

XINXI JIANQUO BAYONG JISHU

信息检索 利用技术

信息检索利用技术编写组 编著

四川大学出版社



责任编辑:徐 燕 张 晶
责任校对:王 平 周 颖
封面设计:罗 光
责任印制:程绍娟

图书在版编目(CIP)数据

信息检索利用技术/信息检索利用技术编写组编著.
成都:四川大学出版社,2001.8
ISBN 7-5614-2171-0

I. 信... II. 信... III. 情报检索-师范大学-教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2001)第 052467 号

书名 信息检索利用技术

作者 信息检索利用技术编写组 编著
出版 四川大学出版社
地址 成都市一环路南一段 24 号 (610065)
印刷 华西医科大学印刷厂
开本 787mm×1092mm 1/16
印张 19
字数 420 千字
版次 2001 年 8 月第 1 版
印次 2001 年 8 月第 1 次印刷
印数 0 001~6 000 册
定价 23.50 元

◆读者邮购本书,请与本社发行科
联系。电话:5412526/5414115/
5412212 邮政编码:610064

◆本社图书如有印装质量问题,请
寄回印刷厂调换。

前 言

一个民族要生存、发展就要创新,要创新就必须充分利用信息和信息资源。20世纪新技术革命的杰出成果不是氢弹、原子弹、隐形飞机等杀人武器,而是以计算机硬、软件技术为标志的信息处理、控制、传输、检索与利用。其中信息检索是信息技术的核心,是开发利用信息资源,进行科技创新的关键。美国将联机检索与网络检索放在技术发展的首位,其根据就在这里。目前世界各国都非常重视信息检索技术的研究、普及与技术。

我国的信息检索与利用教育已有二十多年的历史,其专著与教材数以百计。就编著的内容而言,其专科和文理分开的居多,文理兼容的很少,文理兼容、古今贯通,又具有师范性、综合性的更少。为了弥补这些不足,本书作了一次新的尝试。它的内容包括:概论、事实数据信息检索、综合文献信息检索、专业文献信息检索、计算机信息检索基础、光盘检索、联机检索、互联网检索、信息分析研究与利用等知识。

本书的特点在于既兼顾了传统信息检索方法,又突出了现代信息检索利用技能,古今贯通、文理兼容。在编写手法上,借鉴了国内现有“文献检索与利用”和“信息检索与利用”教材与专著的长处,分合统一、前后连贯、纵横交错,知识衔接上具有现代网络的超链接关系。文理科各专业既相对独立,又相互渗透,有助于扩大学生信息检索与利用的视野,也有助于教师组织教学,给人一种鲜活感和强烈的时代特征感。

本书适于高等师范院校、综合性大学文理各科学学生及其他人员学习“信息检索与利用”时使用。

参加本书编写工作的有黄浩耘、郭黎康、郑惠珍、黄楠、王锦、高晋蜀、张怀绥、张效赤等同志。黄浩耘同志承担了全书大纲,前言,第1章,第3章第1节,第4章第4节、第5节,第7章的编写和全书的统稿工作。郭黎康同志承担了第4章第6节,第5章的编写工作。郑惠珍同志承担了第2章第1节前半部分,第4章第1节、第2节的编写工作。黄楠同志承担了第2章第1节后半部分,第3节,第4章第3节的编写工作。王锦同志承担了第6章的编写工作。高晋蜀同志承担了第3章第2节、第3节的编写工作。张怀绥同志承担了第2章第2节、第4节的编写工作。张效赤同志承担了第3章第4节、第5节、第6节、第7节的编写工作。本书的编写得到了四川师范学院图书馆领导的支持,馆长陈国勇同志审阅了全稿。四川大学西区图书馆研究馆员杜桑海同志为本书写了序言。编写者在此向他们致以衷心的感谢。

限于编写者的水平,本书难免有瑕疵之处,敬请专家和读者批评指正,不胜感激。

信息检索利用技术编写组

2001年6月

读后感—代序

本人有幸作为这本书的第一位读者,谈点个人感想,与大家探讨。

我们所处的时代,是飞速发展的信息时代。信息无处不在,信息无处不需。信息已成为人们生活、事业发展和社会进步的重要因素。如何有效地获取信息、利用信息,关键在于掌握信息检索技术。推广、普及信息检索技术,须从学校教育抓起。高等院校乃至于中学和小学,都需要开设不同层次的与信息检索技术有关的课程,就像重视语、数、外一样,重视这门培养学生能力的课程。师范院校是培养教师的重要阵地,这门课的重要性,就不言而喻了。

四川师范学院文献检索与利用课教研室的老师们编写的《信息检索利用技术》一书,正好适应当前学校教学和社会需要。该书具有以下几个特点:

1)综合性和适用性——本书包括文、理各学科,中外、古今信息检索方面的内容,而且高度概括,融为一体,使读者能较快地掌握获取众多学科领域信息检索技术与知识,适用范围之广,目前在同类书籍中还不多见。

2)全面性和系统性——本书系统地阐述信息检索的原理、方法和利用技巧;同时全面介绍了传统的手工检索方法、现代的计算机检索方法,以及新兴的网络检索方法,完全符合当前高校信息检索课程的教学需要。

3)技术性和实践性——信息检索是一门实践性和技术性很强的课程。本书在信息检索技巧上,下了一番苦功,所举检索实例精当,示范性和可操作性突出,是一种实用性很强的读物。

本书的出版发行,不仅可以满足师范院校“信息检索”课教学需要,同时也适合社会各界了解信息检索技术、利用信息检索技术。编著者们辛勤地编写出版此书,必将对推广与研究信息检索技术起到积极的作用。我深切地祝愿本书的编著者们,在信息检索技术教育的道路上不断努力,开拓创新,继续前进,取得更大的成绩!

四川大学图书馆(西区)研究馆员

杜乘海

2001年7月6日

目 录

第1章 概 论	(1)
1.1 信息	(1)
1.2 信息源	(3)
1.3 信息的序化	(5)
1.4 信息检索的步骤与方法	(22)
第2章 事实数据信息检索	(25)
2.1 字典、词(辞)典的利用	(25)
2.2 手册、表谱、图录、名录的利用	(38)
2.3 年鉴、百科全书的利用	(52)
2.4 类书、政书、资料汇编的利用	(65)
第3章 综合文献信息检索	(73)
3.1 图书信息	(73)
3.2 期刊信息	(84)
3.3 会议、研究报告信息	(90)
3.4 专利文献的检索	(99)
3.5 标准文献	(104)
3.6 政府出版物、产品资料及其检索	(106)
3.7 科技档案的检索	(109)
第4章 专业文献信息检索	(111)
4.1 政治、法律、经济专业信息	(111)
4.2 教育、心理学专业信息	(118)
4.3 语言、文学、历史、哲学专业信息	(122)
4.4 体育、音乐、美术专业信息	(131)
4.5 数学、物理、计算机专业信息	(143)
4.6 化学、生物、地理专业信息	(163)
第5章 计算机信息检索	(191)
5.1 计算机信息检索发展概况	(191)
5.2 计算机信息检索原理	(192)
5.3 计算机检索系统的构成	(193)
5.4 联机检索系统	(195)
5.5 光盘检索	(204)
5.6 计算机检索的方法与步骤	(221)

第 6 章 互联网络检索	(227)
6.1 国际互联网的起源及发展	(227)
6.2 国际互联网的基本功能	(228)
6.3 国际互联网的一些基本概念	(229)
6.4 互联网上的信息资源及浏览	(231)
6.5 国际互联网上的信息资源检索工具概述	(240)
6.6 国内检索工具的使用	(244)
6.7 国外综合性网络检索工具介绍	(261)
6.8 国外集合型网络检索工具	(279)
6.9 其他网络检索工具介绍	(282)
第 7 章 信息分析研究与利用	(283)
7.1 信息搜集与整理	(283)
7.2 信息分析研究的一般步骤与方法	(286)
7.3 信息分析研究成果的撰写	(290)
参考文献	(295)

第 1 章 概 论

1.1 信 息

1.1.1 信息、知识、情报、文献、资料、数据的关系

信息,从广义上讲,是事物运动时发出的信号所带来的消息,是事物存在方式和运动规律的一种表现形式。不同的事物具有不同的存在方式和运动规律,从而构成了各种事物的不同特征。信息普遍存在于自然界、社会界以及人的思维之中,是客观事物本质特征千差万别的反应。

信息分为两大类:自然信息与社会信息。

自然信息,由自然界的物质运动发出,通过声波、电磁波和物质的相互作用表现出来。

社会信息,由人类社会活动发出,通过人的语言、文字、图像、符号、手势以及记录有上述内容的各种载体等形式表现出来。社会信息有语义性。

信息是人们认识事物、了解事物、研究事物的媒介。人们通过感官摄取信息,再利用大脑处理信息,从而获得对事物运动规律的认识。因此,知识可定义为:人们对客观事物运动规律的认识,是经过人脑加工处理过的系统化了的的信息。知识是人类经验和智慧的总结,是人们科学地认识世界、改造世界的力量。知识一经产生出来,就以声像信息、实物信息、文献信息的形式加以传播。人们通过信息获得对事物新的认识,从而产生新的知识。人类知识的完善和丰富的过程,实际上是人类不断获取和利用信息的过程。

情报,是人们为了解决某个具体问题所需的新的知识信息,是知识信息海洋中变化的、流动的、富有生气的一部分。它具有很强的针对性。它在人们的意志、决策和行动的思考中起着参考借鉴作用。

文献,是记录有知识信息的物质载体,它是传播信息、知识、情报的重要手段。记录知识信息的物质载体最早有甲骨、陶瓷、青铜、木板、竹简、绢帛。造纸术发明后,纸张成为记录和传播信息、知识与情报的主要载体。到近现代,知识、信息的载体的形式趋向多元化,除了纸张外,主要还有感光材料和磁性材料。随着声像技术和计算机技术的发展,文献的版本形式越来越多样化,文献的生产速度达到前所未有的程度,大大促进了知识、信息的传播和利用。

资料,从广义上讲,是对人有帮助的物质材料,包括生产资料和生活资料的全部对象。狭义的资料是指能对生产、科研、教学、管理决策起参考借鉴作用的文献,即含有所需情报内容的那部分文献。

数据,是描述事物特征的符号序列。信息的内容靠数据来体现。数据有数字数据、图像数据和文字数据三种类型。

1.1.2 信息与物质、能量的关系

信息不是物质,却源于物质。它是物质系统运动状态及运动变化方向性的表征,是物质系统差异性的表征,是物质系统组织程度、有序化程度的标志。宇宙间的一切物质都在不停地运动、变化,其物质和能量必然要在空间和时间顺序上出现不均匀分布,即出现一定的结构和功能特性,就必然有信息产生。现代科学表明,信息的产生、传递、存储、加工等等,都离不开物质,都要以物质作基础。但信息不是物质。第一,任何一个物体转移到别处去后,原来的地方就不再有这一物体了。而信息则不同,当你拥有关于某一事物的信息传递出去后,你自己拥有的信息并未减少或丢失。第二,任何物体都具有一定的质量,而信息本身却没有质量。

信息的获取、传递和处理离不开能量,但信息也不是能量。第一,信息要靠能量才能传递,但信息的内容及其所起的作用并不取决于传递信息能量的大小。常言道:“有理不在言高”,就是这个意思。第二,能量是守恒的,但信息并不遵守守恒定律。存储在头脑中的信息可能遗忘,信息在传递中受到干扰可能丢失。

1.1.3 信息的特性

信息具有以下特点:

1) 时效性

信息的时效性是指信息发出、接收,到利用的时间间隔及其效率,也包括信息本身更新的速度。时效性是信息的重要特点。如果传递很慢,那么再有用的信息常常也会失去其应有的价值,比如新闻、预告等。

2) 传递性

信息借助于一定的物质载体才能进行传递。传递的过程由6个部分组成:

信源→编码→信息→信息传递渠道(载体)→译码→信宿

编码和译码,是把要传递的信息,用语言、文字、图形、代码、符号等形式,转换成可传递的和可接受的信息,是信息被准确接收的关键。

3) 可扩散性

信息的传递性决定了信息的可扩散性,可扩散性指信息通过各种渠道、媒介进行传播。信息网络的覆盖面越来越大,它的扩散是没有穷尽的。

4) 可扩充性

人们对世界的认识是无限的,因此信息资源的扩充与积累也是无限的。

5) 可替代性

信息的物质载体形态是可以互相转移变换的。

6) 共享性

在信息的扩散和用户分享信息的过程中,信息载体本身的信息量并不因此过程而减少,各用户分享的信息份额也不因分享的人的多少而受影响。

社会进步、信息量的激增和人们对信息需求的激增,使信息成为当今社会生活中的一大支柱,成为一种和能量与物质资源并存的越来越重要的现实资源。信息资源的共享将极大推进人类文明的发展。

1.2 信息源

1.2.1 社会信息源

与信息相对应,信息资源有自然信息资源和社会信息资源之分。前者广泛存在于自然界中,是物质运动和生物生存活动的结果,人们通过自然信息资源的利用认识世界;后者则存在于人类活动之中,是社会的直接产物,是人类从事各种活动的基础。无论是自然信息资源,还是社会信息资源,其开发利用都在社会中进行,因而我们应从社会的资源开发角度研究其结构与开发规律。鉴于人类开发自然信息资源的活动最终将产生相关的社会信息,这说明社会信息资源是信息资源的主体。我们着重研究的也应该是社会信息资源,在此基础上全面探讨信息资源的结构与开发问题。

社会信息源又称为情报资源,以下简称信息源。它存在三种类型:口头信息源、实物信息源和文献信息源。

1) 口头信息源

口头信息源是指人们借以获取口头信息的资源,其基本形式有演讲、交谈、会议讨论、口头广告、信息发布以及人与人通过其他直接接触方式所形成的信息资源(又称为交往信息资源)。口头信息源在利用上具有以下优势:(1)获取费时少,交流速度快、及时,这是其他信息源所不能具备的;(2)具有高度的选择性和针对性;(3)信息反馈迅速、准确,如人们通过直接会晤,对任何问题都可以通过双方的意见交换进行探讨;(4)根据人们接触时的气氛、语气、手势、暗示等,领会文献信息所不能包含的东西,从而易于进行信息利用评价;(5)可以提供其他信息不包含的许多细节,而这些细节往往更具有启发和借鉴作用。但是,口头信息源是有限的,人们接触的机会分布也不均衡,同时对口头信息源的利用难以实行有效的社会监督,不易检验其可靠程度,也不便进行加工和积累,因而口头信息源会随着时间的推移而失真和失效。所以在索取、传递和利用口头信息时,人们又常常将它转换成文献型(如笔记、录音、录像等)。

2) 实物信息源

实物信息源是指固化在实物中的信息来源,属于物质信息源范畴。除实物信息源外,物质信息源还包括自然物质信息源。实物信息源与自然物质信息源的区别在于,它是经过人的加工、生产或创造的产品,是实物中所含的信息财富。实物信息源是人类活动的产物,具有内隐性,其内涵为加工工艺、化学成分、物质参数、设计指标、外观状态等。对这类信息源的利用,只有通过分析研究才能解析出来,从而获取相关工艺或其他信息。实物信息的类型复杂,产品样品及人工创造的所有物质都属于这类信息源。实物信息源的主要优势表现在:(1)真实可靠,通过实物,人们可以直观而生动地掌握其内涵;(2)信息浓度大,内容丰富,人们可以从中掌握生产实物的一系列活动;(3)易于开发利用,通过实物,人们可以充分利用其信息,进行仿制、改进或创新。实物信息源也有其劣势,主要表现在传递不方便,流通渠道不如口头信息源和文献信息源广泛。同时,实物信息源不易于存储,因而往往转换成其他形式的信息源为人们利用。

3) 文献信息源

文献信息源是正规的社会信息源,是人类社会所特有的。凡以文献作为载体形式的信息源统称为文献信息源。文献是人类社会发展到一定阶段的产物,它以文字、图形、符号或其他技术手段记录人类的活动信息和知识信息。按来源状况,文献信息源又分为非出版物信息源和出版物信息源两种。非出版物的文献信息源系指人们的书信、笔记、手书、草图和其他记录品,是一种非正式的文献源;出版的文献信息源对于社会来说是正式的,包括图书、报刊、单件文本、具有特殊功能的印刷出版物、缩微出版物、音像出版物、电子出版物、数据库系统等。文献信息源是一种主要的社会信息源,具有特殊的优势:(1)无论在时间上还是在空间上都是传播信息的最重要的手段;(2)所载信息固定、明确,利于多方面加工和利用;(3)是确认人类活动,进行社会规范的工具。文献信息源的缺陷是表现在传递与交流上不够生动和灵活。

上述三种形式的信息源是可以相互转化的,转化的基本依据是社会需要。口头信息源和实物信息源中的信息如果具有长期开发利用价值,往往会转化成文献形式,而文献信息源中的信息也可用口头形式进行传递,同时文献信息源中的信息细节还可以以实物模式记载,进行模型化利用。

1.2.2 文献信息源的类型

文献信息源因传统习惯,又称文献源。其分类具有如下一些方式。

(1) 按知识记录方式,分为手抄型、印刷型、缩微型、声像记录型、计算机阅读型。

(2) 按流通范围,分为公开文献源、内部文献源以及各种密级文献源等。

(3) 按文献记录的性质,分为资料性、研究性、消息性、法规性、娱乐性和其他特殊性质的文献源。

(4) 按流通的社会方式,分为常规文献源和电传文献源。

(5) 按加工层次,分为一次文献源、二次文献源和三次文献源。一次文献又称为原始文献或信息原件,二次文献又称为检索工具,三次文献是在信息分析研究基础上产生的一类文献(如综述、述评和参考工具书等)。

(6) 按记录知识的学科属性,分为社会科学、自然科学和工程技术文献源,在此基础上还可进一步展开。

(7) 按出版形式,分为图书文献源、报刊和其他连续出版的文献源,报告文献源、标准文献源、产品样本文献源、广告文献源、学位论文文献源、档案文献源等。

(8) 按其他形式(如表达知识的文种、文献拥有者等),还可分为多种文献源。

将文献信息源从不同角度划分成不同类型的目的,在于方便人们有效地开发、管理和利用人类宝贵的文献信息资源。

1.3 信息的序化

1.3.1 信息序化概述

1) 什么叫序

序,即次序,是事物的结构形式,事物或系统组成各要素之间的相互联系,也是人们从时空的角度认识事物、分析事物、研究事物的逻辑基础。它有3个特点:(1)连续性,(2)一致性,(3)规则性。

2) 什么叫有序无序

事物集合中各个事物的排列或联系在某一层次上是有规律的,则称这个事物集合是有序的,否则是无序的。

3) 什么叫整序

将一批事物按某种规则进行组织整理,使其排列和联系具有某种规律性,这种工作叫整序,或者叫序化。

4) 什么叫信息整序

将混乱的无序的信息组织成有序的便于人们利用的信息的工作,叫信息整序,或者叫信息序化。

5) 什么叫文献信息整序

将若干文献信息单元按照某种关键码排成序列,这种工作叫文献信息整序,或者叫文献信息序化。

文献信息单元,依据检索的具体需要划定,小到单个字、词的解释,某一问题的单个数据,大到一篇文献、成套专著或艺术作品的全部内容。

关键码,是能标识文献信息特征的代码。常见的文献信息关键码有文献题名、著者姓名、主题词、类号、登记号、专利号、报告号等。对关键码的要求如下:

- (1) 与原始文献信息单元有直接对应关系
- (2) 能根据自身值构成序列
- (3) 对用户有检索意义

6) 如何对文献信息整序

文献信息整序的具体作法是:对于每一文献信息单元形成其替代文献信息单元的多个复本,将每一文献信息单元抽一个替代文献信息单元复本,按照某一种关键码排成一个序列,然后再将另一批替代文献信息单元复本按照另一种关键码排成另一个序列,直到所需的序列全部产生出来。

替代文献信息单元,是简要描述了原始文献信息单元特征的记录。常见的4种替代文献信息单元是:

- (1) 文摘(或题录)
- (2) 目录中的款目
- (3) 一份文献标引词的集合

- (4) 文献号(登记号、专利号、报告号等)
7) 信息整序的语言学基础和逻辑学基础

(1) 语言学基础

语言是人们进行思维和进行信息交流的符号系统,由词汇和语法构成,以声波和文字两种形式表现其内容。它分为自然语言和人工语言两种。自然语言——人们在进行信息交流中约定俗成的符号系统。人工语言是人类活动中某特定领域使用的符号系统,如旗语、电码、计算机语言和检索语言。检索语言是信息整序和信息检索时使用的符号系统(又称为索引语言、标引语言、系统语言等),通常指的是表示信息内容特征的语言。信息检索的广义含义包括信息存储和检索两个过程。存储信息时,标引人员根据信息的主题概念,通过类号或规范化的语词标示出信息的内容特征存入检索工具或检索系统。检索时,检索人员根据对检索课题的分析,再根据分类表或主题词,确定代表信息需求的类号或语词(类号或语词又称为标识),从检索工具中找出该类号或语词标引的信息。因此检索语言沟通了信息存储和检索两个过程,它是沟通标引人员和检索人员双方思想的桥梁。总的来说检索语言有下列作用:

- ① 保证不同标引人员标引信息的一致性,
- ② 保证检索提问和信息标引的一致性,
- ③ 保证检索结果和检索要求的一致性。

概念是检索语言的基础,它反映客观事物的本质属性和一般特征,但要通过语言表达出来。事物—概念—语言三者之间存在着辩证关系。检索语言与自然语言的根本区别在于,在处理事物、概念、语言三者的辩证关系上,检索语言严格遵守一一对应的原则,即对某一特定事物用某一特定概念、某一特定语言来表达,而自然语言则不需要遵守这一原则。例如:自行车,自然语言可称为单车、脚踏车,但检索语言为了保证这三个概念的一致性,只使用自行车这个概念。检索语言有分类语言,主题词(单元词、标题词、叙词、关键词)语言,代码语言,引文语言等类型。

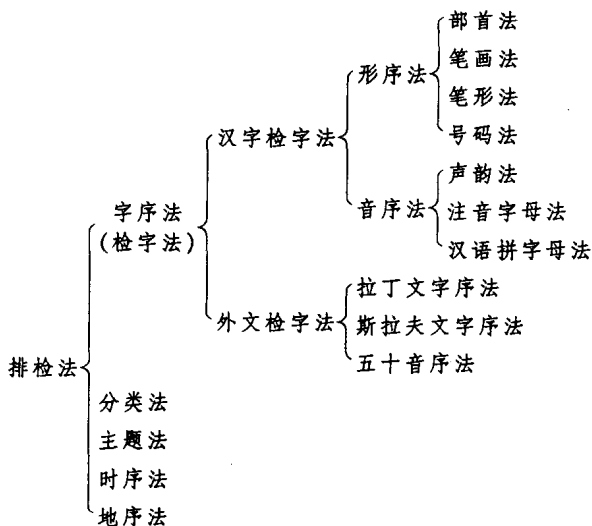
(2) 逻辑学基础

① 概念的划分——一个属概念划分为几个种概念,如物理学划分为力学、热学、电学、光学、原子物理等。被划分的概念为母项,划分出来的概念称为子项。划分的准则有三:a. 每次划分必须按同一划分标准进行;b. 划分的各个子项互不相容;c. 各个子项之和应等于母项。

② 概念的组配(综合)——概念组配不是简单的字面组配,例如:脑+肿瘤→脑肿瘤,河北地方剧+椰子→河北椰子,香蕉味食品+苹果→香蕉苹果。

8) 序化类型

信息的序化主要指文献信息的序化和网络信息的序化,文献信息序化是其他信息序化的基础,这里重点介绍一下。文献信息序化主要用于后面所讲到的检索工具和工具书的排列上,所以有的书籍又称文献信息序化为文献排检法。它的类型大致有5种,即字序法(检字法)、分类法、主题法,时序法和地序法,如下表所示:



以上5类排检法各有各的用途,但并不是每种检索工具或工具书只使用一种排检方法,而是可以同时使用多种排检法。例如《中国大百科全书》,整套书按学科分科出卷,使用的是分类法,而每卷的条目却是按汉语拼音字母法排列;书前编有“条目分类目录”,又使用的是分类法排列;书后附的3个索引:“条目汉字笔画索引”用的笔画法排列,“条目外文索引”按拉丁文字序法排列,“内容分析索引”又按汉语拼音字母法排列。因此,我们要会使用检索工具和工具书,就必须掌握多种排检法。

1.3.2 字序法

1.3.2.1 汉字检字法

它是按一定顺序排列汉字的方法。常用的字典、词典和索引等基本上都采用这种排检法。汉字检字法分为形序法和音序法两种。

1) 形序法

它是按汉字形体差异排检汉字的方法。它具体又分为部首法、笔画法、笔形法和号码法4种。

(1) 部首法

部首法是依据汉字形体结构的特点,按偏旁归类的方法。汉字除少数是“独体字”(如大、人、火)外,大多数是由几部分构成的“合体字”(如花、江、邻)。“艹、氵、阝”就分别是合体字“花、江、邻”诸字的偏旁。把偏旁相同的字归为一部,偏旁就是这一部字的部首。如“江、河、湖、海”诸字的偏旁是“氵”,那么这些字就属“氵(水)部”,“氵(水)”就是部首。

部首法始于《说文解字》一书,为东汉许慎所创。自古以来,许多字典、辞典均采用部首检字法。其中,《康熙字典》代表旧部首法,《辞海》(1979)代表新部首法。

(2) 笔画法

笔画法是根据汉字笔画多少来排列汉字次序的方法。笔画少的字排前,笔画多的字排后。同笔画的字,再按部首或起笔笔形区分。如“托、刚、传、刘、防”诸字,同属六画。若按部首区分,则归入“扌、刂、亻、阝”四个部首中;若按起笔笔形区分,则归入“一、丨、ノ、

、一”五种笔形中。《十通索引》的“单字笔画检字表”，就是按笔画部首法编排的。《辞海》(1979)的“笔画查字表”，就是按笔画笔形法编排的。

笔画法虽然简单易学，但也存在不足，如繁简字体的差异和各人书写的习惯不同，都给确定汉字笔画多少和起笔笔形造成一定的困难。

(3) 笔形法

笔形法是根据汉字起笔笔形的同一性来确定汉字顺序的方法。关于汉字起笔笔形究竟有多少种，说法不一。一般认为，汉字起笔笔形主要有“丶(点)、一(横)、丨(竖)、ノ(撇)、㇇(折)”五种。

将汉字按起笔笔形进行排列的方法有：“元亨利贞”(一、ノ、丨)法，“江山千古”(、丨、ノ、丨)法，“寒来暑往”(、一、丨、ノ)法。上述方法均采用四字成语的各字首笔来排列汉字。

笔形法虽笔形位置固定，方法简单，但终是因人而异，起笔无严格的规范，查检不便。故用此法来编排的工具书很少，多数工具书仅用它作为其他编排方法的补充。如《现代汉语词典》的“部首目录”，先按笔画多少排列部首，同笔画的部首，再按起笔笔形“、一丨ノ”顺序排列。

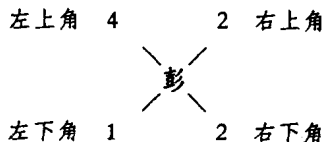
(4) 号码法

号码法是把汉字分解成若干种笔形，用若干个固定数字(代码)表示出来，将这些数字按照一定规则连接成为号码，然后根据号码的大小来确定所代表的汉字的次序。号码法主要有四角号码法和中国字度撇法。其中又以四角号码法流行较广。

① 四角号码法

四角号码法是根据汉字方块字形的特点，将汉字的四角笔形各取号码，按四角号码大小顺序来排列汉字的方法。此法由商务印书馆 1926 年创始，以后又在此基础上产生了新四角号码法。新法与旧法略有不同，因目前大多数工具书均采用新法，故这里主要介绍新法。

新四角号码法把汉字笔形归纳为十种，用“0,1,2,3,4,5,6,7,8,9”十个数字来表示。它按一定规则给汉字的四角笔形取号码，顺序是：左上角、右上角、左下角、右下角。例如：



在上例中，“彭”字的四角号码为“4212”。

四角号码取号口诀是：“横 1 垂 2 点捺 3，叉 4 插 5 方框 6，7 角 8 八 9 是小，点下有横变 0 头。”熟记口诀便于快速应用四角号码法。

在以上取号顺序和取号规则基础上，就按照汉字所代表的四角号码大小次序排列，号码小的排前，号码大的排后。例如“彭”、“郑”两字的四角号码分别是“4212”和“8782”，“彭”排前，“郑”排后。

此外,还有一些比较特殊的取号规则。如缺角取号为0(弓为1702),一笔分角取号(乙为1771),一笔取完后取0(宝为3010),外围是“口、门”的,先取外,后取内,左右两下角取里面的笔形(园为6021,闭为3724)。使用四角号码编排的工具书,要参考书后附的《新旧号码对照表》。在取号上,新法和旧法有些差异(如“大”,旧号码为4003,新号码为4080),不留神容易错。

② 中国字度撷法

中国字度撷法是建国前哈佛燕京学社引得编辑处发明的笔形号码法。“度撷”(guixie)是放入取出的意思,“度撷法”即汉字解剖排列法。

此法将“度撷(撷)”二字笔形分拆开来,即得“丶 一 ノ 十 又 扌 系 丿 目 八”十种笔形,分别用“0 1 2 3 4 5 6 7 8 9”十个阿拉伯数字代表。在对笔形的划分和取号的规则上,与四角号码法不同。它把所有汉字划分成“中、国、字、度、撷”五种体类,分别用罗马数字“I II III IV V”代表。每种字体根据笔形的划分和取号的规则,取出四角号码。各字的取号顺序没有统一,必须根据具体字体而定。

由于该法取号规则复杂,取号次序又不一致,不易掌握,采用者少。现在再版的这一类工具书,大多附有《笔画检字表》作参考。

2) 音序法

音序法是根据汉字读音的语言符号顺序来排列汉字的方法。它分为汉语拼音字母法、注音字母法和声部韵部法三种。

(1) 汉语拼音字母法

汉语拼音字母法是根据1958年公布的《汉语拼音方案》的字母表顺序来排列汉字的方法。这是一种比较科学而又流行的编排方法,已被应用到绝大多数工具书中。如《中国大百科全书》、《新华字典》、《现代汉语词典》等,均按照汉语拼音字母顺序编排。

采用该法编排工具书的条目,首先按汉语拼音的第一个字母排,第一个字母相同时,再按该条目的第二个字母排,其余类推。当读音相同时,则按声调阴平、阳平、上声、去声次序排列。例如:低 dī、敌 dì、底 dǐ、帝 dì。

汉语拼音字母法不受汉字繁简体影响,检索率较高,并且符合国际上工具书的编排规则,同时也适用于计算机检索,故日益受到重视。《辞海》等工具书虽然采用了部首检字法,但同时又附有汉语拼音索引作参考。其不足是,对于不认识或读不准普通话音的字,查检很不方便。

(2) 注音字母法

注音字母法是《汉语拼音方案》公布之前所采用的以注音符号排列汉字的方法。新中国成立前及初期出版的一些工具书,大多按此法编排。查检按此法编排的工具书时,先查声母,后查韵母,同声同韵的则按阴阳上去四声次序查找。自从汉语拼音方案问世,这套注音字母被逐渐淘汰。除港台地区出版的一些工具书仍然采用外,大陆的工具书均不采用。一些原来使用这种方法编排的工具书再版时,也另附有汉语拼音索引。

(3) 声部韵部法

声部韵部法是我国古代按照古声母、古韵母的分类来编排汉字的方法。它分为两种:

① 声部法

声部法是将汉字按古声母分类后进行排列的方法。所谓古声母,是指中世纪人们分析汉语语言构成所产生的用汉字代表的中古时期声母。唐代人根据唇齿舌喉各个发音部位的不同发声方法,归纳出 30 个字母。宋代人增订为 36 个字母。这 36 个字母是:

见	溪	群	疑	端	透	定	泥	知
彻	澄	娘	帮	滂	并	明	非	敷
奉	微	精	清	从	心	邪	照	空
床	审	禅	影	晓	匣	喻	来	日

按古声部法编排的词典性著作有清代王引之编著的《经传释词》等。

② 韵部法

韵部法是将汉字按其读音的古韵母分类后进行排列的方法。古代将汉字按韵母归类,称作“韵部”,按韵部编排的字典称为“韵书”。各个朝代采用的韵部多少不等,其中以“平水韵”最为流行。“平水韵”是南宋末年金人王文郁编的《平水新刊礼部韵略》中所列的韵部。因该书刊行于平水(原平阳府城,今山西临汾市),故名。“平水韵”把汉字按声调上平声、下平声、上声、去声、入声分为五类,每类下按韵目分部。其中上平声 15 韵,下平声 15 韵,上声 24 韵,去声 30 韵,入声 17 韵,共计 106 韵。每一韵部下再按同音字分类排列。按韵部编排的工具书,有《佩文韵府》、《经籍纂诂》、《辞通》等。由于现代大多数人不熟悉这种编排法,查检比较困难,所以再版这类韵书,均附有按笔画或四角号码编排的辅助索引。

1.3.2.2 外文检字法

目前世界上有 2 000 多种语言,多数都有文字,其中使用范围超过 5 000 万人口的语种有 13 种,包括汉、英、法、俄、西、意、日、德、阿拉伯、印度、孟加拉、葡萄牙和印尼语等。为了更好地促进国际交流,掌握各种文献的排检法是十分必要的。

英、法、德、瑞典、荷兰、意大利、葡萄牙、西班牙等国的主要语言都采用拉丁语,即字母从 A 直至 Z 共 26 个字母。德语除上述外,还有 4 个变音字母 Ä, Ö, Ü, B̈。根据国际上的统一规定,它的替代关系是:Ä = AE, Ö = OE, Ü = UE, B̈ = SS。荷兰语也采用拉丁语字母 A 到 Z,在一些外来语中有时 S = Z, F = V, Q = KW。瑞典语也采用拉丁语字母,另加 Ö, Ä, Å 共 29 个字母, Q, W, X 仅用于外来字,由于读音习惯, W 在外来字中可用 V 代替,其替代规则是 Ö = OE, Ä = AE, Å = AA。意大利语也采用拉丁语字母,但除去了 J, K, W, X, Y 5 个字母,只存 21 个,但在表述外来语时又用以上 5 个字母。法语与英语字母完全相同。西班牙语是去掉 W,另加 CH, ñ, 共 28 个字母, W 只用在外来语中。葡萄牙语是去掉 K, W, Y, 共 23 个字母, K, W, Y 只用在外来语中。这类同语系的外文检索工具和工具书都是按拉丁语字母排列的。

日语按五十音图,俄文按斯拉夫语字母次序排列。

1.3.3 分类法

分类法是按照一定的规则,把形式、体裁和内容不同的文献资料,分门别类地组织成科学系统的、便于人们查找利用的文献信息排检方法。它是文献信息管理工作部门常用的传统方法。它具有其他排检法所没有的优点,主要体现在:

第一,便于查找同一科学门类的文献资料,对统计、排架、搜集、出纳、阅览等工作用处较大。

第二,惯用的分类号查寻起来比较快。

第三,分类法是族性检索,检索到所需文献,会顺带找出许多相关文献。

第四,分类法系统性强,无论分类统一不统一,找到有关类,都可以找到有关文献,因而通用性强。

第五,分类法类目有上、下位关系,处理文献能力较强,并且目录较单一,便于管理。

当然,分类法也存在着不少缺点。如不懂分类法的人查找起来会遇到困难,对范围很细小的问题,不一定能反映出来,即使反映出来,也不够突出,容易埋藏在大的类目下,而不受到注意等等。

按其用途,分类法可分为排架分类法和书目分类法。使用分类法将文献资料按照等级体系顺序排列在书架上,称为排架分类法(或书架分类法)。使用分类法将文献资料按照学科的等级体系顺序排列在目录或书目等检索工具中,称为书目分类法。

按其性质,分类法又可分为两类:一类是按学科系统分类,即用等级列举的方法,层层依次展开各学科类目。一类是按事物性质分类,即按事物的属性归列类目。

1.3.3.1 按学科体系排列法

1) 六分法

西汉末年(公元前6年),刘向、刘歆父子编撰的《七略》是我国第一部图书分类目录。它所创立的分类体系,是我国最早的图书分类法。它分为:辑略、文艺略、诸子略、诗赋略、兵书略、数术略、方技略。辑略是概论,不是一类。其余六略则是6个大类。各略下又分小类,共计38个小类。这就是我国古代图书分类法中的“六分法”。

《七略》分类表排列如下:

六艺略:易、书、诗、礼、乐、春秋、论语、孝经、小学。

诸子略:儒家、道家、阴阳家、法家、名家、墨家、纵横家、杂家、农家、小说家。

诗赋略:赋一、赋二、赋三、杂赋、歌赋。

兵书略:权谋、形势、阴阳、技巧。

数术略:天文、历谱、五行、蓍龟、杂占、形法。

方技略:医经、经方、房中、神仙。

2) 四分法

魏晋时期,西晋荀勖在《中经新簿》中改“六分法”为“四分法”,将图书类分为“甲、乙、丙、丁”4部。甲部是经书,乙部是子书,丙部是史书,丁部是文集。后来东晋李充对此作了修订。他在所编的《四部目录》中,把乙、丙部位置互换,就形成了甲(经)、乙(史)、丙(子)、丁(集)的分类序列。从此,按“经、史、子、集”对图书进行分类的“四分法”就成为我国古籍分类的主要方法,一直为后世所沿用。清乾隆时期编撰的《四库全书总目》,则是“四分法”达到极其完善的重要标志。如果要检索和利用古人为我们留下的丰富的文化遗产,就必须熟悉“四分法”,就应该了解“经、史、子、集”所收古籍的范围。

经部收录的“经书”,通常是指为历代封建统治者所推崇的《诗经》、《尚书》、《周易》、《周礼》、《仪礼》、《礼记》、《春秋左氏传》、《春秋公羊传》、《春秋谷梁传》、《论语》、《孝经》、