HistCite™:

A Software Tool for Informetric Analysis of Citation Linkage

Eugene Garfield (Philadelphia, PA), Soren W. Paris (Philadelphia, PA), Wolfgang G. Stock (Düsseldorf)

HistCite™ is a software tool for analyzing and visualizing direct citation linkages between scientific papers. Its inputs are bibliographic records (with cited references) from „Web of Knowledge" or other sources. Its outputs are various tables and graphs with informetric indicators about the knowledge domain under study. As an example we analyze informetrically the literature about Alexius Meinong, an Austrian philosopher and psychologist. The article shortly discusses the informetric functionality of „Web of Knowledge" and shows broadly the possibilities that HistCite offers to its users (e.g. scientists, scientometricans and science journalists).

HistCite™: Ein Softwarepaket für die informetrische Analyse von Zitationsbeziehungen.

HistCite™ ist eine Software für die Analyse und Visualisierung direkter Zitationszusammenhänge zwischen wissenschaftlichen Dokumenten. Als Input benötigt HistCite bibliographische Nachweise (mit Referenzen) aus „Web of Knowledge" oder anderen Quellen. HistCite erarbeitet diverse tabellarische sowie graphische Darstellungen informetrischer Indikatoren, jeweils angewandt auf eine Wissensdomäne. Als Beispiel untersuchen wir informetrisch die Literatur über Alexius Meinong, einem österreichischen Philosophen und Psychologen. Der Artikel diskutiert kurz die informetrische Funktionalität von „Web of Knowledge" und zeigt detailliert die Möglichkeiten auf, die HistCite seinen Nutzern (z.B. Wissenschaftlern, Szientometrikern und Wissenschaftsjournalisten) bietet.

Introduction:

Algorithmic historiography of science

Can a computer system create raw data for writing the history of science? More than 40 years ago, a report about „The use of citation data in writing the history of science" was published (Garfield, Sher & Torpie, 1964). It concludes (p. 33), „it is felt that citation analysis has been demonstrated to be a valid and valuable means of creating accurate historical descriptions of scientific fields, especially beyond the first quarter of the twentieth century when bibliographic citation had become well established as part of scientific publication". This report, produced by the Institute of Scientific Information (ISI) makes use of the *Science Citation Index* to study the history of science. Subsequently, in the early 1970s Henry Small developed his co-citation method for clustering documents (Small, 1973; Small & Griffith, 1974). These methods permitted disciplinary and national mappings a kind of „geography of science" (Small & Garfield, 1986), that is, „scientography" (Garfield, 1986; Garfield, 1994) and the product „ISI Atlas of Science". „Our maps represent where researchers have been. As historical records, then, these maps are surveys of the geography of scientific ideas and discoveries – intellectual gazetteers, if you will – for a given year" (Garfield, 1986, p. 3). Algorithmic historiography makes use of the „axiom", that bibliographic records (including cited references) contain concise information about scientific (or other historically relevant) content. „In our conception of facilitating historiography – that is, writing the history of modern science – we make the basic assumption that the bibliographic information contained in a collection of published scientific articles is sufficient for the purpose of recapturing the historiographic structure of the field" (Garfield, Pudovkin & Istomin, 2003, p. 400).

Analyzing science in this way leads to three types of citation analysis, (1st) to directed graphs of information flows from an information sender to an information receiver or – vice versa – of reputation flow from the receiver to the sender and (2nd) to undirected graphs derived from co-citations or bibliographic coupling. Both methods allow one to monitor (3rd) scientific developments by (a) comparing the science maps year by year or by (b) analyzing the citation history of a knowledge domain. So Henry Small could describe a scientific revolution in the sense of Thomas S. Kuhn (1962) by the means of co-citation analysis (Small, 1977).

HistCite is a software tool for analyzing and visualizing direct citation linkages between scientific papers. Its inputs are bibliographic records (including cited references) from Thomson Scientific's *Web of Knowledge (WoK)* or other sources. Its outputs are various tables and graphs with informetric data about the knowledge domain under study. *HistCite* covers types 1 and 3(b) of the above mentioned methods of citation analysis. It does not utilize co-citation mapping (kind 2 and 3(a)). Here another informetric software tool, e.g. *CiteSpace*, is applicable (Chen, 2004; Chen, 2006).

What is the „philosophy" behind algorithmic historiography of science? What role does *HistCite* play in writing history of science or – more precisely – in analyzing a specific knowledge domain? *HistCite* was designed originally to enable scholars to create genealogical microhistories of authors or topics "automatically" that is, algorithmically. In its further development it became apparent that it could be used for many purposes too numerous to mention here. In the library, given a key word or subject heading one can search *WoK* for a group of papers on a topic and then use *HistCite* to identify the most-cited core papers— usually from 25 to 50. From this *HistCite* generates tables and historiographs showing the evolution of the field.

Since *HistCite* creates a minicitation index of all the references cited in a collection it is possible to identify papers or books ("outer references") whose titles do not contain the topic of the original search. If these references are deemed relevant to the search then they can be added to the original "inner" collection by going back into *WoK* or by manual methods if the article or book is not in the *WoK* collection.

Web of Knowledge

Since *HistCite* works with outputs from *Web of Knowledge*, we should start with a short description of this database! In *WoK* it is possible to perform some informetric analyses. Thomson Scientific's *Web of Knowledge* (Stock & Stock, 2003) consists of many bibliographic databases, including the Science Citation Index, the Social Sciences Citation Index, the Arts & Humanities Citation Index and the Proceedings database (Stock, 1999). These form the bibliographic basis for informetric studies using publication and citation analyses. *WoK* additionally offers specialized analytical databases: the Journal Citation Reports (JCR) with indicators of academic journals (Stock, 2001) and Essential Sciences Indicators (ESI) with indicators of top science by countries, disciplines, institutes, journals and scientists (Stock, 2002). *WoK*'s users have some informetric functions available as part of a search. The *WoK* Analyze function permits the analysis of publications in a set of retrieved records as, e.g., ranking documents by authors, affiliations or countries and sorting of the records by times cited (see table 1).

Table 1: Informetric functionality of Web of Knowledge

Function	
ranking documents in a set of records	by authors (via ANALYZE) by country of authors (via ANALYZE) by affiliation of authors (via ANALYZE) by document types (via ANALYZE) by source titles (via ANALYZE) by languages (via ANALYZE) by publication year (via ANALYZE) by times cited (via SORT) by relevance (via SORT)
creating times series of documents in a set of records	(via ANALYZE)

As an example let us analyze informetrically the literature about Alexius Meinong, an Austrian philosopher and one of the early theorists of Gestalt psychology. Meinong (1853 to 1920) founded one of the first psychological laboratories in Europe (in 1894). His scientific oeuvre consists of about 200 publications, especially about objects and theory of objects, judgments and assumptions, values, representations and psychology. In the early 20th century, Meinong and his findings were discussed by Bertrand Russell in philosophy and by the Gestalt psychologists in psychology (Stock & Stock, 1990, pp. 1264-1265). Our informetric research questions are: Is there a knowledge domain of Meinong research in recent decades? If yes, what are the top journals and authors, institutions, what terminology is used in Meinong research, are there dominating publication languages, and how are the documents of Meinong research connected?

In *WoS* we performed a search for CITED AUTHOR=MEINONG, A*, which led to 257 records in April, 2006. In fig.1 we see a list of documents sorted by times cited which cited at least one work by Alexius Meinong. The user easily identifies the relevant documents with the greatest impact. Keep in mind that the number one article by Gardner was published in 1961, and the article by Salmon (rank three) appeared in 1998. Due to the time difference the chance of being cited is much higher for the Gardner article than for the Salmon article.

Fig. 2 illustrates *WoK*'s Analyze function for ranking by country of the number of articles citing Meinong. 37.7 % of the papers have a USA correspondence address, 7.0 % an address in Canada and 5.4 % in Austria. The user can sort the items by rank (as in fig. 2) or alphabetically or numerically. Using the field „year" and sorting by „selected field" the user creates a time series of the marked documents. The Analyze function has limited resources. A maximum of 100,000 records can be processed.

The sort option and the analyze options of *WoK* are easy to use. They allow informetric analyses for everyone, „informetrics light". *HistCite* can provide analyses of much larger collections and provides additional capabilities including editing functions. Of course, the creation of historiographs with variable thresholds of inclusion was its original raison d'être.

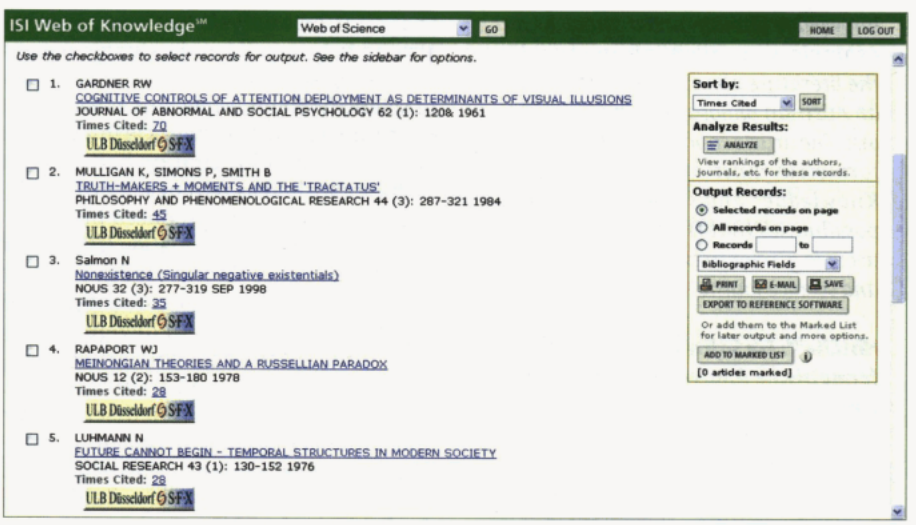

Figure 1: Web of Knowledge's sorting option by times cited

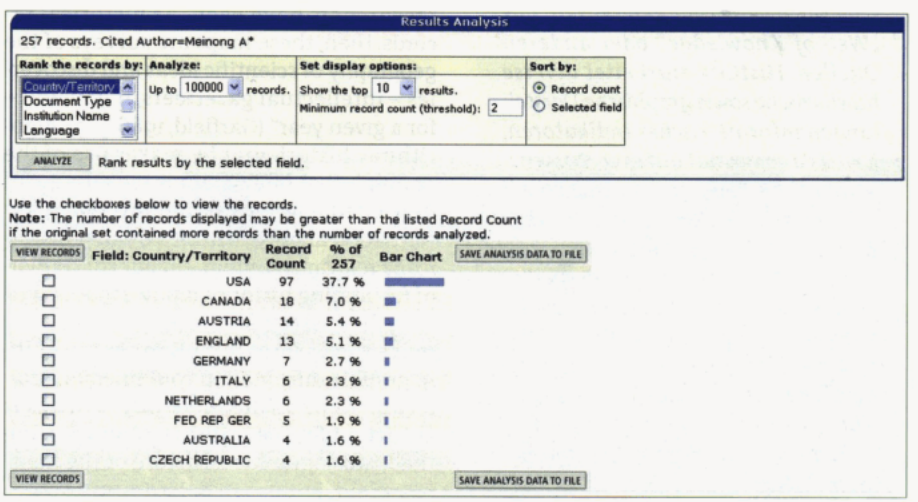

Figure 2: Web of Knowledge's analyze options

Table 2: Main functionality of HistCite (version 2006.02.28)

Function	
ranking documents of a knowledge domain	by local cited references (LCR) by the number of cited references (NCR) by global citation score (GCS) by local citation score (LCS) alphabetically by title alphabetically by author name alphabetically by journal title by date
ranking journals of a knowledge domain	by the number of documents in the domain (PUBS) by total global citation score (TGCS) by TGCS per year by total local citation score (TLCS) by TLCS per year by the total number of cited references (TLCR) alphabetically by journal title
ranking authors of a knowledge domain	by the number of documents in the domain (PUBS) by total global citation score (TGCS) by TGCS per year by total local citation score (TLCS) by TLCS per year by TLCR in the beginning of the collection by TLCR in the end of the collection by the total number of cited references (TLCR) alphabetically by author name
ranking outer references	by local citation score (LCS) alphabetically by title alphabetically by author name alphabetically by journal title by date
ranking title terms of the documents of knowledge domain	by the number of documents in the domain a (PUBS) by total global citation score (TGCS) of the documents by total local citation score (TLCS) of the documents alphabetically
ranking years	by the number of documents in the domain (PUBS) graphically (histogram)
ranking formal aspects (year, document type, language, institution, country) of papers of a knowledge domain	by the number of documents in the domain (PUBS) by total global citation score (TGCS) by total local citation score (TLCS) alphabetically
citation matrix	nodes – (local) references – (local) citations
information flow in a knowledge domain	visualization of the citation matrix

Functionality of *HistCite*

HistCite (Garfield, 2004; Garfield, Istomin & Pudovkin, 2002; Garfield & Pudovkin, 2004; Garfield, Pudovkin & Istomin, 2002; Garfield, Pudovkin & Istomin, 2003a; Garfield, Pudovkin & Istomin, 2003b) is a tool for analyzing and visualizing direct citation linkages between two or more documents. Its inputs can be records saved from citation-based databases including Thomson Scientific's citation indexes in *WoK* (or in *DIALOG*). Records from *Scopus*, *CAS* or *Medline* could be analysed provided the export format is compatible with *HistCite* . However, Medline does not include cited references. These can be added by linking to *WoK*.

If the imported records represent a knowledge domain (collection), the user is able to rank documents, journals, authors, institutions, words etc. of the knowledge domain by the number of local (i.e., inside the knowledge domain) cited references and citations and by the number of global (i.e., inside and outside the knowledge domain) citations. Tab. 2 is a list of the main functionality of *HistCite* .

There are various descriptions of knowledge domains on the *HistCite* 's Web site (www.HistCite.com) including some applications of *HistCite* in scientometric analyses. For example, F. Byrne and S. Chapman (2005) analyzed the scientific literature of tobacco control (using a sample of 9,745 papers) with the help of *HistCite*.

Continuing the Meinong example, all records citing Meinong (N = 257) were collected in a marked list and exported from *WoK* (including all the references cited in each document) and imported into *HistCite*. *WoK* allows only marked lists with up

to 500 records. If there are more items, a searcher has to divide the search question in slices which produce maximal 500 records, and then import slice to slice into *HistCite*, where there is essentially no limitation on the number of records. The actual limit depends upon computational power. We learn from *HistCite*'s main screen (fig. 3), that the knowledge domain of Meinong research, as documented in *WoS*, has a time span from 1959 to 2005 and consists of 198 different authors, who have published their findings in 126 journals using 734 different title terms.

The data in the records are not always „clean". Due to variant spellings, mistakes made by the citing authors or errors during the indexing process, bibliographic citations related to the same item are sometimes expressed differently. *HistCite* provides a list with so called „missed citations" (fig. 4), i.e. references which are similar to other „clean" references. It is possible for the user to join the variants in one node. In fig. 4, all seven missed citations refer correctly to other nodes. This permits the user to make appropriate unifications.

There are many possibilities for unifying and cleaning up variants and errors in the input data. However, if the cited author has published more than one paper in a given year it may be necessary to go into *WoK* to learn which paper has in fact been cited and then that citing record can be corrected. In certain cases it is possible to do „global corrections" in which a repeated variant can be changed with one command. In the case of cited books, the user must decide whether individually cited pages will be retained or unified so that the full citation count for the cited book is obtained.

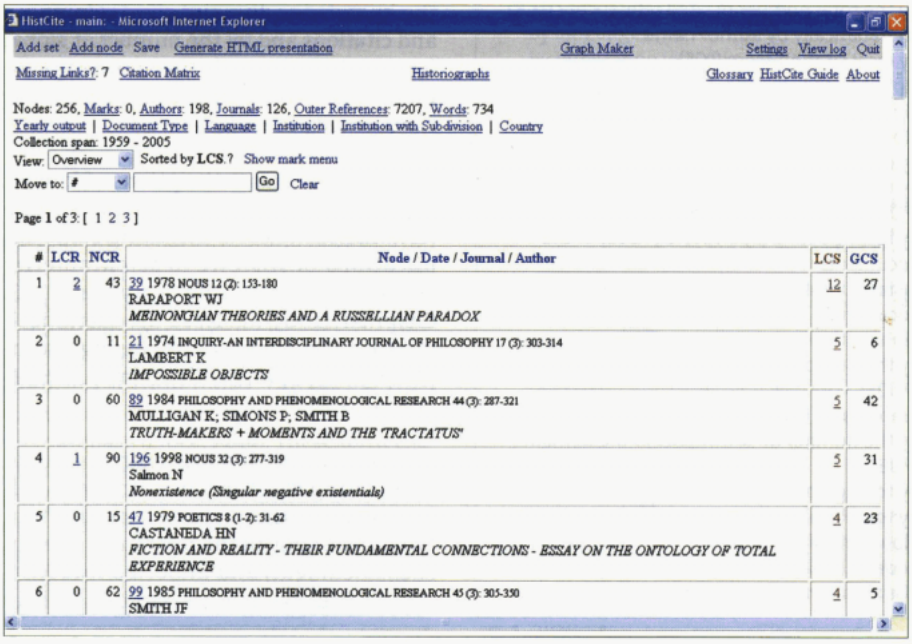

Figure 3: The main screen of HistCite

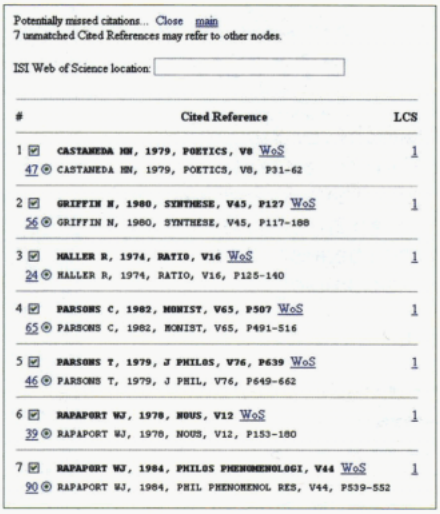

Figure 4: Proposal of adding „missing links" to the node list

Analyzing the knowledge domain by document, journal and author rankings

Every document of the knowledge domain is described by its references and citations inside and outside the domain. For all the documents inside the knowledge domain, the user is able to sort alphabetically by author name, journal and node, chronologically by publication date, by the count of locally cited papers (LCR), by the number of references (NCR), by Local Citation Score (LCS) and by Global Citation Score (GCS). Concerning LCS, the top cited documents inside the Meinong research domain are (fig. 3):

1st Rapaport, W.J., 1978, Nous, 12 LCS : 12,
2nd Lambert, K., 1974, Inquiry, 21 LCS : 5,
Mulligan, K.; Simons, P.; Smith, B., 1984, Philosophy and Phenomenological Research, 44 LCS: 5,
Salmon, N., 1998, Nous, 32 LCS: 5,
5th Castaneda, H.N., 1979, Poetics, 8 LCS: 4,
Smith, J.F., 1985, Philosophy and Phenomenological Research, 45 LCS: 4.

The top locally cited article by Rapaport contains 43 references (NCR), from which two are references to other works inside the collection(LCR).

GCS citations include not only those from inside the collection but also those from outside, i.e. the "times cited" data of *WoS*. The top five GCS records follow:

1st Gardner, R.W., 1961, Journal of Abnormal and Social Psychology, 62 GCS: 70
2nd Mulligan, K.; Simons, P.; Smith, B., 1984, Philosophy and Phenomenological Research, 44 GCS: 42,
3rd Salmon, N., 1998, Nous, 32 GCS: 31,
4th Luhmann, N., 1976, Social Research, 43 GCS: 28,
5th Rapaport, W.J., 1978, Nous, 12 GCS: 27.

Some top documents of the LCS list (e.g., Mulligan et al. 1984, Salmon 1998 and Rapaport 1978) are in the GCS top list as well. Their global citation scores are much higher than the local ones, because they take into account citations from outside the collection. Perhaps they "export" domain-specific insider-knowledge into other domains. The situation is different when highly cited papers (such as Gardner 1961 with a GCS of 70) have only a small LCS (the Gardner article has a LCS of 1). Here we can assume, that there is little relatedness between these articles. Keep in mind that the GCS score is taken directly from *WoS* where it is called "times cited" for that paper. GCS scores will often be quite high when the work in question has multidisciplinary impact.

The source (journal) list ranking options (fig. 5) allow for sorting by the number of publications as well as by the count of the cited papers within the collection (TLCR), by the total LCS and GCS and by TLCS and TGCS per year. The score per year shows the average citation score since the publication date. Fig. 5 is sorted by the Total Citation Score per year of the journal title in the knowledge domain. Here the top journal is „Nous" with an average of 1.32 locally cited documents per year. Counting all locally cited papers, „Nous" is (with a score of 23) ranked first, too. Analyzing the global impact of the journals, „Nous" is ranked 2nd (with a TGSC of 72 and a TGSC/t of 5.57) just behind „Philosophy and Phenomenological Research". The domain specific literature production of „Nous" is (with 10 publications) smaller than that of „Philosophy and Phenomenological Research" (16 publications) and that of „Topoi" (13 publications). In the 10 papers of „Nous",

HistCite - Source list: - Microsoft Internet Explorer

Ranked Source list Close main Glossary HistCite Guide
Total: 126, Nodes: 256, Marks: 0
TLCS: 81, TGCS: 711, TNCR: 8498, mean LCS: 0.64, mean GCS: 5.64, mean NCR: 67.44
View: Bibliometric Sorted by LCS/t ? Show mark menu

#	Name	Pubs	Percent	TLCS	TLCS/t	TGCS	TGCS/t	TLCR
1	NOUS	10	3.9	23	1.32	72	5.57	7
2	PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH	16	6.3	13	0.75	76	5.99	7
3	TOPOI-AN INTERNATIONAL REVIEW OF PHILOSOPHY	13	5.1	9	0.42	54	5.08	8
4	DIALECTICA	6	2.3	2	0.40	9	0.82	1
5	COGNITIVE SCIENCE	2	0.8	4	0.26	29	1.88	9
6	POETICS	2	0.8	6	0.23	26	0.98	0
7	INQUIRY-AN INTERDISCIPLINARY JOURNAL OF PHILOSOPHY	1	0.4	5	0.16	6	0.19	0
8	GERMANIC REVIEW	1	0.4	2	0.10	2	0.10	0
9	ANALYSIS	2	0.8	2	0.07	6	0.23	0
10	JOURNAL OF PHILOSOPHY	1	0.4	2	0.07	6	0.22	4
11	AUSTRALASIAN JOURNAL OF PHILOSOPHY	2	0.8	1	0.07	4	0.25	1
12	SYNTHESE	5	2.0	2	0.06	23	0.77	0
13	PHILOSOPHY AND LITERATURE	2	0.8	1	0.05	3	0.16	1
14	HUSSERL STUDIES	8	3.1	1	0.05	3	0.26	3
15	KODIKAS CODE-ARS SEMEIOTICA	2	0.8	1	0.05	1	0.05	1
16	PSYCHE-ZEITSCHRIFT FUR PSYCHOANALYSE UND IHRE ANWENDUNGEN	1	0.4	1	0.05	5	0.24	0
17	THEORETICAL LINGUISTICS	1	0.4	1	0.05	1	0.05	1
18	THEORY AND DECISION	2	0.8	1	0.04	10	0.32	0
19	AMERICAN PHILOSOPHICAL QUARTERLY	5	2.0	1	0.03	23	0.82	1

Figure 5: Journals by average local citation score per year

there are seven references to documents inside the knowledge domain. What are the top journals in Meinong research , domain specific references and (local and global) citations? There is a clear result: „Nous", „Philosophy and Phenomenological Research" and „Topoi".

Is it possible to identify important authors of Meinong research? Looking at fig. 6 we see two further indicators: the Local Citation Score in the beginning (TLCSb; the time span since the beginning of the collection until an arbitrary cutoff year) and the Local Citation Score in the end (TLCSe; the time span since an arbitrary cutoff year and the last year of the collection). The most productive author in Meinong research is B. Smith with 13 publications, of which four were written in recent years. These are the top authors in the Meinong knowledge domain:

Authors

by publications	by TLCS	by TGCS
Smith B (13)	Rapaport WJ (22)	Smith B (112)
Rapaport WJ (7)	Smith B (11)	Gardner RW (82)
Chisholm RM (6)	Mulligan K (8)	Rapaport WJ (69)
Haller R (5)	Parsons T (8)	Mulligan, K (60)
Jacquette D (4)	Simons P (7)	Simons P (45)
Modenato F (4)	Lambert K (6)	Chishom RM (36)
Parsons T (4)	Salmon N (5)	Salmon N (31).

Most of the top scorers in the three rankings appear in all lists (with the exception of the one who is alien to Meinong research, R.W. Gardner). The main article authors of contemporary Meinong research seen by production and influence are W.J. Rapaport and B. Smith.

#	Name	Pubs	Percent	TLCS	TLCS/t	TGCS	TGCS/t	TLCR	TLCSb	TLCSe
1	Rapaport WJ	7	2.7	22	0.95	69	4.44	18	1	2
2	Smith B	13	5.1	11	0.64	112	10.81	7	0	4
3	Salmon N	1	0.4	5	0.63	31	3.88	1		5
4	Chrudzimski A	3	1.2	2	0.40	2	0.40	3		2
5	MULLIGAN K	3	1.2	8	0.38	60	2.76	2		3
6	SIMONS P	3	1.2	7	0.31	45	2.04	0		2
7	PARSONS T	4	1.6	8	0.30	21	0.78	4	2	1
8	LAMBERT K	2	0.8	6	0.20	7	0.23	2	2	0
9	SMITH JF	1	0.4	4	0.19	5	0.24	0		0
10	Wiebe JM	2	0.8	2	0.18	9	0.91	4		1
11	Varzi AC	1	0.4	1	0.17	21	3.50	0		1
12	HEIDSIECK A	3	1.2	3	0.15	5	0.26	4		0
13	CASTANEDA HN	1	0.4	4	0.15	23	0.85	0	0	1
14	Shapiro SC	1	0.4	1	0.11	7	0.78	3		1
15	FINE K	1	0.4	2	0.08	13	0.54	1		0
16	Kroon FW	2	0.8	1	0.07	1	0.07	5		1
17	LINSKY B	1	0.4	1	0.07	3	0.20	1		1
18	ZALTA EN	3	1.2	1	0.07	7	0.51	3		1
19	CHISHOLM RM	6	2.3	2	0.06	36	1.04	0	0	0

Figure 6: Authors by average local citation score per year

If you refer to www.HistCite.com you can find a number of interesting examples of what can be done including topics like Small World Theory (Milgram), Matthew Effect (R.K. Merton), etc. In fact, as is well known most fields are closely identified with a single author. That is why our inventory of „topics" primarily consists of *HistCite* collections of papers by highly productive scholars and scientists. Thus, there are literally dozens of *HistCites* for Robert K. Merton, the eminent sociologist of science, as well collections for Albert Einstein, Francis Crick of Watson-Crick fame, etc. In addition a field may often be defined by a particular journal. Therefore we assembled *HistCite* collections for journals such as Cell Death and Differentiation, Scientometrics, JASIST, Library Quarterly, etc.

Analyzing outer references

On the level of the single document one can identify those references which are not part of the retrieved knowledge domain, for example monographs (which are not included as sources in *WoK*), old journal documents and journal articles which are not covered by *WoK*. It is possible to sort those "outer references" alphabetically by author name, journal or node, chronologically or by their citation score inside the knowledge domain (LCS). In the Meinong collecton the outer references with the highest LCS are:

1st Parsons, T., 1980, Nonexistent Objects (book) LCS: 30,
2nd Findlay, J.N., 1963, Meinongs Theory of Objects (book) LCS: 27,
3rd Russell, B., 1905, Mind, 14 (article) LCS: 20,
4th Meinong, A., 1960, Realism (book) LCS: 18,
5th Meinong, A., Möglichkeit und Wahrscheinlichkeit (book) LCS: 17.

It is possible to include outer references in the knowledge domain (via „make node"). From fig. 7 we learn that there is another important contemporary author in Meinong research: T. Parsons with his book „Nonexistent Objects". All in all the top outer references are important sources for understanding the philosophy of Alexius Meinong.

One can also include papers that cite the works in outer references. To do this you must go back into *WoK*, perform a cited reference search on that outer reference, and there you will find many papers which may be related to Meinong research.

We have imported all papers citing "Nonexistent Objects" (N=142; outer reference #1) and "Meinong's Theory of Objects" (N=47; #2) to test the differences. The list of the main journals (sorted again by TLCS/t) shows indeed differences to the ranking of fig. 5: Now we see "Nous", Topoi" and – new – "Synthese" as the leading journals. There are differences in the rankings of the main authors as well. Sorting authors by TLCS/t the top author W.J. Rapaport remains first, N. Salmon ranks second (formerly third) and T. Parsons third (formerly seventh). It seems to be a crucial decision to the quality of the research results to include the „right" outer references into the analysis. (The follow-

#	Author / Year / Journal / Reference	Pubs	Percent
1	PARSONS T, 1980, NONEXISTENT OBJECTS WoS make node	30	11.7
2	FINDLAY JN, 1963, MEINONGS THEORY OBJE WoS make node	27	10.5
3	RUSSELL B, 1905, MIND, V14, P479 WoS make node	20	7.8
4	MEINONG A, 1960, REALISM BACKGROUND P, P76 WoS make node	18	7.0
5	MEINONG A, 1915, MOGLICHKEIT WAHRSCHE WoS make node	17	6.6
6	RUSSELL B, 1903, PRINCIPLES MATH WoS make node	13	5.1
7	MEINONG A, 1904, UNTERSUCHUNGEN GEGEN WoS make node	12	4.7
8	MEINONG A, UBER ANNAHMEN make node	12	4.7
9	CASTANEDA HN, 1974, PHILOSOPHIA, V4, P3 WoS make node	11	4.3
10	LAMBERT K, 1983, MEINONG PRINCIPLE IN WoS make node	11	4.3
11	MEINONG A, 1960, REALISM BACKGROUND P WoS make node	11	4.3
12	MEINONG A, MOGLICHKEIT WAHRSCHE make node	11	4.3
13	BRENTANO F, 1924, PSYCHOL EMPIRISCHEN WoS make node	10	3.9
14	CHISHOLM RM, 1960, REALISM BACKGROUND P WoS make node	10	3.9
15	EVANS G, 1982, VARIETIES REFERENCE WoS make node	10	3.9
16	GROSSMANN R, 1974, MEINONG WoS make node	10	3.9
17	HUSSERL E, LOGISCHE UNTERSUCHUN make node	10	3.9
18	PARSONS T, 1974, J PHILOS, V71, P561 WoS make node	10	3.9
19	RUSSELL B, 1919, INTRO MATH PHILOS WoS make node	10	3.9
20	TWARDOWSKI K, 1977, CONTENT OBJECT PRESE WoS make node	10	3.9

Figure 7: Outer references by citations from the knowledge domain

ing figures were calculated on the base of the original set, i.e. without the outer references.)

Analyzing title terms

The ranked word lists of *HistCite* sort the vocabulary taken from the document titles, excluding stop words and all words with two characters and less. The software allows one to rank the words alphabetically, by the count of the publications in which they are title terms and by TLCS / TGSC of their publications.

It is not surprising that the term „Meinong" is the top scorer of the title terms (fig. 8). Ranked second and third, we see „theory" and „objects". Since *HistCite* does not yet recognize phrases, it is possible that there is only one concept „theory of objects". *HistCite* does not yet make use of conflation algorithms, so „objects" and „object" are two separate entries in the list.

The database „Graz school" (Stock & Stock,1990; Stock, 1989) consists of a (more or less) complete collection of publications by and about Meinong and his students (until 1989). Text-word method, developed by Norbert Henrichs (1970; see also Stock 2000), was used for indexing. In the bibliography of the Graz school there is a list with all main topics of the papers of the complete Meinong research (N = 1.210; Stock & Stock, 1990, p. 1297). So we can compare the complete list with the results of *WoK* and *HistCite* . The following list is derived from the first 22 terms of the *HistCite* output. An asterisk marks those words which also occur in the list by Stock & Stock among the first 22 entries.

- Meinong (Meinongian, Alexius)*,
- theory of objects (object, objects)*,
- logic*,
- Husserl*,
- fiction (fictional),
- ontology,
- knowledge,
- psychology*,
- Brentano*,
- existence*,
- philosophy*,
- value*,
- language,
- phenomenology,
- Russell*,
- representation*,
- semantics.

There are six topics in the *HistCite* list, which are not mentioned by Stock & Stock. Since this database was closed with the year 1989,

#	Word	Pubs	Percent	TLCS	TGCS
1	MEINONG	30	11.7	10	20
2	THEORY	25	9.8	13	51
3	OBJECTS	23	9.0	8	61
4	LOGIC	15	5.9	3	37
5	HUSSERL	12	4.7	0	6
6	ONTOLOGY	10	3.9	7	37
7	KNOWLEDGE	9	3.5	1	25
8	PSYCHOLOGY	9	3.5	0	11
9	BRENTANO	8	3.1	3	15
10	EXISTENCE	8	3.1	0	20
11	MEINONGIAN	8	3.1	17	38
12	PHILOSOPHY	8	3.1	0	4
13	VALUE	8	3.1	0	6
14	ALEXIUS	7	2.7	0	0
15	LANGUAGE	7	2.7	2	12
16	PHENOMENOLOGY	7	2.7	1	4
17	RUSSELL	7	2.7	1	19
18	FICTION	6	2.3	6	29
19	FICTIONAL	6	2.3	2	7
20	OBJECT	6	2.3	2	4
21	REPRESENTATION	6	2.3	3	40
22	SEMANTICS	6	2.3	1	8

Figure 8: Title terms by count of publications

these six terms may be from titles of newer papers. All top-five terms of the Stock & Stock list are in the *HistCite* list, too. The *HistCite* list seems to give a good approximation to the „true" term list. *HistCite* would benefit from further (automatic) processing steps like phrase identification and term conflation and from processing not only title terms, but also the words of the abstracts. *HistCite* is a work in progress. The developers have made dozen of changes since its first implementation. Phrase identification is one of many further modifications that we hope to make in the future.

Analyzing formal aspects

There are many options to analyze formal aspects of the literature of the knowledge domain by the count of publications and by (local and global) citation scores. *HistCite* offers tools for analyzing publication years (fig. 9), document types (fig. 10), publication languages (fig. 11), institutions (fig. 12) and countries (fig. 13). For publication years there is a view in form of a table (analogous to figures 10 to 13) and a graphical form of a histogram (fig. 9).

The Meinong literature is mainly written in English (71.5 %) and German (11.3 %), but the English documents gather 93.8 % of all local citations and the German ones only 6.2 %. All documents in other languages are not cited locally. By the count of publications in Meinong research Salzburg University (Austria), Graz University (Austria) and University of Manchester (England) are very productive. But the people of SUNY Fredonia

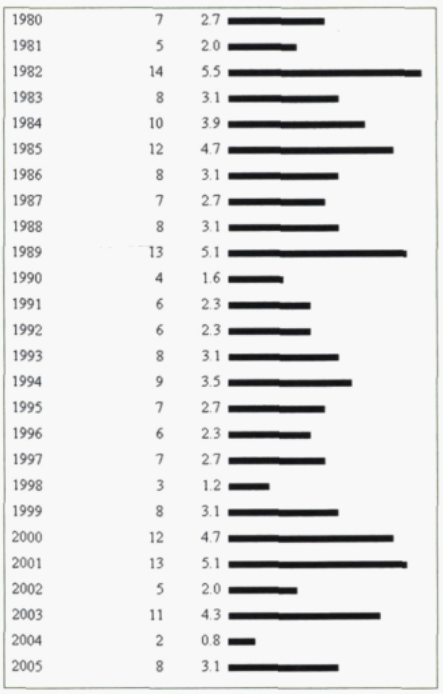

Figure 9: Publication years

#	Document Type	Pubs	Percent	TLCS	TGCS
1	Article	232	90.6	79	689
2	Book Review	13	5.1	0	1
3	Review	5	2.0	0	17
4	Editorial Material	3	1.2	0	0
5	Note	2	0.8	2	3
6	Discussion	1	0.4	0	1

Figure 10: Document types

have written the most influential documents (with a Total Local Citation Score of 16). Following the links to the documents we see that there is only one famous scientist, W.J. Rapaport.

#	Language	Pubs	Percent	TLCS	TGCS
1	English	183	71.5	76	679
2	German	29	11.3	5	23
3	French	15	5.9	0	3
4	Italian	12	4.7	0	1
5	Spanish	7	2.7	0	3
6	Czech	6	2.3	0	1
7	Russian	2	0.8	0	0
8	Flemish	1	0.4	0	0
9	Slovak	1	0.4	0	1

Figure 11: Publication languages

Analyzing formal aspects with the help of *HistCite* is similar to the options of *WoK* (fig. 2). Both present (some) indicators with a bar chart. But there are differences. *HistCite* offers sort options by citation scores which are not realized in *WoK*. In all *HistCite* provides total local and global citation counts TLCS (and TGCS) per publication and rankings by these indicators.

Each new *HistCite* collection presents intriguing perspectives on the topic or scholar involved. We have rarely been disappointed in the results obtained. Almost all of our collections have been shared with the individual scholars where possible. Perhaps the most common comment from highly productive scientists is that the overall collection needs to be broken down into separate categories. Thus, for the work of Bruce Alberts, the recent president of the *National Academy of Sciences*, it was important to place his papers in seven different categories, corresponding to the different fields of his research over many decades.

#	Institution	Pubs	Percent	TLCS	TGCS
1	SUNY COLL FREDONIA	3	1.2	16	32
2	UNIV MASSACHUSETTS	4	1.6	10	23
3	Salzburg Univ	8	3.1	9	50
4	UNIV MANCHESTER	6	2.3	9	54
5	SUNY Buffalo	5	2.0	6	48
6	UNIV CALIF	1	0.4	5	6
7	Univ Calif Santa Barbara	1	0.4	5	31
8	UNIV HAMBURG	1	0.4	5	42
9	Unknown	61	23.8	5	154
10	INDIANA UNIV	4	1.6	4	25
11	UNIV CALIF IRVINE	2	0.8	3	4
12	Univ Geneva	2	0.8	3	10
13	UNIV SO CALIF	2	0.8	3	5
14	BROWN UNIV	5	2.0	2	21
15	UNIV MICHIGAN	2	0.8	2	14
16	Columbia Univ	2	0.8	1	32
17	Graz Univ	6	2.3	1	3
18	MCMASTER UNIV	4	1.6	1	13
19	NEW MEXICO STATE UNIV	1	0.4	1	7
20	STANFORD UNIV	3	1.2	1	7

Figure 12: Top cited institutions

In other cases, as e.g. Einstein or Watson-Crick it is essential to limit results to certain historical periods so that one can see the year by year growth in interest of the topic. And in the case of Watson-Crick we demonstrated the intimate link between their work and that of Avery et al. even though they admittedly did not cite their work in the classic paper on the double helix structure of DNA.

The citation matrix and its visualization

The most impressive feature of *HistCite* is its visualization capability (fig. 15). The Historiograph is the main production from *HistCite* and provides a snapshot of the evolution of the topic and highlights the core works.

A very important feature is the visualization of the information flows in a knowledge domain. In fig. 15 we see all documents of the Meinong knowledge domain which are cited locally at least twice. We can identify both "hub" documents (documents citing many other documents in the domain) as well as "authority" documents (those that are being cited by many other domain-specific documents) (Kleinberg, 1999; for importing Kleinberg's terminology into scientometrics

see Schloegl & Stock, 2004, pp. 1159-1160). Hub documents of the contemporary Meinong research (without consideration of the outer references) are Rapaport 1985 (no. 107), Rapaport 1986 (no. 97) and Parsons 1979 (no. 46), authorities are Lambert 1974 (no. 21), Parsons 1978 (no. 38) and Rapaport 1978 (no. 39).

#	Country	Pubs	Percent	TLCS	TGCS
1	USA	97	37.9	58	352
2	Austria	14	5.5	10	53
3	UK	16	6.3	9	76
4	Germany	7	2.7	6	48
5	Unknown	61	23.8	5	154
6	Canada	18	7.0	4	55
7	Switzerland	2	0.8	3	10
8	New Zealand	4	1.6	1	3

Figure 13: Top cited countries

cited nodes	LCR	NCR	Nodes	LCS	GCS	citing nodes
	0	38	31 1977 HARTMANN K	0	0	
	0	22	32 1977 MARGOLIS J	0	0	
	0	11	33 1977 HUSSERL EG	0	0	
	0	28	34 1977 GRIFFIN N	0	2	
	0	20	35 1977 MARGOLIS J	0	5	
	0	11	36 1978 SMITH Q	0	1	
	0	22	37 1978 ROLLMANN H	0	0	
	0	8	38 1978 PARSONS T	3	7	39 46 74
21 38	2	43	39 1978 RAPAPORT WJ	12	27	46 90 97 107 128 146 147 154 157 161 188 233
	0	22	40 1978 [ANON]	0	0	
	0	10	41 1979 RAPAPORT WJ	2	3	97 107
	0	7	42 1979 SMITH B	0	0	
	0	24	43 1979 ROUTLEY R	0	1	
	0	8	44 1979 STACK GJ	0	0	
	0	8	45 1979 PIETERSMA H	0	0	
5 21 38 39	4	26	46 1979 PARSONS T	2	6	75 157
	0	15	47 1979 CASTANEDA HN	4	23	97 107 157 233

Figure 14: Citation matrix of all papers of a knowledge domain

If we mean by "hub" documents papers such as comprehensive literature reviews, then it is important for the reader to differentiate those from other "core" authority documents which form the basic structure of the field. However, it is important to realize that certain reviews play a critical role in the development of science. In studying the history of a topic, however, it is important to know which papers were possibly critical links but were not necessarily highly cited. This can often be seen in detailed historiographs, as we learned from the very first exercise in 1964 with the History of DNA. If you ranked core documents in small world theory today by citation frequency the founding work of Milgram in 1967 would be drowned out by the more highly cited papers in physics, which have superseded the social network emphasis of that field in recent years.

In *HistCite* 's graph maker there is a provision to export data to the Pajek software (De Nooy, Mrvar & Batagelj, 2005). Pajek is a program for large network analysis and thus for „citation network analysis" (Batagelj 2003). Using Pajek with data from *HistCite* it is possible to analyze citation paths networks, you can perform for example calculations of main paths (Batagelj 2003, 10), of hub and authority papers (Batagelj 2003, 11) and of subnetworks and islands (Batagelj 2003, 13).

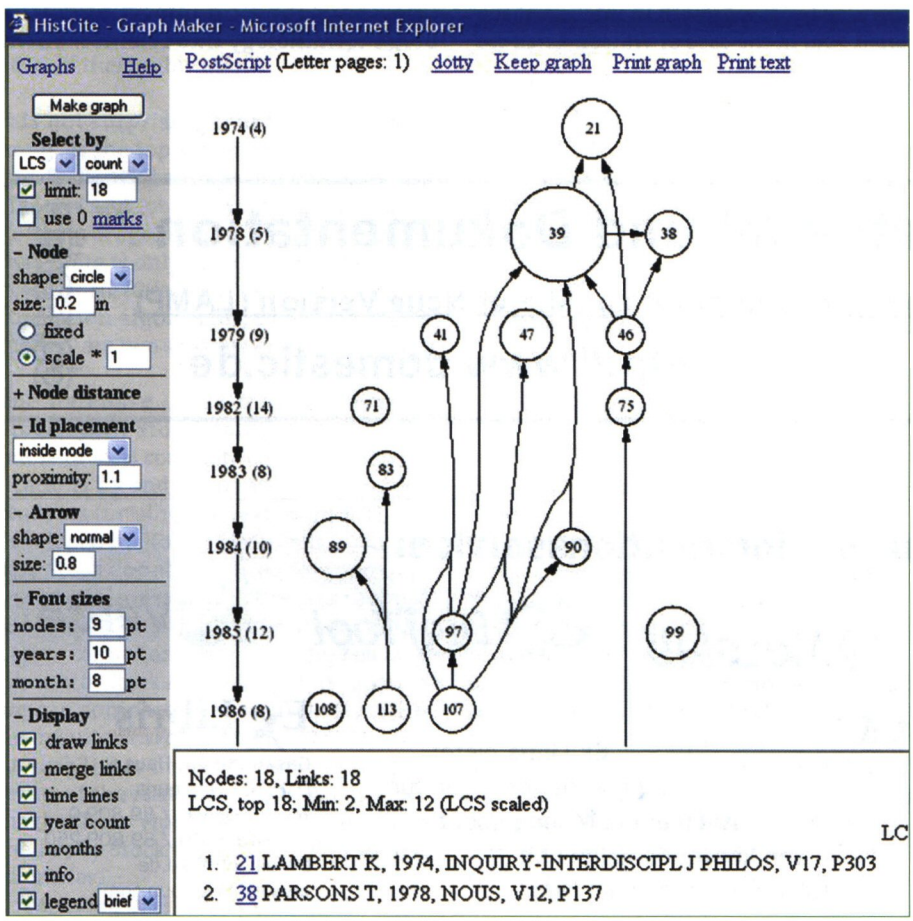

Figure 15: Graph maker

Integrating *HistCite* in the *Web of Knowledge*?

Today, most of the informetric studies are prepared by experts in scientometrics or informetrics. Börner, Chen and Boyack (2003, p. 237) wrote, „despite advances in visualization research, many nonexperts find the use of visualization tools to be alien and nonintuitive". They recommend increasing „the accessibility of domain visualization for nonexperts" (p. 327). With an easy-to-use informetric functionality on *WoK* much more people, i.e. also professional end-users, will profit from informetrics.

There is another reason for integrating *HistCite* into *WoK*. The great competitor of Thomson Scientific's *Web of Knowledge* is Reed Elsevier's *Scopus* (Trkulja, 2005). *Scopus* allows sorting of records by citation count – that is its only informetric functionality. With a wide range of analytical tools *WoK* would have a competitive advantage over other commercial information suppliers.

Thomson Scientific is well aware of *HistCite* 's capabilities. It is one thing to provide an offline capability in the hands of the individual user and another to provide the same capability to thousands of users simultaneously online. Of course eventually ISI can work out an optimum integration of *WoK*'s Analyze function with *HistCite* 's unique capabilities.

While you as an information specialist are interested in detailed informetrics capabilities the average user, scientist or scholar, does not ordinarily need this type of sophistication. Thus, two modes of operation of *HistCite* are included in the software. The first is aimed at the average user, while the second is for the more sophisticated user who is perhaps more "citation conscious" (Garfield & Stock, 2002). In our experience it is only when scientists become more mature that they take a deeper interest in the history of their topics. However, *HistCite* has proven extremely useful even for journalists and is regularly used e.g. at The Scientist magazine in the research on hot papers and other analyses of individual scientists and institutions.

Literature

Batagelj, V. (2003): Efficient algorithms for citation network analysis. Prep. Ser. Univ. Ljubljana, Inst. Math., 41(897), 1-29.

Börner, K., *Chen*, C., & *Boyack*, K.W. (2003): Visualizing knowledge domains. Annual Review of Information Science and Technology, 37, 179-255.

Byrne, F., & *Chapman*, S. (2005): The most cited authors and papers in tobacco control. Tobacco Control, 14(3), 155-160.

Chen, C. (2004): Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Science of the United States of America, 101 (Suppl. 1), 5303-5310.

Chen, C. (2006): CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology, 57(3), 359-377.

De Nooy, W., *Mrvar*, A., & *Batagelj*, V. (2005): Exploratory Social Network Analysis with Pajek. – Cambridge: Cambridge Univ. Pr.

Garfield, E. (1979): Citation Indexing – Its Theory and Application in Science, Technology, and Humanities. New York: Wiley.

Garfield, E. (1986): Towards scientography. Current Contents, no. 43, 3.

Garfield, E. (1994): Scientography: Mapping the tracks of science. Current Contents, no. 45, 5-10.

Garfield, E. (2004): Historiographic mapping of knowledge domains literature. Journal of Information Science, 30(2), 119-145.

Garfield, E., *Istomin*, V.S., & *Pudovkin*, A.I. (2002): Process for creating and displaying a publication historiography. Patent Number US 6,728,725. Priority: May 2, 2002.

Garfield, E., & *Pudovkin*, A.I. (2004): The HistCite system for mapping and bibliometric analysis of the output of searches using the ISI Web of Knowledge. Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology.

Garfield, E., *Pudovkin*, A.I., & *Istomin*, V.S. (2002): Algorithmic citation-linked historiography. Mapping the literature of science. Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology, 14-24.

Garfield, E., *Pudovkin*, A.I., & *Istomin*, V.S. (2003a): Why do we need algorithmic historiography? Journal of the American Society for Information Science and Technology, 54(5), 400-412.

Garfield, E., *Pudovkin*, A.I., & *Istomin*, V.S. (2003b): Mapping the output of topical searches in the Web of Knowledge and the case of Watson-Crick. Information Technology and Libraries, 22(4), 183-187.

Garfield, E., *Sher*, I.H., & *Torpie*, R.J. (1964): The Use of Citation Data in Writing the History of Science. Philadelphia, PA: Institute for Scientific Information.

Garfield, E., & *Stock*, W.G. (2002): Citation consciousness (Interview). Password, no. 6, 22-25.

Henrichs, N. (1970): Philosophische Dokumentation. Literatur-Dokumentation ohne strukturierten Thesaurus. Nachrichten für Dokumentation, 21, 20-25.

HistCite . Bibliographic Analysis and Visualization Software. Online: www.HistCite .com.

Kleinberg, J. (1999): Authoritative sources in a hyperlinked environment. Journal of the ACM, 46, 604-632.

Kuhn, T.S. (1962): The Structure of Scientific Revolutions. Chicago: Univ. of Chicago Press.

Schloegl, C., & *Stock*, W.G. (2004): Impact and relevance of LIS journals: A Scientometric analysis of international and German-language LIS journals – Citation analysis versus reader survey. Journal of the American Society for Information Science and Technology, 55, 1155-1168.

Small, H. (1973): Co-citation in the scientific literature. A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24, 265-269.

Small, H. (1977): A co-citation model of a scientific specialty: A longitudinal study of collagen research. Social Studies of Science, 7, 139-166.

Small, H., & *Garfield*, E. (1985): The geography of science: Disciplinary and national mappings. Journal of Information Science, 11, 47-59.

Small, H., & *Griffith*, B.C. (1974): The structure of scientific literatures. I: Identifying and graphing specialties. Science Studies, 4, 17-40.

Stock, M., & *Stock*, W.G. (1990): Psychologie und Philosophie der Grazer Schule. Eine Dokumentation. Amsterdam, Atlanta: Rodopi.

Stock, M., & *Stock*, W.G. (2003): Web of Knowledge. Wissenschaftliche Artikel, Patente und deren Zitationen: Der Wissenschaftsmarkt im Fokus. Password, no. 10, 30-37.

Stock, W.G. (1989): Datenbank „Grazer Schule". Eine Spezialdatenbank im Bereich der Philosophie- und Psychologiegeschichte. Zeitschrift für philosophische Forschung, 43, 347-364.

Stock, W.G. (1999): Web of Science. Ein Netz wissenschaftlicher Informationen – gesponnen aus Fußnoten. Password, no. 7+8, 21-25.

Stock, W.G. (2000): Textwortmethode. Password, no. 7+8, 26-35.

Stock, W.G. (2001): JCR on the Web. Journal Citation Reports: Ein Impact Factor für Bibliotheken, Verlage und Autoren? Password, no. 5, 24-39.

Stock, W.G. (2002): ISI Essential Science Indicators. Forschung im internationalen Vergleich – Wissenschaftsindikatoren auf Zitationsbasis. Password, no. 3, 21-30.

Trkulja, V. (2005): Elseviers Scopus. Weltgrößte Abstracts- und Zitationsdatenbank aus dem wissenschaftlichen Web. Password, no. 5, 37-40.

Note: This paper emerged from questions (by WGS) and answers (by EG). Then all three authors transformed the result into this article.
Publications by Eugene Garfield are available online via www.garfield.library.upenn.edu/pub.html; publications by Wolfgang G. Stock via www.phil-fak.uni-duesseldorf.de/infowiss/content/mitarbeiter/stock.php.

HistCite, citation, citation linkage, knowledge domain, informetrics, bibliometrics, scientometrics, algorithmic historiography of science, Web of Knowledge, Web of Science, citation analysis, scientific document, scientific journal, scientific author, title term, citation matrix, visualization, Meinong research

The Authors

Eugene Garfield

Chairman Emeritus, Thomson ISI; Founding Editor and President, The Scientist. Eugene Garfield is a pioneer in information retrieval systems and inventor of Current Contents (1955), Index Chemicus (1960), Science Citation Index (1964), SSCI (1970), AHCI (1975), and Journal Citation Reports (1975). He is an eclectic science communicator and founding publisher/editor of The Scientist as well as author of over 1,000 articles and books. His *HistCite* system of algorithmic historiography now maps the output of the Web of Science database searches of 35,000,000 articles.

E-Mail: garfield@codex.cis.upenn.edu
Homepage: www.eugenegarfield.org

Soren W. Paris

graduated from West Chester University in 2001 with a B.A. in English. He has been research assistant in information technology to Eugene Garfield for the past three years. He is currently adjusting to life as a father and hopes to return to school to study language, history and culture.

E-Mail:
Soren.Paris@contractor.Thomson.com

Wolfgang G. Stock

is the head of the information science department of the Heinrich-Heine-University Düsseldorf (Germany) since 2003. Between 1995 and 2003 he worked as professor for information science at the University of Applied Sciences Cologne, from 1989 to 1995 he served as the head of the Information Centre of the ifo Institute for Economic Research in Munich. In the 1980s he worked in Graz (Austria) as a free-lance information scientist and lecturer of the Karl-Franzens-University. About 200 publications.

E-Mail:
stock@phil-fak.uni-duesseldorf.de
Homepage:
www.phil-fak.uni-duesseldorf.de/infowiss