

## Web 引文的可查证性分析

安美荣 姜仁珍 张翠英

(山西大学信息管理系, 太原 030006)

**摘要** 本文以图书情报学科 6 种有代表性的核心期刊为样本, 从中随机抽取了 300 篇 web 引文, 从可获得性、语种、来源网站三个不同的角度研究了其可查证性。此外, 还简要介绍了国外在这方面所作的研究。

**关键词** web 引文 可查证性 可获得性 语种 来源网站

## Web Citation's Verification Analysis

An Meirong, Jiang Renzhen and Zhang Cuiying

(Information Management Department of Shanxi University, Taiyuan 030006)

**Abstract** Based on the six representative samples in the library science and information science, the paper draws 300 web citations randomly and explores their availability from three different aspects—availability, language, original web site. What's more, the author introduces the foreign development in this research.

**Keywords** web citation, verification, availability, language, original web site.

### 1 引 言

随着网络技术的兴起, 互联网的广泛应用和普及, 人类信息资源由早先的实物化、纸质化阶段, 经过电子化、数字化阶段的发展, 开始进入网络化阶段。网络文献在期刊作为引文的现象越来越引起人们的注意。引文可以反映学术的继承性和作者的立论根据, 便于他人鉴别成果的学术价值, 同时它也为后续的相关研究提供线索, 以利于学术的积累和发展。那么作为新生现象的网络引文(以下简称 web 引文)呢? 它是否也可以查证到, 从而为后续研究提供借鉴呢?

国外在这方面的研究比我国要早得多。1999 年, Koehler 随机地抽取了 361 个网页和 343 个网址, 研究他们的长期性和连续性, 以确定他们在内容上

和可获得性方面随时间的变化情况。随后, Koehler 又于 2002 年对以前相同的网页和网址进行了跟踪分析。Bing Tan, Schubert Foo 和 Siu Cheung Hui 三人在一项研究中发现, 在他们研究期间, 他们所跟踪的网页有 44.8% 已经改变, 有 3.8% 已经消失。同时, 他们还发现, 域名为教育类、商业类和娱乐类的网页比别的域名的网页改变的可能性要小, 而数据库是最稳定的网页。John Markwell 和 David W. Book 在 14 个月的时间内, 对科学教育类网页的持续性跟踪研究表明, 有 46.5% 的网页已经改变或消失。此外, 他们还发现, “gov”类网页是最易改变的, 紧跟着的是“edu”、“com”、“org”。Stephen P. Harter 和 Hak Joon Kim 研究了发表在学术电子期刊上的 279 篇文章的 34317 篇引文, 发现有 1.9% 引用了网络文献。Philip M. Davis 和 Suzanne A. Cohen 在对学生的毕业论文的引文研究表明, 在 1996 年 ~ 1999 年期间, 学生们引

收稿日期: 2004 年 1 月 12 日

作者简介: 安美荣, 女, 1975 年生, 山西大学信息管理系 2002 级在读硕士研究生。姜仁珍, 女, 1971 年生, 山西大学信息管理系 2002 级在读硕士研究生, 山东莱阳农学院馆员。张翠英, 女, 1949 年生, 山西大学信息管理系硕士研究生导师, 教授。

用网络文献的比例增长了 12%。但同时,他们也发现,1999 年,不可查证的 web 引文比例为 16%,1996 年此比例为 53%。

由以上对国外研究的回顾来看,国外的学者已从不同角度、不同侧面对 web 引文的可查证性进行了深入、细致的探讨。当然,这可以为我国的学者提供一定的借鉴作用。参照国外的分析研究,笔者仍选用图书馆学、情报学有代表性的 6 种核心期刊《中国图书馆学报》、《大学图书馆学报》、《情报理论与实践》、《情报资料工作》、《情报学报》、《图书情报工作》为研究样本,从中随机抽取了 300 篇(每种期刊取 50 篇)web 引文,然后逐个登录,研究其可查证性。

2 web 引文的可查证性

2.1 web 引文的可获得性分析

为了比较科学地研究 web 引文的可获得性,笔者从 6 种期刊中随机抽取了 300 篇 web 引文,并把这些 web 引文按其所提供的网址分为两大类,一类是提供了网址及题目、作者的,另一类是只提供了网址的。这两类 web 引文的篇数及所占的比例见表 1。

表 1 两类 web 引文的可获得结果表

类 型	篇数	所占比例 %
提供了网址及题目、作者的 web 引文	237	79
只提供网址的 web 引文	63	21
共 计	300	100

在所查证到的 300 篇 web 引文中,最终可以找到的有 184 篇,所占的比例为 61.3%;找不到的有 102 篇,所占的比例为 34%;不能确定的有 14 篇,所占的比例为 4.7%。其中,可以找到就是指:(1)通过所提供的网址和其他信息能找到相关文献;(2)若通过所提供的网址及其他信息找不到相关文献,如出现“找不到该页”、“无法显示该页”等信息,笔者又利用搜索引擎(考虑到 google 强大的搜索功能及其较好的专指度,笔者选用 google 作为搜索工具),或进入其主页进行查找;(3)用这两种方法都不能找到的,一个星期后再重复上述方法检索。用上述三种方法的任一种能找到相关文献,则都算找到。找不到就是指用上述三种方法的任一种都不能找到相关文献。不能确定是指输入网址后,只进入其主页或

出现作者不认识的外文,因此无法确定是否是相关文献。结果见表 2。Mary F. Casserly 和 James E. Bird 在其研究中表明,在他们所统计的 500 篇 web 引文中,有 344 篇可以找到,所占比例为 68.8%,146 篇找不到,所占比例为 29.2%,有 10 篇不能确定。我们的研究结果和国外学者的研究结果基本接近。

表 2 两类 web 引文的可获得结果表

类型	篇数	所占比例 %
找到	184	61.3
找不到	102	34
不能确定	14	4.7
共计	300	100

表 3 提供网址及题目、作者的 web 引文的可获得结果表

结果类型	篇数	所占比例 %
找不到或无法显示	62	26.2
输入网址找不到,通过其他途径找到	114	48.1
输入网址直接找到	61	25.7
共计	237	100

表 4 只提供网址的 web 引文的可获得结果表

结果类型	篇数	所占比例 %
找不到或无法显示	40	63.5
输入网址找到	9	14.3
不能确定	14	22.2
共计	63	100

笔者又仔细研究了两类 web 引文的查证结果,发现第一类 web 引文比第二类 web 引文的可查证比例要高得多。在前者中,237 篇有 175 篇可以查找到,所占比例达到 73.8%;而后者中,63 篇只有 9 篇可以查找到,所占比例只有 14.3%,查找不到的网络文献,由于无法确定其主题,因此无法再用其他方式进行查找。

从以上的分析中可以看出:(1)web 引文可查证到的比例不高。输入网址直接能查找到的只有 70 篇,所占比例才为 23.3%。究其原因,笔者认为主要是由于网站更新的频率快,有的网站又没有历史档案存放这些网页,因此无法查找到过期的网页。(2)两类 web 引文相比较,第一类 web 引文尚可按其

所提供的题目通过搜索引擎或其他途径来查找,第二类 web 引文的可查证比例更低。这就牵扯到 web 引文的著录问题。笔者查阅了所研究的 6 种期刊的文后参考文献著录须知,皆没有详细的或明确的规定电子资源或网络资源的著录格式。没有一个规范的、统一的格式,这就给著者在著录网络文献时带来一定的随意性。进而会影响到为后续相关研究提供的线索的功能。笔者认为,在著录网络文献时,应包括以下几部分:完整的文献信息(即题目、作者、网址)、著者引用日期、被引用网页的创建日期和上一次更新日期。目前,这个问题的解决仍无定论,还需广大学者的进一步探讨。

## 2.2 web 引文的语种

在所查证到的 184 篇 web 引文加上 4 篇其他语种的 web 引文,共计 188 篇 web 引文中,为中文的网络文献有 52 篇,所占比例为 27.7%;为其他语种的网路文献为 4 篇,所占比例为 2.1%;为英文的网络文献有 132 篇,所占比例高达 70.2%。

表 5 web 引文的语种分布表

种类	篇数	所占比例 %
web 引文为中文	52	27.7
web 引文为英文	132	70.2
web 引文为其他语种	4	2.1
共计	188	100

这从另一个侧面反映出一种可观的趋势,即学者们在进行学术研究中,更多地参考、利用外文文献。大家都知道,由于科技水平和经济水平的差异,国外的学者在专业理论及技术方面所取得的成就较国内要超前,因此相比较而言,可借鉴的地方也要多。我国的学者们能融贯中西,对图书情报学科的发展自然是有很大的推动作用。另外,也由于学者们平时接触到的外文专业期刊不多或很少,因特网强大的搜索、检索功能给学者们提供了很大的便利,他们可以自如地使用因特网上丰富的网络资源。

## 2.3 web 引文来源网站的稳定性分析

我们都知道,网站是网络信息资源的重要来源,也中唯一来源。研究网站的稳定性直接关系到 web 引文的可查证性。按其提供的网址的域名类型,我们将 web 引文来源网站分为 7 类。见表 6。需要特

别强调的是,这里的链接到是指输入网址能直接找到相关文献;链接不到就是指输入网址不能直接找到相关文献。和前面提到的找到、找不到是两个不同的概念。

表 6 web 引文来源网站的稳定性分布表

类型	链接到	所占比例 %	链接不到	所占比例 %	总计	所占比例 %
org	23	34.8	43	65.2	66	100
net	4	26.7	11	73.3	15	100
ac	5	23.8	16	76.2	21	100
com	15	18.5	66	81.5	81	100
edu	9	16.7	45	83.3	54	100
gov	2	14.3	12	85.7	14	100
other	12	24.5	37	75.5	49	100
总计	70		230		300	

从表 6 中发现,在域名为 org 的网站能直接链接到相关网络文献的比例最高,达到 34.8%,相对而言,可以认为是比较稳定的。域名为 net 的网站随后。最不稳定的是域名为 gov 的网站。而 Mary F. Casserly 和 James E. Bird 在其研究中表明,最稳定的是域名为 edu 和 org 的网站,两个网站能链接到相关文献的比例达到 64.5% 和 64.2%。我们所得出的结果和国外学者的研究结果不相一致。考虑到所统计数据的局限性,这个问题仍需对此有兴趣的学者作进一步的论证。

## 3 结束语

自从文献计量学在图书馆学、情报学开始应用以来,学者们就有了一种从定性转变为定量的研究方法。采取定性分析方法,仅仅依靠个人的思维判断,带有一定的主观性。采用定量的分析方法,可以揭示事物发展的数量规律,更具有客观性和说服力。而引文分析又是文献计量学中应用最广泛的一种定量分析工具。传统的引文分析都是研究书目和期刊的引用和被引用,因特网的飞速发展从另一个空间给学者们提供了一个观察事物、分析事物的角度,本文从另一个角度研究了 web 引文。

经过上述的分析,笔者提出几点建议与大家共享:(1)在评价网站的好坏时,不应只从其提供的内

容的质量来评价,还应考虑其所提供的内容的持续性或长期性;(2)web 引文的著录问题。各期刊编辑部和出版商等应尽快规定统一的网络资源的著录格式,并要求作者严格履行此格式。从而可以更好地为后续的相关研究提供借鉴。

### 参 考 文 献

- 1 孙永发主编.情报学进展:2002—2003 年度评论.第五卷.国防工业出版社
- 2 何朝晖.网络参考文献研究.图书与情报,2002(4)
- 3 Mary F. Casserly and James E. Bird. Web Citation Availability:

Analysis and Implications for Scholarship. College and Research Libraries, 2003(4)

- 4 Susan Davis Herring. Use of Electronic Resources in Scholarly Electronic Journals: A Citation Analysis. College & Research Libraries, 2002(4): 334 ~ 340
- 5 Stephen P. Harter, Hak Joon Kim. The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis. The Public-Access Computer Systems Review 7, 1996(5). Available online from <http://info.lib.uh.edu/pr/v7/n5.html> 7n5.html

(责任编辑 芮国章)

## 第三届“信息化与信息资源管理”学术研讨会 征文启事

为了加强信息资源管理及信息化领域的理论与应用研究,促进科技创新,推动我国信息业的快速和健康发展,武汉大学信息资源研究中心将主办第三届“信息化与信息资源管理”学术研讨会。本次研讨会将围绕教育部批准立项的第三批人文社会科学重点研究基地重大项目来进行,其主题为:文献计量与内容分析的综合研究、中国企业信息竞争能力的研究与评估。热忱欢迎广大理论研究人员和实际工作者撰文与会。现将会议征文内容及有关事项公告如下:

一、征文内容:1. 国内外文献计量与内容分析的新进展和发展趋势;2. 文献计量与内容分析的综合原理和基础理论研究;3. 文献计量与内容分析的综合方法研究;4. 文献计量与内容分析的综合应用研究及案例分析;5. 文献计量法与内容分析法的比较研究;6. 计算机辅助内容分析法的研究与应用;7. 内容分析软件工具的开发与应用;8. 知识网络与内容分析的研究;9. 基于知识元的内容分析法的研究;10. 内容分析与科学评价的关系研究;11. 企业信息化的体系结构研究;12. 企业信息化与企业竞争能力的关系研究;13. 中国企业信息化竞争能力的理论与应用模型;14. 企业信息竞争能力评价体系的研究;15. 企业信息竞争能力的实施方法研究;16. 国内外企业信息竞争

能力的实证分析与比较分析;17. 与会议主题有关的其他问题研究。

二、征文要求与截止日期:1. 论文须主题突出并未公开发表过,字数为 5000 ~ 10000 字左右,含 300 字以内的中、英文摘要。2. 论文按如下顺序排列:论文标题、作者、单位及邮编、摘要、关键词(其后列出以上各项对应的英文)、正文(小标题以数字编号,如 1.1, 1.1.1……)和参考文献。具体格式可参考《图书情报知识》论文。文尾请附作者详细通信地址、联系电话、传真、E-mail 地址。3. 论文要求以中文 Word97 或 Word2000 格式存入软盘,并提供便于通信、编辑和出版的电子版本。4. 论文请寄:武汉大学信息资源研究中心查荔收(请在文稿左上角标注“研讨会论文”)。邮编:430072;电话:(027) 87665541;传真:(027) 87665541;电子邮箱:irc@whu.edu.cn。5. 征文截止日期:2004 年 9 月 15 日。录用通知:2004 年 10 月 15 日前寄出。本次研讨会录用的论文将在会前正式出版论文集。

三、研讨会时间与地点:时间为 2004 年 11 月,地点在武汉大学。

(邱均平)