

图书馆管理学丛书

文献信息检索 理论与实践

吴 斌 刘和文 编著



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP

安徽大学出版社

图书馆管理学丛书

文献信息检索 理论与实践

吴 斌 刘和文 编著



北京师范大学出版集团

BEIJING NORMAL UNIVERSITY PUBLISHING GROUP

安徽大学出版社

图书在版编目(CIP)数据

文献信息检索理论与实践 / 吴斌, 刘和文编著. —合肥: 安徽大学出版社, 2010. 6

ISBN 978-7-81110-780-7

I. 文... II. ①吴... ②刘... III. ①情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2010)第 103980 号

文献信息检索理论与实践

吴 斌 刘和文 编著

出版发行: 北京师范大学出版集团

安徽大学出版社

(安徽省合肥市肥西路 3 号 邮编 230039)

www.bnupg.com.cn

www.ahupress.com.cn

印 刷: 中国科学技术大学印刷厂

经 销: 全国新华书店

开 本: 170mm×240mm

印 张: 23.25

字 数: 391 千字

版 次: 2010 年 7 月第 1 版

印 次: 2010 年 7 月第 1 次印刷

定 价: 39.00 元

ISBN 978-7-81110-780-7

责任编辑: 卢 坡

装帧设计: 孟献辉

责任校对: 百 川

责任印制: 陈 如 韩 琳

版权所有 侵权必究

反盗版、侵权举报电话: 0551—5106311

外埠邮购电话: 0551—5107716

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 0551—5106311

序

拿到吴斌、刘和文的书稿——《文献信息检索理论与实践》，我的第一感觉是：“文献信息检索”家族中又“添丁”了；读完这本书稿后，我的第二感觉是作者将文献信息检索这个“老话题”赋予了一些探索的“新视角”。

早在20世纪80年代中期原国家教委就要求高校开设文献检索与利用课程，教学内容着重于图书馆利用及文献检索，对培养用户的信息意识和信息技术能力有一定作用。客观而论，我国图书馆学中的文献信息检索学科体系建设较为成熟，因此，要对文献信息检索的学科框架与实践方法进行一些创新存在一定的困难。尽管如此，作者立足于其开设的文献信息检索课程讲义，在借鉴国内外同行优秀成果的基础上，在内容安排、谋篇布局以及对文献信息检索的理论介绍上“同中求异”，让人耳目一新。我认为其意义有二：

第一，通过中外文献信息检索技术的系统介绍，让用户获取了进入信息资源宝库的最佳路径。阮冈纳赞著名的图书馆学五定律的第四定律是：“节省用户的时间。”资料显示，在国外，研究人员只要花20%~30%的研究时间查资料，而把70%~80%的时间用在这些资料的整理和研究上，在国内却正好相反^①。“文献信息检索”的价值和意义，首先在于让用户掌握文献信息检索的技巧，让用户知道什么是最有用的文献信息，到什么地方去查对其最有用的文献信息，怎样更有效地获取对其最有用的文献信息。试想若能让研究人员省下20%查资料的时间，对一个乃至全体研究人员的意义是何等重要！若再进一步做个假设，研究人员用于查找资料的时间占全部研究时间的20%，而用于真正的研究占到80%将是一个什么样的局面？从这个视角上说，“节省用户的时间”是图书馆人的历史使命，并且，在这方面图书馆人“用武之地”的舞台还非常宽广。另据一项问卷调查显示，“对于一个特定的主

^① 黄华：《加强用户时间成本管理略论》，《晋图学刊》，2010年第1期。

题,通常你会从哪里开始查询资料?”用户的回答是:搜索引擎 84%;图书馆网站 2%(OCLC: Perceptions of Libraries and Information Resources 2005)。这一组数据说明一个重要问题:为什么拥有数十年文献信息检索教学实践和理论优势的图书馆吸引用户眼球的“份额”如此不成正比?笔者以为,原因有二:其一,我们的文献信息检索理论在图书馆学的第二定律——“每个用户有其书”的研究上出问题了。我们的理论是否坚持了“以用户为本”的理念?在个性化服务上,是否细化用户群体,根据用户需求区别对待,有针对性地提供信息资源和服务?其二,我们的文献信息检索在图书馆学的第三定律——“每本书有其用户”的实践上也出问题了。我们的实践是否以用户需求为取向,不断地组织和更新信息资源?同时,是否根据用户需求开发新的更有效更人性化的资源?从这个意义上说,我们的文献信息检索理论和实践的发展“任重而道远”。

第二,通过中外相关文献信息检索理论和知识的梳理,让用户获取提升信息素质的持久“动力”。倘若“文献信息检索”的意义仅仅局限于让用户停留在掌握文献信息检索技巧的层面,那么,文献信息检索的理论与实践将难以获得可持续发展的新动力。因此,“文献信息检索”的最高境界是让用户从信息检索技巧向信息素质上提升。信息素质——发现和使用信息的能力,不仅是新时期人才标准(“知识+能力+素质”)的重要衡量指标,也是终身教育的前提。这又让我想起了阮冈纳赞图书馆学五定律的第一定律——“书是为了用的”,这才是“文献信息检索”理论与实践的第一要义。

鉴于此,我们要感谢图书馆界的先贤,是他们创立了文献信息检索技术与理论,尽管这些技术和理论还没有达到尽善尽美的境地;正因为这样,才给予我们这些并不知名的追随者发展的空间,使先辈的事业在后来者的身上得以延续。也才有吴斌、刘和文《文献信息检索理论与实践》的问世。祝愿“文献信息检索”家族中这一“新丁”能够茁壮成长;也期盼更多的“新丁”丰富“文献信息检索”这个大家族。

吴文革

2010年3月31日于合肥

目 录

序	1
第一章 文献信息检索理论	1
第一节 文献、信息资源概述	1
一、文献概念的演变	1
二、与文献相关的几个概念——信息、知识、情报	5
三、信息的类型	8
四、信息资源的概念	10
第二节 文献信息的功能与作用	18
一、文献的功能	18
二、文献学的作用	19
第三节 文献信息检索的基本原理	21
一、信息检索及其发展	21
二、信息检索类型	22
三、信息检索语言	23
第二章 文献信息的整序	25
第一节 文献整序与揭示的基本意义	25
一、文献的整序	25
二、文献的揭示	26
第二节 目录学与图书的有序揭示	28
一、目录	28
二、目录学	28

三、书目对图书的揭示与排序	29
第三节 索引	32
一、字词索引	32
二、专名索引	34
三、句子索引	34
四、篇目索引	35
第四节 文摘	35
第三章 中文工具书的主要排检方法	39
第一节 汉字形序排检法	39
一、笔画法	40
二、笔顺法	40
三、部首法	41
第二节 汉字号码查字法	42
一、四角号码法	42
二、中国字度撷法	43
第三节 汉字音序排检法	44
一、汉语拼音字母排检法	44
二、注音字母排检法	45
三、韵部排检法	45
第四节 分类排检法	47
一、按事物性质分类的排检法	47
二、四部分类法	48
三、按学科体系分类的排检法	48
四、采用自编分类体系的排检法	52
第五节 主题排检法	52
一、主题法和《汉语主题词表》	52
二、主题索引	53
第六节 时序、地序排检法及其他	54
一、时序排检法	54
二、地序排检法	55

第四章 社会科学类工具书举要	56
第一节 社会科学类工具书概观	56
一、社会科学信息检索的特点	56
二、社会科学信息检索工具体系	57
三、我国检索刊物的内容编排	58
第二节 国内检索刊物简介	59
一、图书的检索	59
二、论文资料检索的工具书	78
第三节 国外社会科学检索工具书	87
一、国外社会科学综合性工具书举要	87
二、国外社会科学专业检索工具书举要	89
第四节 社会科学参考工具书介绍	93
一、百科全书	93
二、字典、词典	95
三、名录	100
四、年鉴(Annual 或 Yearbook)	101
五、手册(Handbook 或 Manual)	103
六、图录	104
七、表谱(Tables)	105
第五章 自然科学检索的索引与类型	108
第一节 索引导论	108
一、索引的定义和性质	108
二、索引的用途与种类	110
第二节 索引的结构	111
一、索引款目	111
二、索引的参照系统	113
三、索引档的组织方法	115
第三节 索引编制方法	117
一、索引的编制程序	117

二、索引质量评价	119
第四节 索引的类型	120
一、元词索引	121
二、叙词索引	122
三、关键词索引	125
第六章 自然科学工具书举要	128
第一节 自然科学工具书概貌	128
第二节 中文自然科学工具书	129
一、查找自然科学类图书	129
二、查找自然科学类论文	130
三、专业文献检索工具书	132
四、我国自然科学文献检索工具举要	140
第三节 外文自然科学工具书	144
一、世界三大综合性检索性工具——日本《科学技术文献速报》、法国《文摘通报》和俄罗斯《文摘杂志》	144
二、索引类检索工具	147
三、文摘类检索工具	163
第七章 网络信息资源检索	175
第一节 网络信息资源的基本概念	175
一、几个基本概念	175
二、因特网在我国	178
三、因特网上的服务类型	178
四、因特网上的服务功能	179
五、网络信息资源	182
第二节 搜索引擎的信息检索	183
一、搜索引擎界说	183
二、搜索引擎的发展趋势	184
三、搜索引擎的检索方法	185
四、搜索引擎检索示例	187

第三节 网络学术信息检索工具——搜索引擎举要	195
一、目录式搜索引擎举要	195
二、全文搜索引擎举要	197
三、元搜索引擎举要	200
四、英文搜索引擎举例	202
五、中文搜索引擎举例	203
六、网站举例	203
 第八章 中文数据库检索举要	207
第一节 中文索引数据库	209
一、CSSCI 数据库	209
二、全国报刊索引数据库	211
三、报刊资料索引数据库	212
第二节 中文期刊论文数据库	214
一、中文科技期刊全文数据库	214
二、中国期刊全文数据库	220
三、万方数据化期刊全文数据库	223
第三节 中文图书数据库	228
一、超星数字图书馆	228
二、书生之家数字图书馆	231
三、方正 Apabi 数字图书馆	232
四、中国数字图书馆	233
 第九章 外文数据库检索举要	236
第一节 外文索引数据库	236
一、美国《工程索引》(EI)数据库	236
二、美国《科学引文索引》(SCI)网络数据库	239
三、其他索引类数据库	241
第二节 外文文摘类数据库简介	244
一、英国《科学文摘》(INSPEC)数据库	244
二、美国《化学文摘》(CA)数据库	247

三、美国剑桥科学文摘(CSA)数据库	251
第三节 外文全文数据库	252
一、Elsevier Science(SDOS)期刊全文数据库	252
二、EBSCOhost 系统全文数据库	254
三、Kluwer 期刊全文数据库	257
四、应用科学与技术(ASTP)期刊全文数据库	259
五、IEL 全文数据库	263
第四节 外文图书数据库	266
一、Springer 电子期刊及丛书全文数据库	266
二、OCLC FirstSearch 系统数据库	268
第五节 生物信息学数据库(Bioinformatics database)	271
一、生物信息学的定义与发展	271
二、生物信息数据库种类	272
第十章 特种文献及其数据库检索举要	275
第一节 特种文献的检索	275
一、科技报告	275
二、会议论文	279
三、专利文献	281
第二节 特种文献的网络资源检索	284
一、会议论文网络资料检索	284
二、专利网络资料检索	285
第十一章 文献信息检索服务与检索效果评价	289
第一节 文献信息服务的内容和方法	289
一、定题服务	289
二、回溯检索	292
三、数据(事实)咨询服务	295
四、编制专题书目	295
五、用户培训	296
六、“研究信息”咨询服务	296

七、文献传递服务	297
第二节 检索效果评价	297
一、评价的目的、范围和历史	298
二、评价标准与方法	299
三、影响查全率和查准率的主要因素	301
四、提高检索效果的措施	303
第三节 检索理论与检索技术	304
一、检索理论	304
二、检索技术	308
三、检索策略	310
四、检索方法及其种类	311
第十二章 文献信息的综合处理	314
第一节 学术论文写作	314
一、学术论文概述	314
二、学术论文的准备	316
三、学术论文的撰写	319
第二节 课题资料的检索与利用	324
一、检索步骤与方法	324
二、文献信息的整理	329
附录:文后参考文献著录规则	332
参考文献	357
后记	361

第一章 文献信息检索理论

文献,是一个既熟悉又陌生的名词。一方面,人们平时接触“文献”一词的机会颇多。另一方面,人们隐约感到,在不同的语境中,“文献”有不同的内涵,同时,“文献”一词逐渐与“情报”、“信息”等词交替使用,一时难以有一个确切的定义。“文献”的概念不断演变,“文献学”也有“古典”与“现代”之分;学界对文献和文献学的内涵,本来就见仁见智,解说不一。这里首先要论述的,就是文献与文献学的概念问题。不同时期的文献与情报、信息乃至图书既有相同的内涵,又有不同的含义。

第一节 文献、信息资源概述

一、文献概念的演变

“文”、“献”二字连用,最早见于《论语·八佾》:

子曰:夏礼,吾能言之,杞不足征也;殷礼,吾能言之,宋不足征也。文献不足故也。足,则吾能征之矣。

汉代郑玄注:“献,贤也。”并以“文章贤才不足”解释“文献不足”^①。宋代朱熹

^① 何晏《论语集解》引,商务印书馆1959年版。

《四书章句集注》：“文，典籍也；献，贤也。”^①郑玄、朱熹的解释大体一致：文，指文章、典籍；献，指贤人、贤才，自然包括他们的言论和他们所掌握的史料。以上说法代表了汉、宋学者对“文献”的理解。

宋末元初著名学者马端临编撰了我国第一部以“文献”命名的著作——《文献通考》。这是一部考述上古至南宋中期历代典章制度的典志体著作，马氏在序言中解释了此书的命名来由：

凡叙事，则本之经史，而参之以历代会要，以及百家传记之书。信而有证者从之……乖异传疑者不录，所谓“文”也。

凡论事，则先取当时臣僚之奏疏，次及近代诸儒之评论，以至名流之燕谈，稗官……之纪录。凡一话一言，可以订典故之得失，证史传之是非者，则采而录之，所谓……“献”也。

其载诸史传之纪录而可疑，稽诸先儒之论辨而未当者，研精覃思，悠然有得……窃著己意，附其后焉。命其书曰《文献通考》。^②

可见，马端临秉承了前代学者对“文献”的见解并有所发展。为了将“文”与“献”区分开，马氏对著书的行文格式作了明确区分：“文”（叙事）顶格书写，“献”（论事）低一字书写，互为印证。

元代以来，“文献”一词发生偏义，从兼指典籍与贤人变为偏指典籍。如元代诗人杨维桢《送僧归日本》诗末四句云：“车轮日出扶桑树，笠盖天倾北极星。我欲东夷访文献，归来中土校全经。”^③这里的“文献”是指典籍，具体指中国已散佚的经籍。古人以为，秦始皇时，方士徐福东渡求仙，是随带一批书籍的，去的地方就是日本。由于当时尚未焚书坑儒，这批书籍幸存于日本。欧阳修说：“徐福行时书未焚，逸书百篇今尚存。”^④说的就是这个意思。杨维桢说他也很想东渡日本，寻访中土已散佚的典籍。

明代以来，人们所说的文献，通常是指有参考价值的图书资料。如明代永乐年间，编纂了一部大型类书，辑录了14世纪以前各类图书资料，这部规模宏大的类书就叫《文献大成》，后更名为《永乐大典》。

我国当代学者给“文献”下的定义有多种表述方式，或简或繁，不尽一致。

① 朱熹：《四书章句集注》，中华书局1983年版。

② 马端临：《文献通考》，中华书局1986年版。

③ 杨维桢：《送僧归日本》，顾嗣立《元诗选初集·辛集》，世界书局1988年版。

④ 欧阳修：《日本刀歌》，《欧阳修全集》，中国书店1986年版。

1983年,我国颁布了《中华人民共和国国家标准·文献著录总则》(GB3792.1—83),给“文献”下了简明的定义:“记录有知识的一切载体。”现在,这一定义已为多数学者所接受,并被广泛引用。

文献是“记录有知识的一切载体”,这个定义虽然只有10个字,但内涵是丰富的。

第一,记录有知识的载体,才能称之为文献。指出提供知识是文献的本质属性。从知识的时间跨度看,既包括古书上记载的知识,也包括不断涌现的新知识。从知识的学科归属看,也是没有限制的,如历史文献记载着历史知识,地理文献记载着地理知识,科技文献记载着科技知识等。那种认为“文献”是“古文献”的代名词的看法是片面的。同样,认为“文献”只是文史知识的载体,或认为单指科技文献,都是以偏概全。

第二,所谓“一切载体”,则强调记录知识的物质载体是多种多样的。古代的知识载体有金石、竹帛,后来发明了纸张,现代又有感光材料(如缩微胶卷)、磁性材料(如磁带、磁盘)等。纸质文献历史悠久,曾为人类知识的保存和传播立下不朽的功勋,至今仍被广泛使用。但我们不能把眼光局限于此。我国有不少古籍珍本流传于国外,国内已查不到原件,但国内一些大图书馆收藏着这些珍本图书的缩微胶卷,可供利用,因此,这些缩微胶卷也是文献。

第三,记录知识的手段也是多种多样的,如书写、印刷、录音、录像等。

值得注意的是,20世纪90年代以来,以光盘为载体的多媒体“文献库”已陆续问世,令人耳目一新。读者可以利用电脑进行浏览,也可以进行自由词模糊检索,迅速查到所需要的相关信息,还可以打印、拷贝。

由此可见,文献主要由以下五个基本要素构成:

(1)信息内容。是文献的灵魂所在,是构成文献的最基本要素。

(2)记录符号。即提示和表达知识信息的标识符号,如语言文字、图形、声频、视频、编码等。

(3)载体材料。信息内容存储的依附体,又是信息内容传播的媒介体,如龟甲兽骨、竹木绢帛、金石泥陶、纸张、胶片、胶卷、磁带、磁盘、光盘、穿孔卡片等。

(4)制作方式。文献的制作方式经历了刻画、手写、机械印刷、拍摄磁录、电脑自动输入存储方式等阶段。

(5)载体形态。即文献的外部单元形式,如册装、散装、册等。

情报、信息是知识的一部分,文献是情报、信息的一种载体。文献不仅是

情报、信息传递的主要物质形式,也是吸收利用情报、信息的主要手段。关于信息、情报、知识的关系在本章后面将进一步阐述,此处不作具体介绍。

这里有必要说明“文献”一词的英译。

与“文献”相对应的英文名词主要有两个:document 和 literature。document 一词除了有文献的含义外,还有“文件”、“证件”之意;而 literature 则多指某一学科或某一专题的文献。这两个英文名词还有一点不同:document 包括印刷品以外的文字记录,如碑文、古币图文等,而 literature 一般只指书刊资料。

与古典文献学相比,现代文献学有几个鲜明的特点:

(1)重视计算机技术的运用。现代文献学的诞生和发展,几乎是与计算机技术同步的。在国外,第一台电子计算机诞生于 1946 年,20 世纪 50 年代开始进入市场,60 年代初就正式用于文献的存储和检索。我国学者对此十分敏感,袁翰青于 1964 年发表的《现代文献工作的基本概念》就介绍了美国于 1960 年用电子计算机编制《化学篇名录》(Chemical Titles)之事。此后,“机检”迅速由单机检索发展为联机检索,引起了文献检索的历史性变革。可以说,没有计算机科学,就没有真正意义上的现代文献学。

(2)关注急剧增长的科技文献。现代文献学的视野虽然包括人文社会科学文献,但关注更多的是科技文献。现代文献学本来就“发难”于科技工作者(这与中国古典文献学扎根于文史领域有很大的不同)。现代文献学的早期代表人物布拉德福和袁翰青都是化学专业出身,这决非偶然。因为在各学科领域中,化学化工的文献增长最迅猛。如果不用现代文献学的理论与方法对浩如烟海的文献进行整理和排序,科学家们便无法充分有效地利用现代文献。

(3)重视多语种的文献。现代文献工作者为了及时引进国外先进科学技术,除了关注中文文献外,还以大量精力搜集、处理外文文献。现代文献工作者如果不掌握外语,便寸步难行。

(4)重视文献工作的标准化。文献工作标准化是伴随着大规模的现代化生产而发展的,它是提高文献工作质量的技术保证,更是促进国际文献交流和共享,实现文献工作自动化和网络化的重要基础,因而备受现代文献工作者的重视。

虽然古典文献学与现代文献学存在较大的差异,但又在许多方面有共通之处。

首先,古典文献学与现代文献学的研究对象都是文献。尽管古代文献和现代文献的外观形态和内容不同,但都是知识的载体。

其次,古典文献学 and 现代文献学都要研究文献的整理、传播、利用,都有共同的工作规律可循。例如,古典文献学讲究“互著”、“别裁”,现代文献学讲究“互见”、“分析著录”,两者的精神是相通的。又如,古典文献学重视版本鉴别,现代文献学同样重视版本鉴别,目的都是为了保证科研工作的客观性、科学性,而且在鉴别方法上也有互通之处。

第三,古典文献学 and 现代文献学的根本任务,都是要实现知识的科学组织与有效利用。清人章学诚所说的“辨章学术,考镜源流”,其实质就是知识的组织,目的在于“欲人即类求书,因书究学”,“以便稽检”。^①当代学者强调现代文献工作的基本内容是知识的组织和检索,与古典文献学的宗旨是相通的。

第四,在研究工作中,古典文献学 and 现代文献学也经常是优势互补,难以截然分开。研制古典文献的全文数据库,需要兼有古典文献学 and 现代文献学两方面的知识。整理现代文献,如果不具备古典文献的知识便难以胜任。

二、与文献相关的几个概念——信息、知识、情报

1. 信息

在当今的社会里,信息无时无刻不存在于宇宙的各个角落,无时无刻不包围着人们,信息和人们的生活、学习和工作息息相关,既有人的感官和身体所感受到的信息,又有借助于语言和文字等手段彼此传递和交流的信息。人们在各种活动中,都在自觉和不自觉地使用着普遍约定和个别约定的手势、体态姿势、目光及简短用语等方式交流和传达着彼此的感受和感觉,以达到交流的目的。由此看来,信息联系着世间万物,联系着每一个人。信息无处不在,无时不有,无人不用,它已成为当今人们使用频率极高的词汇之一。信息究竟是什么呢?不同的学科,对信息这个概念有不同的解释。

控制论的创始人维纳说:“信息既不是物质,又不是能量”,信息“是人们

^① 章学诚著,刘公纯标点:《校雠通义》,北京古籍出版社1956年版。