Facets of Informetrics

Subjects and Research Areas of Informetrics

According to Jean M. Tague-Sutcliffe, “informetrics” is “the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists” (Tague-Sutcliffe, 1992, 1). Leo Egghe also defines „informetrics“ in a very broad sense. “(W)e will use the term ‘informetrics’ as the broad term comprising all-metrics studies related to information science, including bibliometrics (bibliographies, libraries, ...), scientometrics (science policy, citation analysis, research evaluation, ...), webometrics (metrics of the web, the Internet or other social networks such as citation or collaboration networks)” (Egghe, 2005b, 1311). According to Concepción S. Wilson „informetrics“ is „the quantitative study of collections of moderate-sized units of potentially informative text, directed to the scientific understanding of information processes at the social level“ (Wilson, 1999, 211). We should add to Wilson’s units of text also digital collections of images, videos, spoken documents and music. Dietmar Wolfram divides „informetrics“ into two aspects, „system-based characteristics that arise from the documentary content of IR systems and how they are indexed, and usage-based characteristics that arise how users interact with system content and the system interfaces that provide access to the content“ (Wolfram, 2003, 6).

We would like to follow Tague-Sutcliffe, Egghe, Wilson and Wolfram (and others, for example Björneborn & Ingwersen, 2004) and call this broad research of empirical information science „informetrics“. Informetrics includes therefore all quantitative studies in information science. If a scientist performs scientific investigations empirically, e.g. on information users’ behavior, on scientific impact of academic journals, on the development of the patent application activity of a company, on links of Web pages, on the temporal distribution of blog postings discussing a given topic, on availability, recall and precision of retrieval systems, on usability of Web sites, and so on, he or she contributes to informetrics. We see three subject areas in information science in which such quantitative research takes place.

 Following Wolfram’s article, we divide his system-based characteristics into the „information itself“-category and the „information system“-category. Figure 1 is a simplistic graph of subjects and research areas of informetrics as an empirical information science.

Nomothetic and Descriptive Informetrics

„Information itself“ can be studied in various ways. Generally, we work descriptively on information, information flows or content topics and try to derive informetric regularities by generalizing the descriptive propositions or using mathematical models (Egghe & Rousseau, 1990). Examples are the laws of Lotka (Egghe, 2005a) or the „inverse logistic“ distribution of documents by relevance (Stock, 2006). Following the Greek notion of „nomos“ (law) we will call this kind of empirical information science „nomothetic informetrics“ (Stock, 1992, 304). Typical
nomothetic research questions are „What kind of distribution is adequate to multi-national authorship?” (which is the topic of Egghe’s article in this issue) or „Are there laws of temporal distribution of, say, blog postings?” In contrast to nomothetic informetrics stands descriptive informetrics which analyses individual items such as individual documents, subjects, authors, readers, editors, journals, institutes, scientific fields, regions, countries, languages and so on. Typical descriptive research questions are „What are the core subjects of the publications of Albert Einstein?” or „How many articles did Einstein publish per year in his whole life time?” If there are known informetric laws, researchers can compare the findings of their descriptive work to these laws. This way they can make a distinction between „typical” individual distributions (if the individual’s data approximate one of these laws) and non-typical distributions. Methods of data gathering in informetrics which are concerned with information itself consist of citation analysis (Garfield, 1972; Garfield, 1979; Cronin & Atkins, Eds., 2000; for problems of citation analyses see MacRoberts & MacRoberts, 1996) and publication analysis (Stock, 2001a), including subject analyses of publications. According to Jürgen Rauter (2006) citation analysis is theoretically related to literary studies, for the concept of „intertextuality” used in the study of literature finds its expression as a reference and citation in the literary genre of academic writings.

The study of information itself has been called „bibliometrics” (Pritchard & Wittig, 1981) too. „Bibliometrics” is sometimes used in the context of scientometrics. However, the concept of „bibliometrics” refers to books (old Greek „biblos” means „book”). It is therefore more appropriate to use the concept of „informetrics” as the broadest term as it includes all kinds of information. The term „informetrics” (in German „Informetrie”) was coined by O. Nacke (1979) in the Federal Republic of Germany and by L. Blackert and K. Siegel (1979) in the German Democratic Republic.

Scientometrics

According to A.P.J. van Raan „(s)cientometric research is devoted to quantitative studies of science and technology” (van Raan, 1997, 205; see also Callon, Courtial, & Penan, 1993). Main subjects of scientometrics are individual scientific documents, authors, scientific institutions, academic journals, and regional aspects of science. Scientometrics exceeds the boundaries of information science. „We see a rapid addition of scientometric-but-not-bibliometric data, such as data on human resources, infrastructural facilities, and funding” (van Raan, 1997, 214). In information-science oriented scientometrics, in contrast to economy, sociology or psychology of science, aspects of information and communication are examined. These aspects may include productivity (documents per year), subjects of the documents (words, co-words), reception (readers of the documents) and formal communication (references and citations, information flows, co-citations) (Juchem, Schúglö, & Stock 2006, 32).

Scientometrics is focused on scientific information only. There are other kinds of special information, above all patent information and news information are very important. Quantitative studies of patent information can be called „patentometrics” or „patent bibliometrics” (Narin, 1994), empirical studies of news „news informetrics”. Patentometrics, scientometrics and news informetrics together form a special field to produce some interesting indicators for economics (for patentometrics, see Griliches, 1990).

Webometrics

In short, webometrics is informetrics on the World Wide Web (Björneborn & Ingwersen, 2001, Cronin, 2001; Thelwall, Vaughan, & Björneborn, 2005). According to Lennart Björneborn and Peter Ingwersen, webometrics consists of four main research areas, (1) Web page content analysis; (2) Web link structure analysis; (3) Web usage analysis (including log files of users’ searching and browsing behavior); (4) Web technology analysis (including search engine performance) (Björneborn & Ingwersen, 2004, 1217). There are definite connections to other informetrics activities. Web page content analysis is a special case of subject analysis, Web link structure study (Thelwall, 2004) has its roots in citation analysis, Web usage analysis is part of a more general user and usage research, and Web technology analysis refers to information systems evaluation. Webometrics meets its subjects on the World Wide Web. But this is only one of the Internet’s services. If we include all those services such as e-mail, discussion groups and chats it is possible to speak about „cybermetrics”. We can define special branches of webometrics. So analyzing the blogosphere informetrically leads to „blogometrics”, that also represents a kind of special information, namely blog postings, podcasts and vodcasts (video podcasts).

There are close relations between general descriptive and nomothetic informetrics and special applications like scientometrics and webometrics. For example, in general informetrics co-citation analysis is a way to map the intellectual structure of a scientific field. In webometrics a co-link analysis also leads to the production of a map, but this map does not necessarily represent intellectual or cognitive structures (Zuccala, 2006). So the application of informetric methods in special fields of empirical information science is not always the same, but sometimes only a procedure by analogy. In the early days of webometrics links between Web pages and citations were seen as two sides of the same coin. Web pages „are the entities of information on the Web, with hyperlinks from them acting as citations” (Almind & Ingwersen, 1997, 400). Today we have to recognize specific differences between links and citations, for example links are time-independent and citations are not. They are „actually measuring something different and therefore could be used in complimentary ways” (Vaughan & Thelwall, 2003, 36).

User and Usage Research

Topics of user research are humans and their information behavior (Wilson, 2000). Information seeking behavior on the Web (especially the usage of search engines) is well studied (Silverstein, Henzinger, Marais, & Moricz, 1998; Spink, Wolfram, Jansen, & Saracevic, 2001; Spink & Jansen, 2004; Spink & Cole, Eds., 2006). Typical research questions are the length of queries, the use of Boolean operators, the sort of questions (e.g. concrete versus problem-oriented), the topics searched for and the number of hits noticed in a search engine’s results set. Similar studies have been conducted on users as well as usage of library services and services of commercial information providers. User research distinguishes between user groups, for example information professionals, professional end-users, and end-users (Stock & Lewandowski, 2006) or between author, reader and editor (Schlägl, 2004; Schlägl &Petschnig, 2005).

Methods of user research comprise observations of humans in information gathering situations, questionnaires and surveys, and analyses of log files. Methods of usage research include log files, statistics of downloads, numbers of interlibrary loan cases and lending numbers (in libraries). The results of user and usage research can be applied to performance and quality studies of information services.

Retrieval Evaluation

Traditional evaluation of retrieval systems (Tague-Sutcliffe, 1996; Harter & Hert, 1997), e.g. within the Text REtrieval Conferences (TREC), makes use of humans’ relevance judgments. „The relevance judgments are what turns a set of documents and topics into a test collection”, Ellen M. Voorhees (2005) says. TREC-like retrieval evaluation has only recognized two states of judgment, relevance and non-relevance, for many years. „TREC usually uses binary relevance judgments – either a doc-
Information Systems Evaluation

One can describe retrieval systems not only by recall and precision or by ADM, but also by functionality (Stock, 2000) or subject of their databases and distribution of terms to hierarchical levels, use and so on (Wolfram, 2003). There are not many retrieval systems on the Web, but millions of Web sites with billions of Web pages. Here test methods of usability (Nielsen, 2000) find broad application. Methods are, among others, task-based user tests and heuristic evaluations (Roßmann, 2002; Röttger & Stock, 2003). Merging user research, usability research and other dimensions, e.g. tests of accessibility, accuracy, relevance, believability, completeness, objectivity and timeliness, information quality research (Parker, Moleshe, De la Harpe, & Wills, 2006) is a new branch of evaluation of Web page content. When taking a broader view on systems, evaluations of libraries and other information services make use of empirical methods. Main subjects are performance and quality measurements of institutions in the library sector (Kantor, 1984; Baker & Lancaster, 1991; Poll & te Boekhorst, 1996; Stock, 1998; Reichmann, 2001) and information industry (Göcke, 1999; Stock, 2001b). Like scientometrics, performance measurements do not only study subjects of information science, but also the relations to other disciplines like business administration or economics.

Informetrics is Growing

For Egghe, quantitative information science is a growing branch of science. „There is a fast multidisciplinary expansion (growth) of the field of informetrics, mainly due to the ‘new’ topics that have been included in informetrics such as quantitative study of networks, including the Internet“ (Egghe, 2006, 1405). To confirm this assumption we performed a request on „Web of Science“ and made use of its ANALYZE-function. We searched for „TS=informetrics OR bibliometrics OR scientometrics OR webometrics OR retrieval evaluation“. This is no elaborated formulation of our topic and there are double counts (articles discussing two or more of the terms), but besides all methodological constraints there is a clear result: Informetrics is a growing scientific field. The majority of articles in the results list is on „bibliometrics“ (409 hits), followed by papers on „scientometrics“ (237), „informetrics“ (108), „webometrics“ (37) and the expression „retrieval evaluation“ (34).

Nearly every day, from September 14, 2005 to September 8, 2006, a blog post concerning our topics appeared. Besides journals with broad coverage of information science which publish many studies on empirical information science such as „Journal of the American Society for Information Science and Technology” or „Information Processing & Management” there are the following specialized journals on our subject:

- „Journal of Informetrics“ (quantitative aspects in general),
- „Scientometrics“,
- „Webometrics“, „Cybermetrics“, „Information Research“ (many articles about user and usage research),
- „Performance Measurements and Metrics“ (library performance),
- „Proceedings of the Text Retrieval Conferences“ (retrieval evaluation).

In this special issue of „Information – Wissenschaft und Praxis“ we publish research articles concerning informetrics/scientometrics/webometrics and retrieval evaluation. Due to space limitations articles about user research or non-retrieval aspects of information systems (as, e.g., usability and performance measurements) have to be left out.

The Papers in this Special Issue

In this special issue of “Information – Wissenschaft und Praxis” we publish seven research articles about informetrics. The first article (“HistCite™: A Software Tool for Informetric Analysis of Citation Linkage”) by Eugene Garfield (Philadelphia, USA), Soren Paris (Philadelphia, USA) and Wolfgang G. Stock (Düsseldorf, Germany) is concerned with the practice field of informetrics. The topic is a software called “HistCite™. By means of this software, direct citation linkages between scientific papers can be visualized and analyzed. Bibliographic records saved from citation based databases (e.g. Web of Knowledge) serve as input and deliver as output different tables or graphics with informetric data of the knowledge domain under study. In order to demonstrate how HistCite works, different analysis modes are shown by analyzing informetrically the literature about Alexius Meinong, an Austrian philosopher and psychologist. Since “Web of Knowledge” also plays an important role in this context, its informetric functionality is discussed shortly. Users of HistCite are scientists, scientometricians and science journalists.
In the next article (“Foundations and Research Areas of Webometrics”, in German language) by Mike Thelwall (Wolverhampton, Great Britain) and Tina Ruschenburg (Bielefeld, Germany) webometrics is the main topic. The authors discuss the use of link count metrics in the broad context of informetrics with special focus on social science link analysis. Just as citation analyses, link analyses can also be used as a tool for social science research. Thelwall and Ruschenburg discuss the relations between webometrics and social network analysis. Link analyses can be used in many different ways since they can be combined with different data sources and provide information about certain coherences that depend on the topic. Besides already existing and known webometric approaches, the authors also discuss future research directions of link analyses with the main focus on blogs.

The article written by Christian Schlögl (Graz, Austria) and Viktoria Pernik (Graz, Austria) also focuses on webometrics (“Possibilities and Limitations of Web Structure Mining. Information Science Departments in German-Speaking Countries as an Example”, in German language). However, strictly speaking it does not only spotlight webometrics but also scientometrics. The article describes the possibilities and limitations of Web structure mining. In this context a webometric analysis of Web sites associated with departments of librarianship and information science from Germany, Austria and Switzerland was conducted. The objectives of this analysis were to show on the one hand a hyperlink network analysis of these different institutions and on the other hand a co-link analysis, that demonstrates their similarities.

Jürgen Rauter’s (Düsseldorf, Germany) paper is about an aspect of nonomotic informetrics (“The Combination of Kleinberg’s ‘Authorities’ and ‘Hubs’ in van Rijsbergen’s ‘Effectiveness Measure’, in German language). The peculiarity of his paper is that various basic approaches of information science are brought together and put into relation. Primarily it is about the algorithm of Jon M. Kleinberg and the effectiveness measure (E-measure) of C.J. van Rijsbergen, which by means of examples is calculated at the end. The theory of the hubs and authorities of Kleinberg, that normally only appears in the context of the Internet or search engines, is now applied to scientific literature. Rauter makes use of the concept of “hypertext” in the sense of literary studies: “Hypertext” is about quoting a text, whereas “Hypotext” stands for a quoted text. It is analyzed how far hubs and authorities have a proximity with the expressions recall and precision in the sense of Gerard Salton and Michael McGill. In the context of explanation of informetric laws, the approaches of Garfield and Bradford are also considered in the analysis. The article by Steven van Impe (Antwerp, Belgium) and Ronald Rousseau (Antwerp and Oostende, Belgium) belongs to the field of descriptive informetrics with reference to webometrics (“Web-to-Print Citations and the Humanities”). Web to print citations and Web to print references are the topics of this article. Web to print citations receive references from the Web and Web to print references are references to printed documents which are made in the Web. The article focuses especially on the impact of Web to print citations. The authors ask whether Web to print citations will replace the classic (print to print) citation indexes in the future. For a better investigation of the structure of Web to print citations, an empirical study is executed which looks at the field of humanities and local history journals. Not only is the practical oriented study described, but also the theoretical background of the citation analysis. It is clarified, for example, how Web to print citations can be distinguished from links.

The article of Leo Egghe (Hasselt and Antwerp, Belgium) belongs to nomothetic informetrics (“Empirical and Combinatorial Study of Country Occurrences in Multi-Authored Papers”). It primarily is about setting up a (new) law. The subject of this article are academic papers, which were written by several authors, who are probably from different countries. Egghe derives a function of the number of multi-authored papers with n countries. It is a Lotka-like power law with a (very) high exponent.

In the area of retrieval evaluation Vincent della Mea (Udine, Italy), Gianluca Demartini (Hannover, Germany), Luca di Gaspero (Udine, Italy), and Stefano Mizzeno della Mea (Udine, Italy), and Gianluca di Gaspero (Udine, Italy) introduce a completely new approach to the measurement of retrieval effectiveness (“Measuring Retrieval Effectiveness with Average Distance Measure”). With the Average Distance Measure (ADM) the old binary approach is queried. The known measure of retrieval effectiveness has binary relevance, i.e. either a document is relevant or not relevant, and binary retrieval, i.e. either a document was found or not. There are no possibilities to get other states. This should change now by using ADM. By showing comparisons with other measurements, it becomes clear that the new basic approach has potential for evaluating Web search engines and other retrieval systems with hits sets ranked by relevance. At first the article describes the general problems of measuring the IR effectiveness in order to show the advantages and properties of the new methodology afterwards.

Most of the articles work with illustrative examples. It is worth to mention that many of the examples are adopted from social sciences (pedagogy – Egghe; information science – Pernik & Schlögl) and, which is very rare and exceptional, from humanities (philosophy – Garfield, Paris, & Stock; medieval studies – Rauter; local history – Van Impe & Rousseau).
Zuccala, A. (2006): Author cocitation analysis is to intellectual structure as Web colink analysis is to... Journal of the American Society for Information Science and Technology, 57, 1487-1502.

**Informations-Retrieval und Dokumentation**

Die komplette Anwendung über das Internet zur Miete! Neue Version (LAMP) Application Hosting

http://www.domestic.de