Newcomer Corner

Mathilde B. Friedländer, Düsseldorf

Der Kanon der Informationswissenschaft

Blaise Cronin hat 2013 die Frage formuliert, ob es so etwas wie einen Kanon der Informationswissenschaft gibt. Er schlägt vor, zur Ermittlung des Kanons auch Namens- und Sachregister von einschlägigen Lehrbüchern informationstrisch auszuwerten. Untersucht wurden englischsprachige Publikationen aus der Kategorie *Information Science & Library Science* der Fachdatenbank *Web of Science* sowie neun ausgewählte Lehrbücher der Informationswissenschaft. Im *Web of Science* wurden über 324.000 Publikationen ermittelt. Im Rahmen der scientometrischen Analyse wurden Zitationen, Publikationen und Themen ausgewertet. Zusätzlich zur klassischen Auswertung der Datenbankdokumente wurde erstmals eine Registeranalyse bei den Lehrbuchdokumenten durchgeführt. Die unterschiedlichen Ergebnisse wurden zusammengefasst, um einen gemeinsamen Kanon der Informationswissenschaft bilden zu können.

Deskriptoren: Kanon, Informationswissenschaft, Informetrie, Bibliometrie, Scientometrie, statistische Analyse, Zitierhäufigkeit, Publikationsanalyse, Themenanalyse, Registeranalyse, Empirische Untersuchung.

Canonicity of Information Science

In 2013, Blaise Cronin expressed the question, whether there is such a thing as a canonicity of information science. He proposed that to determine the canonicity indexes of names and subjects of relevant textbooks have to be evaluated. Basis of the study were publications from the *Web of Science* category *Information Science & Library Science*, as well as nine chosen elementary textbooks of Information Science. The data selection from *Web of Science* contained more than 324,000 publications. During the scientometric analysis citations, publications and topics were evaluated. In addition to the classic evaluation of database records, for the first time an index analysis was performed on the textbooks. The different results were merged to create a common canonicity of Information Science.

Descriptors: Canonicity, Information Science, Informetrics, Bibliometrics, Scientometrics, Statistical analysis,

Citation frequency, Publication analysis, Topic analysis, Index analysis, Empirical study.

Le canon des sciences de l'information

En 2013, Blaise Cronin s'est posé la question de savoir s'il existait quelque chose d'équivalent à un canon des sciences de l'information. Pour déterminer le canon, il proposait d'évaluer par des moyens informétriques des registres de noms et de sujets de manuels spécialisés. Il étudia des publications de langue anglaise dans les domaines des sciences de l'information et de la bibliothéconomie extraites de la base de données spécialisée « *Web of Science* », ainsi que neuf manuels de sciences de l'information. Plus de 324 000 publications ont ainsi été identifiées dans le « *Web of Science* ». Dans le cadre de l'analyse scientométrique, l'auteur a évalué des citations, des publications et des sujets. En plus d'une analyse classique des documents de la base de données, il a mené, pour la première fois, une analyse des registres des manuels. Par la suite, les différents résultats ont été regroupés pour former un canon commun des sciences de l'information.

Descripteurs: canon, sciences de l'information, informétrie, bibliométrie, scientométrie, analyse statistique, citation, analyse de publication, analyse thématique, analyse de registres, enquête empirique.

DOI 10.1515/iwp-2015-0021

1 Einleitung

Als „Kanon“ bezeichnet man in der Literatur die Zusammenstellung aller Werke, die als wesentlicher Bestandteil einer bestimmten Fachrichtung gelten. Er stellt eine „Sammlung von als maßgeblich, musterhaft, bedeutend oder repräsentativ angesehenen Texten, Werken oder Autoren“ (Burdorf, Fasbender & Moenninghoff, 2007, S.372) dar. Als relevante Selektionskriterien für einen möglichen Kanon werden in der Informationswissen-

schaft vor allem bibliometrische Maße angesehen. Sie geben indirekt Auskunft über die Qualität von Publikationen, indem statistische Untersuchungen durchgeführt werden. Da man davon ausgeht, dass eine Publikation umso wichtiger für einen Fachbereich ist, je häufiger sie zitiert wird, gibt sie Auskunft über den wissenschaftlichen Wert einer Publikation. Quellen bibliometrischer Untersuchungen sind häufig wissenschaftliche Informationsdienste wie *Web of Science* oder *Scopus*.

Neben den bibliometrischen Analysen an Datenbanken könnten laut Cronin (2013) ebenso Register von Büchern ausgewertet werden, um zentrale Punkte der Fachrichtung herauszufiltern.

What I am proposing is that we delve into the field's "bibles" to see if information science does in fact have a recognizable canon (Cronin, 2013, S. 2190).

Ziel dieser Studie ist es, empirisch den Kanon der Informationswissenschaft zu bestimmen. Hierfür wurde zunächst eine geeignete Datenbasis ausgewählt, um auf diese einerseits die verschiedenen bekannten datenbankbasierten bibliometrischen Methoden sowie andererseits die neue Methode der Auswertung von Registern aus Lehrbüchern der Informationswissenschaft anzuwenden. Die Ergebnisse dieser Untersuchung liefern Antworten auf folgende Forschungsfragen: Welche Themen sind in der Informationswissenschaft von Bedeutung? Welche Personen haben die Informationswissenschaft vorangebracht? Welche Publikationen sind für die Informationswissenschaft am relevantesten? Die Auswertung der Daten ergibt ein Ranking der untersuchten Publikationen, Autoren und Themen.

2 Forschungsstand

Bereits seit Jahrzehnten werden Statistiken über verschiedene Wissenschaften erstellt, um daraus verschiedene Untersuchungsergebnisse zu gewinnen. Die Bibliometrie profitiert von der Möglichkeit, große Datenmengen online wie offline auszuwerten. Mittlerweile hat sie sich als etablierte Methode zur wissenschaftlichen Bewertung wissenschaftlicher Institutionen bewährt.

Die Idee der bibliometrischen Auswertung der Buchregister zur Erfassung eines Kanons wurde bisher jedoch nur sehr selten verwendet.

Anhand des *Handbook of Information Science* (Stock & Stock, 2013) stellt Cronin (2013) fest, dass man mit Hilfe der Indexeinträge im Register von Lehrbüchern einen wirklichkeitsnahen Querschnitt der Informationswissenschaft erstellen könnte. Dieser sollte sich allerdings nicht

nur über ein einziges Lehrbuch, sondern über eine größere Palette der informationswissenschaftlichen Dokumente erstrecken. Die vorliegende Studie konzentriert sich auf die Erstellung eines informationswissenschaftlichen Kanons, basierend auf informetrisch-statistischen Untersuchungen. Hierbei werden die bekannten bibliometrischen Analysen auf Datenbankdokumente sowie die Analyse von Buchregistern auf Lehrbücher angewendet. Der Kanon wird anschließend als die Menge der Top-zitierten Autoren, Top-zitierten Werken und Top-thematisierten Themen gebildet. Wir erhalten so auch eine methodologische Aussage über die Verwendbarkeit der Buchregister in einer solchen Auswertung.

3 Methoden

Gegenstand der Studie ist das Erfassen der wichtigsten informationswissenschaftlichen Personen, Themen und Publikationen durch quantitative informetrische Methoden. Zu diesem Zweck wurden Artikel aus einer Datenbank und aus Lehrbüchern ausgewertet.

3.1 Datenbank

Die Datenbasis der Artikel ergab sich aus der Kategorie *Information Science & Library Science* der Datenbank *Web of Science*. Die Kategorie umfasst Zeitschriften, welche nach ihrem Einfluss, d. h. nach der Anzahl ihrer Zitationen, sortiert wurden. Die Zeitschriften mit direktem Bezug zur Informationswissenschaft, insgesamt 13, wurden hier als Datenbasis A (siehe Anhang) ausgewählt und via Endnote exportiert. Unser Datensatz enthält rund 324.000 Artikelnachweise. Im Verlauf der Studie wurden auf die Dokumente der Datenbank ausschließlich die bekannten bibliometrischen Methoden angewendet. Die *Publikationsanalyse* bezeichnet die quantitative Analyse wissenschaftlicher Publikationen und wird so „von wissenschaftlichen Instituten oder Fachbereichen benutzt, um deren Produktivität abzubilden bzw. um – vergleichend – Produktivitätsrangordnungen zu erstellen“ (Stock, 2001). Die *Zitationsanalyse* ist eine weitere Methode, welche durch die Auswertung von Zitaten Aufschluss über das Ansehen eines Wissenschaftlers oder die Qualität einer Fachzeitschrift gibt. Konkret beschäftigt sie sich mit den Beziehungen zwischen zitierter und zitierender Publikation.

Ca. 10 bis 30 Prozent aller Zitate bestehen aus Selbstzitationen (MacRoberts & MacRoberts, 1989). Sie beschreiben die Zitierung eigener Werke in neueren Publikatio-

nen und können Einfluss auf den Rang eines Autors und dessen Arbeit in der wissenschaftlichen Community ausüben. Hier können jedoch nicht nur themenrelevante, sondern auch irrelevante Dokumente zitiert werden, wodurch der Rangplatz ebenfalls gesteigert wird (Stock, 2001). Wir haben Selbstzitate in der vorliegenden Studie identifiziert und markiert. Bei der *Themenanalyse* steht die Verteilung von Themen im Mittelpunkt. In der Regel wird hier mit Deskriptoren aus einem Thesaurus oder mit den Notationen eines Klassifikationssystems gearbeitet. Stehen diese nicht zur Verfügung, kann man auf die Titelwörter der Publikationen zurückgreifen (Milojević, Sugimoto, Yan, & Ding, 2011). Die Dokumente aus Datenbasis A wurden auf wichtige Personen, wichtige Publikationen und wichtige Themen (Titeltermine) untersucht. Um die große Datenmenge besser verwalten zu können, wurde Microsoft Access eingesetzt. Die Anwendung ermöglicht es, ohne besondere Programmierkenntnisse eine Datenbank zu erstellen, in der sich Daten unter anderem sammeln, verwalten, aufbereiten und auswerten lassen. Mittels einer Tabelle können Daten erfasst werden und dann durch sogenannte Formulare ausgebaut, beziehungsweise bearbeitet werden. Filter ermöglichen eine Abfrage innerhalb des Datensatzes.

Wichtige Personen

Zunächst mussten die unterschiedlichen Schreibweisen der Autoren, vor allem in Hinblick auf die Groß- und Kleinschreibung, berücksichtigt werden. So findet sich im *Web of Science* zum Beispiel die Schreibweise „Salton G.“, aber auch „SALTON G.“, die sich jedoch beide auf denselben Autoren beziehen. Variationen in der Zitierweise, wie *Price DD*, *Price DJD*, *Price D. J. D.*, *Price de Solla D J* usw. mussten erkannt und vereinheitlicht werden. Weitere Probleme wie Synonyme, Namensänderung, Spitznamen, Tippfehler, verschiedene linguistische Hintergründe etc. können bei der Auswertung von Autorennamen auftauchen und wurden in unserer Studie geprüft und bearbeitet. Zuletzt wurden die Autorennamen und ihre Selbstzitate im *Web of Science* ausfindig gemacht, so dass wir die Anzahl der Zitate mit und ohne Selbstzitate erhielten. So konnte das Auftreten der Autorennamen ermittelt und mittels eines Filters nach Häufigkeit sortiert werden. Besonderes Augenmerk lag auf den Top 50 Autoren und ihrer Verteilung hinsichtlich der bereinigten Zitate.

Wichtige Publikationen

Basierend auf Datenbasis A wurden nun ausgewählte bibliographische Angaben aller Artikel jener Zeitschriften als Textdatei im Format Name, Vorname, Jahr, Titel, Band, Seite aus *Web of Science* exportiert. Diese Textdatei wurde wiederum mithilfe von automatischer Sprachverarbeitung (natural language processing – kurz: NLP) durch ein Python-Programm analysiert. Die Referenzen wurden dabei als regulärer Ausdruck erfasst. Anschließend zählte das Programm jedes Vorkommen derselben Referenz und gab ein Ranking der meist zitierten Publikationen aus.

Wichtige Themen

Durch ein weiteres Python-Programm und dem Modul NLTK (Natural-Language-Toolkit-Processing) wurde eine Sprachanalyse ausgeführt, welche alle Sätze der Publikationen aus Datenbasis A in ihre Bestandteile zerlegte. Anschließend entfernte das Programm die Stoppwörter der standardisierten Stoppwortliste des Moduls und führte Stemming durch, um verschiedene morphologische Varianten eines Wortes auf ihren gemeinsamen Wortstamm zurückzuführen. Als Ergebnis gab das Programm eine csv-Datei aus, welche die Wortstämme mit jeweiliger Häufigkeit enthält, sodass die Häufigkeitsverteilung ermittelt werden konnte.

3.2 Lehrbücher

Die Auswahl der Datenbasis der Lehrbücher wurde von der Forschungsgruppe anhand eigener Kriterien durchgeführt. Diese sind beispielsweise der inhaltliche und zeitliche Aspekt; die Dokumente sollten detailliert Aspekte der Informationswissenschaft behandeln und in den letzten fünf bis zehn Jahren veröffentlicht worden sein.

Time, place, culture, and language all play a part in determining which concepts and theories matter most to us, which works are deemed paradigmatic, which authors are granted canonical status (Cronin, 2013, S. 2190).

Bei der Auswahl der Werke zeigte sich, dass es kaum deutschsprachige Bücher mit brauchbaren Personen- und Sachregistern gibt, welche den oben genannten Kriterien entsprechen. So reduzierte sich die Auswahl auf englischsprachige Literatur. Die endgültige Datenbasis B (siehe Anhang) wurde ebenfalls im Hinblick auf wichtige Personen, wichtige Publikationen und wichtige Themen analysiert. Im Gegensatz zur Auswertung von Datenbasis

A wurden hier bibliometrischen Analysen an Registern durchgeführt. Verfügte jedoch ein Lehrbuch aus der Datenbasis über kein entsprechendes Personenregister, wurde auf die Referenzen zurückgegriffen.

Wichtige Personen

Ein Personenregister ist eine Liste von Referenzen bezogen auf Autoren, denen Inhalte des Dokumentes zugeschrieben werden. Hieraus lässt sich eine gewisse Relevanz der Autoren anhand der Anzahl der zu ihnen gelisteten Seiten schließen. Bei Büchern, die über kein Personenregister verfügen, wurde pro Referenz in der Bibliographie eine Seite gezählt. Letzteres Vorgehen ist durch die Vielfalt der Autoren unübersichtlicher als die erste Methode, weil jeder Autor aus jeder Referenz einzeln identifiziert werden musste. So wurde zu jedem Lehrbuch aus Datenbasis B eine Liste der Top 80 Autoren absteigend sortiert nach ihrer Seitenzuordnung angelegt. Zuletzt wurde die Vereinigungsmenge der erstellten Listen gebildet und absteigend nach ihrer Seitenzuordnung sortiert.

Wichtige Publikationen

Die ausgewählten neun Lehrbücher wurden auf die darin enthaltenen Literaturangaben untersucht, wobei unterschiedliche Zitierweisen zu beachten waren. Diese wurden zur Weiterverarbeitung vereinheitlicht und in ein Textdokument übernommen. Hierbei wurde ein Python-Programm verwendet, welches die Referenzen aus den Literaturverzeichnissen als reguläre Ausdrücke auffasst.

Mit Hilfe der Tool-Box *bibexcel* (siehe Abb. 1) wurden anschließend zeilenweise die csv-Dateien eingelesen, darin enthaltene Dubletten identifiziert und vereinheitlicht. Die entstandene Auflistung von Autoren mit Publikationsjahr wurde nach Häufigkeit sortiert. Durch Autor und Publikationsjahr wurde anschließend die entsprechende Publikation im Literaturverzeichnis identifiziert und so ein Publikationsranking der Top 30 erstellt.

Wichtige Themen

Jedes Buch aus Datenbasis B wurde zunächst einzeln betrachtet: Mit Hilfe der Sachregister wurden die Terme ermittelt, die auf mehr als eine Seite verwiesen und anschließend nach Häufigkeit sortiert. Um auf einen gemeinsamen Kanon zu kommen, wurden die Term-Listen der einzelnen Bücher nun zusammengeführt und zunächst nach Themenkomplexen strukturiert. Dabei wurden die zugeordneten Seitenzahlen addiert. Durch die Sortierung nach Häufigkeit konnten die Top 30 der häufigsten Themen ermittelt werden.

4 Ergebnisse

4.1 Wichtige Personen

Die Zitatanalysen im *Web of Science* ergaben, dass E. Garfield, der als Mitbegründer der empirischen Informationswissenschaft gilt, am häufigsten zitiert wurde. Er vereint 2.497 Zitierungen auf sich. Platz 2 der Auswertung belegt L. Leydesdorff mit 2.206 Zitationen. Er ist Professor der Universität von Amsterdam im Bereich *Communication*

Abb. 1: Tool-Box *bibexcel*.

Abb. 2: Top 30 Autoren: meistzitierte informationswissenschaftliche Personen im Web of Science.

Abb. 3: Top 30 Autoren: meistzitierte Personen in informationswissenschaftlichen Lehrbüchern.

and Innovation in the Dynamics of Science und Szientometriker. Dicht hinter ihm in der Auswertung liegen G. Salton, seine Arbeit mit SMART, System for the Me-

chanical Analysis and Retrieval of Text, ist besonders erwähnenswert, L. Egghe, ein weiterer Szientometriker, der als Bibliotheksdirektor an der Universität Hasselt arbeitet und an der Universität von Antwerpen lehrt. Auf dem 5. Platz des Ranking liegt W. Glänzel, dessen wissenschaftlicher Schwerpunkt die Bibliometrie ist.

An der Spitze der meist zitierten Autoren in den Lehrbüchern steht W.G. Stock, Leiter der Abteilung für Informationswissenschaft an der Heinrich-Heine-Universität Düsseldorf. Zu seinen Forschungsthemen gehören unter anderem *informationelle Städte*, *Power Tags* und *Informationsmärkte*. G. Salton taucht auch in dieser Auswertung unter den Top 3 auf, was seinen enormen Einfluss untermauert. Der Spezialist für *Information Storage* und *Retrieval-Systeme* F.W. Lancaster war Professor für *Library and Information Science* an der Universität von Illinois. K. Sparck-Jones beschäftigte sich unter anderem mit *Information Retrieval* und *Computational Linguistics*, aber auch mit *künstlicher Intelligenz*. Den 5. Platz der Auswertung der Lehrbücher belegt W.B. Croft, ein Professor der Schule für *Computerwissenschaft* der Universität von Massachusetts in Amherst. Er ist Gründer und Direktor des dortigen *Center for Intelligent Information Retrieval*.

Es fällt auf, dass sich unter den Top-zitierten Autoren im *Web of Science* viele Informatiker befinden. Dies lässt sich womöglich auf ihr Wissen um eine gute Platzierung in solch einem Ranking zurückführen. Bei einem Vergleich mit der Ergebnisliste der Lehrbuch-Analysen

Abb. 4: Top 25 Publikationen: meistzitierte informationswissenschaftliche Publikationen im Web of Science.

Abb. 5: Top 30 Publikationen: meistzitierte informationswissenschaftliche Publikationen in Lehrbüchern.

ist festzustellen, dass diese Personengruppe eine wesentlich kleinere Rolle im Gesamtbild spielt. Diese Tatsache wurde jedoch nicht näher untersucht, sollte aber in künftigen Studien eingehend analysiert werden.

Die Autoren der beiden Ergebnislisten sind ausschlaggebend für ihren Teilbereich – die wichtigsten Autoren in Datenbasis A für allgemeine Dokumente der Informationswissenschaft, die wichtigsten Autoren in Datenbasis B für einführende Lehrbücher der Informationswissenschaft.

4.2 Wichtige Publikationen

Die Auswertung der Publikationen des *Web of Science* ergibt eine eindeutige *Power-Law-Verteilung*; *An Index to Quantify an Individual's Scientific Research Output* (Hirsch, J. E. mit 550 Zitierungen) steht auf dem ersten Platz. Den zweiten Platz belegt *Little Science, Big Science* (de Solla Price, D. J. mit 335 Zitierungen), und *Information Retrieval* (Van Rijsbergen, C. J. mit 304 Zitierungen) liegt auf dem dritten Platz. Auffallend ist wie bei den Personen, dass es sich bei den ersten beiden Publikationen um den

Abb. 6: Top 30 Themen: meistzitierte informationswissenschaftliche Themen im Web of Science.

Abb. 7: Top 30 wichtigste Themen: meistzitierte informationswissenschaftliche Themen in Lehrbüchern.

Themenbereich der Informetrie bzw. Szientometrie handelt.

Die Publikationen, die in den Lehrbüchern am häufigsten genannt wurden, sind ausschlaggebend für einführende Themengebiete in die Informationswissenschaft. Besonders hervorzuheben sind hier die Top 3: *Introduction to Modern Information Retrieval* (Salton, G. mit 16 Nennungen), *Information Retrieval* (Van Rijsbergen, C.J. mit 13 Nennungen) sowie *Indexing and Abstracting in Theory and Practice* (Lancaster, F.W. mit 11 Nennungen). Diese drei Publikationen befassen sich mit dem Einstieg in die Informationswissenschaft: dem Suchen und Finden von relevanten Informationen.

4.3 Wichtige Themen

Der Top-Term *information* erreicht 5.295 Nennungen im *Web of Science* und tritt am häufigsten im Zusammenhang mit *information retrieval*, *information science* und *information technology* auf. Diese Themenkomplexe spielen in der Informationswissenschaft eine zentrale Rolle. An zweiter Stelle steht der Term *research* mit einer Häufigkeit von 2.070. Obwohl der Term in den untersuchten Dokumenten häufig in einem allgemeinen Kontext auftaucht, so handelt es sich hier um eine Schlüsselkompetenz der Informationswissenschaft, unter die Recherchestrategien und informetrische Analysen fallen. Einige Terme, die in dieser Ergebnisliste vorkommen, stellen keineswegs thematische Spezifikation der Informationswissenschaft dar; dazu zählen beispielsweise *study*, *model*, und *public*. Sie wurden jedoch nicht herausgefiltert,

da die standardisierte Stoppwortliste des Moduls NLTK angewandt wurde, um ein objektives Ergebnis zu schaffen. Die Begriffe sind Bestandteil einer jeden Wissenschaft und konstituieren somit auch die Informationswissenschaft.

Das Resultat der Auswertung der Lehrbücher enthält viele Terme, die thematisch aufeinander aufbauen, so sind beispielsweise *retrieval systems and models* und *indexing* Anwendungsgebiete des *information retrieval*. *Image* und *multimedia retrieval* stehen für neuere Retrieval-Möglichkeiten. Die Terme *classification*, *thesaurus* und *folksonomy* bilden den Themenbereich der Wissensordnungen ab. Andere Begriffe, wie *internet*, *term* und *computer*, sind als Oberbegriffe der Thematik zu bewerten. Besonderes Augenmerk liegt auf dem Term *user*, welcher einen wichtigen Bestandteil des Web 2.0 bildet. Ebenfalls ist *user* ein zentraler Forschungsgegenstand der Informationswissenschaft, so beispielsweise in Nutzerforschung und in informetrischen Analysen.

Um einen Kanon der Informationswissenschaft basierend auf den wichtigsten Themen zu erkennen, werden die Ergebnisse der Auswertung von Datenbasis A und Datenbasis B miteinander verglichen. Sowohl *information* als auch *retrieval* sind in beiden Ergebnislisten vorhanden. Auch der Term *index* bzw. *indexing* gehört zu den wichtigsten in beiden Auswertungen. Bei genauerem Hinsehen erkennt man thematische Unterschiede, die sich in der Genauigkeit und Qualität stark unterscheiden. Viele Begriffe, die sich im Kanon von Datenbasis B finden lassen, beschreiben konkrete Werkzeuge des Information Retrievals: *retrieval*, *indexing*, *query*, *metadaten*, *citation*, um nur eine kleine Auswahl zu erwähnen, sowie bestimmte

Nomenclature 635–645, 841	(NAICS) 666–668
CAS Registry File 639–642	Notation 647–650, 841
chemical structure 641	hierarchical 647–650
chronological relation 642–644	hierarchical-sequential 649–650
compound 637–638	sequential 649–650
conflation 639–642	NUTS see Nomenclature des unités territoriales statistiques
controlled vocabulary 635–644	
faceted 713–715	
gen-identity 642–644	OBAR see Ohio Bar Automated Research
homonym 638–639	Object 523
disambiguation 638–639	Ofness 522–523, 616, 768, 841
homonymy 638–639	aboutness 522

Abb. 8: Sachregister (Ausschnitt): Stock, W. G. & Stock, M. (2013) *Handbook of Information Science*. Berlin, Boston, MA: De Gruyter Saur.

Q.5 Index of Names

Abbasi, A. 436, 440	Baca, M. 593–594, 596
Abdullah, A. 84–85	Baccianella, S. 438, 440
Abid, M. 386–387	Backstrom, L. 396–398
Abrahamsen, K.T. 602, 606	Bacon, F. 38–42, 46–47, 508
Acharya, A. 345, 358	Baeza-Yates, R.A. 5, 16, 131, 134–135, 139, 157, 162, 165, 285, 287, 299, 420–421, 489, 495
Adamic, L.A. 385–387	Bagozzi, R.P. 485, 495
Adams, D.A. 485, 495	Bahlmann, A.R. 471, 478
Adar, E. 365	Bakker, E.M. 326
Adomavicius, G. 417, 422	Balabanovic, M. 417, 422
Aesop 27, 523	

Abb. 9: Personenregister (Ausschnitt): Stock, W. G. & Stock, M. (2013) *Handbook of Information Science*. Berlin, Boston, MA: De Gruyter Saur.

Modelle und Methoden, die ebenfalls auftauchen (*folksonomy*, *KOS*, *thesaurus*, *classification* usw.). Der Begriff *information science* erscheint auf den ersten Blick irreführend, da das Thema der Untersuchung der Kanon der Informationswissenschaft ist. Allerdings handelt es sich bei Datenbasis B, um eine Auswahl von Lehrbüchern, daher erscheint der Begriff unter den Top 30 als gerechtfertigt. Denn der Begriff dient in den Lehrbüchern als Einführung und Definition in die eigentliche Thematik.

5 Diskussion

In der vorliegenden Untersuchung wurde der Kanon der Informationswissenschaft als Forschungsfrage formuliert. Zur Beantwortung wurden verschiedene bibliometrische Maße auf ausgewählte Publikationen angewandt – einerseits auf die Indexate im *Web of Science* und die darin befindlichen Referenzen, andererseits auf die Personenbeziehungsweise Sachregister ausgewählter Lehrbücher. Bei der Sichtung der Ergebnisse ist festzuhalten, dass diese unterschiedlich zu bewerten sind, weil die unterschiedlichen Datenbasen zu (teilweise) unterschiedlichen Ergebnissen führen. Bei den wichtigsten Personen der Informationswissenschaft dominiert E. Garfield mit 2.497 Zitierungen in der Rangfolge der *Web of Science*-Auswertung, wohingegen W. G. Stock aufgrund von 129 Selbst-

zitationen bei den ausgewählten Lehrbücher den höchsten Wert erzielt. Bei den wichtigsten Publikationen der Informationswissenschaft fällt auf, dass die Autoren G. Salton und E. Garfield mit vielen Werken in den beiden Rankings auftauchen. Dies spricht für einen enormen Einfluss der beiden auf die Informationswissenschaft und auf deren „Kanonität“. Die wichtigsten Themen der Informationswissenschaft umfassen sieben Bereiche: Zitationen, Datenbasen, Auswertung, Verzeichnis, Information, Wissen und Wiederauffinden.

Wie ist die neue Methode der informetrischen Analyse von Buchregistern zu bewerten? Der Vorteil im Auswerten von Buchregistern liegt in der kompakten Auflistung der benötigten Informationen. Ein Personenregister listet alle zitierten Autoren auf, ein Sachregister alle Themenkomplexe. Im Gegensatz hierzu steht die herkömmlich Auswertung der Indexate und der darin enthaltenen Referenzen. Diese klassische Methode erwies sich in der vorliegenden Studie durch die fehlende Struktur als zeitaufwendiger und komplexer.

Die Untersuchung ist durch die verschiedenen Analysevorgänge beschränkt, sodass zwar quantitativ ein Kanon gebildet werden konnte, nicht jedoch qualitativ. Publikations-, Zitat- und Themenanalysen beschäftigen sich mit dem reinen Vorkommen und sagen nicht direkt aus, wie wichtig etwas ist. Eine weitere Begrenzung der Untersuchung erfolgte durch die Wahl der Publikationen, es wurden nur Dokumente aus der Datenbank *Web of Science*, nicht jedoch zum Beispiel *Scopus*, *Google Scholar*, *Mendeley* oder *CiteULike*, untersucht. Die Datenbasis der neun Lehrbücher wurde im Ermessen der Forschungsgruppe gewählt. Hier spielten Aktualität und Inhalt eine Rolle.

Beim Auswerten der Buchregister trafen wir auf ein großes Problemfeld: die Personen- und Sachregister wiesen keine einheitliche Struktur auf, sodass teilweise allein stehende Terme, aber auch Terme mit zugehörigem Oberbegriff gelistet waren. Darüber hinaus tauchten in einigen Sachregistern zusätzlich zu den Termen auch Personennamen auf. Je nach Sorgfalt, mit der die Buchregister erstellt wurden, eignet sich die Methode ihrer Auswertung jedoch hervorragend. Das Sachregister in *Introduction to Information Science and Technology* (herausgegeben von Davis, C. und Shaw, D.) listet beispielsweise zu jedem informationswissenschaftlichen Thema im Buch alle Seitenangaben auf. Es ist gründlich und gut organisiert, sodass Themen leicht zu finden sind. Das *Handbook of Information Science* (von Stock, W. G. und Stock, M.) enthält ebenfalls ein im Sinne der internationalen Norm (ISO 999:1996) erstelltes Sachregister, welches dem von Davis/Shaw gleicht. Darüber hinaus ent-

hält es jedoch noch ein Personenregister. Dieses gibt einen strukturierten Überblick auf die Inhalte, welche anderen Autoren zugesprochen werden.

Zusammenfassend können wir feststellen, dass Cronin (2013) die Informationswissenschaft um zwei Aspekte bereichert hat: (1.) Es ist ein sinnvolles Unternehmen, den Kanon der Informationswissenschaft zu bestimmen; (2.) hierfür (aber auch für andere empirische Fragestellungen) eignet sich die informatrische Auswertung (gut strukturierter und sorgfältig erarbeiteter) Register in Lehrbüchern.

Anhang

Datenbasis A: Zeitschriften

Annual Review of Information Science and Technology
ASLIB Proceedings
Government Information Quarterly
Information Processing & Management
Information System Research
International Journal of Geographical Information Technology
Journal of Documentation
Journal of Information Science
Journal of Informetrics
Journal of the American Society for Information Science and Technology
Library and Information Science Research
MIS Quarterly
Scientometrics

Datenbasis B: Lehrbücher

Baeza-Yates, R. & Ribeiro-Neto, B. Modern Information Retrieval: the Concepts & Technology behind Search (2011)
Chowdhury, G. Introduction to Modern Information Retrieval (2003)
Chu, H. Informational Representation and Retrieval in the Digital Age (2010)
Davis, C. & Shaw, D., Eds. Introduction to Information Science and Technology (2011)
Lancaster, F. W. Indexing and Abstracting in Theory & Practice (2003)
Manning, C. D., Raghavan, P., & Schütze, H. Introduction to Information Retrieval (2008)

Norton, M. J. Introductory Concepts in Information Science (2010)
Stock, W. G. & Stock, M. Handbook of Information Science (2013)
Vickery, B. & Vickery, A. Information Science in Theory & Practice (2004)

Danksagung: Für das Sammeln und Auswerten der Messwerte waren folgende Personen verantwortlich: Audrone Lange, Nina Badura, Mercedes Siewitz, Regina Kasakowskij, Laura Konrath (Gruppenleitung: wichtige Themen – Lehrbücher), Christina Rypalla, Daniel Knapp, Roman Dammer, Marius von der Kohle (Gruppenleitung: wichtige Publikationen – Lehrbücher), Alexandra Karina Roman, Thorsten Brückner, Raphael M. M. E. Katschke (Gruppenleitung: wichtige Personen – Lehrbücher), Julia Fischer, Kevin Pochwyt, Kristina Ort, Valentin Schöps, Marcel Brockmann (Gruppenleitung: wichtige Themen – Publikationsanalysen), Jaqueline Betz, Katharina Gersbersdorf, Katharina Müller, Lina Schorn, Sophie Rosenblatt, Nora Lindenberg (Gruppenleitung: wichtige Personen – Zitationsanalysen), Daniel Lüke, Maximilian Wiese, Sascha Dreesbach, Veronika Idasiak (Gruppenleitung: wichtige Publikationen – Zitationsanalysen). Die Projektleitung und die redaktionelle Arbeit oblagen Denise Ruhrberg.

Literatur

Burdorf, D., Fasbender, C. & Moenninghoff, B. (2007). Kanon. In *Metzler Lexikon Literatur – Begriffe und Definitionen* (S. 372–373). Stuttgart, Weimar: Metzler.
Cronin, B. (2013). Canonicity. *Journal of the American Society for Information Science and Technology*, 64(11), 2189–2190.
ISO 999 (1996). *Information and Documentation. Guidelines for the Content, Organization and Presentation of Indexes*. Genf: International Organization for Standardization.
MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40(5), 342–349.
Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science. Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933–1953.
Stock, W. G. (2001). *Publikation und Zitat: Die problematische Basis empirischer Wissenschaftsforschung*. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, 29). Köln: FH Köln.
Stock, W. G., & Stock, M. (2013). *Handbook of Information Science*. Berlin, Boston, MA: De Gruyter Saur.

(Foto: W. Stock)

Das Autorenteam

Mathilde B. Friedländer steht für eine Studentengruppe der Heinrich-Heine-Universität Düsseldorf. Wir studieren *Informationswissenschaft und Sprachtechnologie* oder *In-*

formationswissenschaft als Ergänzungsfach im Bachelorstudium. Die Sprecherin von Mathilde B. Friedländer ist Denise Ruhrberg (denise.ruhrberg@uni-duesseldorf.de) – (erste Reihe, Mitte rechts im Bild).