

● 刘 晨

Google 解读及图书馆的应对思考

摘 要 在某种程度上,Google 与图书馆、数字图书馆有着理想主义的重合,又有着某些现实主义的优势。Google 最值得图书馆学习的是它的理念和意识。Google 和图书馆可以互相合作,实现共赢。图书馆也有超越 Google 的领域。参考文献 13。

关键词 Google 图书馆 数字图书馆 比较研究

分类号 G250.73

ABSTRACT The author thinks that Google has some similarities with library in its idealism and also has some realistic advantages. Libraries should learn the mission and conception from Google, co-operate with Google, and do something beyond the fields of Google. 13 refs.

KEY WORDS Google. Library. Digital library. Comparative study.

CLASS NUMBER G250.73

1 Google 的发展

Google 两位创始人是 Sergey Brin(谢尔盖·布林)和 Larry Page(拉里·佩奇)。1995 年两人在美国斯坦福大学攻读博士学位时相识,共同设计了 PageRank 计算方法,休学后,于 1998 年 9 月 7 日在加州一个车库中创立 Google。

Google 是一种开放式搜索引擎,索引内容完全是由其联结分析进行自动添加,人工干预排序极少。它的自动搜索技术,目前唯一被公开的只有 PageRank(网页评级)一种算法。PageRank 原理来自学术界的“引用”,即学术论文的引文系统。互联网是松散的建立在引用基础之上,而链接就是某种意义上的“引用”。PageRank 就是一种寻找网络上反向链接的数量,并评估其质量的方法^[1]。它检查整个网络链接结构,通过对超过 5 亿个变量和 20 亿个词汇组成的方程利用稀疏矩阵进行计算,结合超文本匹配分析技术,综合考虑网页的整体重要性及特定查询的相关性,将获得的 PageRank 最高、最相关可靠的搜索结果放在首位^[2]。

与 PageRank 技术的搜索运算系统匹配,Google 采用 shared-nothing architecture(分布式共享体系)的基础结构,具有超强容错和负载平衡能力的巨型分布式文件系统(GFS),大规模机器学习系统,实现海量并行计算的各式 Google 服务。其他 Google 技术还有诸如 Dynamic page summaries;2003 年 11 月 Google “佛罗里达”更新中的 HillTop 算法^[3];Google 还推出了无线搜索技术。

Google 还通过收购其他公司技术的方法迅速发展扩张。2003 年 2 月,收购全球最大 Blog 服务商 Blogger.com;11 月,收购一家网络社区服务公司

Orkut.com。2004 年,收购 Neotonic 公司,获取 Gmail 软件技术;11 月,收购美国 Keyhole 公司,先后推出 Maps 服务和 Google Earth。2006 年 11 月,并购视频共享网站 YouTube。

2001 年,Google 两位创始人分别自任技术和产品总裁,聘请曾任 SUN 和 Novell 公司 CEO 的 Dr. Eric Schmidt 担任新董事长兼 CEO。之后, Mozilla Firefox 浏览器首席开发人员 Ben Goodger,微软 Windows 系统首席架构官 Marc Lucovsky,微软前高级副总裁李开复,全球 Linux 领袖 Andrew Morton 先后加盟。

Google 早已不局限于搜索引擎,而是迅速扩展到互联网的各个角落。创立初始的目标“接管整个互联网”,今天看来已不再是笑话和大话,它要建立一个以 Google 搜索为核心的互联网生态系统。目前 Google 每月全球唯一身份用户为 3 亿 8000 万,全球网民可以使用 100 多种语言,搜索 80 多亿个网页,10 亿多张图片,10 亿多个 Usenet(新闻组)信息。

Google Labs(Google 实验室)是 Google 正在进行的研发计划,有一些功能和服务仍处于研发和实验测试(Beta 版)阶段。Google 拥有一个开放性的数据库,内含 100 多个动态的未来需要实施的项目。它的业务领域甚至包括操作系统、生物和基因研究、太空相关计划等。

Google 在股票上市前,营利模式是强大搜索引擎技术支持下的广告收入和技术授权费。它没有“令人反感”的横幅或弹出式广告,而以文字格式的广告商“赞助性链接”单列出现,成为全球最大的网络广告提供者。它通过授权网站和企业使用其搜索技术获取授权费,用户包括 AOL(美国在线)、Amazon、思科等,约 80% 的互联网搜索是通过 Google 或使用

Google 技术完成的。

布林和佩奇担心上市面临被收购,特别是防止被 Microsoft 收购的风险,直到 2004 年才以荷兰式拍卖方式发行股票上市。它还以一种在 IT 公司中非常少见的两级股权的双重表决结构,使几位创始人拥有更高级别的投票权,以便尽可能保留决策自主权。

Google 于 2004 年 12 月宣布,将与英美五大图书馆合作,将馆藏图书扫描制作成电子版,加入 Google Print 中供读者网上阅读。协议中包含 1500 多万册书刊和文件,包括斯坦福大学和密西根大学图书馆的全部馆藏,牛津大学图书馆的 1901 年前出版的所有图书,哈佛大学和纽约公共图书馆的部分馆藏。处理成本约为每册 10 美元,预计整个项目至少要耗时 10 年。2006 年 8 月和 10 月,美国加州大学和威斯康星大学宣布加入此计划。

与它合作的还有 PubMed; CALIS 也正与它协商,将部分二次文献加入 Google。

它与图书馆员也有合作。Google Answers 是在线专家答疑服务,雇用的答疑人员都是经过严格筛选的专家,研究者超过 500 人,其中不少是图书馆员。

2 对 Google 的解读

在某种程度上,Google 与图书馆、数字图书馆有着理想主义的重合,又有着某些现实主义的优势。

2.1 理想主义的重合

Google 信奉的是一种反商业的理想主义经营哲学和“不要变得邪恶”的座右铭。其使命是整合全球范围的信息,使人人皆可访问并从中受益。它宣扬的十大价值观包括:“以用户为中心,其他一切纷至沓来。”“最好的方式是将一件事情做到极致。”^[4]

国际图联在《格拉斯哥图书馆、信息服务与知识自由宣言》声明:“不受限制地获取、传递信息是人类基本权利”,“图书馆和信息服务机构及工作人员支持并促进获取知识自由原则的落实,提供无障碍的信息检索服务。”^[5]

2.2 现实主义的优势

Google 的梦想是成为全球信息的有效组织者。它具有前所未有广泛的信息搜索范围、快速的搜索速度、较精确的搜索结果,产品和服务实现了质的飞跃,而且基本免费。

将 Google 现有功能与图书馆信息服务相比,无法否认的是 Google 在很多方面领先:技术领先,品牌强大,口碑良好。图书馆界常常自认为在检索广度和精度上,Google 无法与自己相比,但据实证检索分析,80% 以上检索案例表明,Google 中找不到的,图书馆

也找不到^[6]。而且由于 Google 已在用户群体中树立领导性的品牌地位,会导致人们的情性和习惯,即使有些服务深度弱于图书馆,用户也已习惯依赖 Google。OCLC《图书馆与信息资源的理解(2005)》报告显示:84% 的用户使用搜索引擎开始信息搜索,1% 的人从图书馆网页上开始信息的搜索^[7]。以下对比会进一步说明问题。

(1) Google Book Search(原名 Google Print)与图书馆 OPAC。

图书馆最普及的 OPAC,只能查询书目资料,包括文献的标题、作者、主题、关键词等。

搜索引擎的技术基础是全文检索技术。Google Book Search 按全文检索,检索到的图书能直接阅读目录摘要和部分章节;近期已开始以 PDF 格式提供公共领域版权过期的经典图书的完整版下载。Google 在中国的图书搜索计划已直接与清华大学、高等教育出版社等建立伙伴关系,在全球已对 20 多个国家开放,支持 7 种语言,有上万出版社加入此项目。

(2) Google Answers 与图书馆虚拟参考咨询服务(VRD)。

据 2002 年刚推出时估计,Google Answers 在一天半收到的咨询比美国所有图书馆一年参考咨询量的合计还多^[8]。

(3) Google Scholar 与图书馆引文检索。

Google Scholar 检索范围包括学术出版社、专业学会和大学以及互联网可获得的各种论文、图书、预印本、文摘和技术报告等。搜索结果按学术价值排名,特别是还列出了文章的被引次数,其中不仅有各种数据库中的引用文献,还有图书和各类联机出版社中的引用文献。

(4) Google 网页目录与图书馆分类目录。

Google 网页目录是利用 Netscape 公司所主持的大型公共网页目录 Dmoz 的分类架构来设计,由全世界各地的义务编辑人员审核挑选网页,通过内容开放和共享知识的理念实现网页分类目的。而图书馆对网络资源整合一直未找到一个好的解决办法。

(5) Google 个性化服务。

除了 Google Personalized HomePage(个性化主页, My Google) 和 Google Personalized Search(个性化搜索, My Search History), Google News 和 Google Alerts 已经是一种信息格点的雏形,这种信息格点将不同点上的内容“新闻”进行收集加工,按照阅读者的需要,以相对智能的方式推送到读者面前^[9]。

2.3 弊端

(1) 对文化的冲击。

Google 的垄断不同于 Microsoft 和 Intel,它们的垄断

集中在经济与市场层面,而 Google 的垄断形成后,其实质可能是集中在知识与信息层面,经济反而成为一个附加物,因为它实际上控制了互联网访问的自然资源,形成“数字霸权”。法国国家图书馆馆长让内依认为:“Google 数据库中英语文学的统治地位不但有可能对法语造成威胁,甚至会使其他欧洲的学术和传统边缘化。”

所以,Google 网络图书馆是给知识提供了更便捷的传播方式,还是商业对经典、学术和传统的围剿?

(2) 商业企业的风险。

Google 已不再单纯是搜索引擎,而且是商业企业,上市后将面临投资人压力、股市不确定性等问题,更严重的是法律纠纷。目前已有 Google 不少专利权败诉消息,而大多数服务几乎都存在着对个人隐私侵权的质疑,更严重的是版权问题。Google Print 从一开始就激怒了许多作者和出版社,如 Random House 和美国大学出版社协会等。拥有 8000 多名作家的美国最大出版作家组织——美国作家协会曾在美国纽约州法庭集体起诉 Google 在图书扫描数字化过程中存在“对版权法的系统性侵犯”。核心问题是 Google 在服务器中保留完整数字拷贝,并用于商业活动,这就违反版权法最基本原则^[10]。

Google 多元化的发展策略也存在巨大风险,它在 4 年内推出了 20 多种新产品,但据它自己估计,60% 到 80% 产品最终可能失败,这是过于关注自己的创意,但忽略了终端用户的体验。

3 图书馆面对 Google 的思考

Google 化已无法避免,但今天我们面对的不只是一个 Google,也许还有 Microsoft, Yahoo, 明天是阿里巴巴,后天是……实际上这是一个防不胜防,且并不以图书馆人意志所转移的趋势。图书馆唯有做好自己的服务,才能在信息空间中生存下去。Google 最值得图书馆学习的是它的理念和意识。

3.1 学习的领域

(1) 一切以用户为中心。

Google 主页始终非常干净,只有一个简洁朴素的关键词输入框,没有其他任何干扰。面对用户,把搜索引擎做得如此体贴,正是因为它最注重的是消费者的方便性、用户的友好性。它的 Orkut 和 Gmail 所采用的邀请方式,更以独特的用户体验式服务抓住了用户。

图书馆“需要停止从图书馆的角度来看待事情,而要关注用户的看法”^[11]。借鉴 Google 长期对用户行为进行深度研究和用户体验测试,图书馆要像 Google 一样,后台技术精妙无比,前台却是简明易用

的产品,而不是相反的产品^[12]。

(3) 无处不在的激励创新意识。

Google 代表的是一种自由开放的文化,有打破成规的思想起源,员工 20% 时间都由自己决定,可以让自己的创意来改变整个世界。

OCLC 也在网上开展程序设计大赛 (OCLC contest),公布一系列研究成果 xISBN, OAI 登记系统等,希望能够集思广益,为开拓新的服务模式征求意见。而 OCLC 开放元数据给 Yahoo 等搜索引擎后,WorldCat 的使用率大幅度攀升。

3.2 双赢和互利的领域

(1) 与 Google 合作。

Google 最新设立“Google 图书馆员中心”,就是认为图书馆员和 Google 共享一个目标,整合全球信息,并使所有人获益。Google 支持图书馆员进一步达到这一使命。

可将图书馆与 Google 想象中的“角力”变成一种合作式探索。实际上无论哪个图书馆也缺乏将所有馆藏图书数字化的经费。图书馆和出版社可以提供内容,Google 提供资金和技术,馆藏文献数字化后,Google 允许上网用户阅读所有无版权限制图书,而对那些受版权保护的图书,只能在网上阅读其目录及摘要。但 Google 可以与图书馆的 OPAC 相连,与 OCLC 的 WebCat 相连,指引读者转向图书馆实际馆藏。

纽约公共图书馆馆长保罗·勒克莱尔说:“这是将我们的资料提供给其他人的巨大机会,它可以解决一个古老的问题:如果人们无法找到我们,那我们应该如何找到他们呢?”

图书馆尚无能力与 Google 技术专家们抗衡,搜索速度和实用性等都存在问题,而且图书馆的开发研究缺乏共享,对用户需求也不如 Google 了解,所以合作才能共赢。

(2) 对企业风险的防范。

为了防范商业企业风险,图书馆可以要求 Google 签订协议,不仅由图书馆保留数字化文献的拷贝,而且一旦 Google 无法保存数字化文献并提供相应服务,图书馆有权优先获取这些“产品”和技术。

3.3 共存和超越的领域

Google 的“野心”并不是在于“组建数字图书馆”,它也许最终是要全方位改变世界,希望改变人们工作、通信、阅读、甚至看电视的方式,“Google 想出现在任何有人的地方”。

Google 注定要走的是大而全的路线。但图书馆、数字图书馆能够做到的是对信息内容的深度挖掘,精细的分类标引,构建专业领域的知识库体系等,这也正是图书馆超越 Google 的领域。用(下转第 99 页)

员会每三年都要向欧洲议会、欧洲理事会和欧盟经济与社会委员会提交一份该指令执行情况的报告,尤其是要对数据库特殊权利保护进行审查,以证明这种保护是否造成了垄断地位或其他干扰自由竞争情况的发生,研究是否有必要采取其他补救措施,包括对数据库使用实施强制许可。

4 对我国相关立法的启示

数据库特殊权利保护的确立受到一些发达国家尤其是信息产业发达国家的重视。美国虽然未能完成其立法过程,但一直在不断努力。我国是发展中国家,又是一个信息资源的生产与消费大国。与欧美发达国家相比,我国信息技术还比较落后,信息产业正处于爬坡阶段。目前情况下,我国贸然建立数据库特殊权利保护制度,条件确实不成熟。因为这将造成丰富的信息资源被少数发达国家或外国公司垄断的状况出现,制约我国信息产业的发展。但从国际大环境看,对数据库投资者给予更充分的权利保护是大势所趋。例如,《欧盟指令》只对欧盟成员国的国民或公司、企业制作的数据库提供保护,成员国以外的数据库只有在其本国对欧盟的数据库也给予特别保护的情况下才予以保护。

目前,我国在数据保护方面应完善现有的相关立法,跟踪国外尤其是西方发达国家的立法特点与潮流;在条件成熟时,考虑制订数据库保护的专门法律。

参考文献

- 1 许春明. 论数据库的版权保护. 法学杂志, 2002(4)
- 2 孙秋宁. 著作权法传统边缘的创新: 数据库特殊保护体系的建立与完善. 北京: 西苑出版社, 2002
- 3 翟方明, 徐兰. 数据库法律保护模式的比较研究. 市场论坛, 2005(12)
- 4 吴化碧. 数据库的知识产权保护. 图书馆理论与实践, 2005(2)
- 5 颜晓玉. 数据库的法律保护及思考. 图书情报知识, 2003(4)
- 6 刘峻. 论数据库的法律保护. 兰州学刊, 2003(6)
- 7 陈能华等. 论我国数据库生产业的发展战略. 湘潭大学社会科学学报, 2001(1)
- 8 秦珂. 试论数据库的特殊权利保护. 现代图书情报技术, 2001(6)

汤 曼 湘潭大学管理学院副教授. 通讯地址: 湖南湘潭大学. 邮编 411105. (来稿时间: 2006-11-21)

(上接第 90 页) 户既需要 Google 这样的信息平台, 也需要图书馆专业的信息辅助和配套的增值内容支持。数字图书馆还可以与教育和学习相结合, 重组服务内容和流程, 结合数字化学习, 形成新的知识中心, 无论如何, 信息提供或检索引擎都无法代替图书馆的知识传递。另外还有一个不可否认的超越之处是, 图书馆还是一处提供“庇护”的实体场所, 像教堂和其他神圣的地方一样, 可以供人们进行真实的沟通和心灵的交流^[13]。

参考文献

- 1 John Battelle. The Birth of Google. [2006-08-26]. <http://www.wired.com/wired/archive/13.08>
- 2 Google 公司信息: 技术. [2006-09-06]. <http://www.google.com/intl/zh-CN/corporate/tech.html>
- 3 周必水, 张延红. HillTop 算法剖析. 计算机时代, 2005(4)
- 4 我们的价值观. [2006-09-03]. <http://www.google.com/intl/zh-CN/corporate/tenthings.html>
- 5 The Glasgow Declaration on Libraries, Information Services and Intellectual Freedom. [2006-09-03]. <http://www.ifla.org/faife/policy/iflastat/gldeclar-e.html>

- 6 张甲. 斯坦福大学数字图书馆实例介绍. 2005 厦门数字图书馆前沿问题高级研讨班(2005-07-04)
- 7 图林中文译部. 图书馆与信息资源的理解: 给 oclc 成员的报告之结论部分. [2006-09-16]. <http://www.qiantu.org/liblog/?p=11>
- 8 Barbara Quint, Some Advice for Google Answers. Information Today, 2002(v. 19, i. 6)
- 9 再说机器新闻的分类和聚类. [2006-09-26]. <http://www.wespoke.com/archives/000932.html>
- 10 方恺. Google 的诉讼风暴也来了. [2006-09-26]. <http://www.enet.com.cn/article/2005/0924/A20050924456974.shtml>
- 11 OCLC. 审视发展环境认清前景趋势(清华大学图书馆中译本). 2005: 65
- 12 范并思. 信息技术冲击下的图书馆人文思潮. 桂林: 中国图书馆学会 2005 年会, 2005
- 13 12 Ways Libraries Are Good for the Country. [2006-10-16]. <http://www.ala.org/ala/online/selectedarticles/12wayslibraries.htm>

刘 晨 杭州师范学院图书馆副研究馆员, 北京大学管理学硕士. 通讯地址: 杭州下沙高教园区东区学林街 16 号. 邮编 310036. (来稿时间: 2007-01-12)