

●原宏盛

直接满足中文用户科研文献需求的网络搜索工具*

——Google Scholar 初探

摘 要 Google Scholar 即 Google 学术搜索,是建立在 Google 搜索引擎基础上、直接面向科研需要的学术资源的网络检索工具。其搜索内容涉及诸多学科,并经业内专家评审,具有相当的权威性。该搜索引擎具有检索操作便捷化,选题标准学术化,引用搜索智能化,瞬间运行高速化等特点。Google Scholar 及其中文版的出现,对于推动我国的科技进步和学术研究,具有积极意义。参考文献 31。

关键词 Google 学术搜索 搜索引擎 网络学术检索

分类号 G354

ABSTRACT Google Scholar is a network search tool for academic resources. It is based on the Google search engine and oriented to academic researchers and covers a variety of disciplines. In this article, the author analyzes its user-friendliness, strict selection criteria, smart search functions, etc. It will play positive roles in the development of science and technology in China. 31 refs.

KEY WORDS Google Scholar. Search engine. Online academic search.

CLASS NUMBER G354

“文献信息载体的‘虚拟化’及其传递获取的‘网络化’,是现代图书馆基本品格和根本特征最为集中的体现。”^[1]但这一品格和特征的实现并非易事。必备条件之一就是需要强大的网络文献信息资源存取工具作为技术支撑和后盾。2006 年诞生的中文版 Google Scholar,正是当代中文领域最新、最先进、最快捷的全球网上学术科研文献资源搜索工具和检索入口。

1 从 Google、Google Scholar 到中文版 Google Scholar 的演化

1999 年,在美国斯坦福大学的宿舍里,计算机科学专业博士生拉里·佩奇(Larry Page)和谢尔盖·布林(Sergey Brin)在做着一个梦——颠覆网络传统,改变互联网上的信息走向^[2]。他们凭着其数学天赋,成功开发出 Google 搜索引擎。这个基于链接引用次数原理和关键词搜索的 Google,在技术上具有“宇宙震荡波”式的创新,终于使 Google 成为当今世界搜索引擎之王。

所谓搜索引擎,就是指利用搜索机器人自动获取

以网页为主的因特网信息、制作目录供用户查询的网络资源检索工具^[3]。作为其杰出代表,Google 几乎成了搜索引擎的代名词。但对科研工作者来说,它提供的信息仍过于冗杂。为此,在其创始者产生 Google 梦想的 5 年后即 2004 年 11 月 18 日,Google 公司在不断创新出诸多搜索项目产品的过程中,又推出了一种崭新的搜索服务模式:“Google 学术搜索”(Google Scholar)^[4]。

任何产品或服务都是在有需求的情况下催生的。Google 学术搜索的诞生,同样始于一位普通的专业科研用户的查询需求:一名 Google 工程师在日常工作中发现,在用 Google 搜索出的浩如烟海的网上资源中,还是很难精准、迅速地找到自己急需的科研资料。于是,便开始酝酿开发一种专门针对学术科研资料的搜索服务。令人叹服的是,Google 学术搜索的研发,从想法诞生到项目推出只用了 9 个月,并且只有 4 名工程师参与开发和程序编写^[5]。

因此我们可以这样说,所谓 Google Scholar 即 Google 学术搜索,是建立在 Google 搜索引擎基础之上、直接面向科研需要的学术资源的网络检索工具。

* 本文系江苏省哲学社会科学规划基金项目“图书馆数字化、网络化发展模式与管理模式研究”(N2-069)的研究成果之一。

它专门用于帮助用户查找包括期刊论文、学位论文、专业图书、预印本、文摘和技术报告在内的学术文献,内容涉及诸多学科领域并经过业内专家评审,具有相当的权威性。Google Scholar 与计算机协会、国际电工与电子工程师协会、OCLC 开放性 WorldCat 图书馆定位服务等出版机构建立了合作关系。Google 也收录了 Open Archives Initiative 中由 OAIster 收集的上百万篇论文,并提供 CrossRef 的链接服务。

然而对于中文用户来说,虽然通过 Google Scholar 搜索国外学术科研资料非常方便,但那时 Google Scholar 仅有英文版,并不支持中文。为及时填补这一空白,Google 在 2005 年 10 月着手组建其中国公司^[7],于 2006 年 1 月 11 日召开新闻发布会,世界上第一个“非特定人连续语音识别系统”的创造者^[8]、Google 全球副总裁兼中国区总裁李开复宣布正式推出中文版 Google 学术搜索^[9],历史性地使 Google Scholar 扩展至中文学术科研领域,成为网络时代专门满足中文用户学术研究需要的有力的搜索工具和科研助手。

Google Scholar 不仅从 Google 收集的上百亿个网页中,筛选出具有学术价值的内容,而且最主要的方式是,通过与传统资源出版商的合作来获取足够的具有学术科研价值的文献资源。它首先选择的是各国的学术资源出版商的数据库,通过与其签订协议进行数据库的开放合作^[10]。目前,Google Scholar 中文版的合作伙伴主要有,中国大陆的维普中文科技期刊全文数据库、万方数据资源系统等^[11];台湾地区“华艺数位”的中文电子期刊服务(CEPS),也在 2005 年 10 月与 Google 合作,为台湾最早以学术全文资料库加入 Google Scholar 者^[12]。这些合作的基本机制为,数据库商设立专门的服务器或在普通服务器上开设专用通道,Google 通过网页采集机器人自动收集元数据,并自动加入 Google Scholar 及其中文版,实现即时服务^[13]。

2 中文版 Google Scholar 的强大优势及其主要特点

Google Scholar 在 Google 检索主页没有直接的链接,必须从其专用网址 <http://scholar.google.com> 或 <http://www.google.com> 才能进入。虽然从推出至今还不到半年,但已显示出它不可取代的现实作用和独特的强大优势。

2.1 检索操作的便捷化

任何读者在任何时间都可以在世界上任何一台

与因特网互联的计算机上极其方便、免费地利用 Google Scholar 中文版的功能和资源,从而真正实现了超越时空概念、跨越地域界限的学术科研文献资源共享^[14]。并且不论是普通读者还是专业用户,不必引入布尔逻辑运算或其他种种限定设置条件,也不必非要使用高级检索,只要在其主页面的输入框中,通过随意键入关键词的简单方式,就能实现新兴的“点对点”(P to P)的检索需求^[15]。其中,作者检索是 Google Scholar 的鲜明特色,也是查寻具体文献最有效的方法之一。如果已知所需文献的作者,只要输入其姓名,就会立即跳出“你”想得到的,与“他”相关的学术成果的题录、文摘、论文、著述等。同理,亦可把学科热点、研发对象、选题目标、出版物名称等具体易化为关键词,同样可以成为达到检索愿望的各种具体途径。实践证明,只要通过这样的简易操作,就能够在这个对全球开放的文明宝库中任意提取知识,使中文用户轻松地实现图书馆人梦寐以求的,从以“文献”为中心的聚集、组织、管理和服务,到以“知识”为核心的聚集、组织、管理和服务的历史性过渡,满足信息时代对文献、知识检索利用的简单化、方便化的迫切需要。

2.2 选题标准的学术化

与其他商业搜索引擎不同,Google 本身就一直坚持“专业的学术精神而著称”^[16]。Google Scholar 继承、突出了这一传统,把学术性作为其唯一的选题标准和价值取向,使其“在世界学术领域的地位与日俱增”^[17],促使越来越多世界著名的出版商、数据库商甚至图书馆都在高度“重视这一新生势力对未来学术界带来的巨大影响”^[18]。面对这一有利态势,Google 公司用特定技术在数以亿计的、面向互联网的开放网页中自动提取学术内容,还更为主动地通过对外合作途径,从两个方向直接获取专业文献,并以此作为 Google Scholar 学术资源的另外两大组成部分:一方面经过努力,已经取得与 ACM、Nature、IEEE、OCLC 等专业文献数据库(以刊为主)出版商的协作和支持。一方面在推出 Google Scholar 后约 1 个月即 2004 年 12 月 13 日晚宣布,它将与美国纽约公共图书馆以及哈佛大学、斯坦福大学、密歇根大学和英国牛津大学的图书馆合作,把这些世界顶级图书馆的部分馆藏扫描制作成数字资源(以书为主),放到互联网上供全球读者搜索、阅取。如果此举成功,届时 Google 将与图书馆界共同合作建设一座全球最大的虚拟图书馆。在这一过程中,同时“建立学术信息服

务的新体系,打造学术评价的新标准”^[19]。其“合作伙伴正在世界范围快速增加”^[20],一年后就紧锣密鼓地在中国推出了 Google Scholar 中文版。这个“中国化的学术搜索”,秉承学术宗旨,严格选题原则,坚持“一套自成体系的判断标准”^[21]。娱乐性内容、大众读物等消遣性的东西很难进入 Scholar 体系,从而保证了它的专业规范化和学术纯洁性。进入中文版 Google Scholar 的,主要是论文、著述、文摘、技术报告等学术科研文献。它们的最终源头是各学术出版物、专业学会、预印本库、大学及专业学术性网站。

2.3 引用索引的智能化

文献引用检索及索引在学术研究中极其重要。但对文献被引情况全面而准确的搜集整理却十分困难和耗时费力,是学术科研工作极难解决而又急需解决的基础性问题。而从上世纪末,伴随着计算机智能化研究的不断发展,具有适应性和学习性特征的智能检索技术正从试验阶段过渡到实际应用,搜索引擎与“人工智能(artificial intelligent)已成为网络信息检索的关键技术与核心思想。智能检索可以通过检索内容相关性的自动学习,来提高检索结果的查全率和查准率。”^[22]而 Google 的核心技术之一,就是由拉里·佩奇和谢尔盖·布林创造并应用的 PageRank 算法,其原理正类似于科技论文中的引用机制^[23]。Google Scholar 的检索功能则承接了 Google 的这种引用机制^[24],使其能够专门提供文献引用检索和索引,成为搜索引擎观念和技术的“一项创新”^[25]。融入 Google 及其 Scholar 多项创新观念、技术的 Google Scholar 及其中文版,用“特有的按 Cited 值(被引用次数)呈现搜索结果就是对这一创新观念的继承”^[26],其引用的索引工作也由程序自动完成,这是它们与 Google 通用引擎相比的又一突出优势和特有功能。因此,从中文版 Google Scholar 检索得到的每篇文献,都是以“Cited by”链接形式的题录面目出现,每条记录首行的“文献题名”和末行的“被引用次数”(未被引者无)一目了然。只要对这两者进行点击就能分别显示出文献引用与被引的索引链接,两者分别暗涵了“来源文献”与“被引文献”两个逻辑思维检索取向:点击前者就会自动出现相关文摘,在其中的参考文献下,会列出被该文献引用过的来源文献(还可对它继续点击);点击后者,即刻出现该文献被引频率及情况、即自动显示出引用过该文献之文献的题录(又可对它继续点击)。凡在点击这前后两者过程中出现的所有文献,又都可按这两个逻辑取向持续进行交互

式点击,从而形成了以引用索引为中心的文献链。从它们之间相互引证的关系中,既可揭示文献之间的内在联系,又可揭示已知理论和知识的应用、提高、发展和修正过程,从一个侧面揭示学科研究发展的基本走向,并有利于对新知识的有效发现和对学术成果的科学评价,对交叉学科和新学科的研究具有重要参考价值。

2.4 结果排序的科学化

在网络时代知识爆炸的信息海洋中,对于科研用户来说,困难的常常并不是提供而是如何遴选知识,即怎样花最少的时间寻找出最急需的东西。文献在检索结果中的排序状态,就成为他们选择文献信息资源检索工具的一个重要尺度。而 Google Scholar 及其中文版,对搜索到的文献排序取决于3个要素:一是被引因素(Cited 值),二是时间因素^[27],三是优先因素即前者优先于后者。只有 Cited 值相同时,文献才按照时间降序排列。这些原则在 Google Scholar 中文版中表现得更加专业化。

(1)在按检索目的输入关键词经搜索后,作为检索结果呈现的文献集合页面上,左侧实际上显示的为“全部文章”排序,它坚持被引因素优先,按照 Cited 值大小排序,Cited 值越大者排位越靠前;文献被引越多,表明其参考价值越高,故而搜索排名理应越靠前。这保证了尽管索引量极大,但还是能把最佳结果排在搜索结果的最前面。实践证明,这样做常能命中优秀文献,用户也常能方便地搜索到参考价值较高的文献。但由于文献被引频率还取决于其发表时间,一般需2~4年才能进入被引高峰期,而后发表的文献还需通过时间提高被引量来证明其学术价值,因此这种排序也有缺陷。

(2)点击页面右上方处“近期文章”,就可以按照时间后先排序,即发表或出版时间越近者排位越靠前,从而弥补了(1)之缺陷。

(3)显示出的每篇文献都表现为题录形式,其格式固化性地依次排列出题名、作者、出版物、出版年、数据库商、刊期、刊页、摘要提示等符合图书馆规范的编目信息。

(4)仔细分析,还可体会出中文版 Scholar 在所显文献中,暗含着按照文献“三大块”的外在表现形式进行排列的内在总体逻辑顺序:点击“题录”中的题名→显示“文摘”→点击其中的“阅读全文”按其提示操作可获取全文。总之,中文版 Google Scholar 通过对于学术规范的专业化表现,把检索结果排序的科学

性表现得淋漓尽致。

2.5 瞬间运行的高速化

面对网络环境、科技创新时代,对有效信息、知识、文献的获取速度已越来越成为决定科研成败的重要因素之一。“同时,也给人们如何快速有效地获取自己需要的信息提出了更严峻的挑战”^[28]。而万维网“是一个分布式动态快速增长的由各类文档组成的海量信息资源中心”^[29],在网络资源激增的同时,也增加了读者对其有效利用的复杂程度和获取困难程度。即使是图书情报界推出不久的馆内“跨库检索系统”和“一站式”检索,也不能满足学术科研用户对检索效率的要求。而 Google Scholar 及其中文版,既不需用户登录,又不要输入密码,只要输入查询词,仅需 0.05 秒左右(经笔者反复测试)这几乎感觉不到的等待,即可实现时空穿越,成为在国际互联网织就的信息领域和全球信息高速公路的“发动机”与“扳道工”,让包括图书馆、数据库商在内的各种信息机构的门户栅栏烟消云散,瞬息之间便可将用户急需的但却分散于各地的文献信息资源整合加工,集中呈现出来供其利用,使之真正享受到获取文献信息资源的一站式服务。这时的 Google Scholar 及其中文版,就是一个虚拟的,但却真实高速运行的、提供统一检索入口的、全球性的跨库检索平台和学术搜索引擎。之所以能够做到这些,一方面归功于高效的搜索算法,另一方面则归功于 Google 将数以千计的低成本计算机联网到一起,制造出了一部超高速搜索引擎^[30]。

已有学者指出,随着互联网深入到社会的各个角落,网络信息内容的全面性、流逝性和随意性使其成为一种珍贵的社会资源,对互联网信息研究有可能成为研究我们社会的一个独特的途径。在这样一个海量信息资源库之上,人们还能开展各种研究工作,许多领域的“学者都有可能从中发现自己感兴趣的宝藏,发展出新的研究方法,取得创新的成果。”^[31]在此大背景下,直接用于学术科研的 Google Scholar 及其中文版的出现,对于推动我国的科技进步和学术研究,更具有标志性的重大意义。虽然目前它仅是个测试版即 Beta 版,还要经历一段较长时间的完善、发展,但其强大优势表现出来的突出地位,更将对中国的图书情报事业带来深刻影响。

参考文献

- 1,14 原宏盛.现代图书馆“虚”“实”论.中国图书馆学报,2005(5)
- 2 张卫华,王建荣.有了 Google:把斯坦福大学图书馆搬回家. [2005-11-17]. http://news.xinhuanet.com/newmedia/2005-01/11/content_2443159.htm
- 3 毛军.数字图书馆的 Google 之道.图书情报工作,2005(3)
- 4 奇迹报道. Google 推出面向科研人员的搜索服务. [2006-06-20]. <http://www.qiji.cn/news/scinews/2004/11/23/20041123114100.htm>
- 5,9,11 伊豪. Google 中文学术搜索发布 李开复承诺不收费. [2006-06-22]. <http://news.ccidnet.com/art/1032/20060111/412613-1.html>
- 6,15,23,24 张文彦. Google Scholar 与图书馆的未来. 中国信息导报,2005(9)
- 7,26,27 唐磊. 全新医学文献搜索平台——Google 学术搜索. 寄生虫病与感染性疾病,2006(1)
- 8 张岚. Google 中文学术搜索上线 李开复称不求收益. [2006-06-23]. <http://www.zdnet.com.cn/news/net/story/0,3800050307,39435562,00.htm>
- 10,13,16,17,18,19,20,21 毛力. Google Scholar 的出现与期刊评价. [2006-06-24]. http://www.d-library.com.cn/info/info_literature_bq_detail.jsp?id=307
- 12 线上资料库消息报道. Google Scholar——找寻学术资料的新引擎工具. [2006-06-26]. <http://irlab.cis.nctu.edu.tw>
- 22 徐谦. 网络信息检索的智能化趋势. 图书馆理论与实践,2006(2)
- 25 朱佳鸣. Google Scholar Beta 检索性能的初步分析. 图书情报工作,2005(12)
- 28 陈汉深,李卫忠. 基于 C/S 的新一代智能化、个性化搜索引擎. 情报学报,2006(1)
- 29 王继民,彭波. 搜索引擎用户点击行为分析. 情报学报,2006(2)
- 30 Google 公司. 为什么选用 Google. [2006-06-28]. http://www.google.com/intl/zh-CN/why_use.html
- 31 李晓明,阎宏飞,王继民. 搜索引擎——原理、技术与系统. 北京:科学出版社,2005

原宏盛 苏州职业大学图书馆副馆长、副研究馆员。通信地址:苏州市干将西路 1122 号—苏州电大敬德图书馆。邮编 215004。(来稿时间:2006-07-12)