

孙 巍^{1,2}, 张学福³

(1. 中国科学院 国家科学图书馆, 北京 100190; 2. 中国科学院 研究生院, 北京 100049; 3. 黑龙江大学 信息管理学院, 黑龙江 哈尔滨 150080)

基于引文的信息检索可视化相关系统比较分析

摘 要: 分析了基于引文的信息检索系统、可视化系统及检索可视化系统三类系统的评价标准, 从 14 个方面对 6 个典型系统进行深入地比较分析, 找出现有的基于引文的信息检索可视化系统存在的一些不足, 为国内从事相关研究的同行提供参考。

关键词: 引文; 信息检索; 可视化; 比较分析

Abstract: The evaluation standard of the 3 citation-based systems, that is, the information retrieval system, the visualization system and the information retrieval visualization system, is analyzed. The 6 typical systems are compared in depth from 14 aspects. Finally, the paper summarizes the deficiencies of the present citation-based information retrieval visualization system in order to provide some help for relevant researchers in China.

Keywords: citation; information retrieval; visualization; comparative analysis

1 目标对象的选取

笔者从功能的角度将现有基于引文的相关系统分为 3 类: 基于引文的信息检索系统、基于引文的信息可视化系统及基于引文的信息检索可视化系统。在选取目标对象时主要考虑两点: 原型系统及应用系统兼顾, 上述三种类型的系统兼顾。因此, 笔者选取了以下三类 6 个典型系统。

1) 基于引文的信息检索系统。Web of Science^[1]是在尤金·加菲尔德的指导下推出的功能较强大的基于引文的信息检索系统。由科学引文索引、社会科学引文索引、艺术与人文科学引文索引、Index Chemicus 以及 Current Chemical Reactions 共 5 个数据库组成。共收录了 9 000 多种世界权威的、高影响力的学术期刊, 内容涵盖自然科学、工程技术、生物医学、社会科学、艺术与人文等领域。

2) 基于引文的信息可视化系统。CiteSpace^[2]是 2003 年由 Chen Chaomei 开发的用于分析和可视化文献资源关系的原型系统。系统主要的数据来源是 Web of Science, 系统的首要目标就是利用可视化显示技术在知识域中帮助用户进行突发趋势的分析。HistCite^[3]是 2003 年由 Garfield 开发的引文历史可视化分析工具, 它通过生成文献编年图表来实现领域分析功能。

3) 基于引文的信息检索可视化系统。美国的爵士大学自 1998 年以来, 相继开发了基于作者共引分析 (ACA) 理论的 AuthoLink、基于词共现理论的 ConcepLink、集成多种类型的索引和多种视图的 PNASLink 三种基于引文的

信息检索可视化系统。AuthoLink、ConcepLink、PNASLink 让普通用户通过 Web 浏览器就可以完成实时可视化检索, 尤其是 PNASLink 提供了许多检索功能, 可视化的图形包括作者、关键词、杂志或它们的组合^[4]。

2 系统评价标准

显而易见, 上述三类系统在功能、结构、技术等多方面存在交叉性, 因此, 要通过对各类典型系统进行全面、系统的比较分析, 得出现有的基于引文的信息检索可视化系统存在的问题, 需综合考虑各类系统的评价标准。

三类系统的共性在于以引文理论为基础, 因此, 系统的理论基础应作为首要的、基础的评价依据。对基于引文的检索系统的评价除了依据系统的理论基础外, 还应依据传统的检索系统的评价标准^[5], 即数据源、信息检索、结果显示、界面、资源导航功能、特色服务等。

基于引文的信息可视化系统的主要特征是运用可视化技术进行知识域分析, 因此对该类系统的评价除了依据可视化理论基础外, 还应依据信息可视化系统的评价标准, 即分析过程中系统的交互性、可视化的显示形式、可视化的映射方法、用户界面等。

从系统的功能角度讲, 对基于引文的信息检索可视化系统的评价除了要依据理论基础外, 还应依据基于引文的检索系统及可视化系统的评价标准。另外, 对检索接口的评价也很关键。White 和 McCain 提出对可视化信息检索接口进行评估的 6 项标准: 其显示是否对一个简单的列表有

所改进,它是否提供新的性能,它是否能快速被理解,它是否有助于实时化(或能在一个可接受的时间范围内),它对一个重要的文献集是否有限制,它是否可向上升级适应更大容量的文献集。

3 6个典型系统比较分析

基于上述对系统评价标准分析,笔者主要从14个方面对以上6个典型系统进行比较分析,详细比较情况见表1。

1) 引文基础理论。6个系统的引文理论运用深度不同,Web of Science运用了引文年代分析、引文量分析等多种引文分析理论,可以轻松地回溯某一研究文献的起源与历史,或把握某一学科的发展趋势,或者追踪其最新的进展;CiteSpace, HistCite信息可视化系统不但运用引文分析方法,还运用了引文网络理论, CiteSpace主要是对突发事件的分析,而 HistCite主要是进行域分析,分析功能较强; AuthoLink, ConcepLink, PNASLink信息检索可视化系统,只是通过共引网络及共引分析理论来揭示概念词、作者、期刊等之间的关系,分析功能相对较弱。

2) 数据源。数据源标准是针对具有检索功能的系统设定的。Web of Science应用系统,数据库的规模较大、范围较广,收录的数据权威、质量高,数据源可为动态数据; AuthoLink, ConcepLink, PNASLink检索功能较弱,它们的数据源的规模相对较小,范围也存在一定的局限性。

3) 数据的可扩展性。具有检索功能的系统,其数据应具有一定的可扩展性。由表1可知,Web of Science应用系统的数据可扩展性较强, AuthoLink, ConcepLink, PNASLink向上升级适应更大容量的文献集能力较差。

4) 信息组织管理。信息索引方法。Web of Science, AuthoLink, ConcepLink, PNASLink都具有自动、手动索引、标引功能,而三者最快的检索方式是通过在数据库中事先存储好的检索结果的索引实现的; HistCite的索引只是为统计分析功能而设。信息更新周期。具有检索功能的典型系统的数据要定期更新。Web of Science的信息更新快,而 AuthoLink, ConcepLink, PNASLink的信息更新周期要依检索接口的数据库数据更新的周期而定。信息的组织方式。所有的典型系统都是以引文中特征词索引的方式组织信息的,只是索引的类别及规模有所差别,Web of Science的索引规模较大。

5) 信息检索。信息的检索方法。Web of Science提供了快速检索、一般检索、化学结构检索、高级检索等多种检索方法,而 AuthoLink, ConcepLink, PNASLink只具有一般的检索功能,检索功能不够全面、系统。信息的检索方式。所有的典型系统都是以提问式检索为主,辅之以联机检索。Web of Science还提供了网络在线检索。

信息的检索途径。Web of Science检索途径较多,以文献名为主,辅之以文摘、著者、日期等各种途径,提供了多种全文获取途径; AuthoLink, ConcepLink, PNASLink强调检索结果中特征词间的内部关系的展现,辅助用户检索,但检索途径相对较少,只是提供了诸如作者检索、概念检索或期刊检索等,未提供全文检索。

6) 检索结果显示形式。Web of Science以有序列表的形式显示检索结果,形式最清晰、简洁; AuthoLink, ConcepLink, PNASLink只对前 n 个检索结果按照被引频率降序排列,以列表的形式显示,显示信息不够全面。

7) 界面显示形式及帮助功能。由表1可知,原型系统界面提供的帮助功能相对较少,即使提供也不够详细,界面设计较空泛;而 Web of Science应用系统提供了较详细的帮助功能,界面清晰,内容组合的逻辑性强。

8) 资源导航功能。6个系统中,只有 Web of Science具有资源导航功能,依据不同的数据源对数据进行分类,用户可按照自己的需求选择所需信息。

9) 可视化显示形式。CiteSpace采用了空间聚类法和时间区域法两种显示技术,从中很容易看出具有突出贡献的节点,发现研究前沿与其智能基础之间的关系,并通过领域专家的参与最终认定分析的结果,可视化显示形式较丰富,但图形的清晰直观性还有待提高; HistCite则以时序网络以及图表的形式显示作者、期刊、主题词等之间的内部关联,图形、表格较清晰; AuthoLink, ConcepLink, PNASLink仅运用了共引网络的形式对引文中的特征词关系进行可视化显示,但显示的图形不够清晰、直观,可视化显示的形式单调,清晰直观度受节点个数限制。

10) 可视化映射技术。由表1可知,具有可视化功能的典型系统都采用 SOM 及 PFNET可视化映射技术。在 PFNET映射中,都通过设定不同的参数对图形显示的粒度、变化形式等进行控制,相对而言, CiteSpace的控制机制较全面、完善; SOM映射技术提供了一个很好的浏览,强调领域、区域和它们的临近关系,很容易在图形的核心区域增加标签,但其在映射的降维过程中,很容易使图形显现的内容与实际脱节。

11) 降低用户认知负担。AuthoLink, ConcepLink, PNASLink通过用户与可视化图形的交互来降低认知负担,辅助检索; CiteSpace, HistCite通过用户调整可视化图形参数来修正、更新可视化图形,系统依此得出不同的分析结果,辅助用户分析,降低认知负担; Web of Science对检索结果分析、以列表的形式显示,辅助用户检索。

12) 系统实时交互性。Web of Science具有实时检索及交互功能,用户可以在检索结果的启发下,重构检索式,进行更深层次的实时检索; AuthoLink, ConcepLink,

表 1 6个典型系统比较分析表

	基于引文的检索系统		基于引文的可视化系统		基于引文的信息检索可视化系统		
	Web of Science	CiteSpace	HisCite	AuthoLink	ConceptLink	PNASLink	
引文理论基础	运用了多种引文分析理论	各类引文分析及引文网络理论	各类引文分析及引文网络理论	共引分析、共引网络	共引分析、共引网络	共引分析、共引网络	
数据源 (数据库规模、范围及质量控制)	9000多种学术期刊, 内容涵盖广, SCI, SSC, A&HCI, CCR, IC共5个数据库。根据影响选刊, 选择权威、高影响力的期刊	Web of Science及PubMed搜索引擎。没有对数据源进行质量控制	Web of Knowledge。没有对数据源进行质量控制	ISI引文数据库, Diabg。没有对数据源进行质量控制	PubMed搜索引擎。没有对数据源进行质量控制	PNAS资源、Diabg。没有对数据源进行质量控制	
数据的可扩展性及对重要文献的限制	扩展性好, 对重要文献集有限制			扩展性较好	扩展性较好	扩展性较好	
信息管理	信息索引方法	自动索引、自动标引、手动标引	为分析建立的索引, 自动索引	自动索引、事先存储在数据库中的标引	自动索引、事先存储在数据库中的标引	自动索引、事先存储在数据库中的标引	
	信息更新周期	数据每周更新、更新快		依据检索接口连接的数据源所定	依据检索接口连接的数据源所定	依据检索接口连接的数据源所定	
	信息组织方式	作者、来源刊、主题等索引	作者、作者国别、来源文献等索引	作者索引	关键词索引	作者、期刊、关键词索引	
信息检索	信息检索方法	快速检索、一般检索、化学结构检索、高级检索等		简单检索	简单检索	简单检索	
	信息检索方式	以提问检索为主, 辅以在线检索、联机检索		以提问检索为主, 辅以联机检索	以提问检索为主, 辅以联机检索	以提问检索为主, 辅以联机检索	
	信息检索途径	以文献名为主, 辅以文摘、全文、著者、日期、信息类型等检索		作者检索、主题检索	概念词检索	作者、期刊检索、热点主题、“NAS成员”或“引用最多的作者”检索	
检索结果显示形式	清晰简洁的检索结果列表, 辅之以各种排序			按被引频率有序排列	按被引频率有序排列	按被引频率有序排列	
界面显示形式及帮助功能	界面清晰, 直观, 帮助功能详细	界面较清晰, 提供简单的帮助说明	界面较清晰, 无帮助功能	界面较清晰, 无帮助功能	界面较清晰, 无帮助功能	界面较清晰, 无帮助功能	
资源导航功能	按不同的数据源进行分类						
可视化显示技术		空间聚类显示技术、时序区域法显示技术	时序网络显示	共引网络显示	共引网络显示	共引网络显示	
可视化映射技术		PNET映射技术	PNET映射技术	SOM, PNET映射技术	SOM, PNET映射技术	SOM, PNET映射技术	
降低用户认知负担	对检索结果分析、以列表的形式显示	用户通过调整参数, 修正、更新可视化图形, 辅助用户进行统计分析	用户通过调整可视化图形参数, 调整更新可视化图形, 得出不同的分析结果, 以此降低用户认知负担	通过用户与可视化图形的交互降低用户认知负担, 辅助检索	通过用户与可视化图形的交互降低用户认知负担, 辅助检索	通过用户与可视化图形的交互降低用户认知负担, 辅助检索	
系统实时化交互性	提供实时检索, 实时性强	用户与可视化界面的交互性强	用户通过调整可视化图形参数, 调整更新可视化图形, 实现与系统交互, 直到得出满意的分析结果	系统通过自动生成或手工生成(存储在系统中)可视化图形与用户交互, 实时性较差	系统通过自动生成或手工生成(存储在系统中)可视化图形与用户交互, 实时性较差	系统通过自动生成或手工生成(存储在系统中)可视化图形与用户交互, 实时性较差	
个性化服务	我的检索、我的期刊列表等服务						
特色服务	引文跟踪服务、存盘下载、支持OpenURL服务、建立MyLibrary等	统计分析, 将分析结果存盘等	导出分析结果				

PNASLink, Citespace, HiteCite都可以通过可视化技术实时地生成可视化图形, 具有实时可视化及分析功能, 但AuthoLink, ConceptLink, PNASLink实时检索功能较弱, 检索结果是通过事先检索好的数据库直接生成的。

13) 个性化服务。由表1可知, 只有Web of Science应用系统提供了个性化服务, 如我的期刊列表, 管理自己的检索历史、定题Email Alerting服务等功能, 其他系统未提供个性化服务。

14) 特色服务。Web of Science提供了引文跟踪服务、存盘下载、支持OpenURL服务、建立MyLibrary等特色服务; Citespace可以将统计分析结果存盘; HiteCite实现了统计分析结果导出功能。其他系统未提供特色服务。

4 结束语

通过对6个典型系统的比较分析, 不难看出, 现有的基于引文的信息检索可视化系统在某些方面具有一定的优势, 如系统通过用户与可视化图形的交互, 在某种程度上降低了用户的认知负担, 同时系统还具有一定的向上升级适应更大容量的文献集的能力。但系统在以下几方面仍存在不足, 有待进一步完善: 系统的引文理论基础不够完备。系统的数据源的规模较小, 数据范围存在一定的限制, 没有对数据质量加以控制。系统的自动索引及自动

标引功能有待提高, 检索方法相对较单一。系统的检索功能不够全面, 需要进一步扩充检索途径, 提高检索结果显示的清晰度及信息组合的逻辑性。由于目前还处于原型系统研究阶段, 系统的界面设计不够人性化, 帮助功能不够完善。系统不具备资源导航功能, 缺乏对数据源的推荐与评价。

除此之外, 系统缺少个性化服务及特色服务, 因此, 在服务方面还应加强与其他服务方式的集成。

参考文献

- [1] 尤金·加菲尔德. 引文索引法的理论及应用 [M]. 侯汉清, 等译. 北京: 北京图书馆出版社, 2004: 243-246
- [2] Chen Chaomei. CiteSpace: Visualizing Patterns and Trends in Scientific Literature [EB/OL]. <http://cluster.cis.drexel.edu/~cchen/citespace/>
- [3] Garfield E, Paris SW, Stock W G. A software tool for informetric analysis of citation linkage [J]. Information, 2006 (8)
- [4] 林夏. 信息可视化与内容描述 [J]. 现代图书情报技术, 2004 (10): 3-13
- [5] 朱震远. Internet专业导航系统和传统信息检索系统评价标准的比较 [J]. 现代情报, 2002 (9): 59-61

作者简介: 孙巍, 女, 1978年生, 博士生。

张学福, 男, 1966年生, 博士, 教授。

收稿日期: 2008-02-03

(上接第520页)

适当方法, 最后检查结果的有效性。有时馆长必须在冲突各方之间担任仲裁的角色, 仲裁者必须秉持公正的态度, 依照规则来做出裁判, 以免有偏袒的行为发生。

2) 投诉机制。投诉个案的处理, 往往就是服务整改的开始。要建立应诉常规机制, 让图书馆上下、内外形成一个有效的工作整体, 共同承担起完善读者服务的责任。设立受理投诉的专门机构, 同时根据案例做出分析、研究, 对读者意见较集中的问题及时通报相关业务和管理部门, 提供其作为整改的依据。建立读者与管理层便捷沟通的渠道, 设立意见箱、公布投诉电话, 在读者中聘请监督员, 以便随时听取读者意见, 并转达到相关部门; 同时及时反馈读者, 说明或解释情况。

3) 反馈机制^[10]。读者作为服务的对象, 在评价上最具有发言权, 读者评价所占的权重应当较大。读者反馈功能可以使图书馆从用户的角度认识自己的不足, 不断改善自己。设立专门机构和管理人员, 对读者的反馈意见进行收集和分析整理, 组织相关人员对问题进行分析, 寻求解决的方法, 有效排除影响服务质量的障碍, 以进一步改进服务质量; 建立一个服务跟踪机制, 以便对馆员问题解答的全过程进行监控并且可及时向馆员提供反馈信息, 提醒

其工作中的不足, 以保证图书馆提供高质量服务。

参考文献

- [1] 冲突管理 [EB/OL]. [2008-01-02]. http://www.52mba.com/sjhh_news_view.asp?d_id=1457&sort1=29&sort2=66
- [2] 刘桔. 读者服务中的馆读冲突及心理学研究 [J]. 图书情报知识, 2007 (4): 47-50
- [3] 袁红军. 关于高校图书馆读者服务中情绪服务的思考 [J]. 图书馆界, 2004 (2): 21-24
- [4] 周振林. 论管理冲突的三种类型 [J]. 长白学刊, 2005 (4): 13-15
- [5] 陈玉荣. 现代冲突理论与图书馆管理 [J]. 情报理论与实践, 2007 (3): 357-359
- [6] 王庆瑶, 李姝丽, 孙继亮. 高校图书馆读者维权意识的觉醒及读者工作对策 [J]. 大学图书馆学报, 2007 (2): 64-66
- [7] 王世伟. 服务是图书馆存在的理由 [J]. 图书馆论坛, 2002 (5): 42-43, 39
- [8] 袁红军. 对图书馆角色互换的再认识 [J]. 长春师范学院学报: 人文社会科学版, 2005 (5): 173-175
- [9] 王世伟. 论“双不”方针与图书馆的服务与管理 [J]. 图书馆杂志, 1997 (7): 1-3
- [10] 袁红军. 图书馆读者服务中读者意见处理机制分析 [J]. 图书馆建设, 2007 (3): 89-90, 92

作者简介: 袁红军, 男, 1970年生, 硕士, 副研究馆员。

收稿日期: 2008-02-22