WILEY InterScience® DISCOVER SOMETHING GREAT

Scientific Research Activity and Communication Measured With Cybermetrics Indicators

Isidro F. Aguillo, Begoña Granadino, José L. Ortega, and José A. Prieto
Laboratorio de Internet, CINDOC-CSIC, Joaquín Costa, 22, 28006 Madrid, Spain. E-mail: {isidro, bgranadino, jortega, joseaprieto}@cindoc.csic.es

To test feasibility of cybermetric indicators for describing and ranking university activities as shown in their Web sites, a large set of 9,330 institutions worldwide was compiled and analyzed. Using search engines' advanced features, size (number of pages), visibility (number of external inlinks), and number of rich files (pdf, ps, doc, ppt, and xls formats) were obtained for each of the institutional domains of the universities. We found a statistically significant correlation between a Web ranking built on a combination of Webometric data and other university rankings based on bibliometric and other indicators. Results show that cybermetric measures could be useful for reflecting the contribution of technologically oriented institutions, increasing the visibility of developing countries, and improving the rankings based on Science Citation Index (SCI) data with known biases.

Introduction

In the last decades, evaluation of research activity has been done by applying bibliometric techniques that measure the scientific output as published articles in scientific journals (Glanzel & Moed, 2002). The most commonly used indicators are those derived from Citation Indexes databases, and the most widely known is the impact factor (IF) published in the Journal Citation Report (JCR).

IF measures show several shortcomings noted in different studies (Bordons, Fernandez, & Gómez, 2002), and not all scientific output is reflected in Institute of Scientific Information (ISI) journals. Only a few journals have received an IF value, and English-language journals are favored (Rey-Rocha, Martín Sempere, López Vera, & Martínez-Frías, 1999). The scientific disciplines are not represented equally, and ISI does not calculate IFs for journals in the humanities; however, new types of indicators of the input (human and economic resources) and the output (patents) have allowed a more extensive analysis of scientific and academic research activity (Albert & Plaza, 2004).

Received June 9, 2005; revised July 19, 2005; accepted August 22, 2005

 • Published online 25 May 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20433

Currently, the Web offers a new object to apply informetrics and bibliometrics techniques (Björneborn & Ingwersen, 2004; Thelwall, Vaughan, & Björneborn, 2005). The Web itself is continually changing and developing, and its evolution has allowed including a large volume of information that reflects academic and research activities. The availability of hyperlinks between pages seems analog with citation between journal articles (Egghe, 2000; Ingwersen, 1998; Larson, 1996; Rousseau, 1997; but see Prime, Bassecoulard, & Zitt, 2002). The high number of objects (pages, images, links, etc.) in the Web offers the opportunity for a great variety of quantitative analysis, now the working field of cybermetrics and Webometrics (Björneborn & Ingwersen, 2004). An important theoretical and empirical background already exists on the application of metric methods to the analysis of research activity and publications; thus, the application of bibliometric and scientometric techniques seems logical (Thelwall & Vaughan, 2004).

The actors in Research and Development (R&D) are the same, as institutions, scientists, and managers have a presence both in traditional publications and in Web contents, although possibilities of interaction and communication between them and the variety of the content probably are higher in the Web. In this sense, the measure and the evaluation of the presence in the Web of different academic and research institutions (Bar-Ilan, 2004; Li, Thelwall, Musgrove, & Wilkinson, 2003; Smith & Thelwall, 2002; Tang & Thelwall, 2004; Thelwall & Tang, 2003; Thomas & Willet, 2000; Vaughan & Thelwall, 2005) could give complementary indicators to bibliometric standard measurements.

Prior to a wide use of Web data for evaluation purposes, several theoretical and methodological questions should be solved. If cybermetric indicators are added to the list of evaluation criteria, more researchers probably use the Web as a publication channel not only for formal articles but also for their informal production, providing access to a large range of their research activities. This publicly available information will be richer, useful, and easy to access and monitor, but also cheaper to produce and distribute, making scientific publication more universal. The audience not only will be larger but more representative as well, as developing countries could find a place in the scholarly communication network through the Web.

The quantitative analysis of presence on the Web of different institutions related to the scientific-knowledge generation and transfer should address the following questions:

- What is the real visibility and impact of the presence of R&D institutions on the Web?
- Is there any correlation between the most productive institutions according to bibliometric indicators and those with a higher presence on the Internet?
- What is the Web position of the R&D centers with respect to other organizations of the same country?
- Do the technological-oriented centers obtain similar rankings to those of larger, multidisciplinary universities?
- What is the scientific-activity Web visibility in the less-developed countries?
- Besides general information of the institution, is there any scientific information such as theoretical aspects, methods, tools, data, results, transfer of technology, docent activity, divulgation, and so on?
- Which are the most frequent units and at what level? Are personal pages relevant?

In this article, we aim to answer the previous questions by studying different examples that would need future, detailed analysis.

Methods

The first problem is to define a unit. In bibliographic databases, with a formal structure of records and fields, the unit is the document—mostly articles in scientific journals. The Web does not have such a formal structure, although there is an hypertext markup language (HTML) tagging system that can be used for quantitative purposes. Moreover, Web contents are dynamic and frequently updated, and include the so-called rich files: independent and complete units in formats such as pdf, ps, or doc that are actual documents.

In bibliometrics, units are articles and journals; but when institutional and informal communication are prevalent, such as in the Web, the unit of work should reflect an institutional component. Many research groups, university departments, or even individual researchers have an autonomous presence on the Web, usually a set of Web pages hierarchically organized and interlinked with an specific URL address. This group of pages is referred to as a Web site, and several Web sites with the same domain owned by an academic or research organization constitutes an institutional domain. Web sites and institutional domains can be formally (e.g., authors, geographic location, thematic contents) and quantitatively (e.g., Web pages, hypertext links, media objects) described. For the purposes of this article, the units were institutional domains of universities that are easily identified and show a well-defined hierarchy, mostly under a unique Web domain, that represents that institution presence on the Web (Aguillo, 1998).

The standard method for data gathering from the Web is the use of crawlers, directly (Cothey, 2004; Thelwall, 2001) or through commercial search engines (Bar-Ilan, 2003). Custom crawlers are more precise and exhaustive, but for large surveys, search engines are easy to use and powerful enough for automatic Web data extraction. Limitations have been noted by several authors, such as incomplete coverage of the Web (Lawrence & Giles, 1999) or lack of stability or precision (Bar-Ilan, Levene, & Mat-Hassan, 2004). Some of the problems have been solved, and engines are very useful tools for cybermetric analysis (Vaughan, 2004; Vaughan & Thellwall, 2004).

For quantitative purposes, one can use some advanced possibilities offered by a limited number of more comprehensive search engines that extract specific results. Using HTML tags, some search engines build a derived field structure that can be filtered using search strategies that involve operators called delimiters. In this way, the number of pages of a certain domain or site can be recovered using a specific delimiter. The main delimiters and their syntax for the main search engines are described in Table 1. Since 2004, the only current independent search engines that allow delimiters are Google, Yahoo! Search (which bought Altavista and AlltheWeb), MSN Search, and Teoma.

TABLE 1. Cybermetric operators in search engines (March 2005).

Request	Syntax			
	Google	Yahoo! Search	MSN Search	Teoma
Domain	site:xx	site:xx	site:xx	
		domain:xx	domain:xx	
Subdomain/site	site:aaa.xx	site:aaa.xx	site:aaa.xx	site:aaa.xx inurl:aaa.xx
Subdirectory	inurl:www.aaa.xx/bb	url:http://www.aaa.xx/bb		inurl:aaa.xx/bb
Word in URL	inurl:bb	inurl:bb		inurl:bb
Links to domains		linkdomain:aaa.xx	link:aaa.xx	
Links to pages	link:www.aaa.xx	link:http://www.aaa.xx/bb		
Country	form 200 countries	form 24 countries	loc:xx	
Language	form 35 languages	form: 32 languages	language:xx	lang:xx
Filetype	filetype:xxx	originurlextension:xxx		

DOI: 10.1002/asi

TABLE 2. Size (no. of Web pages) according to search engines.

		Google	MSN	Teoma	Yahoo!
Size	No. of pages	389,369,088	364,891,111	129,659,822	129,205,256
	Pages/domain	41,733	39,109	13,897	13,848
Correlation coefficient	Google	1.000	.917**	.731**	.893**
	MSN	.917**	1.000	.692**	.856**
	Teoma	.731**	.692**	1.000	.822**
	Yahoo!	.893**	.856**	.822**	1.000

**The correlation is statistically significant .01.

To build a large target for testing the objectives, a survey of world universities and higher education institutions was launched. Although not all universities, especially small, private ones from developing countries, have a Web presence, it was possible to compile a list after checking several international directories (e.g., Braintrack, World of Learning) and national lists both in printed and electronic form. Institutions without their own institutional domain were excluded, so a final list of domains of 9,330 universities worldwide (ranking of world universities in the Web: see www.Webometrics.info) was compiled.

The list has several biases, as large universities usually have additional independent domains not included in our analysis. There are a few institutions with two equally valid domains and addresses no longer valid but still covered in search engines. The survey probably covers over 90% of the world universities.

Despite the large size of the Web, data compilation can be done automatically (Mayr & Tosques, 2005). For each one of the institutional domains and for each search engine, an automatic extraction procedure was launched in January 2005 using several scripts developed in-house. The strategies used are those derived from Table 1 and explained in the next section.

Results

Presence of Universities on the Web: Real Visibility and Impact

Due to shortcomings of the search engines and the dynamic nature of the Web, the use of absolute values can be misleading, so a ranking could be a better answer for describing relative impact of an institutional Web site.

Previous relevant approaches to Web indicators include the development of the Web Impact Factor (Ingwersen, 1998), which takes into account both visibility (i.e., number of external inlinks) and size (i.e., number of Web pages) in a ratio of 1:1. For our analysis, we decided to increase the weight of the visibility, but also to give relevance to the pages that consist of rich files, thus making a new ratio of 4:3.

For each indicator, there are several sources of data and different search engines with different databases. Prior to the combination of results from these engines, the consistency among them was checked for the three measurements [Size (S), Visibility (V), and Rich Files (R)].

The S of each domain was calculated with the strategies: *site:domain* (Google, MSN Search, and Yahoo! Search) or *site:domain inurl:domain* (Teoma).

Each robot operates with different time schedules, according to programmed criteria and accessing Web sites without a clearly defined timing. Independently of the characteristics of the robot, the crawling of a Web site depends on unexpected factors. Therefore, taking into account the limited overlap among engines, one should consider combining data extracted from several sources.

From results shown in Table 2, the significant positive correlation measured by the nonparametric test of the Spearman's rho suggests the combination of the values of the four search engines. The depth level of indexing and the time spent in each server by the different crawlers can explain the differences in the total number of pages collected by them in the 9,330 universities domains.

The V (i.e., external links received by the domain) can be calculated only in the Yahoo! Search *linkdomain:domain -site:domain* and the MSN Search *link:domain -site:domain*.

In Table 3, we can see that the correlation obtained is positive and statistically significant. The high value supports the use of a combined figure for the visibility index.

The R were obtained in Google with the strategy *site:domain filetype:format*. The correlations obtained, although lower, are statistically significant (Table 4).

With these results, the data were combined by the sum of the values obtained in the three categories: S, V, and R (Table 5).

The score values are substituted by their relative position in a ranking of the domains determined according to these three parameters (Rs, Rv, and Rr). An indicator (I) is defined with the relative contribution of each parameter determined by its relevance and importance according to our subjective

TABLE 3. Visibility (no. of inlinks) according to search engines.

Correlation coefficient		MSN	Yahoo!
	No. of links	71,397,027	266,834,855
	MSN	1.000	.888**

**The correlation is statistically significant .01.

DOI: 10.1002/asi

TABLE 4. Rich files according to search engines.

		PDF	PS	DOC	PPT	XLS
No. of files		43,930,937	4,516,755	7,021,760	2,089,441	668,374
Files/domain		4,709	484	753	224	72
Correlation coefficient	PDF	1.000	.612**	.738**	.734**	.670**
	PS	.612**	1.000	.568**	.638**	.607**
	DOC	.738**	.568**	1.000	.791**	.778**
	PPT	.734**	.638**	.791**	1.000	.773**
	XLS	.670**	.607**	.778**	.773**	1.000

**The correlation is statistically significant .01.

TABLE 5. Correlation among the three cybermetric indicators.

Correlation coefficient	Size	Visibility	Rich files
Size	1.000	.899**	.892**
Visibility	.899**	1.000	.833**
Rich files	.892**	.833**	1.000

**The correlation is statistically significant .01.

TABLE 6. Comparison of university rankings (non-Web and Web-Based).

Correlation coefficient $N = 99$		Times	Shanghai	ESI-ISI	Web
	Times	1.000	.638**	.426**	.379**
	Shanghai	.638**	1.000	.760**	.538**
	ESI-ISI	.426**	.760**	1.000	.609**
	Web	.379**	.538**	.609**	1.000

**The correlation is statistically significant .01.

perception based on our previous experience and some empirical tests.

$$I = 2\ Rs + 4\ Rv + Rr$$

This measure was uses for producing the ranking of world universities on the Web (www.Webometrics.info).

Comparing Web and Bibliometric Rankings

To determine a possible correlation between bibliometric and Webometric indicators, a comparative analysis between the Web Ranking and the other three university rankings was carried out based on traditional bibliometric indicators and other measures of prestige and impact. The Essential Science Indicators (www.esi-topics.com) provides a list of the most productive institutions according to the number of articles published and the received cites in the journals covered by the ISI Web of Knowledge during 1994–2004. The *Academic Ranking of World Universities* (2004), elaborated by the University of Shanghai (ed.sjtu.edu.cn/ranking.htm), and the *World University Ranking* (November 2004), published by the Times Higher Education Supplement (www.thes.co.uk/worldrankings), consider criteria such as research output, peer-review score, quality of education, number of foreign students, staff winners of Nobel Prizes or Field Medals, and others.

For this purpose, the first 100 universities from the Shanghai University Ranking were selected, and the position of these institutions in the other three lists annotated. The correlation between the Web ranking and the other three academic rankings is shown in Table 6. The results indicate a positive association between them, although the values are not very high. There is a correlation between the most productive academic institutions according to bibliometric and cybermetric indicators.

Eighty percent of the first 100 institutions of the Web ranking are from English-speaking countries. A university Web site has two main aspects to exploit: (a) scientific output (mostly publications in scientific journals) produced by the institution that uses English as its official language, and (b) general academic and para-academic information about the structure and activities of the university that uses the mother tongue. This fraction of the site is a target for links from a broad spectrum of pages from other organizations, including non-universities generally using the same language as the cited one. This fact could explain the low representation of French or Japanese universities in the Web ranking. Universities of other countries such as Norway, Sweden, or Germany that also use English for academic purposes are comparatively better represented in the Web ranking. On the other hand, Asiatic universities are absent on the Web but well represented on the Times ranking. This may be due, in part, to the importance of their national prestige and their large size.

Cybermetric Rankings and Bias in Traditional Bibliometric Indicators

Some of the shortcomings of the traditional bibliometric measurements noted in different studies can be targeted by the cybermetric indicators. Webometrics could be useful to evaluate Third World universities with less financial resources and less access to mainstream scientific publications. On the Web, both academic (uses the local language) and scientific information (mainly in English) is published, thus reducing language bias. Moreover, technological universities are usually undervalued by bibliometric indicators due to the low number of articles published in scientific journals in this field. However, their institutional presence in the Web is

DOI: 10.1002/asi

greater, thus improving their evaluation if cybermetric parameters are taken into account (discussed later).

Scientific activity and visibility in developing countries. Combining the data obtained from all the university domains, we analyzed relative contribution of each national domain (cTLD), expressed as the percentage of pages of all institutional domains of that country of the total obtained for the whole set. As relative contribution of each country is different for each search engine, data were normalized against the Google value (100). As can be seen in Table 7, there are important biases for every search engine when languages other than English are considered. Yahoo! Search covers U.S. pages well, but not those from other countries. The Asiatic countries' pages are only covered well by the MSN Search engine.

Web position of the Spanish-speaking R&D centers. We generated a list including the 947 main Web pages that use Spanish, using as the visibility indicator the Google PageRank. The relative position according to the language can be obtained with a strategy that filters Web pages in Spanish with a term semantically neutral (http). In total, 104 Web pages corresponded to universities in Spain and Latin America, suggesting that academic institutions are of great importance in organizing the Hispanic Web. Sixty-two sites corresponded to universities in Spain, 13 to Mexico, and 11 to Argentina (Table 8).

TABLE 7. Normalized contribution of the universities by top-level domains (TLD) showing search engines bias.

Relative contribution	Google	MSN search	Teoma	Yahoo! search
English-speaking countries				
USA (edu)	100.00	108.58	190.10	160.64
Canada (ca)	100.00	84.31	104.19	97.00
India (in)	100.00	135.63	81.90	93.10
United Kingdom (uk)	100.00	69.71	122.01	82.95
Australia (au)	100.00	71.06	100.42	72.11
Israel (il)	100.00	67.67	326.37	64.39
New Zealand (nz)	100.00	62.18	74.19	61.74
South Africa (za)	100.00	69.21	71.26	57.19
Asian countries				
Taiwan (tw)	100.00	141.24	25.38	33.44
South Korea (kr)	100.00	165.99	21.82	24.26
Japan (jp)	100.00	112.22	24.44	21.86
China (cn)	100.00	87.55	5.72	13.69
Eastern Europe				
Russia (ru)	100.00	155.77	16.21	118.74
Ukraine (ua)	100.00	116.81	12.83	90.68
Poland (pl)	100.00	114.84	11.56	79.28
Hungary (hu)	100.00	126.95	20.40	74.27
Czech Republic (cz)	100.00	124.63	30.30	62.68
Other in Europe				
Germany (de)	100.00	99.70	21.76	86.78
Denmark (dk)	100.00	96.01	79.62	78.10
Austria (at)	100.00	97.42	38.90	75.21
Sweden (se)	100.00	70.19	69.25	72.52
Italy (it)	100.00	79.25	31.03	68.51
Spanish-Portuguese-speaking countries				
Brazil (br)	100.00	105.71	34.96	101.76
Mexico (mx)	100.00	84.64	79.89	95.00
Argentina (ar)	100.00	90.48	47.54	83.44
Chile (cl)	100.00	110.29	57.67	82.78
Spain (es)	100.00	83.58	63.53	76.13
Portugal (pt)	100.00	113.74	28.52	73.98

TABLE 8. Number of universities' pages among the top 1,000 Spanish-speaking ones according to Google PageRank.

Countries	Universities
Spain	62
México	13
Argentina	11
Chile	4
Colombia	4
Puerto Rico	3
Venezuela	3
Peru	2
Costa Rica	1
Ecuador	1
Total	104

We extracted data concerning the first 50 universities for Spain. The first 11 are shown in Table 9. We determined the absolute visibility (external links received by a domain) and the relative visibility (Google PageRank), and correlated these values with the number of articles published in SCI journals from 1996–2001 calculated by Gómez et al. (2004). The correlations obtained are statistically significant (Table 10).

R and personal Web pages. Each type of R of a Web site represents a characteristic way of knowledge communication, including scholarly or research activity. Articles in scientific journals are the main way of formal communication, but we also found notes about docent activity, raw data, congress communications, workshop presentation, and so on. Powerpoint (ppt) files represent teaching activities or participation in workshops or meetings. Excel (xls) files could contain raw numbers, which are needed to distribute high volumes of information. Word (doc) and Acrobat (pdf) documents are used to disseminate final versions; however, note that many files in this format also can be used for bureaucratic or administrative purposes. This is not the case for PostScript (ps) documents, which are mainly used for final versions of scientific and technological texts in disciplines such as mathematics, physics, or engineering.

With Google filetype delimiter, we analyzed the types of R of the Web domains of the first 50 universities for Spain (Google PageRank). We showed that some of them contain a large collection of R in a volume exceeding what should be expected for administrative information. This is evident for the pdf files and ps, especially in the technological universities. On the other hand, xls data are very scarcely represented,

DOI: 10.1002/asi

TABLE 9. Rank of Spanish universities according to Google.

PageRank	University	URL
15	Universidad de Sevilla	*http://www.us.es/*
23	Universidad Complutense de Madrid	*http://www.ucm.es/*
27	Universitat de Barcelona	*http://www.ub.es/*
54	Universitat Oberta de Catalunya	*http://www.uoc.edu/*
70	Universidad Autónoma de Madrid	*http://www.uam.es/*
72	Universidad de Granada	*http://www.ugr.es/*
81	Universidad Politécnica de Madrid	*http://www.upm.es/*
82	Universidad de Salamanca	*http://www.usal.es/*
95	Universidad Nacional de Educación a Distancia	*http://www.uned.es/*
96	Universitat Autonoma de Barcelona	*http://www.uab.es/*

TABLE 10. Visibility rankings (citations, inlinks, PageRank).

Correlation coefficient $N = 50$		SCI	Visibility	Pagerank
	SCI	1.000	.750**	.776**
	Visibility	.750**	1.000	.881**
	Pagerank	.776**	.881**	1.000

**The correlation is statistically significant .01.

probably due to the authors' suspicion and distrust about public availability of their data source on the Web. Table 11 shows a positive and significant correlation between all R and scientific productivity as measured by number of articles in the SCI. As technological institutions—the main users and producers of ps documents—are underrepresented in the citation databases, the correlation is lower for that file type.

The quantitative study of the contents of these domains was analyzed with Yahoo! Search using the *feature* delimiter with two options: *index* to identify Web subdomains and *homepage* to identify personal Web pages (Table 12).

The results showed that the large Spanish universities have a great number of subdomains, which indicates the general maturity of the system, as this rich structure is possible only when university departments and facilities assume control of their own Web sites, differentiating them with a subdomain.

The number of personal pages is approximately 10%, which confirms the importance of this typology, although it is lower than what it was considered. Note that some of these pages are duplicates on the Web for some research teams, which consider the Web sites as relevant scientific-communication tools.

TABLE 11. Correlation between rich files and citation data.

Correlation coefficient $N = 50$		PDF	PS	DOC	PPT	XLS
	SCI	.671**	.543**	.674**	.663**	.457**

**The correlation is statistically significant .01.

TABLE 12. Subdomains and personal pages.

Correlation coefficient $N = 50$		Size	Visibility	Index	Homepage
	Size	1.000	.973**	.842**	.692**
	Visibility	.973**	1.000	.837**	.684**
	Index	.842**	.837**	1.000	.644**
	Homepage	.692**	.684**	.644**	1.000

**The correlation is statistically significant .01.

Ranking the Spanish scientific and technological universities. For this analysis, we selected the four most productive universities of Spain (i.e., Universidad de Barcelona, Universidad Complutense de Madrid, Universidad de Valencia, and Universidad Autonoma de Barcelona) and those with a technological focus (Politecnicas), with lower productivity measured as publications in SCI journals, but with visibility indicators close to the other four universities mentioned previously (Table 13).

The results in Table 13 show that the technological universities acquire better ranking according to Yahoo! visibility and Google PageRank than to SCI. The most productive universities without a technological focus maintain almost the same ranking according to the different criteria. Thus, these results suggest that cybermetric indicators are useful to evaluate technological universities whereas traditional bibliometric indicators have limitations.

Conclusions

The cybermetric techniques contribute to the battery of tools available for the evaluation of scientific output. Cybermetrics also plays an important role in the description of processes of scientific communication, but not as formally as the scientific publications. The hypertextual structure of the Web and the institutionalization of its content offer opportunities to describe communication patterns in the science–technology–society system.

The cybermetric indicators showed the possibility to measure Web visibility and impact of R&D institutions, suggesting that these values are related to the scientific and academic output of the institution. University Web sites host different scientific content, and their sites are used to publish scientific results, with the research groups as the main

DOI: 10.1002/asi

TABLE 13. Comparison between most productive and technological Spanish universities.

Most productive universities		SCI (96–01)	Ranking	Yahoo! visibility	Ranking	Google PageRank (country = Spain)	Ranking
Universitat de Barcelona	ub.es	7774	1	185,000	3	19	2
Universidad Complutense de Madrid	ucm.es	7500	2	189,000	2	22	3
Universitat de València	uv.es	4566	3	460,000	1	73	11
Universidad Autónoma de Barcelona	uab.es	4117	4	155,000	8	49	6
Technological universities							
Universitat Politècnica de Catalunya	upc.es	2326	11	177,000	5	52	7
Universidad Politecnica de Madrid	upm.es	1699	15	180,000	4	58	8
Universitat Politecnica de Valencia	upv.es	1128	24	177,000	6	70	10
Universidad Politécnica de Cartagena	upct.es	133	49	11,100	47	500	50

authors of their own Web sites. The positive results obtained for the technological universities open the possibility of a real evaluation of their scientific contribution. It is expected to extend this result to the academic institutions of the Third World countries.

References

Aguillo, I.F. (1998). STM information on the Web and the development of new Internet R&D databases and indicators. Proceedings Online Information Meeting 98, London. Learned Information, 239–243.

Albert, A., & Plaza, L.M. (2004). The transfer of knowledge from the Spanish public R&D system to the productive sectors in the field of biotechnology. Scientometrics, 59(1), 3–14.

Bar-Ilan, J. (2003). The use of Web search engines in information science research. Annual Review of Information Science and Technology, 38, 231–288.

Bar-Ilan, J. (2004). A microscopic link analysis of academic institutions within a country—The case of Israel. Scientometrics, 59(3), 391–403.

Bar-Ilan, J., Levene, M., & Mat-Hassan, M. (2004, May). Dynamics of search engine rankings—A case study. Proceedings of the WWW 2004 Conference, New York. Retrieved from http://www.dcs.bbk.ac.uk/WebDyn3/Webdyn3_ proceedings.pdf

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for Webometrics. Journal of the American Society for Information Science and Technology, 55(14), 1216–1227.

Bordons, M., Fernandez, M.T., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country. Scientometrics, 53(2), 195–206.

Cothey, V. (2004). Web crawling reliability. Journal of the American Society for Information Science and Technology, 55(14), 1228–1238.

Egghe, L. (2000). New informetric aspects of the Internet: Some reflections, many problems. Journal of Information Science, 26(5), 329–335.

Glanzel, W., & Moed, H.F. (2002). Journal impact measures in bibliometric research. Scientometrics, 53(2), 171–193.

Gómez, I., Fernández, M.T., Bordons, M., & Morillo, F. (2004). Obtención de indicadores de producción científica y tecnológica de España (1996–2001). Madrid: CINDOC. Retrieved from http://www.cindoc.csic.es/investigacion/informe1.pdf

Ingwersen, P. (1998). The calculation of Web impact factors. Journal of Documentation, 54, 236–243.

Larson, R.A. (1996, July). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. Proceedings of the annual ASIS Conference. Retrieved from http://sherlock.berkeley.edu/asis96/asis96.html

Lawrence, S., & Giles, L. (1999). Accessibility of information on the Web. Nature, 400, 107–109.

Li, X., Thelwall, M., Musgrove, P., & Wilkinson, D. (2003). The relationship between the links/Web Impact Factors for computer science departments in the U.K. and their RAE (Research Assessment Exercise) ranking in 2001. Scientometrics, 57(2), 239–255.

Mayr, P., & Tosques, F. (2005). Webometrische Analysen mit Hilfe der Google Web APIs [Webometric analysis with the help of Google's API]. Information, 56(1), 41–48.

Prime, C., Bassecoulard, E., & Zitt, M. (2002). Co-citations and co-sitations: A cautionary view on an analogy. Scientometrics, 54(2), 291–308.

Rey-Rocha, J., Martin-Sempere, M., Lopez-Vera, F., & Martínez-Frías, J. (1999). English versus Spanish in science evaluation. Nature, 397(6714), 14.

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics, 1(1), 1. Retrieved from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Smith, A., & Thelwall, M. (2002). Web Impact Factor for Australasian universities. Scientometrics, 54(3), 363–380.

Tang, R., & Thelwall, M. (2004). Patterns of national and international Web inlinks to U.S. academic departments: An analysis of disciplinary variations. Scientometrics, 60(3), 475–485.

Thelwall, M. (2001). A Web crawler design for data mining. Journal of Information Science, 27(5), 319–325.

Thelwall, M., & Tang, R. (2003). Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with mainland China and Taiwan. Scientometrics, 58(1), 153–179.

Thelwall, M., & Vaughan, L. (2004). Webometrics: An introduction to the special issue. Journal of the American Society for Information Science and Technology, 55(14), 1213–1215.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. Annual Review of Information Science and Technology, 39, 81–135.

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of Librarianship and information science. Journal of Information Science, 26(6), 421–428.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. Information Processing & Management, 40(4), 677–691.

Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias: Evidence and possible causes. Information Processing & Management, 40(4), 693–707.

Vaughan, L., & Thelwall, M. (2005). A modeling approach to uncover hyperlink patterns: The case of Canadian universities. Information Processing & Management, 41(2), 347–359.

DOI: 10.1002/asi