

●段宇锋 邱均平

基于链接分析的网站评价研究*

摘要 从排名前50位的美国商学院中,随机抽取20个,把它们主页所在网站作为研究对象,以指向网站的网页数和网络影响因子(Web-IF)作为测定核心网站的依据。研究表明:用这两种依据来测定核心网站,所得结果基本一致;Web-IF对评价网站质量和测定核心网站具有重要价值;在计算Web-IF时应以其他网站指向被研究对象的网页数和网站在该时刻可访问到的网页数作为依据;布拉德福定律可能适用于核心网站的研究。表3。参考文献11。

关键词 核心网站 测定 评价 链接分析

分类号 G250.7

ABSTRACT The authors make a comparative study of 20 websites randomly selected from the websites of the top 50 business schools in U. S. A. , use the numbers of pages pointing to the websites and Web-IF (web impact factors) as the criteria for the determination of core websites, and conclude that the Bradford's Law may be applicable to the study of websites. 3 tabs. 11 refs.

KEY WORDS Core website. Measurement. Evaluation. Link analysis.

CLASS NUMBER G250.7

1 基于链接分析的核心网站测定研究现状

1.1 基于网站被链接频次的核心网站测定研究

就笔者目前所见文献,Allison Woodruff是最早对核心网站开展研究的学者。他在1995年11月利用Inktomi crawler收集了260万个网页,通过对网页链接的分析,筛选出被链接频率最高的网站。这项研究还对源于网站内部和外部的链接进行了区分,并指出以源于网站外部的链接作为研究指标将更有意义^[1]。1999年,崔雷(Lei Cui)按照布拉德福定律确定出25个被链接次数最多的美国医学院网站作为核心网站^[2]。2000年,Robert C. Vreeland用“luminosity”表示某网站指向其他网站的链接的数量,用“visibility”代表某网站被其他网站链接的次数。他对156个法律图书馆网站的影响力进行了研究。结果显示,20%的网站提供了80%的站外链接;luminosity和visibility之间无明显的相关性。研究还根据站外链接频次统计和利用AltaVista的链接分析功能,列出了“可见度”最高的网址和主机^[3]。蒋颖在1999年10月利用AltaVista对我国232个图书情报网站被链接情况进行了统计分析。研究发现,各地网站被链接情况差异很大,网站被链接情况与各地互联网发展程度有极强的相关关系。并且,网站被链接

的频次基本上反映了网站的质量和建站机构的水平^[4]。

对网站被利用情况的统计方法可概括为两种:第一种是借鉴传统文献计量学研究中以布拉德福定律确定核心期刊的方法。以这种方法测定核心网站必须解决3个问题,即合理地选择样本、完整地获取样本网站原始数据、完整地提取网站中的链接,只有这样才能得出准确的结果。第二种方法是依靠具有链接分析功能的商业搜索引擎获得网站被链接的信息,又可分成3种情况:(1)利用搜索引擎获取其他网站指向样本网站的信息,将所有链接到样本网站的网址汇总统计,将这一组数据作为测定核心网站的依据。(2)将样本网站之间相互链接的次数作为网站被链接的次数。(3)将其他网站指向样本网站的数量作为网站被链接的频次。根据所获得数据从高到低进行排序,依据一定的判定标准,将出现频次最高的一组网站确定为核心网站。第一种情况适用于不预先限定核心网站范围的研究,第二、三种情况适用于在指定范围内确定核心网站的情况。利用搜索引擎确定核心网站,要清醒地认识到搜索引擎的覆盖率、检索结果的波动性等方面所固有的缺陷。

1.2 基于Web-IF的核心网站测定研究

1998年,Peter Ingwersen在*The Calculation of Web*

* 本文是国家自然科学基金资助项目“网络信息计量学的理论、方法与实证研究”(批准号70273032)的组成部分和研究成果之一。

Impact Factors 一文中提出了 Web-IF 概念^[5]。他将 Web-IF 定义为:在某一特定时刻,指向特定国家或网站的网页总数与该国或网站中网页数之比。作者还指出指向网站的链接分为来源于外部的链接(external-citations)和源于自身的链接(self-citations),认为“external Web-IF”是测度网站影响力的指标,而“self-linkage”反映的是服务器上网页组织的逻辑结构。这篇文献立即引起学术界的关注,各国学者都围绕 Web-IF 进行了大量研究。在这些研究中,比较有代表性的是 Alastair G. Smith 对澳大利亚和新西兰 42 个大学网站以及 22 个电子期刊 Web-IF 的研究。这项研究表明,大部分大学网站的 self-link WIFs 在 0.5 左右,这与 Ingwersen 的研究结果是一致的;22 种电子期刊只有 5 种被 ISI 收录,external WIF 最高是 PSYCHE,它也是在 ISI 中被引用次数最多的电子期刊^[6]。笔者于 2002 年 10~11 月间,利用 AllTheWeb 对 42 种工程类中文期刊网站的 Web-IF 进行测定,并与中国科学技术信息研究所 2001 年底公布的 2000 年期刊影响因子逐一对应比较,发现期刊影响因子与期刊网站的 Web-IF 具有统计学意义的相关关系。因此,我们认为 Web-IF 可以作为网站评价的重要指标^[7]。

计算网站的 Web-IF 必须获得指向被研究网站的网页数和网站自身的网页数。前者可分为指向被研究网站的网页总数、外部指向被研究网站的网页数和网站内部指向自身的网页数,这些数据一般是利用具有链接分析功能的搜索引擎获得。而网站的页面数则有两种获取途径:一是直接利用搜索引擎索引的指定网站的页面数;二是采用其他软件遍历指定网站,统计其页面数。利用 Web-IF 测定核心网站,在获得样本的 Web-IF 值后,依据一定的规则按 Web-IF 的大小确定。尽管 Web-IF 的提出强有力地推动了网络链接研究的发展,然而,研究者们也发现,在具体研究中所遇到的困难远比想象中的多,甚至有些问题迄今还找不到合理的解决办法。其中最大的问题是缺乏像测定期刊影响因子中 SCI 这样的工具,尽管以搜索引擎作为 Web-IF 测定工具的缺陷是众所周知的,但目前却没有更好的解决办法。

2 两种基于链接分析的核心网站测定方法的相关性研究

目前几乎所有测定核心网站的研究采用的都是上面所讨论的这两类方法。为了明确这两类方法评

价结果的一致性,我们对测定结果进行相关分析,并对网页计数方法对结果的影响进行深入探讨。

2.1 核心网站测定方法

我们分别采用指向样本网站的网页数和网站的 Web-IF 值作为测定核心网站的依据。

2.1.1 样本的选择

根据 U. S. News 在网上公布的研究型院校排行榜(American's Best Graduate Schools2004),获取排名在前 50 位的美国商学院,以这些院校主页所在的主机作为总体,按照随机抽样的原理,从中抽取 20 个作为样本。抽样结果为排名第 1、2、5、7、10、11、12、16、21、22、23、25、30、32、33、37、42、44、45、46 的商学院网站。

2.1.2 核心网站的测定步骤

本研究基于网站被链接频次的核心网站测定,借鉴了核心期刊的确定方法,以在特定范围内样本被链接的绝对次数作为评价的依据。首先,以 MetaProducts 公司开发的 Offline Explorer2.9.1280 作为样本网站原始数据获取工具。其次,利用自主开发的网络链接分析软件 webStat 作为从所下载的原始数据中解析和提取链接的工具。链接解析完成后,相关数据自动存入名为“result.mdb”的数据库中。如果是指向主机所在域的链接,则将该记录在数据库中的 Is_inner 数据项置为“1”,否则置为“0”。第三,统计各网站被链接的频次(X)。利用 WebStat 的“统计链接”功能,对 result.mdb 数据库 LinkStat 表所有 Is_inner 数据项为“0”的记录进行统计,统计结果自动存储在名为“result.txt”的文件。最后,根据各研究对象被链接的次数,由高到低排序,依据“二八”律判定核心网站。

基于 Web-IF 的核心网站测定,是先利用搜索引擎 alltheweb 获取指向被研究网站的网页数。其指令为:

link.all:网站地址 指向该网站的网页总数(L1);

link.all:网站地址 -site:网站地址 网站以外网络资源指向该网站的网页数(L2);

link.all:网站地址 -site:域名 网站所在域以外网络资源指向该网站的网页数(L3);

利用 webStat 获取样本网站的网页数(P1)。webStat 将后缀为.html,.htm,.asp,.jsp,.php,.stm,.stml,.idc,.shtml,.txt,.htx.text,.xsp,.xml,.ixml,.cfm,.wml 的文件视为页面文件,对其进行计数。

利用搜索引擎 alltheweb 获取网站的网页数 (P2)。其指令为: url.all: 网站地址 site: 网站地址。

Web-IF 的计算方法为: 指向网站的网页数/网站网页数。由于指向网站网页数的计数方法有 3 种, 对网站网页数的统计也有 2 种不同的方法, 因此, 共将产生 6 个 Web-IF 值, 即:

$$\begin{aligned} \text{Web-IF1} &= L1/P1 & \text{Web-IF3} &= L3/P1 \\ \text{Web-IF2} &= L2/P1 & \text{Web-IF4} &= L1/P2 \end{aligned}$$

Web-IF5 = L2/P2 Web-IF6 = L3/P2
 然后根据 Web-IF 值的大小判定核心网站。

2.2 核心网站的测定结果

根据上述方法, 我们将所获得的数据整理成表 1。由于技术上的原因, 我们无法获取样本 5, 即网站 http://www.tuck.dartmouth.edu/ 的原始数据, 因而将它从研究对象中剔除。

表 1 研究对象各相关指标的数据

样本编号	网站主机地址	P1	P2	L1	L2	L3	Web-IF1	Web-IF2	Web-IF3	Web-IF4	Web-IF5	Web-IF6	X	排名
1	www.hbs.edu	2734	4164	33214	30839	23674	12.1485	11.2798	8.6591	7.9765	7.4061	5.6854	53	1
2	www.gsb.stanford.edu	4566	1536	27919	23908	16159	6.1145	5.2361	3.5390	18.1764	15.5651	10.5202	19	2
3	www.kellogg.northwestern.edu	12296	0	7437	7437	2370	0.6048	0.6048	0.1927	#DIV/0!	#DIV/0!	#DIV/0!	13	5
4	www.fuqua.duke.edu	952	1154	15419	13628	8878	16.1964	14.3151	9.3256	13.3614	11.8094	7.6932	191	7
6	www.darden.virginia.edu	1058	1116	7748	5801	1234	7.3233	5.4830	1.1664	6.9427	5.1980	1.1057	6	11
7	www.stern.nyu.edu	6411	7880	54903	51477	45519	8.5639	8.0295	7.1001	6.9674	6.5326	5.7765	81	12
8	www.johnson.cornell.edu	588	846	2325	2227	1215	3.9541	3.7874	2.0663	2.7482	2.6324	1.4362	13	16
9	www.goizueta.emory.edu	1412	1323	814	750	129	0.5765	0.5312	0.0914	0.6153	0.5669	0.0975	0	21
10	www.kenan-flagler.unc.edu	2195	759	668	652	321	0.3043	0.2970	0.1462	0.8801	0.8590	0.4229	13	22
11	www.kelley.iu.edu	3586	1059	1819	1524	302	0.5073	0.4250	0.0842	1.7177	1.4391	0.2852	8	23
12	www.mgmt.purdue.edu	4277	4419	11292	9929	7670	2.6402	2.3215	1.7933	2.5553	2.2469	1.7357	28	25
13	www.biz.uiowa.edu	5074	9052	25777	24881	23090	5.0802	4.9036	4.5507	2.8477	2.7487	2.5508	35	30
14	www.olin.wustl.edu	2388	2952	7281	6775	5646	3.0490	2.8371	2.3643	2.4665	2.2951	1.9126	17	32
15	www.smeal.psu.edu	3506	4108	11058	8683	6853	3.1540	2.4766	1.9546	2.6918	2.1137	1.6682	16	33
16	wpcarey.asu.edu	5004	4994	1285	655	339	0.2568	0.1309	0.0677	0.2573	0.1312	0.0679	7	37
17	www.freeman.tulane.edu	634	446	272	279	190	0.4290	0.4401	0.2997	0.6099	0.6256	0.4260	0	42
18	www.rhsmith.umd.edu	6915	6580	22857	16899	14446	3.3054	2.4438	2.0891	3.4737	2.5682	2.1954	19	44
19	mba.vanderbilt.edu	6004	2459	9968	7709	5568	1.6602	1.2840	0.9274	4.0537	3.1350	2.2643	14	45
20	www.gsm.uci.edu	1215	3099	13161	11720	10318	10.8321	9.6461	8.4922	4.2469	3.7819	3.3295	4	46

2.2.1 基于网站被链接频次的核心网站测定结果
 通过对 result.txt 中数据的分析, 我们发现 19 个

商学院网站中共有 212852 个指向域外的链接, 分布于 16388 个网站。其中 80% 的链接集中在被链接频

次最高的623个网站(也就是说,80%的链接指向3.8%的网站,呈现出高度集中的态势)。这623个网站,被链接次数最高的高达15045次,最少的是22次,可以说跨度极大。如果以被链接数22作为核心网站判定的标准,在研究所用的19个样本中,样本4、7、1、13、12的被链接次数分别为191次、81次、53次、35次、28次,可以被判定为核心网站。

但是,由于被链接的网站高度集中,并且绝大部分不属于学术型网站,如果不剔除无关数据,研究结果会出现极大的偏差。因此我们认为,应该以19个样本网站相互链接的次数作为核心网站的测定依据。19个样本网站之间相互链接的次数共计537次,按被链接的频次由高到低排序,我们发现80%的链接分布于被链接次数最多的7个网站,占样本总数的36.84%。这7个网站是样本4、7、1、13、12、2、18。其中,被链接次数最多的是样本4,为191次,最少的是样本2和18,均为19次。

2.2.2 基于Web-IF的核心网站测定结果

本研究计算出了6组Web-IF值。但是,笔者认为指向网站自身内部的链接多为结构性链接,不能作为衡量网站影响力的依据,因而以L1作为计算Web-

IF的依据会影响到研究结果的准确性。同时,由于本研究所考察的对象是独立的主机,而不是主机所在的域,因此采用L2计算Web-IF比L3更合理。本研究采用了两种网页计数方法,但是,由于Mike Thewall的研究表明搜索引擎对不同区域网站的覆盖率存在着明显的差异^[8-11],因而在计算Web-IF时笔者更倾向于采用P1。鉴于上述原因,我们以Web-IF2作为判定核心网站的依据。

由于目前还没有以Web-IF大小确定核心网站的标准,因此,本研究只能将Web-IF2的值由高到低排序,取与前一种方法数量相同的网站,即Web-IF2值最高的7个网站作为核心网站。其结果依次为样本4、1、20、7、6、2、13。将两种方法所确定的核心网站进行比较,我们发现其中有5个是相同的,即样本4、1、7、2、13。

2.3 两种方法测定结果的一致性研究

上述两种方法都可以用于核心网站测定。为了弄清这两种方法研究结果的吻合程度,我们分析了样本网站之间相互链接的频次与Web-IF值的线性相关关系,结果见表2。

表2 样本网站相互之间链接的频次与Web-IF的线性相关分析

XVS	Web-IF1	Web-IF2	Web-IF3	Web-IF4	Web-IF5	Web-IF6
相关系数	0.74178	0.74934	0.68898	0.54865	0.57892	0.59784
P值	0.0003	0.0002	0.0011	0.0184	0.0118	0.0088

表2中的数据显示,X与Web-IF具有显著的相关关系。这些数据还表明,Web-IF1、Web-IF2、Web-IF3与X的相关程度比Web-IF4、Web-IF5、Web-IF6高;并且,Web-IF2与X的相关程度最高,这对我们确定以Web-IF2作为判定核心网站的依据进一步给予了支持。

2.4 网页计数方法对Web-IF值的影响

网页数对Web-IF的计算具有直接影响。本研究发现,P1与P2存在一定的差异,为了澄清这一差异对Web-IF的计算是否构成实质性的影响,我们对6组Web-IF值进行了线性相关分析,结果见表3。

表3 6组Web-IF值的线性相关分析

	Web-IF1	Web-IF2	Web-IF3	Web-IF4	Web-IF5	Web-IF6
Web-IF1	1.00000	0.99654	0.93881	0.68770	0.71013	0.71585
Web-IF1		<.0001	<.0001	0.0016	0.0010	0.0008
	19	19	19	18	18	18
Web-IF2	0.99654	1.00000	0.95812	0.66891	0.69760	0.71965
Web-IF2	<.0001		<.0001	0.0024	0.0013	0.0008
	19	19	19	18	18	18
Web-IF3	0.93881	0.95812	1.00000	0.57016	0.61120	0.70865
Web-IF3	<.0001	<.0001		0.0135	0.0070	0.0010
	19	19	19	18	18	18

续表

	Web-IF1	Web-IF2	Web-IF3	Web-IF4	Web-IF5	Web-IF6
Web-IF 4	0.68770	0.66891	0.57016	1.00000	0.99676	0.94530
Web-IF 4	0.0016	0.0024	0.0135		<.0001	<.0001
	18	18	18	18	18	18
Web-IF 5	0.71013	0.69760	0.61120	0.99676	1.00000	0.96266
Web-IF 5	0.0010	0.0013	0.0070	<.0001		<.0001
	18	18	18	18	18	18
Web-IF 6	0.71585	0.71965	0.70865	0.94530	0.96266	1.00000
Web-IF 6	0.0008	0.0008	0.0010	<.0001	<.0001	
	18	18	18	18	18	18

表3说明,Web-IF1与Web-IF4、Web-IF2与Web-IF5、Web-IF3与Web-IF6虽然具有统计意义上的显著线性相关关系,但相关系数均<0.8,就是说,并没有达到高度相关的程度;而Web-IF1与Web-IF2、Web-IF1与Web-IF3、Web-IF2与Web-IF3、Web-IF4与Web-IF5、Web-IF4与Web-IF6、Web-IF5与Web-IF6表现出高度线性相关关系。因而,总的来说,采用这两种不同方法获取网页数,计算出的Web-IF具有显著的线性相关关系,这对Web-IF可以作为判定核心网站的指标给予了支持。但同时,两组Web-IF并没有表现出高度的线性相关性,这说明P1与P2的差异对Web-IF值有一定的影响,我们在实际应用中必须慎重选择合适的方法。

3 讨论

3.1 几点启示

(1)对样本指向外部的链接分析和统计结果显示,80%的链接指向3.8%的网站,被样本链接的网站高度集中。如果以网站被链接次数为横坐标,具有相应被链接次数的网站数量为纵坐标绘制频数分布图,那么我们得到的将是陡度极大的递减的双曲线在第一象限中的一支,其形状与 $n < 0$ 的幂函数($y = x^n$)在第一象限中的图像极其相似。

(2)在所有被链接的网站中,域名是以.edu结尾的很少,而以.com、.org结尾的占绝大多数。这表明,学术机构网站中的链接与学术期刊中科技文献的引用存在着本质的差异。这还说明学术机构的网站目前并不是开展学术交流和传递学术思想的平台,或者学术信息的组织、开发与利用还没有成为网站设计、建立和维护者的主要目的。

(3)将样本根据相互之间链接次数的多少,由高到低排序;将被链接次数划分成数量相等的三个区

域,相应区域内网站的数量分别为1、3或4、15或14,数量关系基本符合 $1:3.8:3.8^2$ 。就是说,样本相互之间的链接在样本中的分布符合布拉德福定律。

(4)无论是以网站被链接频次,还是以网站的Web-IF值作为核心网站的判定依据,都存在以何种标准计量指向网站的网页数这一问题。我们必须明确是以指向网站的总网页数还是以外部指向网站的网页数为依据。

3.2 核心网站测定应该注意的问题

(1)必须认识到搜索引擎固有的缺陷对研究结果造成的影响。计算Web-IF需要两个参数:指向样本网站的网页数和样本自身所拥有的网页数。目前,前者的获取必须依靠搜索引擎,而对于后者,大部分研究也直接将其等同于搜索引擎索引的该网站的网页数。但是,由于搜索引擎收录范围的局限性、算法的不健全以及系统的波动性,使人们根本没有可能完整而准确地获得这两个参数的实际值。

(2)应该考虑合理地选择网站自身所拥有网页数的计量方法。正如我们前面所提到的,网站网页的计数既可以以特定时刻能够实际访问到的网页数为依据,也有大量研究采用搜索引擎所索引的网页数。表2的数据显示,X与Web-IF1、Web-IF2、Web-IF3的线性相关系数分别比与Web-IF4、Web-IF5和Web-IF6的相关系数大,这表明以实际访问到的网页数计算出的Web-IF值比以搜索引擎索引的网页数计算出的Web-IF值与网站被链接次数的相关度高。

参考文献

- Allison Woodruff. An Investigation of Documents from the World Wide Web. http://www5conf.inria.fr/fich_html/papers/p7/Overview.html,2003.9.23

(下转第41页)

目前网上开放访问期刊主要为外文期刊,而外文期刊因价格昂贵一直是我国所有图书馆收藏中的软肋。据统计,全世界目前发行的学术期刊中约 21000 种为经同行评阅、审稿的高质量期刊,而其中的 6% (约 1200 种) 现已成为开放访问期刊^[12]。图书馆应投入力量,通过以下方式开发利用这部分宝贵的资源,补充、扩大自己的馆藏:关注、收集、整理因特网上的开放访问文献,编制在图书馆的期刊目录和相关数据库中,并在图书馆网页中建立链接,使用户在网上访问图书馆资源时即可方便地了解和使用该部分免费资源;将网上已免费开放访问的电子期刊与图书馆订购的期刊进行比较,整合、删减现有馆藏期刊资源,以节省有限的书刊经费。

图书馆应积极支持、参与开放访问运动,并可具体实施开放访问文献典藏。其一,可利用人力、物力、技术和资源优势,建置本地区、本机构的开放访问电子典藏,如大学图书馆可建立电子预印本典藏服务器,为学校老师开设电子预印本典藏服务,介绍教师、教研人员的研究成果。其二,帮助科研人员将他们的研究论文加入相应学科或相应机构的开放访问典藏。其三,尝试出版开放访问期刊,通过出版开放访问期刊,支持开放访问运动。

参考文献

- 1 Richard J. Roberts, Harold E. Varmus, Michael Ashburner, etc. Information access: Building a "Gen Bank" of the published literature.
- 2 Open Letter to Scientific Publishers. Public Library of Science (PLoS). <http://www.plos.org/about/letter.html>
- 3 Bethesda Statement on Open Access Publishing. <http://www.earlham.edu/~peters/fos/bethesda.htm>.
- 4 IFLA Statement on Open Access to Scholarly Literature and Research Documentation. <http://www.ifla.org/V/cdoc/open-access04.html>.
- 5 Budapest Open Access Initiative. <http://www.soros.org/openaccess/read.shtml>
- 6 Grant Buckler. Open access: Academia's information sharing future. Information Highways, July/August 2004. http://www.econtentinstitute.org/issues/ISarticle.asp?id=152672&story_d=21008102132&issue=07012004&PC
- 7 The green and the gold roads to Open Access. Nature web focus: Access to the literature: the debate continues. <http://www.nature.com/nature/focus/accessdebate/21.html>
- 8 An Open Letter to the U. S. Congress Signed by 25 Nobel Prize Winners. August 26, 2004. <http://www.fas.org/sgp/news/2004/08/nobel082604.pdf>.
- 9 Nature web focus: Access to the literature: the debate continues. <http://www.nature.com/nature/focus/accessdebate>
- 10 The Impact of Open Access Journals—A Citation Study from Thomson ISI. <http://www.isinet.com/media/presentrep/acropdf/impact-oa-journals.pdf>
- 11 Directory of Open Access Journals (DOAJ). <http://www.doaj.org>
- 12 Tenopir, C. Online Scholarly Journals: How Many. Library Journal. 2/1/2004. <http://www.libraryjournal.com/article/CA374956>

- 1 Richard J. Roberts, Harold E. Varmus, Michael Ashburner, etc. Information access: Building a "Gen Bank" of the published literature.
- 2 Open Letter to Scientific Publishers. Public Library of Science (PLoS). <http://www.plos.org/about/letter.html>
- 3 Bethesda Statement on Open Access Publishing. <http://www.earlham.edu/~peters/fos/bethesda.htm>.
- 4 IFLA Statement on Open Access to Scholarly Literature and Research Documentation. <http://www.ifla.org/V/cdoc/open-access04.html>.
- 5 Budapest Open Access Initiative. <http://www.soros.org/openaccess/read.shtml>
- 6 Grant Buckler. Open access: Academia's information sharing future. Information Highways, July/August 2004. http://www.econtentinstitute.org/issues/ISarticle.asp?id=152672&story_d=21008102132&issue=07012004&PC
- 7 The green and the gold roads to Open Access. Nature web focus: Access to the literature: the debate continues. <http://www.nature.com/nature/focus/accessdebate/21.html>
- 8 An Open Letter to the U. S. Congress Signed by 25 Nobel Prize Winners. August 26, 2004. <http://www.fas.org/sgp/news/2004/08/nobel082604.pdf>.
- 9 Nature web focus: Access to the literature: the debate continues. <http://www.nature.com/nature/focus/accessdebate>
- 10 The Impact of Open Access Journals—A Citation Study from Thomson ISI. <http://www.isinet.com/media/presentrep/acropdf/impact-oa-journals.pdf>
- 11 Directory of Open Access Journals (DOAJ). <http://www.doaj.org>
- 12 Tenopir, C. Online Scholarly Journals: How Many. Library Journal. 2/1/2004. <http://www.libraryjournal.com/article/CA374956>

马景娣 浙江大学图书馆副研究馆员。通信地址: 杭州市。邮编 310027。(来稿时间: 2004-09-30)

(上接第 23 页)

- 2 L. Cui (1999), Rating health Web sites using the principles of citation analysis: A bibliometric approach. Journal of Medical Internet Research, 1(1) e4 (ISSN: 1438-8871)
- 3 Robert C. Vreeland. Law libraries in hyperspace: A citation analysis of world wide web sites. Law Library Journal. 2000, 92(1)
- 4 蒋颖. 我国图书情报站点的“引用”分析. 情报资料工作, 2001(3)
- 5 Peter Ingwersen. The calculation of web impact factors. Journal of Documentation. 1998, 54(2)
- 6 Alastair G. Smith. A tale of two web spaces: comparing sites using web impact factors. Journal of Documentation. 1999, 55(5)
- 7 邱均平, 安璐. 中文期刊影响因子与网络影响因子和外部链接数的关系研究. 情报学报, 2003, 22(4)
- 8 Mike Thelwall. A comparison of sources of links for academic

- web impact factor calculations. Journal of Documentation. 2002, 58(1)
- 9 Mike Thelwall. Web impact factors and search engine coverage. Journal of Documentation. 2000, 56(2)
- 10 Mike Thelwall. Results from a web impact factor crawler. Journal of Documentation. 2001, 57(2)
- 11 Mike Thelwall. Conceptualizing documentation on the web: An evaluation of different heuristic-based models for counting links between university web sites. Journal of the American Society for Information Science and Technology. 2002, 53(12)

段宇锋 华东师范大学商学院信息学系教师。通信地址: 上海。邮编 200062。

邱均平 武汉大学中国科学评价研究中心教授。通信地址: 武汉。邮编 430072。(来稿时间: 2004-08-27)