

基于链接分析的学术性核心网站评价

田红梅 李强

(解放军理工大学气象学院图书馆, 南京 211101)

摘要 学术性核心网站的评价对网站建设与发展起着非常重要的作用。文章从“核心网站”概念的提出,分析了确立“核心网站”的意义;根据链接分析方法研究的主要内容,探讨了对学术性核心网站评价的基本内容、原则和基于 WWW 链接分析的评价方法。

关键词 核心网站 网络信息资源 评价 链接分析

中图分类号 G203

文献标识码 A

文章编号 1007-7634 (2004) 09-1078-03

Link-based Quantitative Evaluation Method for Core Websites

Tian Hongmei Li Qiang

(Library of Meteorology College, PLA University of Science and Technology, Nanjing 211101)

Abstract The article bases on the idea of "core website" and the main content of link-based analysis, analyzes the significance of determining "core website", studies the content and the principle and the methods of core website evaluation.

Keywords Core website Network information resources Evaluation Link analysis

们借鉴核心期刊的相关理论提出了“核心网站”的概念。

1 核心网站的提出

“文献集中与离散规律”揭示了核心期刊的存在。布拉德福从期刊相关载文率出发,对独立学科的期刊等级排列,发现科学论文在相关期刊中并非均匀分布,而是呈现出明显的集中—分散现象,从而可以得到期刊的核心区及其相继区;加菲尔德利用期刊引用分析法,把核心期刊推广到多学科文献,他利用美国科技情报所的《科学引文索引》数据库进行的统计表明,所有引文量的75%来自不到1000种期刊,有70%的引文仅来自500种期刊,从而又从另一个角度证实了文献离散规律的存在,揭示了文献情报流的内在规律,反映了社会发展和文献活动的规律性。因特网作为一个开放性的全球分布式的信息网络,它在给人们带来无限信息空间的同时,也带来了巨大的信息污染,网上信息资源质量隐患在很大程度上影响了人们对因特网资源的充分利用,因此,如何从众多网上资源中筛选出符合需要的有用信息或知识,如何评价网络信息资源的质量等越来越受到人们的关注。由于网络信息资源与以书本为载体的文献资源存在着一定的共性,因此,越来越多的人开始探讨和尝试将文献计量学应用到网络信息资源的研究上,网上信息资源的分布同样遵循布拉德福定律,呈现集中—分散现象。为此,人

2 确立核心网站的意义

因特网是开放性的全球分布式网络,是一个极具价值的信息网络。但由于其开放性、分布性、自由性、共享性等特征,使得因特网信息来源极为广泛,信息发布带有很大的随意性,在它为用户提供浩如烟海信息的同时,也造成了信息过载和信息污染。面对源源不断涌来的信息,用户面临着如何加以选择和利用的问题。尽管用户通过网址(URL)中的顶级域名区分各类信息,却不能对信息的准确性、可靠性等方面作出评判。因此,网上信息的可靠性、有效性、相关性以及真实性便成为人们关注的重要方面,并由此引发了对网络质量的评估问题。

然而,对网络信息资源评估实质上是对网络信息资源的信息质量进行合理的网站评价标准和方法,越来越受到人们的关注。核心网站的确立,不仅是开展信息服务和满足用户网络信息需求的需要,同时也是扩大网站的知名度、促进网站自身建设和提高网络信息服务效率的需要。

(1) 有助于节省网络信息检索的时间,提高科研效率。核心网站集中了大量有价值的信息,使用户用最少的时间就可以获得最大的信息量,从而尽

快地完成科研任务。

(2) 有助于人们了解某学科领域的发展方向和最新研究成果。核心网站的信息不仅价值相对较高,而且能反映学术研究的最新成果和动态,便于研究人员了解和跟踪某一科研课题。

(3) 有利于促进网站建设。核心网站的形成同样是符合“成功导致成功”即“马太效应”的,也就是说由于核心网站的信息价值和可信度等都较其他网站高,所以它的访问者多、人气较旺,容易吸引稿源和相关作者加入讨论,形成“累积优势”(Cumulative advantage distribution),从而保持网站的良性循环,进一步促进网站的建设和发展。

“核心网站”的确立和相对认可,必将会扩大网站的知名度,增加网站的利用价值。

3 链接分析研究的主要内容

因特网的基本成分是网页,网页与网页之间由链关系相互连接,链是网页之间相互关联的纽带,它反映了网页之间的各种内在联系。通过对链的数量进行统计和分析,可以了解网页乃至网站的情况。

链接分析的内容主要包括:链接和被链接量,链接网页的类型,链接的频次和变化,链接网页之间的关系。

(1) 链接和被链接量。许多搜索引擎都能检索某一网页被链次数,可直接用于检索链接关系,用搜索引擎检索出的某一网站的链接数能反映该网站被利用的情况。

(2) 链接网页的类型。在“利用链接关系评价网络信息的可行性研究”一文中作者认为新浪网的站外链接有以下几种:推荐链接、合作链接、相关链接、资源链接、通讯链接和广告链接等,其中与主页及频道建立的链接价值更高。如推荐链接直接对被推荐网页质量做了肯定,在评价网络信息上是极为有效的。

(3) 链接的频次和变化。有价值的站点终究会满足人们的需求并被人们利用。一个网页被链接次数越多,从一个侧面证明该网页质量越高,可利用被链的次数和变化对该网页(网站)信息质量进行评价。

(4) 链接网页之间的关系。如果站点 A 与站点 B 相互引用较多,则它们之间的关系较紧密。网站的内容是以网页为单位的,同样,网页之间相互引用较多则说明它们之间的关系较紧密。

4 核心网站的评价

4.1 评价的基本内容

(1) 网站总体设计。主题明确、重点突出,页面设计美观、布局合理,所要展示的内容条理清晰、排列得当,各网页之间风格统一,有整体感。

(2) 网站信息质量。包括内容的权威性、准确性、新颖性、特色性、客观性等方面。这是评价网站的最重要、最基本的指标。

(3) 网站信息组织。主要包括信息内容组织得是否有序,是否针对使用对象的需要、便于检索,提供信息的广度和深度以及时间的连续性等。

(4) 网站信息结构。包括信息是否符合逻辑,链接是否合理并与内容相符、符合浏览规律,链接是否正确稳定。

(5) 网站易用程度。包括资源的利用是否方便有效、用户界面是否友好、检索功能是否齐全,页面的交互性、兼容性和连接响应速度等。这是使用网络信息所需要的最基本的客观条件。

4.2 评价的原则

(1) 用户第一原则。网站是一个单位或机构在网络世界中的代表。建立网站的目的,一是宣传自己,二是为用户服务。而后者更为重要,只有在为用户服务当中网站才能实现自身价值。因此用户第一的思想是否贯穿于网站设计的全过程,是我们评价网站时应坚持的一个原则。

(2) 整体性原则。评价一个网站应看其在网站建设整体设计构思上是否对所提供的信息内容(表现形式)进行全面规划,确立清晰可行、有持续发展能力的信息资源建设思路。

(3) 特色性原则。有特色才有生命力,特色的信息与服务是一个网站区别于其它网站的特征,是其能否吸引用户的关键,是其能否迅速发展的决定性因素。

(4) 动态发展原则。网站建设是一项不断改进、不断完善的工作,网站信息服务的质量与水平(内容和形式)只有不断完善与提高,才能保证网站有旺盛的生命力和竞争力,否则,其位置很快会被竞争者取代。因此,评价网站要看其是不断发展还是停滞不前。

4.3 评价的基本方法

确定或评价核心网站的方法主要有定性和定量评价方法。定性评价法包括用户评价法、专家评价法、评估工具评价法和指标体系评价法等,但我们这里主要讨论定量评价法,亦即基于链接分析的评价方法。

引文分析法是评价期刊质量的经典工具,是对文献进行定量评价最为有效的方法之一。它的主要依据和研究内容是科学文献之间的引证与被引证关系。互联网存在着一种与传统文献引证与被引证关系十分相似的情况,就是站点之间的链接与被链接的关系。Web 网站中的链接可被看作类似于印刷型出版物中的引文,可通过计算其相关的数量指标来计算网站的相对质量。与引文分析法的基本假设相比,一个网页被另一个网页链接的原因有两种:一是提供的资源和服务被利用,二是因为该网页质量高而被推荐。这两种原因都说明一个网页被链接与其质量存在着正向的关系,链接分析法符合引文分析法的基本假设。一个网站被链接次数越多,从一个侧面证明该网站质量越高,说明它的影响力就越大。如果对同一类站点按被其它站点链接次数从大到小的顺序排列,就可以看出各站点之间影响力的差异。如果取一个合适的值作为阈值,大于阈值的站点即为核心站点。

(1) 影响因子评价法。影响因子是一个国际上通行的期刊评价指标,是加菲尔德于1972年提出的。由于它是一个相对统计量,所以可以较公平地评价和处理各类期刊。1996年中国科学院文献情报中心应中国科协和中国科学院的要求,为48种期刊提供了影响因子指标,从中可以看出,影响因子的大小与期刊质量的高低存在着一定的内在联系。借鉴期刊影响因子分析的思路与方法,1998年,Ingwersen 提出可以把文献计量学的期刊影响因子应用到网页的评价中去。网页的影响因子(Web impact factor, WIF)是指某一类型的域名或网页被链接之和与有关域名或网页之和的比例。网页影响因子可以用来分析在一定的时期内相对关注的网页情况。通常,影响因子越大,网站的学术影响力就越大,质量就越高。因此,影响因子从信息的利用角度,体现了信息资源在信息交流中被人们所重视的程度。

(2) 搜索引擎评价法。因为搜索引擎具有能够从所搜索的网页及其链接的大型数据库中识别网页

数量、链接数量、统计网页和链接结果的特点,所以它可有效地用于网络计量研究。许多搜索引擎 Alta Vista、Fast Search、Hotbot、Excite、Lycos 等都能检索某一网页的被链次数。如在 Fast Search 中输入检索式 sina.com 可检索到 Fast Search 收录的网页中有多少页面是"sina.com",也就是 Sina 站点“被引”了多少次。这样就可以比较各站点或页面的“被引”次数,进而了解站点或页面的影响力。一般情况下,“被引”次数越多,站点或页面信息水平和可信度越高,说明站点或页面的影响力就越大,从而对网站的重要程度作出评价。

5 核心网站评价中应注意的几个问题

(1) 核心网站是一个动态的概念,因此对它的评价并不是一成不变的,所以应经常关注网站的信息在内容质量方面的变化,以作出适时调整。

(2) 由于不同类型的网站提供的内容和服务是不同的,因此被链接的机会也不相同,在利用链接次数评价网站时,应选择好比较对象,即同类型网站相比,使其具有可比性。

(3) 在互联网上,作为链接分析法研究对象的网站,其网页“被引”所产生的动机远比印刷型文献中的引文复杂,除引证目的外,还包括随意性和不相关的链接等。因此,研究人员指出,在将引文分析方法完全应用于网络信息资源评价之前,还必须深入了解网络链接产生的机理、链接行为模式和合适的链接统计工具。

参考文献

- 1 杨海平,陆宝益,张天俊. 核心网站论, 图书情报工作, 2001 (12): 68~70
- 2 刘雁书,方平. 利用链接关系评价网络信息的可行性研究. 情报学报, 2002 (4): 401~406
- 3 黄奇,李伟. 基于链接分析的学术性 WWW 网络资源评价与分类方法. 情报学报, 2001 (2): 186~192
- 4 刘君玉. 引文分析的评价. 情报理论与实践, 1998 (2): 87~89
- 5 蒋颖. 因特网学术资源评价: 标准和方法. 图书情报工作, 1998 (11): 27~31
- 6 邱均平,黄晓斌. WWW 网页的链接分析及其意义. 中国图书馆学报, 2221 (6): 44~47
- 7 李长忠. 网络计量学理论与实证研究. 图书情报工作, 2001 (10): 21~23, 32
- 8 张咏. 网络信息资源评价方法. 图书情报工作, 2002 (10): 41~47, 61