

楼。

参考文献:

- [1] 马红月,牛晓芳,李 萍.网络环境下图书馆的流通服务工作[J].中华医学图书情报杂志,2003,12(4):40-41.
[2] 林兴国.无线网络在大学图书馆网络建设中的应用设计[J].四川图书馆学报,2004,(2):55-58.

- [3] 郑振容.无线网络对图书馆网络布线设计思想的影响[J].现代情报,2004,(9):75-76.
[4] 陈 鹏,赵春华.无线网络技术及其应用模式分析[J].中国有线电视,2003,(16):23-26.
[5] 杨嵩山,秦春燕.无限局域网技术及应用[J].邮电设计技术,2002,(4):5-9.

[本文编辑:杜海洲]

常用中文文献数据库引文检索功能的比较研究

王 飞¹,侯跃芳²,崔 雷²

(1.中国医科大学附属第二医院信息科,辽宁 沈阳 110004,2.中国医科大学信息系,辽宁 沈阳 110001)

关键词:中文文献数据库;引文检索;查全率

中图分类号:G252.7

文献标识码:B

文章编号:1671-3982(2005)04-0052-05

1 前言

近年来,我国自行开发建设的光盘和网上数据库获得了很大的发展^[1]。目前,数据库强大的引文检索功能已成为科技人员及信息工作者获取信息的重要途径^[2]。对生物医学工作者来讲,较常用的具有引文检索功能的中文文献数据库有中国科学引文数据库(CSCD)、维普中文科技期刊数据库(引文版)(简称维普)、中国科技论文引文统计分析库(简称万方)和中国期刊全文数据库(简称CJFD)。

这些数据库是进行中文引文统计分析的主要情报源,是科研人员广泛使用的检索工具与评价工具。在这些数据库中,“来源文献”与其所引用的“被引文献”已经用数据库语言关联起来,“来源文献”与“被引文献”之间的检索越来越方便,改变了以往原文查找费时、费力而检索结果又不能令人满意的局面^[3]。

但是以上各数据库在收录范围、检索效果等方面都有一定的差别。本文从引文检索功能方面对上述中文文献数据库进行比较研究,包括收录范围、检索方式、检索结果(查全率)、检索结果输出格式等,以使用户了解它们的优点和缺点,从而更有效地选择和利用。

2 材料与方法

2.1 材料 中国科学引文数据库 2002 版、维普中文科技期刊数据库(引文版)、中国科技论文引文统计分析数据库、中国期刊全文数据库。

2.2 方法 1973 年美国情报学家 Lancaster 和 Fayen 列出了衡量信息检索系统效果的 6 项指标,即收录范围、查全率、查准率、响应时间、用户负担和检索结果输出格式^[4]。因客观因素的限制,本文仅选用收录范围、查全率和检索结果输出格式 3 个指标,并另外补充了检索方式指标进行评价。检索方式指标包括检索字段、检索年限和检索功能。

在用这些评价指标(除收录范围)来评价中文文献数据库时,选取收录的所有学科的期刊。

2.2.1 收录范围的比较。对收录范围(收录期刊的种类、年限和数据量)列表进行比较。

2.2.2 检索方式的比较。对检索字段、检索年限及检索功能(包括布尔逻辑符、特殊功能和二次查询)列表进行比较。

2.2.3 检索结果(查全率)的比较。由于这四个数据库共有的检索字段只有作者、刊名、篇名和机构,且各数据库收录范围不同和在录入篇名时略有差异,本文选择了作者字段进行比较。选取的样本分别选自中国医科大学生物学、基础医学、临床医学、口腔医学、公共卫生与预防医学和药学的 21 个二级学科点的博士生导师 21 名。为了消除数据库间收录年限不统一的因素,将来源文献年限定为 1999-2001 年。分别在四个数据库中作者字段输入这些教师的姓名进行检索。因 CSCD 和万方中只列出第一作者的文献被引用情况,为了保持可比性,从维普和 CJFD 的检索结果中筛选出被检作者为第一作者的源文献作为检索结果。为了避免有非中国医科大学的同名被引作者出现,笔者又通过中国生物医学文献光盘数据库(CBMdisc)核对检出来源文献的参考文献,以保证检出来源文献所引用文献的作者确为中国医科大学博士生导师。

检出结果后计算查全率。查全率理论计算公式为^[5]:查全率=(被检出相关文献数/文献库中相关文献总数)×100%。由于无法计算各数据库中相关文献的总数,故用各个中文数据库中被检出的相关文献的总数^[4]来代替,计算相对查全率。在本次研究中具体为:查全率=(各个中文文献数据库检出的被引次数/被引总次数)×100%(被引总次数为查重后各个中文文献数据库中检出的各被引次数之和)。

比较各中文文献数据库的查全率,进行配伍设计 χ^2 检验。由于本文计算相对查全率有特殊性,缺少全阴性数据,所以无法用统计学方法对四个数据库同时检验。因 CSCD 是国内比较权威的引文数据库,所以只对其他三个数据库进行

收稿日期:2004-10-12

作者简介:王 飞(1980-),女,吉林长岭人,本科,助教。

配伍设计的 χ^2 检验(检验水准 $\alpha = 0.05$)。对查重后的 326 篇文章进行分析,根据各数据库有无某篇文章,记为“1”(有)或“0”(无),列表后进行检验。对三个数据库中每两个数据库进行配对设计的 χ^2 检验,比较每两个数据库之间的差异,检验水准 $\alpha = 0.05/3 = 0.017$ 。

分析各数据库检出的 21 名作者被引次数不同的原因。以 CSCD 为标准对其他三个数据库的优缺点进行了比较。以 CSCD 为标准,分析每篇漏检的来源文献,将漏检的被引次数作为缺点中的次数,将缺点分为期刊收录不全、期刊中文献收录不全和参考文献著录失误三类,将漏检的原因一一归

类。分析每篇多检出的来源文献,以比 CSCD 多检出的被引次数作为优点中的次数,将优点分为期刊收录全、期刊中文献收录全和参考文献著录全三类,将多检的原因一一归类。列表比较,计算出各个数据库中每种优/缺点所占的次数占所有优/缺点次数总和的比例。

2.2.4 检索结果输出格式。列表比较各数据库每页显示条目数、是否提供全文、是否需要专用浏览器和显示字段。

3 结果与分析

3.1 收录范围

表 1 收录范围一览表

	CSCD(2002 版)	万方	CJFD	维普
收录期刊种类	国内出版的 1000 多种中英文重要期刊和优秀期刊,其中核心期刊 633 种	1200 多种科技核心期刊	6100 种期刊,其中生物医学期刊 747 种,含核心期刊近 200 种	科技类期刊 5000 多种,其中包括《中文核心期刊要目总览》中的核心期刊 1500 余种
数据量	到 2002 年止,共收录了 71 万条来源数据和 222 万条引文数据	论文部分数据达 38 万余条,引文部分数据近 140 万余条	积累全文文献 800 万篇,题录 1500 余万条	达 200 万条/年
收录期刊年限	1989 年至今	1989 年至今	1994 年至今	1990 年至今

从表 1 可知,在收录期刊的种类方面,CJFD 要优于其他数据库;在收录文献的年限方面,CSCD 收录文献的年限最长。

3.2 检索方式

3.2.1 检索字段

表 2 检索字段一览表

	CSCD	万方	CJFD	维普
引文检索字段	被引著者、被引文献、被引机构、被引实验室、被引书刊	引文作者、引文题名、引文机构类型、引文基金类型、引文刊名、引文出版年份、引文作者单位、引文作者地区、引文学科分类	引文(包括作者、题名、机构、刊名、年份等)	作者、篇名、刊名、任意字段
来源文献检索字段	论文题名、文献著者、著者机构、期刊刊名、关键词、省市地区、实验室	论文题名、论文作者、全文	篇名、作者、机构、刊名、关键词、摘要、基金、全文、ISSN、主题词	题名、作者、机构、刊名、关键词、第一作者、文摘、分类号、任意字段
总数	12	12	11	13

从表 2 可以看出,维普的检索字段最多(13 个),CSCD 和万方次之(12 个),CJFD 的检索字段最少(11 个)。其中,万方中被引文献的检索字段最多(9 个),而且很多字段是它特有的,如引文基金类型、引文学科分类、引文作者地区等;CJFD 中被引文献的检索途径最少,只有“引文”一个入口。虽然“引文”字段中包括作者、题名、刊名、年份等多个字段,相当于引文中的任意字段,但是正是因为检索引文的途径不明确,不能自由选择被引作者、题名等具体检索途径,用户只能在引文的全部范围内检索,不能直接获得引文结果。CSCD 引文检索字段中的“被引文献”字段又包括文献著者、期刊名称、年、卷、期、页等 6 项信息,给用户提供了更多的选择。这些数据库引文检索共同的字段只有被引作者、被引书刊名、被引文献题名。从来源文献的检索字段可知,只有维普和 CJFD 有文摘或摘要字段。全文检索字段是非检索专业用户最易用

的,CSCD 却没有。

3.2.2 检索年限。截至 2004 年 6 月,各数据库的检索年限见表 3。

表 3 检索年限比较

	CSCD	万方	CJFD	维普
起始年	1989	1994	1994	1999
截止年	2001	2001	2004	2003
跨越年度	12	7	10	4

表 3 说明,CSCD 收录期刊的年限最长而且最早,CJFD 更新速度最快。

3.2.3 检索功能。

表 4 检索功能比较

	CSCD	万 方	CJFD	维 普
布尔逻辑符	或,与,非	或,与,非	或,与,非	或,与,非
特殊功能	限第一著者、限第一机构、字典检索	选择显示格式、记录显示顺序、期刊种类选择、地区选择	相关度或日期排序、中英文扩展、目录选择	同义词库、同名作者库、期刊范围选择、模糊与精确查询、分类导航、刊名导航
二次查询	有	有	有	有

从表 4 中可以看出,这四个数据库都可用布尔逻辑运算符进行组配检索。但万方的命令检索是由用户输入的,其他数据库是由用户选择逻辑符的。

在特殊功能方面,维普的特殊功能最多(6 个)。其中同义词库在用关键词检索时,有选择地扩大检索范围,提高查全率;同名作者库在用作者、第一作者查询时,可以提示同名作者的单位列表供选择,以提高查全率;“模糊”和“精确”检索功能,在用关键词、刊名、作者、第一作者和分类号 5 个字段检索时生效,默认为“模糊”检索;期刊范围选择,选项有“所有期刊”、“重要期刊”和“核心期刊”,默认为“所有期刊”;分类导航和刊名导航,可有效地缩小检索范围,实现精确查询。CSCD 可以在检索时限制第一著者和第一机构,利用字典检索时,还可采用标记关键词方法一次选择多个关键词进行逻辑

“或”操作。万方可按数据库更新日期“升序”或“降序”显示结果,系统默认为“升序”,也可按记录顺序进行排序;可选择期刊种类和地区,提高查全率;可通过全部信息和论文题录两种方式显示检索结果。CJFD 可选择目录,如“理工 A 辑专栏目录”、“医药卫生辑专栏目录”等进行查询,提高查全率;选择“无”、“相关度”和“更新日期”顺序显示结果。

这四个数据库都具有二次查询功能。对于二次检索与一次检索关系,CSCD 和维普可以是“与”、“或”、“非”中任何一种关系,万方可以是“与”、“或”,CJFD 则只有“与”。

3.3 检索结果 在数据库中通过作者字段检索所选的样本博士生导师的被引次数。21 名作者的检索结果(被引次数)见表 5。

表 5 检索结果(被引次数)一览表

作者	维普		万方		CJFD		CSCD		被引总次数
	被引次数	查全率	被引次数	查全率	被引次数	查全率	被引次数	查全率	
滕国玺	0	0	0	0	1	1	1	1	1
孙开来	7	0.64	6	0.55	10	0.91	4	0.36	11
宋今丹	28	0.78	22	0.61	12	0.33	25	0.69	36
石玉秀	5	1	1	0.20	4	0.80	1	0.20	5
吕昌龙	28	0.72	27	0.69	21	0.54	28	0.72	39
罗恩杰	5	1	4	0.80	4	0.80	4	0.80	5
宋继谒	6	0.75	3	0.38	4	0.50	3	0.38	8
贾静涛	2	0.18	11	1	2	0.18	11	1	11
白小涓	1	0.33	0	0	3	1	0	0	3
麻宏伟	3	0.43	5	0.71	6	0.86	4	0.57	7
宋芳吉	5	0.45	7	0.64	9	0.82	5	0.45	11
郭启勇	11	0.58	10	0.53	15	0.79	7	0.37	19
谷春久	4	0.80	4	0.80	5	1	3	0.60	5
尚 涛	17	0.77	9	0.41	15	0.68	7	0.32	22
张劲松	14	0.48	13	0.45	18	0.62	9	0.31	29
于靖寰	0	0	6	0.86	6	0.86	6	0.86	7
徐惠绵	35	0.78	11	0.24	33	0.73	7	0.16	45
许国忠	16	0.89	9	0.50	9	0.50	4	0.22	18
张力平	6	0.67	7	0.78	8	0.89	3	0.33	9
孙贵范	13	0.81	14	0.88	9	0.56	9	0.56	16
王怀良	8	0.42	7	0.37	11	0.58	12	0.63	19
合 计	214	0.66	176	0.54	205	0.63	153	0.47	326

表 5 显示,各数据库检出的被引次数及查全率有较大差异。如:维普中“于靖寰”的查全率是 0,“罗恩杰”的查全率是 1。综合 21 名作者的被引结果进行比较,对查全率进行汇总,

得出结论:维普的查全率最高(0.66),CJFD 次之(0.63),CSCD 最低(0.47)。这可能与 CSCD 对刊种和文献的高度选择性有关。对检索结果进行了 χ^2 检验。检验结果见表 6。

表 6 326 篇来源文献的配伍检验结果

维普	万方	CJFD	篇数
1	1	1	74
1	1	0	47
1	0	1	53
0	1	1	36
1	0	0	38
0	1	0	23
0	0	1	45
0	0	0	10
212	180	208	326

表 6 中,1 表示数据库中有,0 表示数据库中无。经检验,这三个数据库之间的查全率有显著性差异($P < 0.05$)。经配对设计的 χ^2 检验得出,维普和万方之间有显著差异。万方和 CJFD 之间、维普和 CJFD 之间差异不明显。这说明维普检索结果的查全率略高于其他数据库。

以 CSCD 为标准对其他三个中文数据库的优缺点进行汇总,结果见表 7。

表 7 三个中文文献数据库与 CSCD 比较优缺点统计表

	优 点						缺 点					
	期刊收录全		期刊中文文献收录全		参考文献著录全		期刊收录不全		期刊中文文献收录不全		参考文献著录不全	
	次数	比例	次数	比例	次数	比例	次数	比例	次数	比例	次数	比例
万方	11	0.26	28	0.65	4	0.09	5	0.25	1	0.05	14	0.70
维普	27	0.23	83	0.69	10	0.08	35	0.60	11	0.19	12	0.21
FD	29	0.24	92	0.75	1	0.01	4	0.06	5	0.07	58	0.87

注:比例为次数除以优点或缺点中次数的总和。如万方中比例 $0.26 = 11 / (11 + 28 + 4)$ 。

从表 7 可知,与 CSCD 相比,万方、维普和 CJFD 期刊文献收录全的比例较高,分别是 0.65、0.69 和 0.75。这是由于 CSCD 收录的期刊经过严格的筛选,收录期刊文献时具有高度选择性。维普的缺点是期刊收录得不全(0.60),其原因并不是刊种收录得不全,而是收录的卷、期不全。例如:2000 年的《中国医科大学学报》,维普只收录了第 1、2、3、6 期。万方和 CJFD 由于参考文献著录失误而造成漏检的比例较大,分

别是 0.70 和 0.87。著录失误包括著录不全或著录错误。CJFD 著录失误的原因主要是著录不全,而万方则是著录错误较多,如将“宋继渴”著录成“朱继渴”。

3.4 结果输出格式 结果输出格式是帮助用户快速找到自己所要的文献的一个重要方面,也是评价检索系统的一个重要指标。结果见表 8。

表 8 结果输出格式比较

	CSCD	万 方	CJFD	维普
每页显示条目数	不限	20	10	10
是否提供全文	否	否	是	否
是否需下载浏览器	否	否	是	否
显示字段	被引文献、引证文献、题名、出处、第一作者、关键词、引文总数、中文引文、被引文献	记录号、引文题名、引文作者、引文刊名、引文出版年份、引文年卷期、引文页数、引文作者单位、引文机构类型、引文作者地区、引文学科分类、引文基金类型、论文题名、论文作者、论文出版年份、论文年卷期	篇名、作者、刊名(卷、期)、关键词、机构、聚类检索、中文摘要、光盘号	题名、作者、关键词、刊名、机构、分类号、文摘、ISSN 号、光盘号、CN 号、馆藏号、被引文献

从表 8 可以看出,CSCD 采用了滚屏模式,每页显示条目数不限;维普和 CJFD 每页显示条目数相同(10 条);而万方显示 20 条。只有 CJFD 提供全文服务,但需要付费,且需要下载浏览器,并且一次只能浏览一篇原文,每次也只能保存该文后的引文记录,用户使用颇为不便。万方显示字段最多,能给用户提供较多的信息,方便用户判断相关文献。文摘或摘要对用户判断相关文献是很重要的,除 CJFD 外,其他中文数据库均不显示文摘字段。CSCD 引文检索获取的信息还包括引证论文(来源文献)的引文条数和被引的总次数。

在结果输出时,CSCD、维普和 CJFD 的字段数都是固定的,如果某一字段没有著录,就无法显示;而万方却不是这样,不同文献的著录字段并不固定,有时会缺少某些有用字段,这给用户判断文献相关性带来了一定困难,而且在通过这些字段进行检索时也容易造成漏检。CJFD 高级检索的结果输出格式最好,它能够每屏显示 10 条详细信息,而其他数据库只能先显示 10 条的简单信息,点击“题名”才能看到详细信息,而且一屏只显示一条详细记录。但是 CJFD 并非每条记录都可以显示出详细信息,点击“引用文献”常常显示“无

相关记录”或检出一些不相关文献这给用户带来了不便。

4 讨论和结论

4.1 本次研究的不足 一是因客观因素的限制,对各中文数据库的响应时间、用户费用负担等指标没有进行研究。二是在收录范围方面,没有对各数据库收录学科进行分析。三是在检索题目的选择方面,虽然涉及到了医学的众多分支学科,但有可能存在偏差。四是在检索时虽然发现个别数据库在标引方面有些问题(如万方有的记录没有标出作者字段;各个数据库篇名的标引有微小差别),可能会影响检索结果的差异,但并没有深入研究。五是对于被引次数不同的原因分析可能会带有一些主观成分,会影响到分析的结果。

4.2 四个中文文献数据库的不足 能否给用户提供更途径的查询,也是评价数据库的一个重要指标;方便用户检索的字段也同样重要。对非专业检索人员来讲,有许多检索字段,如表 2 所列的基金类型、机构类型、被引实验室等,用处并不大。而全文检索字段是最方便用户使用的,但 CSCD 却没有。这说明这几个数据库的检索字段建设得还不完善。

比较检索结果后发现,这四个数据库的查全率都不高(低于 70%),说明各数据库在引文检索功能方面比较欠缺。CSCD 收录的刊种较少,而维普收录期刊的卷、期不全。

在著录方面,万方、CJFD 和维普都存在一些问题。CJFD 引用文献和万方的一些重要的记录字段标引得不够完整,比如缺少作者字段;维普的篇名标引有出入。这使用户容易漏检一部分相关文献。

在结果输出方面,CSCD 最为稳定,万方、维普和 CJFD 都不稳定。万方对不同文献的著录字段数并不是固定的,有时会缺少某些有用字段,这给用户判断文献相关性带来了一定的困难。CJFD 在使用超过一定时间后要求用户重新登录,并且再次下载检索结果记录最多 10 条,这对于想要保存多个记录的用户来说极为不便。CJFD 并不是每条记录都可以显示出详细的著录内容,引用文献这一字段内容尤为不全,且查

准率低。说明 CJFD 在引文功能方面欠佳。

4.3 结论 收录范围:在刊种方面,CJFD 要略优于其他中文数据库,维普略次之,CSCD 收录的最少;在年限方面,CSCD 和万方收录文献的年限最长,CJFD 最短。

检索方式:在检索途径(检索字段)方面,维普提供的最多,CJFD 最少;仅从被引文献角度看,万方提供的检索字段最多。维普的检索功能最多,而且它的二次查询功能非常方便使用,二次检索词与一次检索词之间可以是“与”、“或”、“非”中任一关系。因此检索方式方面,维普要优于其他数据库。

检索结果(被引次数):维普的查全率最高,收录的文献较全,著录失误较少,要优于其他数据库。CJFD 次之,CSCD 的查全率最低。

结果输出格式:CSCD 最为稳定,CJFD 也比较好。

总的来说,维普的引文检索功能略优于其他数据库。

4.4 建议 万方、维普和 CJFD,应加强对数据库的维护,保证数据库在输出结果的稳定性;万方和维普,应加强文献的标引工作,对用户常用的检索字段尽量做到全部标引;万方和 CJFD 应加强参考文献的著录;CJFD 应加强引用文献字段的著录;维普应注意完整收录期刊中的文献;CSCD 则要加大对中文刊的收录,增加特殊功能,加快更新速度。

参考文献:

- [1] 朱胜坚.国内主要中文生物医学数据库简介[J].医学信息,2000,13(7):387-388.
- [2] 曹学艳,胡文静.我国文献计量学进展研究[J].情报杂志,2004,23(2):67-69.
- [3] 姜福奇,刘海航.四种可检引文的中文数据库浅析[J].现代情报,2004,24(1):66-67.
- [4] 李星逸.《中国学术期刊(光盘版)》检索评价[J].图书馆建设,2000,(6):19-23.
- [5] (美)F. W. Lancaster 著.陈光祚,王知津,王津生译.情报检索系统——特性、试验与评价[M].北京:书目文献出版社,1984.5.

[本文编辑:王 颖]

远程医学的三大支撑技术

田 红 成

(福建省医学科学研究所,福建 福州 350001)

关键词:远程医学;医疗卫生;远程通信技术;信息处理技术

中图分类号:TP37,R319

文献标识码:B

文章编号:1671-3982(2005)04-0056-03

远程医学(Telemedicine)是通过远程通信技术和计算机多媒体技术,来实现对医学信息的远程采集、传输、处理和查询,以达到异地疾病诊治、远程教学、卫生保健服务、共享医疗信息资源等医疗目的。具体地说,远程医学包括远程医学影像、远程医学病理、远程咨询及会诊、远程医学教育、远程

医学信息数据共享等,而医疗保健技术、远程通信技术、信息处理技术便是远程医学的三大关键技术。

1 医疗保健技术

用于远程医学的医疗保健技术包括医疗专业人员的诊疗技术、临床实验室检测技术(如血、尿、体液的各种生化含量指标)、生物学信号数据(如心电图、肌电图、血压、血氧等生理和电生理参数)、医学成像技术(如:B超、CT、断层扫描、核磁共振等医学成像技术)和病人医疗记录(电子病历)。由

收稿日期:2005-04-25

作者简介:田红成(1954-),女,重庆人,大专,馆员,发表论文数篇。