

网上三大中文期刊数据库引文功能之比较

□黄春燕

摘要 针对目前互联网上3种常用的中文期刊数据库:中国期刊全文数据库、维普中文科技期刊数据库、万方数字化期刊子系统,从收录范围、检索途径、检索功能、检索结果的输出处理及检索效果等方面对三者的引文检索功能进行了比较分析,并指出它们各自的优势与不足。

关键词 中文期刊 数据库 互联网 引文检索

引文分析作为一种应用广泛的文献计量学研究方法,早在20世纪20年代就已出现,并随着1961年《科学引文索引》(SCI)的诞生而得到普及。它以科学文献正文后附的参考引用文献作为统计对象,通过对大量引用文献的统计学处理来滤出“噪音”^[1],揭示科学期刊、学科、单位与论文著者等各种分析对象的数量特征及内在规律性。目前引文统计与分析被应用于职称评审、成果申报、机构评估、项目考核等众多领域,成为人们日益关注的一项科研活动。国内许多大型期刊网站都开始重视引文检索功能的开发,网上三大期刊数据库——中国期刊网全文数据库、维普中文科技期刊数据库、万方数字化期刊子系统——目前均已提供引文检索功能,本文就这三大期刊数据库的引文检索特点作一比较分析。

1 三个数据库的简介及收录范围比较

中国期刊网全文数据库(<http://www.cnki.net>,以下简称清华库)是由清华同方光盘股份有限公司在《中国学术期刊(光盘版)》的基础上,发展起来的一个国内最大型的综合信息资源网,已成为广大读者在网上阅读国内学术期刊的重要途径。除了全文数据库的一般功能外,引文检索功能也是其特色之一。该库收录期刊的数量庞大,专业齐全,是国内目前唯一可进行所有学科专业引文检索的综合性数据库。

维普中文科技期刊数据库(<http://www.tydata.com>,以下简称维普库)由中国科学院西南信息中心重庆维普资讯公司研制出版,目前该公司在天元网上提供了信息资源检索服务。《中文科技期刊数据库》(引文版)是该中心近年来开发的新产品,它以《中文科技期刊数据库》(文摘版)为依托,收录1990年以来公开出版的科技类期刊5000多种,是我国目

前唯一正式使用的可免费检索的网上引文数据库。用户在天元网上进行注册获得用户名和密码后可免费登录访问。

由中国科技信息研究所集成的万方数字期刊网以两种方式提供引文检索服务:其一是科技信息子系统下的《中国科技论文引文分析数据库》(该数据库已有较多论文讨论,在此不再赘述);其二是数字化期刊子系统的引文查询功能。万方数字化期刊子系统(<http://periodicals.wanfangdata.com.cn>,以下简称万方库)是万方数据资源系统2001年改版后推出的全文检索系统,该系统集纳了医药卫生、工业技术、农业科学、基础科学、社会科学、经济财政、教科文、哲学政法等8大类的100多个类目的3000多种期刊的全文内容。该系统在原有的引文查询功能基础上,从2003年初起推出功能更为强大的期刊论文的引文关联检索服务模式,在网上提供免费检索服务。

三个数据库收录范围比较见表1。

表1 数据库收录范围比较

	清华库 ^[2]	维普库 ^[3]	万方库 ^[4]
学科范围	综合	自然科学、工程技术、农业科学、医药卫生、经济管理、教育科学、图书情报	理、工、农、医、人文
收录时间	1994-	1990-	2001-
来源期刊	6000多种	5000多种	3039种
引文数量	790多万篇	140多万篇	490多万篇
更新周期	日更新	年更新	季度更新

注:以上数据库统计截止至2003年8月15日。

从表1可以看出:

(1)清华库收录的学科范围最广,专业门类齐全,涵盖了自然科学、工程技术、农业、医药卫生、文史哲、经济政治、法律、教育、社会科学等各个方面,特别是在考古、哲学、宗教、心理、政治、党建、外交、军事、法律等社会科学方面的内容比其他两个数据

库多得多。万方库收刊偏重于科技领域,人文科学方面则以学报为主,期刊来源以中华医学会系列杂志、大学学报、中国科学系列和科学普及期刊为主,并从2001年底开始,收录了我国所有科技核心期刊。维普库收录的内容主要有自然科学、工程技术、农业科学、医药卫生、经济管理、教育科学、图书情报等方面。对地方性、行业性期刊收录较其他两个数据库全。

(2)清华库收录了1994年以后的数据,维普库收录期刊的年限最长,收录了1990年以来的数据,比较适合于进行回溯性检索。万方库自建立以来,发展速度虽比较快,上网期刊数量也从1998年10月的120种增加到现在的近3000种,但是与其他两个库收录的文献数量相比,它目前提供的资源还不够充分。因此若想查找较早的引文信息,可选择维普库,但对上个世纪90年代以前的引文数据,暂无网上数据库可查。

(3)清华库不仅收录期刊种类多,且数据量大,有利于统计不同学科之间的互引量,研究交叉学科的内部联系。另外,清华库为日更新,时效性强,并且引文与现刊同期出版,而其他两个数据库更新周期较长,因此用户在查检最新的引文数据时,可考虑选择该库。

2 检索途径比较

三个数据库均提供了多个检索途径,见表2。

表2 数据库检索途径比较

	清华库		维普库		万方库	
	源文献	被引文献	源文献	被引文献	源文献	被引文献
关键词	√		√			
刊名	√		√			√
作者	√	√	√	√		√
第一作者			√			
机构	√	√	√			√
题名	√	√	√	√	√	√
文摘	√		√			
分类号			√			
全文	√					
基金	√					
出版年	√	√				

从表2可以看出:

(1)维普库的检索系统平台提供了源文献模式和被引文献模式两种检索界面,两种界面可通过切换键方便切换。源文献模式提供了关键词、刊名、作者、机构、题名、文摘和分类号等检索入口,可查检

索到的文章引用的所有文章;被引文献模式有被引作者、被引题名、被引刊名三个检索入口,可从检索到的文章查引用了该文章的所有文章。

(2)万方库的期刊引文检索界面提供了论文题目、参考文献题目以及引文数据库三种检索入口,前两种检索入口可分别查找到论文或参考文献的题目中包含有特定检索词的文章,引文数据库模式的引文功能最为灵活,可查找出含有特定内容(被引作者、被引刊名、被引篇名、被引机构、出版年、卷、期、页码等)的文章,并可进一步得到这些文章的引用(或被引用)信息。

(3)清华库的论文查询共有13个检索入口,引文只是其中的一个检索点,但其功能相当强大。由于其引文具有全文检索的功能,因此在文献引文条目中出现的任何信息都可作为检索点,满足对参考文献的作者、题名、刊名、机构和出版年等检索项目的引文检索需求。

(4)三个数据库均能从多个作者入手查找被引文献,清华库和万方库可以查到前三个著者的被引文献,维普库则只能查到前两个著者的被引文献。

(5)当被引文献有外文论文时,维普库只收中文引文;万方库则中外引文同时收录,检索时只需输入检索词的中文名称,就可以把符合检索条件的中英文引文全部检索出来;清华库则应选中检索界面上的“中英扩展”复选框,就能保证中英文引文的检索,但因著者的拼写变化多样,最好分别采用作者姓名的不同拼写形式进行检索,这样才能保证没有遗漏。

(6)三个数据库既能查到论文引用期刊、图书、学位论文及会议文献的情况,也可以通过被引刊名字段查找到以上文献类型的被引情况,但万方库查找图书、学位论文以及会议文献的被引结果欠准确。

3 检索功能

检索功能是评价一个数据库的重要指标,检索功能是否完善,直接关系到引文的查全率和查准率。三个数据库的引文检索功能对比见表3。

从表3可以看出:

(1)万方库的检索功能较弱,每次只能在一个字段内输入一个检索词进行检索;清华库和维普库由于支持逻辑检索,因此都可以实现复杂条件的组

表3 数据库的引文检索功能比较

	清华库	维普库	万方库
检索方式	单项检索、组合检索	单项检索、组合检索	单项检索
逻辑检索	支持	支持	不支持
二次检索	支持	支持	支持
全文检索	支持,即引文条目录中出现的所有信息都可作为检索点	不支持,但其“任意字段”检索入口默认为引文条目录中“篇名”、“刊名”和“作者”任意一个都可作为检索点	支持,即引文条目录中出现的所有信息都可作为检索点
检索历史	系统自动将历次检索词(式)保留在历史列表框中,双击某一检索词(式),即可进行再次检索	不保留	不保留
限定检索	年代限定	源文献模式下可进行检索年代限定	无

合检索。维普库的逻辑检索最为灵活,可以直接在检索词输入框内输入用“*”、“*+”、“-”连接起来的布尔逻辑运算式,若检索词不在同一字段内,则需要先在检索词前注明其字段符号;清华库的逻辑检索是在高级检索界面下实现,字段间或字段内的不同检索词之间可进行“*”、“*+”、“-”任意组配操作。

(2)三个数据库均支持二次检索,用户可根据需要逐次接近检索目标;清华库和万方库还支持全文检索,维普库不支持全文检索,但可进行“任意字段”检索;清华库和维普库可进行年代限定检索,但维普库的年代限定检索需在源文献模式下;三者均不具备检索策略和检索历史保留功能,但清华库可保留检索词。

(3)清华库由于引文检索与其他字段检索处于同一检索界面下,因此可通过组合引文检索直接获得作者、机构和刊物的自引或他引情况;另外两个数据库则不具备以上功能,需通过对检索结果的一一甄别、排除后才能获得自引或他引的情况。

4 检索结果的输出处理比较

从表4可以看出:

(1)三个引文数据库中,维普库对检索结果的处理最为科学合理,用户检索起来概念明确,检索结果以大多数用户习惯的思维方式显示,用户可直接获得简明直观的引文统计报告;万方库则将论文和引文作为一个关联体系,在获得检索结果后还可以分别追溯论文或引文各自的引用与被引信息,有利于提示学科和同行间的引用与被引用、继承与被继承关系;而清华库的引文功能毕竟只是全文检索中的一个字段检索功能,检索结果以来源文献形式显示,初学者不易理解,并且对检索结果的浏览筛

选过于繁琐。

(2)清华库的引文检索结果可按来源文献的相关度或更新日期排序,维普库和万方库的检索结果排序无明显规则。

(3)清华库提供了在线帮助,维普库则提供了动画演示教程,而万方库在网上未提供任何的联机帮助或学习内容。

5 检索效果比较

为了更直观地比较3个引文数据库的检索效果,笔者从作者、期刊及篇名三个方面进行试检索,检索结果如表5所示。

从表5可以看出,利用清华库检索“周厚高”、《广西农业生物科学》及“器官移植”1999年的发文量及在2000~2002年的被引情况,查全率明显高于其他两个数据库。在相同条件下,三者的检索效果依次为清华库>万方库>维普库,原因是由于三者在

表4 数据库的检索结果输出处理比较

	清华库	维普库	万方库
检索结果	系统只能根据检索条件检索出来源文献篇数及其全文,据其后的参考文献能浏览引文情况	系统根据检索条数及被引作者、题名、(书)名及出版信息,单击相应题名,可链接到其引用的详细情况	系统根据检索条件自动记录并逐条列出,对引文题目,点击进一步显示该文献及被引信息
检索结果处理	在文末参考文献处找到被引文的条目,粘贴保存	检索结果可按记录或全部记录输出,形成引文统计报告	检索结果只能按显示页面,直接保存,或为html格式
检索结果排序方式	可按来源文献的相关度或更新日期排序	无规则	无规则
在线帮助或联机学习	在线帮助	Flash动画演示教程及培训教室	无

表5 数据库的引文检索效果比较

	作者		刊名		篇名		
	周厚高	广西农业生物科学	器官移植	1999	2000-2002	1999	2000-2002
	发文量	被引频次	发文量	被引频次	发文量	被引频次	被引频次
清华库	15	8	94	50	58	481	
维普库	3	5	45	16	11	24	
万方库	4	5	75	26	13	170	

收录合著者、语种及收刊范围上的差异所致。

6 结论

通过对以上三个数据库的引文比较及笔者的检索实践,得出以下结论:

(1)三者作为国内大型的网上中文期刊数据库,均能提供引文检索功能,并且其引文检索功能各有特点,在收录的学科专业范围及年限上各不相同,互相不能替代,用户可根据实际需要选择使用。维普中文科技期刊数据库(引文版)和万方数字化期刊的引文数据库在网上提供免费的检索服务,而中国期刊全文数据库已为国内许多高校和科研单位提供镜像或包库服务,因此用户也可方便地使用其引文检索功能。

(2)维普中文期刊数据库引文收录的时间最长,适合于进行引文回溯检索。但该库在收录引文合著者及文种等方面的局限性,大大影响了引文检索的效果,建议该库今后在这些方面加强和完善。

(3)万方数字化期刊虽然推出时间不长,但发展势头强劲。由于该库收录了科技领域的全部核心期刊和大学学报,因此更适合于决策者和科研管理者进行引文分析。同时该库将论文与引文作为一个关联体系,类似于SCI的编制体系,因此它完全适合未来引文数据库的发展方向。

(4)中国期刊全文数据库无论是在收录的数据量、学科范围、检索功能或检索效果等方面,在三者中均占有优势,因此它能满足不同专业、不同类型用户的检索需要。到目前为止,它仍然是网上检索国内人文社科类引文信息的主要检索工具。该库引文检索的时效性最强,数据更新最及时,因此用户可用它来检索最新的引文信息。但由于中国期刊全文数据库毕竟不是一个专门的引文数据库,因此它在许多方面尚不能完全满足引文检索的需要,尤其在检索结果的处理上有严重不足。中国期刊网有着庞大的资源和技术优势,完全有能力筹建国内第一个综合性引文网络数据库,有消息说中国期刊网正在开展这方面的工作,我们期待早日与其在网上见面。

参考文献

- 1 缪建兰,陈永忠.《中国科学》A辑(中文版)引文分析.情报科学,2003,21(1):75-76
- 2 <http://www.cnki.net>
- 3 <http://www.tydata.com>
- 4 <http://periodicals.wanfangdata.com.cn>

作者单位:广西大学图书馆,南宁,530005

收稿日期:2003年8月9日

A Comparative Study on the Citation Search Function of Three Online Chinese Periodical Databases Huang Chunyan

Abstract: By comparing three online Chinese periodical databases, this paper analyzes the differences and similarities of citation search function in source data scope, search methods and search results. It furthermore summarizes the advantages and disadvantages of each database.

Keywords: Chinese Periodicals, Online Database, Citation Search

(上接第65页)

- 4 俞君立. 中国文献分类法百年发展与展望. 武汉大学出版社, 2002.12
- 5 赵英莉. 信息网络环境中情报检索语言兼容问题研究. 四川图

书馆学报,1998(4)

作者单位:国防科技大学图书馆,长沙,410073

收稿日期:2003年9月30日

On the Advantages of Nature Language and the Fading Away of the Artificial Language Wang Qun, Jing Qing

Abstract: This paper discusses respectively the current status, as well as advantages and disadvantages, of nature language and artificial language. It concludes that artificial language will be replaced by nature language in the application of information retrieval.

Keywords: Information Retrieval, Nature Language, Artificial Language