Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 75, No. 3 (2008) 591–605
DOI: 10.1007/s11192-007-1880-z

The *h*-index of *h*-index and of other informetric topics

JUDIT BAR-ILAN

Department of Information Science, Bar-Ilan University, Ramat Gan (Israel)

In this paper we examine the applicability of the concept of *h*-index to topics, where a topic has index *h*, if there are *h* publications that received at least *h* citations and the rest of the publications on the topic received at most *h* citations. We discuss methodological issues related to the computation of *h*-index of topics (denoted *h-b* index by BANKS [2006]). Data collection for computing the *h-b* index is much more complex than computing the index for authors, research groups and/or journals, and has several limitations. We demonstrate the methods on a number of informetric topics, among them the *h*-index.

Introduction

The paper entitled "An index to quantify an individual's scientific input" by Jorge Hirsch appeared exactly one year ago in the Proceedings of the National Academy for Science [HIRSCH, 2005B]. Even though it was submitted to Arxiv.org a little earlier (in August 2005, [HIRSCH, 2005A]) the *h*-index defined in the above-mentioned paper has already succeeded in attaining considerable interest both among informetricians and the research community in general.

Received August 8, 2007

Address for correspondence:
JUDIT BAR-ILAN
Department of Information Science, Bar-Ilan University
Ramat Gan, 52900, Israel
E-mail: barilaj@mail.biu.ac.il

0138–9130/US $ 20.00

The *h*-index of a researcher *R* is defined by Hirsch as the unique number *h* such that *R* has *h* publications that received *h* or more citations and all *R*'s other publications received at most *h* citations.

The research community was informed about the *h*-index through a series of news items (e.g., [BALL, 2005; PETERSON, 2005; ROEDIGER, 2006]) and immediately lists of researchers with high *h*-indices began to appear. HIRSCH [2005A, 2005B] listed researchers in physics; PALSBERG [NO DATE] calculated *h*-indices for computer scientists, based on Google Scholar data; and partial lists were produced for physical chemists and life scientists [H-INDEX, 2006]. Of course informetricians [GLÄNZEL & PERSSON, 2005; BAR-ILAN, 2006] and information scientists in general were not left out either [CRONIN & MEHO, 2006]. BUTLER & MCALLISTER [2006] studied the applicability of the *h*-index for researchers in the Social Sciences; SAAD [2006] compared data obtained from the Web of Science (WoS) and Google Scholar for consumer scholars, BORNMANN & DANIEL [2005] studied the relation between the *h*-index and the acceptance of post-doctoral grants; BAR-ILAN & LIN [2006] computed the *h*-indices of highly cited Israeli researchers.

In parallel other types of *h*-indices were being suggested. BRAUN & AL. [2005, 2006] introduced the *h*-index for journals, where instead of counting citations to scientists' works; citations to journals are being counted. SCHUBERT & GLÄNZEL [2006] further developed this idea and found through regression analysis a definite relationship between the IF (impact factor) and the *h*-index for journal. ROUSSEAU [2006A] calculated the *h*-index of *JASIS*. VAN RAAN [2006] and MOED [2005] prefer to consider research groups instead of individual researchers.

BANKS [2006] introduced what he calls the *h-b* index, the *h* index for compounds and topics. Computing the *h*-index for compounds in solid state physics is rather straightforward, since usually the names are unique and the name of the compound will appear in publications related to the compound (see the section on Methodological issues for our reservations regarding this assumption). For topics the issue is more complex, but it is not elaborated in the paper and only one or two word topics are demonstrated there. BANKS [2006] also computes the *m* value suggested by HIRSCH [2005B] under the assumption that researchers publish papers at a steady rate, and each paper gains a constant number of citations each year. Under these assumptions $h \sim mn$, where *n* is the number of years under consideration (the researcher's scientific age for Hirsch and the year when the specific compound was introduced for Banks). The value *m* is the gradient of the assumed linear relation between the *h*-index and the number of years of publishing. Thus *m* seems to be a better measure when comparing researchers or compounds of different scientific age. The topics with the five largest *m* values were published in *Nature* (2006) and reported on PhysicsWeb [DUME, 2006], as the "top five in physics". It is not clear how Banks decided to compute the *h* and *m* values for the specific list of topics and whether by some oversight he has not excluded an even

"hotter" topic (see [NATURE NEWSBLOG, 2006] for comments). In this paper we will elaborate on methodological issues related to computing the *h-b* index (some methodological issues are raised by Meho in the [NATURE NEWSBLOG, 2006]).

To complete our literature review related to the *h*-index, we should mention works discussing the time dependency of the *h*-index (see [EGGHE, 2007; KELLY & JENNIONS, 2006; LIANG, 2006]); its correlation with other measures [SCHUBERT & GLÄNZEL, 2006; COSTAS & BORDONS, 2006; MILLER, 2006]; models for the *h*-index [EGGHE & ROUSSEAU, 2006; GLÄNZEL, 2006; ROUSSEAU, 2006B; BURRELL, 2006]; and suggested improvements [BATISTA & AL., 2006; EGGHE, 2006; SIDIROPOULUS & AL., 2006; IGLESIAS & PECHORROMAN, 2006] and of course some criticism regarding the new measure [PURVIS, 2006].

Methodological issues

How should we define the *h*-index of a topic (denoted *h-b* index by BANKS [2006])? Banks does not provide a formal definition, he simply states: "The *h-b* index is found by entering a topic or compound in the Thomson (ISI) Web of Science database and then ordering the results in terms of citations, by largest first. The *h-b* index is defined as above in the same manner as the *h* index" [BANKS, 2006, P.162].

Here, we shall provide our, more formal definition to the concept of the *h-b* index. The *h-b* index is an analog of the *h*-index, therefore:

> The *h-b index* of a set of documents is the unique number *h-b*, such that there are *h-b* items in the set that received at least *h-b* citations, and all other items in the set received at most *h-b* citations.

This is the easy part of the definition; the major question is how to obtain the set of publications on the given topic? First we see no specific reason to use the Web of Science only for the citation data; any citation database can be used for the task. There are several options for obtaining the set of publications for which the *h-b* index is computed: The set may be compiled by experts (e.g. a bibliography for the topic), it may be the result set for a specific query retrieved from a bibliographic database, or the two methods can be combined (e.g. the set retrieved from a database undergoes peer-review by experts). Note that the database does not have to provide the citation data; this can be obtained separately from a citation database.

In case we decide to search bibliographical databases, there may be several options to choose from. BANKS [2006] used the "Topic" search box on the General Search page of the Web of Science (http://portal.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame), where searches are conducted within titles, keywords and abstracts. Scopus (http://www.scopus.com) also has a comparable search option, the "search for … in article title, keywords, abstracts" option. When using this option, one

has to take into account that not all items in these databases have abstracts (see Meho's comment in [NATURE NEWSBLOG, 2006]) and the keywords assignment is not uniform (they can be assigned by the bibliographical database, chosen freely by the author(s), chosen from a controlled vocabulary or thesaurus used by the publisher, etc.) and keywords are not assigned to all papers. These differences between and within the queried databases may result in missing important publications related to the topic. Note that Google Scholar (http://scholar.google.com/), the third currently available comprehensive citation database does not have an option to search within title, keywords and abstracts only, one can search either only in the title only or in the full text of the publication. Searching in the title of the publication only is too restrictive, while searching in the full text is often too noisy, as we will demonstrate in the section on Google Scholar.

To sum up, it is highly recommended to carry out extensive searches (not only in citation databases), and then ask experts to delineate the set of documents for which the *h-b* index will be computed. In this paper we demonstrate the limitations and capabilities of the Web of Science, Scopus and Google Scholar for the computation of the *h-b* index, and provide values for several informetric topics.

Search terms for the topic

Suppose we decide to collect publications on the topic by searching bibliographical databases. In this case we have to decide on the set of search terms that cover the topic and introduce minimal noise. Let us consider, for example the *h*-index. This seems to be a simple concept, thus it might be enough to search for this term only. First it turns out that the term "*h*-index" is used not only in the informetric sense (Web of Science retrieved papers in geophysics, economics, chemistry, ecology, neurology as well). The situation (at least when looking only for highly cited papers) somewhat improves when the publication years are limited to 2005 and onwards. In this case, Web of Science as of November 24, 2006, retrieves one relevant item with ten citations – [BORNMANN & DANIEL, 2005], three papers with two citations each [BATISTA & AL., 2006; GLÄNZEL, 2006; VAN RAAN, 2006], three items with one citation each and nine items without citations. But where are the original paper of HIRSCH [2005] and the BRAUN & AL. [2005, 2006] papers on the *h*-index for journals? Surely they are relevant to the topic.

To answer this question, we have to take a look at the titles and abstracts of the above-mentioned papers. HIRSCH [2005B] has a short abstract: "I propose the index *h*, defined as the number of papers with citation number $\geq h$, as a useful index to characterize the scientific output of a researcher" and three keywords provided by the author: citation, impact, unbiased. The citation count of this paper is 19. The Braun–Glänzel–Schubert in the *Scientist* [BRAUN & AL., 2005] has no abstract and no keywords, thus one can only search in its title: "A Hirsch-type index for journals"

(which is cited ten times according to WoS). Their *Scientometrics* paper [BRAUN & AL., 2006] has a short abstract: "We suggest that a *h*-type index – equal to *h* if you have published *h* papers, each of which has at least *h* citations – would be a useful supplement to journal impact factors" and two keywords: scientists and rankings (added by WoS, since the original paper does not have keywords) but still no mention of the term "*h*-index". Thus to cover these articles as well, the query should be extended to: *h* index OR Hirsch-type (luckily the WoS syntax had been changed recently, and *h* index covers all occurrences of both 'h' and 'index' in the title, keywords and abstract, including *h*-index). This search retrieved 3,902 items (most of them non-relevant in our context) as compared to 35 items in the previous search. However even this extended search did not succeed in covering the BALL [2005] paper entitled: "Index aims for fair ranking of scientists". The paper has no abstract or keywords, and it has been cited 18 times (all the searches on the topic were carried out on November 24, 2006). Probably the best strategy for covering this item as well is to add its title as a phrase to the search, but this demands thorough acquaintance with the literature on the topic. This is not very difficult in case the topic is around less than a year, but much more difficult for well-established topics, e.g., citation analysis. One can claim that the reason for the non-uniform use of terminology in this case is because the topic is novel, and therefore we had difficulties locating only the first papers in this topic. The preliminary version of Hirsch's paper was submitted to arxiv.org in August 2005, whereas the paper in PNAS was published only in November 2005. The preliminary version of the paper has been cited as well in publications indexed by Thomson (ISI). Since Thomson(ISI) does not index publications from the preprint server at arxiv.org, we can only find the citation count using the cited reference search interface, where this publication appears under different versions of the "cited work": ARXIV ORG E PRINT AR, ARXIVPHYSICS0508025, INDEX QUANTIFY INDIV, INDEX QUANTIFY INDIV 2005, PHYSICS0508025 and PHYSICS0508025 2005 – altogether eight citations. After extensive investigation, using WoS the *h-b* index of *h*-index is 5 as of November 2006 (see Table 1), because papers 1–5 were cited five times or more. All other papers on the *h*-index were cited four times or less in WoS.

Next we turned to Scopus with the same search terms, here out of the 41 documents retrieved for *h*-index published in or after 2005, we found the Bornmann-Daniels paper (cited 8 times), Batista et al. and Glänzel cited twice, four papers cited once and seven items that are yet to be cited. The extended search: *h index OR Hirsch-type* produced 4,432 results. Scopus does not index the news item from *Nature News* [BALL, 2005], it does not index arxiv either, and unlike WoS, currently it does not have a cited reference interface, thus we have no information on the number of times these items were referenced in Scopus-indexed items. The *h-b* index of the topic *h*-index is 3 based on citation data retrieved form Scopus (papers 1, 3 and 4 in Table 1).

Table 1. Papers included in the computation of the *h-b* index of the topic "*h*-index"

	Authors(s)	Title	Source	#cits. WoS	#cits. Scopus	#cits. GS
1.	HIRSCH	An index to quantify an individual's scientific research output	*PNAS*	19	16	21
2.	BALL	Index aims for fair rankings of scientists	*Nature*	18		
3.	BRAUN, GLÄNZEL, SCHUBERT	A Hirsch-type index for journals	*Scientist*	10	8	11
4.	BORNMANN, DANIEL	Does the *h*-index for ranking of scientists really work?	*Scientometrics*	10	8	11
5.	HIRSCH	An index to quantify an individual's scientific research output	*Arxiv*	8		10
6.	GLÄNZEL	On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact	*Scientometrics*			7

Google Scholar searches in the full text of the articles indexed by it, which is a considerable difference when comparing it with the other two, previously discussed citation databases. Both Hirsch papers appear in Google Scholar results already for the query "*h index"* (phrase search), because the term *h* index has 18 occurrences in the *PNAS* paper (and 19 in the preprint), the first one in the second paragraph of the introduction: "Here, I would like to propose a single number, the 'h index,' as a particularly simple and useful way to characterize the scientific output of a researcher." For similar reasons, the Braun, Glänzel and Schubert paper in the *Scientist* is also retrieved for the *"h index"* search among the 308 reported results. In this case the expanded query *h index OR Hirsch-type* produced 58,300 results, and browsing the first 100 results does not seem to add any new items to the list. The article of BALL [2005] is not indexed by Google Scholar, but instead we found that the paper of GLÄNZEL [2006] has been cited seven times according to Google Scholar (see Table 1). Thus, when computing the *h-b* index based on Scopus data items 1,3-6 have to be taken into account. All other items are cited less than 5 times as of November 2006.

To sum up, the *h-b* index of *h*-index is 5 based on WoS and Google Scholar and 3 based on Scopus as of November 24, 2006. In this section we only wanted to demonstrate the complexity and the difficulty in creating appropriate search terms for covering a given topic. Next we discuss the specific pros and cons of WoS, Scopus and Google Scholar for computing the *h-b* index.

Google Scholar

The limitations and the inconsistencies of Google Scholar have been extensively discussed in previous studies [BAR-ILAN, 2008; MEHO & YANG, 2007; JACSO 2005, 2006]. Here we only discuss specific problems related to topic search. When searching for authors, using the *author:"J Doe"* format, the results are listed in decreasing order of citations the listed publications received (with a few exceptions). However when searching for topics, there must be other factors in addition to the number of citations that influence the rankings. Consider, for example a search for "Web impact factor" (the results page is from November 25, 2006) as can be seen in Figure 1. The query is not intended to fully cover the topic; it only illustrates the retrieval problems when using Google Scholar. The first item retrieved is cited 51 times, while the third item 84 times (although the exact phrase "Web impact factor" is not in the title of this item). It is listed before item 5 with the exact phrase in the title, but with seven citations only. The tenth item (see Figure 2) contains the search term only in the reference; however the 11th item contains the term in the text and has received 63 citations compared to the 38 citations of the tenth item. Especially striking is result 13 ("Perspectives on webometrics") with 102 citations and the exact term mentioned in the text 5 times, compared with the Vaughan and Hysen paper, listed as no. 7 with 28 citations and a single appearance of the search term in the text (we chose to compare these two items, since neither of them contain the search phrase in the title). Thus it seems that the ranking is influenced among other factors by the placement and frequency of occurrence of the search terms, together, perhaps with the number of citations and maybe the number of links pointing to the article. Search engine ranking algorithms are secret, thus we can only guess the "ingredients" of the ranking algorithm. Whatever the specific algorithm is, it makes the work of the informetrician extremely challenging. What items should be included and/or excluded? We used this example to illustrate the difficulties in computing the *h-b* index using Google Scholar. For the calculation of the *h-b* index for Web impact factors, we extended the query to "Web impact factor" OR "Web impact factors", downloaded the full results list (237 items), and rearranged the results in decreasing order of citations (Note that this is only possible when there are less than 1,000 results, since Google Scholar only displays 1,000 results). Table 2 lists the publications with 25 citations or more as listed by Google Scholar, the rank of the publication in Google Scholar's result list, the placement of the first time one of the phrases "Web impact factor" or "Web impact factors" appears in the publication and the number of times the phrases appear. The question, which of these publications are closely related to the concept and whether no significant publications are missing can only be decided by experts, unless we decide to consider occurrences in title, abstract and keywords only. We conclude that at this point of time Google Scholar cannot be utilized for exhaustive topic searches.

Figure 1. Results of the search "Web impact factor" – results 1–7

Figure 2. Results of the search "Web impact factor" - results 8–14

Table 2. Results of the search "Web impact factor" OR "Web impact factors" on Google Scholar on November 25, 2006

Authors	Title	Source	Publ. year	TC	Orig. order	Term first appears	# occ. of terms
INGWERSEN	The calculation of Web Impact Factors	*J. Doc*	1998	187	1	title	11
INGWERSEN, BJORNEBORN	Perspective of webometrics	*Scientometrics*	2001	102	47	abstract	9
THELLWALL	Extracting macroscopic information from Web links	*JASIST*	2001	90	85	abstract	8
SMITH	A tale of two web spaces: comparing sites using web impact factors	*J. Doc*	1999	84	2	title	16
CRONIN	Bibliometrics and beyond: some thoughts on web-based citation analysis	*J. Inf. Sci.*	2001	84	164	ref	1
BAR-ILAN	Data collection methods on the Web for infometric purposes: A review and analysis	*Scientometrics*	2001	65	115	text	4
THELWALL	Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting kinks between university web sites	*JASIST*	2002	63	65	text	8
THOMAS, WILLET	Webometric analysis of departments of librarianship and information science	*J. Inf. Sci.*	2000	59	58	text	4
EGGHE	New informetric aspects of the Internet: some reflections-many problems	*J. Doc*	2000	58	76	abstract	3
THELWALL	Web impact factors and search engine coverage	*J. Doc*	2000	56	4	title	7
CRONIN & AL.	Invoked on the Web	*JASIS*	1998	55	193	ref	1
WILKINSON & AL.	Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication	*J. Inf. Sci.*	2003	54	73	ref	3
SMITH, THELWALL	Web Impact Factors for Australasian universities	*Scientometrics*	2002	53	3	title	28
VAUGHAN, THELWALL	Scholarly use of the Web: What are the key inducers of links to journal Web sites?	*JASIST*	2003	53	69	text	8

Table 2. (cont.)

Authors	Title	Source	Publ. year	TC	Orig. order	Term first appears	# occ. of terms
ROUSSEAU	Daily time series of common single word searches in AltaVista and NorthernLight	*Cybermetrics*	1998	52	56	text	2
THELWALL	A web crawler design for data mining	*J. Inf. Sci.*	2001	52	63	ref	4
THELWALL	A comparison of sources of links for academic Web impact factor calculations	*J. Doc*	2002	51	5	title	9
THELWALL	Evidence for the existence of geographic trends in university Web site interlinking	*J. Doc*	2002	38	45	ref	6
THELWALL	The top 100 linked-to pages on UK university web sites: high inlink counts are not usually associated with quality scholarly content	*J. Inf. Sci.*	2002	38	80	text	4
THELWALL	What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation	*Inf. Res.*	2003	37	106	ref	3
THELWALL	Results from a Web Impact Factor crawler	*J. Doc*	2001	31	6	title	7
HERSH	Information Retrieval: A Health and Biomedical Perspective		2003	31	225	text	6
THELWALL, SMITH	Interlinking between Asia-Pacific University Web sites	*Scientometrics*	2002	29	101	ref	2
VAUGHAN, HYSEN	Relationship between links to journal Web sites and impact factors	*Aslib Proc.*	2002	28	44	text	4
THELWALL, HARRIES	The connection between the research of a university and counts of links to its web pages: An investigation based upon a classification of the relationships of pages to the research of the host University	*JASIST*	2003	25	68	ref	6
THELWALL	The responsiveness of search engine indexes	*Cybermetrics*	2001	25	161	ref	1

Web of Science

When searching Web of Science, even for recent publications (1996 an onwards), there are several cases where the publication has an abstract but the abstract is not indexed by the Web of Science, for example Adams' publication "The counting house"

which appeared in *Nature* in 2002 is indexed by Thomson (ISI), but not its abstract which contains the term "citation analysis". This is a highly cited item, cited 57 times as of November 22, 2006, but it is not retrieved for the search "citation analysis" in WoS. Another issue is the keywords added by Thomson (ISI), Keywords Plus where "Keywords Plus are words or phrases that frequently appear in the titles of an article's references, but do not necessarily appear in the title of the article or in a list of author keywords." (http://apps.isiknowledge.com/WoS//help/h_fullrec.htm#keywords_plus). Articles where the search terms appear only in Keywords Plus are retrieved when carrying out a Title, Abstract, Keywords search, for example, when searching for "citation analysis", the paper "Why the impact factor of journals should not be used for evaluating research" by Seglen was retrieved by WoS, even though the search term appears in the references only – it was added as Keywords Plus to the bibliographic record. This item was cited 215 times according to WoS as of November 2006, and is the most cited item for "citation analysis" in the period between 1996 and 2006; however the term "citation analysis" appears only in the references. The paper is about impact factors and studies citations of individual papers versus the impact factor of the journal in which the papers are published. Should this paper belong to the topic?

One can complement the searches through the tedious process of using the Cited Reference interface, which covers citations to unindexed items as well, but to locate them one must know the publication source and has to carry out extensive "detective" work to discover all the citations. As an example consider Jon Kleinberg's paper "Authoritative sources in hypertext environments" that has first appeared in the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA) in 1998 and in 1999 in the *Journal of the ACM*. The small but important difference between the two versions is that the term "citation analysis" appears in the abstract of the proceedings paper, but was removed from the journal paper. We searched both Kleinberg JM (the exact name) and Kleinberg J through cited reference search and came up with the following citations and citation counts which we identified as citations of the above-mentioned proceedings paper (the number of citations are in parenthesis): 9 ANN ACM SIAM S DIS (5), 9 ACM SIAM S DISCR M (2), 9 ANN ACM SIAM S DIS (2), ACM S DISCR ALG ONL (1), ACM SIAM S DISCR ALG (4), ACMSIAM S DISCR ALG (1), P 9 ACM SIAM S DESCR (1), P 9 ACM SIAM S DISCR (60), P 9 ACM SIAM SODA (1), P 9 ANN ACM SIAM S D (114), P ACM SIAM S DISCR A (31), SODA (2), SODA SAN FRANC US (1) – altogether 225 citations. Thus searching cited references is possible (but cannot search for titles only for authors and publication sources), but very tedious and all sorts of variations of publication sources have to be taken into account.

Scopus

We complained about the complexity of searching for cited references in WoS. Scopus did not provide any information on citations that are not indexed by it as of the time of data collection and this is a shortcoming for informetric purposes. It also adds keywords of its own, called *index keywords* – "controlled vocabulary terms from an index or thesaurus added to the database during abstracting and indexing" (from Scopus help). It is not clear from which sources these keywords are selected.

Data collection

Data was collected from two sources, the Web of Science, Scopus using carefully designed queries that are supposed to cover the specific topic. Google Scholar was excluded both because of the difficulty of using it for searches of this type and because of all the specific problems mentioned above. We ran "title, abstract, keyword" searches in both databases, and report here the values of the *h-b* indices for two cases: 1) the searches were limited to title, abstract and author keywords and 2) the searches included KEYWORDS PLUS and/or "index keywords". The usefulness of keywords added by the database and the consistency of these added keywords should be further investigated. All computations were based on publications that appeared after 1995, in order to have a fair comparison between the two citation databases. Scopus provides full citation data only from 1996 and onwards.

Results

We have already calculated the *h-b* index of *h*-index (Table 1). Here we only demonstrate the concept, thus in many of the cases the queries do not cover all aspects of the topic, especially for the queries informetrics, bibliometrics, scientometrics and webometrics, where we only tried to assess the relative popularity of each term.

For example instead of "Web impact factor" OR "Web impact factors" the query "Web impact factor" OR "Web impact factors" OR WIF OR WIFs would have covered more relevant papers. But the extended query included many non-informetric papers as well.

The results appear in Table 3.

Among the topics appearing in Table 2, "impact factor" seems to be most visible, and there are more highly-cited publications that include the term "bibliometrics" than highly cited publications with "scientometrics", "informetrics" or "webometrics".

Table 3: The *h-b* indices of selected topics

Topic	*h*-WoS	*h*-WoS with keywords PLUS	*h*-Scopus	*h*-SCOPUS with index keywords
(journal OR journals) AND ("impact factor" OR "impact factors")	27	28	29	29
"citation analysis" OR "cocitation analysis"	22	26	25	25
(Bradford OR Lotka OR Zipf) AND (law OR laws)	22	23	24	24
"Web impact factor" OR "Web impact factors"	8	9	10	10
web AND "link analysis"	8	8	13	13
bibliometric OR bibliometrics	23	24	24	30
scientometric OR scientometrics	13	15	14	14
informetric OR informetrics	13	13	12	13
webometric OR webometrics	8	10	10	10

In most cases there are almost no differences between WoS and Scopus. There is one exception, "link analysis" – here the *h-b* index based on Scopus is much higher, because it indexes the newsletter SIGIR Forum (http://www.acm.org/sigs/sigir/forum/), whereas WOS does not index this publication. Scopus seems to be using the index keyword "bibliometrics" rather extensively as can be seen from the difference between the *h-b* index with and without the index keywords.

Conclusion

Computing the *h-b* index for topics is a promising direction, since it seems to be a good measure of visibility. However the computation of this measure is far from being straightforward.

References

BALL, P. (2005), Index aims for fair rankings of scientists. *Nature News*, 436 : 900.

BANKS, M. G. (2006), An extension of the Hirsch index: Indexing scientific topics and compounds. *Scientometrics*, 69 (1) : 161–168.

BAR-ILAN, J. (2006), H-index for Price medalists revisited. *ISSI Newsletter*, 2 (1) : 3–5.

BAR-ILAN, J. (2008), Which *h*-index? – A comparison of Web of Science, Scopus and Google Scholar. *Scientometrics*, 74 (2) : 257–271.

BAR-ILAN, J., LIN, A. (2006), Which *h*-index? – A comparison of WoS, Scopus and Google Scholar. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 6–8.

BATISTA, P. D., CAMPITELI, M. G., KINOUCHI, O., MARTINEZ, A. S. (2006), Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68 (1) : 179–189.

BORNMANN, L., DANIEL, H. (2005), Does the *h*-index for ranking of scientists really work? *Scientometrics*, 65 : 391–392.

BRAUN, T., GLÄNZEL, W., SCHUBERT, A. (2005), A Hirsch-type index for journals. *The Scientist*, 19 (22) : 8.

BRAUN, T., GLÄNZEL, W., SCHUBERT, A. (2006), A Hirsch-type index for journals. *Scientometrics*, 69 (1) : 169–173.

BURRELL, Q. L. (2006), Hirsch's h-index: a preliminary stochastic model. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 26–28.

BUTLER, L., MCALLISTER, I. (2006) The Hirsch index: Is it applicable to the Social Sciences? In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 31–32.

COSTAS, R., BORDONS, M. (2006), H-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 189–191.

CRONIN, B., MEHO, L. (2006), Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57 (9) : 1275–1278.

DUME, B. (2006), Hottest topics in physics revealed. *PhysicsWeb*. Retrieved November 4, 2006, from: http://physicsweb.org/articles/news/10/5/4

EGGHE, L. (2006), Theory and practise of the g-index. *Scientometrics*, 69 (1) : 131–152.

EGGHE, L. (2007), Dynamic h-index: the Hirsch index in function of time. *Journal of the American Society for Information Science and Technology*, 58 (3) : 452–454.

EGGHE, L., ROUSSEAU, R. (2006), An informetric model for the Hirsch-index. *Scientometrics*, 69 (1) : 121–129.

GILES, J. (2006), Top five in physics. *Nature News*, 441 : 265

GLÄNZEL, W. (2006), On the h-index – A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67 (2) : 315–321.

GLÄNZEL W., PERSSON, O. (2005), H-index for Price medalists. *ISSI Newsletter*, 1 (4) : 15–18.

h-INDEX. (2006), In: *Wikipedia, The Free Encyclopedia*. Retrieved 16:32, November 20, 2006, from http://en.wikipedia.org/w/index.php?title=H-index&oldid=88067648

HIRSCH, J. E. (2005A), *An Index to Quantify an Individual's Scientific Research Output.* Retrieved November 4, 2006, from: http://arxiv.org/abs/physics/0508025

HIRSCH, J. E. (2005B), An index to quantify an individual's scientific research output. *PNAS,* 102 (46) : 16569–16572.

IGLESIAS, J. E., PECHARROMAN, C. (2006), *Scaling the H-index for Different Scientific ISI Fields*. Preprint. Retrieved November 4, 2006, from http://arxiv.org/ftp/physics/papers/0607/0607224.pdf

INGWERSEN, P. (1998), The calculation of Web impact factors. *Journal of Documentation*, 54 (2) : 236–243.

JACSO, P. (2005), As we may search – Comparison of major features of Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89 (9) : 1537–1547.

JACSO, P. (2006), Deflated, inflated and phantom citation counts. *Online Information Review*, 30 (3) : 297–309.

KELLY, C. D., JENNIONS, M. D. (2006), The h-index and career assessment by numbers. *TRENDS in Ecology and Evolution*, 21 (4) : 167–170.

LIANG, L. (2006), H-index sequence and h-index matrix: Constructions and applications. *Scientometrics*, 69 (1) : 163–169.

MEHO, L. I., YANG, K. (2007), Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the Amercian Societu for Information Science and Technology,* 58 (13) : 2105–2125.

MILLER, C. W. (2006), *Superiority of the h-index over the Impact Factor for Physics*. Preprint. Retrieved November 4, 2006, from http://arxiv.org/PS_cache/physics/pdf/0608/0608183.pdf

MOED, H. F. (2005), *Hirsch Index is a creative and appealing construct but be cautious when using it to evaluate individual scholars.* Retrieved November 4, 2006, from: http://www.cwts.nl/hm/Comments_on_Hirsch_Index_2005_12_16.pdf

NATURE NEWSBLOG (2006), *Top Five in Physics*. Retrieved November 4, 2006, from: http://blogs.nature.com/news/blog/2006/05/top_five_in_physics.html

PALSBERG, J. (NO DATE), *The H-index for Computer Science.* Retrieved November 4, 2006, from: http://www.cs.ucla.edu/~palsberg/h-number.html

PETERSON, I. (2005), Rating researchers. *Science News Online*, 168 (23), Retrieved November 4, 2006, from: http://www.sciencenews.org/articles/20051203/mathtrek.asp

PURVIS, A. (2006), The h-index: Playing the numbers game. *TRENDS in Ecology and Evolution*, 21 (8) : 422.

ROEDIGER, H. L. (2006), The h index in science: A new measure of scholarly contribution. *The Academic Observer*, 19 (4), Retrieved November 4, 2006, from: http://www.psychologicalscience.org/observer/getArticle.cfm?id=1971

ROUSSEAU, R. (2006A), *A Case Study: Evolution of JASIS' Hirsch Index*. Preprint, Retrieved November 4, 2006, from: http://eprints.rclis.org/archive/00005430/01/Evolution_of_h_JASIS_rev.pdf

ROUSSEAU, R. (2006B), *Simple Models and the Corresponding H- and G-index*. Preprint. Retrieved November 4, 2006, from http://eprints.rclis.org/archive/00006153/01/Rousseau_Dalian.pdf

SAAD, G. (2006), Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, 69 (1) : 117–120.

SCHUBERT, A., GLÄNZEL, W. (2006), A systematic analysis of Hirsch-type indices for journals, preliminary results. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 251–253.

SIDIROPOULOS, A., KATSAROS, D., MANOLOPOULOS, Y. (2006), Generalized h-index for disclosing latent facts in citation networks. In: *Proceedings of LinkKDD06*. Retrieved November 4, 2006, from http://arxiv.org/PS_cache/cs/pdf/0607/0607066.pdf

VAN RAAN, A. F. J. (2006), Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67 (3) : 491–502.