

ZUR METRIE VON REFERENZEN

ALOIS KEMPF

ABSTRACT

Die Quellenangabe in Form eines Literaturverzeichnisses oder durch Fussnoten hat in der wissenschaftlichen Kommunikation eine lange Tradition. Die Hypertextstruktur des World Wide Web hat dazu beigetragen, dass diese Beziehungen zwischen zitierender Publikation und referenziertem Werk eine Beschleunigung erfahren. Im Sinne des Citation Indexing können solche Literaturnachweise (cited references) als Ausgangspunkt der Informationssuche oder als Navigationspfad zum Aufspüren von Forschungsnetzwerken (communities) eingesetzt werden. Neuerdings scheint freilich das quantitative Auszählen von Publikations- und Zitationshäufigkeiten zwecks Imagebildung an Bedeutung zu gewinnen. Da ist es wichtig zu verstehen, was und wie gezählt wird, wenn statistische Vergleiche und Rangierungen vorgenommen werden. Der Beitrag versucht eine begriffliche Klärung anhand von ausgewählten Beispielen.

1. EINLEITUNG ZUM BEGRIFFLICHEN UMFELD

Von der Sprache her betrachtet ist der Titel dieses Beitrages nicht ganz korrekt. Die Bezeichnung Metrie lässt eher eine angelsächsische Ableitung von metrics ins Deutsche vermuten. Doch bei Wikipedia – für einige Zeitgenossen heute der rasche, wenn auch unterschiedlich zuverlässige Zugang zum enzyklopädischen Wissen – finden wir den Eintrag Metrie, mit der Erläuterung: „Das Suffix -metrie (von griech. metron Mass) im Wort bedeutet, dass etwas gemessen wird“ [1].

Im vorliegenden Fall geht es ums Messen von Referenzen, um bibliometrische und informetrische Kennzahlen sowie deren Deutung und Bedeutung in der Wissensgesellschaft. Informetrie (oder seltener Infometrie) wird in der Regel als Oberbegriff verwendet, wengleich Bibliometrie und Ableitungen davon (bibliometrisch, bibliometrics) in der Literatur häufiger sind [2]. Ebenso begegnet man Bezeichnungen wie Mediametrie, Literametrie, Webometrie oder Szientometrie, die jeweils eine bestimmte Ausprägung von „Information“ zum Messgegenstand haben [3]. Vereinfacht ausgedrückt handelt es sich bei der Bibliometrie um die Analyse der vornehmlich schriftlichen Wissenschaftskommunikation, gestützt auf die Beziehungen von zitierenden VerfasserInnen und deren als Quellen zitierten Werken.

Auch Literaturdatenbanken und Informationsdienste nutzen für die Online-Navigation zunehmend das Reference Linking, um Verknüpfungen zwischen Einträgen im Literaturverzeichnis, den dazu verfügbaren Volltexten oder weiteren relevanten bibliografischen Angaben anzuzeigen. Beispiele dazu findet man unter anderem in PubMed, in CSA-Datenbanken unter der Oberfläche Illumina, in EBSCO Academic Search Premier, bei Ovid SilverPlatter-Recherchen oder in ScienceDirect von Elsevier. Thomson Scientific (ISI) bietet überdies in Web of Science eine Vergleichs- und Rangierungsfunktion Related Records an, die auf dem Prinzip der Kozitation, also gemeinsamer Referenzen, beruht. Die genannten Linkmechanismen erlauben viele nützliche neuartige Suchstrategien, um elektronisch verfügbare Dokumente zu durchstöbern und abzusuchen. Mit den Google-Befehlen link: und related: aus der „seitenspezifischen Suche“ [7] lassen sich Referenz-Zitations-Analysen beispielsweise auch in Google Scholar, dem auf wissenschaftliche Dokumente ausgerichteten Datenbanksegment, einsetzen.

2. WISSENSCHAFTLICHE VERÖFFENTLICHUNGEN UND DEREN REZEPTION

Nicht die oben genannten Navigationspfade sind jedoch Hauptaspekt dieses Beitrags, sondern die zunehmende Diskussion um die Interpretation von bibliometrischen Kennzahlen. Ein Kürzel, das in diesem Zusammenhang stets auftritt, lautet ISI. Früher als Abkürzung für das Institute for Scientific Information von Eugene Garfield [8] bekannt, sind die drei Buchstaben inzwischen zum Grundwortschatz für Anhänger und Gegner von Rankings auf der Basis bibliometrischer Daten geworden. Heute im Besitz von Thomson Scientific verfügt der Zitationsindex mit den verschiedenen Teildatenbanken (Science Citation Index Expanded SCIE, Social Sciences Citation Index SSCI und Arts and Humanities Citation Index A&HCI) über eine langjährige Messreihe bibliografischer Datensätze, die eine quantitative Auswertung der Relation Referenz / Zitation erlauben. Die Produkte solcher Analysen stammen einerseits von Thomson Scientific selber, andererseits beziehen verschiedene Forschungsgruppen entsprechende ISI-Datenreihen für ihre je spezifisch entwickelte Methodik, so etwa CEST in Bern/CH, CWTS in Leiden/NL oder das deutsche Hochschulranking [9]. Dabei wird nicht bloss gezählt, sondern häufig auch ein „Impact“ errechnet, um damit die Wirkung von Veröffentlichungen als Indikator für Forschungstätigkeit zu deuten.

An einem konkreten Beispiel soll das Beziehungsgeflecht von zitierendem Autor und zitiertem Autor in Erinnerung gerufen werden. Ich wähle dazu eine Publikation von Heinz Hauffe zum hier behandelten Thema Zitationsanalyse (siehe Tabelle 1, Ebene L2). Auf weiter gehende Erläuterungen zu Begriffen wie bibliografische Koppelung

von Veröffentlichungen, Kozitation von Autoren in Literaturverzeichnissen oder Selbstzitation wird verzichtet. Ebenfalls vernachlässigt wird die Mehrdeutigkeit von Beziehungen zwischen zitierendem Werk und zitiertem Werk, wenn gleichzeitig Print- und Online-Version erschienen sind oder elektronische Dokumente in unterschiedlichen Formaten (PDF, HTML) und Archiv-Kopien genutzt werden können.

Ebene L3 späterJahr ≥	Hauffe, Heinz: The role of citation analysis in the history and evaluation of science. Online-Mitteilungen Nr. 79, Juli 2004, S. 7-9 Bericht über einen Vortrag von Eugene Garfield (Wien, 26. Mai 2004)	Hauffe 2004 referenziert Hauffe 1994 sowie Seglen 1989
Ebene L2 Vergleichs- punkt	Hauffe, Heinz: Is citation analysis a tool for evaluation of scientific contributions? <i>Pteridines</i> 5 (1994), p. 73-74. [siehe auch: http://www.uibk.ac.at/ub/mitarbeiter_innen/publikationen/hauffe_is_citation_analysis_a_tool.html , mit Kommentar Eugene Garfields]	Hauffe 1994 wird zitiert von Hauffe 2004. Hauffe 1994 referenziert Seglen 1989.
Ebene L1 früherJahr ≤	Seglen P.O.: From bad to worse: evaluation by journal impact. <i>Trends in Biochemistry</i> 14, Aug. 1989, p. 326-327.	Seglen 1989 wird zitiert von Hauffe 1994 sowie Hauffe 2004

Tabelle 1: Relation von Referenz und Zitation, eine Frage der Sichtweise

Im Sinne des „Schneeball-Prinzips“ lautet die Referenzkette in unserem Beispiel: Hauffe 2004 zitiert (referenziert) Hauffe 1994, worin der Autor Hauffe bereits einmal dem Werk von Seglen 1989 die Referenz erwiesen hatte. Die Technik des *Citation Indexing* folgt dem umgekehrten Suchpfad: Bekannt ist die Arbeit von Seglen 1989 (siehe Abbildung 2). Gesucht werden im Zitationsindex Arbeiten, die sich im jeweiligen Quellenverzeichnis darauf bezogen haben (Vgl. Abbildung 3). Hauffe 1994 ist jedoch nicht unter den 16 Zitierenden; denn die Zeitschrift *Pteridines* gehörte gemäss Web of Science erst ab 1995 zu den ausgewerteten ISI-Quellen. Seglen hätte sonst eine Zitation mehr, nämlich jene von Hauffe 1994. Die Beziehung zu Hauffe 2004 fehlt ebenfalls, da die „Online-Mitteilungen“ nicht zu den für den Zitierindex ausgewerteten internationalen Fachzeitschriften gehört und damit das deutschsprachige Editorial für Seglen 1989 wiederum keine Zitation ausweist. In bibliometrische Studien, die sich auf diese elektronisch verfügbaren und rasch zu verwertenden ISI-Datenreihen stützen, muss dies verständlicherweise die errechneten Verflechtungen verfälschen. Das Auffinden eines bibliografischen Datensatzes in Web of Science garantiert übrigens nicht, dass dieser Datensatz zur engeren Auswahl jener Core-Journals zählt, die Bibliometriker für ihre Analysen berücksichtigen (Differenz von SCI zu SCIE).

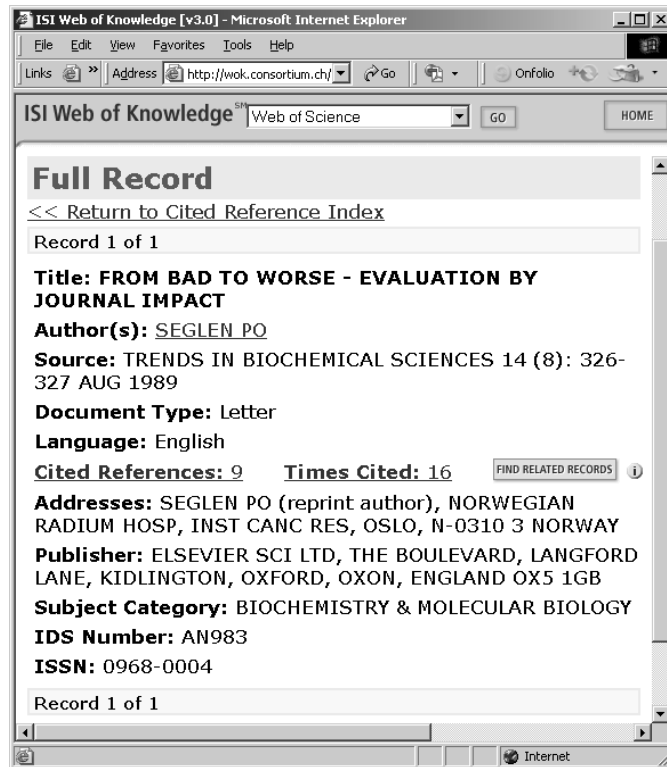


Abbildung 2: Bibliografischer Nachweis in Web of Science, General Search zu Seglen P.O. 1989

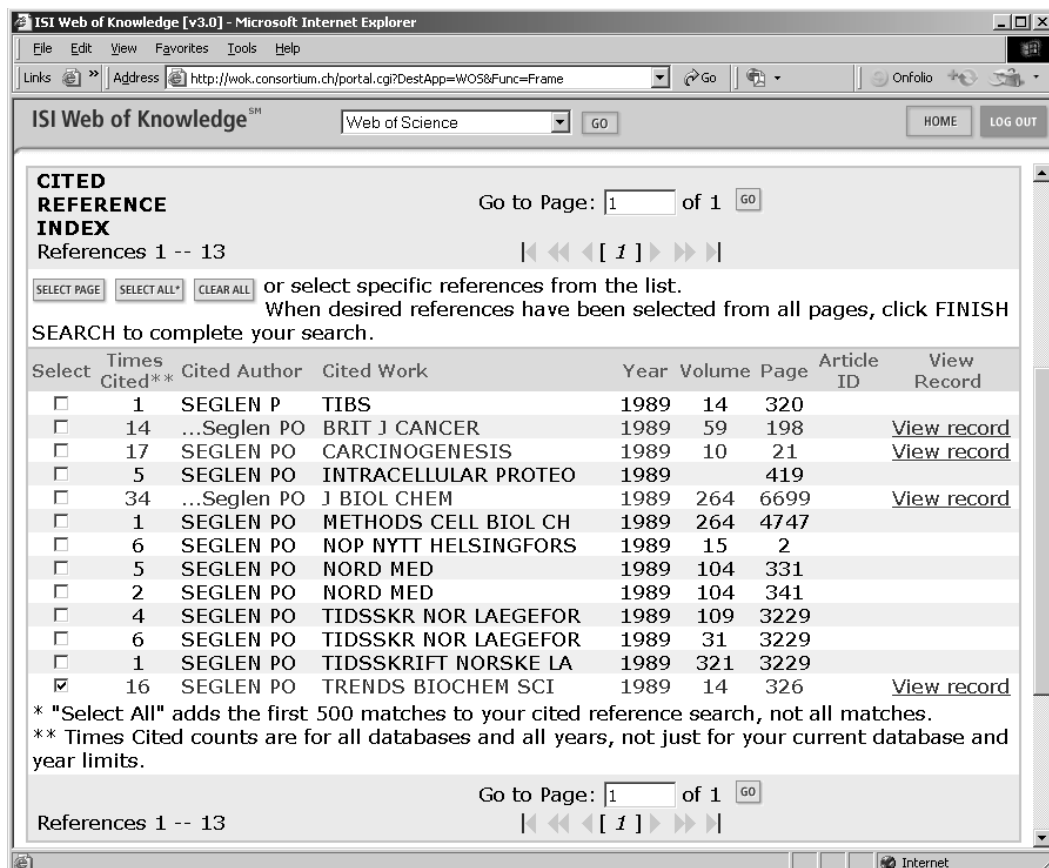


Abbildung 3: Zitationsindex in Web of Science, Cited Reference Search zu Seglen P.O. 1989

3. VON BIBLIOGRAFISCHEN ANGABEN ZU BIBLIOMETRISCHEN KENNWERTEN

Als Rohmaterial für Auswertungen sind die Datenfelder Autor, Werk, Publikationsjahr und zitierte Referenzen (d.h. zitierter Autor, zitiertes Werk, zitiertes Jahr) wichtig. Von Bedeutung sind zudem die Angaben im Adressfeld der ISI-Datenbank. Aus der Anschrift des Autors und allfälliger Mitautoren lassen sich Deutungen zur Person und deren Affiliation (Zugehörigkeit zu Institution und Land) ableiten. Die Auswertungsmerkmale Autor, Werk, Institution, Zeitschrift und Land kommen auf der Startseite von ISI Essential Science Indicators (ESI) und auf der Werbeseite www.in-cites.com (siehe Abbildung 4) gut zum Ausdruck. Vergleiche und Ranglisten folgen solchen Gruppierungen. Wegen vielfältiger Abkürzungs- und Sprachvarianten ergeben sich bei den Referenzen und Adressen einige Unsicherheiten.

Etwas komplexer gestaltet sich die thematische Zuordnung, die bei Diskussionen um die Bewertung von Publikationsleistungen und Forschungsoutput relevant wird. Bei Thomson Scientific erfolgt die Klassierung einzelner Beiträge (Papers) durch Zuweisung der betroffenen Zeitschrift zu einem konkreten Fachgebiet (vgl. Feld Subject Category in Abbildung 2). Die Klassifikation ist in Zeitschriftenlisten von ISI nachzuschlagen und erscheint dann u.a. im ISI-Produkt Journal Citation Reports (JCR) als Gruppierungselement [10].

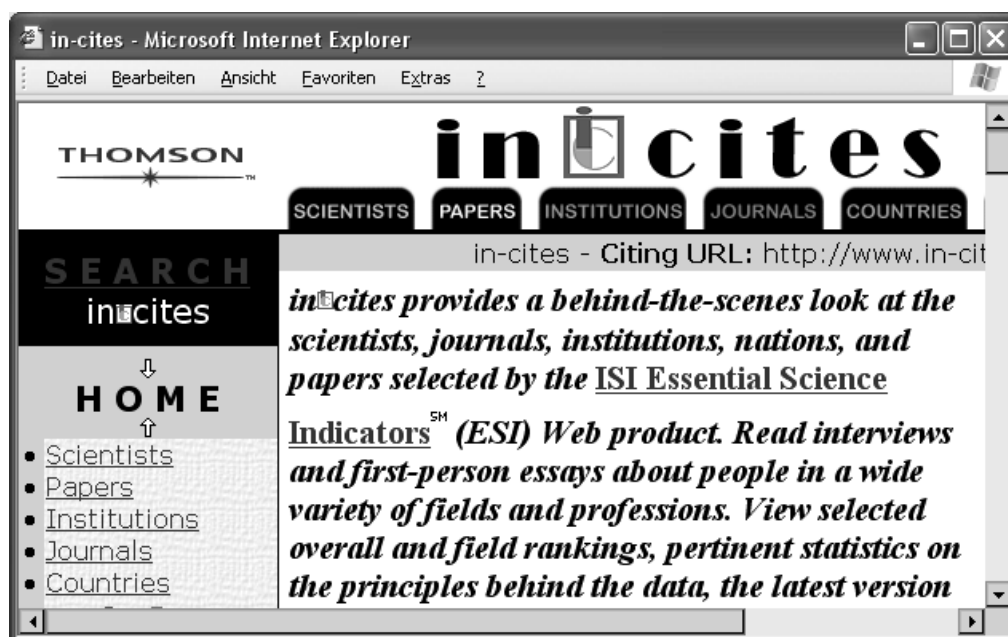


Abbildung 4: Auswertaspekte bibliometrischer Studien

Auswahlverfahren und Quellenlage sind in der Regel im Methodenkapitel von bibliometrischen Untersuchungen ausführlich beschrieben. Oft kann man dort auch die Empfehlung nachlesen, dass das bibliometrische Instrumentarium mit Vorteil im Verbund mit anderen Bewertungsverfahren, insbesondere dem Peer Review, eingesetzt werden

sollte. Das eigentliche Problem tritt dann jedoch beim Interpretieren der Ergebnisse zu Tage, wenn die quantitativen Auszählungen als Indikator für qualitative Aussagen, sprich Qualität der Forschungsleistung, benutzt werden. Aus dem Blickwinkel ökologischer und langfristig angelegter Forschungsprojekte sind etwa die Zweckmässigkeit eines Aktualitätsindex (laufendes Jahr) oder eines zweijährigen Zeitschriften-Impacts (in JCR) kritisch zu hinterfragen (vgl. Abbildung 5). Sind die ausgewerteten Zeitschriftenbeiträge und die darin enthaltenen Referenzen des Quellenverzeichnisses das adäquate Mass für die betroffene Fachdisziplin? Zum anerkannten Grundwissen der Bibliometrie und Szientometrie gehört nämlich, dass die Zitationsmuster je nach wissenschaftlicher Disziplin variieren, der Verlauf der Zitationskurve (siehe Abbildung 5) folglich unterschiedlich verläuft. Neben der zweijährigen Beobachtungsperiode für JCR errechnet Thomson Scientific für Essential Science Indicators (ESI) eine gleitende 10-Jahresperiode [11]. Da in Ranglisten manchmal Werte aus beiden Quellen aufgeführt sind [12], sollte man den Hinweis auf deren unterschiedliche Herleitung beachten. Die ausgewerteten Zeitschriften von ESI und JCR überlappen weit gehend.

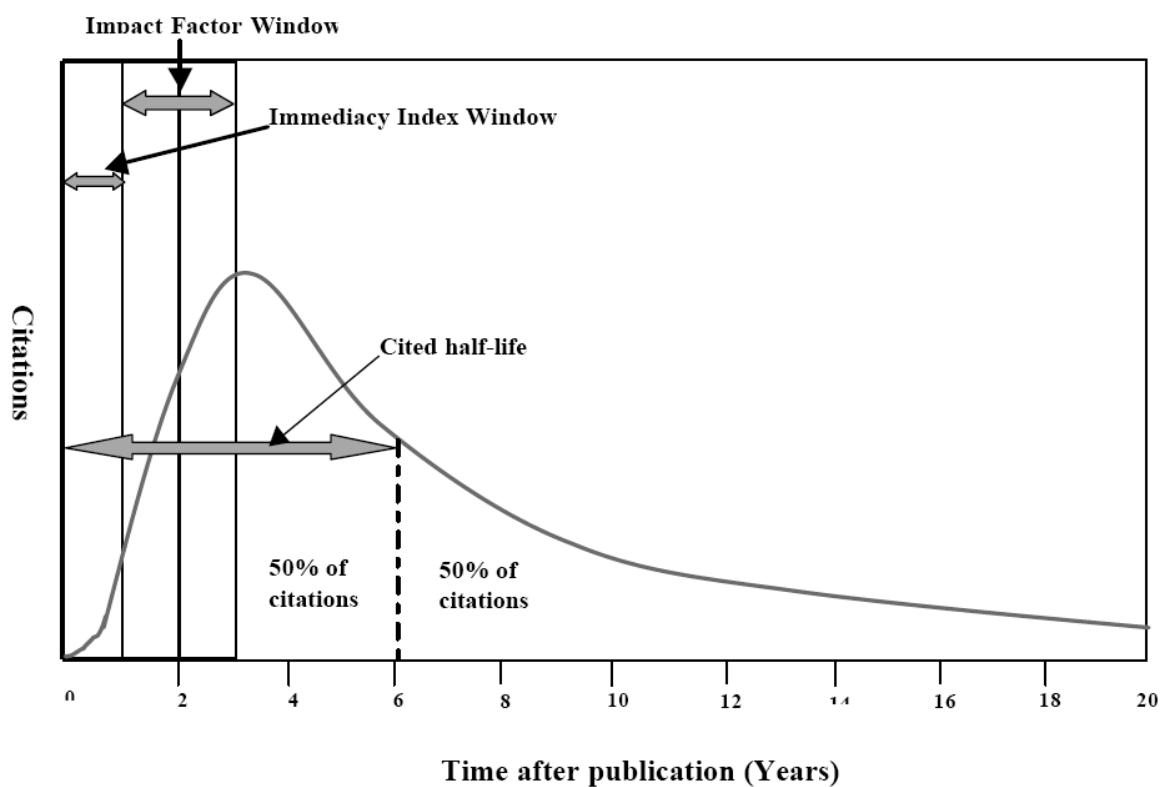


Abbildung 5: Ablauf von Zitationen in Zeitfenstern

Eine Interpretation ist nicht zuletzt auf Sprache angewiesen. Dies gilt auch für wissenschaftliche Analysen, wo Sprachbilder oder bestimmte Formulierungen den Gang der Dinge zu beeinflussen vermögen. Der Begriff „Champions League“, von CEST für überdurchschnittliche Werte bei der Rezeption (impact) verwendet, kann hier als Beispiel erwähnt werden [13]:

Seit 2002 erstellt das CEST mit Hilfe bibliometrischer Indikatoren ein internationales Ranking von Forschungsorganisationen, bekannt unter dem Titel „Champions League“. (...)Der gewählte Titel „Champions League“ führte zudem zu Interpretationen der Studie, welche weit darüber hinaus gingen, was mit dem Ranking von Forschungsorganisationen mit Hilfe bibliometrischer Indikatoren ausgesagt werden kann. Aus diesen Gründen verzichtet das CEST darauf, die betreffenden Studien weiterhin öffentlich zugänglich zu machen.

Das Beispiel verweist ferner auf den möglichen Konflikt, wenn wissenschaftlich erarbeitete Analysen in Kurzfassungen (Executive Summary) auf den Punkt gebracht werden sollen. Da Entscheidungsträger selten auch das Kleingedruckte zum methodischen Vorgehen nachlesen werden, kann eine politisch orientierte Einschätzung ziemlich weit entfernt vom Kern des laufenden Forschungsgeschehens stattfinden.

P.S. Von der Thematik her nachzutragen ist nach der ODOK'05 vom September der so genannte h-Index von J.E. Hirsch [14]. Dabei handelt es sich um einen neuen Vorschlag zur Bewertung wissenschaftlicher Publikationsleistungen einzelner Forscher auf der Basis von Publikationen (Output) und Zitationen (Rezeption): „A scientist has index h if h of his/her N_p papers have at least h citations each, and the other ($N_p - h$) papers have no more than h citations each.«. Der Wert h wird also definiert durch jene Stelle in einer Rangliste von Publikationen, an der die Anzahl der Zitierungen noch mindestens dem Rang entspricht. Ob damit ein allgemein anerkannter Bewertungsmaßstab gefunden wurde, bleibt abzuwarten. Die Diskussion ist freilich nur scheinbar neu lanciert, so lange die Grundlagendaten „auf die Schnelle“ den gleichen Quellen entnommen werden, die man bei anderen Produkten zuvor als nicht repräsentativ beanstandet hatte.

ANMERKUNGEN

- 1 Siehe <http://de.wikipedia.org/wiki/-metrie>
- 2 Hood, W.W. & C.S. Wilson: The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics* 2001, 52(2): S. 291-314.
- 3 Sengupta, I.N.: Bibliometrics, scientometrics and librametrics; an overview. *Libri* 1992, 42(2): S. 75-98.
- 4 Lafouge, T., Le Coadic, Y.F. & C. Michel: *Éléments de statistique et de mathématique de l'information : infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie; cours avec exemples et exercices corrigés.* Villeurbanne: Presses de l'enssib, 2002, 319 S. (Les cahiers de l'enssib; 1)
- 5 Siehe <http://www.isinet.com/essays/citationindexing/1.html/>
- 5 Siehe <http://www.touchgraph.com/TGGoogleBrowser.html>

- 6 Muschke, P.: Autorennetzwerke: Netzwerkanalyse als Mehrwertdienst für Informationssysteme. ISI 2004. In: Schriftenreihe zur Informationswissenschaft (Konstanz), 2004, 42: S. 141-162.
- 7 Vgl. http://www.google.ch/advanced_search?hl=de sowie <http://scholar.google.com>
- 8 Siehe <http://www.garfield.library.upenn.edu/> sowie <http://www.thomsonscientific.com> und <http://scientific.thomson.com/mjl/>
- 9 Siehe <http://www.cest.ch>, <http://www.cwts.nl> und <http://www.che.de>
- 10 Siehe <http://scientific.thomson.com/products/categories/evaluation/>
- 11 Kempf, A.: Der Impaktfaktor, ein Begriff im Plural. ISI 2004. In: Schriftenreihe zur Informationswissenschaft (Konstanz) 2004, 42: S. 471-473.
- 12 Siehe SCI-Bytes unter <http://www.in-cites.com/research/2006/index.html>
- 13 Siehe http://www.cest.ch/de/archiv/Kommentar_CEST_CL.pdf
- 14 Siehe Preprint unter <http://arxiv.org/abs/physics/0508025>

ADRESSE DES AUTORS

Dr. Alois Kempf

Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft WSL

CH-8903 Birmensdorf, Zürcherstrasse 111

E-Mail: alois.kempf@wsl.ch