

普通高等教育“十二五”应用型规划教程

信息资源检索与利用 实用教程

主 编 马春晖

副主编 周晓丽 李艳芬



國家圖書館出版社
National Library of China Publishing House

普通高等教育“十二五”应用型规划教材

信息资源检索与利用实用教程

主 编：马春晖

副主编：周晓丽 李艳芬



國家圖書館出版社
National Library of China Publishing House

图书在版编目(CIP)数据

信息资源检索与利用实用教程/马春晖主编. --北京:国家图书馆出版社,2016.9
ISBN 978-7-5013-5903-5

I. ①信… II. ①马… III. ①信息检索—高等学校—教材 IV. ①G254.97

中国版本图书馆CIP数据核字(2016)第191246号

书 名 信息资源检索与利用实用教程
著 者 马春晖 主编
责任编辑 金丽萍

出 版 国家图书馆出版社(100034 北京市西城区文津街7号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真) 66126156(门市部)

E-mail nlcpress@nlc.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京玥实印刷有限公司

版 次 2016年9月第1版 2016年9月第1次印刷

开 本 787毫米×1092毫米 1/16

印 张 20

字 数 280千字

书 号 ISBN 978-7-5013-5903-5

定 价 48.00元

前 言

信息检索课是培养大学生信息素质的重要途径,也是一门知识性和实践性较强的课程。随着信息技术的广泛应用、数字信息资源的迅速增加,信息检索课的内容也越来越丰富和充实。如何使高校学生更快更好地适应日新月异的数字化新环境,改变学生的学习观念和教育模式,培养学生的信息能力和创新能力,是我国大学课程与教学改革必须着重研究和思考的重大课题。基于此,在充分调研现有同类教材和高校信息资源检索与利用教学实际的基础上,我们编写了这部《信息资源检索与利用实用教程》。

《信息资源检索与利用实用教程》以提升信息素养和强化信息检索技能为目标,通过对信息与信息资源相关概念以及信息检索原理的介绍,使读者掌握信息检索的基础知识和基本方法。同时,从广大师生的实际需求出发,尽可能地提供国内外最新的信息资源的发展情况,介绍其形式与内容、提供检索与利用的方法,并注重将基本理论的叙述与实际操作应用有机结合起来,列举大量检索实例,图文并茂。本教材区别于一般的同类教材,将数据库按照文献内容所属学科分为化学、食品科学、计算机、商管财经、法律、艺术传播等专业学科数据库,结合实例详细介绍了数据库的检索步骤和方法,更加突出实用性、指导性和时效性。此外,本教材不仅对于开放获取资源和学术发现系统等进行了重点推介,而且对于现代图书馆的信息资源平台、网络学科与课程导航等服务进行了系统阐述,尽可能地反映近年来知识发现、数据挖掘及其应用所取得的新成果,期望能为相关的研究者和实践工作者迅速获取所需的知识和信息提供帮助。每章附有思考题,既启发读者进行探究式学习,又便于教师开展教学工作。

本教材共分十二章。主要包括信息检索概述、数字图书馆信息资源利用、全文型文献数据库检索、引文数据库、电子图书检索、特种文献检索、多媒体学习资源检索、自然科学类和人文社科类文献数据库检索、网络免费学术资源检索与利用、常用个人文献管理工具、文献检索综合利用等内容。既可作为高等院校文献检索课程的教材,又可作为图书馆与信息管理从业人员及科研人员使用的指南。

本教材由马春晖任主编,周晓丽、李艳芬任副主编,分别负责全书的组织策划、大纲拟定、体例设计和修改定稿工作。本教材的内容由工作在信息检索教学和信息参考咨询服务第一线的老师合力撰写完成:宋维平编写第一章,程宏伟编写第二章,熊欣欣编写第三章,宁兆椿编写第四章,刘芳编写第五章,刘亮编写第六及十章,宰姣姣编写第七章,张南编写第八章,滕蔓编写第九章,张焯编写第十一章,周晓丽编写第十二章。熊欣欣负责全书的统稿工作。

本教材在内容组织、整理和写作过程中,参阅了大量的相关文献,包括同行学者的学术

论著以及相关数据库与检索系统的培训课件和使用指南等,在此致以谢忱!文后参考文献或未罗列详尽,尚请见谅。

由于编写匆忙,加之教学急用,限于编者水平,本书难免有疏漏和不妥之处,真诚欢迎各位专家和读者批评指正,以便在本书修订时加以补充、更正和完善。

编者

2016年6月于北京

目 录

前 言	(1)
第一章 信息检索概述	(1)
第一节 信息与信息资源	(1)
第二节 信息检索原理	(6)
第三节 检索工具类型与检索方法	(10)
第四节 信息检索策略的制定与调整	(14)
第二章 数字图书馆信息资源	(18)
第一节 数字图书馆概述	(18)
第二节 文献检索系统	(21)
第三节 数字图书馆服务	(31)
第四节 数字资源利用中的知识产权保护	(39)
第三章 全文型文献数据库	(43)
第一节 中文期刊全文数据库	(43)
第二节 外文期刊全文数据库	(62)
第三节 中外文报纸数据库	(75)
第四章 引文数据库	(85)
第一节 国内著名引文索引数据库	(85)
第二节 国外著名引文索引数据库	(92)
第五章 电子图书	(105)
第一节 电子图书数据库概述	(105)
第二节 常用中文电子书数据库	(106)
第三节 常用英文电子书数据库	(114)
第六章 特种文献	(118)
第一节 常用学位论文数据库	(118)
第二节 常用会议论文数据库	(120)
第三节 标准文献数据库	(123)
第四节 专利文献数据库	(133)

第七章 多媒体学习资源 ·····	(141)
第一节 多媒体学习资源概述·····	(141)
第二节 多媒体学习考试平台·····	(142)
第三节 视频音频资料·····	(149)
第八章 自然科学类文献数据库 ·····	(161)
第一节 化学类文献数据库·····	(161)
第二节 食品类文献数据库·····	(176)
第三节 计算机类文献数据库·····	(190)
第九章 人文社科文献数据库 ·····	(203)
第一节 商管财经类文献数据库·····	(203)
第二节 常用法律数据库·····	(224)
第三节 艺术传播类文献·····	(236)
第十章 网络学术资源 ·····	(244)
第一节 开放获取资源·····	(244)
第二节 学术门户网站·····	(248)
第三节 慕课·····	(251)
第四节 其他网络学术资源·····	(255)
第十一章 常用个人文献管理工具 ·····	(261)
第一节 EndNote 文献管理软件·····	(261)
第二节 NoteExpress 文献管理软件·····	(271)
第三节 其他文献管理软件概览·····	(277)
第十二章 文献检索综合利用 ·····	(280)
第一节 本科毕业论文与文献检索·····	(280)
第二节 研究生论文选题与开题前的文献检索·····	(289)
第三节 利用 Web of Scence 平台选择投稿期刊·····	(304)

第一章 信息检索概述

第一节 信息与信息资源

一、信息与信息资源基本概念

(一) 信息的概念与基本特征

1. 信息的概念

信息,在我们的生活中无处不在。根据《现代汉语词典》的解释,信息可以表述为“音信、消息”,这也是最常见的用法。作为学科术语,信息是一个争论颇多的尚无定论的概念,但又是一个对许多研究领域,包括自然科学和社会科学领域,都有着十分重要意义的概念。因其形式与内容的复杂性,在不同的学科领域,有不一样的解释。

(1) 图书情报领域的认识。美国学者巴克兰德(M. Buckland)(1991)认为,当我们今天谈到信息系统的时候,我们是在“事物(thing)”的意义上使用“信息”一词的,因为信息系统是围绕记录、文本、数据等事物而运行的。也就是说,信息可以定义为事物或记录(record)。巴克兰德进一步归纳,许多事物都可以是信息,文本固然是信息,图片、录音磁带、博物馆陈列品、自然物体、实验、事件等也是信息。^① 总之,只要环境条件许可,任何事物都可以是信息。巴克兰德的信息定义很实用,但过于宽泛,它未能区别信息与信息载体以及信息与信息源。

另一位美国学者萨克利夫(J. Tague-Sutcliffe)在其专著《信息测度》(1995)中从信息服务的角度对信息进行了界定。他认为,信息是人和人所生产的记录跨越时空与其他人所交流的内容。信息是依赖于人类的概念化和理解能力的无形的东西,对于记录而言,它所包含的信息对于读者(或用户)则是相对的。信息是读者通过阅读或其他认知方法处理记录所理解的东西,它不能脱离外在的事物而独立存在,它是与文本和读者以及记录和用户之间的交互行为相关的,是与读者大脑中的认知结构相对应的东西。^②

西班牙学者库拉斯(Emilia Curras)则在给 *International Forum on Information and Documentation* (《国际信息和文献工作论坛》)(1993)的一封信中谈了自己对信息的认识。她首先用文学的语言对信息进行了描述:信息什么也不是但又是任何事物,没有它生活是不可能的;它是人类活动的内驱力和能量。信息可以被传递、被感知和被理解,它需要有形的载体以变为实实在在的信息。信息是一种现象和一个过程,前者是指无意识感知的信息,用来调

^① Buckland M. Information as thing. *Journal of American Society of Information Science*, 1991, 42(5): 351—360.

^② Tague-Sutcliffe J. *Measuring Information: An Information Services Perspective*. San Diego: Academic Press, Inc., 1995: 18—21.

整我们的知识状态和态度;后者是我们需要和寻求的信息,是从文献中的数据经处理而来的。^①

(2)通信和信息科学领域的认识。信息的概念最早出现在通信领域,指通信系统传输和处理的对象,一般指事件或资料数据。1988年,钟义信在其出版的《信息科学原理》一书中指出,信息是事物运动的状态与方式,是事物的一种属性。信息不同于情报,情报通常是指秘密的、专门的、新颖的一类信息,可以说所有的情报都是信息,但不能说所有的信息都是情报。信息也不同于知识,知识是认识主体所表述的信息,是序化的信息,而并非所有的信息都是知识。^②

(3)心理学领域的认识。从某种意义上说,心理学所研究和处理的也主要是信息现象。加拿大学者桑盖特(W. Thorngate)(1995)认为,心理学家在定义什么是信息方面,与其他人同样感到困难,但他们至少能说明信息不是什么——信息不是知识。信息是存在于我们意识之外的东西,它存在于自然界、印刷品、硬盘以及空气之中;知识则存在于我们的大脑之中,它一般与不确定性(uncertainty)相伴而生,我们一般用知识而不是信息来减少不确定性。信息是外在的,是不依附于人的意志而存在的;知识则不然,它是人类大脑的产物。

(4)信息资源管理领域的认识。信息资源管理是现代信息技术在管理领域的应用所激发的一种新的信息理论,其对信息的认识更多地受计算机领域的影响。美国学者史密斯(A. N. Smith)和梅德利(D. B. Medley)在其所著的《信息资源管理》中认为,信息是数据处理的最终产品。具体地说,是经过收集、记录、处理,以可检索的形式储存的事实或数据。^③

通过比较各学科领域的定义,我们认为,作为与物质、能量同一层次的信息的定义,取钟义信所归纳的信息是事物运动的状态与方式最为适宜。因为这个定义具有最大的普遍性,不仅能涵盖所有其他的信息定义,而且通过引入约束条件,还能转换为所有其他的信息定义。^④此外,不同领域对信息及其相关概念的认识不是对等的,心理学家的信息定义不同于图书情报领域的信息定义,也就是说,我们在应用不同的理论时,首先应弄清有关的概念并予以区别。

2. 信息的基本特征

(1)普遍性。信息是普遍存在的,是自然界、人类社会和思维领域中客观存在的基本现象。只要有事物存在,有事物运动,就会有信息存在。特别是现代社会,人的一切行为和社会活动,都离不开信息的传递和交换。

(2)客观性。信息是事物及其运动状态和规律的特征,是客观现实的反映。信息是物质的一种属性,如果被人人为地篡改了,那么信息就会失去它的价值。信息本身的产生和存在都不受人的主观意识而改变,但是它可以被感知、储存、传递和利用。

(3)无限性。物质的世界是无限的,人们对于物质的认识也是无限的,认识过程中产生

^① Emilia Curras. The influence of systems science on the concept of information. International Forum on Information and Documentation, 1993, 18(2): 32.

^② 钟义信. 信息科学原理. 福州:福建人民出版社, 1988: 41—45.

^③ Smith A N, Medley D B. Information Resource Management. Cincinnati: South Western Publishing Co., 1987.

^④ 孟广均. 信息资源管理导论. 北京:科学出版社, 2000: 33—36.

的信息还是无限的。信息是一切事物运动的状态和方式,事物的发展变化是无限的,因此信息具有无限性。

(4)时效性。信息的实效性是信息的重要特征。信息具有使用价值和交换价值,它可以为人们所认识和利用,但其价值会随着时间的流逝而减少。因此只有及时将信息加工、收集、传递,才能发挥信息的时效性,使之对人类生活和社会发展产生积极的作用。

(5)依附性。信息本身不是实体,没有体积和重量。因此信息在产生和传输的过程中,必须依附在一定的载体上存在(如纸张、声波、电磁波、化学材料、磁性材料等),并且同一信息可以有不同载体。信息离开语言、文字、图像、符号等记录手段就不能表达,但信息的内容却不会因为记录手段或物质载体的改变而发生变化。

(6)可传递性。信息可以跨越时空进行传递,能够从源物质转移到另一物质,并能脱离源物质而存在。实际上,信息的获取必须通过各种介质或载体的转换在不同的主体之间传递。语言、文字、表情、动作等都是人类常用的信息传递方式。

(7)共享性。这是信息的一个重要特征。信息在传递过程中,并不是此消彼长。同一信息不仅可以被多个主体同时共有,而且还能够被无限地复制、传递。

(二)信息资源的概念与基本特征

1. 信息资源的概念

信息已与能源、材料并列为当今世界三大资源。信息资源(information resources)在现代社会当中起着举足轻重的作用,谁掌握了信息,谁就能更有效地利用物质和能量资源,在竞争中处于优势地位。鉴于信息资源的重要性,我们有必要了解一下信息资源的概念。

关于信息资源的概念,由于学科的不同或研究者本身的理解不同,学术界尚未形成统一的想法,但一般从狭义和广义两种角度来认识和理解信息资源的含义:狭义的信息资源只限于信息本身,而不包括其他因素,它指人类社会经济活动中经过加工处理的、有序化并大量积累的有用信息的集合,包括文字、声像、印刷品、数据库、网络信息等。广义的信息资源是指信息活动中各种要素的总和,既包括信息本身,也包括与信息相关的人员、技术、设备和资金等各种因素。

信息与信息资源之间,既有联系,又有区别。首先,信息是构成信息资源的根本要素,但信息并不等同于信息资源,它的外延比信息资源更广。其次,信息在经过有序化加工处理后,大量累积才能成为信息资源。没有控制、没有组织的信息是没有价值的,也不能够被称为信息资源。

2. 信息资源的基本特征

当今社会,科技的发展和人们的日常生活学习,都离不开对信息资源的有效利用。与自然资源、物质资源相比,信息资源具有流动性、多样性、无序性等特点,能够被重复使用,人们对其检索和使用不受时间、空间的制约。信息资源是有用信息的集合,它具备信息的基本特征。同时,信息资源作为一种经济资源和具有特殊配置形式的社会资源,还有以下特性:

(1)共享性。信息资源是人类智慧的结晶,可以被共享和重复使用。爱尔兰剧作家萧伯纳曾说过:“你有一个苹果,我有一个苹果,彼此交换一下,我们仍然是各有一个苹果。如果你有一种思想,我也有一种思想,我们相互交流,我们就都有了两种思想,甚至更多。”这段话形象地说明了信息的共享性。显而易见,信息在传递过程中,并不是此消彼长。同一信息可以被多个主体同时共有,而且还能够被无限地复制、传递。

(2)时效性。信息资源比其他任何资源都更具有时效性。在情报科学领域,常用“半衰期”来衡量某学科文献老化的速度。例如,美国人巴尔顿(R. E. Burton)和凯普勒(R. Kebler)利用引文数据计算出医学学科的半衰期仅为3年,这意味着现有正在被使用的该学科文献中有50%是在3年内发表的。3年后,该学科发表的文献就会因内容老化而失去参考价值。^①

(3)知识性。信息资源是人类所开发与组织出来的信息,是脑力劳动和认知过程的产物。人们在一定的知识水平条件下,吸收外来的信息,引起思考,对其进行加工和整理,并加以利用。可见,信息资源是人类知识利用的集中体现。不同的信息资源反映了不同社会或者地区的知识水平。

(4)不均衡性。信息资源的不均衡性表现在人们对其掌握程度和其区域分布的差异。由于不同的个体存在认识能力、知识储备和信息环境等方面的不同,导致所掌握的信息资源也存在差距;而随着社会发展不平衡的加剧,经济发达地区占有了大量的信息资源,在一定程度上造成了知识鸿沟的形成。

二、文献信息资源类型

(一)文献的概念

“文献”一词,在古代指典籍及熟悉掌故的人(贤者),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,宋不能征也。文献不足故也。”南宋朱熹《四书章句集注》认为:“文,典籍也;献,贤也。”所以这时候的“文”指典籍文章,“献”指的是古代先贤的见闻、言论以及他们所熟悉的各种礼仪和自己的经历。随着时代的发展,文献的含义仅侧重于“文”,一般泛指具有史料价值的图书和档案,“贤”的意义逐渐消失了。

现在的文献有狭义和广义之分。狭义的文献指具有历史保存价值和现实使用价值的书刊文物资料,包括各学科主要的书刊资料,以及历史文物档案资料。广义的文献指记录、传递有知识和消息的一切载体。目前,关于文献的权威性定义有两个:中国国家标准《文献著录总则》(标准号:GB3792.1—1983)指出,文献是记录有知识的一切载体。国际标准化组织制定的《文献情报术语国际标准》(标准号:ISO/DIS5127)对文献的定义是:“在存储、检索、利用或传递记录信息的过程中,可作为一个单元处理的在载体内、载体上或依附载体而存储有信息或数据的载体”。

(二)文献的要素

简言之,文献就是记录有知识和信息的一切载体。根据定义可以看出,文献是由知识、载体、记录3个要素组成的。知识是文献的信息内容,是文献的核心与灵魂;载体是文献的存在形式,如龟甲、兽骨、纸张、胶片、光盘等,知识信息必须依附于一定的物质载体才能得以保存和传播;记录是二者之间的纽带,知识信息要借助于文字、图形、音频、视频记录下来才能被人认知。因此,离开3个基本要素中的任何一个,均不能构成文献。

(三)文献的分类

文献的类型多种多样。按编写或出版形式,可分为图书、连续出版物、会议文献、专利文献、科技报告、学位论文、政府出版物、技术标准和规范、产品样本说明书、技术档案等。按文

^① 王绍平. 图书情报词典. 北京:汉语大词典出版社,1990:33.

献的加工深度,可分为零次文献、一次文献、二次文献、三次文献。按文献的载体形式,可以划分为印刷型、微缩型、视听型、机读型文献等。

三、文献信息源的获取途径

(一) 文献信息源的概念

文献信息源是指用一定的记录手段将系统化的信息内容储存在纸张、胶片、磁带、磁盘和光盘等物质载体上而形成的一类信息源。

(二) 文献信息源的获取途径

1. 文献阅读

文献阅读是了解文献信息源基本信息的一条主要途径。通过阅读的方式,我们可以从各种类型的图书、报纸、杂志、简报及其他文献资料中搜集文献信息源的相关信息。如杂志中的推荐书目或书评通常会提供一些图书的出版信息及内容简要;某些文章会推荐和介绍一些学科的信息资源,这些资源可能是一本工具书、一个数据库或者一些网站;各种综合和专业的文献检索教材中,也有大篇幅的信息资源及检索工具的介绍。

2. 浏览图书馆等信息服务机构的网站

图书馆等信息服务机构一般都会在自己的网站上通过各种方式向读者揭示本机构所拥有的各种信息源。例如印刷型的书刊信息可通过馆藏书目系统检索获取,也可通过新书通报等方式了解;数据库则一般会按语种、字顺或学科分类列出,并附上简单的资源介绍,方便读者选用。此外,许多信息服务机构还精选了互联网上优秀的免费资源,加以组织之后推荐给读者使用。因此,浏览这些信息服务机构的网站将能快速地获得众多文献信息源的相关信息。

3. 参加会议、展览或培训

了解文献信息源的基本信息,我们还可以通过参加会议、展览和培训的方式。如参加学术会议,掌握本学科的最新研究成果及其出版情况;参加图书展销会,了解相关图书的基本信息;参加图书馆举办的培训活动,了解某些数据库资源的基本情况和使用方法。

4. 咨询专业人员

通过阅读和参加会议的方式获取的文献信息源信息毕竟有限,有的时候,咨询专业的检索人员,全面了解某个领域的文献信息源基本情况,有针对性地收集资料,也是一条比较快捷的途径。咨询专业人员,可以通过三种途径:一是咨询本学科的专业人员、老师、同行或朋友,二是咨询图书馆的专业检索人员,三是参加图书馆或数据库供应商针对学科举办的专业资源介绍讲座。

5. 文献检索

通过各种手工检索和网络检索工具,获取信息源的基本信息。如通过《全国总书目》可以查找国内出版的各类图书的基本情况;通过《全国期刊联合目录》,可以了解期刊的出版信息及馆藏信息;通过网上搜索引擎,可以搜索各学科信息源的基本情况。

另外,还可以通过调查、参观等多种形式,搜集文献信息源的基本信息。

第二节 信息检索原理

检索,源于英语词汇“retrieval”,本意为查询、找回。随着信息技术的发展,信息检索(information retrieval)开始广泛在图书馆的参考咨询工作中展开。

一、信息检索基本原理

(一)信息检索定义

信息检索,是指将信息按一定的方式组织和存储起来,并根据信息用户的需求找出所需信息的过程和技术。也就是说,广义的信息检索包括“存”和“取”两个环节,又称“信息存储和检索(information storage and retrieval)”。

信息存储是指收集大量无序的信息,根据数据源的外部特征和内容特征,由信息标引人员依据检索语言进行分类、标引、编制,使之有序化、系统化,成为可以使用的信息集合。

从狭义上讲,信息检索仅指检索这一环节,本质上是一个匹配的过程,即从信息集合中找出所需信息的过程。通常所说的信息检索一般指狭义的概念,本文所涉及的也是指狭义的信息检索。

(二)信息检索原理

在储存信息时,信息著录和标引人员首先要对各种原始信息进行分析,把原始信息中包含的信息内容分析出来,形成若干能代表该原始信息主题的概念,并用检索语言(即标引标识)把这些概念标引出来,然后按一定规则存入检索工具或系统,形成信息集合。信息检索的过程是信息存储的逆过程。用户检索时,首先对自己的所需信息进行分析,形成若干主题概念,把这些概念转换成检索语言,用检索语言表达信息提问,然后从检索工具或系统中查找相关信息进行匹配。

存储时,文献信息工作者将大量的文献、数据、事实资料以一定的格式输入到计算机软件系统中,通过系统的多种分类检索功能组成可供检索的数据库。检索时,将检索提问词组成检索式输入计算机内,计算机将检索提问词与存储系统的数据进行匹配运算,输出符合需要的检索结果。

(三)信息检索的类型

信息检索的划分标准多种多样,可以根据检索对象、检索方式、检索性质等进行划分。

1. 根据检索对象或检索内容划分

(1)文献检索(document retrieval)。指以文献为检索对象的信息检索,它是利用检索工具和系统查找文献线索、获取文献信息的过程,是传统上信息检索的核心部分。文献检索的结果可以是文献线索也可以是原文全文,但最终目的是得到原文,或至少得到能反映原文的题录或文摘。因此可以将文献检索分为书目检索和全文检索。书目检索指以文献线索为检索对象的文献检索;全文检索指以文献的全部信息作为检索内容的文献检索,即检索系统存储的是整篇文章、整部图书的全部内容,而不仅是关于文献的线索或其主要内容的介绍。随着计算机信息检索的发展,全文检索的地位越来越重要,大量数据库提供文献全文或至少提供订购、获取原文的快捷途径。

(2)数据检索(data retrieval)。指以数据为特定检索对象的检索,即查找文献中的某一数据、公式、图表以及某一物质的化学分子式。数据检索是一种确定性检索,检索的结果是经过核实、整理的数据信息,是用户可以直接利用的信息,例如电话号码、银行账号、各种参数、统计数据等。可利用的工具主要有百科全书、词典、年鉴、手册、黄页以及其相对应的数据库和网络资源。

(3)事实检索(fact retrieval)。指以客观事实为对象的检索活动,利用参考工具书、数据库等检索工具从存储事实的信息系统中查找特定事实的过程,包括检索事物的性质、定义、原理以及发生时间、地点、过程等。事实检索与数据检索一样,是一种确定性的检索,检索的结果可以供用户直接利用,例如查找某一历史事件、某产品的性能、某机构的情况等。可利用的工具主要有百科全书、字典、辞典、年鉴、手册、人名录、地名录、机构指南及其相对应的数据库和网络资源。

2. 根据检索方式划分

(1)手工检索(manual retrieval)。指以手工操作的方式,利用传统的印刷型检索工具,如文摘、目录、索引、百科全书等进行信息检索。手工检索的优点是直观、灵活、无须各种设备和网络,检索条件简单;检索过程中可随时获取反馈信息,及时调整检索策略;查准率高。缺点是漏检严重,检索速度慢。

(2)计算机检索(computer retrieval)。指利用计算机技术,对已经数字化的信息进行存储和检索的过程,包括脱机检索、联机检索、光盘检索和网络检索。计算机检索是在手工检索的基础上发展起来的。随着计算机技术和通信技术的不断发展,计算机检索日益成为信息检索的主流方式。与手工检索相比,计算机检索具有速度快、效率高、查全率高、不受时间和空间限制、检索结果输出方式多样化等优势。

3. 根据检索的性质划分

(1)全文检索(full text retrieval)。指将存储于数据库中文章、著作的任意内容,查找出来的检索。它可以根据需要,获得全文中有关章节、段落、甚至词句,也可以进行各种统计和分析。

(2)超文本检索(hyper text retrieval)。指对每个节点中所存信息及信息链构成的网络信息的检索。这种检索方式强调中心节点之间的语义链接结构,靠系统提供的复杂工具进行图示穿行和节点展示,提供浏览方式查询,可以进行跨库检索。

(3)超媒体检索(hyper media retrieval)。指对存储的文本、图像、声音等多种媒体信息的检索。它是多维存储结构,有向的连接,与超文本检索一样,可以提供浏览方式查询和跨库检索。

二、信息检索语言

(一)检索语言的定义

检索语言又称标引语言、索引语言,是用来描述文献内容特征、外部特征和表达信息提问的一种人工语言。它是协调统一信息标引人员与检索人员检索用语一致性的规范,是沟通信息存储与信息检索过程的桥梁。检索语言的基本构成单元是检索词(索引词、标引词)。检索词不仅包括名词术语、专业词汇,也可以使用符号、代码等。检索语言除表达事物的不同概念外,有时还要揭示概念间的逻辑关系。

(二) 检索语言的功能

检索语言在信息检索中起着十分重要的作用,其功能可以概括为:①表达描述信息内容和信息需求。存储信息时,对信息内容进行分析,概括出代表信息主要内容的若干主题概念,用检索语言对这些概念标引,再存入信息系统中;检索信息时,对课题进行分析,形成能代表信息需求的若干语词,并根据检索语言转换为提问标识,然后在信息系统中匹配出标识相同的文献信息。②词汇或术语的规范和控制。对标引、检索用词进行控制和管理,便于将标引用语和检索用语相符性比较,保证标引和检索用词的一致性。③信息的组织和整序。使同一主题概念的信息或与主题相关的信息有序集中或得到揭示,保证信息存储的系统化和组织化。

(三) 检索语言的类型

按照不同的划分方式,检索语言具有不同的类型,通常的划分方式如下:

1. 按检索词的规范程度划分

(1)受控语言,又称人工语言,也叫规范化语言,是通过对标引词词型、词义、词量及词间关系的规范化控制,消除或减少其歧义,达到标引词表达文献概念的单一性。受控语言主要用来编制参考工具书、手工检索工具的分类目录与各种索引以及网上数据库中的分类导航表等,如标题语言、元词语言、叙词语言和分类语言等。

(2)非受控语言,又称自然语言,是采用未经人工控制的词语或符号作为检索标识的一种检索语言。该语言对主题概念中的同义词、多义词等不加处理,取其自然状态,有较大的灵活性。它能及时反映最新出现的词汇,反映规范词难以表达的特定概念或新概念。检索中常用的关键词语言就属于非受控语言。

2. 按检索标识的组配方式划分

(1)先组式语言。指检索文献之前,表达文献内容的标识已事先固定组配好,如分类语言、标题词语言等。先组式语言有较好的直接性和专指性,但灵活度差。

(2)后组式语言。指表达文献主要概念的标识,在编制词表和标引文献时,都没有预先固定组配,而是在进行检索时,根据检索的需要,按组配原则临时组配起来,如单元词语言、叙词语言等。后组式语言的组配方式灵活,在计算机检索中应用广泛。

3. 按描述信息的特征划分

描述信息内容特征的语言,主要有主题语言、分类语言和代码语言。

描述信息外部特征的语言,主要有题名、责任者、编号、机构名称或引文等。它是采用文献信息的名、责任者、编号、机构名称和引文等作为著录和标引的对象,检索时可将它们作为检索标识进行检索。

三、信息检索方法、途径与步骤

(一) 信息检索的基本方法

采用合适的检索方法和检索途径,可以花费较少的人力、物力和时间,达到最佳的检索效果,使用户努力最小、速度最快、费用最低地获取信息,查全率和查准率都很高。检索方法是一个广义的概念,在信息检索过程中包含了一系列检索技术和原理。常见的检索方法主要有以下几种。

1. 手工检索方法

手工检索是指直接利用印刷型检索工具进行信息检索的方法,包括直接检索一次信息源的“直接检索”方式和利用手工检索工具获取信息线索的“间接检索”方式。直接检索方式费时、费力且效果差。间接检索方式是利用传统检索工具先获取信息线索,然后按信息线索获取原始信息。根据检索要求的不同,这种方法又可分为追溯法、工具法和综合法。

(1)追溯法。追溯法又称回溯法、引文法,是以某一篇文章末尾所附的参考文献为线索,由近及远地进行逐一追踪查找的方法。这种方法直观、方便,不断追溯可查到某一专题的大量参考文献。缺点一是费时,二是原文文献引用的参考文献是很有限的。

(2)工具法。工具法也称常规法,就是直接利用传统检索工具先获取所需信息线索,再根据所得线索获取原始文献全文的方法。工具法根据课题需求和对文献时限的要求,又分为顺查法、倒查法和抽查法3种。

(3)综合法。综合法又称分段法、交替法或循环法。具体地讲,就是利用检索工具查找出一批相关文献后,再利用这些文献后所附的“参考文献”的指引进行追溯查找,以获得更多的相关信息。简而言之,就是先工具法,后追溯法,两者不断交替使用。同时也可以根据已经掌握的参考文献线索,先追溯查找出一批相关信息,再利用检索工具扩大查找信息的线索和范围。

2. 计算机检索方法

计算机检索的核心是数据库,包括目录数据库、文摘数据库、全文数据库、超文本数据库和多媒体数据库等,所以计算机检索方法就是指在特定计算机检索系统的检索界面下,从数据库中检出所需信息的过程。计算机检索的一般性方法和技术,将在本章第三节中详细介绍。

(二)信息检索的途径

在开始信息检索前,必须先确定从什么检索途径着手。使用一定的检索途径,就是把所需信息的某种特征标识转换为检索标识。一般来说,检索途径可以按照信息资源外部特征途径和内容特征途径分为两大类,如图1-1所示。

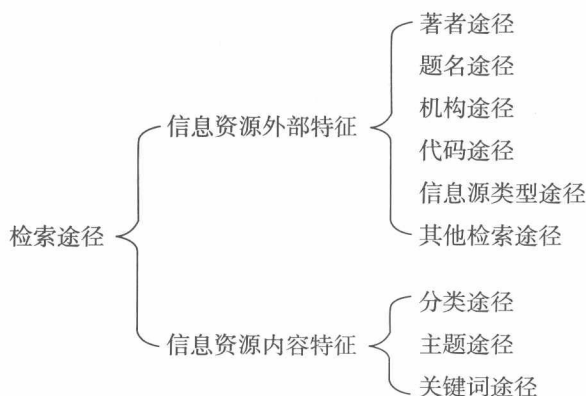


图 1-1 信息检索途径

(三)信息检索的步骤

无论是手工检索还是计算机检索,检索的程序都是相同的,都是经过仔细分析,并通过

实践逐步完善查找方法的过程。信息检索的步骤,一般包括课题分析、选择检索工具、确定检索途径、制定检索表达式、结果输出和用户评价等几个程序。

1. 课题分析

分析检索课题是检索过程中的首要环节,其目的在于明确检索要求和范围,做到有的放矢。目的越明确、范围越具体、掌握的线索越多,检索所需信息的准确性就越大。课题分析一般从课题的主题内容、研究要点、学科范围、语种范围、时间范围等几个方面进行。

2. 选择检索工具

在课题分析的基础上,需明确选择适当的检索工具或系统。不同的检索工具或系统,其收录范围、文种、编排形式等性能参数方面各不相同。如果所查课题内容较多,涉及面较广,可选用综合性检索工具,否则应选用专业性检索工具。如果对课题有特定的要求且条件具备,则应选用联机检索系统和网络数据库进行检索。

3. 确定检索途径

检索途径又称检索入口、检索点或检索项等。选择合适的检索途径,能够更加快捷、准确、全面地查找到所需信息。一般优先选用主题词或关键词途径,其次是分类途径。如果明确所需信息的特定代码、著者、题名或机构名称等,则直接依据信息的外部特征途径进行检索。

4. 制定检索表达式

检索表达式是人与检索系统进行交流的人工语言,它是检索步骤中最重要的一环,直接影响检索效率。如果制定的检索表达式准确度差,那么检索出的无关信息或冗余信息就多,检索表达式分为两种:简单表达式和复合表达式。

简单表达式指单独使用一个检索词进行检索。一个检索词可以为著者、题名、关键词、分类号等。

复合表达式指两个或两个以上的检索词按现代检索技术所构成的复杂检索字符串。它是专门应用于计算机系统的检索技术。制定复合检索表达式要应用到布尔逻辑检索技术、截词检索技术、限制检索技术、加权检索技术和全文检索等现代检索技术。

5. 结果输出和用户评价

制定了检索表达式之后,对于手工检索系统而言,要获得所需信息的线索和原始信息全文,要靠手工操作去处理;对于计算机检索系统而言,其执行检索和结果输出是自动完成的。如果检出的信息满足了用户需求,那么对该课题的检索过程就圆满结束。如果用户认为在查全、查准或有效性等方面没有获得满意效果,那么就需要及时调整检索方案。

第三节 检索工具类型与检索方法

一、检索工具的定义和类型

(一)检索工具的定义

检索必须借助于一定的检索工具。检索工具有广义和狭义之分。广义的检索工具是指用来报道、存储和查找各类信息的一切工具和设备,可分为手工检索工具、计算机检索工具等。广义的检索工具与检索系统的含义是一致的,彼此并没有严格的区分界定。狭义的检