

# 数字出版网络影响因子研究

刘英<sup>1)</sup> 曾莉<sup>2)</sup>

1) 华中科技大学附属协和医院杂志社《临床心血管病杂志》编辑部; 2) 华中师范大学管理学院, 430022, 武汉

**摘要** 网络影响因子是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, 可以用于对期刊网站进行评价。搜索引擎有 AltaVista、Fast、Google。通过开发专业搜索引擎来完善网络影响因子的计算。

**关键词** 科技学术期刊; 网站; 网络影响因子; 搜索

**Web impact factor in website of academic periodicals** // LIU Ying, ZENG Li

**Abstract** Web impact factor is the ratio from the page number linked to other website or region divided by all page number for a moment. It can be used on evaluation of journal website. Searching engines are AltaVista, Fast, Google. We improve the calculation of web impact factor through development of specialized searching engines.

**Key words** sci-tech academic periodical; website; web impact factor; searching

**First-author's address** Journal office of Union Hospital, Huazhong University of Science and Technology, 430022, Wuhan, China

网站是科技期刊数字出版的重要组成部分, 通过整合计算机技术、信息技术和网络技术来提供服务。它反映着网络社会中目标期刊的定位、特点及文化内涵。目前期刊数字出版网站主要可分为 2 类: 一类是大型期刊网站, 如中国期刊网、万方数据库数字化期刊、中文科技期刊全文数据库等, 集中了数千种学术性期刊; 另一类是独立的期刊网站, 数量庞大, 有上万种中、外文电子期刊。针对期刊网站的规划和设计, 需要一个客观、科学的评价指标<sup>[1]</sup>。

## 1 网络影响因子的提出及其作用

网络影响因子(web impact factor, WIF)是从文献计量学中的影响因子引申而来的。在文献计量学中, 影响因子就是在 2 年内某期刊上发表的所有文章被引用的总次数与这 2 年该期刊上所发表文章总数的比值。网络影响因子则把链接网页总数看作是文章的引用次数, 把网页总数看作是文章总数而形成的。

Ingwersen 于 1998 年提出的网络影响因子的定义如下: 网络影响因子可以表示为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。就像传统的影响因子可以

对期刊进行评价一样, 网络影响因子也可以用于对期刊网站进行评价。根据链接网页性质的不同, 链接网页数可以分为总链接数、外部链接数和内部链接数, 相应地网络影响因子也就据此分为总网络影响因子、外部网络影响因子和内部网络影响因子<sup>[2]</sup>。

网络影响因子作为一个纯粹的定量指标, 比定性指标更为客观, 可以直接纳入到网站评价的指标体系当中。网络影响因子在网站评价上的应用发展迅速, 许多类型的网站都被作为评价对象。大多数研究结果表明, 网络影响因子对评价网站质量和测定核心网站具有重要价值。网站的影响力能够在一定程度上反映期刊的影响力。

1999 年, 国外研究对 22 种电子期刊的网络影响因子进行了研究, 网络影响因子最高的期刊《PSYCHE》, 同时也是 ISI 中被引用次数最多的电子期刊。2003 年国内一项研究中, 对 42 种工程类中文期刊网站的网络影响因子进行测定, 并与中国科学技术信息研究所 2001 年底公布的 2000 年影响因子逐一对应比较, 发现传统影响因子与网络影响因子具有统计学意义的相关关系<sup>[3]</sup>。

## 2 网络影响因子的搜索

能用作网络影响因子测度的工具必须具备以下条件<sup>[4]</sup>。

1) 数据库的覆盖范围足够大, 这样才能保证结果的准确性。

2) 数据库经常更新, 因为网络几乎每天都有成千上万的新网页产生和旧网页消失。

3) 能搜索特定网站的全部网页。

4) 能搜索所有包含到某网站链接的网页。

5) 能够很好地支持布尔逻辑运算。

6) 性能稳定。

目前针对网络影响因子的搜索引擎很多, 比较著名的有 AltaVista、Fast、Google 等。AltaVista 和 Fast 能够提供多种类型的限制检索, 如主机名限制、超链接限制、域名限制、文件类型限制、新闻组限制、主题限制等。此外, AltaVista 还提供布尔逻辑检索、截词检索、字段限制检索、日期限制检索、范围限制检索、动态分类检索、指定语种检索、位置检索, 等等。Ingwersen、

Alastair G1Smith、Mike Thelwall 和 OwenThomas 所使用的工具都是 AltaVista 搜索引擎。

国内有些研究者还采用 Fast 搜索引擎对影响因子进行了分析。如杨涛利用 Fast 搜索引擎,对中国大陆 20 所大学网站的总网络影响因子、外部影响因子、总科研影响因子、科研网络影响因子和教育网影响因子进行了统计分析。虽然 Google 搜索引擎因其高级搜索功能有限而限制了它在网络影响因子方面的应用;但是,由于它具有强大的网络信息搜索能力,所以,对于计算总网络因子的研究来说,也是一种有力的工具<sup>[2]</sup>。

这些搜索引擎基本上都属于商业搜索引擎,都不是为了网络计量研究而设计的。为了克服搜索引擎的局限性,研究者们还采用了很多其他的办法。最典型的例子就是“学术网络爬虫”。借助这种专用工具,研究者可以更深入地分析一个网站,可以获得更准确的链接数据;但是,“学术网络爬虫”最大的局限性在于应用复杂,效率较低,只能检索较小网域内的链接,在较大网络范围内则无能为力。除“爬虫”之外,还有一些其他的工具出现在网络影响因子的研究当中。但从总体上看,这些工具有的流通范围很窄,有的还处于起步阶段,目前都只能起到有限的作用;因此,我们在相当长的时间内仍把这些商业搜索引擎作为主要的网络数据收集工具<sup>[3]</sup>。

### 3 发展趋势

从目前情况来看,最被看好的突破方向是开发“专业搜索引擎”。这种工具既能够有效地克服商业搜索引擎的种种局限性,又能够充分利用搜索引擎的技术优势,无疑是最理想的选择,因而得到许多搜索引擎厂商的重视。

例如,在搜索引擎中引入 XML 和元数据。XML 全称是 extensible markup language,即可扩展标记语言,它的最大优点是使文件的内容和格式分离,更便于揭示文件的内容。元数据是关于数据的数据,有很多种,此是其中比较常见的一种。它包含多个基本元素,如题名、主题、内容描述、关联、来源、语种、覆盖范围、创建者、其他创作者、出版者、权限、日期、格式、标志符和类型。用来对网络信息进行组织,可以使网络信息更加规范,便于对它们进行计量。特别是关联和来源都可以用在网络影响因子的测度中。引入 XML 和元数据,可以使网络影响因子的测度深入到内容层次中。

再如,引入引文分析中的引文类型和引文动机,区

分网页的权重。我们在网络影响因子的计量中,将所有的网页都看作是等价的;但是,从引文分析中可以知道,不同的引文类型和引文动机,其价值是不同的。在网络中也应该一样,要根据链接的不同类型和不同动机给出网页的不同权值,使网络影响因子的值更加准确<sup>[4]</sup>。

### 4 存在的问题

要进行网络影响因子的测度,首先需要解决的就是网页的处理问题,也就是在测度中什么样的网页才能算在内以及如何去计算。

我们知道,在网络中除了有格式的网页外,还有其他格式,特别是在学术论文中广泛采用的是 PDF 格式的网页。那么,在计算这些格式的网页时,是全部都算还是只选取其中的 1 种或几种? 还有,框架网页是比较常见的,但在计算时是将其算成 1 个还是多个网页? 在影响因子的计算中,人们一般只考虑论文的数量,不考虑论文的篇幅;但在网络影响因子的测度中,则必须考虑这方面的情况,因为一篇论文有时候用 1 个网页,有时候要用多个网页,在用多个网页时,是不是要将这些网页合并起来计算? 网络上有很多内容是重复的,同一内容可以在网络的多处出现。这种情况在影响因子的计量中是不会出现的,因为学术期刊都不允许一稿多投、一稿多发;但在网络中这种情况很常见,要不要对它进行特殊处理?<sup>[4]</sup>

### 5 结束语

网络影响因子是目前重要的期刊网站计量因子,能比较客观地评价网站的建设与维护。尽管还存在不足,但相信随着网络技术的日益进步和期刊数字化发展的逐渐深入,网络影响因子的计算一定会更加完善和客观。

### 6 参考文献

- [1] 顾东蕾. 期刊网站的评价指标体系[J]. 现代图书情报技术, 2005, 21(11): 68-71
- [2] 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
- [3] 张洋. 网络影响因子研究综述[J]. 中国图书馆学报, 2010, 36(1): 63-79
- [4] 杨涛, 曹文娟. 网络影响因子及其测度[J]. 图书情报工作, 2009, 12(9): 24-27

(2010-07-08 收稿; 2010-07-28 修回)