

Wissenschaftsindikatoren im Zeitalter digitaler Wissenschaft¹

von Rafael Ball

Digitale Bibliotheken entwickeln sich zum Standard der Literatur und Informationsversorgung in Wissenschaft und Forschung. Printmedien werden besonders im STM-Segment zu einem Nischenprodukt. Dabei sind die Angebote von Digital Libraries Bestandteil eines ganzheitlichen digitalen Ansatzes, wonach die elektronische Informations- und Literaturversorgung integraler Bestandteil von E-Science (Enhanced Science) oder Cyberinfrastruktur darstellt. Hierbei verschmelzen dann Produktion, Diskussion, Distribution und Rezeption der wissenschaftlichen Inhalte auf einer einzigen digitalen Plattform. Damit sind dann nicht nur die Literatur- und Informationsversorgung (Digital Libraries), sondern auch die Wissenschaft selbst digital geworden. Diese dramatische Veränderung in der Wissenschaftskommunikation hat direkte Auswirkungen auf die Messung der Wissenschaftskommunikation. Bisherige Systeme der Wissenschaftsvermessung basieren hauptsächlich auf bibliometrischen Analysen in international anerkannten Datenbanken. Die Digitalisierung der Wissenschaft in ihrem kompletten Lebenszyklus erfordert und ermöglicht zugleich den Nachweis von Output und Rezeption durch neue bibliometrische Formen, etwa der Webometrie (Webmetrics). Im vorliegenden Paper haben wir hierzu Analysen durchgeführt und stellen eine Abschätzung vor, wie sich der Anteil von webometrisch erfassbarer und zugänglicher wissenschaftlicher Literatur im Vergleich zu Literatur, die mit den Standardsystemen nachgewiesen werden kann im Laufe der letzten Jahre verändert hat.

Science indicators in the age of digital science

The use of digital libraries is becoming the norm for the provision of literature and information in science and research. Print media are becoming a niche product in the STM sector in particular. What digital libraries have to offer forms part of a comprehensive digital approach, whereby the provision of electronic information and literature represents an integral part of e-science (enhanced science) and cyberinfrastructure. Here, the production, discussion, distribution and reception of scientific content merge together on one single digital platform. This means that it is not just the provision of literature and information that is digital (digital libraries) but that science itself has also „gone digital“. This dramatic change in scholarly communication has direct implications for the way that scholarly communication is measured, in other words on the evaluation of scientific output. Systems used up until now to survey science have been based predominantly on bibliometric analyses. The digitalisation of science in its complete life cycle, the increasing use and acceptance of document repositories, institute servers and other forms of electronic publication within the framework of e-science demand and simultaneously allow the assessment of output and reception using new forms of bibliometrics such as webmetrics. We have carried out such analyses and in the following paper, we present an evaluation of how the proportion of readily accessible scientific literature that can be measured webmetrically has changed over the course of the past few years compared to literature that can be accounted for using standard systems.

Les indicateurs de science dans l'âge de la science numérique

La disponibilité des bibliothèques numériques viennent progressivement de la norme pour la provision des littératures et des informations de la science et la recherche. Des médias imprimés sont alors un produit spécialisé notamment dans le secteur STM. Les bibliothèques numériques offrent des services numériques compréhensifs et la mise à disposition des informations et littératures électroniques fait partie intégrante d'e-science (enhanced science) et de la cyberinfrastructure. La production, la discussion, la distribution, et la réception du contenu scientifique forment une plate-forme numérique unique. Ce développement ne signifie seulement que la provision des littératures et des informations est ainsi numérique (des bibliothèques numériques) mais aussi que la science même devient numérique. Cette transformation dramatique dans la communication savante a des conséquences directes pour la mensuration de la communication savante, autrement dit pour l'évaluation de la production de la science. La numérisation du cycle de vie de la science, l'utilisation et l'adoption en hausse des référentiels de documents, des serveurs dans les institutions et des autres formes de la publication électronique dans le cadre d'e-science exigent et permettent à la fois l'évaluation de la production et de la réponse par des nouvelles méthodes bibliométriques, par exemple par la webométrie. Dans cet article nous menions ces analyses et essayer de vous expliquer comment la proportion des informations scientifiques disponibles ce qu'on peut évaluer par la webométrie a changé au cours des années dernières en comparaison avec des littératures ce qu'on peut examiner par les systèmes standards.

1 Einleitung

Die Bereitstellung und Nutzung digitaler Bibliotheken entwickelt sich allmählich zum Standard der Literatur und Informationsversorgung in Wissenschaft und Forschung. Ganzen Disziplinen genügt oftmals die verfügbare digitale Information, Printmedien werden besonders im STM-Segment zu einem Nischenprodukt.² Digitale Texte können beliebig eingebaut, kopiert und nachgenutzt werden, die Verlinkung zwischen Metadaten und Volltexten bringt weitere Nutzungsvorteile. Dabei sind die Angebote von Digital Libraries Bestandteil eines ganzheitlichen digitalen Ansatzes, wonach die elektronische Informations- und Literaturversorgung integraler Bestandteil von E-Science (Enhanced Science, Terminologie in Europa) oder Cyberinfrastruktur³ (Terminologie in USA) oder e science, grid application (Asia-Pacific)⁴ darstellt. Hierbei verschmelzen dann Produktion, Diskussion, Distribution und Rezeption der wissenschaftlichen Inhalte auf einer einzigen digitalen Plattform⁵. Damit ist dann nicht nur die Literatur- und Informationsversorgung (Digital Libraries), sondern auch die Wissenschaft selbst digital geworden. Diese dramatische Veränderung in der Wissenschaftskommunikation hat direkte Auswirkungen auf die Messung der Wissenschaftskommunikation, also auf die Evaluation von wissenschaftlichem Output.

Bisherige Systeme der Wissenschaftsvermessung basieren hauptsächlich auf bibliometrischen Analysen, d.h. der Quantifizierung des Outputs und dessen Rezeption (Zitierhäufigkeit). Basis dafür sind insbesondere im STM-Bereich die international anerkannten Datenbanken des ISI⁶ (Thomson Scientific) insbesondere der Science Citation Index (SCI), oder vielleicht zukünftig das Konkurrenzprodukt SCOPUS⁷ des Wissenschaftskonzerns Reed Elsevier. Die Digitalisierung der Wissenschaft in ihrem kompletten Lebenszyklus, die zunehmende Nutzung und Akzeptanz von Dokumentenrepositorien, Institutsservern und anderen elektronischen Publikationsformen im Rahmen von E-Science erfor-

¹ Vortrag gehalten auf der internationalen Konferenz „Digital Libraries and Semantic WEB“ in Bangalore, 19.-21.2.2007.

² Siehe z.B.: Jones, Maggie: Archiving E-Journals Consultancy- Final Report. Report Commissioned by the Joint Information Systems Committee (JISC) October 2003.

<http://www.jisc.org/02-summer/refereed.html>,

<http://www.eso.org/gen-fac/libraries/lisa4/Storey2.pdf>

³ hier der Report der National Science Foundation (NSF) http://www.communitytechnology.org/nsf_ci_report/report.pdf

⁴ http://www.gridatlasia.net/component/option,com_frontpage/Itemid,1/ ;

<http://www.gridtoday.com/gridtoday.html>

⁵ Zur Definition siehe: „What is Cyberinfrastructure?“ Report der NSF http://www.communitytechnology.org/nsf_ci_report/appendices.pdf

⁶ <http://scientific.thomson.com/products/wos/>

⁷ <http://www.scopus.com/scopus/home.url>

dern und ermöglichen zugleich den Nachweis von Output und Rezeption durch neue bibliometrische Formen, etwa der Webometrie (Webmetrics). Diese kann die bisherigen traditionellen Formen der Output- und Resonanzmessung ergänzen und erweitern, hat aber auch ihre Grenzen.

2 Zum Verhältnis von Bibliometrie, Szientometrie, Informetrie und Webometrie

Obwohl der Begriff Bibliometrie am häufigsten in jener Literatur erscheint, die sich mit der Anwendung mathematisch-statistischer Methoden bei der Quantifizierung geschriebener Kommunikation befasst, benutzt eine immer größere Anzahl von Autoren auch die Begriffe Szientometrie oder Informetrie. Als Folge der Veröffentlichung wissenschaftlicher Arbeiten in elektronischer Form und deren ausschließlichem Zugang über das Netz, verbreitet sich seit der Mitte der 90er Jahre des 20. Jahrhunderts zunehmend der Begriff Webometrie. Der Begriff **Bibliometrie** wurde von Alan Pritchard geprägt, vor allem um den irreführenden Begriff „statistische Bibliographie“, den E.W. Hulm bereits 1922 einführte, zu ersetzen⁸. Pritchard bestimmte den Begriff als Anwendung mathematischer und statistischer Methoden auf Bücher und andere Medien der wissenschaftlichen Kommunikation („...application of mathematical and statistical methods to books and other media of communication“). Narin und Moll griffen Prichards Bibliometrie-Definition auf und bestimm-

ten sie als Quantifizierung von Prozessen geschriebener Kommunikation. Diese Definition verleiht der Bibliometrie den Status eines Teilgebietes der Szientometrie, Informetrie und Webometrie.⁹

Über die Bestimmung des Begriffs Bibliometrie gibt Osareh einen ausführlichen Überblick.¹⁰

Debackere und Glänzel bezeichnen die Bibliometrie als Wissenschaftsfeld, während sie „evaluative bibliometrics“ und „science mapping“ für Subdisziplinen halten, die für

des 20. Jahrhunderts stetig zunimmt und benutzt wird, um Aspekte der Wissenschaft kennen zu lernen: das Wachstum, die Struktur, gegenseitige Beziehungen und die Produktivität. Die Szientometrie ist eng verbunden mit der Bibliometrie und Informetrie und überschneidet sich teilweise mit ihrer Definition.¹⁴

Der Begriff der **Informetrie** wurde erstmals von O. Nacke als Teilbereich der Informationswissenschaft im Jahre 1979 ausgelegt und als Anwendung mathemati-

Preprint	Autoren-Vertrag	Peer Review	Erstveröffentlichung	Zweitveröffentlichung (Postprint)
Instituts- oder fachbezogene Preprint-server, ohne Qualitäts-sicherung.	Zu diesem Zeitpunkt entscheidet der Autor, welche Rechte er/sie abtreten will.	Qualitäts-sicherung	Print/online in traditioneller Zeitschrift (Zugang nur bei kostenpflichtiger Subskription)	Elektronisch in institutsbezogenen Repositorium (Institutional Repository).
			Open Access Journal mit diversen Geschäftsmodellen zur Finanzierung	

Abbildung 1: Lebenszyklus einer Publikation

die Wissenschaftspolitik von grundlegender Bedeutung seien.¹¹

Der Begriff der Szientometrie bezieht sich auf die Anwendung quantitativer Methoden in der Analyse der Wissenschaft als Informationsprozess. Pritchard zufolge benutzten die russischen Autoren Dobrov und Korennoi den Begriff zum ersten Mal, während E. Garfield von der Wichtigkeit der Zitatanalysen ausgehend, Derek de Solla Price als Begründer der Wissenschaftswissenschaft bezeichnet.¹²

Obwohl der Unterschied dieser beiden Termini ziemlich deutlich erscheint, kommt es oft zu einer synonymen Verwendung in der Literatur. Eine akzeptierbare Erläuterung und inhaltliche Abgrenzung dieser beiden Bereiche bieten Brown et al.¹³ Die Bibliometrie untersucht demnach Bücher, Zeitschriften und andere Informationsquellen als formale Dokumente um Bibliothekssammlungen und Dienstleistungen quantitativ zu analysieren und die wissenschaftliche Dokumentation, Informations- und Kommunikationsaktivitäten zu verbessern. Die Szientometrie hingegen analysiert quantitative Aspekte der Entstehung, Verbreitung und Benutzung wissenschaftlicher Informationen um zu einem besseren Verständnis der Mechanismen wissenschaftlicher Forschung als sozialer Aktivität beizutragen. Ziel der Bibliometrie ist es demnach, mittels quantitativer Methoden Eigenschaften der Wissenschaft und wissenschaftlicher Forschungen zu bestimmen. Hood und Wilson weisen darauf hin, dass die Popularität des Begriffs Szientometrie seit ihrer Prägung in den 60er Jahren

schwerer Methoden auf die Sachverhalte des Informationswesens bestimmt.¹⁵ In der ehemaligen Sowjetunion initiierte FID (*Federation Internationale de la Documentation*) die Benutzung des Ausdrucks Informetrie als Gattungsbezeichnung für Bibliometrie und Szientometrie.

Den Begriff Webometrie riefen Almind und Ingwersen ins Leben und bestimmten ihn als Anwendung informetrischer Methoden auf das World Wide Web (WWW).¹⁶ Da das Netz heute zu einer selbstverständlichen Informationsquelle geworden ist, weisen Björneborn und Ingwersen auf die Wichtigkeit der Untersuchung des Webs hin.¹⁷

Einer der produktivsten Autoren auf dem Gebiet der Webometrie, Mike Thelwall, behandelt in einer neueren Arbeit die Rolle akademischer Webressourcen in der wissenschaftlichen Kommunikation.¹⁸ Das Web kann unter verschiedenen Aspekten untersucht werden, z. B. im Hinblick auf verfügbare Informationsquellen oder als Medium der wissenschaftlichen Kommunikation. Bislang jedoch werden hauptsächlich bibliometrische Prinzipien angewendet, allen voran Zitatanalysen aufgrund von Verlinkungen.

Ein weiterer Grund ist in der Tatsache zu suchen, dass sich verschiedene Experten mit der Problematik der Publikationsauswertung befassen: Bibliothekare und Informationsspezialisten um ihre Bestände auszuwerten und Erwerbungsentscheidungen zu treffen, Wissenschaftler und Forscher, um ihre eigene wissenschaftliche Produktion auszuwerten und ihren Einfluss in der

⁸ Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 25(4), 348-349.

⁹ Narin, F., & Moll, J. K. (1977). Bibliometrics. *Annual Review of Information Science and Technology*, 12, 35-58.

¹⁰ Osareh, F. (1996a). Bibliometrics, citation analysis and co-citation analysis. A review of literature I. *Libri*, 46(3), 149-158.

¹¹ Debackere, K., & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*, 59(2), 253-276.

¹² Garfield, E. (1998). From Citation Indexes to Informetrics: Is the Tail Now Wagging the Dog? *Libri*, 48, 67-80.

¹³ Brown, T., Glänzel, W., & Schubert, A. (1985). *Scientometric indicators: A 32-country comparative evaluation on publishing performance and citation impact*. Singapore-Philadelphia: World Scientific.

¹⁴ Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2), 291-314.

¹⁵ Nacke, O. (1979). Informetrie: ein neuer Name für eine neue Disziplin. *Nachrichten für Dokumentation*, 30(6), 212-226.

¹⁶ Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: methodological approaches to 'Webometrics'. *Journal of Documentation*, 53(4), 404-426.

¹⁷ Björneborn, L., & Ingwersen, P. (2001). Perspective of webometrics. *Scientometrics*, 50(1), 65-82.

¹⁸ Thelwall, M., & Gareth, H. (2004). Do the web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology*, 55(2), 149-159.

Community zu messen und Vergleiche anzustellen, Geldgeber und Unterhaltsträger, um Entscheidungsunterstützung zu erhalten, und Bibliometriker, Szientometriker und Informatiker als Experten, die sich mit der theoretischen und pragmatischen Entwicklung dieser Subdisziplin befassen.

3 Der Lebenszyklus einer Publikation

Der Lebenszyklus einer Publikation ist in Abbildung 1 dargestellt. Traditionelle Veröffentlichungswege gehen über das Peer Review eines Verlages, das Paper wird dann vom Editor angenommen und im Journal publiziert. Bereits seit vielen Jahren gibt es einzelne Disziplinen, wie etwa die Hochenergiephysik, die ihre Veröffentlichungen vorab auf einem Preprint Server ablegen.¹⁹ Doch dies hat noch keine fundamentalen Auswirkungen auf die Wissenschaftskommunikation und ihre Messung. Erst die Möglichkeit von „Green Road“ Veröffentlichungen im Rahmen der Open Access Initiative bedeutet das Aufweichen traditioneller Veröffentlichungsformen und der Wahrnehmungsmessung. Nun sind Dokumente erstmals zur bekannten „Journalveröffentlichung“ (meist Zeit versetzt) auch auf Dokumentenrepositorien nachgewiesen und wahrnehmbar. Wenn Autoren sogar die „Golden Road“ beschreiten, existiert das Paper nur noch auf dem Dokumentenserver.

Der Nachweis von wissenschaftlichem Output und seiner Wahrnehmung erhält vor diesem Hintergrund eine neue Dimension. Offensichtlich genügt es nicht mehr, nur noch die einschlägigen Nachweisdatenbanken zu konsultieren, um Output (Publikationshäufigkeit) und Wahrnehmung (Zitierhäufigkeit) zu bestimmen, sondern es ist notwendig, auch auf die beschriebenen Web-Ressourcen zu rekurren. Gleichzeitig sind Redundanzen in den beiden Systemen zu vermuten, da ein Abgleich nicht erfolgt.

Wir haben deshalb im folgenden die Nachweismöglichkeiten traditionell erzeugter Wissenschaftsproduktion in Web im Science (WoS) und die von Web-Ressourcen durch den Web Citation Index (WCI)²⁰, SCOPUS²¹ und Google Scholar²² untersucht und Vergleiche angestellt.

Erschließungsfelder des SCI

Field Tags	Booleans
(General Search only)	AND
TS=Topic	OR
TI=Title	NOT
AU=Author	SAME
GP=Group Author	
SO=Source	
PY=Publication Year	
AD=Address	
OG=Organization	
SG=Suborganization	
SA=Street Address	
CI=City	
PS=Province/State	
CU=Country	
ZP=Zip/Postal Code	

Erschließungsfelder des WCI

Field Tags	Booleans
(General Search only)	NEAR
FT=Full Text	AND
TS=Topic	OR
TI=Title	NOT
AU=Author	SAME
PY=Year	
RE=Repository	

Abbildung 2: Suchbare Felder im Science Citation Index (SCI) und im Web Citation Index (WCI)

4 Vergleich von Web Citation Index und Science Citation Index

„Nearly 50 years after Dr. Eugene Garfield began indexing scholarly research bibliographic data, Thomson Scientific continues his emphasis on quality and relevance – carefully selecting and indexing the core literature published in peer-reviewed scholarly journals, books, and proceedings.

The Web Citation Index is a multidisciplinary citation index of Web-accessible, scholarly research papers (including articles, preprints, theses, dissertations, proceedings, technical reports, and other grey literature).

The Web Citation Index, on the other hand, contains carefully selected content. The Web Content Editors of Editorial Development serve as the content curators for the Web Citation Index, choosing content that meets defined selection criteria.”²³

So beschreibt es der Datenbankhersteller selbst und verspricht damit einen hochqualitativen Zitationsindex mit digitalen Inhalten. Quelle dieser Daten sollen e-Print Archive wissenschaftlicher Einrichtungen sein.

Das klingt zuerst gut und kann durchaus als Ergänzung zum klassischen Dienst des Science Citation Index gesehen werden, schaut man sich aber die Zusammensetzung der Daten und die Erschließungstiefe genauer an, so stellt man fest, dass gewohnte Qualitätsstandards aufgegeben werden (Abb. 2).

Die linke Abbildung lässt erkennen, in welcher Tiefe Dokumente für die Advanced Search im Science Citation Index erschlossen werden. Dagegen offenbart die Abbildung rechts, welche Felder im Web Citation Index durchsucht werden können: Der Web Citation Index bietet durch seine Struktur und inhaltliche Zusammensetzung vor allem den Vorteil, Dokumente im Volltext zu

The screenshot shows a database record for a paper titled "Design and Test of a Carbon-Tolerant Alkaline Fuel Cell". It includes the author(s) "Urquidí-Macdonald Mirna; Sen Ayusman; Grimes Patrick; Tewari Ashutosh; Sarbhay Varun", the repository "arXiv.org e-Print Archive", and citation statistics: "Cited References: 3", "Times Cited: 0", and "Associated Records: 0". An abstract follows, discussing the development of fuel cells for cars and the challenges of carbon poisoning. A WCI Number of 1395777 is also listed.

Abbildung 3: Screenshot des Datensatzes „Design and Test of a Carbon-Tolerant Alkaline Fuel Cell“

durchsuchen. Die präzise feldgenaue Suche und systematische Aufbereitung der Datensätze fehlt ihm allerdings. Im Science Citation Index ist es beispielsweise möglich, nachgelagert zur Suchanfrage eine Verfeinerung des Suchergebnisses nach „Subject categorieres“ vorzunehmen. Die Einteilung dieser Kategorien wird nach den Journals, in denen die Beiträge veröffentlicht sind, vorgenommen.

Durch die Zusammensetzung des Inhalts und deren fehlende Struktur muss auf derartige Auswertungen verzichtet werden. Aber auch die vorhandenen Felder produzieren oft nicht das gewünschte Ergebnis: Führt man beispielsweise eine Suche mit zeitlicher Eingrenzung unter Verwendung des Field Tags „PY“ (Publishing Year) durch, erhält man nur sehr wenige Treffer, weil selbst dieses einfach zu ermittelnde Feld nur sehr fragmentiert aus den Daten extrahiert wird.

Im dargestellten Datensatz „Design and Test of a Carbon-Tolerant Alkaline Fuel Cell“ (Abb. 3) fehlt beispielsweise jeglicher Hinweis auf ein Erscheinen dieser Arbeit, auf den Umfang und die veröffentlichende Einrichtung. Es wird nicht deutlich, welchen Status diese Arbeit hat (beispielsweise ein Preprint, ein bereits veröffentlichter Artikel, eine Powerpoint-Präsentation, ein Discussion-Paper o.ä.):

Auch regionale Eingrenzungen sind nicht darstellbar im Web Citation Index: Da fast jedes Dokument seine eigene Struktur hat, ist an eine differenzierte Suche, beispielsweise nach Dokumenten bestimmter Einrichtungen oder bestimmter Dokumentarten, nicht zu denken. Es fehlt die Extraktion derartiger Daten in eigenständige Felder. Das bedeutet, dass die Bewertungsmöglichkeiten in der digitalen Volltextwelt verloren geht.

¹⁹ <http://www.archive.org>

²⁰ <http://portal.isiknowledge.com/portal.cgi?DestApp=WCI&Func=Frame>

²¹ <http://www.scopus.com/scopus/home.url>

²² <http://scholar.google.de/schhp?hl=de>

²³ Martello, Angela: SELECTION OF CONTENT FOR THE WEB CITATION INDEX(TM): INSTITUTIONAL REPOSITORIES AND SUBJECT-SPECIFIC ARCHIVES, zitiert von: <http://www.scientific.thomson.com/free/esays/selectionofmaterial/wci-selection/> <ausgeführt am 25.10.2006>

Derartige Daten und Suchmöglichkeiten sind aber eigentlich die Grundlage einer professionellen Datenbank. Vor allem fehlt dem Nutzer die Möglichkeit, die Qualität des Dokuments auf fundierter Basis einzuschätzen, ohne sich jedes Dokument einzeln anschauen zu müssen.

Das dargestellte Dokument im Volltext ist beispielsweise sehr umfangreich und umfasst 78 Seiten. Ein Publikationsjahr sucht man aber selbst im Original-Dokument vergeblich. Man weiß als Nutzer also nicht, welchen wissenschaftlichen Stand das Dokument verkörpert. Bei einer Recherche zum State-of-the-art einer technischen Entwicklung kann eine genaue Zuordnung zum Stand der Entwicklung nicht erfolgen.

Als Einrichtung geben die Autoren „Penn State University“ an. Keine Angabe zu einem Land oder einer Stadt. Es bleibt in der Aufgabe des Nutzers, nachzurecherchieren, dass die genannte Universität die „Pennsylvania State University“ ist. Recherchiert man das Paper im Netz nach, kann man herausfinden, dass es vom April 2005 ist. Auch in CiteBase²⁴, einer kleineren Datenbank, die ebenfalls wissenschaftliche Literatur verzeichnet und frei im Netz zugänglich ist, findet man den Artikel mit korrekter Angabe des Datums. Ein sehr interessantes Feature für webometrische Auswertungen ist die Möglichkeit, sich in CiteBase die Download-Häufigkeit und die Anzahl der Zitate (beides auch in zeitlichem Verlauf) ausgeben zu lassen. Leider ist die Datenbank noch sehr wenig umfangreich und erst im Testbetrieb, was zu Lasten der Aussagekraft dieses Features geht. Im Prinzip wird hiermit aber erstmals das möglich, wonach Webometriker suchen: der Versuch, einen Zusammenhang zwischen Download (Hit) und Zitation herzustellen.

Beispielhaft für einen Artikel der Physik²⁵ ist die Grafik der Abbildung 4.

Erst wenn diese Datenbank eine kritische Masse erreicht hat werden derartige Aussagen dann auch auf einer validen Grundlage getroffen.

Scopus

Die Datenbank Scopus von Elsevier als direkte Konkurrenz zum Science Citation Index bietet bereits die Berücksichtigung von Webressourcen an. So kann man aus einer Suchoberfläche heraus eine Suche absetzen und dann wahlweise als Ergebnis wissenschaftliche Veröffentlichungen aus der konventionellen Datenbank oder aus dem Web wählen.

Die genaue zeitliche Abdeckung des Web-Content ist sowohl im WCI des ISI als auch bei Scopus nicht feststellbar, da eine

Vielzahl an unterschiedlichen e-print Archiven und anderen Dokumentenrepositorien ganz individuelle Zeiträume umfassen. Eine verlässliche Angabe der Datenbankanbieter zu diesem Punkt gibt es nicht.

Google Scholar

Die Angaben zu Google-Scholar sind am ungenauesten. Es wird nicht offen gelegt, welche Inhalte verfügbar sind und aus welchen Quellen sie stammen. Auch thematische Festlegungen zu den Web-Inhalten gibt es ebenso wie bei Scopus und WCI nicht. Zudem ist die Anzeige der Ergebnisse bei Google Scholar für bibliometrische Zwecke wenig geeignet: so etwa wer-

im WCI sind nur die Nachnamen abgefragt, ohne Spezifizierung des Vornamens oder der Initialen des Vornamens. Die Darstellung der Schreibweise der Namen ist dem Abfrageformat der jeweiligen Datenbank angepasst.

2 Themensuche

Geht man mit einer thematischen Suche die unterschiedlichen Treffermengen von WCI, SCI und Scopus an, sind die Unterschiede noch deutlicher. In Abb. 7 haben wir eine komplexe Suche nach Gentechnik (Gentechnik I), eine Einfach-Suche (Gentechnik II), eine Einfach-Suche nach „Schnellem Brüter“ sowie drei komplexe Suchen nach „Brennstoffzellen“, „Robotik“ und „Stammzellen“ durchgeführt.

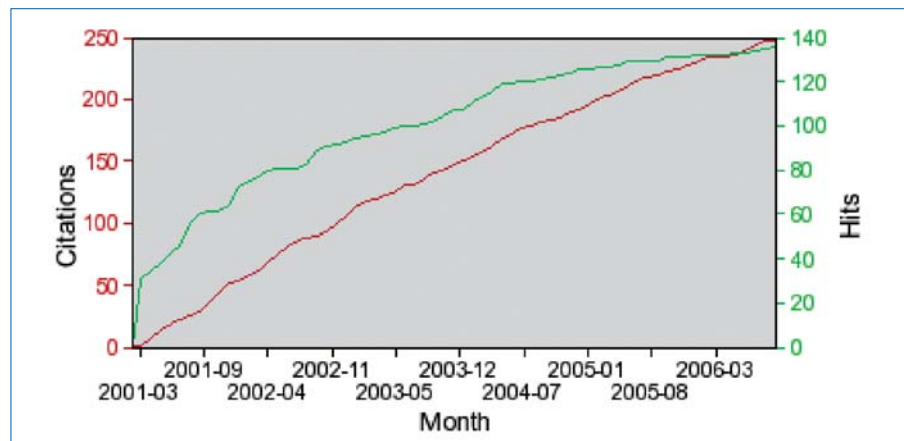


Abbildung 4: Zusammenhang zwischen Zitierhäufigkeit und Downloads eines Physik-Artikels im Laufe der Zeit (aus, V.A.: Large and infinite extra dimensions, <http://arxiv.org/abs/hep-ph/0104152>)

den Fußnoten zu einer Arbeit und die Zitate zu dieser Arbeit in einer einzigen Ergebnisliste angezeigt.

Beispielanalysen

1 Personensuche

Bei einer Quick and dirty Analyse zur Tendenzbestimmung der Abdeckung von bibliometrischen Webdatenbanken haben wir hochbewertete Wissenschaftler aus dem Bereich der Physik und der Lebenswissenschaften auf ihre Webpräsenz überprüft. Die Ergebnisse sind in Abb. 5 und Abb. 6 dargestellt. Es ist schnell offensichtlich, dass die Webressourcen meist nur einen Bruchteil der im SCI nachgewiesenen Beiträge darstellen. Lediglich bei einem Autor ergab die Suche im WCI eine größere Treffermenge als die des SCI. Die Abdeckung bei Scopus scheint (zumindest von Größendimensionen her) ähnlich der des SCI. Völlig überraschend sind die hohen Treffermengen bei Google Scholar. Sie liegen durchweg über den Mengen des SCI und weit über denen des WCI.

Eine ausführliche Interpretation dieser „Schnellanalyse“ verbietet sich. Durch die unterschiedliche zeitliche Abdeckung der Datenbanken sind die zu Grunde liegenden Datenbestände unterschiedlich groß,

Die Ergebnisse sprechen eine überdeutliche Sprache: der WCI konnte bei allen Fragen nur einen Bruchteil der Ergebnisse liefern, den der SCI geliefert hat. Scopus hingegen hat meist höhere Treffermengen, kommt allerdings mit komplexeren Suchanfragen weniger zurecht.

Obwohl auch diese Analyse den Stil einer Quick and dirty Analyse hat, zeigen die Trends klar, dass für thematische Anfragen der WCI noch längst nicht geeignet ist. Nimmt man als weiteres Beispiel das Massachusetts Institut of Technology (MIT) so findet man von dieser renommierten Einrichtung im WCI lediglich 7997 Dokumenten, während der SCI mehr als 100.000 Dokumente verzeichnet.

5 Zusammenfassung

Die Berücksichtigung von Web-Ressourcen und den darauf abgestellten Suchmaschinen und Datenbanken ist auch heute noch kaum von Vollständigkeit geprägt. Einerseits sind die Web-Ressourcen weitgehend uneinheitlich, zum zweiten sind die Suchergebnisse, die man mit Datenbanken erhält, die Web-Ressourcen verzeichnen, noch immer Zufallsprodukte.

Das Auffinden und die Feststellung von verwandten Arbeiten (related records) in Datenbanken als Mittel zur Steigerung po-

²⁴ <http://www.citebase.org/> <ausgeführt am 25.10.2006>

²⁵ Rubacok, V.A.: Large and infinite extra dimensions, <http://arxiv.org/abs/hep-ph/0104152> <ausgeführt am 25.10.2006>

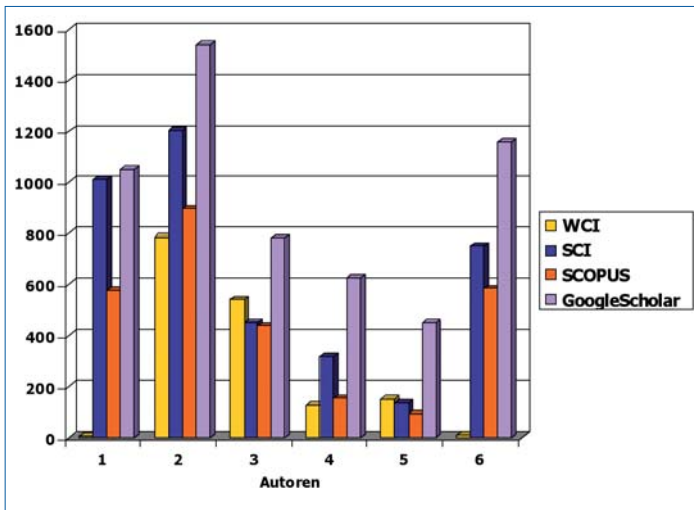


Abbildung 5: Anzahl der Downloads bei der Namensuche hochbewerteter Wissenschaftler aus dem Bereich der Physik und ihre Webpräsenz in verschiedenen Datenbanken

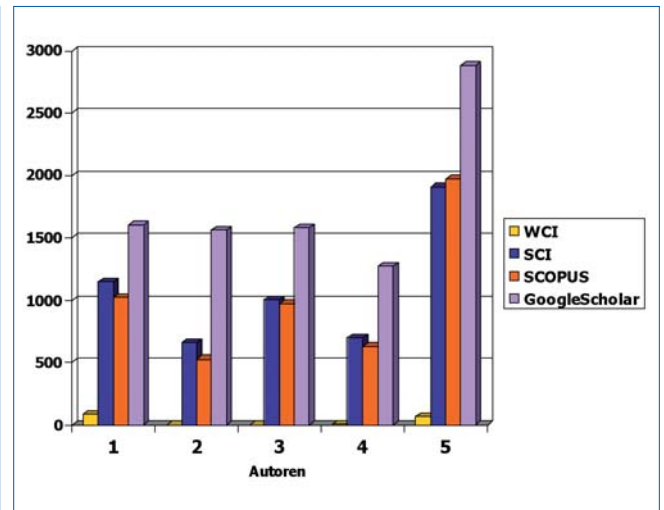


Abbildung 6: Anzahl der Downloads bei der Namensuche hochbewerteter Wissenschaftler aus dem Bereich der Lebenswissenschaften und ihre Webpräsenz in verschiedenen Datenbanken

tenzieller Relevanz bei der Recherche birgt auch ein großes Potenzial für bibliometrische Zwecke, die in der Bedeutung weitaus tief greifender sind als Kozitierungen oder bibliographische Paare. Theoretisch sind die Effekte bei Volltextdatenbanken noch ausgeprägter, da Analysen der vollständigen, in elektronischer Form vorliegenden Texte gemacht werden können. Wie dabei der Grad der Relevanz von Referenzen bestimmt wird, hängt von der Struktur der Datenbank und von Softwaremerkmalen ab. Neben Zitaten, die traditionell ein Verbindungsglied von Arbeiten darstellen und Relationen in der Wissenschaft aufschlüsseln, bieten sich Relationsindikatoren wie Titel- und Schlüsselwörter an.

Van Raan hält es für wenig wahrscheinlich, dass *hypertextlinking* (bei dem sich elektronische Dokumente direkt referieren bzw. nur elektronische Informationsquellen zitiert werden) die klassischen Zitierweisen verdrängen wird.²⁶ Das *Hyperlinking* entbehrt festgesetzter Elemente des

Auswertungsprozesses, was die Grundlage des in der Wissenschaft vorherrschenden Reputationssystems bildet. Auch Egghe vertritt die Meinung, dass konzeptuelle Unterschiede zwischen dem Hyperlinking und klassischen Zitaten vorliegen, besonders im Hinblick auf eine mangelnde quantitative Bestimmung des „Einflusses“²⁷. Van Raan erwartet, dass elektronische Dokumente die Möglichkeiten von klassischen Zitierweisen erweitern werden.²⁸ Elektronische Versionen gedruckter Zeitschriften oder wissenschaftliche Zeitschriften, die ausschließlich in elektronischer Form vorliegen, ermöglichen einen schnellen Zugriff auf den Volltext, und somit bessere Möglichkeiten für die Analyse, weshalb auch Verbesserungen von bibliometrischen Untersuchungen zu erwarten sind.

Zusammenfassend kann man schließen, dass durch die Webometrie neue quantitative Aussagen möglich sind (etwa hits per paper, number of downloads), aber gleich-

zeitig durch die nicht standardisiert abgelegten und/oder referierten Paper und Dokumente eine notwendige, qualitativ hochwertige Erschließungs- und Bewertungstiefe nicht möglich ist. Die Webometrie dringt nicht in das Deep-WEB ein und bleibt folglich an der Oberfläche des visible Web hängen. Die Konsequenzen hieraus sind Zufall und Beliebigkeit, auf die ein seriöses wissenschaftliches Informationsmanagement nicht setzen kann.

Zum Autor:

Dr. Rafael Ball
ist Leiter der
Zentralbibliothek des
Forschungszentrums
Jülich GmbH
D-52425 Jülich
E-Mail:
r.ball@fz-juelich.de

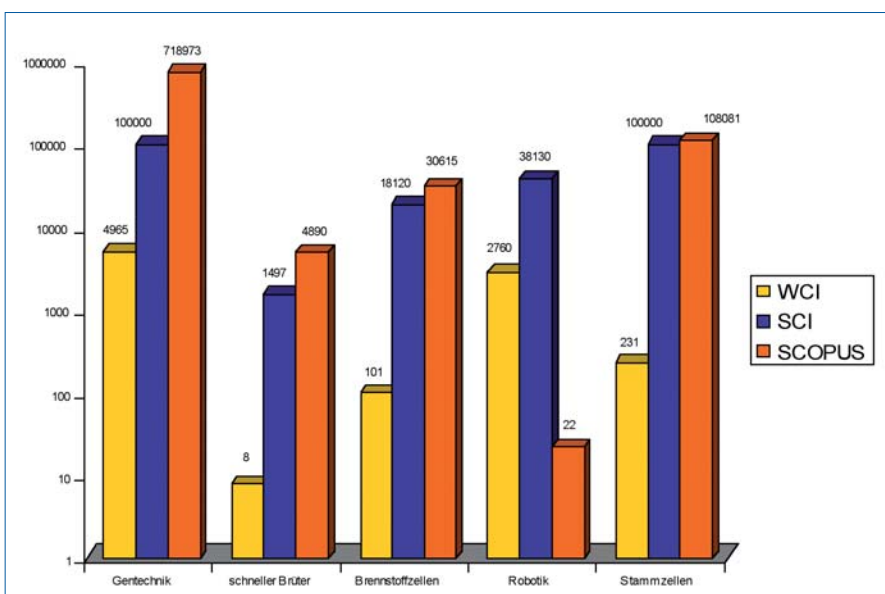


Abbildung 7: Ergebnisse einer komplexen Themensuche: Anzahl der Downloads

²⁶ an Raan, A.F.J. (a). (2001). Bibliometrics and internet, some observations and expectations. *Scientometrics*, 50(1), 59-63.

²⁷ gghe, L. (2000). New informetric aspects of the Internet, some reflections – many problems. *Journal of Information Science*, 26(5), 329-335.

²⁸ an Raan, A.F.J. (b). (2001). Competition amongst scientists for publication status, toward a model of scientific publica*Scientometrics*