What Does the German-language Information Science Community Cite?

An Analysis of the German Information Science Handbook "Grundlagen der praktischen Information und Dokumentation"

Dirk Lewandowski[1], *Stefanie Haustein*[2]

[1] Hamburg University of Applied Sciences, Department of Information
Finkenau 35, 22081 Hamburg, Germany
dirk.lewandowski@haw-hamburg.de

[2] Université de Montréal, École de bibliothéconomie
et des sciences de l'information (EBSI)
3150 Rue Jean-Brillant, Montreal, H3T 1N8 (Canada)
stefanie.haustein@umontreal.ca

Abstract

We present a bibliometric analysis of the "Grundlagen der praktischen Information und Dokumentation", the major German-language information science handbook. Using the bibliographic and citation data from the handbook, basic statistics and bibliometric indicators such as the number of papers, citations, citation rates as well as citing half-lives were computed. We found that German-language information science is not a close-knit community. Authors reference their colleagues' work only moderately. From the cluster analysis, we can conclude that German-language information scientist work in areas rather distant from one another.

Keywords: Information science, Citation analysis, Bibliometrics, Germany

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 93–104.

1 Introduction

The performance of their own actors is of interest to research communities. It is no wonder then that information science, especially as it is a discipline also considered with bibliometrics, produced some work on its own performance (Borgman & Furner, 2002), whether considering the top scientist and institutions in the field (Larivière, Sugimoto, & Cronin, 2012), collaboration patterns (Levitt & Thelwall, 2009) or specific journals (Haustein & Larivière, 2014). Our work adds to the work on the citation behavior of German-speaking information scientists, focusing on the central handbook of information science in German language.

"Grundlagen der praktischen Information und Dokumentation" (Kuhlen, Semar, & Strauch, 2013) is the major handbook of information science and practice in the German-speaking countries. Published in its 6th edition since 1972, the 2013 edition contains 54 chapters, written by 46 authors. It covers most aspects of information science and practices. The handbook is also known as "KSS" (derived from the first letters of the editors' last names). KSS is authored by the German-language information science community. As the editors remark in their preface, the authors documented the standard of knowledge in information science and practice[1] (Kuhlen et al., 2013: VIII). Therefore, the KSS handbook makes for a good object of study when it comes to citation patterns of German-language information science. To what degree do German-language information scientists cite each other? What role, in the context of the handbook chapters, do German-language information scientists play in the discourse and in how far do the authors build upon the work of their international colleagues?

To answer these questions, we collected all references from the KSS handbook and performed citation and co-citation analyses as well as social network analysis to determine central documents and authors cited in the handbook.

1 "[…] AutorInnen, deren Artikel in der Gesamtheit den Wissensstand der Informationswissenschaft und -praxis beeindruckend dokumentiert […]"

2 Literature review

There is some prior work on the publication and citation behavior of German-language information science. Schlögl and Stock (2004) found that German-speaking information scientists prefer reading German-language works and preferably publish their work in German. Friedländer (2014) performed a citation analysis of scientist working at German-language research universities using Web of Science, Scopus and Google Scholar data. She identified the leading scholars and institutes (according to citations) and concludes that the visibility of German information science could be increased if authors would publish their work in English rather than German. Based on Web of Science and Google Scholar data, Gärtner (2013) considered the publication output of all full professors at German-language information science institutes from research universities as well as universities of applied sciences, which play an important part in the German-speaking information science community. Gärtner also found that a large degree of the publications are written in German, and that the visibility of German information science could be increased by publishing more work in English. Schlögl (2013) found that the visibility of information science research from Germany and Austria is relatively low, when considering journal articles indexed in Web of Science. Prior research found that authors are biased towards citing documents written in their own language (Leeuwen, Moed, Tijssen, Visser, & Raan, 2001; Yitzhaki, 1998) as well as authors from their own country (Glänzel & Schubert, 2005; Jaffe, 2011; Pasterkamp, Rotmans, de Kleijn, & Borst, 2007).

All reported research leads to the conclusion that a problem with information science from the German-speaking countries may lie in its low international visibility, which results from authors publishing their work in German, and not in English. Further reasons, e.g., whether the quality of the work itself may be a factor, are not discussed.

3 Methods

As the German-speaking information science community has low international visibility and to a large extent publishes in German journals, analyses based on commonly used citation indexes such as Web of Science and Scopus, would only capture a part of its scientific output and impact (Leeuwen et al., 2001). Hence, references were collected manually from all KSS chapters based on the assumption that the handbook largely reflects the German-speaking information science community. It should be noted that Stock and Stock (2013) represents another handbook of German information science but we limit the analysis to KSS to include a broader range of citing authors. Extracting the bibliographic information of all reference cited in the 54 KSS chapters, 1,868 unique documents were identified. For each of these documents, author names, title, year and document type (journal article, book chapter, monograph, etc.) were recorded together with the information in which of the chapters these were cited. Author names were disambiguated based on last and first names provided in the reference lists, complemented by online searches, if necessary. As the 45 authors of the 54 chapters were included in the disambiguation, author self-citation rates could be computed. Using the bibliographic and citation data, basic statistics and bibliometric indicators such as the number of papers, citations, citation rates as well as citing half-lives were computed.

A 2-mode network of direct citations from chapters to cited authors was computed and visualized with UCInet and Netdraw. The spring embedding algorithm was used, which, trying to find a global optimum for the network layout, positions more central nodes – e.g., authors cited in a large number of chapters – in the center and more specialized – those cited in only one chapter – in the periphery of the network graph (Freeman, 2000). Based on the asymmetrical direct citation matrix of citing chapters and cited authors, author co-citation relationships were extracted. Based on the symmetric square matrix of co-cited authors the landscape of important authors was mapped and clustered using VOSviewer choosing particular parameters for optimization[2] (van Eck & Waltman, 2009, 2010). Similarly to the 2-mode network,

2 To optimize cluster and layouting, the following parameters were chosen: min. cluster size = 3; clustering resolution = 0.3; mapping attraction = 5; mapping repulsion = 0. For details regarding these parameters refer to the VOSviewer manual at http://vosviewer.com/download/f-y2.pdf.

authors that are frequently co-cited are positioned close to each other in the network graph. Clusters, indicated by node color, group related authors based on similar co-citation patterns.

4 Results

The 54 chapters were authored on average by 1.4 authors and contained 35.8 cited references. As the chapters are supposed to cover different topics, we did not expect too much overlap regarding cited documents. In fact, 97% of the 1,868 documents received only one of the 1,931 citations, while 49 (2.6%) and 7 (0.4%) publications were cited two and three times, respectively. The seven most cited works[3] are almost all monographs, written in English and published – with the exception of Salton (1963) – during the last ten years.

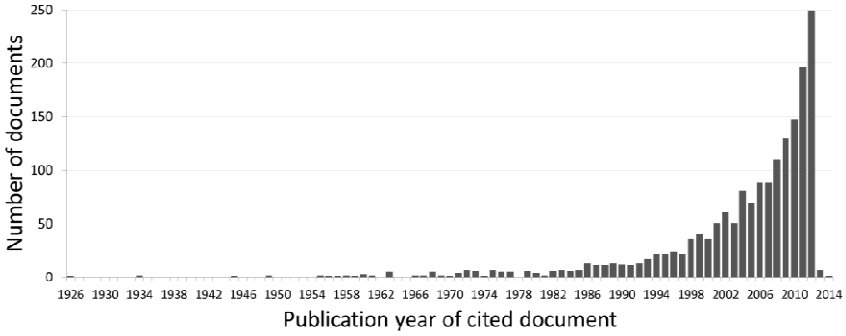

Figure 1. Number of cited documents per publication year.

The literature cited by the chapters is quite recent. For example, of the 1,740 documents for which a publication year could be determined (fig. 1), the majority of papers (53.3%) were published between 2007 and 2014. The overall citing half-life, i.e. the age of references, of the 54 chapters was 5.6 years, which for a handbook, intended to give an overview of the basics of

3 Hobohm (2013), Ingwersen (2005), Kuhlen (2012), Leckie (2010), Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2012), Pariser (2011), Salton (1963)

information science, is quite low. However, the citing half-life varied between chapters from 0.2 (D14[4]) to 19 years (B13[5]), as did the range of publication years of the references from 7 (D15[6]) to 86 years (C10[7]). These differences can partly be explained by the developments of certain topics, such as the chapter about Web Science (C7, half-life = 2.4) compared to that about information retrieval models (B15, 18.0).

In terms of document types, the handbook mostly cited journal articles (34.8%), monographs (29.3%) and to a lesser extent conference papers (12.6%), book chapters (10.4%) as well as online documents (7.3%). Other document types (5.5%) include, for example, standards, newspaper articles, reports and press releases. Particular differences can be observed between chapters such as C9, which cites almost only monographs, B17, B16 and C10, which are largely based on journal articles, D14 and D11, which mainly cite online documents, and D5, which, given its topic, references mostly norms and standards.

Table 2. Most-cited authors (excluding self-citations)

Rank	Author	Times cited	Affiliated country
1	Berners-Lee, T.	11	UK
	Kuhlen, R.	11	Germany
3	Salton, G.	10	USA
	Stock, W. G.	10	Germany
5	Bates, M. J.	9	USA
6	Hjørland, B.	8	Denmark
	McKeown, K.	8	USA
	Shneiderman, B.	8	USA
	Voorhees, E.M.	8	USA
10	Ingwersen, P.	7	Denmark
	Mani, I.	7	USA

4 “Transformation von Buchhandel, Verlag und Druck”

5 “Maschinelle Übersetzung”

6 “Patentinformation und Patentinformationssysteme”

7 “Sziento- und bibliometrische Verfahren”

On the level of cited authors, Kuhlen, Hobohm and Berners-Lee are the most cited considering all citations. Excluding self-citations, the top cited authors are Kuhlen, Berners-Lee, Stock, Salton and Bates (table 2). Among the 20 most-cited authors, there are only two authors from Germany, one of which is the editor of the handbook. Almost half of the most cited authors are from the US.

When looking at the KSS authors, we find that except for Kuhlen, information scientists from the German-language countries are cited only moderately. When excluding self-citations, one author receives 4 citations, three authors 3 citations, seven authors 2 citations, and ten authors one citation, respectively. 24 KSS authors are not cited in the handbook at all.

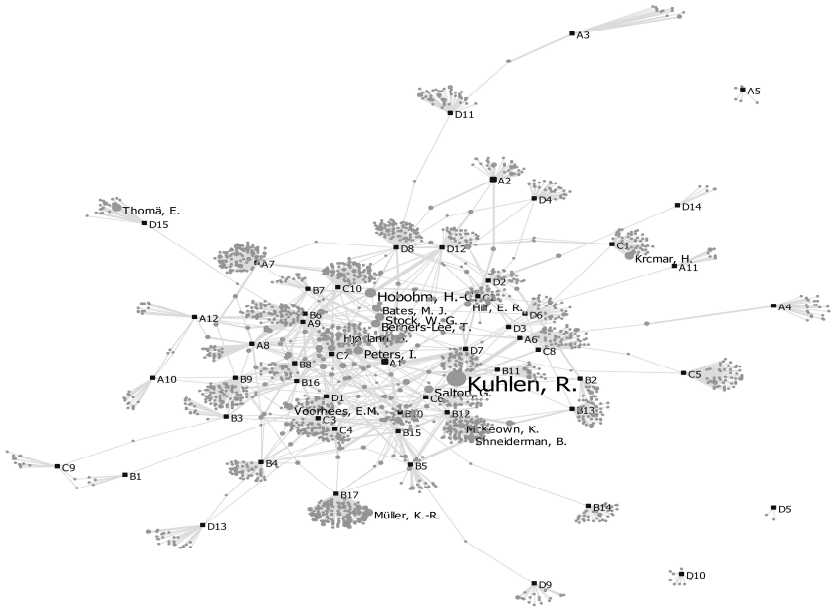

Figure 2. 2-mode network of KSS chapters (blue squares) and cited authors (orange nodes). Node size of cited authors corresponds to number of citations.

Figure 2 visualizes the direct relations between the 54 chapters and 2,491 authors cited in the handbook. It can be seen that the 2-mode network consists of one large component, which links 51 chapters through 2,476 cited authors, and three isolated chapters (A5, D5, D10), which cite authors that have not been included in the other chapters' reference lists. This isolation seems to be either caused by the topics' remoteness from other chapters, such

as norms and standards (D5) and archives (D10), or by national particularities in the case of A5 (copyright and Internet law in Austria). As to be expected, A1 which defines information science and serves as an introduction to KSS, is positioned in the center of the 2-mode network citing 131 authors (degree). The chapter on information behavior by Hobohm (A9) has the same degree, while chapters B17 (machine learning) and B12 (automatic abstracting) obtain even higher scores, citing 194 and 164 different authors, respectively. As shown by the node size, Kuhlen is clearly the most cited author with 26 citations from 10 different chapters.

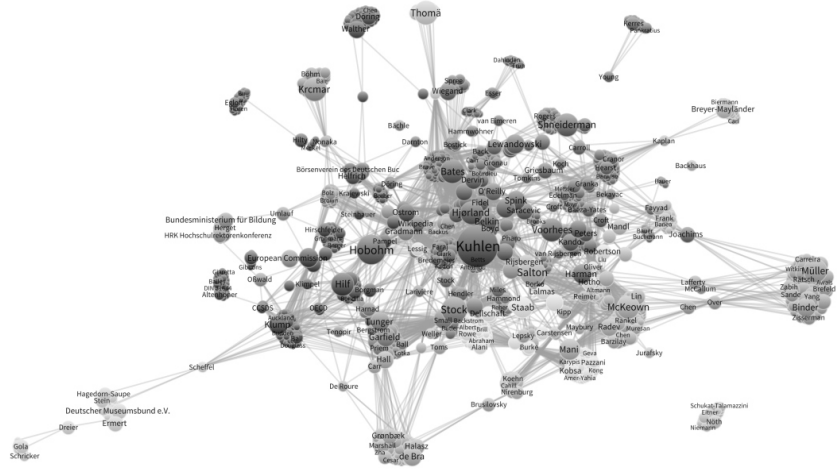

Figure 3. Co-citation network of cited authors. Node size represents number of citations and node color indicates clusters of similar authors.

Figure 3 visualizes 2,476 of 2,491 authors who are connected through co-citations in the largest component of the co-citation network, while 15 authors isolated from the largest component were not included in the visualization and clustering analysis. Choosing the particular parameters described in the Method section, 15 clusters of similar authors were identified. These clusters – identified through color coding in figure 3 and listed in table 3 – grouped between 19 and 355 related authors based on co-citation linkages and thus researchers working on similar topics cited in the KSS handbook. With eight KSS authors each, Clusters 1 and 3 incorporate the largest number of KSS authors, namely Hammwöhner, Reiterer, Jetter, Griesbaum, Lewandowski, Spree, Kerres and Preussler (Cluster 1) and Kuhlen, Seadle, Hilf,

Severiens, Altenhöner, Klump, Bertelmann and Wittenzellner (Cluster 3), while most of the clusters include two to three KSS authors and Clusters 10 and 14 contain none.

Table 3: Number of (KSS) authors and most cited authors per cluster.

Cluster #	Number of authors (KSS)	Most-cited (KSS) authors
1	355 (8)	Shneiderman, Lewandowski, O'Reilly, Wiegand
2	320 (5)	Hobohm, Bates, Hjørland, Voorhees
3	276 (8)	Kuhlen, Hilf, Klump, DFG
4	224 (1)	Staab, Sparck Jones
5	193 (3)	Müller, Binder, Kolmogorov, Kawanabe
6	190 (3)	Breyer-Mayländer, Mandl, Rijsbergen
7	180 (3)	Salton, McKeown, Mani, Radev
8	152 (3)	Krcmar, Gradmann, European Commission, McGill
9	142 (3)	Peters, Stock WG, Garfield, Haustein, Tunger
10	115 (0)	de Bra, Hall, Grønbæk, Halasz
11	114 (2)	Ostrom, Walther, Döring, Helfrich
12	106 (2)	Berners-Lee, Weller, Horrocks, Hendler, Miles, Stock M
13	57 (2)	Ermert, Deutscher Museumsbund e.V.
14	33 (0)	Nöth, Bocklet, Eysholdt, Haderlein, Jurafsky
15	19 (2)	Thomä, Emmerich, Tribiahn

As authors tend to specialize in particular research areas within the field of information science, they can be used as concept markers for particular topics (Larivière et al., 2012). Each of the clusters thus represents particular topics. When analyzing the clusters shown in fig. 3 (with some authors representing them shown in table 3), we can see that, for instance:

1. Cluster 1 represents information retrieval, search engine optimization, web 2.0 and information visualization.
2. Cluster 13 is focused on museums. This cluster is outside the core of information science and has only one author connecting it to the rest of the authors.
3. Cluster 9 is focused on informetrics. With three KSS authors (and four of the top-five authors being German information scientists), this is the only cluster with a clear dominance of German-language authors.

4. Cluster 10 ("hypertext") does not contain any of the KSS authors. This is particularly striking, as hypertext is an important topic for information science and there is even a chapter on it in KSS.
5. Some clusters are quite heterogeneous and some KSS authors are connected to several clusters, functioning as bridges between authors and topics. This is presumably due to these authors publishing widely instead of focusing on one rather narrow area of specialization.

5 Discussion and conclusion

Our research adds to our knowledge on the state of German-language information science. From our analysis of the KSS handbook, we can see that German-language information science is not a close-knit community. Authors reference their colleagues' work only moderately. So, apart from the findings from prior work, that German-language information science is focused on the regional, as well as on German-language publications (Friedländer, 2014; Gärtner, 2013; Schlögl & Stock, 2004; Schlögl, 2013), we found that the work of German-language information scientists is not well cited within the community of KSS authors. From the cluster analysis, we can conclude that German-language information scientist work in areas rather distant from one another. We were only able to identify one cluster with at least three KSS authors (the "informetric cluster").

However, the question remains to what degree a rather small community as that of German-language information science could contribute to overview chapters such as the ones presented in KSS, i.e. whether the large influence of information science research mainly from the US might show that information science *is* in fact an international community. In contrast, if the ratio of citations to authors from the German-speaking countries were high, one could argue that the community does not let in influences from the international community.

The data from the handbook is certainly limited. It may be worthwhile to perform similar studies on other German-language information science publications, like the *ISI Proceedings* and *Information Wissenschaft und Praxis*, the major German-language information science journal.

References

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 3–72.

Freeman, L. C. (2000). Visualizing social networks. *Journal of Social Structure*, 1 (1). http://www.cmu.edu/joss/content/articles/volume1/Freeman.html.

Friedländer, M. B. (2014). Informationswissenschaft an deutschsprachigen Universitäten – eine komparative informetrische Analyse. *Information – Wissenschaft & Praxis*, 65 (2), 109–119. doi:10.1515/iwp-2014-0018

Gärtner, C. (2013). *Produktivität und Impact bibliotheks- und informationswissenschaftlicher Institute im deutschsprachigen Raum: Ein Vergleich wissenschaftlicher Publikationsaktivitäten an Hochschulen*. Hamburg University of Applied Science.

Glänzel, W., & Schubert, A. (2005). Domesticity and internationality in co-authorship, references and citations. *Scientometrics*, 65 (3), 323–342. doi:10.1007/s11192-005-0277-0

Haustein, S., & Larivière, V. (2014). A multidimensional analysis of Aslib Proceedings – using everything but the Impact Factor. *Aslib Journal of Information Management*, 66 (4), 358–380.

Jaffe, K. (2011). Do countries with lower self-citation rates produce higher impact papers? Or, does humility pay? *Interciencia*, 36 (9), 694–698. http://atta.labb.usb.ve/Klaus/art191.pdf.

Kuhlen, R., Semar, W., & Strauch, D. (Eds.) (2013). *Grundlagen der praktischen Information und Dokumentation*. Berlin: De Gruyter Saur.

Larivière, V., Sugimoto, C. R., & Cronin, B. (2012). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology*, 63 (5), 997–1016. doi:10.1002/asi.22645

Leeuwen, T. N. Van, Moed, H. F., Tijssen, R. J. W., Visser, M. S., & Raan, A. F. J. Van (2001). Language biases in the coverage of the Science Citation Index and its consequencesfor international comparisons of national research performance. *Scientometrics*, 51 (1), 335–346. doi:10.1023/A:1010549719484

Levitt, J. M., & Thelwall, M. (2009). Citation levels and collaboration within library and information science. *Journal of the American Society for Information Science and Technology*, 60 (3), 434–442. doi:10.1002/asi.21000

Pasterkamp, G., Rotmans, J. I., de Kleijn, D. V. P., & Borst, C. (2007). Citation frequency: A biased measure of research impact significantly influenced by the

geographical origin of research articles. *Scientometrics*, 70 (1), 153–165. doi:10.1007/s11192-007-0109-5

Schlögl, C. (2013). International visibility of European and in particular German-language publications in library and information science. In: H.-C. Hobohm (Ed.). *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013)* (pp. 50–62). Glückstadt: Verlag Werner Hülsbusch.

Schlögl, C., & Stock, W. G. (2004). Impact and relevance of LIS journals: A scientometric analysis of international and German-language LIS journals – Citation analysis versus reader survey. *Journal of the American Society for Information Science and Technology*, 55 (13), 1155. doi:10.1002/asi.20070

Stock, W. G., & Stock, M. (2013). *Handbook of Information Science*. Berlin, Boston: De Gruyter Saur.

Yitzhaki, M. (1998). The “language preference” in sociology: Measures of “language self-citation”, “relative own-language preference indicator”, and “mutual use of languages.” *Scientometrics*, 41 (1–2), 243–254. doi:10.1007/BF02457981