

# 网络关键信息筛选与影响力评估研究

曹锦梅, 李韶山, 孟祥娟

(新疆医科大学 高职学院, 新疆 乌鲁木齐 830054)

**摘要:** 网络时代提高信息服务质量的关键是信息筛选。网络信息影响力评价方法主要有搜索引擎收录情况、用户访问量统计和链接分析法, 这些方法通过定量评价网络信息, 有效地弥补了同行评议法等定性评价方法的不足。结合这些方法为用户检索关键有效信息提供了一些思路。

**关键词:** 网络; 关键信息; 筛选; 影响力; 评估

中图分类号: TP393.07

文献标识码: A

## Research on critical information network and the assessment of its influence

CAO Jin Mei, LI Shao Shan, MENG Xiang Juan

(Vocational Institute, Xinjiang Medical University, Urumqi 830054, China)

**Abstract:** The key to improve the quality of in the network information service is information filtering. The influence of network information evaluation methods include major search engines situation, user traffic statistics and link analysis, which are through quantitative evaluation of these methods through the network information, making up for the peer review and other methods of qualitative assessment effectively. The paper provides some ideas for the effective retrieval of key information users by combining these methods.

**Key words:** network; key information; screening; influence; assessment

如何从 Internet 资源库中迅速准确地获取所需信息成为目前一个亟待解决的问题。搜索引擎成为目前人们获取信息的重要途径。面对着大量的信息, 如何进行有效处理? 如何利用网络筛选获得自己所需要的信息? 如何才能有效地利用这些信息? 这些问题是能否高效利用现代化技术的至关重要的环节。然而, 即使像 Google 和百度这样技术先进的通用搜索引擎巨头仍然面临一些棘手问题尚未解决<sup>[1]</sup>。

(1) 低查准率: 查准率和查全率一直难以两者兼得, 通用搜索引擎往往以牺牲查准率来获得较高的查全率, 而这种做法对有特定信息需求的人群越来越显得无能为力。

(2) 搜索的“垂直度”问题: 如何针对专业领域的行业需求, 更精确地、深入地挖掘和获取用户所需信息既是一个难度很高也是一个亟待解决的现实问题。

### 1 网络信息检索的发展及其特点

网络信息检索系统的核心是搜索引擎, 搜索引擎的发展史就是网络信息检索的发展史。网络信息检索呈现

出以下几方面的特点:

#### (1) 多用户

网络时代的信息资源服务从一开始就是多用户的, 在此模式下, 信息检索的服务必须注重快速反应, 注重对并发访问的支持、对公共数据的共享、对临时工作数据的清理等。

#### (2) 海量数据

在网络环境下, 信息源众多、数据量巨大。大数据量的后果就是垃圾数据增多, 而寻找 1 种甄别数据有效性的高效信息检索方法就显得尤为重要。

#### (3) 非专业化

网络环境下使用信息检索系统的用户大多都不具备专业的信息检索技能, 这些非专业用户不能像图书情报专家那样熟练地使用主题词、分类号、逻辑组合和严格的查询语言来检索网上的信息内容。

### 2 筛选

提高信息质量的关键是由于网络信息资源数量繁多, 包罗万象, 所以对其应当有选择、有甄别地利用。网络信息

资源的选择应遵循以下原则:

(1)针对性原则。有针对地编制文摘、综述、题录等,通过各种交流途径进行传递报道,及时提供给用户。

(2)科学性原则。信息资源的选择需要采用科学的方法,高质量的信息资源意味着它已经过质量的控制,已经得到组织加工并以1种系统而统一的方式存储,在1个可靠的、用户友好的检索系统中提供对它们的存取。

(3)计划性原则。对网络信息的选择要统筹计划、归类整合、制定目标、合理管理。

(4)对特定用户进行跟踪服务、主动及时地提供新颖、准确、有价值、实用的专业化网络信息资源。

### 2.1 筛选出真实信息,增强准确性

网络上收集到的信息大都来源广泛、涉及面复杂,往往混杂着一些不真实、不准确、虚假的、夸张的成份,对这一部分信息如不加以鉴别,就必然导致信息的不准确。所以在筛选信息时,要努力做到两点:首先搞好鉴别,判断其真实和准确程度;其次是追根溯源,一定要筛选出来重新核实,最大限度地剔除其不确定因素,确保所选的资料科学、周密、准确。

### 2.2 筛选出核心信息,增强服务性

在整理筛选核心资料时,一要找准位置,二要把握需求,根据要求,有目的地收集和选择信息资料,并加以全面系统地归纳、综合和概括;三要抓住中心,不可孤立地看待信息的“好坏”,而必须以实际需求为准绳,才能把握信息的质量。

### 2.3 筛选出典型资料,增强代表性

在文献资料库中,大部分信息是分散的、零星的、不系统的,反映的只是表层与局部,缺乏代表性。而要揭示事物本质及其发展变化规律,就必须选择具有代表性的典型材料。把收集到的大量信息加以浓缩,这样的信息才具有代表性。

### 2.4 筛选出简洁资料,增强实用性

收集的信息资料有许多是重复多余的,只有剔除了这部分信息才便于加工整理、利用和存储。在整理此信息资料时,一要抓住根本点;二要把握好尺度,有些信息资料虽然在内容上有一定重复,但也反映了事物的一些特点,筛选时可以剔除重复部分,综合整理出有特点的部分。

## 3 网络信息影响力评价方法

网络信息的急剧增长,需要对其进行评价以帮助人们更好获取与利用信息。目前对网络信息的评价研究主要集中在内容评价方面,主要评价标准有可信度、准确度、客观性、及时性、报道范围、链接、文章的目的性及高效性等。

内容评价主要采取同行专家评审法,此方法主观性大,耗时长,专家的代表性与公正性难以得到有效保证。

而网络信息内容是否有价值,最终都会在其对用户行为的影响上表现出来,质量较高的网络信息会导致更多的浏览次数,随之而来的是更多的被链次数,但可以通过一些具体的指标进行测量,定量评价网络信息,有效地弥补同行评议法等定性评价方法的不足。利用影响力对网络信息进行评价是目前一个热点课题。

目前,国内学者认为网络信息评价主要有3种方法。第3方评价法、网络信息服务用户评价法、由文献计量学引申和发展而来的网络计量法。这3种方法的评价过程,实际上就是信息评价从定性方法到定量方法的过程。定性方法主要依靠评价者的主观判断,用于评价网络新体系时难免会影响结果的客观性,降低结果的可信度,同时也会使评价过程花费时间较多。定量方法为人们提供系统客观、规模宏大的数量分析方法,结果更加直观、具体。由于目前网络发展变化快、稳定性差,通过定量方法得到的结果只能代表某一阶段的状况。所以,评价网络信息资源时仍不能偏废定性方法的使用,最佳方案是把定性与定量方法结合起来,这也是未来网络资源评价的一个趋向。

网络信息影响力评价实际涉及2个层面,一是单个网影响力评价,二是网站影响力评价。因为网络信息是以网站为单位发布的,网络信息影响评价多是针对网站进行的。网站影响力评价的标准与方法关系极为密切,通常是一种评价指标就是由一种具体的方法来测定的,因此,就将评价标准与方法结合在一起,对网站影响力评价的方法主要有以下几种。

### 3.1 搜索引擎收录情况

搜索引擎是网民在互联网中获取所需信息的基础应用,从1998年起,中国互联网络信息中心(CNNIC)于每年1月和7月发布《中国互联网络发展状况统计报告》。第23次调查统计数据截止日期为2008年12月31日,该次调查表明,目前搜索引擎的使用率为68.0%,在各互联网应用中位列第4。2008年全年搜索引擎用户增长了5100万人,年增长率达到33.6%<sup>[2]</sup>。印刷型文献影响评价有1个重要指标就是被著名检索工具(如Medline、BA、CA、Ei等)收录情况,被其它资源收录情况也是网络信息影响力的1个重要指标。不被搜索引擎收录的网站被访问、被利用的机会是很小的。

所以根据搜索引擎收录情况可以评价网站的影响力,另外被链次数、被点击次数、被评议过的站点也能促进网站的排名。1个站点被链接的次数越多,证明该站点影响越大,质量也越高。

### 3.2 用户访问量统计

用户访问量是对网络信息进行评价的1个客观指标,访问人数多是网络信息价值高、可用性好的结果。目前,对用户访问量进行统计主要有2种方法:用户行为

监测法与服务器登录日志(Server Access Log)分析法。用户行为监测法的代表是 Media Metrix 的网站及数字媒体访问量排名。Media Metrix 选择有代表性的家庭及工作单位上网用户进行调查,在被调查用户计算机的操作系统中安装“Media Metrix Meter”监测装置,对用户浏览网页的每一次行为都进行统计,以此为基础公布网站及数字媒体的访问量排名。服务器登录日志分析法是对服务器的访问日志进行分析,所有访问者对某一网站的访问情况都被记录在网站服务器的“登录日志”中,登录日志显示了访问者在网站上的全部访问行为。

### 3.3 链接分析法

链接分析法是利用某一站点的被链次数对其质量进行评价。被链次数是网络信息评价重要指标。1个站点被链接的次数越多,证明该站点影响越大,质量也越高。某一站点被链次数是搜索引擎搜索结果排序的一项重要指标。

网络关键信息的筛选与影响力评估所带来的益处是显而易见的,必将对用户的实际需求与科学研究等方面的发展起到积极的作用。

#### 参考文献

- [1] 赫建营,晏海华,金茂忠,等.结合本体筛选和文本挖掘的垂直搜索引擎研究[J].计算机科学,2008,35(2):188-190.
- [2] 中国互联网络信息中心.中国互联网络发展状况统计报告:第23次.http://www.cnnic.net.cn,2009.1.
- [3] 杨海涛.网络信息检索的困境及发展趋势[J].科技情报开发与经济,2008,18(1):13-14.
- [4] 刘雁书,方平.网络信息影响力评价方法[J].高校图书馆工作,2002,22(88):16-19.
- [5] 韦彤.论高校图书馆情报咨询部门的新职能-网络信息资源搜索评估与咨询[J].高校图书馆工作,2002,22(87).

(收稿日期:2009-06-09)