

21世纪普通高等教育规划教材

信息资源检索

理论·技巧·训练

李谋信 编著

第3版



为使用本书作教材的教师提供

配套PPT

cmp9721@163.com

 机械工业出版社
CHINA MACHINE PRESS

21 世纪普通高等教育规划教材

信息资源检索:

理论·技巧·训练

第 3 版

李谋信 编著

机械工业出版社

本书以信息检索能力培养为目的,构建了理论解析、技巧运用、模块训练三位一体的教学体系,体现了检索技能专业通用性、教学内容社会实用性、训练素材普遍可获性的特点。本书第3版在结构和内容上都做了大量调整和补充,主要体现在:将检索理论部分拆分细化,整合了作者近年来的新认识,并将其系统化;将实验指导书纳入教材体系,增加了10个实验,这些实验均取材于真实课题;检索资源部分,各章选择两种重点资源详细介绍、其他相关资源简略介绍,并增加了开放存取的内容;针对每章内容给出了15道测验题,包括判断题、单选题和多选题,可供教学使用。

本书可作为各类高校“信息检索”课程的教材,也可供其他人员自学参考。

图书在版编目(CIP)数据

信息资源检索:理论·技巧·训练/李谋信编著. —3版. —北京:机械工业出版社, 2017.4

21世纪普通高等教育规划教材

ISBN 978-7-111-56370-9

I. ①信… II. ①李… III. ①信息检索-高等学校-教材
IV. ①G252.7

中国版本图书馆CIP数据核字(2017)第052701号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:易敏 责任编辑:易敏 刘静 任正一

责任校对:张薇 封面设计:马精明

责任印制:李昂

北京瑞德印刷有限公司印刷 (三河市胜利装订厂装订)

2017年5月第3版第1次印刷

185mm×260mm·16·5印张·439千字

标准书号:ISBN 978-7-111-56370-9

定价:35.80元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线:010-88379833

机工官网:www.cmpbook.com

读者购书热线:010-88379649

机工官博:weibo.com/cmp1952

教育服务网:www.cmpedu.com

封面防伪标均为盗版

金书网:www.golden-book.com

前言

信息检索发轫于目录学,成形于索引理论,在吸收了逻辑学和计算机科学的内容后逐渐形成一门独立的学科。自第二次世界大战以来,信息检索工作随着科学文献的爆炸性增长得到了快速发展,经历了书本式检索工具书、穿孔卡片、缩微阅读设备、远程联机登录、局域网光盘检索、网络资源学术化和社会网络化几个阶段。随着社会生活信息化和互联网的普及,信息检索如今已成为各领域专业人员的必备技能之一。信息检索在科研立项、专利审查、技术引进、人才猎取等许多方面的前期工作中都有着重要作用,并成为技术中介、市场调查、论文写作中收集整理和核实数据的手段之一。信息检索还有许多新用途,例如,据清华大学伊斯雷尔·爱泼斯坦对外传播研究中心王君超研究员的介绍^①,信息检索可用于识破网络谣言,通过关键词搜索、比对,可以识别出虚化了时间、地点的“旧闻翻新”和情节离奇的消息。信息检索在美国等经济发达国家自20世纪90年代以来发展很快,逐渐形成了一个新兴行业——信息经纪业。信息检索已不仅仅是图书情报人员的业务内容,它正吸引着工程技术、律师、公共关系等行业的从业人员不断加入其中。

信息检索工作极富挑战性,它除了要求检索工作者具备丰富的专业知识、熟练的语言能力(包括中外文专业文献阅读能力和词汇联想能力)、娴熟的检索技巧和对文献资源分布情况的了解外,更需要勇于探索的精神和人际交往能力。对信息检索的教学同样也富有挑战性。自改革开放之初“科技文献检索”被引入我国高校以来,“信息检索”经历了在教学内容上从侧重于情报学到逐渐倾向于信息科学,在检索工具上从以手工检索为主到逐渐以互联网检索为主,在教育对象上从面向理工科专业扩展到面向所有专业这样一个过程。“信息检索”课程在我国40余年的发展历程表明,它仍是一门年轻的、快速发展的学科。为了使信息检索这门课程适应新时代发展的需要,教材应具有科学性、时代性、通用性和实用性,以帮助读者提高情报素养、获得检索经验、掌握典型信息资源。本书在这方面做了初步的努力。

关于构造一个以实践为特征的信息检索课程体系结构的问题,多年来一直萦绕在笔者心头。信息检索行为和技巧是否需要理论来说明、解惑?哪些技巧需要归纳提炼?信息检索训练目标应该由哪些子目标组成?这些子目标通过什么题型实现?这些疑问在本书中都有了初步回答。

本书由13章组成。第1~5章是基础理论部分,其中,第1章介绍了一些便于后续内容理解的术语概念和规律以及传统文献的类型等;第2章介绍了传统检索工具书的类型和构成要素、检索系统的数据结构和索引类型以及检索结果的显示形式;第3~5章从检索人员的角度介绍了以分类和主题检索语言为主要内容的信息组织理论和复杂概念的分析表达方法。第6~12章是网络资源搜索的概念、工具与资源部分,其中,第6章介绍了网络检索的一些基本概念,网络资源查找工具的类型、搜索引擎的工作原理和检索功能;第7章重点介绍了

^① 见《人民日报》2013年8月3日第4版《四件“武器”识破谣言》。

维普资讯和中国知网两种国内综合性全文数据库,并简要介绍了一些国内的开放存取资源;第8章重点介绍了EBSCOhost和SpringerLink两种国外全文数据库,并简要介绍了一些国外的开放存取资源;第9章重点介绍了国家知识产权局网与WIPO知识产权和技术数据库,并简要介绍了一些知识产权概念和部分国外知识产权检索系统;第10章重点介绍了中国技术性贸易措施网和中国标准服务网,并简要介绍了一些技术性贸易措施的概念和其他相关网站;第11章重点介绍了国家统计局网和中国经济信息网,并简要介绍了一些与统计数据相关的概念和数值型资源;第12章重点介绍了Getty Images网和视觉中国网,并简要介绍了图片库的含义、特点和相关网站。第13章将本课程所学内容与毕业论文的撰写相结合,进行针对性训练,主要介绍了学术训练的概念、学术论文的结构,并列举了两个撰写实例。

本书自2006年出版以来,重印和修订了多次,本版(第3版)修订内容主要体现在以下几方面:

- (1) 突出了实践性。在分类主题概念分析、网络资源和学术论文撰写等章节增加了实验内容、目的、要求和实例,以及实验报告项目说明、错误分析。
- (2) 增加了课题分析和主题概括的步骤和分析实例。
- (3) 专类网络资源前增加了相关专业基础知识介绍。
- (4) 增加了检索系统基本概念的介绍。
- (5) 将“标准法规资源”调整为“技术性贸易措施检索资源”。
- (6) 增加了“学术训练基础”一章。
- (7) 去掉了太过陈旧的内容或已消失的资源,增加了新的内容。
- (8) 将各章后的思考题改为判断题、选择题等。

本书是笔者多年教学和实践经验总结。笔者努力将具有规律性、基础性、共性和典型性的内容,以及经验技巧归纳、提炼出来,以利于读者举一反三,触类旁通。本书的结构体系和内容遴选是笔者多年“信息检索”课教学探索的结果,但仍有不妥之处,有待继续修订,真诚地希望同行和读者不吝赐教,在此先行谢过。

李谋信

于上海松江大学城

作者制作了与本书配套的PPT课件等补充资料,使用本书授课的教师可联系本书编辑索取(cmp9721@163.com)或与作者联系(5385@suibe.edu.cn)。本书习题答案仅供授课教师参考。

目 录

前言

第1章 预备知识..... 1

- 1.1 基本概念 1
- 1.2 信息资源的选择和检索效果的评价 3
- 1.3 检索系统中的代码 5
- 1.4 传统文献的类型 10
- 1.5 测验题 13

第2章 检索系统概述 15

- 2.1 传统检索工具书 15
- 2.2 计算机检索系统 16
- 2.3 检索结果的显示格式及其内容 24
- 2.4 检索过程 30
- 2.5 测验题 32

第3章 分类检索语言及其运用 34

- 3.1 文献分类法的体系结构 34
- 3.2 文献分类号的结构 35
- 3.3 文献分类原则 36
- 3.4 类目含义的识读 38
- 3.5 分类规则 39
- 3.6 国际专利分类 42
- 3.7 国际外观设计分类 46
- 3.8 实验一 46
- 3.9 测验题 47

第4章 主题检索语言及其运用 50

- 4.1 基本概念 50
- 4.2 主题检索标识的特点 51
- 4.3 概括主题词的一般方法 54
- 4.4 主题概括的步骤 55
- 4.5 课题分析实例 57
- 4.6 主题分析小结 59
- 4.7 测验题 60

第5章 组配检索及其运算符 62

- 5.1 逻辑运算符 62

- 5.2 字段限制符 65

- 5.3 通配符 65

- 5.4 关系运算符 66

- 5.5 位置符 66

- 5.6 逻辑组配原则 67

- 5.7 构造检索策略实例 67

- 5.8 课题分析小结 69

- 5.9 实验二 70

- 5.10 测验题 71

第6章 网络资源搜索 73

- 6.1 网络资源搜索基础 73

- 6.2 独立搜索引擎：百度 84

- 6.3 学术搜索引擎：读秀 87

- 6.4 开放存取及其一站式检索平台 93

- 6.5 实验三 94

- 6.6 测验题 95

第7章 国内综合性检索平台 97

- 7.1 维普资讯 97

- 7.2 中国知网 105

- 7.3 国内开放存取资源简介 114

- 7.4 实验四 115

- 7.5 测验题 118

第8章 国外全文数据库 120

- 8.1 EBSCOhost 120

- 8.2 SpringerLink 137

- 8.3 其他国外全文数据库简介 142

- 8.4 国外开放存取资源 143

- 8.5 实验五 144

- 8.6 测验题 146

第9章 知识产权检索资源 148

- 9.1 知识产权基础知识 148

- 9.2 国家知识产权局网 151

- 9.3 WIPO 知识产权和技术数据库 166

9.4 其他国外知识产权检索系统简介	180	11.5 实验八	223
9.5 实验六	181	11.6 测验题	224
9.6 测验题	184	第12章 图片库	226
第10章 技术性贸易措施检索资源 ...	186	12.1 概述	226
10.1 基础知识	186	12.2 Getty Images	227
10.2 中国技术性贸易措施网	192	12.3 视觉中国	231
10.3 中国标准服务网	195	12.4 相关网站简介	237
10.4 其他相关网站简介	204	12.5 实验九	238
10.5 实验七	205	12.6 测验题	239
10.6 测验题	207	第13章 学术训练基础	241
第11章 统计数据资源	209	13.1 学术训练	241
11.1 统计数据概述	209	13.2 学术论文的结构	244
11.2 国家统计局网	210	13.3 撰写实例	246
11.3 中国经济信息网	217	13.4 实验十	251
11.4 其他数值型资源简介	222	13.5 测验题	253
		参考文献	255

第1章 预备知识

本章重点

1. 信息、知识、文献、情报的概念
2. 文献老化和分散规律在信息检索中的作用
3. 检索效果评价指标的含义
4. 常见国际标准代码的种类和结构
5. 传统文献的类型和特点

1.1 基本概念

1.1.1 信息、知识、文献、情报和信息链

1. 信息

物质、能量、信息是构成自然界和人类社会的三大要素。对于信息，迄今还没有一致的定义。一般认为，信息是物质运动的状态与方式，是事物间相互联系、相互作用的中介。自然界、生物界、人类的行为或体态中都会反映出某种信息，使信息具有普遍存在性、形态多样性和内容杂乱性的特点。由于以上特点，使少量有用信息淹没于大量无用信息中，使之难以识别和直接利用。

2. 知识

知识是人类认识的结晶，是有用的信息，它是来源于实践并经记忆的过滤保留下来的有用的经验或教训。知识借助于一定的语言形式或物化为某种劳动产品的形式传递给下一代，成为人类共同的精神财富。记录知识的形式主要有语音、图形图像和文字。以语音表示的知识的缺点是不易积累和传播，以图形图像表示的知识不易准确理解，唯有用文字记录的知识具有含义稳定明确、便于传播的特点。

3. 文献

根据国家标准定义，图书情报学范畴的“文献”指的是记录有知识的一切载体，即物化了的知识。文献的特点是：以有形的物质为载体，以知识为内容，以文字图形图像为表征知识的符号，可以累积和重复使用。狭义的文献是指静态的、含有文字图表的载体。广义的文献不仅包括传统的印刷品，还包括记载动态的音频、视频信息等非书资料，以及被称为零次文献的手稿、讲话、录音、实验记录等。由于文献的知识信息被固化，记录形态稳定，因而具有便于传播、积累和重复使用的特点。

文献的类型有多种划分方法，如按载体划分、按记载形式划分、按出版类型划分等。图书情报单位一般根据出版类型将文献分为三大类：普通图书、连续出版物和非书资料。属于普通图书的文献有图书、产品资料、技术档案等；属于连续出版物的文献有期刊、报纸、科技报告、会议文献、政府出版物、专利文献、标准文献、学位论文等；属于非书资料的文献有缩微阅读

品、计算机可读产品、声像产品, 以及记录有文字的实物。

4. 情报

有关情报的定义非常多, 我们采纳钱学森的定义, 认为情报是为了解决一个特定的问题所需要的激活了的知识^①。具有定向意识的知识是信息转化为情报的条件, 即人们在接收某些信息时, 只有具备了有准备的头脑, 包括专业知识和意识准备, 这些才有可能被转化为情报。大多数情报信息以文献的形式存在, 具有记录形态稳定可靠的特点。非文献型情报是指听到、看到或通过仪器探测到的信息, 是文献型情报的重要补充。情报具有效果大、时效短、针对性强的特点。信息、知识与情报的关系如图 1-1 所示。

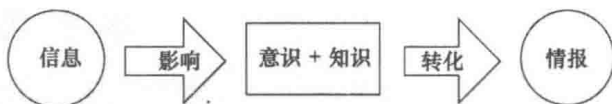


图 1-1 信息、知识与情报的关系

5. 信息链

在你需要信息的时候, 大多数情况下最初面对的是没有经过整理的、散乱的信息。但在这些纷乱的信息中有一些是具有启发意义的信息, 可作为进一步搜索信息的踪迹, 或称为入口信息。这些原始信息具有目标信息的某些特征, 以这些特征为起点继续搜索更接近目标特征的信息, 从而形成一个信息链。情报搜集中的碎片拼图就反映了信息链的特征。信息链会使搜索者的目标图像逐渐清晰起来, 并最终找到需要的信息。比如已知某产品, 想了解该产品在哪些国家申请过专利, 可以去查《世界专利索引》(WPI) 的专利权人索引, 找到相关记录后再根据记录中的存取号查存取号索引, 这样, 经过专利权人代码、存取号两个特征信息就可找到该专利申请在各地的分布情况。

1.1.2 文献老化和分散规律

1. 文献老化的规律

所谓“文献老化”, 是指随着出版年代的远去, 文献逐渐失去了作为情报源的价值, 或者被人们利用得越来越少。文献老化的速度通常用“半衰期”来描述。某一学科文献的“半衰期”是指专业文献在统计年限内 50% 被引文献的发表年代, 即文献量的中值所对应的年份称为“半衰期”。例如, 机械工程文献的中值为 5.2 年, 表示在统计期间内所引用的机械工程文献中有一半是在 5.2 年内发表的, 这个学科的半衰期就是 5.2 年。如果以文献的出版年龄为横轴, 被参考文献量为纵轴, 可用一条负指数曲线来说明文献老化的过程, 如图 1-2 所示。由于可以将参考文献量看作文献利用率的缩影, 而文献利用率是知识活跃度的标志, 当一个专业的文献被利用得越频繁半衰期就越短, 表明该专业知识更新速度很快, 同时表明该专业很年轻;

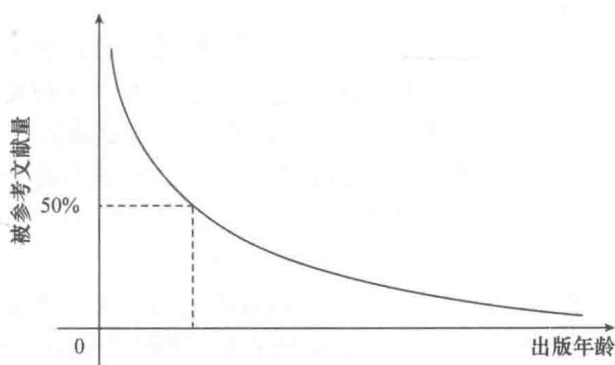


图 1-2 文献老化曲线

① 钱学森. 新技术革命与系统工程——从系统科学看我国今后 60 年的社会革命 [M] // 上海交通大学. 智慧的钥匙——钱学森论系统科学. 上海: 上海交通大学出版社, 2005: 48-63.

反之,半衰期越长,则知识更新速度越慢,专业活力越低。因此可以说,文献老化规律揭示了一个专业的文献利用率与一个专业成熟度的关系。

文献老化规律在信息检索中有重要的作用。首先,它揭示了要重点阅读新出版的文献的内在原因,即新研究成果通常建立在原有结论的基础上,是对旧的理论、观点的修正或补充;其次,它揭示了不同专业成熟度差异的内在原因,启示我们检索中要重视时间条件限定的作用。

统计分析表明,不同学科、不同时期、不同类型的文献半衰期是不一样的,但它们都遵循负指数曲线形态。读者通过统计历年的引文频次、累积引文频次和累积相对比率,可以自行计算出某一学科的文献半衰期。进行文献老化分析除了采用引文统计法外,还可以采取图书馆流通统计方法,但这两种方法都有一定的局限性,需要综合起来考查。

2. 文献分散的规律

论文发表有一个普遍的现象,即关于某一主题的论文往往刊登在多种专业杂志上,其中有少数几种期刊刊载得比较多,大多数期刊只零星登载一些。统计表明,这种文献分散的现象具有一定的规律性。英国文献学家布拉德福最早从定量的角度分析文献分散现象,并发现了文献分散规律,简称布拉德福定律。

布拉德福将相关论文刊载数以递减的顺序排列某专业的期刊,然后采用区域分析、图像观察和数学推导3种不同方法进行分析,得出:可以把期刊分为专门面向这个学科的核心区和包含着与核心区同等数量论文的几个区,这时核心区与各区的期刊数量成一定比例。这个比例系数称为布拉德福分散系数,一般用 a 表示。文献分散规律揭示了期刊利用率与期刊品种的关系。统计表明,不同专业的 a 值是不一样的,一般为5。

假如在统计期间内某一学科的所有660篇论文共刊登在248种期刊上,则期刊分布情况如图1-3所示。

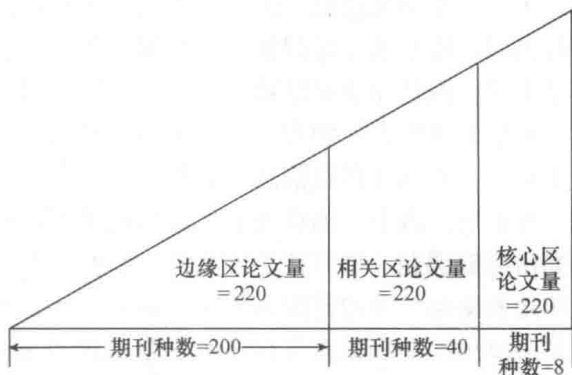


图 1-3 布拉德福期刊分布规律

从图1-3可以看出,核心区期刊论文刊登率最高,每种期刊平均载27.5篇论文,相关区期刊平均载5.5篇论文,边缘区期刊平均刊载

相关论文1.1篇。在各区相关论文数相等的条件下,各区的期刊数与布拉德福分散系数有如下比例关系(其中 n_1 表示核心区期刊数, n_2 表示相关区期刊数, n_3 表示边缘区期刊数):

$$n_1 : n_2 : n_3 = a^0 : a^1 : a^2$$

例如,核心期刊基数为8, a 取5,则各区的比值为 $8 : (8 \times 5) : (8 \times 5^2)$ 。从理论上讲,分区可以不止3个,但超过3个区将与实际情况有较大的误差。

文献分散规律对信息检索的指导意义在于:文献分散规律揭示了论文的层次和相关论文的集聚度与期刊分区的对应关系,即专业期刊利用率与期刊品种的关系,使我们选择阅读核心期刊论文的收获要高于相关期刊,远高于边缘期刊。

1.2 信息资源的选择和检索效果的评价

1.2.1 数据库检索效果的评价

信息检索的效果可从技术效果和社会经济效益两方面来评价。社会经济效益是指检索结果

对用户的技术经济方面的影响,它与用户花费的检索时间和费用有关。技术效果是指对数据库所收文献的覆盖面和摘贮率的满意度,包括查全率和查准率两项指标,它与用户的投入无关。查全率和查准率是评价检索效果的主要标准,它们从不同侧面反映了影响系统检索质量的因素。

“查全率”是指对于相关文献而言,查出的相关文献量占数据库中的全部文献量的百分比。例如,数据库中共有10万条记录,其中有100条相关记录,现在查到20条记录,其中10条是相关记录,即查全率为10%。

“查准率”是指在查出的文献中,查出的相关文献量与查出的全部文献量的比率。根据上例检索结果,查准的比例为10:20,即查准率为50%。

站在检索者的角度看,查全率强调的是覆盖面,要求以一定的误检为代价把需要的文献尽可能地网罗进来;查准率强调的是精准度,要求以遗漏少量相关文献为代价,使查出的文献尽可能多地符合要求。正确理解查全率和查准率之间的关系,对指导具体的检索实践工作有积极的促进作用。最理想的检索效果是查全率、查准率均为100%,而漏检率、误检率均为零。但在实际检索中这是不可能的。英国学者克勒维在1957—1968年间进行的检索试验证明,在检索的查全率和查准率之间存在着互逆关系:提高检出的查全率,就会降低查准率;提高检出的查准率,就会降低查全率。这是因为影响查全率和查准率的因素主要是标引文献主题的广度和揭示文献主题的深度。如果对文献的主题分析得很全面,概括出的关键词很多,相关主题的文献容易被检索出来,查全率就较高,但是检索出的文献有许多不适用,因而查准率就会降低。相反,如果文献标引时,给出的关键词概念很专深,都是固定词组的话,那么检索出的结果当然会很对口,查准率很高,但是命中文献数量会很少,很多相关文献将被过滤掉,从而降低了查全率。反之亦然。有学者认为,在物理、技术科学范围内,当提高查准率1%,将导致查全率降低3%。实践经验证明,在现代科技信息检索系统中,当查全率为60%~70%,查准率则为40%~50%。

在检索实践中,侧重查全率还是侧重查准率与用户课题研究的类型和研究阶段有关。对那些要研究新课题、撰写评述性文章与学术著作,以及申请专利的用户来说,他们往往需要全面、系统地收集某一专题范围内的文献资料,因此对查全率要求较高;而对于在研究过程中需要解决某一具体问题或正在进行技术攻关的用户来说,他们往往只需要了解某一关键内容或一些数值的情报,只要求检索结果能解决他们的实际问题,不一定需要很多文献,因此对查准率要求较高。也有一种观点认为,对于检索失误来说,漏检产生的负面影响比误检更严重,因为误检损失的是筛选时间,降低检索的效率,但并不影响相关文献的利用。因此,漏检是影响检索质量的最主要因素,必须把它减到最低限度。

1.2.2 网络信息资源的选择标准

网络信息资源处于开放的环境中,影响网络信息质量的因素较之传统信息资源更加复杂,并且不同类型网站影响因素也不同,例如新闻类网站和以数据库为主要内容的网站就不同。从影响网站权威性的因素来看,依次是:内容、设计、功能、可查检性、专业化程度、原创性等,其中内容可靠和操作便利是用户利用网络资源最关心的问题。从用户角度看,检索功能、导航系统、链接数量与质量、上网和下载速度、网络资源安全稳定性、上网费用与环境条件、界面友好性等是最为关心的内容。若单就网络数据库来看,信息质量、收录范围、信息来源、资源类型、更新频率、作者标识,以及是否有注释、参考文献和引证资料是重要的评价标准。

自20世纪90年代以来,国内外对网络信息资源评价指标的研究提出了多种方案,大致认为,一个好的网络资源的评价标准至少具有以下特征:①易辨识,即检索范围与信息价值等评价特征标识容易区别界定;②可量化,即能够量化,可以统计;③可操作,即能够以参数和公式的形式进行

计量和比较。表 1-1 是在对国内外网络资源评价标准进行归纳的基础上建立的评价指标体系。

表 1-1 网络信息资源评价指标体系

	一级指标	二级指标		一级指标	二级指标
定性指标	内容	客观性	定性指标	编排设计	导航系统
		实用性			信息资源组织
		全面性			用户界面
		准确性			美感度
		权威性		成本	技术支持
		新颖性			连通速度
		独特性		其他	网络安全
		针对性			信息伦理
		时效性			合法性
		连通性			评论
	易用程度	传输速度	定量指标	数字统计	链接
		稳定性			学科资源分配数量
		连续性			访问次数
		链接正误			订购人数
		检索功能			

(摘自：肖琼. 图书馆网络信息资源评价标准体系探讨 [J]. 情报杂志, 2007 (6).)

1.3 检索系统中的代码

1.3.1 国际标准代码

1. 国际标准书号

国际标准书号 (International Standard Book Number, ISBN) 是用于管理印刷型图书及小册子、缩微资料、盲人读物、混合型媒介出版物、机读制品、幻灯片等专题出版物的唯一性标识，用于标识特定出版社发行的一种著作的一个发行批次。

ISBN 基于国际标准 ISO2108，由国家 ISBN 机构分配，在我国是一个强制性的标准，但国际上是一个非强制性标准。为了与国际物品条码编码 EAN - UCC (EAN 是 European Article Number 的缩写，EAN - 13 是其标准版，EAN - 8 是缩短版) 系统接轨，国际标准化组织 (ISO) 从 2007 年开始将国际标准书号升级为 13 位，为了与原 10 位 ISBN 相区别，简称 ISBN - 13。此后每一种正式出版的专题出版物既有标准书号，又有与商品条码格式一致的标准书号条码，并且还将其使用范围由传统的印刷型出版物扩展到电子出版物等非书资料，特别是互联网出版物上。标准书号与标准书号条码是书号编码的两种识读形式，两者的统一为出版物的发行、国际流通以及馆藏信息化建设提供了基础保证。原有的出版单位在其

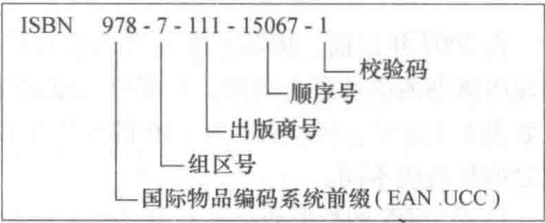


图 1-4 国际标准书号的结构

原有 ISBN 前加上前缀“978”，新成立的出版单位加上前缀“979”，该前缀号由国际物品编码协会分配。国际标准书号的结构如图 1-4 所示。

第一组数字是国际物品编码系统前缀，用于标识各成员单位的代码，例如我国的商品前缀在 690 ~ 695 之间，但图书和期刊作为特殊商品被给予了专门的编号。第二组数字是厂商识别代码，在 ISBN - 13 中称为组区号，主要用于表示文献的出版国、世界重要的地区、国际组织或出版地所处的语区，有些多语种国家则有多个号码（例如瑞士、加拿大等国家）。表 1-2 是部分组区号的含义。从表中可以看出，随着出版量的变化或国家的分合，组区号会有变化。例如苏联的组区号是 5，解体后许多国家有了自己的代码。又如伊朗原来的组区号是 964，现在提升为 6。也有因出版量突破了原有限额而增配一组组区号的，例如中国香港原来是 962，现在又增加了一组 988。

表 1-2 组区号的含义

语区	组区号	国家/地区	语区	组区号	国家/地区
英语	0	英国		5 和 966	乌克兰
		美国		6 和 964	伊朗
		澳大利亚		7	中国大陆
		新西兰		80	捷克
	1	加拿大		81 和 93	印度
		南非		82	挪威
		津巴布韦		83	波兰
法语	2	法国		84	西班牙
		加拿大		85	巴西
		比利时		86	南斯拉夫
		瑞士		87	丹麦
		卢森堡	意大利	88	意大利
德语	3	德国			瑞士
		奥地利		89	韩国
		瑞士	荷兰	90	荷兰
	4	日本			比利时
俄语	5	俄罗斯		91	瑞典
		阿塞拜疆		92	联合国教科文组织
		塔吉克斯坦		957 和 986	中国台湾
		土库曼斯坦		962 和 988	中国香港
	5 和 985	白俄罗斯		99905 和 99954	玻利维亚

在 2007 年以前，国际标准书号用前 9 位数字唯一地表示一种书。如果组区号的位数较短，则说明该书的语种是大语种，从而可容纳的出版商的位数也较多。组区号的位数与出版量的关系如表 1-3 所示。校验码由 0 ~ 10 的数字组成，数字 10 用 x 表示，13 位 ISBN 的校验码与原 10 位时的校验码不同。

13 位 ISBN 的校验码的加权算法是：先用前 12 位数的偶位数之和乘以数字 3，加上奇位数之和求出 S_1 ，然后再用大于或等于结果 S_1 且为 10 的最小整倍数的值减去 S_1 求出 S_2 ，若 S_2 与第 13

位数相同，则前 12 位数输入正确。

表 1-3 组区号与出版量的关系

组区号位数	组区号取值范围	分配组数	每组的出版量
1	0 ~ 7	8	1 亿
2	80 ~ 93	14	1000 万
3	950 ~ 989	40	100 万
4	9940 ~ 9989	50	10 万
5	99901 ~ 99955	54	1 万

2. 国际标准连续出版物编号

国际标准连续出版物编号（International Standard Serial Number，ISSN）是为不同国家、不同语言、不同机构（组织）间各种媒体的连续出版物或集成性资源（包括报纸、期刊、动态指南、年鉴、年报等）信息识别、控制、交换、检索而建立的一种标准的、简明的、唯一的识别代码。ISSN 与识别题名一起唯一标识了一种特定的连续出版物或集成性资源。它基于国际标准 ISO 3297，由 ISSN 网络中心分配，由国际连续出版物数据系统（International Serial Data System，ISDS）的国际中心负责管理和分配号码。ISDS 每次分配给申请国一段数字，这段数字用完之后再重新申请一段。凡是参加 ISDS 的国家或地区向本国或本地区的中心或国际中心提出申请，经过核对和登记后给予 ISSN。申请国获得一段数字序列后，集中按申报刊物的刊名字头给号，每个 ISSN 唯一地与一种期刊的标识刊名相对应，因此一个国家的 ISSN 号可能是不连续的。ISSN 中国国家中心设在国家图书馆内，负责经国家新闻出版广电总局正式批准出版的连续出版物 ISSN 的分配、管理、使用和咨询，中国连续出版物书目数据送交 ISSN 国际中心数据库等工作。

期刊与图书一样，也被纳入了国际物品条码编码 EAN - UCC 系统。基于 EAN13 的 ISSN - 13 条码由 3 组 4 部分数字组成，共 11 位数字。第一组是国际物品编码系统前缀（期刊是 977），第二组数字共 4 位数按刊名字顺排列，第三组前 3 位数字是在字头基础上的流水号，最后一位数字是计算机校验码。在不使用条码时，习惯上仍使用原 8 位 ISSN 编码，因为当不作为普通商品时，8 位编码成为刊物的唯一标识。

11 位 ISSN 校验码的加权算法是：①右起第 2 位（偶数位）开始，所有偶数位的数字之和乘以 3 得 S_1 。②右起第 3 位（奇数位）开始，所有奇数位数字之和乘以 1 得 S_2 。③将 S_1 和 S_2 相加得 S_3 。用 10 减去 S_3 的个位数，所得余数为校验码。当所得的个位数为 0 时，检验码为 0。

3. 国际标准音像制品号

国际标准音像制品号（International Standard Recording Code，ISRC）为识别唱片、录音带、录像带、激光视盘等音像制品的国际标准编号。我国于 1992 年 5 月基于 ISO3901 批准了强制性国家标准《中国标准音像制品编码》（CSRC），自 1993 年起中国所有音像出版社都必须在其生产的每一种音像制品上，对所录入的节目以及节目中每一项可以独立使用的部分编加一个以“ISRC”为标识的国际标准音像制品编码。

4. 数字化对象识别符

数字化对象识别符（Digital Object Identifier，DOI）是由美国出版商协会于 1997 年推出的一种电子出版物内容标识系统，目前 DOI 系统由非营利性的国际 DOI 基金会管理。这里说的数字化对象泛指所有计算机可读文件，包括电子文档、可执行文件、图像文件、音频文件、视频文件，或其他任何文章、图书、课堂实验等在计算机中有地址可寻的数字化资料。设计 DOI 的目的是保护知识产权并实现版权所有者的商业利益。DOI 系统为出版商提供了一个既能用于出版物

的内部管理,又能进行数字化商务的工具。它对文件检索、票据交换、出版许可等电子出版日常的业务程序提供了自动化管理的工作基础。

DOI是数字化对象生存期内具有唯一性的标签,它只与该对象的知识内容有关而与该对象的存储位置无关。一个DOI记录同该数字化对象的具体位置的信息一起被送往DOI中心服务器登记、存储。这种被集中存储起来的数据形成了一个数据分辨库。利用特殊的软件,这一数据库可以链接或分辨某个DOI,使其与同其有关的数字化对象的位置联系起来。当用户寻找一个数字化对象或有关这一对象的信息时,DOI问询就会被传送到DOI服务器,该服务器寻找到DOI的记录及同其有关的对象的地址,将两者链接起来,将其位置(大部分是统一资源定位符)送回给用户浏览器。然后浏览器自己检索这两个对象或有关该对象的信息,将结果显示给用户。

DOI由前缀和后缀两部分组成,中间用一道斜线分开。前缀由识别符管理机构指定,后缀由出版商自行分配。

前缀又由两部分组成,中间用一个圆点分开。第一部分有两个字符,代表分配该DOI的管理机构。目前因为只有一个识别符管理机构,所以总是1、0两个数字。以后会有多家管理机构,例如一个国家一个,或一个行业(如出版、摄影、音乐、软件等行业)一个。前缀的第二部分代表被分配使用该DOI前缀的出版机构,或者是在识别符管理机构登记的版权所有人。

后缀由出版商或版权所有人自行给定,是一组唯一的字符,用来代表特定的数字化对象。许多出版商选用已有的识别符号作为后缀,如ISBN、ISSN等。

实例一:10·1000/1

这是DOI系统在Web上的主页识别符,前缀是10·1000,后缀是1。

实例二:10·1016/S0921877797000232

这是Elsevier出版的一份杂志上一篇文章的识别符。其后缀用的是出版商款目识别符(PII)。

实例三:10·1111/j.1475-7599.2006.00203.x

这是EBSCOhost的期刊论文,1475-7599是ISSN号,2006是论文收稿年份。

实例四:10·1016/S0892-6875(07)00030-1

这是EBSCOhost的期刊论文,0892-6875是ISSN号,07是期刊的出版发行年。

实例五:10·1002/pon.1047

这是EBSCOhost的期刊论文,pon是刊名缩写。

一个出版商既可登记一个前缀,也可为其不同的子公司注册不同的前缀。当出版商将其产品版权转让或出售时,数字化对象的位置就发生了变化,该出版商要将这一对象位置改变的信息送到DOI服务器,以保证DOI文献追踪的有效性。为了保证在被链接的版权所有者的出版时间和地点发生变化的情况下,其识别符仍能保持一种持久性,DOI系统提供了一种识别途径,为读者或用户将相关的资料链接在一起。DOI系统的这种能力,可以使用户从一个出版商的产品(如书目数据库或文章引文)链接到另一出版商对这一数字化对象所做的文摘或全文中,尤其在使用那些在服务器之间来回转换的文件、新版的软件、不同文档格式的数字化音乐作品,或经过修改或增加内容的科学文章中特别有用。

1.3.2 压缩键

1. 压缩键的设计

在进行计算机检索时,输入检索词往往需要频繁击键,采用某种技术后只要键入少数几个字母即可实现原来的操作。这种把完整的检索标识信息压缩成几个键的技术叫“压缩键”技术。

压缩键常用于刊名、公司名、商品名等数据重复出现较多的字段，有时也用于书名、题名、人名等字段。在数据库中运用压缩键技术可以减少计算机储存空间、提高数据排序和检索速度，明显减少检索人员敲击键盘的次数。以下举例说明压缩键的取码原理。

(1) 简易取码法：按每个字的首字母取码，字段设置较长。这种取码方式的好处是方法简单易记，压缩码本身也可以进行前方一致检索，图书馆书目检索系统中的书刊名字段通常具有简易压缩键检索功能，也有将该功能单独设置成书名拼音和作者拼音字段的。其缺点是占用存储空间较多，对虚词赋码增加了无效信息（例见图 1-5）。

(2) 四位取码法：码长固定 4 个字母，虚词不给码。这种方法的好处是，由于压缩为 4 个码，每个码所含的信息高度浓缩，所需存储空间大大减少。固定长度的压缩码要遵循一些规则。例如：4 个字以上的名称，取前 3 个字及最后一个字的首字母；4 个字的名称，取每个字的首字母；3 个字的名称，取第一个字的前 2 个字母和后 2 字的首字母；2 个字的名称，取每个字的前 2 个字母；1 个字的名称，取前 4 个字母或前 3 个字母加最后一个字母（例见图 1-6）。

汕头科技和教育发展公司	SHKJHJYFZGS
上海科技教育出版社	SHKJJYCBS
上海科健软件公司	SHKJRJGS
上海肯捷咨询公司	SHKJZXGS

图 1-5 简易取码法实例

大学图书馆学报	DXTB
图书馆杂志	TSGZ
图书馆工作	TSQZ
图书馆情报知识	TSQS

图 1-6 四位取码法实例

(3) 偏重取码法：码长固定 6 个字母，虚词不给码。取码规则是 321，即第一个字取其汉语拼音的前 3 个字母，第二个字取其前 2 个字母，第三个字取其首字母，当第一个字不足 3 个字母时，前三个字各取 2 个字母。这种取码方法的理论依据是：在字符串中，越是靠前的字重复的可能性越大，同时越是靠前的字所含的信息越重要。这种取码方法的好处是，规则简单易记，重码相对较少，码长有限，占用存储空间不多。图 1-7 是按偏重取码法得到的压缩码。

2. CODEN 码

CODEN（Code Number）是国际上通用的刊名识别代码，是一种用于管理连续性和专题性出版物的编码方法，广泛应用于文献数据库和检索系统。1975 年起，CODEN 由美国化学文摘服务社接手管理，成为国际 CODEN 中心，出版《国际 CODEN 目录》（*International CODEN Directory*），并对各国的重要期刊分配号码。CODEN 码具有唯一性，主要用于期刊联合目录。CODEN 码由 6 个字母或数字组成，前 4 个字母是刊名压缩代码，遵循 Bishop Rule 刊名缩写规则，第 5 个字母是与其他同类刊物的关联字母（一般为期刊类别代码），第 6 位是计算机校验码。Bishop Rule 刊名缩写规则为：刊名为 1 个词时取头 4 个字母，2 个词时取每个词的头两个字母，3 个词时按 112 的规则取字母个数，4 个词时按 1111 规则取字母个数，若一个编码已被占用，则按重码处理规则处理。图 1-8 为 CODEN 码实例。

大学图书馆学报	DAXUTU
情报学报	QINBAX
上海工程技术大学学报	SHAHAG
中国图书馆学报	ZHOGUT

图 1-7 偏重取码法实例

材料保护	CAIBE3
电力自动化设备	DZSHFK
化学工业与工程技术	HGGJFD
人工晶体学报	RJXUEN
软件学报	RUXUEW
石油大学学报（自然科学版）	SDXZE7
太原理工大学学报	TLDXF3

图 1-8 CODEN 码实例

1.4 传统文献的类型

在数字媒体蓬勃发展的今天,传统文献仍以其丰富的积累、众多公开和内部出版的文献类型,以及简便的阅读条件而表现出顽强的生命力。不同来源的文献信息功用互补,能满足不同的文献需求。

1.4.1 文献型出版物

1. 专业图书

专业图书是对现有研究成果、生产技术、劳动经验,或者某一领域知识的整理和概括,其出版特点是内容比较系统、成熟。专业图书包括大中专教材、学术专著、论文集等,供阅读之用。教材一般只介绍基础知识和公认的见解。专著一般就某一主题进行比较深入的探讨。论文集是许多论文或报告的汇编。

2. 专业期刊

专业期刊是一种定期或不定期出版的连续出版物,其出版特点是:每期有若干个主题(除专辑外)、每年至少一期、有连续编号。专业期刊记录了正式的、公开的科学活动,期刊论文对内容的独创性有严格要求,需要经过专家评议。专业期刊是科学交流和情报传递的主要载体和重要的信息源。期刊按内容性质可分为学术性刊物、技术性刊物、快报性刊物、消息性刊物、资料性刊物、检索刊物等几类。

1.4.2 数值型出版物

数值型出版物主要是指各种社会生产活动中积累、汇总起来的统计资料,它是进行经济决策和制订经济计划的重要依据。各国、各地区、各行业甚至各单位一般都出版统计资料,有些是以年鉴形式出版,有些则以期刊或图书的形式出版,或在政府网站或国际组织的网站上公开。例如商业性统计资料《上海生产资料价格信息》《2016年中国机电产品报价目录》,更常见的是读者面较窄的各种专业数据手册,如《建筑施工分项工程常用数据资料与计算速查手册》《简明建筑电气工程师数据手册》等。政府出版的统计年鉴具有综合性、连续性的特点,例如《2016上海统计年鉴》《中国统计摘要—2016》等。统计资料的来源主要有公开出版发行的研究报告、官方统计资料、公共组织统计资料、信息调查机构和咨询公司数据库等。

1.4.3 事实型出版物

事实型出版物包括百科全书、年鉴、名录、辞典、表谱、图录等,其作用是供读者寻找答案,所以统称为参考工具书。

(1) 百科全书。百科全书是概述人类一切门类或某一门类知识的完备工具书,是知识的总汇。它是对人类已有知识进行汇集、浓缩并使其条理化的产物。百科全书一般按条目(词条)字顺编排,另附有相应的索引,可供迅速查检。百科全书大致可分为专科性和综合性两大类。一些著名的百科全书除出版印刷版和光盘外,还推出了在线百科全书,如不列颠百科全书(britannica.com)。特别值得一提的是维基百科全书(www.wikipedia.org),这是一种由网民自发参与编辑、修改的开放式百科全书,由专家进行审核,条目内容解释和外部知识链接丰富,涉及文本、音频、视频等各种文件格式,是更新最快的在线百科全书。

(2) 年鉴。年鉴按年度汇集一定范围内的重大事件、新进展、新知识和新资料,供读者查