

Literatur im Wikiversum – Eine praktische Annäherung über API-Abfragen und Wikipedia-Metriken

Illmer, Viktor J.

v.illmer@fu-berlin.de
Freie Universität Berlin, Deutschland
ORCID: 0000-0002-7334-781X

Soethaert, Bart

bart.soethaert@fu-berlin.de
Freie Universität Berlin, Deutschland
ORCID: 0000-0002-3845-605X

Welz, Lilly

l.welz@fu-berlin.de
Freie Universität Berlin, Deutschland

Fischer, Frank

fr.fischer@fu-berlin.de
Freie Universität Berlin, Deutschland
ORCID: 0000-0003-2419-6629

Jäschke, Robert

robert.jaeschke@hu-berlin.de
Humboldt-Universität zu Berlin, Deutschland
ORCID: 0000-0003-3271-9653

Wikipedia als Repräsentant für populäres Wissen über Literatur

Die kollaborativ erstellte Online-Enzyklopädie Wikipedia bietet mit derzeit über 60 Millionen Artikeln in über 300 Sprachversionen (Wikimedia, 2023; Wikipedia-Beitragende, 2023) Informationen zu den unterschiedlichsten Wissensbereichen. Nach dem Wiki-Prinzip werden die Beiträge von einer weltweiten Gemeinschaft freiwilliger Redakteur*innen erstellt, bearbeitet und gepflegt. Diese Community überprüft aktiv die Referenzen und Änderungen in bestehenden Artikeln sowie die Veröffentlichung neuer Artikel auf die Einhaltung der redaktionellen Richtlinien. Aufgrund des offenen Redaktionsmodells ist Wikipedia ein enzyklopädisches Gemeinschaftsprojekt, das seinesgleichen sucht (Danowski et al. 2005; Lovink et al. 2012; Tkacz, 2015).

Neben der individuellen Lektüre der Fließtexte (und ihrer jeweiligen Versionshistorie) über die Website bietet die

Online-Plattform über eine API weitere Möglichkeiten zur Analyse der enzyklopädischen Inhalte und des Community-Engagements. Die Vielzahl an Metadaten, sowohl zu den einzelnen Themen selbst als auch zur Bearbeitung und Nutzung durch die aktiv partizipierende Community bzw. die Leser*innen sowie die semantischen Verknüpfungen lassen sich auch mit digitalen Methoden sammeln, quantifizieren und auswerten.

Auch die rezeptionsorientierte Literaturwissenschaft hat das Projekt inzwischen als Forschungsgegenstand und Datenressource entdeckt (vgl. Hube et al., 2017; Chiu, 2022; Fischer et al. 2023b), da es viele enzyklopädische Beiträge und Metadaten zur Literatur und zum literarischen Leben versammelt, zu Autor*innen, literarischen Werken, Genres, Epochen und weiteren literaturgeschichtlich relevanten Kategorien.

Jüngste Untersuchungen in diesem Bereich werten die inhaltliche Reichweite von Wikipedia etwa im Hinblick auf die Aufnahme und Darstellung von einzelnen Autor*innen (Blakesley, 2018; Bronner, 2018; Fischer et al., 2019; Blakesley, 2022b), Gruppen (Blakesley, 2020; Carrillo-Jara, 2023), literarischen Werken (Blakesley, 2022a), literarischen Figuren (Picard et al., 2023; Wojcik et al., 2023), Gattungen (Figlerowicz, 2023) und Kanones (Miller et al., 2016; van der Deijl et al., 2018; Wojcik et al., 2019; Lippolis, 2023) aus. Unterschiede in der Verteilung enzyklopädischer Artikel zu bestimmten Themen in verschiedenen Sprachen geben Aufschluss über das unterschiedliche Interesse und die attestierte Relevanz dieser Themen für bestimmte Sprachgemeinschaften. Darüber hinaus können Veränderungen der Seitenaufrufe, der Überarbeitungen und der Beitragenden auch im Zeitverlauf analysiert werden, um das sich entwickelnde Interesse an und die Auseinandersetzung mit bestimmten literarischen Autor*innen und Werken zu verfolgen. Die datenanalytische Auswertung anhand solcher Wikipedia-Metriken ermöglicht es somit, die Auseinandersetzung mit Literatur in Wikipedia evaluierbar zu machen und Aussagen über literarische Kanonizität, Wertungspraktiken und Popularität im Kontext offener Enzyklopädieprojekte weiter zu diversifizieren. In kritischer Auseinandersetzung mit der Kanon- und Popularitätsforschung in globaler Perspektive wird unter anderem besonders deutlich, dass sich in der Wikipedia kein monolithischer Kanon zeigt, sondern viele, sich zudem dynamisch verändernde Kanones manifestieren.

Im Zentrum des Hands-On-Workshops steht die Wikipedia-API, mit deren Funktionsweise die Teilnehmer*innen vertraut gemacht werden. Sukzessive werden Abfrageskripte in Form eines Jupyter Notebooks erarbeitet. Um eine benutzerfreundliche Programmierumgebung anzubieten und langwierige Installationsprozesse zu umgehen, wird für das Ausführen des Notebooks auf Google Colaboratory zurückgegriffen. Im Folgenden werden drei Typen von Abfragen kurz vorgestellt, die im Workshop jeweils im Hinblick auf eigene, von den Teilnehmer*innen mitgebrachte Fragestellungen und Forschungsinteressen modifiziert werden können. Der Programmiercode wurde um Annotationen ergänzt, die es auch Python-Anfänger*innen

ermöglichen, über die bereitgestellten Formulare eigene Anfragen auszuführen.

Autor*innenzentrierte Abfragen

Eine mögliche Abfragestrategie ist die autor*innenzentrierte Abfrage, wie am Beispiel von Theodor Fontane demonstriert werden soll.

Über die Wikipedia-API lässt sich herausfinden, wie viele der über 300 Sprachversionen der Wikipedia einen eigenen Artikel über den Autor bereithalten – die Anzahl dieser Sitelinks gilt in der Forschung als »a simple measure of canonicity« (Kukkonen 2020). Diese können dann etwa diagrammatisch auf ihre Artikelgröße hin verglichen werden (Abb. 1). Auf diese Weise können ebenfalls die Anzahl der Überarbeitungen des Artikels, die Anzahl der Bearbeiter*innen, die Backlinks oder das Datum der Artikelerstellung untersucht werden. Diese Datenpunkte können sprachübergreifend Aufschluss über etwaige Konjunkturen der Fontane-Rezeption geben. Es lassen sich Rückschlüsse auf Anlässe ziehen, die eine Erweiterung der Informationsbasis in der Wikipedia ausgelöst haben könnten (Preise, Jubiläen, Übersetzungen, Schulstoff).

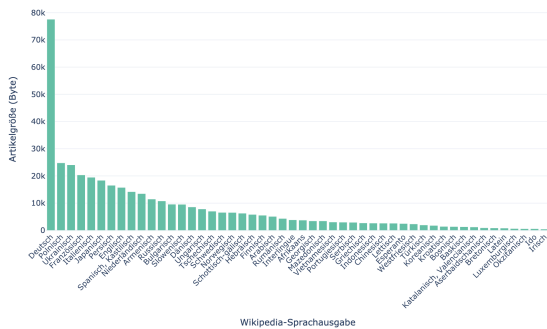


Abb. 1. Größe des Artikels *Theodor Fontane* nach Wikipedia-Sprachausgabe.

Werkzentrierte Abfragen und breitere Kontextualisierung

In ähnlicher Weise kann vorgegangen werden, wenn das Augenmerk auf die Werke Fontanes gerichtet wird. Auch hier können die Anzahl der Überarbeitungen des Artikels, die Anzahl der Bearbeiter*innen, der Backlinks oder das Datum der Artikelerstellung abgefragt und visualisiert werden.

Im Folgenden kann als digitale Entsprechung der literaturhistorischen Praxis ein Vergleichspool anderer Autor*innen zusammengestellt werden, um Fontane und sein literarisches Werk mit denen zeitgenössischer Kolleg*innen zu vergleichen.

Fontanes Geburtsjahr ist 1819, eine mögliche Operationalisierung von Zeitgenossenschaft wäre etwa die Zusammenstellung anderer deutschsprachiger Autor*innen, die bis zu 20 Jahre vor und nach Fontane geboren wurden. Diese Festlegung ist natürlich kontingent und kann individuellen Informationsbedürfnissen angepasst werden.

Die Visualisierung der Artikellänge und ein entsprechendes Ranking ergeben dann beispielsweise, dass Wilhelm Busch unter den zeitgenössischen literarisch schreibenden Autor*innen den aktuell umfangreichsten Artikel vorweisen kann, Fontane aber immerhin in den Top-5 rangiert (Abb. 2). Auch wenn sich diese Artikellängen, die oft über mehr als 20 Jahre gewachsen sind, größenordnungsmäßig nicht so schnell ändern, sind diese Werte in einer communitybetriebenen digitalen Enzyklopädie natürlich durch die Zeit variabel.

Dass Wilhelm Wundt, als Psychologe und Philosoph ebenfalls einflussreicher Autor, bei diesem Ranking mit dem umfangreichsten Artikel ganz vorn steht, zeigt auch, dass es eines weiteren Schritts bedürfte, wollte man die Ergebnisliste auf vorderhand literarisch schreibende Zeitgenoss*innen eingrenzen. Außerdem zeigt sich ein systematischer Bias in der Artikellänge: Zu umfangreiche »Werk«-Abschnitte werden oft in eigene Artikel für Werke ausgliedert, während die Werke »kleinerer« Autor*innen oft Teil der Personenartikel bleiben. Die genaue Kenntnis der Gepflogenheiten innerhalb von Wikipedia erweist sich daher als Voraussetzung für eine sinnvolle Einschätzung der Quantifizierungen.

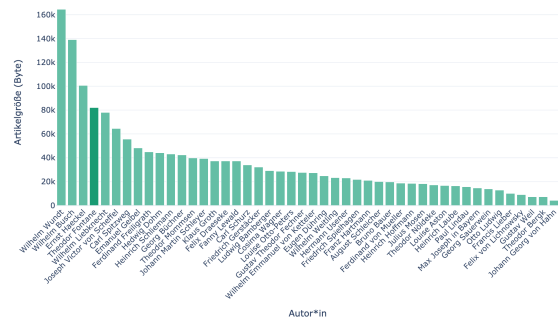


Abb. 2. Umfang der Artikel zu Fontanes Zeitgenoss*innen in der deutschsprachigen Wikipedia.

Abfragen zur Ermittlung der relativen Bedeutung von Artikeln innerhalb der Wikipedia

Neben den inhaltsbezogenen Informationen, die über Wikipedia direkt bezogen werden können, lassen sich über die internen Verweise zwischen Artikeln (Wikilinks) auch Netzwerkmetriken wie der PageRank berechnen (vgl. Thahammer, 2016; Hube et al., 2017). Ähnlich wie dies Such-

maschineralgorithmen zur Bestimmung der Rangfolge von Ergebnissen tun, können diese Zahlen dazu verwendet werden, den Wert und die Bedeutung eines Wikipedia-Artikels und seines Themas innerhalb des Hyperlink-Netzwerks der Plattform einzuschätzen. Der Wert, der einem Wikipedia-Artikel auf der Grundlage dieser Metriken zugewiesen wird, gibt Aufschluss über die relative Bedeutung und Konnektivität eines Themas innerhalb der vernetzten Informationen von Wikipedia.

Ausblick

Die Abbildung der Literatur in Wikipedia und in darauf aufbauenden oder damit verwandten Projekten (DBpedia, Wikidata) wird in den Literaturwissenschaften zunehmend als möglicher Forschungsgegenstand wahrgenommen. Einen repräsentativen Überblick über solche Zugänge und Fragestellungen bietet der Sonderband des *Journal of Cultural Analytics*, »Wikipedia, Wikidata, and World Literature« (Fischer et al., 2023a). Die Teilnehmer*innen des Workshops sollen in die Lage versetzt werden, daran anzuschließen und diese entstehende Praxis für ihre eigenen Forschungsfragen und -interessen produktiv zu machen. Während des interaktiven Workshops werden die Teilnehmer*innen praktisch erfahren und üben, wie sie auf relevante Metadaten in der Wikipedia zugreifen und eigenständig API-Abfragen durchführen können, um Einblicke in die Prozesse, den Umfang und die Inhalte der kollaborativen und selbstregulierten Informationsaggregation über Literatur im Wikiversum für die eigene Forschung nutzen zu können.

Format und Aufbau des Workshops (4h)

- 00:00–00:30 Einführung zum Thema anhand aktueller Forschungsbeiträge
- 00:30–01:00 Einführung in die Wikipedia-API
- 01:00–01:45 Allgemeine Erarbeitung des Jupyter Notebooks
- 01:45–02:00 Pause
- 02:00–03:00 Individuelle Anpassungen des Jupyter Notebooks
- 03:00–03:30 Präsentationen der Ergebnisse
- 03:30–04:00 Abschlussdiskussion und Perspektiven

Zielgruppe und notwendiges Vorwissen

Der Workshop zielt auf Literaturwissenschaftler*innen, aber auch auf Kolleg*innen angrenzender Gebiete. Vorkenntnisse der Programmiersprache Python sowie zu Programmierschnittstellen (APIs) sind hilfreich, aber keine

Voraussetzung zur Teilnahme am Workshop. Den Teilnehmer*innen soll das nötige Praxiswissen vermittelt werden, um eigenständig weiterzuarbeiten.

Zahl der möglichen Teilnehmer:innen

max. 20 Personen

Benötigte technische Ausstattung

Die Teilnehmer*innen benötigen einen eigenen Laptop. Für die Durchführung des Workshops wird ein Beamer benötigt; Flipchart und Stifte wären hilfreich. Als Grundlage für eigene Abfragen können Workshopteilnehmer*innen eine Liste mit Artikelnamen einer einzelnen Wikipedia-Sprachausgabe oder eine Liste mit Wikidata-IDs mitbringen. Der Raum sollte über ausreichend Lademöglichkeiten verfügen und ein verlässliches und schnelles WLAN bieten.

Beitragende

- Viktor J. Illmer
- Bart Soethaert
- Lilly Welz
- Frank Fischer
- Robert Jäschke

Fördernachweis

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder innerhalb des Exzellenzclusters Temporal Communities: Doing Literature in a Global Perspective – EXC 2020 – Projekt-ID 390608380.

Bibliographie

- Blakesley, Jacob.** 2018. » The Global Popularity of William Shakespeare in 303 Wikipedias « . *Memoria Di Shakespeare. A Journal of Shakespearean Studies* 5 (Dezember): 149–71. <https://doi.org/10.13133/2283-8759/14509>.
- . 2020. » World Literature According to Wikipedia Popularity and Book Translations: The Case of Modern Italian Poets « . *Comparative Critical Studies* 17 (3): 433–58. <https://doi.org/10.3366/ccs.2020.0373>.
- . 2022a. » The Global Popularity of Dante's ‚Divina Commedia‘: Translations, Libraries, Wikipedia « .

Journal of Dante Studies 5. <https://repository.upenn.edu/bibdant/vol5/iss1/8>.

———. 2022b. » The Wikipedia Popularity of James Joyce « . *James Joyce Quarterly* 59 (2): 289–313. <https://doi.org/10.1353/jjq.0.0164>.

Bronner, Friedrich Georg. 2018. »WHWP – Walter Höllerer bei Wikipedia«. Doctoral Dissertation, Berlin: Technische Universität Berlin. <https://doi.org/10.14279/depositonce-6865>.

Carrillo-Jara, Daniel. 2023. » Escritor / Qillqaq: The Representation of Peruvian Literature in the Spanish and Quechua Wikipedias « . *Journal of Cultural Analytics* 8 (2). <https://doi.org/10.22148/001c.73258>.

Chiu, Kuei-fen. 2022. » World Literature in an Age of Digital Technologies: Digital Archive, Wikipedia, and Goodreads.com « . In *The Making of Chinese-Sinophone Literatures as World Literature*, herausgegeben von Kuei-fen Chiu und Yingjin Zhang, 1. Aufl., 217–36. Hong Kong University Press. <https://doi.org/10.2307/j.ctv2j6xdjn>.

Danowski, Patrick und Voss, Jakob. 2005. » Das Wissen der Welt – die Wikipedia « . In *Open Source Jahrbuch 2005*, herausgegeben von Matthias Bärwolff, Robert A. Gehring, und Bernd Lutterbeck, 393–405. Berlin: Lehmanns Media. http://www.opensourcejahrbuch.de/download/jb2005/chapter_06/osjb2005-06-05-danowskivoss.

van der Deijl, Lucas und Smeets, Roel. 2018. » The Canon of Dutch Literature According to Wikipedia: A Network Analysis of 2286 Wikipedia-Entries on Dutch Authors. « In *Proceedings of EADH2018: » Data in Digital Humanities «*. Galway: National University of Ireland. https://eadh2018.exordo.com/files/papers/104/final_draft/Abstract_Long_Paper_EADH_2018_Van_der_Deijl__Smeets_final_version.pdf.

Figlerowicz, Matylda und Mertehikian, Lucas. 2023. » An Ever-Expanding World Literary Genre: Defining Magic Realism on Wikipedia « . *Journal of Cultural Analytics*, 8 (2). <https://doi.org/10.22148/001c.73249>.

Fischer, Frank, Jacob Blakesley, Paula Wojcik, und Robert Jäschke, Hrsg. 2023a. *Wikipedia, Wikidata, and World Literature*. Special Issue of the *Journal of Cultural Analytics*, 8 (2). <https://culturalanalytics.org/issue/7259>.

Fischer, Frank, Jacob Blakesley, Paula Wojcik, und Robert Jäschke. 2023b. » Preface: World Literature in an Expanding Digital Space « . *Journal of Cultural Analytics*, 8 (2). <https://doi.org/10.22148/001c.74598>.

Fischer, Frank, und Robert Jäschke. 2019. »Fontane im Wikiversum. Ein Beitrag zur digitalen Rezeptionsgeschichte«. Conference Talk gehalten auf der Kongress »Fontanes Medien«, Potsdam, Juni 14. bit.ly/2wShKUT.

Hube, Christoph, Frank Fischer, Robert Jäschke, Gerhard Lauer, und Mads Rosendahl Thomsen. 2017. » World Literature According to Wikipedia: Introduction to a DBpedia-Based Framework « . *ArXiv*. <https://doi.org/10.48550/ARXIV.1701.00991>.

Kukkonen, Karin. 2020. » Does Cognition Translate? « *Poetics Today*, 41 (2): 243–59. <https://doi.org/10.1215/03335372-8172556>.

Lippolis, Anna Sofia. 2023. » Italian Nostalgia: National and Global Identities of the Italian Novel « . *Journal of Cultural Analytics*, 8 (2). <https://doi.org/10.22148/001c.68341>.

Lovink, Geert, und Nathaniel Tkacz, Hrsg. 2012. *Critical Point of View: A Wikipedia Reader*. INC Reader 7. Amsterdam: Institute of Network Cultures. <https://doi.org/10.2139/ssrn.2075015>.

Miller, Ben, Cindy Berger, Sayan Bhattacharyya, Tommaso Caselli, David Kelman, Jennifer Olive, und Jay Rajiva. o. J. » Contextualizing Receptions of World Literature by Mining Multilingual Wikipedias « . In *Digital Humanities 2016: Conference Abstracts*, 282–85. Kraków: Jagiellonian University & Pedagogical University. <https://dh2016.adho.org/abstracts/83>.

Picard, Sophie, Paula Wojcik, und Sina Zarriëb. 2023. » Zirkulation und Wertschöpfung am Beispiel literarischer Figuren « . In *Der Wert der literarischen Zirkulation / The Value of Literary Circulation*, herausgegeben von Michael Gamper, Jutta Müller-Tamm, David Wachter, und Jasmin Wrobel, 431–53. Globalisierte Literaturen. Theorie und Geschichte transnationaler Buchkultur / Globalized Literatures. Theory and History of Transnational Book Culture 3. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-662-65544-3_26.

Thalhammer, Andreas, und Achim Rettinger. 2016. » PageRank on Wikipedia: Towards General Importance Scores for Entities « . In *The Semantic Web: ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 – June 2, 2016, Revised Selected Papers*, herausgegeben von Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, Dunja Mladenić, Sören Auer, und Christoph Lange, 227–40. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47602-5_41.

Tkacz, Nathaniel. 2015. *Wikipedia and the Politics of Openness*. Chicago/London: University of Chicago Press.

Wikimedia. 2023. »Wikimedia Statistics – All Wikipedias – Pages to date«. 2023. https://stats.wikimedia.org/#/all-wikipedia-projects/content/pages-to-date/normal|line|2-year|page_type~content|monthly.

Wikipedia-Beitragende. 2023. » List of Wikipedias « . In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_Wikipedias&oldid=1161354568.

Wojcik, Paula, Bastian Bunzeck, und Sina Zarriëb. 2023. » The Wikipedia Republic of Literary Characters « . *Journal of Cultural Analytics*, 8 (2). <https://doi.org/10.22148/001c.70251>.

Wojcik, Paula, und Sophie Picard. 2019. » Klassiker@wikipedia: Klassikforschung und Digital Humanities. Ein Kommentar zur Studie World Literature According to Wikipedia « . In *Klassik als kulturelle Praxis. Funktional, intermedial, transkulturell*, herausgegeben von Paula Wojcik, Stefan Matuschek, Sophie Picard, und

Monika Wolting, 149–64. spectrum Literaturwissenschaft /
spectrum Literature 62. Berlin/Boston: de Gruyter. <https://doi.org/10.1515/9783110615760-010>.