

中国文献信息资源 与检索利用

李国新 编著

ZHONGGUO WENXIAN XINXI ZIYUAN
YU JIANSUO LIYONG



北京大学出版社
PEKING UNIVERSITY PRESS

中国文献信息资源 与检索利用

李国新 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

中国文献信息资源与检索利用/李国新编著. —北京:北京大学出版社, 2004. 8

ISBN 7-301-07677-0

I. 中… II. 李… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2004)第 073603 号

书 名: 中国文献信息资源与检索利用

著作责任者: 李国新 编著

责任编辑: 郑月娥

标准书号: ISBN 7-301-07677-0/G · 1267

出版发行: 北京大学出版社

地 址: 北京市海淀区中关村北京大学校内 100871

网 址: <http://cbs.pku.edu.cn>

电子邮箱: zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752038

排 版 者: 兴盛达打字服务社 82715400

印 刷 者: 世界知识印刷厂

经 销 者: 新华书店

890mm×1240mm A5 9.125 印张 270 千字

2004 年 8 月第 1 版 2005 年 2 月第 2 次印刷

定 价: 16.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究

前 言

“中国文献信息资源与检索利用”课程是在传统的“中文工具书”课程的基础上改造而成的,是传统的中文工具书课程的时代发展。

北京大学信息管理系自其前身图书馆学系 1947 年创建伊始,就开设了中文工具书课程。我国著名学者王重民先生、该系资深教授朱天俊先生等人先后担任过该课程的主讲教师。20 世纪 80 年代中期到 90 年代末,课程教材较大的修订进行过 3 次。90 年代中期以后,随着国内基于计算机和网络环境的数字资源及其检索系统的迅速发展,给传统的中文工具书课程带来了挑战。挑战绝不仅仅是文献信息资源及其检索工具载体形式上的变化,而且也牵涉到了对资源及其检索工具基本概念、基本类型、基本原理、基本特点、使用方法、检索思路等一系列理论和实践问题的重新解释和定位,牵涉到了对基本文献信息资源及其检索工具体系的重新整合与确认,牵涉到了课程的教学思想、教学内容、教材建设、教学方法、实习环境、绩效评价的重大变化。面对挑战,中文工具书课程自 90 年代中期以来在北京大学一系列课程、教材建设项目的支持下,进行了持续不断的建设、研究与改革。经过对原有课程的体系改造、内容整合、手段变革,使课程从形式到内容都发生了较大变化,基本完成了从传统的“中文工具书”向适应文献信息资源及其检索工具体系时代性变化的“中国文献信息资源与检索利用”的转变。这本教材,就是课程自 90 年代中期以来持续进行的建设和改革的成果之一。

90年代中期以来所进行的建设与改革给课程和教材带来了什么主要变化?形成了什么新的特点?概略说来,大约有如下数端。

第一,构筑了“文献信息资源—文献信息资源的检索查考—文献信息资源的有效利用”三位一体的内容体系。

文献信息资源是人类文明演进、思想成果的记录。具有充分地、理性地、有针对性地获得和利用文献信息资源的能力,是形成终生学习能力和思想创造、知识创新的起点,是发现问题进而分析问题、解决问题的基本途径,是了解学科前沿与新成果、新趋势、新信息的主要方法。因此,获得和利用文献信息资源的能力历来受到重视。

获得和利用文献信息资源能力的基本要素至少包括三个方面:对文献信息资源本身的熟悉,对文献信息资源检索工具的熟悉,对海量文献信息资源的检索、选择、分析能力。三者具有内在的有机联系。对文献信息资源本身的熟悉是学养积淀,是基础培育;对检索工具的熟悉是形成能力的桥梁和途径,是必须掌握的方法和技能;而能力的最终体现,则是文献信息资源的获得、选择、分析、利用的综合水准。基于这样的思想,为适应研究型大学本科生未来工作、学习的需要,课程整体内容设计的改革,主要是突破了单纯的检索工具介绍的模式,围绕着“文献信息资源—文献信息资源的检索查考—文献信息资源的有效利用”而展开,强调以文献信息资源获得为中心的知识、方法、思路拓展,尽量避免课程内容变为缺少有效知识含量、缺少原理介绍和共性抽象的纯粹检索工具罗列。这一思路,也决定了课程和教材的名称没有采用习见的“文献信息资源检索”,旨在强调内容体系的基点是“文献信息资源”与“文献信息资源的检索和利用”,而不仅仅是“文献信息资源的检索”。

为什么又放弃了过去长期使用的“中文工具书”的名称?主要原因是,时代发展到今天,不论是文献信息资源还是检索工具,都已超越了传统意义上“书”的概念,故用“工具书”来指称检索工具,已经名不副实。从这个角度看,改,似乎是必需的。但就我本人而言,对“中文工具书”有特殊的感情,放弃这一称谓,委实有些遗憾。

经过改革,中国文献信息资源与检索利用课程形成的目标定位是:培养学生对中国既有的文献信息资源基本状况的认识和了解;培

养和训练学生检索与查考、选择与鉴别、挖掘与处理文献信息资源的意识和能力;培养和训练学生科学合理地利用文献信息资源的学术伦理与学术规范。与此目标相适应,并考虑到课程分工,形成的课程内容定位是:主要讲授产生在中国的文献信息资源及其检索利用知识,以综合性和人文社会科学类别的资源为主。在时间上,通贯古今;在载体形式上,既包括传统的纸质资源和检索工具,又包括 90 年代中期以后在我国迅速发展的数字资源及其检索系统;在资源类型上,包括迄今为止在中国积淀形成的主要文献信息资源类型。

第二,强化对有关文献信息资源、文献信息资源检索工具基本知识、基本原理、基本特点的阐述与介绍。

为什么要这样做?因为如果把本课程的中心使命定位于培养和训练学生检索、获得和利用文献信息资源的能力上,那么,熟悉检索工具的内容、特点、使用方法无疑是重要的。但是,所谓检索工具是文献信息资源的检索工具,因此,资源本身的特点在某种程度上就决定着检索工具的基本特点,从学理的角度看,更为根本的问题是对文献信息资源本身的熟悉。另外,具体的检索工具数量可以很多,规模可以不同,表现形式可以有异,质量也有高下之分,但某一类型的检索工具,一定有共同的特点和功能,只有真正了解了某一类型检索工具的原理、结构和特点,才能举一反三。所以,强化基本知识、基本原理的介绍,实际上是强化课程的基础性、学理性,着眼于学生知识积淀的宽度和厚度,正所谓“授人以筌”。

强化基本知识、基本原理、基本特点的阐述与介绍的另一方面的含义,是作为在有限学时内完成的课程,强调所涉及的资源类型一定要具有普遍适用性,所涉及的具体的检索工具一定要具有典型性、代表性和可以预期的生命力,尽量减少数量,防止“中药铺”式的罗列检索工具。经过最新一轮课程建设和改革,体现在目前这本教材中,涉及的资源类型和检索工具的数量与过去相比都大为减少。长期的教学实践和研究使我确信一条,想要在数量上穷尽文献资源类型和检索工具品种,既不可能,也无必要。检索工具在不断更新,今天重要的,明天可能不重要;今天实用的,明天可能被淘汰。所以,重要的是基础知识、基本原理,是共性和规律,是意识和能力,而不是以数量取

胜。

第三,跟踪新发展,吸收新成果,形成了立足于纸质资源与工具书和电子资源及其检索系统并存现状的内容体系和理论解释。

传统的中文工具书课程是基于纸质文献资源和印刷版工具书展开的,电子资源及其检索系统的出现,改变的不仅仅是载体形式。比如,电子图书系统和报刊资源数据库的出现,使书目检索工具、索引工具与原始资源走向了合一;网站的出现,使传统意义上分门别类的工具书类型走向了集合;计算机对资源的整合功能,使传统意义上工具书的编纂思路、编纂方法发生了变革;计算机环境下常用的组配检索、模糊检索、位置检索等方法,使资源检索的策略和方法发生了重大变化。面对这些变化,要求课程必须作出与时俱进的成果吸收和理论解释。近年来,这一问题一直是本课程研究和改革的重点之一。到目前,整个课程体系和内容已经将纸质资源和电子资源融为一体。对传统工具书,具有经典意义、目前仍然没有替代产品的则保留,其余的则作了大幅度淘汰;对新兴的电子资源及其检索系统,首先是审慎选择,其次是重点引入,在此基础上,注重原理性、规律性的介绍与概括。对于电子资源及其检索系统的出现给原有的理论、方法所带来的冲击,注重作出理论解释、分析和概括。如目前课程中对电子词典、网络词典特点的阐述,对电子图书系统、报刊检索系统与传统的书目工具、索引工具联系与区别的分析等,都体现了这一思路。

有了电子资源及其检索系统,是不是就可以抛弃纸质资源与工具书?不是。因为首先,并不是所有领域、所有方面都具有了可以完全替代印刷型产品的电子产品;其次,经典性的印刷版工具书本身就是科学文化的时代总结,作为大学的课程,不能只见树木,不见森林,不能不体现出学科历史的厚重,不能失去了系统、宽厚的知识积淀;第三,人们的文献信息资源需求永远是多样化的,文献信息资源的获得能力,需要体现在多途径、多角度、多载体获得的能力上,因此,电子资源及其检索系统的迅速发展,并不意味着纸质文献和工具书的全面终结。基于这一思想,本课程没有走上纯粹讲授电子资源及其检索利用的道路,而是强调立足于二者并存的现状来设计内容构成,建立理论解释,使学生获得的知识和能力是全面的、立体的,而不

是局部的、侧面的。

第四,引入文献信息资源利用的学术伦理与学术规范的内容和训练。

学术伦理是学术活动中重要的素质养成,学术规范是学术伦理的重要表现形式。学术伦理和学术规范涵盖的内容、方面很广泛,但学术活动中利用文献信息资源的准则、方法,始终是学术伦理和学术规范的重要方面。所谓获得和利用文献信息资源的能力,最终目的在于利用,而利用,就存在一个是否符合共同规范、是否科学合法、是否能体现出受过规范训练的问题。基于这一思想,本课程引入了有关文献信息资源利用的学术伦理与学术规范方面的内容,主要是概括了文献信息资源利用学术规范的基本原则,介绍了几种目前国内通行的、重要的文献信息资源利用的引证标注方法。作为教师,传道、授业、解惑并重,教书与育人双赢理应是追求的崇高境界,体现在一个专业课程的教学上,我希望能够通过传达和阐述在学术活动中应该谨守的学术伦理和学术规范,让学生们体会出一些更广泛意义上的操守和准则。

第五,引入对获得的文献信息资源进行有效管理、多样化表达的意识和方法。

在计算机和网络环境下,文献信息资源的检索和获得相对容易。就个人来说,只重视检索和获得,不注意对检索获得的资源进行有效的管理,同样难以做到真正有效的利用。鉴于此,本课程在改革过程中引入了利用通行、易得的数据库软件(如 Access)管理个人资源的内容,目的在于引导学生建立起个人资源必须有效管理才能方便利用的意识,引导学生掌握有效管理个人资源的基本思路和方法。

计算机环境为人们多样化地表达检索获得的事实、数据提供了方便,如利用 Excel 电子表格处理获得的数据,自动生成各类示意图等。这实际上是一种对检索获得的文献信息资源进行分析和再处理的能力,是一种创造性地利用文献信息资源的能力。本课程结合目前国内一些资源数据库捆绑实用处理软件的现状,示例性地阐述和介绍了多样化表达检索结果的思路和方法,目的在于启发学生建立起创造性、多样化地表现和利用文献信息资源的意识,使学生利用文

献信息资源的能力能够体现出技术进步所带来的时代特色。

课程的建设与改革是无止境的事情。时代在前进,学科在发展,与时俱进是课程建设的永恒主题。我曾经看到过,国外著名大学经典基础课的教材,修订的版次可以达到六十多次,这是一代又一代教师持之以恒接力传递的结果。与此相比,我们在这方面还初级阶段得很。目前的这本教材,体现了课程最新一轮建设和改革的成果,但它也是新的建设和改革的起点。为此,我热切期盼这本教材的使用者提出意见和建议。

本教材的编写和出版,得到了北京大学主干基础课教材出版基金的资助。感谢北京大学出版社副总编辑段晓青关注本教材的出版事宜,感谢北京大学出版社胡双宝先生、郑月娥编辑为本教材最后把关,感谢所有为本教材的编写、出版提供了帮助的人们。

李国新

2004年7月于北京西二旗寓所

目 录

第一章 概说	(1)
第一节 中国文献信息资源.....	(2)
第二节 中国文献信息资源检索工具体系.....	(4)
第三节 使用检索工具的基本技术方法.....	(7)
第四节 文献信息资源利用的学术规范	(24)
第二章 辞书资源与汉语字词的查考	(40)
第一节 中国辞书的演进源流	(40)
第二节 中国辞书的结构体例与编纂传统	(47)
第三节 代表性中国辞书	(56)
第三章 古籍资源与基本古籍的查考	(72)
第一节 中国古籍资源	(72)
第二节 基本古籍的全文检索系统	(76)
第三节 古籍流传与古籍版本	(99)
第四节 古籍丛书与类书.....	(112)
第四章 近代以来的图书资源与查考	(143)
第一节 近代以来图书的形态变革与内容变化.....	(143)
第二节 书目检索工具.....	(148)
第三节 电子图书系统.....	(155)
第四节 现代百科全书.....	(169)
第五章 报刊资源与论文资料的查考	(181)

第一节	数字化报刊资源及其检索系统·····	(181)
第二节	报刊引文检索系统·····	(208)
第三节	个人资源管理数据库的创建·····	(222)
第六章	时事信息资源与事实、数据、法规的查考·····	(229)
第一节	网站资源及其检索利用·····	(229)
第二节	年鉴资源及其检索利用·····	(240)
第三节	法律资源及其检索利用·····	(261)
主要检索工具题名索引·····		(272)
主要概念、语词、知识点索引·····		(275)
主要参考书目·····		(280)

第一章 概 说

本教材涉及的“文献信息资源”，主要是产生在中国的文献信息资源。在时间上，通贯古今；在载体形式上，包括纸质文献和电子资源；在资源类型上，包括迄今为止积淀形成的主要文献信息源。

本教材所说的文献信息资源检索工具，以现有的中文检索工具体系中的主要类型和主要品种为主，既包括传统的印刷版检索工具，又包括基于计算机和网络环境的电子版、网络版检索工具。

文献信息资源是人类文明演进、思想成果的记录。充分地占有、理性地利用文献信息资源，是自主学习、启发思想、知识创新的起点，是发现问题进而分析问题、解决问题的基本途径，是了解学科前沿与新成果、新趋势、新信息的主要方法。在文献信息生产能力极大提高的信息社会，充分占有、理性利用文献信息资源的前提，是掌握文献信息资源的检索查考方法；而比较熟练地检索查考文献信息资源的前提，则是对既有文献信息资源本身及其检索查考工具的熟悉。这样，有关“文献信息资源”——“文献信息资源的检索查考工具”——“文献信息资源的利用”的知识与技能，就形成了一个有机联系在一起的具有广泛适用性的知识领域。

第一节 中国文献信息资源

一、资源数量

中国有五千年的文明史,有三千多年的文字记录史,文献信息资源浩如烟海。以下一些数字,可以大体反映我国文献信息资源存量的概貌。

一般估计,流传至今的古代典籍有 10 万种左右。

民国时期国内出版的各类图书有 15 万种左右,其中人文社会科学方面的占 85% 以上。

近代以来至 1949 年国内出版的报纸有 1000 多种,杂志有 35000 多种。

1950—1979 年的三十年间,国内出版各类新书近 25 万种,其中人文社会科学方面的占 60% 以上。出版报纸近 400 种,杂志近 1500 种。

20 世纪 80 年代以来,我国文献信息资源的生产能力超常规、跨越式发展。从 1980 年到 2002 年,我国年出版新书的数量由 13000 多种增加到近 10 万种以上,人文社会科学方面的图书所占份额大体上在 70% 左右。年出版报纸的数量由不足 200 种增加到 2000 多种,平均期印数 1.8 亿份以上。年出版杂志的数量由 2100 多种增加到 9000 多种,平均期印数超过 2 亿册。音像出版物、电子出版物从无到有,规模迅速扩大。2002 年全国由出版社公开出版的音像、电子出版物达 3 万多种,总数量 3 亿件以上。^①

20 世纪 90 年代以来,我国的数字化文献信息资源异军突起。截至 2004 年 6 月底,全国 CN 下注册的域名数达到 38 万多个,网站总数达到 62 万多个。2003 年底时,网页总数近 3.2 亿个,网页字节数达到 60 多亿 KB,在线数据库数迅速增加到近 17 万个,增长速度

^① 2002 年的数字,据《中国出版年鉴》(2003)。

是上一年的 104.8%。数字化信息资源在我国的发展势头十分强劲。

上述数字,大体上反映了我国目前文献信息资源生产的数量水平。时代发展到今天,中国的文献信息资源的总体构成,已经突破了千百年来纸质文献一统天下的局面,进入了纸质文献和电子资源共存的阶段。

二、资源类型

依据不同的标准,文献信息资源可以划分为不同的类型。比如,依载体,可以划分为纸质资源和电子资源;依语种,可以划分为中文资源和外文资源;等等。在实际应用层面,有两种划分类型的方法经常被使用。

一是按照文献资源的形态所作的划分。主要的文献形态有:

- 图书
- 期刊
- 报纸
- 会议论文
- 学位论文
- 研究报告
- 政府出版物
- 档案文献

一是按照文献资源的级次所作的划分。包括:

一次文献。依据作者本人的研究或研制成果创作的文献,即通常所说的“原始文献”、“第一手资料”。一次文献在形态上具有多样性,在内容上具有原创性,在出处上具有分散性。

二次文献。按一定的方法对一次文献进行整理加工,以使之有序列化而形成的文献,主要包括目录、索引、文摘等。二次文献在内容上并不具有原创性,只是对一次文献的整理加工,提供有关一次文献的内容线索。但通过整理加工,它使大量的分散、无序的一次文献实现了集中化和有序化,从而方便了人们对一次文献的检索查考。因此,二次文献通常又被称为“检索性文献”、“线索性文献”。

二次文献的最大优势是,提供的文献线索集中、系统、有序。

三次文献。对大量相关文献进行综合分析而形成的评述研究性文献,如综述、述评等。三次文献一般是围绕某个专题或出于特定目的,利用二次文献提供的线索,选用一次文献的内容,经过分析比较、评述研究等深度加工而形成的,因此,被称为“参考性文献”、“研究文献的文献”。

三次文献在内容上具有综合性,在功效上具有参考性。

对文献区分级次,并没有增加文献所负载的知识总量,但却揭示了文献信息资源通过加工整理可以由分散变为集中、由无序变为有序、由个别提升为综合的演进规律。科学合理地利用好二次文献和三次文献,对一次文献的形成和再生产,对提高文献信息资源的利用效率具有重要意义。

第二节 中国文献信息资源检索工具体系

一、检索工具体系的构成

传统的文献检索工具体系由印刷版工具书构成。20世纪90年代中期以来,中国的文献信息检索工具体系进入了印刷版工具书和电子版检索工具并存的阶段。

传统上所说的工具书,是一种将汇集、编著或译述的材料,按特定的方法加以编排,以供解疑释难时查考之用的图书。

中国是世界上最早编纂刊印工具书的国度之一。从汉代出现的系统字书《尔雅》、《方言》和《说文解字》以及系统目录《别录》和《七略》算起,已经有了两千多年的历史。经过长期的积累和发展演变,逐渐形成了主要包括以下一些类型的工具书体系。

(1) 字典、词典 字典是主要用来解释汉字的形、音、义的工具书。词典是主要用来解释词语的意义、概念、用法的工具书。

(2) 类书 辑录典籍中的资料,按类别或按韵目等形式加以编排,主要供寻检、征引、辑佚历史文献之用的资料工具书。

(3) 百科全书 概述一切门类知识或某一门类知识的工具书,历史上科学文化成就的总结性著述。

(4) 年鉴 全面记录事业年度发展,系统汇集年度重要时事文献信息,逐年编辑、连续出版的资料工具书。

(5) 手册 汇集或记录经常需要参考的知识、文献、资料的工具书。

(6) 书目 一批相关文献的揭示与记录。

(7) 索引 揭示文献内容出处的线索型工具书。

(8) 表谱 以编年或谱系形式反映发展变化的工具书,主要包括年表、历表、专门性表谱(如职官表、地理沿革表等)。

(9) 图录 提供形象化、空间性资料的专门工具书,包括图录和地图。

(10) 名录 以名称为线索,提供简略事实的便捷性工具书。

工具书类型的划分标准及结果不是惟一的。在上述工具书类型中,字典和词典习惯上称为语言性工具书;百科全书、年鉴、手册、表谱、图录、名录习惯上称为资料性工具书,或称参考性工具书;书目、索引习惯上称为线索性工具书,或称检索性工具书。也有人将书目、索引称为检索工具书,其他各类统称为参考工具书。

西方从 20 世纪 50 年代开始进行传统印刷版工具书计算机化的研究与开发。我国在 70 年代初就有过研制电子词典的尝试。但中文工具书计算机化的真正起步,是 80 年代初以来的事情。到 90 年代中期,依托计算机及网络环境的中文检索工具开始较多地出现。今天习惯上所说的“电子版检索工具”,实际上是一种统称。从载体形式上说,有光盘版和网络版之分,早期还有软盘版;从检索方式上说,有单机检索工具、联机检索工具和网络检索工具之分。电子版检索工具的出现,并没有改变工具书的基本属性,但它却丰富了中文检索工具体系的品种,使检索工具体系的构成发生了重大变革。

传统的印刷版工具书类型都可以转化为电子版,供人们在计算机和网络环境下使用,因此,有所谓“电子词典”、“电子百科全书”、“电子年鉴”之称。但电子化的检索工具是以数据库的形式出现的,而数据库是在数据库系统的集中控制下,以一定方式组织存贮在一

起的相关数据的集合。在组织、控制、检索文献信息的技术方法上,数据库方式与传统的印刷版图书有明显不同,因此,人们对电子版检索工具类型的划分更多地不是照搬印刷版工具书的成例,而是从数据库具有的特点出发。以下就是一种较为常见的数据库类型划分方法。

(1) 数值型数据库 存贮并提供数值型数据信息的数据库。对应到印刷版工具书,就是统计信息。数值型数据库除了存贮数值数据外,还可以存贮对应的运算规则,并进行运算。

(2) 事实型数据库 存贮并提供对基本知识和事实的历史、现状、进展进行记录、描述信息的数据库。这类数据库中的数据,既有数值型的,也有文本型的。对应到印刷版工具书,是“参考性工具书”或“资料性工具书”。

(3) 文献型数据库 存贮并提供文献性、资料性信息的数据库。既包括二次文献数据库,如书目数据库,又包括一次文献数据库,即全文数据库。随着技术水平的提高,全文数据库越来越多,如近年来大量出现的 e-book,就是全文数据库。

在今天,不同载体、不同类型、不同功用、不同使用环境的检索工具共同构成了中国文献信息资源的检索工具体系。尽管目前电子版检索工具发展势头迅猛,但可以预料,在今后相当长的时期内,二者并存的格局仍将延续。

二、检索工具的特点

不论是以什么载体形式、什么类型品种出现的检索工具,与一般的图书相比,都有一些共同的特点。

从内容上看,检索工具强调比较全面系统地汇集、记录某一方面的知识或资料;从编排上看,检索工具强调必须按照特定的方法对内容加以组织编排;从编纂目的上看,检索工具强调主要是供人们有目的地检索查考,而不是系统阅读。简单地概括,检索工具在内容上具有完整性,在编排方法上具有易检性,在目的功用上具有查考性。这些特点,并不因为检索工具由印刷版发展到电子版而改变。

但是,电子版检索工具的出现所带来的变革也绝不仅仅是载体