网络链接分析论文的计量研究

邱均平，矫翠翠

（武汉大学 中国科学评价研究中心，湖北 武汉 430072）

摘 要： 本文利用 Web of Science 数据库检索了以网络链接分析为主题的论文，从载文量、著者、关键词、地区分布以及引文等角度对检索到的论文进行定量分析，探讨世界范围内的网络链接分析的发展历史和研究现状，总结其研究热点并预测未来走向。

关键词： 网络链接分析；引文分析；h 指数；被引数

中图分类号：G203　　**文献标识码：**A　　**文章编号：**1007－7634(2008)08－1130－05

Quantitative Research of Papers about Web Link Analysis

QIU Jun－ping, *JIAO Cui－cui*

(*Research Center of Chinese Science Evaluation*, *Wuhan University*, *Wuhan* 430072, *China*)

Abstract: We got some papers about web link analysis from Web of Science data base, then did some quantitative analysis of these papers from the aspects of paper quantity, key words, district distributing and citation. This paper discusses the history and current conditions of web link analysis, and summarizes the research hotspots and forecasts the trend of web link analysis.

Key words: web link analysis; citation analysis; hindex; citation frequency

所谓网络链接分析方法，就是运用网络数据库、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[1]。对网络链接分析的研究最初来自于传统的引文分析法，可以说，网络链接分析的产生以及发展都无不与传统的引文分析法密切相关。

引文分析是上个世纪 20 年代出现的，1927 年 P.L.K.Cross 等人进行了文献史上第一次引文分析，他们统计了化学专业的某些期刊论文的参考文献并进行了分析，得出了化学教育方面的核心期刊。而所谓引文分析（Citation Analysis），就是利用各种数学及统计学的方法和比较、归纳、抽象、概括等逻辑方法，对科学期刊、论文、著者等各种分析对象的引用与被引用现象进行分析，以便揭示其数量特征和内在规律的一种文献计量分析方法[2]。

网络是基于链接的基本结构，网页之间的链接相应地看作文献之间的引用，这个崭新的思路一下子引领超链分析、链接分析成为了学术界以及工业界的研究热点，自 Almind TC 于 1997 年提出网络信息计量学（Webometrics）概念后，网络信息计量作为一门新兴学科异军突起，而网络链接作为评判网络资源的一个有效标准，被国内外学者重视并进行探索性的研究；与此同时，在工业界，Google，百

收稿日期： 2008－01－07

作者简介： 邱均平（1947－），男，湖南涟源人，教授、博士生导师，从事信息管理与科学评价、信息计量与科学计量研究；矫翠翠（1985－），女，山东烟台人，硕士研究生，从事信息计量与科学计量研究.

度等搜索引擎纷纷申请专利，开始利用链接分析的方法提高搜索效果。时至今日，世界范围内的链接分析有哪些发展历史和研究现状呢，为了探索这个问题，本文从载文量、著者、关键词、地区以及引文等角度分析了 Web of Science 中网络链接分析相关论文，希望能够对网络链接分析的研究有所帮助。

1 论文数量分析

论文在检索时段上选取 1999 年到 2007 年 10 月 4 日间的一段数据，选择美国 ISI 公司的权威数据库 Web of Science 作为数据来源，因为它可用于查找最新的研究成果（文摘和所引用的参考文献）。经过反复试验和比较，选用较理想的检索式 TS = (web link) OR TS = "link analysis" 进行文献检索，设定 Document Types 为 "Article"，设定 Subject Categories 为包含 "computer science" 及 "information science& library science"，共检索出 554 篇相关论文，去除其中不相关的 83 篇论文，最后得出 471 篇相关论文。由于检索出来的 554 篇论文不都是相关的，为了结果的准确性，笔者并没有运用该数据库自带分析工具，而是人工对这 471 篇论文的年载文量、著者、关键词、论文被引情况等进行统计，分析网络链接分析的研究现状。

某领域的文献数量在一定程度上反映了一门学科的研究水平和发展程度。表 1 反映了 1999 - 2007 年 Web of Science 中关于网络链接分析论文的年发文量。

表 1 1999 - 2007 网络链接分析论文的年发文量

年份	1999	2000	2001	2002	2003	2004	2005	2006	2007	总数
篇数	28	23	28	42	61	86	90	86	27	471

从表 1 中可以看出，1999 - 2001 年论文量基本维持稳定状态，这段时间是网络链接分析开始的时期，处于事物生命周期的初期；2001 - 2004 年是论文量快速增长的时期，处于生命周期的发展期；2004 - 2006 年论文量有少许波动，但基本处于平稳中略有上升的状态，这基本符合事物产生发展的生命曲线。20 世纪 90 年代互联网迅速发展，网络上的信息量以惊人的速度增长，不计其数的网络链接将这个网络世界联系在一起，将这些信息关联在一起，这就促使了网络链接分析的诞生。可以说网络链接分析是随着互联网的诞生而诞生的，表中的数据即反映了网络链接分析的诞生以及发展的这个生命曲线。

2 论文著者分析

在检索到的 471 篇论文中，作者共有 1008 位（包括合作者），其发文情况见表 2。

表 2 作者发文情况

篇 数	1	2	3	4	5	6	7	8	30
人 数	859	94	30	12	4	3	2	1	1
比 例	85%	9.3%	3.0%	1.2%	0.4%	0.3%	0.2%	0.1%	0.1%

从作者人数与所著论文数之间的关系来看，471 篇论文的 1008 位作者中，发表 1 篇论文的作者有 859 位，发表 2 篇论文的作者有 94 位，发表 3 篇论文的作者有 30 位。发表 1 篇论文的作者数量约占所有作者数量的 85%，这与洛特卡定律所描述的发表 N 篇论文的作者数量约为发表 1 篇论文的作者数量的 1/n2 以及发表 1 篇论文的作者数量约占作者总数的 60%并不吻合，表明网络链接分析的研究尚未成熟，写一篇论文的作者群体过大，在未来的发展中发表多篇文章的作者将会增加。

由表 2 可知，Web of Science 中关于网络链接分析的论文作者中 85%的作者只发表了一篇文章，9.3%的作者发表了 2 篇文章，发表 3 篇的作者有 3.0%，发文多于 3 篇的作者比例逐渐降低，发表论文最多的作者发文量达 30 篇，说明存在这样一些核心作者。同时我们看出，研究网络链接分析的专家发表的论文还是有限的，这与网络链接分析的发展还没有走向成熟有关。这里列举了发文数量在 6 篇以上的（包括 6 篇）的作者情况，见表 3。

赫希认为 h 指数能够比较准确地反映一个人的学术成就。一个人的 h 指数是指他至多有 h 篇论文分别被引用了至少 h 次。一个人的 h 指数越高，则表明他的论文影响力越大[3]。由表 3 知 Wilkinson, D 及 Ma, WY 研究主题有多个，并非仅是本领域的专家，因此 h 指数特别的高[3]。以网络链接分析为主要研究领域的核心作者主要有 Thelwall, M; Vaughan, L; Harries, G; Lempel, R; Kitsuregawa, M 等。这些核心作者研究的主题大部分是信息科学与图书馆科

学、计算机科学、信息系统，可见网络链接分析与这几个领域密切相关。网络链接分析是在计算机科学的基础之上发展起来的，在信息科学与图书馆科学，信息系统领域得到广泛关注和应用。计算机科学已经是一门比较成熟的学科，而网络链接分析只是刚刚起步，但是有这样一个成熟的基础，加上已经很成熟的引文分析等方法的借鉴以及互联网技术的发展，网络链接分析一定会加速发展。

表3 1999－2007年网络链接分析论文核心作者统计表

排 名	作者	论文量	h指数	研究主题
1	Thelwall, M	30	13	信息科学与图书馆科学
2	Vaughan, L	8	11	信息科学与图书馆科学
3	Wilkinson, D	7	21	信息科学与图书馆科学、药学
4	Lempel, R	7	2	计算机科学、信息系统
5	Ma, WY	6	34	生物化学和分子生物、计算机科学、理论与方法
6	Kitsuregawa, M	6	2	计算机科学、信息系统
7	Harries, G	6	4	计算机科学、信息系统

3 论文关键词分析

通过对论文的关键词进行分析，可以了解网络链接分析研究的方向和热点，从而对该领域的研究有一个比较准确、全面的把握。

3.1 根据论文所属的学科领域分类

纵观Web of Science中收录的关于网络链接分析的471篇论文，可以发现关于网络链接分析的研究论文按学科性质分类比重最大的前四个领域是：计算机科学，理论与方法；计算机科学，信息系统；信息科学与图书馆科学；计算机科学，人工智能。其中，计算机科学，理论与方法方面的论文有188篇，计算机科学，信息系统方面的论文有178篇，信息科学与图书馆科学方面的论文有84篇，计算机科学，人工智能方面的论文有56篇。以上数据体现了网络链接分析研究的核心的领域。

3.2 从论文的关键词看研究热点

通过对检索出相关论文的关键词进行统计分析，可以了解网络链接分析研究的方向和热点，这里列举了出现频数最多的一些关键词，见表4。

在检索到的相关论文471篇中含有关键词702个，其中出现频数最多的依次是互联网、链接分析、信息检索、语义网、搜索引擎、网页挖掘、PageRank算法、网页搜索、本体、实验法、网页图像、语义链接及链接结构。可以看出，这些年对于网络链接分析的研究主要集中在以上这些方面，涵盖了网络链接分析的算法、方法、应用等几个方面[4]。“互联网”出现的频次最高，为47次，因为网络链接分析是以互联网为基础的。“信息检索”出现23次，可见研究网络链接分析在信息检索中的应用是非常多的，是重点也是热点。另外，“语义网”出现14次，“搜索引擎”出现12次，“网页挖掘”出现10次，“本体”出现6次，这些都是对网络链接分析深入分析的领域和热点，有很大的潜力和广阔的前景。“PageRank”出现7次，“实验法”出现6次，这些是关于网络链接分析算法和方法的研究，PageRank自从被提出后在搜索引擎和其他网络链接分析应用中得到广泛应用，但是，相对于对网络链接分析应用的研究，目前对于网络链接分析的方法研究还远远不够，还有很长的路需要走，今后的研究方向需要更多得关注到对更有效的算法和更广泛的可行的方法的研究上来。

表4 1999－2007年网络链接分析论文关键词统计表

排 名	关键词	出现频次	比例
1	World Wide Web	47	4.9%
2	link analysis	30	3.1%
3	information retrieval	23	2.4%
4	semantic web	14	1.4%
5	search engines	12	1.2%
6	web mining	10	1.0%
7	PageRank	7	0.7%
8	web search	6	0.6%
9	ontology	6	0.6%
10	experimentation	6	0.6%
11	web graph	5	0.5%
12	semantic link	5	0.5%
13	link structure	5	0.5%

4 论文地区分析

1999－2007年关于网络链接分析研究的471篇论文共来自39个国家和地区，覆盖面比较广，各个国家和地区的论文量具体情况统计见表5，由于数量比较多，这里只列出论文较多的12个国家和

地区。

表5 1999－2007年网络链接分析各国家和地区论文量统计表

排名	地区	论文量	比例
1	USA	119	25.8%
2	England	60	13%
3	China	44	9.5%
4	Japan	31	6.7%
5	Canada	22	4.8%
6	Germany	21	4.5%
7	Israel	17	3.7%
8	South Korea	14	3%
9	Taiwan	13	2.8%
10	Greece	11	2.4%
11	Australia	1	02.2%
12	France	10	2.2%

从表5可以看出，美国的研究论文量达119篇，占世界总数的25.8%，表明美国在网络链接分析研究领域处于绝对的领先地位。互联网最早在美国发展起来，Google最早利用网络链接分析提高搜索效果。紧随其后的是英格兰地区，网络链接分析发文最多的两位作者Thelwall，M，Vaughan，L分别是英格兰的Wolverhampton Univ和Univ Western Ontario两所著名的高校的教授。处于第三位的是中国，研究的论文量占到了世界的9.5%。日本和加拿大分列第四和第五，德国紧随其后。另外，以色列、韩国、中国台湾、希腊、澳大利亚、法国也取得了相当的成就。值得注意的是，一些比较小的国家如以色列、韩国等在网络链接分析的研究中也有一定的进展，这充分说明网络链接分析这一研究方向在世界范围内都有了发展，引起了各国科研工作的关注。此外，从表5还可以看出，网络链接分析的研究形成了北美、西欧、东亚等几个主要的核心区域，并在这几个区域不断向外扩散，由此可以看到网络链接分析的广阔前景。

通过以上分析可以看出，网络链接分析在因特网普及率较高的欧美国家发展较快，无论是在算法研究还是应用研究方面都具有相对优势。中国由于人口基数大，虽然因特网普及率不高，但因特网的使用人口较多，在网络链接分析领域也已取得了一定的成果，但是与美国，英格兰还存在较大的差距，尤其在算法，方法的研究上，但随着互联网的普及和各国学术的交流，一定会缩小差距，取得很大的发展的。

5 引文分析

文献被引用次数是衡量该文献学术水平和科研价值的重要尺度之一，对文献进行被引分析可以让我们了解该学科领域的经典文献和成果。同时也可以通过对每年被引数的统计来预测该学科的发展趋势[5]。表6是对1999－2007年关于网络链接分析的论文被引数的统计。表7是对1999－2007年关于网络链接分析论文被引次数最多的前10篇的统

表6 1999－2007年关于网络链接分析论文被引数统计表

年份	2000	2001	2002	2003	2004	2005	2006	2007
被引数	20	45	110	250	320	360	370	240

表7 1999－2007年关于网络链接分析论文被引次数最多的前10篇统计表

序号	论文	作者	被引次数	年代
1	Authoritative sources in a hyperlinked environment	Kleinberg JM	305	1999
2	*Mining the web's link structure*	*Chakrabarti S, Dom BE, Kumar SR, et al.*	68	1999
3	Extracting macroscopic information from Web links	Thelwall M	66	2001
4	Bibliometrics and beyond: Some thoughts on web－based citation analysis	Cronin B	63	2001
5	Trawling the Web for *emerging cyber－communities*	*Kumar R, Raghavan P, Rajagopalan S, et al.*	48	1999
6	Scholarly use of the Web: What are the key inducers of links to journal Web sites	Vaughan L, Thelwall M	43	2003
7	Conceptualizing documentation on the Web: An evaluation of different heuristic－based models for counting links between university Web sites	Thelwall M	42	2002
8	Web impact factors for Australasian universities	Smith A, Thelwall M	42	2002
9	Web－based analyses of e－journal impact: Approaches, problems, and issues	Harter SP, Ford CE	36	2000
10	Readers, authors, and page structure: A discussion of four questions arising from a content analysis of Web pages	Haas SW, Grams ES	28	2000

在本文统计被引次数最多的前10篇论文中，1999年关于网络链接分析的论文篇被引次数最多的论文的作者是Kleinberg JM；2000年是Harter SP，Ford CE；2001年是Thelwall M；2002年是Thelwall M；2003年是Vaughan L，Thelwall M。之所以要做这样的统计，是因为1999年网络链接分析刚刚起步，研究的作者及发表的论文都非常少，随后几年的作者几乎会参考前几年的大部分论文，所以单单看被引次数多少决定学者的权威性会造成不准确和不可比性，而将被引次数与年代相结合则比较客观。由以上的分析可以看出，Thelwall M在5年中有3年的被引次数最多，堪称网络链接分析研究领域的经典学者。

通过表6可以看出，2000－2004年关于网络链接分析论文被引数几乎呈比例增长，2004到2006年有较小幅度的增长。网络链接分析从20世纪90年代末到现在一直处于高速发展期，随着互联网技术的发展，网络链接分析将一步步走向成熟[6]。

6 结 语

综上所述，本文通过从几个不同角度进行分析，主要研究了网络链接分析的进展，提出了网络链接分析的研究热点，并预测了其未来发展的趋势。网络链接分析近年来发展迅速，随着互联网的新发展，会越来越彰显网络链接分析的必要性，同时，互联网技术的发展，更能为网络链接分析提供新的思路和方法，网络链接分析的应用领域也会越来越广泛，其潜力非常大。由于种种原因，本文在研究角度上还不够细致，比如可以分国家和地区进行关键词统计，以此来分析各国家和地区的研究热点等，因此，本课题在今后还需要进一步研究，希望本文对网络链接分析的进一步发展有所帮助。

参考文献

1 董江山，胡吉祥，邱均平.链接分析法及其应用[J].情报科学，2004，(9)：1081－1086.

2 邱均平.信息计量学[M].武汉：武汉大学出版社，2007：315－319.

3 方舟子.美国学术评价的新招——h指数[EB/OL].http://edu.people.com.cn/GB/8216/4016420.html，2007－10－24.

4 张 洋，邱均平，文庭孝.网络链接分析研究进展[J].图书情报知识，2004，(6)：3－8.

5 杜友桃，何 琳.基于引文分析法的web超链接分析新进展[J].情报探索，2006，(9)：29－33.

6 陈凤娟，邵 波.跨语言信息检索文献的计量分析[J].中国信息导报，2007，(8)：35－38.

(责任编辑：赵立军)

(上接第1129页)

例如广东佛山禅城区政府语音网站的就是一个地区级的面向残疾人的特殊网站[8]，该网站在改版的过程中，设立了国内首个为视障人士服务的政府语音网站，它附设于禅城区政府网，只需点击禅城区政府网最上端显示的“语音”栏就可进入语音网站，一切操作都可以根据语音提示进行操作，使用者只需要控制键盘上的数字键就可以选取收听内容。

就目前的状况，政府可以采取分阶段实施政府网站的信息无障碍整改与建立专门针对特殊需要人群的特殊网站同步进行的方式，来逐步推进政府网站信息无障碍获取，以便最终实现政府网站的全面无障碍获取的目标。

参考文献

1 何 川.国内信息无障碍的现状及展望[J].现代电信科技.2007，(3)：4－8.

2 中国联通.信息无障碍标准的研究[EB/OL].http://www.chinaunicom.com.cn/profile/xwdt/txjs/file1186.html，2007－05－30.

3 任铁民.信息无障碍是残障人士贫困人口等弱势群体的基本发展权[J].中国信息界，2007，(4)：31－34.

4 IT Accessibility & Workforce Division (ITAW) - Office of Governmentwide Policy，Section 508 Standards[R].http://www.section508.gov/index.cfm? FuseAction = Content&ID = 12，2007－11－10.

5 钱小龙，邹 霞.美国信息无障碍事业发展概况“WCAG1.0解读”[J].中国特殊教育，2007(6)：70－74，69.

6 武晓鹏.政府门户网站 如何实现无障碍[EB/OL].eNet硅谷动力.http://www.xinxuyao.com/access/information/200605222099.shtml，2007－09－04.

7 World Wide Web Consortium[EB/OL].http://www.w3.org/TR/WCAG20/，2007－09－20.

8 王 鹰，张智轩.广东佛山盲人也能上网冲浪[EB/OL].佛山日报.2007－03－16，http://www.echinagov.com/echinagov/redian/2007－3－16/12759.shtml，2007－09－23.

(责任编辑：赵立军)