



面向 21 世纪 课 程 教 材

高等学校信息管理与信息系统专业系列教材

信息检索

主 编 符绍宏

副主编 赵 荣 王 琼 茹海涛

Xinxi
Jiansuo



高等教育出版社

HIGHER EDUCATION PRESS

面向 21 世纪 课程教材
高等学校信息管理与信息系统专业系列教材

信 息 检 索

主编 符绍宏

副主编 赵 荣 王 琼 茹海涛

高等教育出版社

内容提要

本书是教育部“高等教育面向 21 世纪教学内容和课程体系改革计划”的研究成果,是面向 21 世纪课程教材。

信息检索是信息管理与信息系统专业的主干课程之一,同时也是一门面向多学科专业的公共课程。面对大量纷繁的信息,用户如何准确、及时、有效地获取自身需求的信息已十分重要。本书针对这一特点,从信息检索的理论基础开始,详细地介绍了信息检索的方法,文献、数据与事实、计算机、因特网等多种形式的信息检索,并在每章后设计了检索实习题,供学习实践。

本书既可作为信息管理与信息系统、图书馆学、情报学专业教材,也可作为高等学校信息检索类公共课教材。

图书在版编目(CIP)数据

信息检索 / 符绍宏主编. —北京: 高等教育出版社,
2004.7
ISBN 7-04-014624-X

I. 信... II. 符... III. 情报检索—高等学校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2004)第 059375 号

策划编辑 刘 艳 责任编辑 耿 芳 市场策划 韩 飞
封面设计 于文燕 责任印制 陈伟光

出版发行	高等教育出版社	购书热线	010-64054588
社 址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-82028899		http://www.hep.com.cn

经 销 新华书店北京发行所
印 刷 涿州市星河印刷有限公司

开 本	787×960 1/16	版 次	2004 年 7 月第 1 版
印 张	26.25	印 次	2004 年 7 月第 1 次印刷
字 数	480 000	定 价	28.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

前 言

信息检索是信息管理学领域重要的研究课题,信息检索课程也一直是信息管理与信息系统专业教育中的主干课程之一,同时也是一门面向所有学科专业学生开设的公共课程。本教材是基于编者近年来在北京师范大学信息技术与管理系从事信息检索专业课教学及相关公共课教学改革实践,并在编者近年来主编或参与编写的两部“面向 21 世纪课程教材”的基础上修改和完善的。

近年来,因特网的广泛应用和发展构成了人类信息环境变化的重要层面。因特网使世界范围内的信息交流、资源共享成为可能,从而大大拓展了人类的信息空间。同时它也对传统的信息组织、信息检索和获取方法构成了很大冲击。传统的信息检索系统、检索工具都发生了很大变化,许多著名的信息检索服务系统,如 Dialog、OCLC 等也都在大力开发建设以因特网为平台的信息服务。网络信息空间的无限、无序、优劣混杂,缺乏统一的组织与控制等特点给人们有目的地查找和利用信息造成了很大的不方便。用户面对大量纷繁复杂的信息都会感到由信息过载(Information Overloaded)引发的困惑和茫然,要在浩瀚、动荡的信息海洋中,准确、及时、有效地找到、获取与自身信息需求相关,切合需要和适用的信息,对所有用户来说都是十分重要的,同时也具有一定的挑战性。

本书紧密结合数字化、网络化信息环境的特点和实际的检索需要,综合全面地介绍广泛分布的信息资源、多样化的信息检索系统、检索技术方法和检索手段等。在内容组织上,本书的特点如下:

- 针对原有相关教材中检索理论大多基于文献检索和传统的封闭、静止的检索系统,而对目前以因特网为中心的开放、动态的检索系统的理论总结则较为薄弱,导致检索理论与检索实践的脱节等情况,本书尝试以统一的理论涵盖文献检索、数据检索、事实检索以及手工检索和计算机检索,以追求效益的原则,注重介绍网络环境下的检索系统理论知识。

- 对因特网信息检索进行新的、更全面的阐述。变化是因特网永恒的主题,网络信息检索资源、网络检索工具的发展也非常快。本教材力求全面地诠释这种不同以往的信息检索模式,更深入地分析、介绍有关最新研究和发展动态,使学生对网络检索资源的分布、利用乃至开发都能有良好的知识储备,培养其较强的网络信息查找与利用能力,提高其在网络信息环境下的生存和发展水平。

- 对数据、事实检索部分的检索工具进行全面的介绍。当代的信息环境和社

会发展条件下,人们更注重用较小的投入,迅速地获知问题的答案。本教材较为全面地介绍了网络上提供特色数据、事实检索服务的网站,指导人们利用网络上的“参考源/工具书”,例如,网上百科全书、人物信息、地理信息、统计信息等检索网站。

• 注重设计各具特色的检索实习。检索课是一门实践性很强的课程,为使学生在检索实习中能充分接触和利用各种信息机构、信息渠道,在大量亲身的检索实践中充分地感受和接触各种最新的信息资源、检索技术、检索手段和检索服务,本书在每个单元均设计和提供了检索实习题。

本教材既可作为信息管理 with 信息系统专业、图书馆学、情报学专业信息检索专业课程的教材,也可作为高校信息检索与利用公共课的教材。

本书编写分工如下:第1~3章由赵荣负责,第4章由王琼负责,第5章由茹海涛负责,第6~8章由符绍宏负责,参加编写的人员还有于超、郑鑫、张丽、苏建华、夏方、林小露、袁理。全书的筹划、内容大纲、编写组织和统稿等由符绍宏负责。

结合专业课的学习和研究,参与本书资料搜集的人员还有北京师范大学管理学院信息技术与管理系的李瑞玲、宋慧敏、康晓伶、黄婧、赵刚、刘漫、孙青、罗璇、马建华、杨程、郑璐、施天翼。北京师范大学图书馆的龙世彤、黄燕云也为本书提供了重要的资料。

应该说明的是,虽然本书的编者大多都是长期担任信息检索课教学的专业教师或承担信息检索、参考咨询工作的图书馆专业人员,但由于工作负担,学习压力导致的投入编写过程的时间有限,而且网络环境下信息资源、检索平台又经常发生变化,加上本书编者的水平和视野有限,本书中定有不少疏漏之处,欢迎业界同行批评和指正,也恳请各位读者反馈宝贵意见。

符绍宏

2004年5月

于北京师范大学

目 录

第1章 绪论	(1)	3.1.1 信息检索语言的类型及特点	(57)
1.1 信息检索	(1)	3.1.2 检索途径及其选择	(69)
1.1.1 信息检索的概念	(1)	3.2 信息检索技术	(71)
1.1.2 信息检索的类型	(3)	3.2.1 布尔检索	(71)
1.2 信息资源	(7)	3.2.2 截词检索	(72)
1.2.1 信息资源的概念	(7)	3.2.3 限制检索	(74)
1.2.2 信息资源的分类	(9)	3.2.4 原文检索	(75)
1.2.3 文献信息资源	(11)	3.2.5 加权检索	(78)
1.3 信息检索需求	(16)	3.2.6 聚类检索	(80)
1.3.1 信息检索的现实需求	(16)	3.2.7 检索技术的发展	(82)
1.3.2 信息检索需求的类型	(20)	3.3 信息检索过程与策略	(90)
思考题	(21)	3.3.1 信息检索一般步骤	(90)
第2章 信息检索基础理论	(22)	3.3.2 检索策略	(91)
2.1 信息检索基本原理	(22)	思考题	(101)
2.1.1 信息检索的基本原理	(22)	第4章 文献信息检索	(102)
2.1.2 信息检索的相关性问题	(26)	4.1 图书信息检索	(102)
2.1.3 信息检索的效果评价	(31)	4.1.1 书目检索工具概述	(102)
2.2 信息检索系统和工具	(35)	4.1.2 主要的书目检索工具	(105)
2.2.1 信息检索系统的类型及特点	(35)	4.1.3 网络环境下书目信息的检索	(106)
2.2.2 印刷型检索工具的类型和结构	(37)	4.1.4 电子图书及其检索	(111)
2.2.3 计算机检索系统的结构及工作原理	(40)	4.2 期刊信息检索	(119)
2.3 信息检索模型	(47)	4.2.1 概述	(119)
2.3.1 信息检索经典模型	(48)	4.2.2 中国期刊信息检索	(120)
2.3.2 结构化文本检索模型	(52)	4.2.3 国际性期刊检索系统	(123)
2.3.3 浏览模型	(54)	4.2.4 电子期刊	(132)
思考题	(56)	4.3 学术性文献信息核心检索工具	(135)
第3章 信息检索基本方法	(57)	4.3.1 概述	(135)
3.1 信息检索语言与检索途径	(57)	4.3.2 美国《工程索引》	(135)

4.3.3 美国《化学文摘》····· (142)	5.1 参考工具概论····· (211)
4.3.4 美国《生物学文摘》····· (150)	5.1.1 数据与事实信息检索
4.3.5 英国《科学文摘》····· (155)	概述····· (211)
4.3.6 《科学引文索引》····· (161)	5.1.2 参考工具及其功能····· (211)
4.4 政府出版物检索····· (167)	5.1.3 数据与事实参考工具的主要
4.4.1 政府信息及政府出版物·· (167)	类型····· (212)
4.4.2 中国政府信息检索····· (168)	5.1.4 参考工具的编排结构····· (213)
4.4.3 国际组织及政府机构出	5.1.5 参考工具的评价····· (214)
版物及相关信息检索····· (170)	5.1.6 参考工具的书目控制····· (215)
4.5 科技报告检索····· (176)	5.2 百科全书····· (216)
4.5.1 概述····· (176)	5.2.1 概述····· (216)
4.5.2 美国政府四大科技报告	5.2.2 重要的国内外百科全书·· (219)
检索····· (177)	5.2.3 百科全书检索网站····· (222)
4.5.3 中国科技成果数据库····· (180)	5.3 传记资料····· (225)
4.6 会议文献检索····· (181)	5.3.1 概述····· (225)
4.6.1 概述····· (181)	5.3.2 重要的传记资料检索
4.6.2 会议消息检索····· (182)	工具····· (226)
4.6.3 会议文献检索····· (183)	5.3.3 传记资料检索网站····· (228)
4.7 学位论文检索····· (188)	5.4 地理资料····· (230)
4.7.1 概述····· (188)	5.4.1 概述····· (230)
4.7.2 中国学位论文的检索····· (188)	5.4.2 重要的地名工具书····· (230)
4.7.3 美国博/硕士学位论文的	5.4.3 地图集····· (231)
检索····· (189)	5.4.4 地理信息的检索网站····· (232)
4.8 专利文献检索····· (192)	5.5 年鉴····· (234)
4.8.1 概述····· (192)	5.5.1 概述····· (234)
4.8.2 中国专利文献检索工具·· (194)	5.5.2 重要的年鉴····· (235)
4.8.3 英国德温特专利文献检索	5.5.3 相关检索网站····· (236)
工具····· (195)	5.6 机构名录····· (237)
4.8.4 专利信息检索网站····· (199)	5.6.1 概述····· (237)
4.9 标准文献检索····· (202)	5.6.2 重要的政府、学术机构
4.9.1 概述····· (202)	指南····· (238)
4.9.2 中国标准文献检索····· (204)	5.6.3 重要的商业性机构名录·· (238)
4.9.3 国际标准信息检索····· (207)	5.6.4 相关数据库和检索网站·· (239)
4.9.4 其他重要标准信息机构及	5.7 词典····· (241)
网站····· (209)	5.7.1 概述····· (241)
思考题····· (210)	5.7.2 重要的词典工具书····· (242)
第5章 数据与事实信息检索····· (211)	5.7.3 相关检索网站····· (244)

5.8 统计资料	(245)	7.2.3 因特网信息检索工具	(303)
5.8.1 概述	(245)	7.3 网络信息检索工具(1)——目	
5.8.2 重要的统计信息检索		录型网络检索工具	(311)
工具	(247)	7.3.1 概述	(311)
5.8.3 相关检索网站	(249)	7.3.2 目录型检索工具	(312)
思考题	(250)	7.4 网络信息检索工具(2)——索引型	
第6章 计算机信息检索	(251)	网络检索工具及搜索引擎	(322)
6.1 概论	(251)	7.4.1 概述	(322)
6.1.1 计算机信息检索系统的		7.4.2 英文搜索引擎	(326)
发展	(251)	7.4.3 中文搜索引擎	(334)
6.1.2 各类型检索系统及其		7.5 网络信息检索工具(3)——集合	
选择	(252)	型及元搜索引擎	(342)
6.1.3 数据库类型和结构	(253)	7.5.1 概述	(342)
6.2 商业联机检索系统	(256)	7.5.2 元搜索引擎	(345)
6.2.1 概述	(256)	7.6 其他类型网络信息检索工具	(360)
6.2.2 我国联机检索系统与数		7.6.1 FTP 资源检索工具	(360)
据库	(256)	7.6.2 用户组 Listserve、Mailing List 和	
6.2.3 国外著名联机检索系统	(258)	Usenet 检索工具	(364)
6.3 光盘检索系统	(268)	7.6.3 多媒体资源检索工具	(367)
6.3.1 概述	(268)	7.6.4 专科型网络检索工具	(374)
6.3.2 我国主要的光盘数据库		思考题	(385)
及其检索方法	(272)	第8章 信息服务及其利用	(386)
6.3.3 国外主要的光盘数据库		信息检索与咨询服务	(386)
及其检索方法	(280)	8.1.1 文献信息服务机构	(387)
思考题	(289)	8.1.2 信息检索咨询服务方式	(390)
第7章 因特网信息检索	(290)	8.1.3 信息服务的利用	(393)
7.1 因特网信息资源	(290)	8.2 信息素质	(395)
7.1.1 因特网信息资源特点及		8.2.1 概述	(395)
价值	(290)	8.2.2 信息素质的构成	(398)
7.1.2 因特网信息资源的种类	(291)	8.2.3 信息素质的养成	(401)
7.1.3 因特网信息资源的评价	(295)	思考题	(406)
7.2 因特网信息检索	(300)	主要参考文献	(408)
7.2.1 因特网信息检索的特点	(301)		
7.2.2 因特网信息检索的一般			
方法	(301)		

第1章 绪 论

本章要点

信息检索及相关概念;信息检索类型划分;信息资源的载体形式、级次、类型划分以及各类型文献的特点和检索利用价值;面向各类用户需要的信息检索需求特点。

1.1 信息检索

1.1.1 信息检索的概念

通俗地说,信息检索(Information Retrieval)就是信息用户为处理解决各种问题而查找、识别、获取相关的事实、数据、知识的活动及过程。作为人类社会活动必不可少的一部分,信息检索有着悠久的历史,而随着信息社会的到来,其重要地位日益突出。伴随着信息检索活动的开展,信息检索研究应运而生,其主要研究内容包括:信息检索理论,信息检索语言,信息检索工具或信息检索系统的构建及评价,信息检索技术与方法等。

20世纪中叶以前,信息的存储和传播主要以纸介质为载体,信息检索活动主要围绕着相关文献的获取和利用而展开,因此“文献检索”成为信息检索的同义词被更为广泛地使用。20世纪50年代以后,社会信息传播与存储载体呈现多元化发展,人们不再拘泥于载体研究信息检索,于是开始广泛使用“情报检索”一词。近年来,由于汉语中“信息”一词较“情报”一词的含义更为宽泛,加上英文词汇Information可以概括“情报”与“信息”,因此,人们越来越倾向于将文献检索和情报检索统称为信息检索这一更具兼容性的概念。

随着信息检索理论和实践的更新发展,人们对信息检索的认识也在不断深入。目前对于信息检索这一概念,学术界尚无公认一致的定义,国内外有关专家从不同的角度解释信息检索,较有代表性的观点主要有以下几种。

1. 从通信的角度认识信息检索

莫尔斯(Calvin N. Mooers)在1950年发表了《把信息检索看做是时间性的通信》一文,不仅首次提出了“信息检索”这个概念,并认为“信息检索是一种时间性的通信形式”,“此种通信是时间性的,在时间上从一个时刻通往较晚的时刻,而在空间

上可能还在同一地点”。也就是说通过信息检索得到了一些文献,从而使得信息发布者与信息用户之间建立起了一种通信。这一观点旨在说明,信息传递是一种特殊的通信形式,正是这种通信促使了人类思想、文化、科学上的交流。莫尔斯强调在通信双方中,信息发送者必须尽可能发送一切信息,是时间性通信的被动一方;而信息接收者是主动活跃的一方,正是接收者才决定什么时候接收以及接收什么信息。因此,信息检索的问题就在于,如何把一个可能的用户指引向所存储的信息。

这种认识突出了用户需求的重要性,对于如何做好信息检索服务具有很强的理论和实践意义。

2. 从信息处理的角度认识信息检索

从信息处理的角度来看,信息检索的基本问题,是如何处理信息和信息的结构。这种认识偏重于信息管理领域,认为信息不仅限于传统“文献”的范围,图像、声音、数据等也都能反映信息,并把信息检索视为计算机科学技术的一个分支。

这种认识强调如何构造以及利用什么形式来构造信息结构的问题。在当今因特网迅速发展,网络信息极大丰富的情况下,这种认识对于信息检索工具的设计和组建,具有指导意义。

3. 从信息检索过程的角度认识信息检索

这种观点认为,信息检索就是查找出含有用户所需信息的文献的过程。这是一种传统的主流观点,支持者较多。美国信息检索专家兰卡斯特的经典表述是“信息检索系统并不检索信息”。因为信息是无形的,必须依附于文献而存在。虽然信息检索的最终结果是满足用户的信息需求,但检索的直接对象还是文献,当用户阅读文献并理解其内容时,用户的信息需求才被满足。

4. 全息检索

我国的王永成教授认为,全息检索就是“可以从任意角度,从存储的多种形式的信息中高速准确地查找,并可以任意要求的形式和组织方式输出,也可仅输出人们所需要的一切相关信息的电脑活动”。

这里所谓的任意角度,是指要求检索系统以用户可能的任何检索需求作为出发点,并把这些出发点都设计成“检索入口”;所谓多种形式的信息,指的是在现代多媒体技术能够存储并输出文本、图像和声音信息的条件下,继续发展直至能输出超声频与超视频信息;所谓任意要求的信息组织形式,是指按用户需求对已检索到的信息加以组织并输出,从而真正实现人机检索过程中“以人为本”的服务宗旨;所谓输出一切“相关信息”,从存储方面看,系统对存储的文本的外部特征,文本的内涵特征以及其他辅助性特征等的描述与信息本身存在不可避免的差异,应尽量缩

小这种差异;从检索一方看,用户对信息需求的认知能力、表述能力也同样影响到其能否检索出与真正的信息需求相关的结果。因此,“相关性”不仅是传统文献检索,也是全息检索的基本特征和评价检索系统的重要参考指标。以上这些工作都是由计算机来完成的。

5. 概念信息检索

Chunk 等专家认为,概念信息检索是基于自然语言处理中对知识在语义层次上的析取,并由此形成知识库,再根据对用户提问的理解来检索其中的相关信息。它与传统文献检索的不同之处在于,后者是基于关键词为核心的标引与检索,而关键词在很多情况下并不能确切表达文献信息的概念和内容,因此误检与漏检在所难免;而概念信息检索要对输入的原文内容中的概念,而不是关键词来进行组织和安排,在对其进行语义层次上的自然语言处理基础上,来获取相关的概念和范畴知识,然后通过记忆机制将它们存储到知识库中以备检索。概念信息检索系统一般由记忆机制、语义分析机制、知识库、人机接口等部分组成。

从实际检索工作的角度出发,对信息检索这样表述:信息检索具有广义和狭义两重含义。广义地讲,信息检索是指将信息按一定的方式组织和存储起来,并根据用户的需要找出相关信息的过程。其中包括存与取两个环节,“存”即信息存储,是对信息进行收集、标引、描述及组织,并对其进行特征化表达集加以整序,形成信息检索工具或检索系统的过程。“取”即信息查找,是通过某种查询机制从检索工具或检索系统中查找出用户所需的特定信息或获取其线索的过程。狭义的信息检索仅仅指信息查找(Information Search)的过程,也即“取”这一环节。信息检索的过程可以用图 1-1 来简略表示。

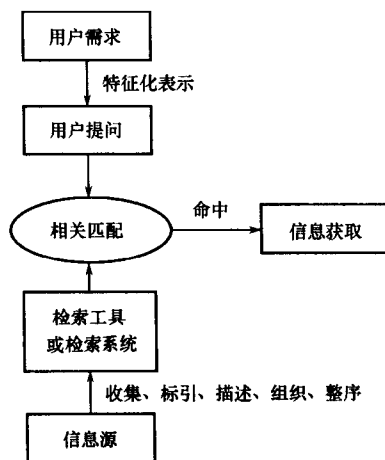


图 1-1 信息检索过程示意图

1.1.2 信息检索的类型

根据不同的标准,信息检索可以划分为不同的类型。

1. 按检索对象的内容区分

按检索对象的内容区分,信息检索可分为文献检索、数据检索和事实检索。

(1) 文献检索(Document Retrieval)

文献检索是以文献作为检索对象,查找含有用户所需信息内容的文献。文献

检索是一种相关性检索而非确定性检索,系统不直接解答用户所提出的问题本身,只提供与之相关的文献或文献的属性信息与来源指示供用户参考和取舍。所以其检索对象是包含特定信息的各类文献。

(2) 数据检索(Data Retrieval)

数据检索是将经过选择、整理、鉴定的数值数据存入数据库中,根据需要查出可回答某一问题的数据的检索。这些数值型数据各种各样,包括物理性能常数、调查数据、统计数据、人口数据、国民生产总值、外汇收支等。数据检索是一种确定性检索,即直接提供用户所需要的确切的数据,而且检索结果一般也是确定性的,要么是有,要么是无;要么是对,要么是错。有些数据检索系统不仅能查出数据,还提供一定的运算、推导能力。用户获得各种经过整理、计算过的量化信息,从而为定量分析提供依据。例如,它可以回答“9.11事件前美国纽约世贸大厦有多高?”,“2002年中国经济增长率是多少?”之类的提问。

(3) 事实检索(Fact Retrieval)

事实检索是存储关于某些客体(如机构、人物等)的指示性描述,或关于某一事件发生的时间、地点、经过等信息并将其查找出来的检索。有人认为这种检索实际上也可归入数据检索,因为事实也可以是一种数据,即非数值性数据。这里所说的事实检索,是指数值信息和系统数据信息混合的检索。这种检索既包含数值数据的检索、运算、推导,也包括事实、概念等的检索、比较、逻辑判断。一般先从系统中检索出所需信息,再加以逻辑推理才能给出结论。这样的事实检索已经超出传统信息检索的范畴,实质上是一种“问题求解”过程或专家系统技术。近年来开发的某些“智能信息检索系统”就初步具有这样的特征和功能。

其实,用户所需的数据或事实不能脱离文献而单独存在,因此,数据或事实检索也是以文献检索为依托的。它们之间有许多共同之处,文献检索用的大多数技术方法都适用于数据检索或事实检索。在信息服务过程中,二者也常常是相互配合、相辅相成的。例如,美国化学文摘社(CAS)生产的数据库中就包括文献型数据库和非文献型数据库,检索时这两种数据库可以配合起来使用。

2. 按检索方式区分

(1) 手工信息检索

手工信息检索指以手工操作的方式,利用检索工具书进行信息检索。手工信息检索是信息检索的传统方式,已经历了一个多世纪的发展历程。其优点是直观、灵活,便于控制检索的准确性;缺点是查找较复杂,检索速度慢,工作量较大。

(2) 机器信息检索

机器信息检索主要指计算机信息检索,是通过机器对已数字化的信息,按照设计好的程序进行查找和输出的过程。按机器检索的处理方式分类,有脱机检索和

联机检索;按存储方式分类,有光盘检索和网络检索等。机器检索不仅大大提高了检索效率和检索的全面性,而且拓展了信息检索领域,丰富了信息检索的研究内容。

计算机信息检索是从手工信息检索的基础上发展起来的,日益成为信息检索的主流方式,但手工信息检索工具的价廉、灵活、便携等特征使其仍有存在和发展的空间。为了满足用户的多种需求,给予用户更多的选择,向用户提供多样化的服务形式,手工信息检索还将与计算机信息检索共同存在,互相补充、互为促进。

3. 按检索要求区分

(1) 强相关检索

强调检索的准确性,向用户提供高度对口信息的检索,也称为特性检索。这种检索注重查准,只要检索得到的文献信息能够满足用户的需求即可,通常对于检索结果的数量多少不做要求。

(2) 弱相关检索

强调检索的全面性,向用户提供系统完整的信息的检索,也称为族性检索。这种检索注重查全,要求检索出一段时间期限内有关特定主题的所有信息。为了尽可能避免漏检相关信息,对于检索的准确性相对要求较低。

要注意的是,这是两种检索要求比较极端的检索类型。实际上更多的时候,用户对于检索的要求介于两者之间,既要求查找对口的信息,又希望得到所有的信息。但在信息检索的实际过程中,查准和查全常常不能兼顾。关于这一点将在第2.1.3小节信息检索的效果评价中进一步讲述。

4. 按检索的时间跨度区分

(1) 定题检索

查找有关特定主题最新信息的检索,又称为SDI(Selective Dissemination of Information)检索。其特点是只检索最新的信息,时间跨度小。定题检索在文献信息库更新时运行,即每当文献信息库加入新的文献信息时,就用根据用户检索需求拟定的检索提问检索一遍,查找出特定主题的最新信息,分析整理检索结果并以一定的方式提供给用户。这种检索模式非常适合于信息跟踪,便于及时了解有关主题领域的最新发展动态。用户一旦向检索服务机构订购定题检索,一般就会在较长时间内多次运行,由检索服务机构持续地向用户提供最新信息。

(2) 回溯检索

查找一段时期内有关特定主题信息的检索,也称为追溯检索。其特点是既可以查找过去某一段时间的特定主题信息,也可以查找最近的特定主题信息。与每个定题检索需要多次运行有所不同,每个回溯检索一般只运行一次,从已有的文献信息库中找出某个时间段内特定主题的信息,并提供给用户。

到目前为止,用户利用最多的是回溯检索,大多数的检索课题都属于回溯检索。但定题检索发展很快,科研课题进行中需要 SDI,近年来在工商经贸领域,定题检索也日益得到企业用户的广泛运用。

5. 按检索对象的形式区分

(1) 文本检索

查找含有特定信息的文本文献的检索,其结果是以文本形式反映特定信息的文献。这是一种传统的信息检索类型,在信息检索中至今依然占据着主要地位。

(2) 多媒体检索

多媒体是相对于单媒体而言的,是指将声音、图像、通信等在内的多种媒体的功能有机结合在一起,并用某种新媒体来代替传统媒体的多种媒体。

多媒体信息检索是指根据用户的需求,对文字、声音、图像、图形等多种媒体信息进行组织、存储从而识别、查找并获取所需信息的过程。多媒体信息检索包括两层含义:其一是对离散媒体的检索,如查找包含某种颜色或色彩组合的特定图像;其二是指对连续媒体的检索,如查找包含某一特定场景的视频资料。

在因特网迅速发展的今天,网上存在着大量的多媒体文献,用户常常需要查找特定的图像、声音、动画等。多媒体文献的信息组织与处理,如标引、描述和有序化编排等与传统的文本处理截然不同,其检索需求及检索途径也别具特色,这就向传统的信息组织、检索方式及理论提出了挑战。

多媒体信息检索与文本信息检索相比,其主要特点是:信息类型复杂,存储与检索技术复杂,具有交互性、同步性,在技术的实现上亟待研究和突破。

6. 按检索对象的信息组织方式区分

(1) 全文检索(Full Text Retrieval)

全文检索是将存储于数据库中的整本书、整篇文章中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、句、词等的信息,也可进行各种统计和分析。例如,它可以回答“《红楼梦》一书中‘林黛玉’一词共出现多少次?”这样的问题。

(2) 超文本检索(Hyper Text Retrieval)

超文本系统是将诸多文本信息通过超链接联系起来而形成的一种非线形的文本结构。从组织结构上看,超文本的基本组成元素是节点(Nodes)和节点间的逻辑链接(Links),每个节点中所存储的信息以及信息链被联系在一起,构成相互交叉的信息网络。与传统文本的线形顺序检索不同,超文本检索强调中心节点之间的语义联系结构,靠系统提供的复杂工具做图示穿行和节点展示,提供浏览式查询。而传统的文本检索系统则强调文本节点的相对自主性。

(3) 超媒体检索(Hyper Media Retrieval)

超媒体检索是对超文本检索的补充。其存储对象超出了文本范畴,融入了静、动图像(形)以及声音等多种媒体信息。信息的存储结构从单维发展到多维,存储空间范围不断扩大。

对于超文本和超媒体检索,两者的链都是有向的,均面向浏览式查询。

7. 按检索途径的特点区分

(1) 常用法

常用法也称工具法,是检索中最为常用的方法,指利用检索工具或系统中常设的检索入口(如主题、分类、著者、题名、号码等)查找文献信息的方法。其具体操作又可分为顺查、倒查、抽查3种。

顺查方式是根据有关课题的起始年代,利用选定的检索工具由远及近地进行逐年查找。倒查方式与顺查方式相反,是利用选定的检索工具由近及远地进行逐年查找,直到满足信息检索的需要为止。抽查方式是针对有关学科专业的发展特点,根据检索的要求,重点抓住学科专业发展兴旺,文献发表数量较多的年代,抽出一个或几个时间段进行查找。

(2) 回溯法

回溯法也称引文法,是利用文献末尾所附的参考文献或引文为检索入口,由点到面地滚雪球般查找到更多的来源文献的方法。这需要编制和使用引文索引这种较为特殊的检索工具或检索系统。

(3) 循环法

循环法也称分段法,是综合常用法和回溯法的检索方法,即在查找文献信息时,既利用一般的检索途径,又利用原始文献后所附的参考引用文献作为检索入口,分阶段按周期的交替使用两种方法。

1.2 信息资源

1.2.1 信息资源的概念

在当代信息社会中,信息已成为人所共知的流行词,信息资源的建设和使用对国家、组织、个人的生存和发展都具有至关重要的意义。信息资源同能量和物质一起被称做人类信息社会生产与生活必不可少的3大资源。

战后日本经济的崛起突出显示出了信息资源的重要作用。日本自然资源贫乏,基础科学也不如欧美发达国家,但是它把大量的人力、物力、财力集中在科技信息资源的获取、开发和应用技术的研究上。当欧美等国出现一种高科技研究成果后,日本马上想方设法引进,并研究如何将其运用到生产之中。例如,日本的电视

机技术是从美国引进的,日本投入了大量的改良研究费用后取得了质量和价格的优势,日产电视机销往美国和全世界,占领了市场。日本的汽车、钢铁、电子产品和机械产品占领世界市场采取的都是这种模式。日本就是依靠掌握第一手的科技信息资源,走出了一条科技振兴经济的道路。

那么究竟什么是信息,什么是信息资源呢?

“信息”对应的英文单词是 Information,它来源于拉丁字 Informatio,意思是解释、陈述。在中国港台地区,Information 常被译为“资讯”。

信息作为科学的概念,首先是在信息论中得以专门研究的。信息论的创始人申农(C. E. Shannon)认为,“信息是用以消除信宿对信源发出何种消息的不确定性的东西”,简单地说,“信息是指有新内容、新知识的消息”。控制论的创始人维纳(N. Wiener)认为,“信息是人们同外部世界进行交换的内容的名称”。我国学者钟义信指出:“信息是事物存在的方式或运动的状态,以及这种方式或状态的直接或间接的表述”。由于人们接触信息的角度不同,所以不同的人对信息给出了不同的定义和描述。据不完全统计,有文可考的“信息”定义已不下一百个。

近年来,我国信息学者多从哲学角度来理解信息,认为信息是事物运动的状态和方式。信息不是事物本身,但它反映了事物的特征与特性。不同的事物有不同的特征,并在不同的条件下发生变化,这种特征与变化就是信息。人们正是通过获取和识别这些信息来认识不同事物。信息可以通过数据、资料、图形、语言、声音等来反映、表现、承载或传递。它既包括经过提炼的知识,也包括人类未知的知识。

信息资源(Information Resource 或 Information Resources)最早也是由国外学者提出的,时间大约在 20 世纪 60 年代末 70 年代初。

信息资源的定义与信息定义一样,目前仍是众说纷纭。美国著名信息资源管理专家霍顿在其 1979 年出版的著作中给出了信息资源的两种解释:单数概念的信息资源是指某种内容的来源,即包含在文件和公文中的信息内容;复数概念的信息资源是指支持工具,包括供给、设备、环境、人员、资金等。

从 20 世纪 80 年代中后期起,中国一些学者也开始关注和研究信息资源管理,先后给出了各自的信息资源定义。例如,乌家培教授认为,“对信息资源有两种理解:一种是狭义的理解,即指信息内容本身;另一种是广义的理解,指的是除信息内容外,还包括与其紧密相关的信息设备、信息人员、信息系统、信息网络等”。又如,中科院霍国庆认为:“信息资源也就是可以利用的信息的集合……换言之,信息资源是经过人类开发与组织的信息的集合”。

尽管在认识上存在一定的分歧,但国内外研究信息资源管理的大多数专家都认为,应该从狭义和广义两种角度来认识和理解信息资源的涵义。

1. 狭义信息资源

狭义信息资源是指人类社会活动中经过加工处理的、有序化的并大量累积后的有用信息的集合。例如,科学技术信息、政策法规信息、金融信息等都是常见的狭义信息资源。在这个定义中需要注意以下3点。

① 信息资源是信息的集合。俗话说“独木难成林”。一棵树构不成森林资源,一滴水构不成水资源,同样,一条信息或几条信息也构不成信息资源。只有当信息达到一定的丰度和凝聚度时,才能成为信息资源。

② 信息资源是经过人类选择、获取的有用信息的集合。有用性是一切资源的本质属性,信息资源也不能例外。从信息海洋中挑选出有用信息,并将之与无用信息区分开来,正是信息管理人员的基本任务之一。

③ 信息资源是经过人类组织序化的信息的集合。与非信息资源相比,信息资源最显著的特征就是有序性。对水资源、石油资源、矿产资源等自然资源来说,无所谓有序、无序,只要具备一定的丰度和凝聚度,值得人们开采、获取即可。信息资源却不然,无序的信息不仅无法利用,还会造成信息通道的“栓塞”,阻碍信息的传播、交流、开发和利用。因此,组织、序化的信息才能成为信息资源,而没有控制的、未经组织的信息不能成为资源。信息检索系统的建立正是组织、序化信息,从而使之成为可利用的资源的过程。

2. 广义信息资源

广义信息资源是信息和它的生产者以及信息技术的集合。也就是说,广义信息资源一般由3部分构成:一是人类社会活动中经过加工处理有序化并大量累积后的有用信息的集合;二是为某种目的而生产有用信息的信息生产者的集合;三是加工、处理和传递有用信息的信息技术的集合。

1.2.2 信息资源的分类

按照不同的标准可以将信息资源划分为不同的种类,下面分别介绍。

1. 按信息资源所依附的载体划分

信息资源按照其依附的载体可以分为体裁信息资源、实物信息资源、文献信息资源、网络信息资源4种类型。

(1) 体裁信息资源

体裁信息资源指以人体为载体并能为他人识别的信息资源。按其表述方式又可分为口语信息资源和体语信息资源。口语信息资源是人类以口头语言表述出来但被记录下来的信息资源。例如,谈话、授课、讲演、讨论、唱歌等。体语信息资源是以人的体态表述出来的信息资源,例如,表情、手势、姿态、舞蹈等。