

网络计量学在医学网站评价中的应用研究

蔡筱英 胡 伟

(华中科技大学医药卫生管理学院 武汉 430030)

〔摘要〕 应用网络计量学方法, 通过 All The Web 和 Alta Vista 两种搜索引擎在不同时间段收集两轮数据, 分析国内主要医学网站的链接数、网络影响因子与网络评价排名的相关关系。结果显示医学网站的外部链接数和总链接数与网络评价中的排名具有显著的相关性, 而网络影响因子和网站排名无相关性, 同时指出链接分析法存在的不足。

〔关键词〕 网络计量学; 链接分析法; 网络影响因子; 网站评价

Application Research on Webometrics on Evaluating Medical Websites CAI Xiao - ying, HU Wei, Health and Medicine School of Management of Huazhong University of Science and Technology, Wuhan 430030, China

〔Abstract〕 Adopting webometrics method, the paper collects two groups of data in different time through two search engines; all the web and alta vista, analyzes the correlation among link number, web impact factor and website evaluation ranking of domestic main medical websites. The result shows that number of external links and total number of links have significant correlation to the ranking, but the web impact factor is not related to the ranking. It also points out the insufficient of link analysis.

〔Keywords〕 Webometrics; Link analysis; Web impact factor; Website evaluation

网络信息计量学是通过对网上信息的计量分析, 为网上信息的有序化组织和合理分布, 及网络信息资源的优化配置和有效利用提供必要的定量依据。网络计量学主要利用搜索引擎工具对网络资源进行分析, 通过对网络链接的数量、分布等方面的分析研究, 揭示网站的信息量、性质等特征, 并反映网站的质量。网络链接分析中基本涉及到的指标主要有 4 项: 总链接数、内部链接数、外部链接数、总网页数。利用网站的链接和被链接情况, 可以确定网站的学术影响力和社会影响力。

网站的链接数和网络影响因子被看作是评价网

站的重要指标。通过网络影响因子 (WIF) 的计算、网站或网页点击率的统计, 可以确定类似于核心期刊的“核心网站”或“核心网页”。“核心网站”和“核心网页”的确定有助于用户快速查找和选择利用网络信息, 增强上网的目的性, 提高获取所需信息的准确性和快捷性。

1 研究内容

应用网络计量学方法, 从国内主要医学网站的链接情况与医学网站排名入手, 通过对两种搜索引擎在不同时间段收集到的两轮数据进行分析, 来研究以下几项内容: 探讨国内主要医学网站的链接数、网络影响因子与这些医学网站在国内网络评价中的医学网站排名的相关性; 比较利用不同的搜索

〔修回日期〕 2009 - 11 - 13

〔作者简介〕 蔡筱英, 副教授, 发表论文 30 余篇, 主编、参编论著 5 部。

引擎得到的结果是否一致，如果不一致，哪一个更可靠；就某一搜索引擎而言，比较不同时间段收集到的两轮数据得到的结果是否一致，如果不同时期收集的数据不一致，是否进行多轮数据的收集有利于得到一个较稳定的结论。

2 资料与方法

2.1 研究对象

选取了在有关医学评论网站网络评价中比较有影响的医学网站 50 个。包括医学院校网站 15 个，其排名根据网络评价中国内医学院校排名情况从排名第 1 的北大医学院到排名第 15 的哈尔滨医科大学依次收集；国内综合性医院 15 个，其排名也是依据网络评价中国内综合性医院综合实力的排名情况从排名第 1 的北京协和医院到排名第 15 的上海长征医院依次收集；医药公司网站 10 个和健康卫生网站 10 个，这 20 个是根据网络评价中声誉较好的网站进行排名依次收集的（排名表略）。

2.2 研究方法

链接分析法是目前网络计量学中应用最多的一

种方法。链接分析与网络计量学几乎相伴而生，链接分析也是网络计量学催生的动力。它是运用网络数据库、数学分析软件等工具，利用数学（主要是统计学和拓扑学）和情报学方法，对网络链接自身属性、链接对象、链接网络等各种对象进行分析，以便揭示其数量特征和内在规律，并用以解决各方面问题的一种研究方法。

目前网络信息计量研究主要利用商用搜索引擎（如 Alta Vista）获取相关数据。为避免搜索引擎给搜集结果带来误差，选用 Alta Vista 和 All The Web 两种搜索引擎对研究对象进行搜索。All The Web 和 Alta Vista 是著名的搜索引擎，具有检索网页的站内外链接功能，能实现复杂的检索任务，是国外链接分析的主要工具，其可用性在国外研究中得到了证实。链接数据的计算采用直接计算法在 SPSS15.0 上进行。

2.3 数据搜集方法

链接数的收集方法：下面以上海交通大学医学院网站（www.shsmu.edu.cn）为例来介绍链接数的收集方法，见表 1。

表 1	链接数的收集方法		
	外部链接	网页	总链接
All The Web	Link; http://www.shsmu.edu.cn - domain; www.shsmu.edu.cn	domain; www.shsmu.edu.cn	Link;http:// www.shsmu.edu.cn
Alta Vista	Link;www.shsmu.edu.cn - domain; www.shsmu.edu.cn	domain; www.shsmu.edu.cn	Link; www.shsmu.edu.cn

在收集数据中发现部分医学网站拥有多个域名，因此对于此类网站分别收集了其网站各子域名的各链接数数值，然后进行求和作为该医学网站的数据。如上海交通大学医学院目前的域名有 www.shsmu.edu.cn 和 www.mechina.org，那么上海交通大学医学院的外部链接数就应该是这两个子域名的外部链接数之和，网页数和总链接计算方法同理。为了验证搜索引擎的稳定性，本文对两个搜索引擎都进行了两轮数据收集，时间间隔为 30 天，分别是 2008 年 4 月 3 日和 2008 年 5 月 3 日，但由于在研究后期 Alta Vista 不能访问，所以只对 All The

Web 进行了两轮的数据收集。

计算各网站的网络影响因子：对各医学网站分别计算了以下几种网络影响因子：（1）总网络影响因子（WIFt），即该网站（或域）空间的所有链接程度， $WIF_t = \text{总链接数} / \text{网页总数}$ ；（2）外部网络影响因子（WIFe），即外部链接到该网站（或域）空间的链接程度， $WIF_e = \text{外部链接数} / \text{网页总数}$ ；（3）内部网络影响因子（WIFI），即该网站（或域）空间内部自身页面链接程度， $WIF_i = (\text{总链接数} - \text{外部链接数}) / \text{网页总数}$ 。

2.4 数据统计分析

采用 SPSS 中的 Spearman 等级相关分析来检验医学网站链接数与网站排名中的得分之间的相关性。

3 结果与分析

表 2 前 15 个医学院校网站链接数

医学院校	All The Web 的总链接数	All The Web 的网页数	Alta Vista 的总链接数	Alta Vista 的网页数
北京大学医学部	9 630	10 100	6 000	9 970
复旦大学上海医学院	242	1 500	250	1 500
上海交通大学医学院	393	1 130	336	1 130
中南大学湘雅医学院	145	1 620	224	1 620
四川大学华西医学中心	1 510	1 210	1 010	1 200
华中科技大学同济医学院	2 710	8 890	1 880	6 510
首都医科大学	10 600	8 790	3 760	12 200
中国医科大学	23 900	14 600	12 400	18 800
吉林大学白求恩医学院	15 800	10 400	31 600	10 400
浙江大学医学院	1 250	68 700	1 250	68 700
天津医科大学	2 210	22 500	6 010	22 400
南京医科大学	7 050	8 470	7 220	8 490
山东大学医学院	605	1 450	615	1 450
武汉大学医学部	82	1 250	71	1 260
哈尔滨医科大学	6 700	10 100	7 190	10 200

3.2 医学网站链接数与医学网站排名的相关性

由于不同类型的医学网站排名差异很大，所以只分析医学院校的网站链接数及其影响因子与这些医学网站排名的相关关系，见表 3，表 4。

表 3 All The Web 网站链接数与排名的相关关系

名称	外部链接数	总链接数	WIFt	WIFe	WIFI
相关系数	0.709	0.685	-1.05	-1.26	-3.15
P 值	<0.001	<0.001	0.354	0.396	0.036

表 4 Alta Vista 网站链接数与排名的相关关系

名称	外部链接数	总链接数	WIFt	WIFe	WIFI
相关系数	0.726	0.700	-1.35	-1.49	-3.08
P 值	<0.001	<0.001	0.362	0.399	0.045

链接分析类似于文献计量学的引文分析，也是采用数学方法和逻辑方法，同样也有一些基本假设，可以概括为以下几点：某网站（网页）被链接与该网站（网页）的质量有正向（肯定）的联系；被链接者与链接者在内容上是相关的；所有的链接

3.1 医学网站链接数

由于医学院校、医院、医药公司和健康卫生网站的性质不同，在链接数上有较大的差异，所以分别计算其链接数，比较其链接数与排名之间的相关性，阐述网络计量学在其评价中的应用。在此只列出前 15 个医学院校的网站链接数，见表 2。

都是等价的。从上面的数据中可以看出，无论是 All The Web 还是 Alta Vista，医学网站的总链接数和外部链接数与医学网站排名之间都存在显著的相关性。

3.3 网络计量学研究中搜索引擎的稳定性验证

为了验证搜索引擎的稳定性，利用 All The Web 收集到两轮数据，将其合并，合并的具体方式是分别取外部链接数、网页数和总链接数的平均值，并重新计算 3 种网络影响因子，最后得出总链接数和外部链接数以及 3 种网络影响因子与医学网站排名中的总得分之间的相关性，见表 5。

表 5 利用 All The Web 收集的两轮数据的相关性

名称	外部链接数	总链接数	WIFt	WIFe	WIFI
两轮合并相关系数	0.708	0.691	-1.08	-1.23	-3.20
两轮合并 P 值	<0.001	<0.001	0.355	0.409	0.050

可见两轮数据的结果非常接近，而且在对两轮数据进行合并后，医学网站的外部链接数和总链接

数与医学网站排名之间的相关性更显著。同时，本研究还对来自 All The Web 的两轮数据进行了一个相关性测试，见表 6。可以看出利用 All The Web 收集到的两轮数据具有非常显著的相关性，意味利用 All The Web 在两个不同时间段收集到的数据是非常相似或一致的。

表 6 All The Web 收集的两轮数据的相关性

名称	外部链接数	网页数	总链接数
相关系数	0.969	0.971	0.989
P 值	<0.001	<0.001	<0.001

4 结论

4.1 网站链接数与网站排名的相关性

根据统计分析，医学网站的总链接数和外部链接数与医学网站排名之间都存在较显著的相关性。这意味着排名较高的医学网站会吸引更多的链接数，因此，网站的总链接数和外部链接数都可以作为评价网站的指标。这说明医学网站的实力越强、声誉（尤其是学术声誉）越高，指向其网站的链接就越多。

4.2 网络影响因子与网站排名的相关性

总网络影响因子和外部影响因子与医学排名无相关性，这说明用评价学术网站的方法计算的网路影响因子并不适用于医学网站，这是因为目前国内医学网站主要是介绍性、发布性的内容，而真正纯学术性的内容还比较少。此外，商业搜索引擎的搜索能力是有限的，所以无法保证搜索引擎能够搜索到全部的链接数和网站下的所有网页，用平均每页的被链接率来评价医学网站就目前来说并不合适。专业网络影响因子和人均网络影响因子在两种搜索引擎中的相关性表现极不一致，可能需要加大样本进行再一次的研究分析。

4.3 医学网站链接分析法存在的不足

一些医学网站虽有很高的外部链接数，但很多链接来自公告、通知、新闻等方面，这些都不足以体现一个医学网站的科研水平和影响力。本文改进

的结果体现了与医学网站排名的一致性，但这种改进仍存在不足。首先，医学网站的网络评价也应分类，只有同种类型的医学网站才能进行比较；其次，健康卫生网站的学术交流链接更频繁，这种比较也存在不公平之处。不论是从 All The Web 和 Alta Vista 等搜索引擎的稳定性还是从 WIF 本身客观性角度考虑，WIF 在实际应用中仍然存在着不少问题。现存的网络影响因子的概念在实际运用中还只是一个相对粗浅的提法，因而所得出的结论也存在一定的问题。

参考文献

1 付威风, 郑春厚. 网络计量学常用工具比较研究 [J]. 图书馆学研究, 2009, (7): 24-27.

2 倪素云, 许培扬. 网络计量学研究的现状与进展 [J]. 医学信息学杂志, 2007, (3): 217-220.

3 王知津, 张收棉. 网络计量学方法在网站评价中的应用及其理论模型 [J]. 图书与情报, 2006, (3): 34-39.

4 杨木容. 搜索引擎在网络链接分析中的应用分析 [J]. 图书情报工作, 2006, 50 (11): 92.

5 杨涛. 网络信息计量学实证研究: 对国内 20 个大学网站的分析 [J]. 图书情报工作, 2003, (9): 63.

6 张翠玲, 吴东红. 网络计量学赢法初探 [J]. 情报杂志, 2005, (1): 17-18.

7 李伟超. 链接分析法——情报学研究新方法 [J]. 情报资料工作, 2003, (3): 11.

8 Openheim. The Correlation between Citation Counts ant the 1992 Research Assessment Exercise Ratings for Birtish Library and Information Science University Departments [J]. Journal of Documentation, 1995, 51 (1): 18-27.

9 Thomas and Willett P. Webometrics Analysis of Departments of Librarianship and Information Science [J]. Journal of Information Science, 2000, 26 (6): 421-428.

10 Heting Chu, Shaoyi He, Mike Thelwall. Library and Information Science Schools in Canada and USA: a webometrics perspective [J]. Journal of Education for Library and Information Science, 2002, 43 (2): 110-125.

11 Sminth, A and Thelwall, M. Web Impact Factors and University Research Links [C]//Proceedings of 8th International Conference on Scientometrics & Informatics, Vol 2, Sydney Australia, 2001: 657-664.