

●陈光祚

论“中国图书情报学书目数据库”的设计与实现

ABSTRACT Bibliographical Database of Library and Information Science in China includes papers in more than 50 titles of journals in library and information science in China and some papers in other journals, conference proceedings, etc. It has now 60,000 records covering all titles since 1989. 7 refs.

KEY WORDS Library and information science Bibliographical database Entry criteria

CLASS NUMBER G258.94

出于教学和科研的需要,近3年来,我一直在从事“中国图书情报学书目数据库”的建设。采用CDS/ISIS软件作为建库与检索的工具,通过数据套录、格式转换装库和原始建库录入书目数据,建立了相应的数据库检索系统。目前该库已有6万条记录,借助于软件的功能,可以提供120万个检索点。并具有数据的分析、重组、排序的能力。通过这次建库的实践体会,愿就中国图书情报学书目数据库的设计与实现问题,谈谈本人的一些思考,并就正于同行同志。

1 建立图书情报学书目数据库是一项必要的基础性建设

我国图书馆及其研究尽管有悠久的历史,但作为一门科学来从事系统探讨,却是本世纪初以来的事。20年代成立了文华图书馆学专科学校(当时称图书科),出现了专门的图书馆学刊物,沈祖荣、刘国钧等先生一批我国图书馆学的先驱开始发表论著。50年代后又出现了情报学研究论著。一个世纪以来,文献量逐渐增加,30年代和90年代是两个文献发表的高峰。据有关的书目索引的统计,

1949~1980年的31年间,中国图书情报学(包括档案学)的文献达11000多条,1981~1985年的5年间文献量为16000多条,1989年的文献量为4196条,1990年为3584条,1991年为4336条,1992年为5070条,1993年为6175条,1994年为8055条,1995年为9966条,1996年为8122条(不完全统计)。这种迅速增长的文献量表明了图书情报学研究的繁荣。在文献类型方面也有很大的发展。作为图书(专著、论文集、教材等)出版的已达2000多种。专业刊物达数十种之多。发表图书情报学文章的大专院校学报达100多种,图书情报工作的标准文件也公布出版。此外,从80年代起出现了硕士学位论文,90年代起出现了博士学位论文。到目前为止,学位论文累计达数百篇。从90年代起出现了专门的研究与发展报告。有关图书馆自动化和计算机检索的专利文献也开始产生。图书情报学计算机数据库和电子出版物也在逐步发展。然而,对于量多、类型复杂的图书情报学文献,却没有建立统一的经常性的书目控制,缺乏相应的专科学目数据库。

因此,建立和维护一个收录文献类型齐全、时间跨度较长、信息含量充分、具有强大

检索和计量功能的图书情报学书目数据库是一项必要的基础性工作,应该提上议事日程。这种数据库的作用不仅在于对研究文献进行目录学的总结,而且可作为文献检索的工具,作为文献计量分析的工具,作为刊物编辑部审稿查重的辅助工具,作为进一步生产专题书目、个人著述书目、各种期刊累积索引等多种书目产品的信息源。

作为文献检索工具的作用:图书情报学的科学研究也像其他科学研究一样,在选题和写作时需要占有系统的资料,以便在吸取已有成果的基础上进行更进一步的探索,这是不言而喻的。然而当前缺乏一个收录文献齐全、收录时间跨度较长、包含文献类型多样的图书情报学书目数据库,研究者在查阅资料时不得不去查寻单文献类型的检索工具(如报刊资料索引或重庆与上海的光盘系统),甚至直接查阅近年来的书刊资料。这表现在许多文章的引文追溯年代仅为数年,一些重要有关文献被漏引。其结果,往往导致无视前人已有成果的倾向,甚至导致研究结论的偏差。有的引文仅限期刊论文,忽略了其他类型的有关文献;有的仅限于核心期刊的文献,即使是学报之类的重要文献也被排除在学术交流的主渠道之外。这些现象的产生,除作者的研究态度之外,一个重要的原因是文献的可获得性限制了作者的视野,而文献可获得性的高低,又主要取决于检索工具。因此,建立一个理想的图书情报学专科书目数据库是推动学术研究的必需。

作为文献计量工具的作用:文献计量学的理论已成为图书情报学的重要组成部分,图书情报学中各种专题和时间范围的文献计量分析研究不断涌现。例如核心期刊、核心作者、作者群落等等的统计分析文章,就发表过数十篇之多。但是由于缺乏一个专科书目数据库,致使这种计量分析不得不依靠手工统计的方式来进行。限于人力,只统计数种期刊、只统计数年的文献量,以此得出的分析结

论难免有以偏概全之嫌。有的人虽然利用计算机化的数据库(如《人民大学复印报刊资料数据库》)来作计量分析,手段较为先进,但是由于库中文献类型单一,且文献量少,所得出的结论也不能完全令人信服。文献计量分析的可信度主要取决于统计对象的完备性和典型性,因此,图书情报学专科书目数据库将是提高这种计量分析准确性的重要工具。同时,检索系统软件中的许多功能可用于计量分析。例如:建立子库法、分别统计第一作者和非第一作者法、各种排序法、关键词聚类法、某一作者或论题文献分布统计法等等,都可作为计量手段而加以应用。因此专科书目数据库既提供计量分析的信息源,又提供了自动化的计量分析手段。

作为期刊编辑部审稿查重辅助工具的作用:在图书情报学刊物上,一稿多投、一稿多登的现象时有发生。有的同一文章在两种甚至三种刊物上同时发表,或时隔半年至一年相继发表。发表时,有的题名完全相同,有的稍改一二字。这给刊物和作者的声誉都带来不良影响。造成这种状况的原因是多方面的,其中的一个原因是当前没有一种有效的查重手段进行把关。编辑人员虽时常浏览其他刊物,但仅凭记忆来避免重登是不可靠的。此外,文章抄袭现象也不能说完全没有。要提高刊物的编辑质量,就需要图书情报学专科书目数据库。如果该库更新及时,编辑部可以有效地制止已发表的文章再次发表。即使当时没有发现,也可事后在库中进行检索,将重复发表的文章题名及作者进行曝光,从而也可以避免今后的一稿多投现象。在数据库中,借助检索软件功能,发现重复刊登的文章是很容易做到的。至于审查内容抄袭的文章,借助数据库也是可能的。在审查稿件时,可从库中检索同类文章,通过浏览进行比较。

作为编制专题书目、个人著述书目等书目产品的信息源的作用:一个较完备的专科书目数据库,可以对其进行二次开发,编制专

题书目、个人著述书目等多种书目产品。这种二次开发主要是依靠检索编辑软件的功能来实现的。其步骤具体来说是: 拟定合适的检索式(例如某一专题的有关分类号关键词等或某一作者的姓名、笔名等)进行检索, 对命中文献进行套录(下载), 指定排序方法(如按作者姓名排, 按题名排, 按出处名称排, 按发表年代排或以上数种排序的分层实现等), 实施计算机排序, 人工审查修改。这种二次开发的效率很高, 所开发的书目信息产品的成本很低。写文献述评的作者, 也可用此方法进行专题文献的普查与排序, 以节省大量的时间。

总之, 这种图书情报学专科书目数据库相当于 LISA 数据库。我们需要中国的 LISA。

2 建立图书情报学书目数据库已具备的初步条件

现在建立图书情报学专科书目数据库, 有不少可资应用的已有条件。

我国图书情报界已编辑出版了一批图书情报学论著目录、资料索引。例如华东师大的《图书馆学情报学档案学论著目录》(1949~1980, 1981~1985) 上下两册; 武汉大学图书情报学院的《图书馆学情报学档案学出版发行学论文索引》(1949~1985); 北京大学的《图书馆学情报学论文题录》(1989~1992 年分年排印) 等等。这些书目索引网罗资料相当完备。例如华东师大的论著目录收录文献 27000 余件, 武汉大学的论文索引收录文献 20500 余件, 北京大学的题录共收文献 19886 件。但以上 3 种由于是印刷型的检索工具(北京大学的数据录入在 dBASE 库中, 输出打印而成, 未形成数据库检索系统), 所提供的检索途径仅为粗略分类和著者两种(华中师大的目录尚缺著者索引), 不能多维地提供题名、关键词、分类号、出处与出版年月及其组合的检索。而且出版之后, 未作增补更新或累

积, 因而缺乏数据库的多维检索、动态更新、库中记录的统计排序重组等功能。

90 年代以来, 我国对图书情报学专科书目数据库也做过努力。例如中国国防科技信息中心于 1993 年编辑了《情报学核心期刊数据库》, 山西省图书馆编辑并于 1997 年发行了《图书馆学论文文摘数据库》。但收录面偏窄(如前者仅限于 10 种核心期刊的文献), 收录数量较少(前者为 10000 多篇, 后者为 20000 多篇)。当前有一些综合性的书目数据库(光盘)包含了图书馆学情报学领域的文献, 如中国科技信息研究所重庆分所编辑发行的《中文科技期刊篇名数据库》和上海图书馆文达光盘《中国社会科学文献数据库》等。作为图书情报学书目资源的检索来说, 这两种光盘是很有价值的, 它们所收录的期刊品种多, 仅图书情报学的专业刊物就收录 50 多种, 此外还收录了大量非专业期刊上的文章, 尤其是高校学报。这较之华东师大、武大、北大的题录索引是一个突出优点。重庆分所的光盘数据库在著录事项方面, 有的文献还包括有摘要。但是作为图书情报学专科书目索引来说, 它们仍然不够理想。首先, 它们仅收录期刊文献, 文献类型不全面。其次, 作为综合性数据库, 它们对文献的标引不完全适用于专科目录索引的标引要求。例如, 不少记录中标引有“图书馆”、“情报工作”之类的关键词。作为综合性数据库做这种标引是合适的, 而作为专科书目数据库来说却专指度太低。第三, 收录材料的时间跨度较短, 重庆分所的光盘仅以 1989 年为上限, 上海文达光盘则从 1993 年开始, 因而对追溯检索存在局限。第四, 检索项较少, 仅提供分类、主题(关键词)、著者、人名(被论及的人名, 仅上海文达光盘有此项著录)、文献出处及发表年代等检索入口。而且对于 3 个或 3 个以上的著者, 仅著录前两名, 因而著者信息不充分, 导致某些作者被埋没。第五, 著作方式往往不著录, 致使原著和译文不易区分。重庆分所的数据库对于

纯文科性质文献往往不收,例如书史、藏书史等方面文献不予收录,但又未作申明,这可能导致检索的失落。上海文达光盘数据库则对一些刊名有明显科技性质的期刊(如核科技情报工作、农业图书情报工作、医学图书情报工作方面的刊物)不予收录。这种收录政策是前者以“科技期刊”为库名,后者以“社科”为库名所决定的。但对于图书情报学文献的根本属性来说,没有必要进行如此严格区分。但是,这两种光盘数据库可以作为建立图书情报学专科书目数据库的信息源。当然,如要进行套录(下载),版权问题需要事先解决。同时要编制相应的数据格式转换等软件,并需改正其录入的文字错误和不适当的标引,进行增补标引,改正重庆分所光盘中不完整的年代表示法。这种套录建库是一种方便、迅速的建库方法。

3 图书情报学书目数据库文献收录政策

(1) 在文献收录的学科范围方面,宜以面向文献情报的学科群为范围,即采取“大学科”的方案。图书馆学、情报学、档案学、文献学、目录学应作为核心部分收录,编辑出版发行学也应予以收录,与信息技术在图书情报方面应用有关的边缘性文献,例如汉字编码、信息载体、编程技术、计算机网络等文献也需要酌情收录。这种广泛的学科范围,正是图书情报学几十年来进展的反映。然而,这种“大学科”的方案对数据库的命名却产生困难,以“图书情报学”命名似有以偏概全之嫌。在学科集成化问题未取得共同认识之前,我们不得已用“图书情报学”这一名称。

(2) 在收录出版物的地域范围方面,可以兼收大陆、台、港、澳地区的专业书刊和其他类型的资料。过去上述地区间的交流不够,但在图书情报学研究领域内,上述各地区的文献有彼此借鉴的意义,都是中华文化的组

成部分。目前香港已回归,澳门也即将回归,海峡两岸的文化交流日趋频繁。顺应这种形势,通过这种“大地域”的专科书目数据库的建设,正是促进学术交流的一个举措。

(3) 在收录文献的类型方面,应对期刊论文、图书、研究报告、会议文献、学位论文、报纸上的有关文章、专利文献、标准文献、电子出版物以及软件产品等都予以收录。这种“多类型”的收录政策,改变了过去图书情报学书目索引仅以书刊为限的狭窄性,可为用户开拓更多的信息资源,也有利于对文献作者的成果进行更全面的评价,使文献计量分析的结论更趋合理。此外,由于各类型文献是可以转化的,例如会议文献可以转化为期刊论文,期刊论文可以选入论文集,因此多种文献类型的兼收政策,也有利于读者获得多种原始文献的渠道,当某一类型的原始文献不易获得时,可用等价的其他类型文献作为代替。

(4) 文献收录的时间范围应以20世纪以来发表、出版的文献为界。本文前面说过,虽然我国文献学、目录学、版本学、校勘学、藏书学有悠久的历史,并代有著述。然而,图书馆学的系统研究文献出版在20世纪初期,情报学文献的出现更是50年代以后的事,因此以20世纪为界是较为适宜的。当然20世纪编辑出版重印的历代著作,也符合收录范围。从本世纪初开始收录材料,更能完整地反映图书情报学研究的历史全程。这就是“长时间跨度”的收录政策。

(5) 在具体文献的选择收录标准方面,宜采取尊重事实、尊重历史、尊重著者劳动的态度。这个书目数据库应从整体上展示我国图书情报学研究的各个阶段的进展。一般来说,收录文献应力求齐全,使每篇文献在库中有其存在的地位,使每个作者的劳动得到尊重。即使是一些从现在来看是“时过境迁”失去现实意义的文献,也应予以收录,因为它们在当时起过作用,成为图书情报学发展史上的特定阶梯,并且对现在和今后也具有学术

发展史研究的价值。这种建库政策正是体现书目具有文化史意义的所在。不要以观点的分歧和作者的知名度高低作为选择的标准。当然,对于某些内容实在浅薄的文献可以不收,但这也要慎重。因为书目的编制既能揭示、宣扬文献,也能埋没、扼杀文献。目录学史上的“兼收并蓄、抱残守缺”的传统不无道理。

(6) 对会议录、论文集、纪念集中的文章和百科全书中条目,宜作分析著录,以提高文献揭示的深度。

(7) 对专业刊物上发表的英文文献也应予以收录,但需注明中文题名及中文作者姓名。

4 图书情报学书目数据库的著录与标引

除了遵循有关编目标引的准则外,拟提出下列几种看法。

(1) 在文献款目中增设“著者所在单位”项(字段)。在一般的书目中缺乏这一著录项。其实这是一个重要的信息项目。著录此项的好处是:首先可以提供文献作者研究写作的环境,有助于判别文献内容的可靠性和先进性,当然这也不能绝对化。其次,可以根据这一著录项普查各单位的作者群状况、科研成果全貌,并借以判断其整体实力和研究特色。第三,鉴于当前作者中同名同姓者增多的现象,著录作者单位,有助于进行区分。当然,目前作者调动工作单位的现象增多,甚至单位本身改名的现象也增多,加之90年代以前的刊物上少有标明著者单位,也给著录增加难度。尽管如此,仍有必要增设此项。近年来重庆分所的光盘数据库已增加此项,这是一个很好的改进。

(2) 标引是揭示文献内容的关键,标引项是构成数据库的检索点的重要组成部分。因此也需认真研究。

首先是分类法的研究。分类以《中图法》

为准,这不仅因为它是国内主导性的分类法,而且目前重庆和上海的光盘数据库也都采用此分类法,经过套录(下载)的记录,分类号可直接使用。当然,作为综合性分类法中的图书情报学部分的类目设置,对于图书情报学专科学目数据库来说会显得过于粗泛。但在计算机化的数据库中,可用标引关键词、题名中的关键词作为组配检索,以实现更大的检索专指度。因此类目粗泛是可以接受的。如另编专科类法,则工程太大。

其次,借助于检索软件,可以将题名作单汉字标引的处理,这将大大提高标引深度,并改善标引关键词(或主题词)的固定先组配的缺陷。例如题名为“联机数据库的套录问题”的文献,在单汉字检索中,可从“数据库”、“联机数据库”、“数据库套录”、“套录”、“联机套录”等入口均能检索出这一文献来。这就大大提高了检索词表达的自由性,适应不同读者的用词习惯。由于是计算机自动进行这种单汉字标引和组配检索的,因此并不增加标引的劳动。将这种题名单汉字检索和标引关键词检索结合使用,是提高查全性能的理想办法。这也是检索语言向自然语言靠拢的一种表现。

对于某些关键词采用通用的英文缩略字,如SD I、CD-ROM、UAP、UDC、IFLA等等作为标引词。这样可以节省检索时输入检索词的时间。对于题名中夹杂有这些英文缩略字的文献,单汉字标引与检索也能直接加以利用。

5 软件的选择

作为数据库,必须有相应的软件作为建库、维护、检索的支持。软件是检索系统的重要组成部分。对于图书情报学书目数据库来说,对软件的选择,除一般要求外,还应是:记录采用可变长方式,以节省空间开销;可定义的字段与记录长度应较大,以容纳摘要;能处理重复字段和子字段;在录入中应能实现定

义隐含值、字串剪粘、格式纠错等功能;能实现单汉字标引与检索;能进行后控词的检索;能自动产生全库的标引词词典,并加注文献登录数;能进行按用户指定的一键或多键进行检索结果的排序;能实现多种格式的输出;能实现与外部数据的交换,所采用的中间格式应是国际标准的交换格式(ISO 2709);一个数据库中记录总数应能突破10万条以上甚至百万条,能为库的增长留有足够的余地;具有网络功能;软件的升级改版有持续的保证;软件的费用低廉;等等。

笔者比较了若干有关软件,认为CD S/ISIS系列包括(Quick Ins)软件较为适宜。

6 中国图书情报学书目数据库情况

根据上述建库考虑,作者初步建立了中国图书情报学书目数据库。当然这仅仅是初步尝试。现将该库有关情况,介绍如下:

(1) 收录文献的类型

收录50多种图书情报学专业刊物上的论文。

收录高校学报、学位论文、非图书情报学专业刊物上的有关论文。

收录会议录及论文集中的单篇论文。

收录图书情报学专著、教材及标准和专利文献。

(2) 收录文献的时间范围

1989年以来的有关文献基本收录。

1989年以前的文献有选择地收录,尤其被90年代论著引用的早期文献重点收录。

收录文献的年代以90年代为主,尽可能追溯早期文献,直至本世纪初。

(3) 数据项

本库每条款目除题名、著者姓名、出处、分类号、关键词摘要、发表年卷期外,增设了“著者所在单位”项。这有利于帮助鉴别文献写作背景,进行作者群落的分析。

(4) 检索点

①关键词。②著者姓名。③著者单位。④分类号。⑤出处(刊名、出版社名称)。⑥发表年代。⑦题名中的单汉字的位置组合。以上各检索项可进行任意的逻辑组配(与、或、非)检索。

未建索引的字段,如摘要,可进行顺序扫描检索。

(5) 建库方法

套录有关光盘中的数据,进行格式转换,改正其录入文字错误,改变其不合适标引词,增标关键词。

根据原始文献或论著的引用书目、参考书目进行录入建库。

(6) 采用软件

采用CD S/ISIS系列软件包括Quick Ins软件,实现可变长记录,节省空间容量,并能提供多种检索手段,提供应用户要求而定的多种输出格式。检索结果既可编辑排序,也可打印套录。

此软件允许用户自行增加或修改数据,以增补本库未收录的资料。因而系统是开放性的。

(7) 本库的用途

检索有关资料。

用计算机作为手段,进行各项文献计量学统计分析,并能自动排序,列出各种“学术榜”。从而有利于主题、核心刊物、人才、机构、地域等方面的文献计量分析研究。

供刊物编辑部进行文献查重。

可据此生产专题库、个人著述书目或打印成书本式目录。

(8) 本库规模

目前有近6万条记录。

联系人: 武汉大学图书情报学院研究所 陈光祚。

参考文献

- 1 华东师范大学图书馆学系、图书馆编,董秀芬主

●赵达雄

我国公共图书馆面临的挑战与对策

ABSTRACT Public libraries are faced with many competitors, such as publishers, book distributors, book lending services, reader's clubs, family collections, TV, consultation services and modern information technologies. Public libraries should develop themselves in the new environment. 6 refs.

KEY WORDS Publishing Book sales Network Information services

CLASS NUMBER G250

公共图书馆姓“公”为公的根本属性注定了其与社会文明程度的提高和民众精神文化需求保持着极为密切的关系。在我国改革开放大潮中涌现出的诸多新事物、新现象往往对公共图书馆产生影响,既构成竞争与挑战,也造就活力与机遇。剖析这类现象的来龙去脉,探究其本质与规律,采取适当对策,对促进我国公共图书馆事业的发展不无裨益。

1 现象胪陈与内涵辨析

1.1 出版业空前繁盛,令图书馆望“书”兴叹 我国出版界一派欣欣向荣景象:仅1996

年就推出新书112813种,报纸2163种,正式期刊7916种,内刊万余种,声像资料万余种且呈逐年上升趋势。单全国性图书奖就有精神文明建设“五个一”工程奖、中国图书奖、全国图书金钥匙奖、全国优秀党建读物奖、青年读物奖、少儿读物奖、畅销书奖等。出版界虽适时提出树立精品意识,施行名牌战略,实现从规模数量型向质量效益型转变之口号,但书刊价格居高不下,煌煌巨著,洋洋丛书动辄数千乃至数万元的势头难以遏制。并生的负面效应是图书总体数量与书价倘若超过图书馆与读者的购买承受能力必然导致大量积压和无人问津。1996年全国库存书达27.76亿

编《图书馆学情报学档案学论著目录(1949~1980)》。上海:上海人民出版社,1984

2 华东师范大学图书馆学系、图书馆编,董秀芬主编《图书馆学情报学档案学论著目录(1981~1985)》。上海:上海人民出版社,1989

3 武汉大学图书情报学院编《图书馆学情报学档案学出版发行论文索引》。北京:科学技术文献出版社,1991

4 北京大学图书馆学系资料室编《图书馆学情报学论文题录(1989、1990、1991、1992)》。该系印刷(共4册)

5 情报学核心期刊数据库通过鉴定《情报科学技术》1993(3)

6 周静华《关于建立中文图书馆学情报学篇名文献数据库的设想》。《江苏图书馆学报》,1996(1)

7 陈光祚《基于图书馆学情报学书目数据库主题计量分析选择核心课程》。《北京大学学报(信息管理系建系五十周年专辑)》,1997

陈光祚 武汉大学图书情报学院教授。通讯地址:武汉市。邮编430072。

(来稿时间:1998.2.12。编发者:刘喜申)