

国外 Google Scholar 研究的文献计量分析

潘 玣

(无锡商业职业技术学院 艺术设计学院, 江苏 无锡, 214153)

摘要: 以 Web of Science 为文献源, 采用文献计量学与信息可视化技术相结合的方法, 借助可视化分析软件 CiteSpace II, 针对 2004 年至今的以 Google Scholar 为研究主题的文献, 从文献量、作者、期刊、重要文献和研究主题等方面进行分析, 以揭示 Google Scholar 的研究现状。

关键词: Google Scholar, 定量分析, 文献计量, CiteSpace II

中图分类号: G350 **文献标识码:** A **文章编号:** 1007-7634(2012)01-77-04

Bibliometric Analysis of Research on Google Scholar Abroad

PAN Heng

(School of Art Design, Wuxi Institute of Commerce, Wuxi 214153, China)

Abstracts: Using the database of Web of Science as information origins, the paper takes the quantitative analysis method and information visualization technology, and uses CiteSpace II to analyze the documents on Google Scholar since 2004, from the aspect of document counts, author, journal, important documents, and research topics, hoping to demonstrate the status quo of the research on Google Scholar.

Keywords: google scholar;quantitative analysis;bibliometric ;citespace II

2004 年 11 月, 搜索引擎国际巨头 Google 推出新的搜索服务—Google Scholar(Beta 版), 为用户搜索免费数字学术信息资源提供了一条新途径。Google Scholar 提供了多种类型、多个学科领域的文献信息。Google Scholar 一经推出, 便迅速受到了广泛关注, 尤其是图书馆学情报学领域, 对于这个新生信息检索与服务工具给予了重点关注, 为了了解近几年来学者对于 Google Scholar 的研究现状, 笔者选择国际权威引文数据 Web of Science (SCI-EXPANDED, SSCI, A&HCI) 为文献源, 通过采用传统的文献计量方法, 并结合利用 CiteSpace II 进行可视化分析, 以期揭示 Google Scholar 研究现状, 并对今后研究有所启发。

收稿日期: 2011-09-23

作者简介: 潘 玣(1978-), 男, 江苏人, 主要从事信息检索与网站评价研究。

1 文献量分析

笔者在 Web of Science 中进行检索, 得到 272 篇文献, 删除不相关文献后, 得到文献 254 篇, 具体如图 1 所示。

Google Scholar 在 2004 年 11 月才由 Google 推出, 当年即有 4 篇相关文献出现, 2005 年的文献量更是到 50 篇, 为 7 年文献数量最高的年份, 这充分说明了 Google Scholar 的推出确实迅速受到了关注。而从图中也可以看出, 从 2005 年至今, Google Scholar 研究文献数量基本保持稳定状态, 据此可以预计, 在未来一段时间, Google Scholar

的相关问题仍将是界内比较关注的问题。

此外,需要指出的是,254篇文献中,最早的文献是来自Nature的一则新闻报道“Science searches shift up a gear as Google starts Scholar engine”。Google Scholar的推出能登上Nature,说明其推出确实引起了学术界的重视。

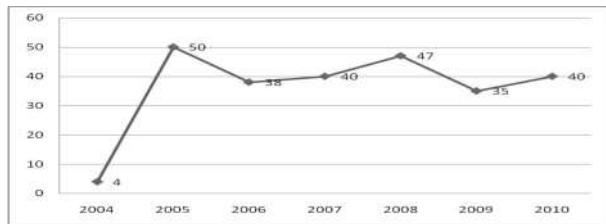


图1 Google Scholar研究文献数量分布

2 作者分析

2.1 发文数量多的作者

对检索得到的254篇文献进行作者的统计分析,发现除去16篇文献没有作者以后,共涉及到作者287位作者。其中,发表文献最多的是来自美国夏威夷大学(University of Hawaii)的Jacso P,共发表文献14篇,Jacso P关注的是Google Scholar的检索功能、引文功能等内容,如比较分析Web of Science、Scopus以及Google Scholar计算h指数的区别等;排在第二的是来自伊朗德黑兰大学(University of Tehran)的Kousha K,发表6篇文献,关注的重点是Google Scholar的引文分析功能,尤其是Google Scholar与Web of Science引文分析功能的比较;此外,BarIlan J以及Chillingworth M等也是发文较多的作者。

2.2 被引频次高的作者

由于Web of Science中提供了文献的引文信息,通过这些信息可以明确在某学科领域中被引频次高的作者。近年来,信息可视化技术成为了信息分析中的前沿技术,通过信息可视化,可以将知识和信息中引人注目的最前沿领域和学科制高点,以可视化的图像直观地展现出来。其中,由美国德赛尔大学(Drexel University)陈超美博士所开发的CiteSpace II^[1]是可视化软件中的翘楚。本文利用该软件,对Web of Science中的作者进行可视化分析。

选择关键路径算法,网络节点为共被引作者(Cited Author),时间区为1年,阈值为(4,3,16),(4,

3,18)和(4,3,20),绘制作者共被引的知识图谱,并统计关键节点数据,具体情况如图2和表1所示。

表1 Google Scholar研究领域作者共被引关键节点统计结果

序号	频次	中心度	作者
1	91	0.10	Jacso P
2	62	0.02	Meho L
3	57	0.06	BarIlan J
4	49	0.11	Garfield E
5	45	0.01	Hirsch J
6	42	0.02	Kousha K
7	38	0.03	Cronin B
8	38	0.03	Bauer K
9	36	0.10	Neuhaus C
10	35	0.04	Bakkalbasi N



图2 Google Scholar研究领域作者共被引知识图谱

从图2和表1可以看出,Google Scholar研究领域单篇共被引频次最高的为Jacso P,达到了91次,其次是Meho L,排名第十的也达到了35次。而在中心性方面,Pauly D和Kessler RC表现最为出色。综合发文数量和被引频次,可以看出Jacso P、BarIlan J、Kousha K、Meho L、Pauly D以及Kessler RC等人在Google Scholar占据着重要地位,是这个研究领域的领军人物。

3 期刊分析

据统计,254篇文献共刊载在145份期刊上,其中,刊载最多的是Journal of Library Administration,刊载文献15篇,其次是Online Information Review。刊载数量Top 5的期刊如表2所示。

表2 刊载数量Top 5的期刊

期刊名	文献量	国别
Journal of Library Administration	15	美国
Online Information Review	13	英国
Journal of the American Society for Information Science and Technology (JASIST)	10	美国
Scientometrics	9	匈牙利
Internet Reference Services Quarterly	8	美国

从表2可以看出,5份期刊中有3份是来自美国,这也表明美国是研究Google Scholar的主要阵地。而JASIST等则是界内最为知名的期刊,从一个侧面反映出Google Scholar的研究具备了较高的水准。

4 重要文献分析

评价论文学术影响力的一个重要指标即是被引频次,被引频次越高,说明其影响力越大。通过检索,“Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google Scholar”被引频次最高,被引106次;被引频次Top 10的文献如表3所示。

看出,Jacso P和Kousha K分别有2篇文献进入被引高前10的文献,而JASIST和Scientometrics则分别刊载了2篇文献。被引次数最高的是由Meho L等于2007年发表的“Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar”,该文献主要是比较Web of Science, Scopus 和 Google Scholar统计科研人员科研成果被引频次的区别,并以此对科研人员进行排序^[2]。Jacso P在“As we may search Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases”对Web of Science, Scopus 及 Google Scholar的学科收录范围与组成、数据库资源数量、检索界面、检索途径以及检索某一特定文献的能力等情况^[3]。排在第三的论文“Which h-index? – A comparison of WoS, Scopus and Google Scholar”比较了利用WoS, Scopus 以及 Google Scholar计算h指数所存在的区别^[4]。

5 研究主题

通过深入文献内部,对文献主题进行分析,可以

较好地反映出某一个研究领域的研究进展和研究热点,据统计,254篇文献共涉及到Google Scholar的功能分析与评价,尤其是引文功能的比较分析与评价,Google Scholar对图书馆发展的影响以及Google Scholar的应用等方面。

5.1 Google Scholar功能分析与评价

Google Scholar不仅是一个学术搜索引擎,同时也是一个引文分析工具,任何网络用户都可以通过Google Scholar检索得到某篇文献的被引频次以及施引文献的详细信息。Google Scholar的引文分析功能,无疑是其最大的亮点,这也给了学者打破Web of Science在引文分析垄断局面的想象空间。因此对Google Scholar的引文分析功能进行分析与评价,尤其是与Web of Science等引文数据库进行比较就成了众多学者广泛关注的对象。如Kayvan K等人利用Google Scholar和Web of Science对自然科学和人文社科领域的开放存取的学术影响力进行比较分析,结论认为Google Scholar比Web of Science更能真实反映文献的被引情况,尤其是被开放存取资源引用的情况,这对于数字环境下的学术研究来说是至关重要的^[5]。而对于Google Scholar的综合表现,如检索界面、检索功能、检索结果等, Jacso P给予了比较多的关注,结论认为,Google Scholar既有界面友好、亮点突出、免费使用等多种优点,同时也存在着数据不够准确、更新不够及时、缺少人工干预等不足^[6]。

5.2 Google Scholar对图书馆发展的影响

对于图书馆来说,Google Scholar的推出,既是机遇,又是挑战。因此,不少图情领域的学者,从图

表3 Google Scholar研究领域中被引频次Top 10的文献

序号	作者	题名	年份	期刊	被引次数
1	Meho L, Yang K	Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar	2007	JASIST	106
2	Jacso P	As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases	2005	Current Science	80
3	BarIlan J	Which h-index? – A comparison of WoS, Scopus and Google Scholar	2008	Scientometrics	56
4	Falagas M等	Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses	2008	Faseb Journal	53
5	Jacso P	Google scholar: The pros and the cons	2005	Online Information Review	51
6	Noruzi A	Google Scholar: The new generation of citation indexes	2005	Libri	47
7	Kousha K, Thelwall M	Google scholar citations and google Web/URL citations: A multi-discipline exploratory analysis	2007	JASIST	44
8	Kousha K, Thelwall M	Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines	2008	Scientometrics	28
9	Walters W	Google Scholar coverage of a multidisciplinary field	2007	Information Processing & Management	26
10	Neuhaus C等	The depth and breadth of Google Scholar: An empirical study	2006	Portal–Libraries And The Academy	26

书馆的角度探讨Google Scholar对于图书馆发展的影响,思考如何有效利用Google Scholar,推动图书馆的发展。如Arendt J认为Google Scholar的出现对于图书馆的数字资源建设是一个较好的机遇,如果用户可以通过Google Scholar免费获取所需的数字资源,则图书馆可以在数据库的购买上提升自己的主动性。该作者将现有主流商业数据库与Google Scholar进行比较,分析各自的优势与不足,并对图书馆数据库的采购给出了两点建议:将Google Scholar与准备购买的商业数据库进行比较,确定商业数据库的优势。如果商业数据库确实有其独特的优势,那么相比其价格,是否值得购买^[7]。Bowering M L等人则对图书馆是否有效利用Google Scholar进行较为全面的调查研究。2005年,该作者对美国100多所高校图书馆进行了调查,调查涉及到高校图书馆是否在网站主页上提供Google Scholar的链接、Google Scholar是否出现在图书馆的数字资源列表中、Google Scholar是否出现在图书馆的OPAC中等9个问题^[8],结果表明,已有不少图书馆已经采取措施,利用Google Scholar开展信息服务。

5.3 Google Scholar的应用研究

从本质上来说,Google Scholar是一个检索工具与引文分析工具,因此,如何有效利用Google Scholar,充分发挥其作用,成为了学者关注的一个焦点话题。不少学者探讨了如何利用Google Scholar进行学术评价,包括对科研人员、科研机构、科研成果、期刊等多个角度的学术评价,如Onyancha O B就利用Google Scholar对撒哈拉沙漠以南非洲国家的图书馆学情报学期刊的学术影响力进行了评价,评价指标包括期刊数量、刊载论文数量、被引频次、篇均被引频次、h指数、g指数以及影响因子等。结果认为,利用Google Scholar可以用来评价未被ISI收录的期刊的学术影响力,Google Scholar也使得综合采用引文分析、文献出版数量等多个指标进行评价成为了可能^[9]。Cesare R D等人则利用Google Scholar来评价了灰色文献(Grey Literature)的学术影响力,结论认为,Google Scholar的出现,大大提升了灰色文献的可见性,同时也可以通过不同版本的全文来反映文献发展变化的生命周期。并且,Google Scholar最大的优点是对其收录

的不同语言的文献记录都是一视同仁^[10]。

6 结语

综上所述,国外已经对Google Scholar展开了较为广泛的研究,文献量较大且整体呈现稳中有升的态势,已形成了一批核心作者和核心期刊,有较为稳定的研究主题。因此,对于国内研究人员来说,如何借鉴国外现有研究成果,为国内图书馆利用Google Scholar更好地开展信息服务,是我们值得思考的问题。

参考文献

- 1 CiteSpaceII: Visualizing Patterns and Trends in Scientific Literature[EB/OL], <http://cluster.cis.drexel.edu/~cchen/CiteSpace/>,2011-06-10.
- 2 Meho L, Yang K, Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar [J]. Journal of the American Society for Information Science and Technology, 2007, 58(13): 2105–2125.
- 3 Jacso P, As we may search Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases [J]. Current Science, 2005, 89(9):1537–1547.
- 4 Barllan J, Which h-index? – A comparison of WoS, Scopus and Google Scholar [J]. Scientometrics, 2008, 74(2): 257–271.
- 5 Kayvan K, Thewall M. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines [J]. Scientometrics, 2008, 74 (2):273–294.
- 6 Jacso P, Google Scholar: the pros and the cons [J]. Online Information Review, 2005, 29(2):208–214.
- 7 Arendt J, Imperfect Tools: Google Scholar VS. Traditional Commercial Library Database[J] Against the Grain, 2008, (4): 26,28.
- 8 Bowering M L, Karen H A, Google Scholar and the Library Web Site: The Early Response by ARL Libraries [J]. College & Research Libraries, 2006, 67(2): 106–122.
- 9 Onyancha O B, A Citation Analysis of Sub-Saharan African Library and Information Science Journals using Google Scholar [J]. African Journal of Library, Archives and Information Science, 2009, 19(2): 101–106.
- 10 Cesare R D etc. The impact of Grey Literature in the web environment: A citation analysis using Google Scholar [J]. The International journal on grey literature, 2008, 2(3): 83–86.

(责任编辑:孙晓明)