uk/our-research/understanding-promise-and-limits-automated-fact-checking>
- Gupta, P., Saxena, K., Yaseen, U., Runkler, T., & Schütze, H. (2019). Neural Architectures for Fine-Grained Propaganda Detection in News. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda* (pp. 92–97). Association for Computational Linguistics.

- Jwa, H., Oh, D., Park, K., Kang, J. M., & Lim, H. (2019). exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT). *Applied Sciences* 2019, 9(4062).
- Khan, S. A., Alkawaz, M. H., & Zangana, H. M. (2019). The use and abuse of social media for spreading fake news. *IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)* (pp. 145–148). Cham: Springer Nature.
- Liu, C., Wu, X., Yu, M., Li, G., Jiang, J., Huang, W., & Lu, X. (2019). A Two-Stage Model Based on BERT for Short Fake News Detection. In Douligieris, C., Karagiannis, D., & Apostolou, D. (Eds.), *Knowledge Science, Engineering and Management* (pp. 172–183). Cham: Springer International Publishing.
- Mahid, Z. I., Manickam, S., & Karuppayah, S. (2018). Fake news on social media: Brief review on detection techniques. In *2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA)*. [Piscataway, NJ]: IEEE.
- Mao, J., & Liu, W. (2019). Factuality Classification Using the Pre-Trained Language Representation Model BERT. In *IberLEF@SEPLN*, pp. 126–131.
- Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A Survey on Natural Language Processing for Fake News Detection. [arXiv:1811.00770](https://arxiv.org/abs/1811.00770) (version 1, 2018).
- Rana, D. P., Agarwal, I., & More, A. (2018). A Review of Techniques to Combat the Peril of Fake News. *4th International Conference on Computing Communication and Automation (ICCCA)*. [Piscataway, NJ]: IEEE.
- Rodríguez, À. I., & Iglesias, L. L. (2019). Fake News Detection Using Deep Learning. [arXiv:1910.03496](https://arxiv.org/abs/1910.03496).
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y. (2019). Combating Fake News: A Survey on Identification and Mitigation Techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 37(4), III:0–III:41.
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. *CoRR*, abs/1809.01286.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017a). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22–36.
- Shu, K., Wang, S., Liu, H. (2017b). Exploiting Tri-Relationship for Fake News Detection. [arXiv:1712.07709](https://arxiv.org/abs/1712.07709)
- Slovikovskaya, V. (2019). Transfer Learning from Transformers to Fake News Challenge Stance Dection (FNC-1) Task. [arXiv:1910.14353](https://arxiv.org/abs/1910.14353)
- Soleimani, A., Monz, C., & Worring, M. (2019). BERT for Evidence Retrieval and Claim Verification. [arXiv:1910.02655](https://arxiv.org/abs/1910.02655)

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. ... Polosukhin, A. (2017). Attention Is All You Need. [arXiv:1796.03762](https://arxiv.org/abs/1796.03762)
- Yang, K.-C., Niven, T., & Kao, H.-Y. (2019). Fake News Detection as Natural Language Interference. [arXiv:1907.07347](https://arxiv.org/abs/1907.07347)
- Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research, Detection Methods and Opportunities. *ACM Computing Surveys*, 1(1), Art. No. 109.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 422–431. DOI: doi.org/10.5283/epub.44959.

Design and Development of an Emoji Sentiment Lexicon

Fabian Haak

Institute for Information Management, Technische Hochschule Köln
Claudiusstraße 1, 50678 Cologne, Germany

ORCID <https://orcid.org/0000-0002-5982-7104>, fabian.haak@th-koeln.de

Abstract

Emojis represent an essential means of expressing sentiments such as opinions and attitudes in computer-mediated communication, especially in chats and social media. To effectively capture these sentiments, the sentiments associated with the emojis used must be known. Previous approaches to determining the sentiments expressed with emojis require a large amount of manual annotation. For many emojis, especially less frequently used platform-specific emojis, studies on expressed sentiments do not yet exist. Therefore, these emojis cannot be considered in sentiment analyses so far. In this work, a method for effective and efficient determination of emojis' sentiments and their compilation in a sentiment lexicon was developed. The determined sentiments are compiled as a sentiment lexicon. For this purpose, software was created in Python to process collections of texts into a corpus. The software derives the emojis' sentiments as valence values based on the sentiments of the texts in which the emojis appear. The lexicons produced by the method can be used in lexicon-based sentiment analysis approaches. The method also derives other information on the emojis and their usage that can be used to assess the sentiment lexicon produced and the usage of the emojis. Using the developed method, two analyses were conducted with corpora of different text sources. The results and subsequent comparisons with existing sentiment lexicons have shown that the developed method is able to efficiently produce similar results as sentiment lexicons produced with manual annotation.

Keywords: sentiment analysis; emojis; computer-mediated communication; natural language processing

1 Introduction

Sentiment analysis is an essential tool for natural language processing. *Sentiments* are opinions, feelings, or tonality expressed in texts and can be described in various ways. Sentiment analysis describes the process of extracting these sentiments from texts. Usually, sentiment is derived by sentiment words recorded in lists, so-called *sentiment lexicons*, or by machine learning processes. The capture of sentiments as a measure of polarity and intensity or subjectivity, so-called *valence-based* sentiment analysis, is a fundamental method for analyzing texts. Especially in social media and chats, many opinions are expressed. These are valuable sources of information for social and information science and other academic disciplines. However, the short length of texts from these sources is a challenge for performing practical sentiment analysis. Conventional approaches to capture valence-based sentiments, both lexicon-based and machine learning, have had problems effectively determining the valence of short texts.

An important feature of text communication in social media and chats is the heavy use of emojis. In these informal text forms, they take over many functions of nonverbal communication. Thus, they are also an essential means of expressing sentiments in computer-mediated communication. Opinions and tonality are often conveyed exclusively through emojis. In order to capture these sentiments effectively, the sentiments associated with the emojis used must be known. Previous approaches to determine the sentiments expressed with emojis require a great deal of manual annotation. For many emojis, especially platform-specific emojis, no studies on expressed sentiments exist. Besides, domain-specific knowledge is required to determine the emojis' expressed sentiments. Therefore, these emojis cannot yet be included in sentiment analysis, and determining their valences would be very labor-intensive, as demonstrated by the development of the emoji sentiment lexicon by Kobs et al. (2020). Furthermore, new Unicode emojis are introduced every year. Emoji sentiment lexicons, such as the Emoji Sentiment Ranking (ESR) by Kralj Novak et al. (2015), lose their relevance for sentiment analysis over time and would ideally be updated regularly. Due to these challenging factors, emojis are rarely considered in lexicon-based sentiment analysis. To overcome them, a solution is presented that aims to

1. minimize the need for manual annotation for emoji sentiment determination and

2. improve the results of lexicon-based sentiment analysis, especially of emoji-rich, domain-specific texts such as chats.

2 Data and method for emoji sentiment lexicon generation

In this master thesis on Market and Media Research at TH Köln, a novel method for the effective and efficient automatic determination of emojis' sentiments is introduced. It substantially facilitates the determination of emoji sentiment for the application in lexicon-based sentiment analysis. A software implementation realized in Python derives the emojis' sentiments as valence values based on the sentiments of texts in which the emojis occur. This methodology utilizes emojis' property that they are usually utilized to reinforce sentiment expressed in messages. As Hu et al. (2017) stated, "[...] expressing sentiment, strengthening expression, and adjusting tone are the top three most popular intentions of using emojis". Therefore, we assume that their occurrence in sentimental messages can identify emojis used for expressing sentiment. Furthermore, we assume that these emojis feature the same sentiment as the texts containing them. Previous research, such as by Hu et al., has shown that emojis are used to express irony and sarcasm by adding emojis with valences opposite to the message's. Therefore, a second assumption is that a bimodal distribution of the valences of texts, in which an emoji appears, with local maxima in both polarities, is an indicator for emojis that express irony and sarcasm.

For the implementation of the developed method, two text collections were compiled. The first corpus consists of 2.86 million messages taken from chat logs of the video streaming platform Twitch.tv (2009). Characteristic for chat communication on Twitch is the frequent use of platform-specific emojis, called *Twitch emotes*. Hence, the Twitch corpus ought to demonstrate the developed method's ability to determine platform-specific emoji valences. Due to the large number and their ambiguous nature, the determination of sentiments expressed by Twitch emotes would be very complex using conventional methods. To better evaluate the effectiveness of the developed method and the quality of the assigned valences, a second corpus consisting of 1.48 million publicly accessible Twitter messages (2018) was generated.

This way, valences of Unicode emojis are derived, which can be compared with existing emoji sentiment lexicons valences.

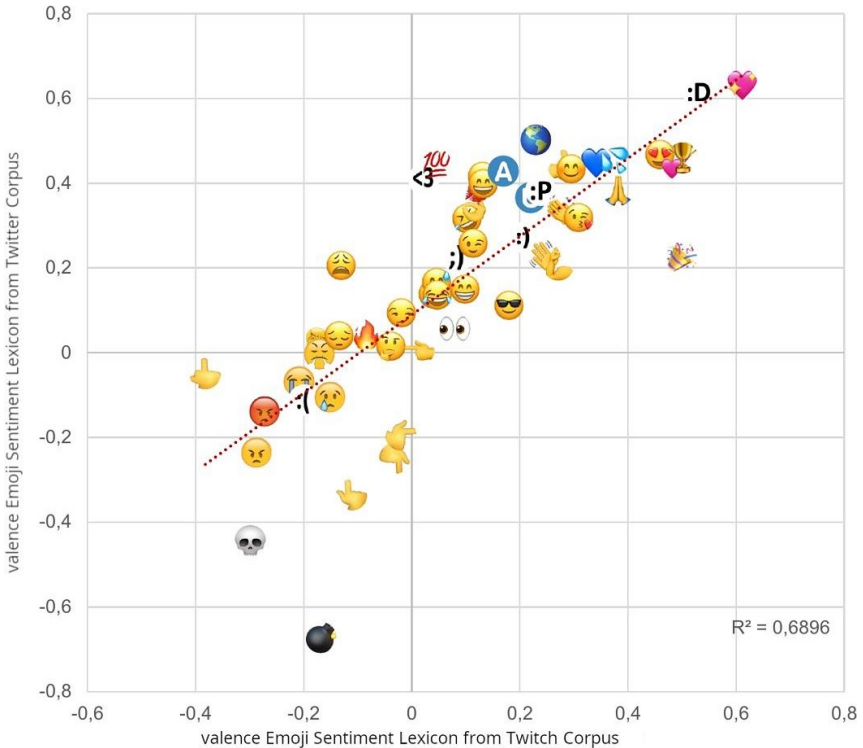


Fig. 1 Comparison of determined emoji valences of emojis, for which both produced lexicons contain valence scores

3 Results

The valences of the emojis are derived by the valences from the texts containing the respective emojis. These are analyzed using the VADER method developed by Hutto and Gilbert (2014), which currently represents the best lexicon-based approach for determining the valences of short informal texts. The VADER method has been modified so that all emojis whose valence must be determined have first been removed from the VADER sentiment

lexicon. This is to prevent that the valences of these emojis are derived from the valences recorded for them in the VADER dictionary. For each emoji, the valence values of all texts containing the emoji are analyzed, and the standard deviation and arithmetic mean of these values calculated. All emojis whose standard deviations are below a certain threshold (0.625) and for which there are enough occurrences (more than ten) are considered valid. These emojis and their corresponding valence values (mean valence of texts they appear in) are exported as a sentiment lexicon. The produced lexicon can easily be inserted into lexicon-based sentiment analysis models. By doing so, the model is extended by information on platform-specific emoji valences, thus increasing the effectiveness of sentiment analysis of texts containing these emojis. The developed method also determines additional information such as TF-IDF values and n-grams about the emojis. Analyses of the two corpora (Twitch, Twitter) were performed, producing two emoji sentiment lexicons. In both analyses, close to 200 valid emoji valences were assigned (Twitter corpus: 198, Twitch corpus: 185, 107 of which are Twitch emotes). Both produced lexicons are publicly available at Zenodo (2020), along with a script that incorporates them into the VADER sentiment lexicon to use them in a standard sentiment analysis toolchain.

3 Evaluation

To evaluate the quality of the produced sentiment lexicons and to be able to judge the effectiveness of the developed method for determining emoji valences, the two lexicons produced were first compared with each other. The comparison of the valence values of the mutual emojis contained in both of the two dictionaries (see Fig. 1) and their Spearman's rank and Pearson correlation show that there is a statistically significant correlation between the two dictionaries. This is a strong indication that the determined valences represent the actual emoji valences and that the method is reliable. For a better evaluation of the quality of the generated lexicons and the developed method, the lexicons were compared with the ESR, and a sentiment lexicon for Twitch emotes developed by Kobs et al. (2020), both produced with manual annotation. The emoji sentiment lexicon generated using the Twitter corpus showed a strong positive Spearman's rank correlation ($r = 0.74$, $p < 0.01$) and Pearson correlation ($r = 0.723$, $p < 0.01$) with the ESR. Comparing the

emoji sentiment lexicon generated by analyzing the Twitch corpus with the *emote lexicon* produced by Kobs et al. also shows a strong and highly significant Pearson correlation coefficient ($r = 0.717$, $p < 0.01$) as well as rank correlation according to Spearman ($r = 0.744$, $p < 0.01$).

4 Conclusion and outlook

The results and subsequent comparison with existing sentiment dictionaries have shown that the developed method can efficiently produce results similar to those of sentiment dictionaries produced with manual annotation. Thus, the approach of determining emojis' valences via the texts in which they occur is a promising method to improve the precision of sentiment analysis of short informal texts. A test with the emoji sentiment lexicon produced by the Twitch corpus analysis, integrated into the VADER model, identified twice as many sentiment expressing texts as the stand-alone VADER model.

In his current work in the ESUPOL project as a Ph.D. student at TH Köln, Fabian Haak's research is related to bias detection in search query suggestions and the impact of query suggestions for search queries of political topics on opinion formation. Sentiment analysis of short texts is highly relevant for bias detection, showing the developed method's potential and future use-cases.

References

- Haak, F. (2020). Emoji Sentiment Lexicons [Data set]. <http://doi.org/10.5281/zenodo.4159573>
- Hu, T., Guo, H., Sun, H., Nguyen, T., & Luo, J. (2017). Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (pp. 102–111). Palo Alto, CA: AAAI Press.
- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. <http://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vaderSentiment.py>

- Kim, J. (2019). Twitch.tv Chat Log Data. Harvard Dataverse [Data set]. <https://doi.org/10.7910/DVN/VEoIVQ>
- Kobs, K., Zehe, A., Bernstetter, A., Chibane, J., Pfister, J., Tritscher, J., & Hotho, A. (2020). Emote-Controlled: Obtaining Implicit Viewer Feedback through Emote based Sentiment Analysis on Comments of Popular Twitch.tv Channels. *ACM Transactions on Social Computing*, 3(2), 1–34.
- Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PLOS ONE* 10(12), e0144296. <http://dx.doi.org/10.1371/journal.pone.0144296>
- Kramer, S. (2018). Social Media Bot Detection by Paragon Science [Data Set]. <https://data.world/drstevkramer/social-media-bot-detection-by-paragon-science>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 432–438. DOI: doi.org/10.5283/epub.44960.

Reflektive vs. intuitive Ästhetikwahrnehmung von Homepages

Eine Analyse kultureller Unterschiede zwischen
Testpersonen aus Deutschland und Lateinamerika

Neele Lange

Stiftung Universität Hildesheim,
Universitätsplatz 1, 31141 Hildesheim, Deutschland

neele_lange@web.de

Abstract

Die Masterarbeit befasst sich mit kulturellen Unterschieden zwischen deutschen und lateinamerikanischen Teilnehmenden hinsichtlich der reflektiven und der intuitiven Ästhetikwahrnehmung von Homepages. Ziel war es, neben kulturellen Unterschieden auch allgemeine Divergenzen zwischen den Wahrnehmungsarten intuitiv und reflektiv bezüglich der Ästhetik von Homepages zu untersuchen.

Die Ergebnisse der Arbeit bestätigen dabei die Wichtigkeit einer kulturellen Adaption von Websites, da vor allem in der intuitiven Wahrnehmung kulturelle Unterschiede aufgezeigt werden konnten. Sie zeigen außerdem, dass Websites insbesondere in Aspekten der Kunstfertigkeit optimiert werden sollten, da diese kulturunabhängig am stärksten mit der Ästhetikbewertung korrelierten.

Die Arbeit wurde am Institut für Informationswissenschaft und Sprachtechnologie der Universität Hildesheim im Masterstudiengang *Internationales Informationsmanagement – Informationswissenschaft* verfasst.

Keywords: Mensch-Maschine-Interaktion; intuitive Ästhetikwahrnehmung; reflektive Ästhetikwahrnehmung; Kulturvergleich; Website

1 Einleitung

Die ästhetische Gestaltung hat einen enormen Einfluss auf die gesamte Wahrnehmung einer Website (vgl. Thielsch, 2008, S. 262). So ist die Ästhetik ein verlässliches Mittel, um eine gewisse Zufriedenheit bei den Nutzern zu erreichen (vgl. ebd.). Zudem wird die Ästhetik schneller beurteilt als andere Aspekte einer Website, wie die Benutzerfreundlichkeit oder der Inhalt, und ist somit insbesondere entscheidend, um die Aufmerksamkeit von Usern zu erlangen (vgl. Sutcliffe, 2002, S. 1839 f.).

Durch die globale Vernetzung und länderübergreifende Nutzung von Websites entstehen internationale Anforderungen an Webauftritte (vgl. Meidl, 2013, S. 1). Beim Lokalisierungsprozess sind kulturelle Faktoren – also Wissen, Präferenzen, Eigenheiten und Weiteres, das innerhalb einer Kultur geteilt wird – von enormer Wichtigkeit, um die Akzeptanz und Nutzbarkeit sicherstellen zu können (vgl. Sun, 2001, S. 96).

Diese Begebenheiten sowie Erkenntnisse aus der Forschung innerhalb dieser Thematiken – wie kulturelle Unterschiede bezüglich Farbpräferenzen (vgl. Cyr et al., 2010) oder visueller Komplexität auf Websites (vgl. u. a. Reinecke/Gajos, 2014), dem Einfluss von Ästhetik auf die wahrgenommene Usability (vgl. u. a. Kurosu/Kashimura, 1995) und auf die Glaubwürdigkeit (vgl. u. a. Robins/Holmes, 2008) oder auch der Zusammenhang zwischen der Betrachtungsdauer und Ästhetik (vgl. u. a. Lindgaard et al., 2006) – sind Grundlage sowie Motivation für die Fragestellung der Arbeit.

Die Arbeit behandelt die folgenden Forschungsfragen:

- Welche Aspekte von Homepages sind in der deutschen und in der lateinamerikanischen Kultur am ausschlaggebendsten für eine Bewertung als ‚ästhetisch‘?
 - Gibt es Unterschiede zwischen der deutschen und der lateinamerikanischen Kultur bei den ausschlaggebendsten Aspekten von Homepages für deren Bewertung als ‚ästhetisch‘?
- Bestehen Unterschiede zwischen der intuitiven und der reflektiven Ästhetikwahrnehmung?
 - Gibt es innerhalb der deutschen sowie innerhalb der lateinamerikanischen Kultur Unterschiede zwischen der intuitiven und der reflektiven Ästhetikwahrnehmung?

- Bestehen innerhalb der intuitiven sowie innerhalb der reflektiven Ästhetikwahrnehmung Unterschiede zwischen den deutschen und den lateinamerikanischen Teilnehmenden?

Die Forschungsfragen kombinieren die verschiedenen Komponenten Ästhetik, zwei verschiedene Formen der Wahrnehmung, die Mensch-Maschine-Interaktion (MMI) sowie einen kulturvergleichenden Aspekt.

2 Methode

Zur Datenerhebung wurden zwei unterschiedliche Studien durchgeführt, eine zur Messung der intuitiven und eine zur Erfassung der reflektiven Ästhetikwahrnehmung, mit einem *Between-subject*-Studiendesign, um einen *mere exposure effect*, also eine positivere Wahrnehmung der Homepages beim erneuten Betrachten (vgl. Moreland/Zajonc, 1982, S. 410), zu verhindern.

Im Rahmen einer Laborstudie sollte das intuitive Ästhetikwahrnehmen von Homepages untersucht werden. Dieses wurde durch eine Präsentationsdauer der Stimuli von 50 ms sowie eine sofortige Ästhetikbewertung operationalisiert. Die Vorgehensweise bezüglich der Präsentationsdauer basiert auf den Erkenntnissen von Lindgaard et al., die eine stabile Ästhetikbewertung ab eben dieser 50 ms nachweisen konnten (vgl. Lindgaard et al., 2006, S. 124). Als Stimuli dienten Screenshots von Homepages, da Nutzer in der Regel als erstes auf die Homepage einer Website gelangen und innerhalb der ersten Betrachtung von 50 ms ohnehin keine Interaktion stattfindet. Die Bewertung der Homepages erfolgte durch das Ausfüllen siebenstufiger Likert-Skalen mit den Polen 1 = *überhaupt nicht ästhetisch* und 7 = *sehr ästhetisch*. An der Laborstudie nahmen insgesamt 20 deutsche und 20 lateinamerikanische Testpersonen teil.

Im Rahmen einer Onlinestudie sollte das reflektive Ästhetikwahrnehmen von Homepages untersucht werden. Die reflektive Ästhetikwahrnehmung wurde zum einen durch eine unbegrenzte Präsentationsdauer operationalisiert. Zum anderen diente auch die Detailliertheit der Ästhetikbewertung zur Steuerung und somit Unterscheidung zwischen intuitiver und reflektiver Ästhetikbewertung. Daher wurde für die Bewertung der reflektiven Ästhetikwahrnehmung der VisAWI (Visual Aesthetics of Websites Inventory), ein von Moshagen und Thielsch entwickelter Fragebogen zur Erfassung von

Ästhetik auf Websites, mit insgesamt 18 Aussagen zu den vier zentralen Aspekten der Webästhetik (Einfachheit, Vielfalt, Farbigkeit und Kunstfertigkeit), dessen Zutreffen auf die vorliegende Website jeweils auf siebenstufigen Likert-Skalen beurteilt werden, eingesetzt (vgl. Thielsch/Moshagen, 2011, S. 261). Weiterhin sollte mit der Onlinestudie und dem Einsatz des VisAWI herausgefunden werden, welche gestalterischen Aspekte einer Website ausschlaggebend sind für deren Bewertung als ‚ästhetisch‘. Als Stimuli wurden die gleichen Screenshots von Homepages verwendet wie in der Laborstudie, um die Vergleichbarkeit sicherstellen zu können. Insgesamt konnten Datensätze von 100 deutschen und 100 lateinamerikanischen Teilnehmenden in die Auswertung einbezogen werden.

Zusätzlich wurde in beiden Studien eine *Card sorting*-Aufgabe durchgeführt, um herauszufinden, welche visuellen Aspekte der Website-Ästhetik von den Nutzern subjektiv als am ausschlaggebendsten für die wahrgenommene Ästhetik der Website empfunden werden. Die Testpersonen sollten die vier Begriffe, die Moshagen und Thielsch zufolge die vier gestalterischen Aspekte der Website-Ästhetik darstellen (s.o.), nach der geschätzten und subjektiv empfundenen Wichtigkeit für die Ästhetik einer Website sortieren. Auf diese Weise konnten die mithilfe des VisAWI berechneten Werte zusätzlich mit den jeweiligen Eigeneinschätzungen abgeglichen werden.

3 Ergebnisse

Die Korrelationen zwischen den Bewertungen der visuellen Aspekte innerhalb des VisAWI und den geschätzten Ästhetikbewertungen ergaben, dass in beiden Kulturen der visuelle Aspekt ‚Kunstfertigkeit‘ durch die stärkste Korrelation am ausschlaggebendsten für die Ästhetikbewertung ist. Kulturelle Unterschiede konnten diesbezüglich nicht festgestellt werden. Die Korrelationen zwischen den Gruppen zeigten keine signifikanten Unterschiede. Innerhalb der *Card sorting*-Aufgabe wurde in beiden Kulturen der visuelle Aspekt ‚Einfachheit‘ als am ausschlaggebendsten für die Bewertung als ästhetisch eingeschätzt, gefolgt von der ‚Kunstfertigkeit‘. Auch hierbei lagen keine Unterschiede zwischen den deutschen und den lateinamerikanischen Teilnehmenden vor.

In der kulturunabhängigen Betrachtung konnten neben den Unterschieden zwischen den intuitiven und reflektiven Bewertungen zudem allgemein posi-

tere reflektive als intuitive Ästhetikbewertungen festgestellt werden. Die Ergebnisse der kulturvergleichenden Betrachtung zeigen, dass sich die lateinamerikanischen und die deutschen Teilnehmenden darin unterscheiden, bei welchen Homepages Differenzen zwischen den reflektiven und den intuitiven Ästhetikwahrnehmungen bestehen. Weiterhin ergab die Untersuchung, dass die Kulturen sich nicht in den reflektiven, jedoch in den intuitiven Ästhetikwahrnehmungen der Homepages unterscheiden. Es konnte eine bessere Bewertung der Homepages bei der intuitiven Wahrnehmung von den lateinamerikanischen gegenüber den deutschen Teilnehmenden festgestellt werden.

Zusammenfassend bestätigen die Ergebnisse der vorliegenden Arbeit daher aufgrund der festgestellten kulturellen Unterschiede die Wichtigkeit der Adaption von Websites an unterschiedliche Kulturen. Außerdem kann geschlossen werden, dass Websites allgemein hin stärker hinsichtlich Aspekten der Kunstfertigkeit optimiert werden sollten, um kulturübergreifend eine positivere Wahrnehmung zu erzielen.

Die Ergebnisse tragen dazu bei zu verstehen, wie die Ästhetik von Websites intuitiv wie auch reflektiv in der deutschen und in der lateinamerikanischen Kultur wahrgenommen wird. Die Arbeit liefert somit wichtige Erkenntnisse für künftiges Webdesign und unterstreicht die Wichtigkeit der Kulturadaption von internationalen Online-Auftritten.

Literatur

- Cyr, D.; Head, M.; Larios, H. (2010): Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies*, 68 (1–2), 1–21.
- Kurosu, M.; Kashimura, K. (1995): Apparent Usability vs. Inherent Usability: Experimental analysis on the determinants of the apparent usability. In: J. Miller (Hrsg.): *Conference Companion on Human Factors in Computing Systems*. New York, NY: ACM Press, S. 292–293.
- Lindgaard, G.; Fernandes, G.; Dudek, C.; Brown, J. (2006): Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25 (2), 115–126.
- Meidl, O. (2013): *Global Website: Webdesign im internationalen Umfeld*. Wiesbaden: Springer Gabler [Zugl.: Masterthesis, Alpen-Adria-Univ., Klagenfurt, 2013].

- Moreland, R. L.; Zajonc, R. B. (1982): Exposure effects in person perception: Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology* 18 (5), 395–415.
- Reinecke, K.; Gajos, K. Z. (2014): Quantifying visual preferences around the world. In: M. Jones, P. Palanque, A. Schmidt, T. Grossman (Hrsg.): *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM Press, S. 11–20.
- Robins, D.; Holmes, J. (2008): Aesthetics and credibility in web site design. *Information Processing and Management* 44 (1), 386–399.
- Sun, H. (2001): Building a culturally-competent corporate web site. In: S. R. Tilley (Hrsg.), *SIGDOC '01: Proceedings of the 19th Annual International Conference on Computer Documentation*. New York: ACM Press, S. 95–102.
- Sutcliffe, A. G. (2002): Assessing the reliability of heuristic evaluation for Web site attractiveness and usability. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE Comput. Soc., S. 1838–1847.
- Thielsch, M. T. (2008): *Ästhetik von Websites: Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.
- Thielsch, M. T.; Moshagen, M. (2011): Erfassung visueller Ästhetik mit dem Vis-AWI. In: H. Brau, A. Lehmann, K. Petrovic, M. C. Schroeder (Hrsg.): *Tagungsband UP11*. Stuttgart: German UPA, S. 260–265.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 439–444. DOI: doi.org/10.5283/epub.44961.

Towards the Identification of Information Needs in Conversational Search Dialogues

Alexander Frummet

University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany
alexander.frummet@ur.de

Abstract

As conversational search becomes more pervasive, it becomes increasingly important to understand the user's underlying needs when they converse with such systems in diverse contexts. We report on an in-situ experiment to collect conversationally described information needs in a home cooking scenario. A human experimenter acted as the perfect conversational search system. Based on the transcription of the utterances, we present a coding scheme comprising 27 categories to annotate the information needs of users. Moreover, we use these annotations to perform prediction experiments based on random forest classification to establish the feasibility of predicting the information need from the raw utterances. We find that a reasonable accuracy in predicting information need categories is possible.

Keywords: conversational search; information needs; cooking

1 Introduction and motivation

Voice-based interaction systems are changing the way people seek information, making search more conversational. Spoken queries are very different to typed queries and by mining spoken interaction data, intelligent assistance can be provided. Information Science has a rich tradition of studying conversations to learn about user information needs (Taylor, 1962) and their behavior to address these (Belkin, 1987). Voice-based digital assistants such as Amazon Echo and Google Home show that information seeking conversations now take place in diverse situations embedded in users' everyday lives. One crucial feature for this kind of assistant is the ability to understand and

infer user needs. With conversational search tipped to dominate search in the future (Culpepper, 2018), it is crucial to understand how conversations vary in these diverse domains. If systems could move beyond simply providing results to offering intelligent assistance, complex tasks for which this mode of interaction is suited (Radlinski & Craswell, 2017) can be resolved better. According to Culpepper et al. (2018), accurately eliciting information needs is a key challenge for conversational search. The focus of this thesis is on need elicitation – specifically on understanding and predicting user information needs, which are important for systems to conversationally identify what a user requires, facilitate appropriate retrieval and attain relevance feedback (Trippas, 2018). We study information needs in the domain of home cooking, which, based on the literature, we believed would be a fertile context for the kinds of complex needs suited to conversational search (see Elweiler, 2017; Cunningham & Bainbridge, 2013) and a situation where users simultaneously perform practical, sometimes cognitively challenging tasks that make searching in the traditional sense problematic.

We perform an in-situ study that facilitates a naturalistic cooking situation resulting in the organic development of information needs. We do this with respect to the following research questions:

RQ1: Which information needs occur in cooking situations?

RQ2: Can these needs be identified using conversational utterances from the user?

To answer these questions, we analyze the collected data qualitatively to learn about the diverse types of information needs which can occur in this context, and we utilize machine learning approaches to classify needs using the raw transcription of participant utterances.

2 Methods

2.1 Data collection

To establish a corpus of naturalistic conversational data large enough to perform machine learning prediction, we devised an in-situ user study. We simulated a natural cooking situation by gifting a box of ingredients to participants in their own kitchen. Participants were tasked with cooking a meal

which they had not cooked before based on as many of the contained ingredients as possible. To assist the process, they could converse with the experimenter who would answer questions and needs using any resource available to him via the Web. The experimenter communicated answers in a natural human fashion (arguably the optimal behavior for a conversational system). The advantage of this method was that it provided naturalistic speech utterances – we uncovered real information needs, described by participants in their own words while cooking in their own kitchen. Moreover, the method achieved a corpus large enough to run machine learning experiments.

2.2 Ingredients

To ensure divergent recipes and conversations the ingredient boxes varied across participants. The ingredients typically had a value of around €10 and were chosen based on guidelines by the German Nutritional Society (DGE). Typically, the box contained some kind of grain or starch (e.g., potatoes or rice), a selection of vegetables and a source of protein (e.g., eggs). Participants prepared diverse meals using the ingredients, a selection of which can be found in Figure 1.



Fig. 1 Some example meals cooked during the experiments

2.3 Participants

45 Participants (22 females, $M_{age} = 24$ years, $Range = 19-71$ years, 20% non-students) were recruited using a snowball sampling technique. This method offers two advantages. First, it generates a basis for trust among the participants and the experimenter which Castella et al. (2000) claim leads to more informal and open speech. Second, it allowed a relatively large sample to be achieved.

2.4 Transcription and identification of needs

In total, 38.75 hours of material were collected with the language spoken being German. The recorded conversations were transcribed and annotated by a trained linguist, who was also the experimenter, using the recommendations by Dresing and Pehl (2015). Thereafter, the utterances were split into queries whereby one query could represent one or more information needs. In total, trials yielded on average 36.93 queries ($Range = 7-73$, $Q1/Q3 = 22/50$, $MED = 36.0$, $SD = 17.48$, $skewness = 0.26$, $curtosis = 2.19$). The overall number of queries extracted was $N_q = 1662$. Utterances could have the form of direct and indirect questions. They could also represent implicit and explicit actions.

3 Analyses

We analyzed the collected data both qualitatively and quantitatively. First, we examine the information needs identified to establish the variation of needs that occurred. The resulting set of information needs annotated with an appropriate category was used for quantitative experiments, to establish the feasibility of automatically categorizing the queries (information needs) using machine learning with the raw utterance text.

3.1 Coding scheme for information needs

The starting point for the coding scheme was the set of categories derived for cooking related questions posted on the Google Answers forum in Cunningham and Bainbridge (2013). Then, in an iterative process akin to the coding process by Strauss and Corbin (1996), each query was taken in turn, and a category from the existing scheme was attempted to be applied. When none of the existing categories were suitable, a new category was derived. The outcomes of this classification were 27 different information need categories. The distribution of these can be observed in Figure 2.

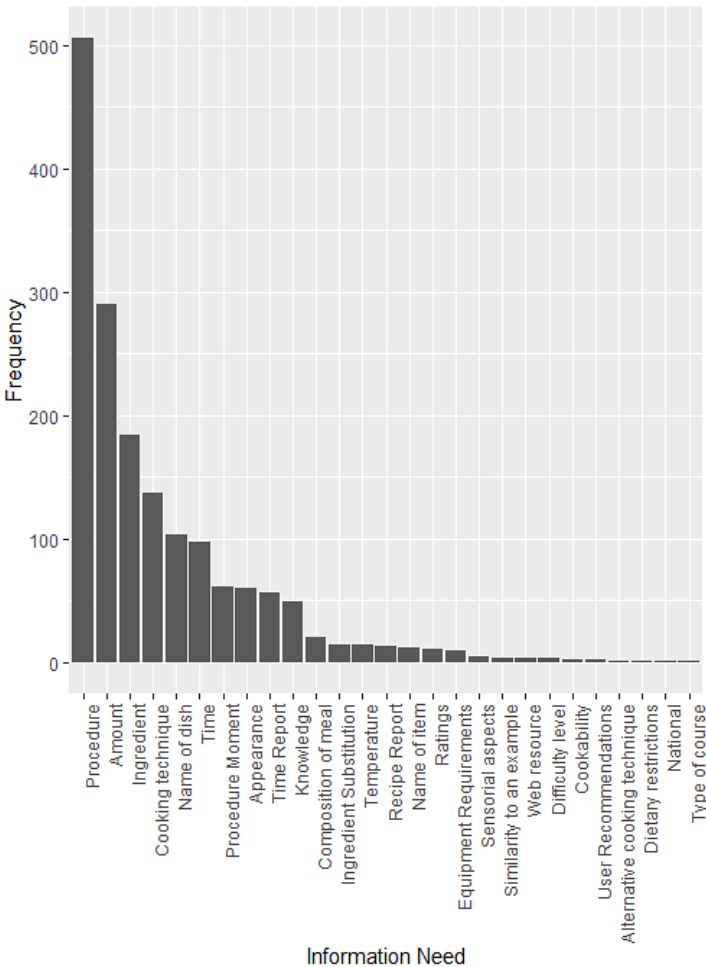


Fig. 2 Information need frequencies

3.2 Predicting information needs

The quantitative analysis was formulated as a prediction task, i.e., given a set of features derived from the raw conversational utterances and context information, is it possible to predict the category of information need. Since most information need categories contained only few turns, the prediction experiments are only concerned with the top 10 categories. We employed a random forest classifier for this purpose because it turned out to be an effec-

tive approach in Shiga et al.'s (2017) work. As the use of word embeddings was shown to be beneficial for predicting information need categories in Shiga et al., we used 200-dimensional word embeddings as a baseline feature for all classification experiments. By doing this, we achieved an average accuracy of 40%. Using resampling methods, we could significantly increase average accuracy up to 64%.

4 Conclusions

The results obtained by the prediction experiments show that the queries issued to the conversational search system are useful for distinguishing different information needs. Our results suggest that information need categories during conversation can be predicted with average accuracy values achieved of up to 64% when resampling is used. Even the non-resampled performance of approximately 40% is significantly larger than chance (which would be 10% with ten classes).

In the thesis, we discuss in detail the implications that our work has for the design of spoken conversational search systems in particular in specific domains such as the kitchen, thereby, making a contribution to the Information Science literature generally.

References

- Belkin, N. J., Brooks, H. M., & Daniels, P. J. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies*, 27(2), 127–144.
- Castellà, V. O., Zornoza Abad, A. M. Alonso, F. P., & Silla, J. P. (2000). The influence of familiarity among group members, group atmosphere and assertiveness on uninhibited behavior through three different communication media. *Computers in Human Behavior*, 16(2), 141–159.
- Culpepper, J. S., Diaz, F., & Smucker, M. D. (2018). Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum*, 52(1), 34–90.

- Cunningham, S. J., & Bainbridge, D. (2013). An Analysis of Cooking Queries: Implications for Supporting Leisure Cooking Ethnographic Studies of Cooks and Cooking. In Schamber, L. (Ed.), *iConference Proceedings* (pp. 112–123). iSchools.
- DGE: Die deutsche gesellschaft für Ernährung e.V. (2018). <https://www.dge.de/wir-ueber-uns/die-dge/>
- Dresing, T., & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende* (6th ed.). Marburg: Dr. Dresing und Pehl, Marburg.
- Elsweiler, D., Trattner, C., & Harvey, M. (2017). Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 575–584). New York, NY: ACM Press.
- Radlinski, F., & Craswell, N. (2017). A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway* (pp. 117–126). New York, NY: ACM Press.
- Shiga, S., Joho, H., Blanco, R., Trippas, J., & Sanderson, M. (2017). Modelling information needs in collaborative search conversations. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 715–724). New York, NY: ACM Press.
- Strauss, A., & Corbin, J. (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz.
- Taylor, R. S. (1962). The process of asking questions. *American Documentation*, 13(4), 391–396.
- Trippas, J., Spina D., Cavedon, L., Joho, H., & Sanderson, M. (2018). Informing the Design of Spoken Conversational Search: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA* (pp. 32–41). New York, NY: ACM Press.

Information Behavior von unbegleiteten, minderjährigen Geflüchteten unter Berücksichtigung der Rolle und Nutzung von Smartphones

Leyla Dewitz

University of Applied Sciences Potsdam (FHP), Germany

ORCID-ID: <https://orcid.org/0000-0003-1333-9209>

leyla.dewitz@fh-potsdam.de

Die Masterthesis wurde im Rahmen des Studiengangs Informationswissenschaft M.A. am Lehrstuhl Information Behavior des Instituts für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin angefertigt.

Abstract

Diese explorative Studie untersucht das Information Behavior von unbegleiteten, minderjährigen Geflüchteten¹ vor, während und nach ihrer Flucht nach Deutschland und inwieweit dabei die Nutzung von Smartphones eine Rolle spielt. Ziel der Untersuchung ist es, Verhaltensweisen jugendlicher Geflüchteter in Bezug auf das Suchen, Finden, Evaluieren, Teilen und Vermeiden von Information zu erforschen. Zur Beantwortung der Forschungsfrage wurde die Methode des qualitativen Leitfadeninterviews genutzt. Die befragten Jugendlichen des Samples ($n = 10$) waren männlichen Geschlechts und bildungsbenachteiligten Milieus zuzuordnen. Ergebnisse zeigen strategisches Vorgehen der Jugendlichen bei der Suche nach Informationen weniger im Heimatland und auf der Flucht, jedoch im Exil in Deutschland. Flüchtiges oder passives Aufnehmen bis hin zu aktivem Vermeiden von Informationen zeigen sich als hervorstechende Merkmale. Smartphones werden nur begrenzt zur Übersetzung und Navigation auf der

1 Kinder und Jugendliche, die unter 18 Jahren ohne Begleitung eines für sie verantwortlichen Erwachsenen in einen Mitgliedsstaat der EU einreisen oder dort ohne Begleitung zurückgelassen werden, gelten nach dem deutschen Asylverfahren als unbegleitete Minderjährige.

Flucht eingesetzt, gelten jedoch als Hilfsmittel zur Befriedigung von sozial-emotionalen Grundbedürfnissen und zum Spracherwerb.

Keywords: Information Behavior; unbegleitete Minderjährige; Geflüchtete; Smartphone

1 Einleitung

Die Zahl der Menschen, die vor Kriegen, Konflikten und Verfolgung fliehen und in den Staaten der Europäischen Union Schutz suchen, war zu keinem Zeitpunkt so hoch wie in den vergangenen Jahren. Dabei wurde in den Jahren 2015 und 2016 der Einsatz mobiler Kommunikationstechnik zu einem Hauptmerkmal der Fluchtbewegungen und Thema medialer Berichterstattungen. Smartphones bzw. die Verwendung mobiler Software-Applikationen – darunter die Nutzung sozialer Netzwerke, Kommunikations-Apps sowie Navigations- und Übersetzungsprogramme – galten für viele Menschen als essenzielle Hilfen auf ihrer Flucht nach Europa. Diese explorative Studie untersucht das Information Behavior von unbegleiteten, minderjährigen Geflüchteten² vor, während und nach ihrer Flucht nach Deutschland und inwieweit dabei die Nutzung von Smartphones und Apps eine Rolle spielt. Ziel der Untersuchung ist es, Verhaltensweisen jugendlicher Geflüchteter in Bezug auf das Suchen, Finden, Evaluieren, Teilen und Vermeiden von Information zu erforschen. Die Studie fokussiert dabei auch die Frage, welche Relevanz der Zugang zu digitalen Ressourcen für die Jugendlichen bei der Planung und Durchführung ihrer Flucht sowie nach ihrer Ankunft in Deutschland besitzt. Dabei kann angenommen werden, dass durch existenzielle Bedrohungssituationen vor, während und nach der Flucht spezielle Bedürfnisse und Probleme entstehen sowie Herausforderungen bewältigt, Entscheidungen getroffen und Strategien gefunden werden müssen, um diesen entkommen zu können.

2 Geflüchtete stellen keine homogene Gruppe an Menschen dar; die Begriffszuschreibung ist aufgrund von einem gemeinsamen kontextuellen Faktor, der Flucht, bestimmt. Die Zuschreibung zur Kategorie des Geflüchteten beruht daher auf einem gemeinsamen und verbindenden Aspekt einer ähnlichen Erfahrung in der Vergangenheit. Die Heterogenität der zu untersuchenden sozialen Gruppe ist daher immer zu berücksichtigen.

2 Vorgehen der Studie

In Kooperation mit der Erst-Inobhutnahme-Einrichtung des Trägers ZEFIE – Zentrum für individuelle Erziehungshilfen auf dem Gelände des Patrick-Henry-Village in Heidelberg – wurden zehn qualitative Interviews mit männlichen³, unbegleiteten Minderjährigen zwischen 15 und 17 Jahren mithilfe von Übersetzern in der jeweiligen Muttersprache durchgeführt (Farsi, Paschtu, Dari, Arabisch). Die befragten Jugendlichen des Samples ($n=10$) waren überwiegend bildungsbenachteiligten Milieus zuzuordnen und aus Städten im Iran, in Syrien und kleineren Provinzen in Afghanistan nach Deutschland geflohen. Die Vorauswahl der potenziellen Interviewteilnehmenden wurde durch die pädagogischen Fachkräfte der Einrichtung vorgenommen und erfolgte nach Kriterien der physischen und psychischen Stabilität und Belastbarkeit der Jugendlichen. Die Beteiligung an dieser Studie geschah dabei auf vollkommen freiwilliger Basis. Die Studie wurde für den emotionalen Verarbeitungsbedarf der Befragten vor, während und nach den Interviews durch eine pädagogische Betreuung unterstützt. Zudem nahm die ethische Reflexion des Forschungsvorhabens einen wesentlichen Teil der Arbeit ein.

3 Information Behavior von jugendlichen Geflüchteten

Der Definition nach wird *Information Behavior* auf Dimensionen des Nutzens, Suchens, Findens, Teilens, Evaluierens und Vermeidens von Information betrachtet (Bates, 2009; Case/Given, 2016). Forschung zum Information Behavior fokussiert meist die *Information Needs* und das *Information Seeking Behavior* eines Individuums bzw. einer bestimmten sozialen Gruppe.

3 Kriterien geschlechtsspezifischer Diversität konnten für die Auswahl des Samples nicht berücksichtigt werden. Ein gleicher Anteil weiblicher Jugendlicher wäre im Hinblick auf die Kontrastierung der Ergebnisse empfehlenswert. Der Feldzugang zu unbegleiteten minderjährige Mädchen ist eingeschränkt, da der Anteil der männlichen Geflüchteten bei den Inobhutgenommenen überwiegt und Mädchen einer meist größeren psychischen Belastung im Rahmen der Flucht ausgesetzt sind. Es wurde deswegen entschieden, das Sample nicht um weibliche Jugendliche zu erweitern.

Information Seeking Behavior wird dabei als eine menschliche Aktivität verstanden, die ein Bedürfnis, sich zu informieren, voraussetzt bzw. als eine Konsequenz daraus entstehen kann. Im Fokus des Information Need steht dabei eine zuvor ins Bewusstsein gerufene Informationslücke, die eine Reaktion des Verhaltens auslösen kann (Stimulus), wobei der Mensch einen Informationsbedarf erkennt und diesen möglicherweise zu stillen versucht. Die Erkenntnis der Notwendigkeit, Informationen zu suchen, impliziert ebenso die Anerkennung, dass der eigene Wissensstand nicht ausreichend ist, um ein latentes oder reales Bedürfnis nach Information zu befriedigen (Case/Given, 2016). Informationsbedürfnisse stellen somit einen Ausgangspunkt dar, der zu einem Informationsverhalten (Information Behavior) führt. Information Behavior fokussiert neben den Vorgängen der bewussten Informationssuche auch die Gesamtheit anderer unbeabsichtigter und passiver Verhaltensweisen. So werden z.B. Handlungen, die im Kontext von aktivem Vermeiden von Informationen stehen (*Information Avoidance*), auch als Teil des Information Behaviors betrachtet. Daher werden in dieser Arbeit alle Dimensionen, sowohl aktive als auch passive Verhaltensweisen, die das Information Behavior umfassen, untersucht.

Erkenntnisse zum Information Behavior von adolescenten Geflüchteten und zum Einsatz von Smartphones bzw. der Nutzung digitaler Ressourcen sind in der LIS-Forschung unterrepräsentiert. Es zeigt sich gegensätzlich dazu ein hoher Bedarf für die Informationsarbeit, um Angebote entsprechend ausrichten zu können. Studien zum Information Behavior von Adolescenten zeigen, dass diese hauptsächlich informelle Quelle, wie Eltern oder Peers einbeziehen, um latente oder reale Informationsbedürfnisse zu befriedigen (u. a. Shenton/Dixon, 2004; Agosto/Hughes-Hassell, 2006). Zudem sind Adolescenten in ihrem Information Behavior durch das Aufwachsen in einer digitalen Welt geprägt (u.a. Abbas/Agosto, 2013; Bilal/Beheshti, 2014). In Studien zum Information Behavior von Geflüchteten konnte festgestellt werden, dass sich kulturelle und linguistische Barrieren, sozioökonomische Gegebenheiten und die Umgebung auf das Information Behavior von Geflüchteten auswirken (Koo, 2016; Fisher et al., 2016). Zusätzlich zeigt sich der Einsatz mobiler Kommunikationstechnik als wesentlicher Bestandteil informations-

geleiteter Prozesse von Menschen vor allem für die Fluchtroutenplanung (u. a. Mason/Buchmann, 2016; Kaufmann, 2016; Kutscher/Kreß, 2015).⁴

4 Methoden

Die Studie verwendet die Methode des leitfadengestützten Interviews zur Datenerhebung. Die Auswahl dieses Erhebungsinstruments begründet sich mit dem explorativen Ansatz der Arbeit und gründet sich zudem auf die limitierte wissenschaftliche Erkenntnislage im Hinblick auf die angestellte Forschung. Weiterhin ist festzustellen, dass individuelle Bedürfnisse und Erfahrungen der Jugendlichen im Fokus der Untersuchung stehen. Diese zu untersuchen, bedarf einer Methode, die es den Befragten ermöglicht, offen über persönliche und individuelle Erfahrungen und Gefühle zu sprechen und ihnen mit einer größtmöglichen Offenheit zu begegnen. Der Interviewleitfaden gliedert sich nach Heimat (vor der Flucht), Flucht (während der Flucht) und Inobhutnahme (nach der Flucht). Dabei ist ein variabler Fragenkomplex zur Smartphone-Nutzung enthalten. Mittels einer Landkarte wurde hier die Aufgabe gestellt, die Fluchtroute zu skizzieren. Durch die Zuordnung von Memory-Kärtchen (blanko zum selbst Beschriften oder mit verschiedenen App-Icons versehen) konnten Fragen zur Nutzung diverser Apps folgen (s. dazu Abb. 1). Dieses partizipative motivierende Mittel half, ein spielerisches Element einzubringen, um die Frage-Antwort-Situation aufzulockern und auf eine haptisch-interaktive Aktivität zu erweitern. Die Auswertung des empirischen Materials erfolgte durch den Prozess der *Thematic Analysis* (Braun/Clarke, 2006). Für den Auswertungsprozess wurde die Software MAXQDA zur Transkription der Audiodateien, während des Kodierens und zum Verfassen analytischer Memos verwendet. Der Kodierungsprozess nach Saldaña (2016) ermöglichte, Inhalte der einzelnen Interviews im Hinblick auf Ähnlichkeiten, Unterschiede, Muster und Beziehungen zu kontrastieren.

4 Die Masterthesis wurde im Dezember 2016 eingereicht und im Februar 2017 verteidigt, daher beinhaltet der Stand der Forschung ausschließlich Literatur, die bis zum Jahr 2016 zu dem Thema veröffentlicht wurde.

5 Ergebnisse

Im Rahmen der Studie wurde das Information Behavior von unbegleiteten, minderjährigen Geflüchteten vor, während und nach ihrer Flucht nach Deutschland untersucht. Zudem wurde erforscht, welche Relevanz der Zugang und die Nutzung zu digitalen Ressourcen (Smartphones und Apps) spielen.⁵ Ergebnisse zeigen vor allem intuitive und informell geleitete Prozesse im Rahmen von informationsbezogenen Verhaltensweisen der Befragten. Strategisches Vorgehen der Jugendlichen bei der Suche nach Informationen konnte weniger im Heimatland und auf der Flucht, jedoch im Exil in Deutschland festgestellt werden. Verhaltensweisen wie das flüchtige oder passive Aufnehmen bis hin zu aktivem Vermeiden von Informationen traten dabei häufig auf. So kann im Zuge der Analyse im Bereich des Information Seeking festgestellt werden, dass zwar Situationen entstehen, in denen latente oder reale Informationsbedürfnisse aufkommen, diese aber nur in wenigen Fällen dazu führen, dass Informationen aktiv eingeholt werden. Die Stimuli, wie das Bewusstwerden einer Informationslücke, setzen keine aktiven Prozesse in Gang. Die informationsbezogenen Verhaltensweisen der Adoleszenten sind vor allem durch kontextuelle Faktoren wie Zeit, Umgebung und Bildung massiv eingeschränkt. Eine Evaluation von Informationsquellen sowie die Prüfung des Informationsgehalts durch das Kontrastieren verschiedener Quellen finden nicht statt. Dies führt an vielen Stellen zu Fehl- und Falschinformation. Beispielsweise zeigte sich während der Flucht, dass der eigene Wissensstand nicht ausreicht, um eine Flucht zu planen. Diese Informationslücke führt jedoch nicht zu einem aktiven Einholen von Informationen über Strecken und Wege, sondern zu einer Übertragung aller Routenentscheidungen an Schlepper. Die Jugendlichen waren durch das Vermeiden der Einholung von (zusätzlichen) Informationen, nach eigenen Aussagen, ernsthaften, lebensbedrohlichen Gefahren ausgesetzt. Von den Jugendlichen favorisiert werden allen voran interpersonelle Informationsquellen, z.B. Menschen, denen eine Expertenfunktion zugeschrieben wird (Mitarbeiter offizieller Stellen, Schlepper, Eltern), Menschen in einer ähnlichen Situation oder Menschen mit ähnlichem kulturellem Hintergrund. Wenn latente Informa-

5 Illustrative Zitate aus den Interviews können auf den Folien des Vortrags am Berliner Bibliothekswissenschaftlichen Kolloquiums nachgelesen werden: <https://rs.cms.hu-berlin.de/bbk/pages/view.php?ref=181> (06.01.2021).

tionsbedürfnisse entstehen, werden gezielt Personen gesucht, die einen niedrigschwiligen Zugang suggerieren.

Es konnte festgestellt werden, dass die Jugendlichen ihre Smartphones nur begrenzt zur Übersetzung und Navigation auf der Flucht einsetzen, da der limitierte Zugang zu Datenvolumen und Strom die Nutzung der Geräte fast unmöglich machte. Angesichts dieser kontextuellen Faktoren und mangelnder Informationskompetenzen aufgrund etwaiger Bildungsdefizite findet das Smartphone als Hilfsmittel der Fluchtroutenplanung bei den Befragten kaum Anwendung. Das Smartphone dient eher zur Ablenkung von den Strapazen der Flucht oder zur Kommunikation mit der Familie. Ein zentrales Ergebnis dieser Studie ist, dass das Smartphone von den unbegleiteten Jugendlichen vor allem als Hilfsmittel zur Befriedigung von Grundbedürfnissen eingesetzt wird und nicht, wie andere Studien zeigten, als Medium, um die Flucht strategisch zu organisieren. Es werden vor allem diverse Kommunikations-Apps verwendet, um Kontakte mit Peers und Familie konstant in allen drei Lebensabschnitten (Heimatland, Flucht, Deutschland) aufrechtzuerhalten.

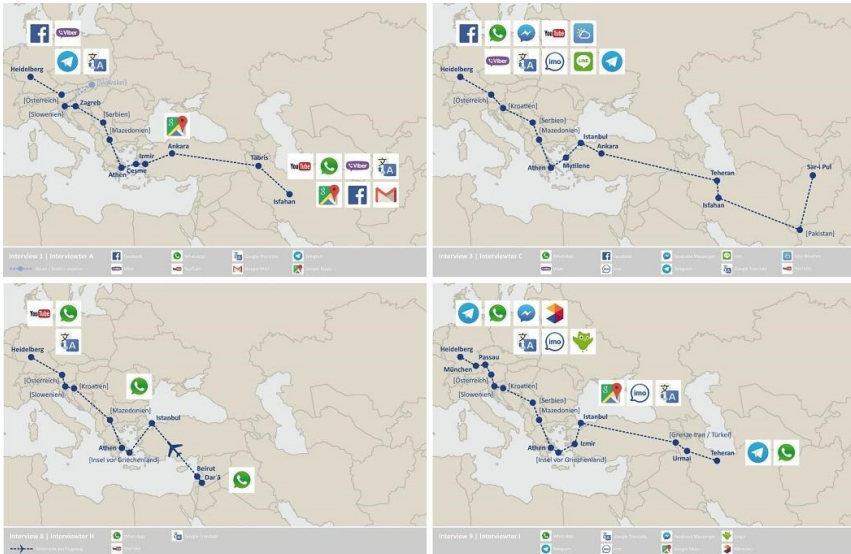


Abb. 1 Fluchtrouten und App-Nutzung der Interviewten I, A, C, H

Jugendliche, die nicht über ausreichend Lese- und Schreibfähigkeiten verfügen, nutzen bestimmte Apps, um per Sprachnachrichten oder Videotelefonie ihre Familien zu erreichen und mit diesen zu kommunizieren. Analpha-

betismus stellt dadurch für die Jugendlichen keine zwingende Barriere bei der Verwendung der Kommunikation-Apps dar. Das Erlernen der deutschen Sprache zeigt sich als ein hervorstechendes kognitives Bedürfnis innerhalb informationsgeleiteter Prozesse der Jugendlichen in Deutschland, wofür diverse Sprach- und Übersetzungs-Apps verwendet werden (s. Abb. 1). Es können hier inklusive und anregende Potenziale für die Förderung aktiver Informations-Interaktionen bei der Verwendung von Informations- und Kommunikationstechnologie durch die geflüchteten Jugendlichen festgestellt werden.

Zusammenfassend untersuchte die Studie das Information Behavior von jugendlichen Geflüchteten aus einer ganzheitlichen Sicht und in Bezug auf die Rolle und Nutzung von digitalen Ressourcen. Die Kernaussage ist dabei, dass aktive informationsgeleitete Verhaltensweisen aufgrund diverser kontextueller Gegebenheiten oft nicht nachgegangen werden kann, jedoch Informationsbedürfnisse und aktive Informations-Interaktionen vor allem im Zielland Deutschland stark hervortreten und der Einsatz digitaler Ressourcen hierfür als eine essenzielle Hilfe gilt.

Referenzen

- Abbas, J.; Agosto, D. E. (2013): Everyday life information behavior of young people. In: J. Beheshti, A. Large (Hrsg.): *The information behavior of a new generation. Children and teens in the 21st century*. Lanham: Scarecrow Press, S. 65–91.
- Agosto, D. E.; Hughes-Hassell, S. (2006a): Toward a Model of the Everyday Life Information Needs of Urban Teenagers: Part 1: Theoretical model. *Journal of the American Society for Information Science and Technology* 57 (10), 1394–1403. doi: [10.1002/asi.20451](https://doi.org/10.1002/asi.20451)
- Agosto, D. E.; Hughes-Hassell, S. (2006b): Toward a Model of the Everyday Life Information Needs of Urban Teenagers: Part 2: Empirical Model. *Journal of the American Society for Information Science and Technology* 57 (11), 1418–1426. doi: [10.1002/asi.20452](https://doi.org/10.1002/asi.20452)
- Bates, M. J. (2009): Information Behavior. In: M. J. Bates, M. N. Maack (Hrsg.): *Encyclopedia of Library and Information Sciences*. New York: CRC Press, S. 2381–2391. doi: [10.1081/E-ELIS3](https://doi.org/10.1081/E-ELIS3)
- Bilal, D.; Beheshti, J. (Hrsg.) (2014): *New Directions in Childrens' and Adolescents' Information Behavior Research*. Bingley: Emerald Publishing.

- Braun, V.; Clarke, V. (2006): Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (2), 77–101. doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)
- Case, D. O.; Given, L. M. (2016): *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Bingley, UK: Emerald.
- Fisher, K. E.; Yefimova, K.; Yafi, E. (2016): Future’s Butterflies: Co-Designing ICT Wayfaring Technology with Refugee Syrian Youth. In: *iConference 2016 Proceedings*. New York, NY: ACM, S. 25–36. doi: [10.1145/2930674.2930701](https://doi.org/10.1145/2930674.2930701)
- Kaufmann, K. (2016): The Empowered Refugee: The Smartphone as a Tool of Resistance on the Journey to Europe. In: *AoIR Selected Papers of Internet Research* 6. <https://journals.uic.edu/ojs/index.php/spir/article/view/9034>
- Koo, J. H. (2016): Information-Seeking within Negative Affect: Lessons from North Korean Refugees’ Everyday Information Practices within PTSD. *Journal of the Korean Society for Library and Information Science* 50 (1), 285–312. doi: [10.4275/KSLIS.2016.50.1.285](https://doi.org/10.4275/KSLIS.2016.50.1.285)
- Kutscher, N.; Kreß, L.-M. (2015): Internet ist gleich mit Essen. Empirische Studie zur Nutzung digitaler Medien durch unbegleitete minderjährige Flüchtlinge. Deutscher Kinderhilfswerk, Universität Vechta. https://images.dkhw.de/fileadmin/Redaktion/I.I_Startseite/3_Nachrichten/Studie_Fluechtlingskinder-digitale_Medien/Studie_digitale_Medien_und_Fluechtlingskinder_Langversion.pdf
- Mason, B.; Buchmann, D. (2016): *ICT4Refugees. A report on the emerging landscape of digital responses to the refugee crisis*. Bonn: Deutsche Gesellschaft für Internationale Zusammenarbeit. <http://www.betterplace-lab.org/wpcontent/uploads/ICT4Refugees-Report.pdf>
- Saldaña, J. (2016): *The Coding Manual for Qualitative Researchers*. Los Angeles, CA: Sage.
- Shenton, A. K.; Dixon, P. (2004a): The development of young people’s information needs. *Library and Information Research* 28 (89), 25–34.
- Shenton, A. K.; Dixon, P. (2004b): The development of young people’s information-seeking behaviour. *Library & Information Research* 28 (90), 31–39.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 452–460. DOI: doi.org/10.5283/epub.44963.

Der Einfluss des Verständnisses von Suchmaschinenwerbung auf das Rechercheverhalten am PC und am mobilen Endgerät

Eine Nutzerstudie

Sebastian Schultheiß

Hochschule für Angewandte Wissenschaften Hamburg (ROR-ID: [oofkqwx76](#))

Fakultät DMI, Department Information. Finkenau 35, 22081 Hamburg

ORCID: [0000-0003-2704-7207](#), sebastian.schultheiss@haw-hamburg.de

Abstract

Repräsentative Untersuchungen ergaben, dass Suchmaschinennutzer zum Großteil weder das von Werbung bestimmte Geschäftsmodell Googles kennen noch bezahlte von organischen Ergebnissen unterscheiden können. Daran knüpft die vorliegende Masterarbeit an und untersucht, welchen Einfluss der Kenntnisstand über Suchmaschinenwerbung auf das Rechercheverhalten am PC sowie am Smartphone ausübt. Dafür wurde ein Methodenmix aus Interview, Eye-Tracking-Experiment und Fragebogen mit 50 studentischen Probanden eingesetzt. Die Probanden verfügten über ausgeprägte Kenntnisse im Bereich der Suchmaschinenwerbung, die nicht mit ihrem Rechercheverhalten korrelierten. Hinsichtlich des bei der Suche verwendeten Endgerätes zeigten sich zum Teil große Unterschiede im Blickverhalten. Diese Unterschiede lassen sich auf den Einfluss der direkten Sichtbarkeit eines Suchergebnisses an beiden getesteten Geräten zurückführen. Zukünftige Studien bieten sich dahingehend an, einerseits das Experiment mit einem anderen Sample zu wiederholen sowie andererseits zu untersuchen, wann und mit welcher Motivation Anzeigen ausgewählt werden und welcher Einfluss daraus auf den Wissenserwerb entsteht.

Keywords: Suchmaschine; Google; Suchmaschinennutzer; Suchmaschinenwerbung; SEA; Anzeigen; Ads; Eye-Tracking; Experiment

1 Einleitung

Suchmaschinen nehmen in unserer Gesellschaft eine bedeutende Rolle ein. Sie zählen zu den am meisten verwendeten Diensten des Internets, wodurch sie einen großen Einfluss auf den Wissenserwerb der InternetnutzerInnen ausüben (Beisch/Schäfer, 2020). Der Suchmaschinenmarkt wird im Wesentlichen von Google beherrscht, was sich durch einen Marktanteil von etwa 90% innerhalb Europas ausdrückt (European Commission, 2017). Neben organischen Ergebnissen finden sich auf der Google-Suchergebnisseite bezahlte Ergebnisse (Anzeigen) wieder, insofern ein Werbetreibender ein Gebot für die Platzierung einer Anzeige abgegeben hat. Werbetreibende bezahlen jedoch nicht bereits für die Darstellung ihrer Anzeigen auf der Suchergebnisseite (*search engine result page*; SERP), sondern für jeden Klick, der darauf erfolgt (Google.com, 2020). Alphabet Inc., der Mutterkonzern Googles, generiert seine Einnahmen hauptsächlich durch Werbung für Produkte und Dienstleistungen (Alphabet Inc., 2020) und profitiert somit unmittelbar davon, wenn eine möglichst große NutzerInnenzahl auf die geschalteten Anzeigen klickt.

Untersuchungen ergaben, dass SuchmaschinennutzerInnen zum Großteil weder das von Werbung bestimmte Geschäftsmodell Googles kennen noch in der Lage sind, die organischen von den bezahlten Suchergebnissen verlässlich zu unterscheiden (Lewandowski et al., 2018). Als ursächlich dafür gilt eine unzureichende Kennzeichnung der Anzeigen. Diese wird durch Google regelmäßig verändert, wobei die Entwicklung hin zu unauffälligeren Kennzeichnungen geht. Dies geht wiederum mit steigenden Klickzahlen auf den Anzeigen einher (Edelman, 2014). Google kann somit vorgeworfen werden, das Problem der mangelhaften Unterscheidbarkeit im Sinne höherer Einnahmen zumindest in Kauf zu nehmen (Lewandowski, 2016). Zusätzlich profitiert Google auch vom Vertrauen, das die NutzerInnen in Suchmaschinen haben, ihnen die relevantesten Ergebnisse zu liefern (u. a. Purcell et al., 2012). Dieses Vertrauen bringen sie aufgrund ihrer unzureichenden Unterscheidungsfähigkeit auch den Anzeigen entgegen. Da Anzeigen jedoch stark durch die Werbetreibenden beeinflusst werden, ist das nutzerseitige Vertrauen im Falle der bezahlten Suchergebnisse zu hinterfragen (Lewandowski, 2017).

In dieser Masterarbeit wurde untersucht, (1) wie sich die Kenntnisse über Anzeigen auf das Blick- und Klickverhalten am Desktop-PC sowie am

Smartphone auswirken und (2) ob sich das Verhalten an beiden Endgeräten voneinander unterscheidet.

2 Forschungsfragen

- F1:** „Gibt es einen Zusammenhang zwischen dem Kenntnisstand der SuchmaschinennutzerInnen über Anzeigen und ihrem Blickverhalten auf Anzeigen?“
- F2:** „Gibt es einen Zusammenhang zwischen dem Kenntnisstand der SuchmaschinennutzerInnen über Anzeigen und ihrem Klickverhalten auf Anzeigen?“
- F3:** „Unterscheidet sich das Blick- und Klickverhalten der SuchmaschinennutzerInnen auf den Anzeigen bei der Desktopversion der Google-Suche von dem der mobilen Version der Google-Suche?“

3 Methodik

Zur Beantwortung der Forschungsfragen wurde eine Methodenkombination aus Eye-Tracking-Experiment, Interview und Fragebogen eingesetzt.

Das Eye-Tracking-Experiment stellt den Kern der Studie dar. Anhand vordefinierter Rechercheaufgaben wurde das Blick- und Klickverhalten auf SERPs mithilfe eines stationären (am Desktop-PC) sowie eines mobilen Eye-Trackers (am Smartphone) erfasst. An beiden Endgeräten bearbeiteten die ProbandInnen insgesamt 20 Aufgaben, davon zehn informations-, sechs transaktions- sowie vier navigationsorientierte (Broder, 2002). Ein Beispiel einer informationsorientierten Aufgabe ist: *„Sie wollen sich über Sehenswürdigkeiten in Wien informieren. Google präsentiert Ihnen die folgende Ergebnisseite. Bitte klicken Sie ein Ergebnis an.“*. Zielsetzung war somit bei allen Aufgaben ein einzelner Ergebnisklick. Alle SERPs beinhalteten organische Ergebnisse und Textanzeigen sowie im Falle der transaktionsorientierten Anfragen zusätzliche Shoppinganzeigen.

Vor dem Eye-Tracking-Experiment fand ein kurzes Interview zur Erfragung demografischer Daten und der Suchmaschinennutzung statt. Im Anschluss an das Eye-Tracking-Experiment wurde mithilfe eines Fragebogens

der Kenntnisstand der ProbandInnen hinsichtlich Google-Anzeigen erfasst. Der Fragebogen beinhaltete Wissensfragen, beispielsweise zur Haupteinahmequelle Googles, sowie Aufgabenstellungen, bei denen auf SERP-Screenshots je nach Fragestellung entweder Anzeigen oder organische Ergebnisse zu markieren waren. Insgesamt konnte im Fragebogen eine Gesamtpunktzahl zwischen 0 und 100 Punkten erzielt werden, wobei die Maximalpunktzahl dann erreicht wurde, wenn sowohl alle Wissensfragen korrekt beantwortet als auch alle Anzeigen und organischen Ergebnisse fehlerfrei identifiziert worden sind.

Zur Teilnahme an der Studie wurden 50 Studierende aus verschiedenen Fachrichtungen akquiriert, mit der Zielsetzung, dadurch ein möglichst breites Spektrum an Anzeigenkenntnissen in der Stichprobe vorzufinden.

4 Ergebnisse

Anzeigenkenntnis

Die ProbandInnen erreichten in der Mehrheit sehr hohe Punktzahlen im Fragebogen zum Anzeigen-Kennntnisstand ($MW = 83,4$, $SD = 18,9$). Lediglich sieben der 50 TeilnehmerInnen erzielten weniger als 70 Punkte.

Zusammenhang zwischen Anzeigenkenntnis und Blick- sowie Klickverhalten

Rangkorrelationsanalysen nach Spearman konnten keine signifikanten Zusammenhänge zwischen den Variablen „Fragebogenpunktzahl“ und „Fixationszahl auf Anzeigen“ ($r_s = -,030$, $p = ,067$) sowie zwischen „Fragebogenpunktzahl“ und „Klickzahl auf Anzeigen“ ($r_s = ,056$, $p = ,351$) nachweisen. Zwischen dem Anzeigenkenntnisstand sowie dem Blick- und Klickverhalten bestehen demnach keine Zusammenhänge.

Unterschiede im Blick- und Klickverhalten zwischen Desktop-PC und Smartphone

Abbildung 1 stellt die Fixationsraten auf den Anzeigen im Vergleich beider Endgeräte dar. Mittels Chi-Quadrat-Tests wurden die Fixationsraten beider Endgeräte miteinander verglichen, wobei vier signifikante Ergebnisse entstanden sind. Die obere Textanzeige an erster ($\chi^2(1) = 8,423$, $p = ,004$) Stelle sowie die erstplatzierte Shoppinganzeige ($\chi^2(1) = 4,548$, $p = ,033$) wurden am Smartphone signifikant häufiger fixiert, wohingegen die drittplatzierte obere

Textanzeige ($\chi^2(1) = 7,608, p = ,006$) und die drittplatzierte Shoppinganzeige am Desktop-PC mehr Fixationen erhielten ($\chi^2(1) = 7,145, p = ,008$).

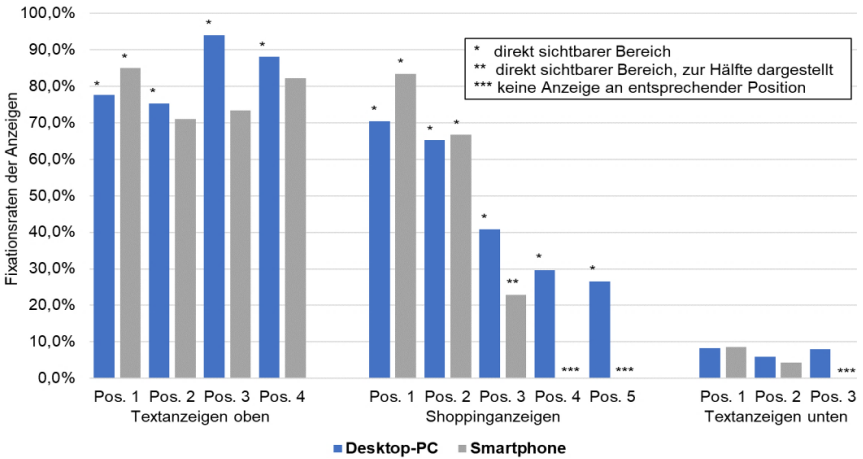


Abb. 1 Fixationsraten der Anzeigen

Mittels Chi-Quadrat-Tests wurden die Fixationsraten beider Endgeräte miteinander verglichen, wobei vier signifikante Ergebnisse entstanden sind. Die oberen Textanzeigen an erster ($\chi^2(1) = 8,423, p = ,004$) und dritter ($\chi^2(1) = 7,608, p = ,006$) Stelle sowie die erstplatzierte Shoppinganzeige ($\chi^2(1) = 4,548, p = ,033$) wurden am Smartphone signifikant häufiger fixiert, wohingegen die drittplatzierte Shoppinganzeige am Desktop-PC mehr Fixationen erhielt ($\chi^2(1) = 7,145, p = ,008$).

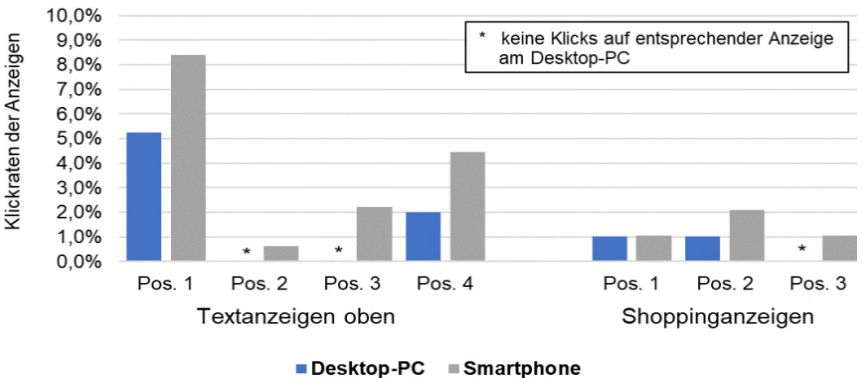


Abb. 2 Klickraten auf Anzeigen

Die generell niedrigen Klickraten auf den Anzeigen werden in Abbildung 2 abgebildet. Nicht in der Abbildung finden sich die Textanzeigen am unteren Ende der SERP wieder, welche durch die ProbandInnen an beiden Endgeräten nicht ausgewählt wurden. Mittels Chi-Quadrat-Tests wurden die Klickraten am Desktop-PC und Smartphone miteinander verglichen. Dabei konnten keine signifikanten Unterschiede nachgewiesen werden.

5 Fazit

Anhand einer Methodenkombination aus Eye-Tracking-Experiment, Interview und Fragebogen wurde der Einfluss der Anzeigenkenntnis auf das Rechercheverhalten untersucht. Hierfür wurden 50 studentische ProbandInnen akquiriert. Deren umfangreiche Anzeigenkenntnisse konnten nicht mit ihrem Blick- und Klickverhalten auf Anzeigen korreliert werden. Hinsichtlich des bei der Suche verwendeten Endgerätes (Desktop-PC und Smartphone) zeigten sich zum Teil große Unterschiede im Blickverhalten. Diese lassen sich auf die Bildschirmgrößen beider Endgeräte zurückführen. Da am Smartphone ausschließlich Anzeigen im direkt sichtbaren Bereich platziert waren, wurden diese wesentlich intensiver betrachtet als die Anzeigen am Desktop-PC, jedoch nicht häufiger ausgewählt. Die Ergebnisse zeigen zum einen, dass eine Wiederholung der Studie mit nicht-studentischen Probanden zielführend wäre, um in der Stichprobe eine größere Vielfalt unterschiedlicher Kenntnisstände hinsichtlich Suchmaschinenwerbung vorzufinden. Zum anderen bieten die Ergebnisse die Gelegenheit zu weiterer Forschung hinsichtlich der Frage, wann und mit welcher Motivation Anzeigen ausgewählt werden und wie sich diese unterschiedlich informierten Auswahlentscheidungen auf den Wissenserwerb der SuchmaschinennutzerInnen auswirken.

Die Masterarbeit ist frei zugänglich (Schultheiß, 2019).

Literatur

Alphabet Inc. (2020): Alphabet Announces Fourth Quarter and Fiscal Year 2019 Results. https://abc.xyz/investor/static/pdf/2019Q4_alphabet_earnings_release.pdf

- Beisch, N.; Schäfer, C. (2020): Internetnutzung mit großer Dynamik: Medien, Kommunikation, Social Media. *Media Perspektiven* 9, 462–481. https://www.ard-zdf-onlinestudie.de/files/2020/0920_Beisch_Schaefer.pdf
- Broder, A. (2002): A taxonomy of web search. *ACM SIGIR Forum* 36 (2), 3–10. <https://doi.org/10.1145/792550.792552>
- Edelman, B. (2014): Google's Advertisement Labeling in 2014. <http://www.benedelman.org/adlabeling/google-colors-oct2014.html>
- European Commission (2017): Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service – Factsheet. http://g8fipikplyr33r3krz5b97di.wpengine.netdna-cdn.com/wp-content/uploads/2017/06/Google-MEMO-17-1785_EN.pdf
- Google (2020): Werbekosten bei Google: Pay-per-Click – Google Ads. https://ads.google.com/intl/nl_de/home/pricing/
- Lewandowski, D. (2016): Die Suchergebnisseite als Dauerwerbesendung? In: H. C. Mayr, M. Pinzger (Hrsg.): *Informatik 2016 : 26.–30. September 2016, Klagenfurt*. Bonn: Gesellschaft für Informatik, S. 183–193. http://searchstudies.org/wp-content/uploads/2016/11/Lewandowski_Die-Suchergebnisseite-als-Dauerwerbesendung.pdf
- Lewandowski, D. (2017): Users' Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice* 5 (4), 6–25. <https://doi.org/10.1633/JISTaP.2017.5.4.1>
- Lewandowski, D.; Kerkmann, F.; Rümmele, S.; Sünkler, S. (2018): An empirical investigation on search engine ad disclosure. *Journal of the Association for Information Science and Technology* 69 (3), 420–437. <https://doi.org/10.1002/asi.23963>
- Purcell, K.; Brenner, J.; Rainie, L. (2012): *Search Engine Use 2012*. Pew Internet and American LifeProject. https://www.eff.org/files/pew_2012_o.pdf
- Schultheiß, S. (2019): *Der Einfluss des Verständnisses von Suchmaschinenwerbung auf das Rechercheverhalten am PC und am mobilen Endgerät: eine Nutzerstudie*. Masterarbeit, HAW Hamburg. <https://doi.org/10.13140/RG.2.2.21818.03526>

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 461–467. DOI: doi.org/10.5283/epub.44964.

Weitere Bände der „Schriften zur Informationswissenschaft“

I. Dorsch, K. Fietkiewicz, A. Ilhan, C. Meschede, T. Siebenlist (Hg.): Facetten von Wolf Stock und ihre Bedeutung für die Informationswissenschaft Festschrift zu Ehren von Wolfgang G. Stock (Bd. 73; 2020) ISBN 978-3-86488-167-1, 27,80€

Gabriele Julia Irle: Gefühlserleben bei der Informationssuche im Internet Eine qualitative Studie zur Individualität und Alltäglichkeit der Sucherfahrung (Bd. 72; 2017) 236 S., ISBN 978-3-86488-126-8, 29,50 €

Julia Maria Struß: Multilinguales aspektbasiertes Opinion Mining Entwicklung eines ressourcenarmen Extraktionsverfahrens und Untersuchung von Nutzerperspektiven (Bd. 71; 2017) 408 S., ISBN 978-3-86488-123-7, 34,90 €

Maria Gäde, Violeta Trkulja, Vivien Petras (Eds.): Everything Changes, Everything Stays the Same? Understanding Information Spaces Proc. ISI 2017 (Bd. 70; 2017) 392 S., ISBN 978-3-86488-117-6, 33,90 €

Terje Tüür-Fröhlich: The Non-trivial Effects of Trivial Errors in Scientific Communication and Evaluation (Bd. 69; 2016) 164 S., 978-3-86488-104-6, 24,80 €

Ben Heuwing: Usability-Ergebnisse als Wissensressource in Organisationen (Bd. 68; 2015) 460 S., ISBN 978-3-86488-084-1, 35,80 €

Ulrich Herb: Open Science in der Soziologie Eine interdisziplinäre Bestandsaufnahme zur offenen Wissenschaft und eine Untersuchung ihrer Verbreitung in der Soziologie (Bd. 67; 2015) 500 S., ISBN 978-3-86488-083-4, 36,80 €

Franjo Pehar, Christian Schlögl, Christian Wolff (Eds.): Re:inventing Information Science in the Networked Society Beiträge ISI 2015 (Bd. 66; 2015) 596 S., ISBN 978-3-86488-081-0, 38,50 €

Anke Reinhold: Forschungsdaten in der videobasierten Unterrichtsforschung Benutzerzentrierte Modellierung und Evaluierung einer Domänen-Ontologie (Bd. 65; 2015) 280 S., ISBN 978-3-86488-061-2, 29,90 €

Norbert Henrichs: Menschsein im Informationszeitalter Informationswissenschaft mit Leidenschaft und missionarischem Eifer [Pioniere der Informationswissenschaft | Norbert Henrichs] (Bd. 64; 2014) 404 S., ISBN 978-3-86488-061-2, 33,90 €

Hans-Christoph Hobohm (Hg.): Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten Beiträge ISI 2013 (Bd. 63; 2013) 502 S., ISBN 978-3-86488-035-3, 34,50 €

Saskia-Janina Untiet-Kepp: Adaptive Feedback zur Unterstützung in kollaborativen Lernumgebungen (Bd. 62; 2012) 322 S., 978-3-86488-023-0, 30,90 €

Carola Carstens: Ontology Based Query Expansion Retrieval Support for the Domain of Educational Research (Bd. 60; 2012) 396 S., 978-3-86488-011-7, 34,90 €

Matthias Görtz: Social Software as a Source of Information in the Workplace Modeling Information Seeking Behavior of Young Professionals in Management Consulting (Bd. 59; 2011) 314 S., ISBN 978-3-86488-006-3, 31,90 €

Joachim Griesbaum, Thomas Mandl, Christa Womser-Hacker (Hg.): Information und Wissen: global, sozial und frei? Beiträge ISI 2011 (Bd. 58; 2011) 536 S., ISBN 978-3-940317-91-9, 36,50 €

Information between Data and Knowledge

Digital humanities as well as data science as neighboring fields pose new challenges and opportunities for information science. The recent focus on data in the context of big data and deep learning brings along new tasks for information scientist for example in research data management. At the same time, information behavior changes in the light of the increasing digital availability of information in academia as well as in everyday life.

In this volume, contributions from various fields like information behavior and information literacy, information retrieval, digital humanities, knowledge representation, emerging technologies, and information infrastructure showcase the development of information science research in recent years.

Topics as diverse as social media analytics, fake news on Facebook, collaborative search practices, open educational resources or recent developments in research data management are some of the highlights of this volume.

For more than 30 years, the International Symposium of Information Science has been the venue for bringing together information scientists from the German speaking countries. In addition to the regular scientific contributions, six of the best competitors for the prize for the best information science master thesis present their work.

www.vwh-verlag.de

Verlag Werner Hülsbusch
Fachverlag für Medien-
technik und -wirtschaft

ISBN: 978-3-86488-172-5



9

783864

881725

ISSN: 0938-8710

ISSN: 978-3-86488-172-5

32,80 € (D)

33,80 € (A)

39,80 CHF