

# 外部网络影响因子在大学网站评估的应用

吴茵茵

(华南农业大学, 广州 510642)

**〔摘要〕** 通过对网络影响因子的计算方法进行改进, 着重介绍外部网络影响因子对网站的影响力进行评估并提出建设性意见。利用外部网络影响因子和总体网络影响因子对中国大学网站的影响力进行计算和对比, 并通过引入链接效率对两种网络影响因子的有效性进行回归分析, 结果表明外部网络影响因子能够比总体网络影响因子更为有效和客观地评价网站的影响力。

**〔关键词〕** 网络计量学; 网络影响因子; 回归分析

**〔Abstract〕** This paper was to investigate the application of external web impact factor (WIF) in the evaluation and development of the websites based on the improved calculating method of WIF. The websites of Chinese Universities were evaluated by using overall WIF and external WIF, and the validity of these two WIFs was tested through the regression analysis. The results indicated that the external WIF can reflect the impact of the website more efficiently and more objectively than the overall WIF.

**〔Key words〕** Webometrics; Web impact factor; regression analysis

**〔中图分类号〕** TP393.1/4 **〔文献标识码〕** C **〔文章编号〕** 1008-0821(2008)04-0066-03

上世纪末以来计算机网络技术得到了异常迅猛的发展, 信息的存储也由纸质存贮、模拟设备存放发展到全方位的数字化存储。在此过程中, 信息资源存贮的数字化和信息资源传播网络化的进程不断加快, 如何对海量的数字化信息进行计量直接促成了网络计量学的产生。网络计量学(Webometrics)由 T.C.Almind 与 Peter Ingwersen 在 1997 年正式提出, 其基本原理是将因特网看作引文网络, 将网页看作引文, 利用信息计量学方法对网络信息进行计量研究<sup>[1]</sup>。由于网络环境的复杂性和新颖性, 网络计量学已经逐渐从传统的信息计量学分离出来而形成了一门新兴的独立学科。

网络影响因子(Web Impact Factor, WIF)作为网络计量学研究中的一个重要分支, 最早由 Ingwersen<sup>[2]</sup>于 1998 年提出: 某一时刻链接到网络上某一特定网站或区域的网页数与这一网站或区域本身所包含的网页数的比值就是网络影响因子。Mike Thelwall 在其 2000 年发表的研究论文中则指出网络影响因子中的分母应该是该网站或区域的大小<sup>[3]</sup>。该网站或区域的大小并不一定局限于该网站的网页数, 也可以用其他指标来进行衡量, 如对学术机构的网站进行研究时也可以是该机构的研究人员数、该机构的研究经费或者全日制学生的数目。由定义可以看出, 网络影响因子的提出就是为了对网络环境下的网站影响力进行评估。根据链接网页性质的不同, 可以分为链接总数、外部链接数、内部链接数, 由此网络影响因子也就分为了总体网络影响因子、外部网络影响因子和内部网络影响因子。

国外对于网络影响因子的研究较早, Ingwersen 在 1998 年选择了 7 个国家、4 个顶级域名和 6 个学术机构的网站得到了网络影响因子的数值<sup>[2]</sup>; Smith 在 1998 年对选取的澳大利亚的大学或研究机构网站和电子期刊网站的网络影响

因子进行了评价<sup>[4]</sup>; Smith 与 Thelwall 对英国、澳大利亚和新西兰大学之间的相互连接部分进行了统计<sup>[5]</sup>; Thomas 和 Willett 在 1999 年也对英国大学图书情报学系网站的网络影响因子做了分析<sup>[6]</sup>。他们使用的检索工具都是 AltaVista 搜索引擎。国内近几年来也对网络影响因子进行较多的研究<sup>[7-12]</sup>, 其中多数使用的也是 AltaVista 搜索引擎对中国大学网站进行网络影响因子分析。在这些研究中, 单一的网络影响因子并不能客观代表网站的影响力已成为共识, 同时邱均平等人也指出了外部链接数对网站评价的有效性<sup>[9]</sup>。但是, 尚未有人对于外部网络影响因子对网站评价的有效性做出完整深入的评估。本文中笔者拟用回归分析的方法来对中国大学网站的外部网络影响因子进行分析, 探讨其在网站评估中的作用。

## 1 研究方法和结果

在此次试验中, 我们选取的 14 所中国大学, 包括清华大学、北京大学、浙江大学、复旦大学、南京大学、华中科技大学、武汉大学、西安交通大学、吉林大学、上海交通大学、四川大学和中山大学。

因为不同的搜索引擎使用不同的搜索语法, 所以下面以清华大学为例来说明 AltaVista 搜索引擎的搜索语法以及网络影响因子的计算方法(见表 1)。

表 1 AltaVista 搜索引擎的数据搜索方法与网络影响因子计算方法

搜索项目	搜索命令/计算方法
网页总数	D = domain: tsinghua.edu.cn
链接总数	A = linkdomain: tsinghua.edu.cn

收稿日期: 2008-01-11

基金项目: 本文为华南农业大学校长基金项目“基于网络计量学的华南农业大学科研实力研究”(2005K049) 研究论文之一。

作者简介: 吴茵茵(1979—), 女, 华南农业大学图书馆馆员, 已发表论文 4 篇, 合作著作 3 部。



续表 1

外部链接总数	$B = \text{linkdomain: tsinghua.edu.cn NOT host: tsinghua.edu.cn}$
总体网络影响因子	链接总数/网页总数 = $A/D$
外部网络影响因子	外部链接总数/网页总数 = $B/D$
链接效率	外部链接总数/链接总数 = $B/A$

对以上的各项测量指标,笔者需要做出三点说明:

1.1 在本文中我们并没有考虑内部链接总数以及内部网络影响因子,因为对于网站影响力的评注中,最重要的是外部链接的数量。内部链接数量不仅能说明网站内部的运行效率高、使用需求量大,但却无法反映出网站对于外界的影响,自然也就无法反映网站的社会影响力;

1.2 在链接命令中,笔者并没有使用 domain: www. tsinghua.

edu.cn 的命令而是使用了 domain: tsinghua. edu. cn, 这是因为多数网页的地址并不是以 www 开头,如果使用前者进行搜索将会错过绝大多数的网页链接。例如用 domain: tsinghua. edu. cn 进行搜索,得到的网页总数为 928 000,而用 domain: www. tsinghua. edu. cn 进行搜索,仅得到 27 100;

1.3 之所以引入链接效率,是因为外部链接总数的绝对量并不能完全反映网站的影响力,而其所占链接总数的比例才能够较为客观地说明网站对于外部社会的影响力。外部链接所占的比例越高,其链接效率就越高,那么网站的相对影响力就越大。因此,链接效率是反映网站影响力大小的一个重要指标。

笔者在 2007 年 8 月 12 日下午 3 时至 5 时利用 AltaVista 搜索引擎对中国 14 所著名大学进行网络影响因子各项指标的测定,具体结果如表 2 所示。

表 2 中国大学网站的网络影响因子测定

大 学	链接总数 (A)	外部链接总数 (B)	网页总数 (D)	链接效率 ( $IE = B/A$ )	总体网络影响因子 ( $WIF = A/D$ )	外部网络影响因子 ( $WIF - r = B/D$ )
清华大学	655 000	561 000	928 000	86%	0.71	0.60
北京大学	861 000	330 000	911 000	38%	0.95	0.36
浙江大学	482 000	292 000	2 560 000	61%	0.19	0.11
上海交通大学	493 000	288 000	2 480 000	58%	0.20	0.12
西安交通大学	83 700	45 100	95 600	54%	0.88	0.47
复旦大学	224 000	208 000	266 000	93%	0.84	0.78
南京大学	191 000	68 900	264 000	36%	0.72	0.26
华中科技大学	110 000	55 800	99 500	51%	1.11	0.56
武汉大学	203 000	93 400	601 000	46%	0.34	0.16
中山大学	53 100	32 700	589 000	62%	0.09	0.06
中国科技大学	272 000	162 000	1 810 000	60%	0.15	0.09
哈尔滨工业大学	89 600	30 500	386 000	34%	0.23	0.08
吉林大学	117 000	92 400	155 000	79%	0.75	0.60
四川大学	53 600	51 700	41 300	96%	1.29	1.25

从表 1 中可以看出:(1)从网页总数的绝对数量来看,浙江大学、上海交通大学以及中国科技大学居前三位,但是它们的链接总数和外部链接总数并没有明显高出其他的高校,而这直接导致了它们的网络影响因子较其他高校为低。四川大学和华中科技大学的总体网络影响因子大于 1,这于它们较小的网页总数以及链接数量有关,与其他高校相比至少小了一个数量级。中山大学由于其网页链接数量和网页总数相差过大,导致其网络影响因子小于 0.1,这可能于搜索引擎的误差有关;(2)通过比较外部网络影响因子和总体网络影响因子,可以发现一方面一些高校的网络影响因子前后发生了较大变化,比如北京大学由 0.95 降为 0.36,西安交通大学由 0.88 降为 0.47,哈尔滨工业大学由 0.23 降为 0.08,武汉大学由 0.34 降为 0.16,降幅多数超

过了 50%。而其他高校如清华大学、浙江大学、四川大学等则未有明显变化;(3)14 所高校中 10 所高校的链接效率超过了 50%,而有 4 所高校的链接效率低于 50%,这说明它们的网络影响力多数来源于自身的内部链接而对社会的影响力明显偏小。这种偏差在比较两种网络影响因子的时候体现得特别明显,而这也就能说明总体网络影响因子由于没有区分内部和外部网页链接从而无法客观的反映真正的网络影响力;(4)在对比链接效率和两种网络影响因子的时候,可以发现链接效率较高的大学,其外部网络影响因子也比较高,反之则较低。而当链接效率低于 50% 时,两种网络影响因子的偏差就会较大,反之则会相对稳定;(5)通过与笔者以前研究的比较,可以发现搜索命令的差异会导致搜索结果的显著不同。在文献 [10] 中,利

用相同的搜索引擎但是用domain: www.tsinghua.edu.cn 命令搜索到的网页总数2 257, 而本文中利用domain: tsinghua.edu.cn 得到的结果为928 000, 排除时间的影响, 由于搜索命令不同导致搜索结果的差异仍然是十分明显的。

为了进一步地确认链接效率和两种网络影响因子的关系, 笔者运用回归分析的方法对链接效率和总体网络影响因子与外部网络影响因子进行了相关分析, 其结果如图1和图2所示。对链接效率和总体网络影响因子的回归分析结果为:  $F(1, 12) = 1.501$ ,  $p = 0.244$ , 皮尔逊相关指数  $r = 0.33$ 。由于  $p$  大于 0.05, 此结果表明链接效率和总体网络影响因子之间并无线性相关性, 也就是说链接效率的变化并不会引起总体网络影响因子的变化。因此, 在对网站进行总体性评价时, 总体网络影响因子并不是可以值得信赖的指标。对链接效率和外部网络影响因子的回归分析结果为:  $F(1, 12) = 13.290$ ,  $p = 0.003$ , 皮尔逊相关指数  $r = 0.73$ 。由于  $p$  明显小于 0.05, 并且  $r$  的数值大于零且接近 1, 所以两者存在显著的线性正相关性, 图2清晰地表现出了这一点。这说明链接效率的变化能够直接地反映出外部网络影响因子的变化, 链接的效率越高, 外部网络影响因子就越大, 其社会影响力就会越大, 而这种变化是比较符合实际情况的。

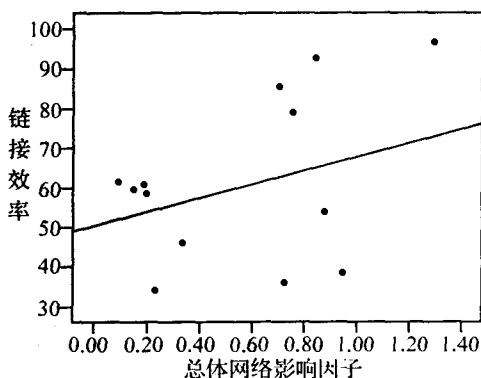


图1 链接效率与总体网络因子的回归分析曲线

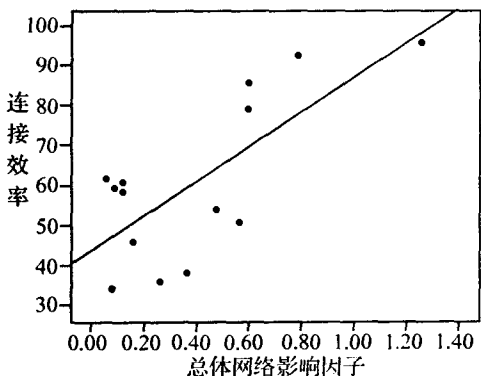


图2 链接效率与外部网络因子的回归分析曲线

## 2 讨论

网络影响因子的初衷是利用网络时代下的搜索引擎来对网站的社会影响力进行有效地评估, 所以它的基本要求是能够客观地反映网站的影响力变化。而实际上由于众多复杂的因素, 网络影响因子从一面世就存在对其客观性的质疑。从本文以及以前的研究来看, 单一的总体网络影响

因子指数无法客观和有效地反映网站的社会影响力, 其原因在于: (1) 总体网络影响因子没有考虑对内部和外部的网页链接进行区分, 导致将内部影响力和外部影响力混为一谈, 从而从原理上就没有完全符合网络影响因子的定义; (2) 网络影响因子的计算定义是链接总数和网页总数的比值, 无论是总体、外部和内部网络影响因子, 它们都抹煞了网页总数和链接总数对网站影响力的绝对影响。也就是说, 网站的影响力的一个重要因素就是网页和链接的绝对数量, 在一定程度上这两者的绝对数量越大表示其社会影响力越大。而网络影响因子恰恰是没有如实的反映出这一点, 网络影响因子经常表现出与绝对数量变化相反的趋势; (3) 搜索引擎的差异以及搜索命令的不同也会导致总体网络影响因子的差异。虽然 AltaVista 被认为进行网络计量学研究的最佳搜索引擎, 但它的稳定性和可靠性仍待提高, 特别是其布尔逻辑性的稳定性。以前的研究已经证明了 AltaVista 的布尔逻辑性有着一定程度的不稳定性, 比如  $A \cap B$  和  $B \cap A$  并不总是返回相同的搜索结果。在本文的研究中还发现 linkdomain: xxx.edu.cn 返回的搜索结果有时会比 linkdomain: xxx.edu.cn NOT host: xxx.edu.cn 少, 而从逻辑上而言后者是前者的一部分, 因此外部链接总数一定会小于链接总数。这种逻辑上的不稳定性会对搜索结果造成不可预测的影响, 并且会干扰对结果的正确解释。在这种情形下, 总体网络影响因子的脆弱性就显露无疑, 从而其他分析参数的补充或者替代就显得更加重要。

要克服这些缺陷, 基本上有两种方法可以选择: (1) 添加辅助性的分析参数来弥补总体网络影响因子的不足, 建立根据各种参数进行综合性评价的分析体系, 以有效的弥补各方面的不足; (2) 彻底抛弃网络影响因子的分析, 转而寻找更为科学有效的分析方法。从目前的情况来看, 由于存在种种的技术障碍以及理论探索的不足, 完全抛弃网络影响因子分析方法并不现实, 而建立综合评价体系的选择是较为现实的。而这需要对各种参数的有效性进行深入分析, 以确定它们各自在评价体系中的作用, 便于将来在系统中能够较为客观的进行权重分配。

有鉴于此, 笔者对于外部网络影响因子的有效性进行了系统分析。通过引入与链接效率的回归分析, 本文的研究结果表明外部链接数量占网页总数的比重变化能够明显地引起外部网络影响因子的变化, 这是一种线性传导的正相关关系。由于外部链接数量更能从根本上体现出其他站点网站的链接程度, 因此就客观性而言, 外部网络影响因子能够比总体网络影响因子更为客观地反映网站的实际社会影响力。而链接效率与外部网络影响因子的正相关性说明要想提高网站的影响力, 就必须降低内部链接数量。过高的内部链接数量来自于网站内部的公告、通知、论坛乃至后勤管理等, 这些链接并不利于提高网站的外部影响力, 特别是不利于提高高等院校的科研影响力。因此, 改进网站内部的架构对于提高链接效率并进一步提升网站影响力有着重要的影响。对于大学网站而言, 增加学术科研类内容而降低管理通知类内容, 增加内容丰富性而降低管理复杂性, 对于提高网站的链接效率和外部网络影响因子会有直接的帮助, 从而能够进一步地提高网站影响力。

(下转第71页)



同时提高他们的文化素质。目前,还可就公众关心的社会热点问题,如再就业、市场物价、房地产、股票、证券等开展广泛的信息咨询服务活动。另一方面可充分发挥图书馆窗口作用,因地制宜地开展读书活动,进一步推进社区青少年读书活动。

### 3.5 加强网络建设,实现城乡资源信息共享

高校图书馆应发挥自身的现代化网络设备和高素质的信息服务人才优势,广泛开展网上资源信息服务,使图书馆成为真正意义上的社会信息中心与重要基地。重庆高校图书馆除了自身想方设法扩大信息源外,还必须有全局观念,加强同有关单位与其它图书馆的联系,组织联合目录,实行馆际互借,建立统一的信息网体系。图书馆要利用自己占有信息资源,快速进行信息加工的能力的优势。利用网络环境,对电子资源进行分析与重组。图书馆服务人员要掌握信息源,了解信息资源分布状况,信息网络的分布,熟悉网上信息机构,更好地组织网上信息,为读者服务。开展有组织、相互协调的信息咨询服务,这是扩大图书馆信息咨询服务功能的重要途径,也是实现资源共享的有效方法。在加强网络服务的同时,开展网上信息咨询服务。

### 3.6 加大科研力度,创新科技信息服务模式

为了适应重庆统筹城乡信息化建设需要,重庆高校图书馆应积极加大科研力度,加强信息化建设的研究,努力创新图书馆信息服务模式。

#### 3.6.1 建立和提供富有特色的信息资源

在信息化时代,图书馆的职能价值不仅以其馆藏专业文献的规模和质量来体现,而更要以其提供全面、科学、创新、快捷、准确的专业信息的能力来体现。在现阶段,还应坚持以城乡经济建设和信息化建设为新的服务方向,要对原有的文献和信息资源充分挖掘,并重新整合和优化

(上接第68页)

## 3 结论

根据上述研究结果,我们可以得出以下结论:

- 3.1 与总体网络影响因子相比,外部网络影响因子更能够较为客观和有效地反映出网站的实际影响力;
- 3.2 链接效率的变化与外部网络影响因子的变化有着显著的线型正相关关系,说明链接效率能够作为一个辅助性的指数来客观地评价网站的影响力;
- 3.3 单一的总体网络影响因子不能完整客观的反映网站的影响力,而外部网络影响因子、链接效率以及外部网页链接数量的综合运用能够更加有效地展现网站的社会影响力,并且能够对网站的进一步的完善与发展提供建设性意见。

## 参考文献

- [1] Almind, T.C. and Ingwersen, P. Informetric analyses on the World Wide Web: methodological approaches to "Webometrics" [J]. Journal of Documentation, 1997, 53 (4): 404-426.
- [2] P. Ingwersen. The calculation of web impact factors [J]. Journal of Documentation, 1998, 54 (2): 236-243.
- [3] Thelwall, M. Web impact factors and search engine coverage [J]. Journal of Documentation, 2000, 56 (2): 185-189.
- [4] Smith A.G. A tale of two web spaces: comparing sites using web impact factors [J]. Journal of Documentation, 1999, 55

资源配置,扬长补短,根据自身资源上的优势以及周边社会环境的特点,形成具有一定特色的服务方式和服务内容,为社会提供方便快捷、新颖独特的信息服务。

### 3.6.2 走“馆企合作”道路

走“馆企合作”是对校企合作的延伸与拓展。高校图书馆可以尝试企业和单位合作,发挥自己信息收集和处理的特长以及在情报信息知识上的独特优势,参与企业和单位开发新产品、进行科研攻关的工作,积极配合企业,真正实现“馆企合作”,使得图书馆的工作高效率、高质量地直接为社会服务。像重庆大学、西南大学、重庆交通大学等技术力量强的高校图书馆,甚至可以开展一些专业性更强的定题服务、专题服务和跟踪服务,深度加工出索引类、汇编类、预测类的信息产品,从更高的层面上,为社会提供支持和服务。

总之,在重庆统筹城乡经济配套改革和信息化建设中,重庆高校图书馆应加强自身现代化建设,充分发挥信息服务功能,通过多种途径为信息化建设服务。

## 参考文献

- [1] 朱萍.论高校图书馆服务社会化的必然性及实现途径[J].高校图书馆工作,2005,(2).
- [2] 董文华.论高校图书馆的社区服务功能[J].图书情报知识,2005,(3).
- [3] 李娟娟.高校图书馆为地方经济建设服务思考[J].高校图书馆工作,2003,(6).
- [4] 童丽娜.高校图书馆信息咨询服务的优势与发展[J].图书馆论坛,2005,(2).
- [5] 陈能华,刘昆雄.信息化环境下我国图书馆业务管理模式变革研究[J].中国图书馆学报,2004,(3).
- [5] 577-592.
- [5] Smith, A. and Thelwall, M. Web impact factors and university research links [J]. Proceedings of the 8th International Conference on Scientometrics & Infometrics, Vol 2, Sydney Australia, 2001, 16-20 July, 2001, 657-664.
- [6] Thomas, O. and Willett P. Webometric analysis of departments of librarianship and information science [J]. Journal of Information Science, 2000, 26 (6): 421-428.
- [7] 杨涛.网络信息计算学实证研究:对国内20个大学网站的分析[J].图书情报工作,2003,(9): 61-66.
- [8] 邱均平,陈敬全,段宇峰.中国大学网站链接分析及网络影响因子探讨[J].中国软科学,2003,(6): 151-155.
- [9] 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[J].情报学报,2003,(4): 398-402.
- [10] 吴茵茵.网络影响因子实证研究:基于不同搜索引擎的大学网站影响力分析[J].图书情报工作,2005,(4): 107-111.
- [11] 吴茵茵.不同搜索引擎在网络影响因子分析中的比较研究[J].情报科学,2005,23(3): 431-435.
- [12] 蒋心瑜.网络影响因子及其在网络计量学中的应用[J].图书情报工作,2006,50(4): 106-109.