

马张华

我国大型图书馆机检系统检索特点研究

摘要 国家图书馆、中国科学院文献情报中心和北京大学图书馆这三大图书馆的大型检索系统在检索操作上都重视了对国内外机检系统已有的经验和方法的汲取,并在检索功能的设置和使用的便利方面进行了许多努力,形成了各自的特点。但三大检索系统对检索语言在计算机检索系统中的应用等问题,还需要改进。表1。图1。参考文献4。

关键词 大型图书馆 计算机检索 计算机检索系统 图书馆自动化

分类号 G254.364

ABSTRACT In this paper, the author analyzes the characteristics of computer retrieval systems in the National Library of China, the Library of the Chinese Academy of Sciences and Peking University Library, and recommends to improve the application of retrieval languages in computer retrieval systems. 1 tab. 1 fig. 4 refs.

KEY WORDS Large-scale library. Computer retrieval. Computer retrieval system. Library automation.

CLASS NUMBER G254.364

如何根据图书馆文献检索的特点和使用需要,建立适用的检索系统,是目前我国图书馆界普遍关心和探索的问题。通过长期努力,我国图书馆领域已经出现了一批计算机检索系统。分析这些检索系统的特点,对了解我国计算机检索系统取得的进展和存在问题,发现计算机检索系统编制的规律,非常必要。

鉴于此,本文选择北京大学图书馆、中国科学院文献情报中心、国家图书馆这3个大型图书馆的机检系统作为对象,通过对这3个具有较大影响的大型检索系统的性能特点和使用问题的分析,对图书馆计算机检索的使用状况和问题进行概要的分析和讨论。

1 三大图书馆机检系统检索特点分析

北京大学图书馆、中国科学院文献情报中心、国家图书馆的检索系统都是大型的集成系统,涉及多种资源和检索方式,如北大和科学院馆系统除可以查阅本馆图书、电子文献外,还可以查阅高校系统、科学院系统的多种资源,涉及到各种不同的检索操作方法。为了便于讨论,本文的分析仅限于这些馆的书目检索系统。从整体情况看,这些系统尽管在检索操作上各有特点,但都重视了对国内外机检系统已有的经验和方法的汲取,在检索功能的设置和使用的便利性方面进行了许多努力,具有以下相同

的特点。

(1) 提供多种功能的检索入口。检索入口是否充分,直接关系到系统的检索能力和检索的便利性。为了能提供多种检索入口,方便用户从不同角度进行查找,3个图书馆都把题名、作者、主题词(或关键字)、分类号或索书号作为主要的检索入口,国家图书馆和科学院文献中心还分别将国际标准书刊号、出版者、出版地等作为检索点,北大图书馆则在“全面检索”项中将各种可检字段的检索集合在一起,提供利用所有字段检索的功能,使用户可以根据所有的线索进行检索(详见表1)。

(2) 提供组配检索。组配检索是计算机检索系统用于扩大或缩小检索范围、改进检索效果的基本手段。上述检索系统基本上都提供了这一检索形式。为了增加组配检索的灵活性,这些系统往往还允许采用结合不同字段组配的方式。其中,中国科学院文献情报中心将这一功能设置在关键词检索中,规定结合表示不同字段的词符(如“a”表示作者、“t”表示题名等)进行操作;国家馆和北大馆则均采用文本框的形式,通过选择逻辑组配符号,加强组配的易用性。比如北大馆目录“复杂检索”下提供的综合查找界面,可以同时结合主题、著者、题名等多种成分,选择相应的逻辑组配关系进行检索查找,极大降低了用户的操作难度。

表1 三大图书馆检索入口对照

图书馆	检索入口
国家图书馆	题名、责任者、关键词、分类号、索书号、题名责任者拼音首字母、国际标准书刊号、出版者、出版地、出版国等
中国科学院文献情报中心	题名、作者、主题、关键字、科图法分类号、中图法分类号、书号、ISSN/ ISBN 号等
北京大学图书馆	全面检索(包括所有可检字段)、著者、题名、主题、索书号、短语检索等

(3) 提供检索条件选择。这一形式可以通过调整检索要求,增加检索的选择性和灵活性。3个系统一般都提供了对匹配方式的选择,以改进系统的适用性。如:北大的检索系统在检索匹配规则中,提供了精确匹配、部分匹配等选择形式;国家图书馆在系统中设置了前方一致、任意一致、等于等匹配手段。使用这些形式,用户可以根据情况选择恰当的匹配方式,取得较理想的检索效果。例如在只知道作者或主题部分数据的情况下,可以利用部分匹配,增加检索入口,提高检全率;在知道检索对象确切数据的情况下,可以选择精确匹配,提高检准率。

这些系统还都提供了限制检索范围的选择,用于提高检准率。例如,北大馆机检系统在其高级检索界面上提供了对图书馆、语言、文献资源类型、文献类别、出版年代等因素的检索选择,国家图书馆也有类似的选择项目。中科院文献中心对于限制检索则采用了二次检索的形式。例如在输入“计算机”检出1171条文献后,系统可以直接以已检出的集合为对象,通过对各种限定因素的选择以及排序方法的确定,进一步限定检索范围,增加检索的准确性。而国家馆和北大馆基本上只提供通过调整检索选择再次检索的形式。

(4) 排序显示功能。即根据需要对输出结果进行排列。这是影响检索效果的一项重要功能,特别在检出结果数量巨大的情况下,这一性能直接影响到使用效果。在3个图书馆的检索系统中,北大馆系统提供的排序选择最丰富,设置了按作者、主题、题名、出版年顺序、反出版年顺序等多种排列方式,并提供了跳跃查找的功能,即可以在需要时直接跳向指定的检索页,相当灵活。不足的是,这一功能只有在检索结果少于200条时才有效。中科院文献情报中心则仅提供了按照出版年代的次序排列一种功能,但处理能力比较强,不受检索结果数量多少的影响。只有国家图书馆并未对检索结果提供选择方式,它目前提供的排序方式,从检索结果看,似考虑了分编时间等因素,基本上按照从近到远的次序排,

但不很严格。

(5) 按层次显示书目数据的功能。这些系统的检索结果通常包括:题录、收藏情况、基本编目数据以及文摘或内容提要等。一般首先显示命中篇数,输出首页命中对象的概要数据;用户可根据需要,进一步获取所选定文献的详细信息。3个系统都在检索结果显示中提供了分类号、题名、主题词、关键词和收藏地等基本的书目数据,通过它,用户不仅可以了解比较完整的书目记录,而且还可以将它作为扩大查找范围的手段。例如在根据作者或题名查出一文献基本数据后,即可根据MARC数据中的分类号或主题词,进行同类文献的检索。但在这3个系统中只有北大馆提供了相应的内容提要。

(6) 浏览检索功能。浏览检索是指系统能提供与检索对象有关的类目或标题的概要显示,供用户进一步选择查找。这一功能有助于用户了解检出资料的整体情况,节省用户时间,提高检准率。北大馆、科学院馆都在一定的检索界面下提供了这一功能。例如:在北大图书馆“浏览检索”项下,可以在输入主题词后输出系统中包括的所有与该主题词有关的标题形式,供用户选择使用。中科院文献情报中心在使用语词检索时也会提供类似显示,如在使用“侯”字进行作者检索时,系统会列出多个侯姓作者的文献记录情况,供选择使用。图1即为在北大图书馆联机检索系统输入“网络”一词后,系统根据其浏览功能显示的相关标题。用户可以根据这一显示,选择对应的标题进一步查找。

(7) 重视界面的友好。3个系统都不同程度地注意使用友好界面,并提供各种检索帮助。例如,三者都提供了对于检索系统的概要介绍和检索方法的说明;都尽可能提供简练的检索界面,并可以方便地进入新的检索选择、后退或返回主页等;都使用了文本框选择等在计算机检索中可以减轻用户智力负担的形式。中科院文献情报中心还提供检索历史记录,可以根据需要查看已经进行过的个人检索的信息,以供重复使用或作为进一步检索的参考等。这些努

力使得用户只要通过简短的了解就能够进行基本的检索操作。

- 1) 网架结构:平板结构 结构设计[1]
- 2) 网络[7]
- 3) 网络 ,NOVELL[1]
- 4) 网络 安全 软件 ,CISCO PIX[1]
- 5) 网络 电话[1]
- 6) 网络 管理 软件 ,CISCO DEBUG[1]
- 7) 网络 管理 软件 ,CISCO WORKS[1]
- 8) 网络 技术[4]
- 9) 网络 数字电路[2]
- 10) 网络 软件设计[1]
- 11) 网络 商品流通[1]
- 12) 网络安全[2]
- 13) 网络操作 ,X WINDOW[1]
-

图 1 北京大学图书馆标题浏览显示样例

2 检索系统应改进之处

尽管这些检索系统有许多优点,但也有明显需要改进的地方,特别是对检索语言在计算机检索系统中的应用问题,尤其值得讨论。

(1) 分类检索能力没有得到充分开发。上述检索系统目前主要有两种进行分类检索的形式:一种是直接使用分类号检索;另一种是在利用其他途径检出文献后,再依靠标引数据中的分类号进行检索。前一种方式必须事先了解分类号,一般用户很难做到,因为即使是专业人员,也没有能力熟练背诵分类号。第二种方法的不足是要拐一个弯,必须先利用其他途径检出适合的文献后,才能使用。这种状况与我国图书馆长期来将分类目录作为主要目录的传统相比,显然是一个退步。因为即使在卡片式手检工具中,分类目录一般也都同时标出分类号和类名。在当前网络搜索引擎中已经设计非常友好的界面,可以通过类目体系的展开,方便地显示类目结构。

(2) 缺乏同义控制。图书馆标引的一个突出特点是,数据是依据控制规范如主题词表、规范文档等,在同义控制的基础上进行操作的。通过规范控制,一个内容或特征的对象可以集中在同一检索标识下,并可以从不同语词的角度进行检索。但在目前的系统中,3 个图书馆的主题检索(或关键词检索)都只提供了简单的字面匹配,不能利用同义词控制

进行检索。例如,在科学院文献情报中心检索系统中,输入“计算机”检出的文献为 1171 条,而使用“电脑”检出的只有 13 条。其他馆检索系统也存在类似情况。这表明,这些图书馆未能在主题或关键词检索中提供概念检索的能力。同样,除中科院文献情报中心外,北大馆、国家馆也都没有为责任者的不同名称建立同义控制,如用“周树人”不能查出“鲁迅”的文献。在依据控制词表、进行规范编目的情况下,系统不能利用已有的检索语言成果提供基本的同义控制,是对编目资源的一种浪费。

(3) 主题标识的相关揭示有待改进。相关揭示是主题检索系统帮助用户有效进行主题检索的重要手段。计算机检索系统中常见相关揭示形式包括:提供词汇浏览功能;显示字顺相近的词汇,供用户参考使用;利用主题词表的参照系统提供检索主题对象的相关词,帮助扩大检索范围,降低选词难度;利用字面成族的特点,提供系统中与检索词字面接近的词,供用户选择使用等。目前这些检索系统中显示概念联系的功能比较弱,应加以改进,同时还应探索其在检索界面的显示形式,逐步将主题词表中的词间关系控制成果引入计算机检索系统。

(4) 没有进行分类主题的结合。分类主题两者在功能上是互补的,如果能加强两者的结合,可以促进检索系统整体功能的发挥。因此在检索语言层面上往往通过两者的结合改进系统的使用效果。比较常见的有:设置类目索引,以便可以从主题的角度查找类目,方便用户对分类系统的使用;为字顺主题表编制范畴索引,帮助从分类的角度进行主题词的查找;建立类目与相关主题词之间的联系,引导用户在分类和主题两者的联系中同时从不同角度结合检索。网上已经有不少系统探索了分类主题的结合应用问题,并取得了良好效果。从目前三大检索系统的情况看,虽然同时包括了分类标识和主题标识,但并没有在两者之间直接建立起相互联系的系统,只能在检索结果编目数据的基础上间接利用两者的联系,结合的水平比较低,应设法改进。

(5) 数据的完善提供问题。完整的书目数据是用户进行检索选择的依据。传统的检索工具,无论是卡片式目录还是各种书目工具,一般都尽可能提供比较完整的基本数据,包括文献内容的数据。上述 3 个系统对 MARC 数据的提供都比较充分,但除北大图书馆外都没有提供图书的内容提要,对用户检索选择不利。特别是目前的大型馆目录多属于联

机目录,它们提供的书目数据是馆际互借的依据,因此,包括内容提要在内的完整书目数据,对于用户进行检索判断尤其重要。建议有关图书馆增加向用户开放书目提要的功能。

(6) 加强和完善检索结果的浏览和显示功能。浏览功能可以在检出文献数量比较大的情况下,提供检出的子集,节省用户时间,提高检准率。虽然目前中科院文献情报中心和北大馆都已经设置了这一功能,但后者在提供这一功能时受检索数量限制比较大,往往在文献数量较大,最需要该功能时却无法使用。国家图书馆则未提供这一功能。这一情况应设法改变。至于在检出文献的排序显示上,北大馆提供的选择比较充分,主要仍然是增加系统的处理能力问题。中科院文献情报中心和国家馆则主要是确定是否需要提供或增加新的选择排列形式的问题。此外,3个馆都对检索结果提供了跳跃选页的功能,表面上看,这种方式可以改进对检索结果查找的灵活性,但在用户无法判断所需文献究竟在哪一页的情况下,作用并不大。可改为网络检索工具中采用的方法,列出前10页页码或在按字顺排列时以拼音字母作为选择标志,也许更为直接,效果可能更好。

此外还存在其他一些可以改进的方面,如系统的稳定性问题,检索界面的进一步改进问题等等。

从以上讨论中看出,三大图书馆在机检系统建立方面进行了许多努力,有许多成功之处。文中提到需要改进的内容中,不少是与检索语言在计算机系统中的应用,或者说如何将知识结构引入检索系统有关的。要在这方面取得进展,一是应善于汲取已有的经验和方法,包括国内外图书馆检索系统的成功形式和网络检索系统中各种好的形式;另一方面,则是在对检索语言特点、用户需求和计算机检索环境综合考虑的基础上,探索新的适用方法。

参考文献

- 1 国家图书馆检索系统 <http://nlc.gov.cn/main.htm>
- 2 中国科学院文献情报中心检索系统 <http://autolas.las.ac.cn/>
- 3 北京大学图书馆检索系统 <http://www.lib.pku.edu.cn/>
- 4 马张华. 信息组织. 北京:清华大学出版社,2001

马张华 北京大学信息管理系副教授。通讯地址:北京大学。邮编 100871。(来稿时间:2002-10-21)

(上接第 29 页)

(3) 网络服务平台需要升级。NSTL 网络服务平台经过两年的运行,吸引了大量用户,但由于访问量增大,时常造成拥塞,检索速度太慢,并经常发生服务器响应不稳定的现象,影响了用户使用 NSTL 网络服务平台的信心。很显然,网络服务平台是 NSTL 的核心组成部分,是重中之重,所以应该加大技术投资,不断提升网络服务平台的软硬件建设。同时,还应考虑其他方案,如为了保证收费用户的正常使用,可将免费的二次文献检索作一些访问限制,或建立镜像站点,以分散流量。

参考文献

- 1 李德宁编译. 不列颠图书馆——英国国家图书馆. 见:世界图书馆事业资料汇编. 北京:书目文献出版社,1990
- 2 Allen Kent. Library Resource sharing. New York: M. Dekker, 1976
- 3 Savage Noel. A National Periodicals Center: the debate in Arlington. Library Journal, 104(10), 1979
- 4, 9, 10, 11 Mary Biggs. The Proposed National Periodicals Center, 1973-1980: Study, Dissension and Retreat. Resource Sharing and Information Networks, 1(3/4), 1984

- 5 CISTI 主页: http://www.nrc.ca/cisti/cisti_e.shtml
- 6, 7, 8 Kathryn J. Mikoski. Document Delivery at CISTI. The Serials Librarian, 26(3/4), 1995
- 12 叶继元等. 全国外文期刊协调工作的宏观思考——兼及全国期刊中心的建立. 中国图书馆学报, 1984(5)
- 13 肖希明. 文献资源共享理论与实践研究. 南宁:广西教育出版社, 1997
- 14 张文玲等. 我国文献资源合理布局的思考. 情报资料工作, 1995(4)
- 15 史学智. 中国科学院文献情报系统结构性调整与优化的思考. 见:中国图书情报工作文库第一卷. 北京:中央编译出版社, 1996
- 16 张复华. 坚持走资源共享的道路. 见:中国图书情报工作文库第一卷. 北京:中央编译出版社, 1996
- 17 张青勇. 浅谈地区性图书馆资源共享. 青海图书馆, 1990(4)
- 18 袁海波, 孟连生. 网络环境下科技文献信息服务的范例——记 NSTL 及其网络服务系统. 见:第四届海峡两岸科技信息交流研讨会论文集, 2002

强自力 西安交通大学图书馆副研究员。通讯地址:西安。邮编 710049。(来稿时间:2002-11-15)