

»»» 高等学校教材

金秋颖 韩颖 © 主 编

社科信息 检索与利用



SHEKE XINXI
JIANSUO YU LIYONG

石油工业出版社

高等学校教材

社科信息检索与利用

金秋颖 韩 颖 主编

石油工业出版社

内 容 提 要

全书共分六章,主要介绍了信息检索的基本理论知识以及计算机信息检索系统和检索技术,中文常用的检索工具和检索数据库,大型的英文检索系统,电子书报、数据和事实、网上搜索引擎的检索方法和技术,信息利用以及论文写作的有关知识。

本书可作为高等学校文科专业信息检索与利用课程的教学用书,亦可作为社会科学科研人员信息检索的学习指导书。

图书在版编目(CIP)数据

社科信息检索与利用/金秋颖,韩颖主编.

北京:石油工业出版社,2006.6

(高等学校教材)

ISBN 7-5021-5557-0

I. 社…

II. ①金… ②韩…

III. 社会科学-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(2006)第 064426 号

出版发行:石油工业出版社

(北京安定门外安华里 2 区 1 号 100011)

网 址:www.petropub.cn

总 机:(010)64262233 发行部:(010)64210392

经 销:全国新华书店

排 版:北京乘设伟业科技排版中心

印 刷:石油工业出版社印刷厂

2006 年 6 月第 1 版 2006 年 6 月第 1 次印刷

787×1092 毫米 开本:1/16 印张:9

字数:232 千字 印数:1—3000 册

定价:18.00 元

(如出现印装质量问题,我社发行部负责调换)

版权所有,翻印必究

前 言

随着信息技术的飞速发展,数字图书馆、大型数据库、网上搜索引擎的不断开发和建设,高校信息检索与利用课程对培养大学生信息素质起到了至关重要的作用,尤其对培养大学生的自学能力、获取信息的能力、创造能力和动手能力具有积极的作用。这门课程注重实际的信息查找方法和技能的训练,强化课程中的实践检索环节,把理论知识联系到操作应用,落实到课题检索的实际过程,本书编者在多年从事信息检索课程教学的基础上,以查找文科信息为主要目的编写了本书。

全书共分六章:第一章信息检索基础知识,对信息资源及其结构、基本检索方法进行了概述;第二章介绍了计算机信息检索系统和检索技术基本知识;第三章介绍了中文常用的检索工具和数据库,包括全国报刊索引、中国人民大学复印报刊资料、CNKI 系列数据库中文科技期刊数据库、万方数据资源系统、高校财经数据库系统、中文社会科学引文索引等检索系统;第四章重点介绍了 EBSCO、Springer Link、EI、PQDD 等外文检索系统;第五章介绍了电子图书、电子报纸、数据和事实以及网上搜索引擎的检索方法和技术;第六章主要介绍信息分析与利用以及论文写作的有关知识。

本书有以下几个显著的特色:

(1)本书是文科信息检索工具书,它从形式上突破了手工检索方式,以数字图书馆为主要形式对网上信息资源及检索方法进行了介绍,是科研人员信息检索的工具书,并且可作为大学生的信息检索与应用课程教材。

(2)在编排体系上,针对典型的信息检索系统,将传统的手工检索工具、光盘数据库与网络数据库结合在一起进行介绍,力求做到深入浅出,给读者创造直观、鲜明的阅读环境。

(3)介绍了国内外重要数据库、电子图书和电子报纸、事实和数据以及网上搜索引擎的检索与使用。

(4)本书包括网上信息检索系统等技术内容介绍,对一般上网用户有重要指导作用。

(5)针对文科大学生的学科特点,重点放在文科信息资源检索技术上,有针对性地编写信息检索的有关内容,主要涉及国内外人文社科、经济、管理等领域。

全书的写作分工为:第一章由韩颖编写;第二章由金秋颖编写;第三章由金秋颖、孙宁编写;第四章由韩颖、金秋颖编写;第五章由金秋颖、王洪群编写;第六章由韩颖编写。全书由金秋颖、韩颖任主编,并负责内容的策划和汇总统稿,孙宁、王洪群任副主编。本书在编写过程中得到了王文广教授的关心和指导,并由王文广教授担任主审。

本书错谬之处,欢迎读者批评指正。

编 者

2006 年 4 月

目 录

第一章 信息检索基础	(1)
第一节 信息资源概述	(1)
第二节 信息资源构成	(6)
第三节 信息检索概述	(9)
第四节 信息检索工具	(12)
第五节 信息检索语言	(15)
第六节 信息检索程序与策略	(19)
第七节 信息检索效果评价	(21)
复习思考题	(22)
第二章 计算机检索系统	(23)
第一节 计算机检索概述	(23)
第二节 数据库及其种类	(23)
第三节 计算机检索技术	(26)
第四节 国际联机检索	(29)
第五节 光盘检索系统	(36)
复习思考题	(41)
第三章 中文检索系统	(42)
第一节 国内检索工具	(42)
第二节 全国报刊索引	(43)
第三节 中国人民大学复印报刊资料	(44)
第四节 CNKI 系列数据库	(45)
第五节 中文科技期刊数据库	(53)
第六节 万方数据资源系统	(55)
第七节 高校财经数据库系统	(61)
第八节 中文社会科学引文索引	(65)
复习思考题	(68)
第四章 国外检索系统	(70)
第一节 工程索引	(70)
第二节 Springer Link 检索系统	(74)
第三节 EBSCO 数据库检索系统	(78)
第四节 PQDD 博(硕)士论文数据库	(83)
复习思考题	(86)

第五章 电子书报和网上搜索引擎	(87)
第一节 电子图书	(87)
第二节 电子报纸	(92)
第三节 数据、事实检索	(94)
第四节 搜索引擎	(104)
复习思考题	(108)
第六章 信息的综合利用	(109)
第一节 信息的整理与利用	(109)
第二节 开题和结题报告的写作方法	(113)
第三节 毕业论文的写作方法	(117)
第四节 文献综述和述评的写作方法	(123)
第五节 学术论文的写作方法	(125)
第六节 学术论文投稿及相关信息的获取	(132)
复习思考题	(137)
参考文献	(138)

第一章 信息检索基础

现代信息技术迅猛发展,信息已成为人类社会发展的—种驱动力,人们越来越重视对信息资源的有效开发与利用。信息是一种极其重要的社会财富,信息同物质、能量构成人类社会的三大重要战略资源。物质提供材料,能量提供动力,信息提供知识与智慧。因此,信息已成为促进科技、社会、经济发展的新型资源,它不仅有助于人们不断地揭示客观世界,深化人们对客观世界的科学认识,消除人们在认识上的某种不定性,而且还源源不断地向人类提供生产知识的原料。

第一节 信息资源概述

一、信息

1. 信息的概念

信息作为一个科学术语最早出现于通信领域,20 世纪中叶后被引入哲学、信息论、系统论、控制论、情报学、经济学、管理学、计算机等领域。不同的学科的学者、专家以及有关领域对信息的定义都是从信息的受体、内涵和控制论等角度对信息的属性进行的描述。因此,信息是事物属性的再现。信息不是事物本身,而是由事物发出的、体现它存在和运动状态的信号、消息、指令和数据等所包含的内容。即广义的信息可定义为“信息是事物属性的表征”;而狭义的信息则可定义为“信息是指系统传输和处理的对象”。

2. 信息的属性

(1) 客观性与普遍性。

世界上—切事物都是在运动中,都有一定的运动状态和状态方式的改变,因而—切事物随时都在产生信息,即信息的产生源于事物,是客观事物普遍性的表征,它可以被感知和共享、被处理和存储、被传递和利用。

(2) 多样性与综合性。

信息在不同的领域具有多种不同的特性或表现形式,如客观事物中的各种自然属性,人工设备的技术特征,人类社会的各种社会特征,人脑中反映客观事物认识的思想、知识,人类交流信息过程中的声音、文字、图像以及用各种编码形式记录下来的数据、新闻、情报和消息等。各种形式的信息又常常以综合的方式表现事物的特征,所谓“多媒体”正是信息多样性和综合性的集中表现。

(3) 流动性与传递性。

信息在事物之间的相互联系必定在信息的流动中发生。信息的传递性表现在人与人之间的消息交换,人与自动机、自动机与自动机之间的信息交换,动物界和植物界的信号交换,同时,人类进化过程中的细胞、遗传也被看作是信息的传递与交换。

(4)相对性与有效性。

从信息作为事物相互联系的反映角度看,信息源不确定的程度或者信息源接受信息量的多少,均与信宿的状态有关。这一特征在人作为信宿接受信息的过程中表现得尤为明显。同一信息对具有不同认知水平的人所产生的作用和有效性也不同。

(5)积累性与价值性。

信息通过人脑思维或人工技术设备的综合、加工和处理,不断积累、丰富,提高其质量和利用价值。信息的质量和价值,实际上是对客观事物属性反映的深度和真实程度的认识。虽然信息是人类的一种重要资源,但信息只有被利用才会产生价值,否则,其价值或随时间的流逝而减少,或成为“信息垃圾”。

(6)中介性与共享性。

信息源于事物,但不是事物本身,是人们用来认识事物的媒介。信息能够共享是区别信息不同于物质和能量的最主要特征,即同一内容的信息在同一时间、同一地域可以被两个以上的用户分享,其分享的信息量不会因分享用户的多少而受影响,原有的信息量也不会因之而损失或减少。

3. 信息的类型

在人类社会和自然界里,不同领域对信息有不同的定义范围,根据不同的定义范围,从不同的角度来划分信息的种类。

(1)按信息产生的客体性质来划分,可分为自然信息与社会信息。

(2)按信息的社会属性划分,可分为政治信息、军事信息、经济信息、科技信息、管理信息和生活信息等。

(3)按信息传递范围划分,可分为公开信息、半公开信息和非公开信息。

(4)按信息的加工程度划分,可分为一次信息、二次信息、三次信息和高次信息。

(5)按价值观念划分,可分为有价值信息和无价值信息。

(6)按运动状态划分,可分为动态信息和静态信息。

(7)按信息所依附的载体形式划分,可分为文献信息、口头信息、电子信息等。

二、知识

1. 知识的概念

知识是人们在改造世界的实践中所获得的认识和经验的总和,知识的本质则是认知活动中的主体与客体的动态关系。知识是人类在认识和改造世界的社会实践中获得的对事物本质认识的成果和结晶,是人的主观世界对于客观世界的概况和如实反映,是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识,并通过人的大脑进行思维重新整合使信息系统化而构成知识。因此,人类不仅要通过信息感知世界、认识世界和改造世界,而且还要根据所获得的信息组成知识。由此可见,知识是信息的一部分。

2. 知识的属性

(1)意识性。

知识是一种观念形态的东西,只有人的大脑才能产生它、认识它、利用它,知识通常以概念、判断、推理、假说和预见等思维方式和范畴体系表现自身的存在。

(2)信息性。

信息是产生知识的原料,知识是经人类认识、理解并经思维重新整合后的系统化信息,知识是信息中的一部分。

(3) 实践性。

社会实践是一切知识产生的基础,也是检验知识的标准,科学知识对实践有重大的指导作用。

(4) 规律性。

人们对实践的认识,是一个无限的过程,人们在这种无限的过程中所获得的知识从一定的层面上揭示了事物及其运动过程的规律性。

(5) 继承性。

每一次新知识的产生,既是原有知识的继承利用、深化与发展,又是更新的知识产生的基础与前提。并且知识被记录或物化为劳动产品后,可以世代相传利用。

(6) 渗透性。

随着人类认识世界的不断深化,各种门类的知识可以互相渗透,形成了许多新的知识门类,构成了科学知识的网状结构体系。

国际经济合作与发展组织(OECD)出版的《以知识为基础的经济》报告中系统地提出人类现有的知识可分为四大类,即:

① 知道是什么的知识(Know—what),是指关于事实方面的知识,这类知识通常被近似地称为信息。

② 知道为什么的知识(Know—why),是指自然原理和规律方面的科学理论,这类知识的生产往往是专门的研究机构形成。

③ 知道怎么做的知识(Know—how),是指技艺或能力方面的知识,被称为技术诀窍或专有技术。许多企业的技术情报和商业秘密被归入这一类信息。

④ 知道是谁的知识(Know—who),是指谁知道和谁知道如何做某些事的信息,这在社会高度分工的经济时代中显得尤为重要,这类知识比任何其他种类的知识都更隐藏在企业内部。

三、文献

在《文献情报术语国际标准(草案)》(ISO/DIS 5127)中对文献是这样定义的,即“为了把人类知识传播开来和继承下去,人们用文字、图形、符号、声频、视频等手段将其记录下来,或写在纸上,或晒在蓝图上,或摄制在感光片上,或录到唱片上,或存储在磁盘上。这种附着在各种载体上的记录统称为文献”。

在“文献著录总则”(GB 3792.1—1983)中对文献是这样定义的,即“文献,记录有知识的一切载体”。

权威部门对文献颁布的这两种定义揭示了文献的特征。由此可知,文献由四要素构成,即文献信息、文献载体、符号系统和记录方式。四位一体不可分割,缺少任何一个都不能构成文献。

文献的属性主要有:

(1) 知识信息性。

这是文献的本质属性。任何文献都记录或传递一定的信息知识,离开知识信息,文献便不存在。传递信息、记录知识是文献的基本功能,人类的知识财富正是依靠文献得以保存、继承和传播的。

(2) 客观物质性。

文献所表达的知识信息内容必须借助一定的信息符号,依附一定的物质载体,才能长时期保存和传递。

(3) 人工记录性。

文献所蕴涵的知识信息是通过人们用各种方式将其记录在载体上的,而不是天然荷载于物质实体上的。

(4) 动态发展性。

文献并非处于静止状态,而是按新陈代谢的规律运动着。随着人类记录水平的不断提高,信息交流频繁,文献的数量日趋庞大,形式日益多样。与此同时,文献的老化速度也在加快,生命周期日益缩短,形成了有规律的运动。

由信息、知识和文献三者的概念可知,三者之间的关系是密切相关的。信息是物质存在的方式、形式和运动规律的表征。人脑对事物属性的感知形成信息,人们对信息集合加工、整理形成系统化表现形成人类社会实践的知识,知识被记录在载体上形成文献,文献被人类广泛传播、运用在理论和实践中又产生新的信息、知识和文献。如此循环不断创新形成各种各样的新知识,从而推动人类社会前进。由此可见,信息、知识、文献在社会系统中表现出一种不间断的延续性。

四、信息资源

在人类社会和自然界的运动发展过程中,产生着各种各样的信息。每一天都会有各种形式的信息层出不穷,这种大量的客观存在的人们直接或间接开发利用的信息集合总称为信息资源。

数字信息资源是信息资源的一种,是数字化了的信息资源,即以数字的形式,把文字、图形、图像、声音等多种形式的信息存放在光、磁等非印刷型介质上,以数字信号的形式传输,并通过相应的计算机和其他外部设备再现出来的一种信息资源。支持人们识别、接收、传递、加工、处理、利用信息的技术称之为信息技术。传感技术、通信技术、计算机技术和控制技术是信息技术的四大基本技术,其中计算机技术和通信技术是信息技术的两大支柱。

数字信息资源按不同的标准可分为若干类型。

1. 按信息的载体分类

(1) 联机网络信息资源。20 世纪 60 ~ 70 年代,世界上发达国家和地区相继建立起计算机联机信息服务系统,如美国的 Dialog、欧共体的 ESA 和德国的 STN 系统等,都为全世界联机用户提供了丰富的电子信息资源。但长期以来,由于费用昂贵,许多用户不敢问津。90 年代以来 Internet 迅猛发展,网络资源十分丰富,价格相对低廉,且有许多免费资源,这使得人们越来越多地从网络上检索、获取信息,通过网络来共享全球的信息资源。

(2) 单独发行的信息资源,以光盘出版物为主。

2. 按信息的媒体形式分类

(1) 文本信息资源。普通的文本信息资源的知识单元按线性顺序排列。阅读时,人们跟随文本的线性流向吸收其中的养分,遇到不懂的地方或想要知道详细情况时,就得暂时中断阅读,去查阅有关参考资料,这就打乱了文本固有的线性配置格局,在读者的头脑中形成了相互参阅的知识单元网状结构。然而,用户不易掌握和追踪这种网状结构,更难以对其修改和补充,仅靠手动、眼看、心记是具有极大的局限性和片面性的,超文本的出现为解决这一问题提供了手段。

(2) 超文本信息资源。超文本是一种人一机交互的友好系统。用户利用计算机可以增删内容,用户的想法可随时存入数据库中,也可随时检索、调用。超文本是一种通信,它可以组织许多研究人员交流思想,沟通情况,这相当于开一个小型电子会议。超文本信息资源是按知识

单元及其关系建立的知识结构网络。其数据库由节点和链路组成,查阅超文本信息资源时,以知识片段及其关系作为追踪、检索的依据。除了处理一般的文字信息外,还包括图片、地图和其他直观信息,超文本信息资源能够把文字信息和图像信息有机地结合在一起。

(3)多媒体信息资源。多媒体是包括文本、图像和声音在内的各种信息或传播形式的总称。多媒体信息能针对用户的需求提供各种形式的信息。它们可以是文本、图像(图表、图画、照片、动画或活动影视)、声音(语言、音乐或其他音响)以及它们的结合。由于计算机软硬件技术的限制,相当长时间以来计算机信息检索系统只限于存储和检索书目、文摘等线索型文献,多媒体的出现使得人们接受的信息资源不但图、文、声并茂,而且丰富多彩。

(4)超媒体信息资源。超媒体是超文本与多媒体两种技术的结合。一般说来,当超文本节点中的信息是多媒体信息时,即在信息浏览环境下,超文本的信息管理方式与多媒体的信息表现方法结合在一起时,就称为超媒体,它是超级媒体的简称。近几年来,超媒体技术发展迅速,在 Internet 上超媒体应用系统不断涌现。在超媒体信息系统中,不同类型的媒体信息能高度综合和集成,空间上图、文、声并茂,时间上媒体信息同步实现,有超文本和多媒体两种信息资源的特点,具有高度的交互性。

五、信息素质

信息素质(Information Literacy)一词最早是由美国信息产业协会主席 Paul Zurkowski 在 1974 年给美国政府报告中提出来的。他认为信息素质是人们在工作中运用信息、学习信息技术、利用信息解决问题的能力。

美国图书馆协会认为,信息素质是人们知道什么时候需要信息并找到、评价及有效地利用所需信息的能力。信息素质能力较强的人知道如何学习,因为他们了解知识是怎样组织的,知道如何找到信息,他们能够终生学习,因为他们能够发现所有与自己职责相关的决策所需要的信息。信息素质的内涵具体包括能意识到准确和完整的信息是决策的基础;了解信息需求及问题所在;制定信息检索策略;掌握信息检索技术;能评价信息;能根据实际用途组织信息;将新信息融会到现有知识结构中。

具有信息素质的人主要具有以下几个方面的能力:

(1)运用信息工具:熟练使用各种信息工具,特别是网上传播工具。

(2)获取信息:根据自己的学习目标有效地收集各种学习资料和信息,能熟练地运用、阅读、访问、参观、实验、检索等获取信息的方法。

(3)处理信息:能对收集的信息进行归纳、分类、存储记忆、鉴别、筛选、分析综合、抽象概括和表达等。

(4)生成信息:在信息搜集的基础上,准确地概述、综合、改造和表述所需要的信息,使之简洁明了,通俗流畅并富有个性。

(5)创造信息:在多种收集信息的交互作用的基础上,迸发创造思维的火花,产生新信息的生长点,从而创造新信息。

(6)信息增效:善于运用接受的信息解决问题,让信息发挥最大的社会效益和经济效益。

(7)信息协作:使信息和信息工具作为跨越时空的交往和合作的中介,成为高效延伸自己的手段,同外界建立多种和谐的协作关系。

(8)信息免疫:浩瀚的信息资源往往良莠不齐,需要正确的人生观、价值观、判别能力以及自控、自律和自我调节能力,能自觉地抵御和消除垃圾信息及有害信息的干扰和侵蚀,并完善

合乎时代的信息伦理素养。

因此,信息素质是指从各种信息源中检索、评价和使用信息的能力,是信息社会劳动者必须掌握的终身技能。信息素质的内涵具体包括能意识到准确和完整的信息是决策的基础;了解信息需求及问题所在;制定信息检索策略;掌握信息检索技术;能评价信息;能根据实际用途组织信息;使用信息,将新信息融会到现有知识结构中。

21 世纪是网络信息和知识大发展的世纪,在信息化社会中,无论是个人还是企业,信息素质是谋生存、求发展的重要因素。对于现代研究人才来讲,只有具备信息素质才懂得在信息化社会中如何去获取、加工、存储、检索和利用信息,使其拥有不断学习和持续发展的能力。

第二节 信息资源构成

信息资源的构成可从不同的层面和角度来划分。

一、按照信息的出版或加工形式划分

信息出版类型一般是指记录有知识的文献出版类型。一般将出版物文献划分为图书、报刊、会议文献、专利文献、科技报告、学位论文、技术档案、产品资料、标准文献和政府出版物。

1. 图书

图书大多是对已发表的成果、生产技术和经验或者基本知识领域系统地论述或概括,它往往以期刊论文、会议论文、研究报告及其他第一手资料为基本素材,经过作者的分析、归纳、组织而编写成的。不少科技图书的内容还包含一些从未发表过的研究成果或资料。

科技图书的特点是:内容比较系统、全面、成熟、可靠,具有一定的新颖性;但编辑出版时间过长,传递信息的速度太慢,包含的内容一般只是反映 2~5 年以前的研究水平。

科技图书是综合、积累和传递科技知识,教育和培养科技人才的一种重要工具,它可以帮助人们比较全面系统地了解特定领域的历史和现状,可以将人们正确地领入自己所不熟悉的领域,还可以作为一种经常性的查考工具。从信息检索角度来看,科技图书一般不作为主要检索对象。研究人员利用图书的比重比较小。美国有的信息专家曾经对美国各大学的科学家和英国电气工程师们进行过调查,发现在他们所阅读的各种科技文献中,图书的比重分别占 19% 和 14%。

2. 报刊

报刊是一种以印刷形式或其他形式逐次刊行的,通常有数字或年月日顺序编号的,并打算无限期地连续出版下去的出版物。

广义的报刊则包括一切定期刊行或不定期刊行的连续性出版物,如杂志、报纸、年度报告、年鉴、丛书以及学会的会议录、学报和纪要等。

报刊在科学技术活动中一直起着非常重要的作用,是交流的主要工具。报刊具有以下特点:数量大、品种多、内容丰富多样;出版周期短,报道速度较快;发行、流通广泛、连续性强,伴随着相应的学科领域发展而发展。

3. 会议文献

会议文献是指在国内外各种学术会议上交流的论文,以及由此汇编成册内部交流或公开出版的文献。

会议文献的主要特点是:传递信息比较及时,传递的信息针对性较强,它反映了某学科、专业的最新成果和发展现状及趋势,是研究工作不可缺少的情报源。

4. 专利文献

专利是用法律来保护科学技术发明创造的制度。专利文献是专利制度的产物,是一切与专利制度有关的各种专利文件的统称,包括发明说明书、专利说明书、专利局公报、专利文摘、专利分类与检索工具书,申请专利时提交的各种文件(如请求书、权利要求书、有关证书等)与专利有关的法律文件和诉讼资料等。狭义的专利文献一般指专利局颁布出版的各种发明说明书或专利说明书及其所派生的各种二次文献。

专利文献的特点:数量巨大,覆盖面广;格式统一,措辞严谨;描述对象具体,单一;技术内容新颖、可靠;文件类型多,重复量大,是重要的技术经济信息源。

5. 科技报告

科技报告是研究或设计单位向提供经费的上级部门提供的关于某项研究或设计任务完成情况 & 财务消耗情况的总结报告。

科技报告的特点:从形式上看,科技报告的出版形式比较特殊,每份报告自成一册,篇幅长短不等,有连续编号,装订简单,出版发行不规则。从内容上看,科技报告的内容比较新颖、详尽、专深。

6. 学位论文

学位论文是高等学校或研究机构培养的学生为获得某种学位而撰写的科学论文,一般有学士论文、硕士论文和博士论文。学位论文中除了少数可能发表在期刊或其他出版物以外,多数是不出版的。每篇学位论文有一复本保存在授予学位的学校的图书馆里,可供查阅。

7. 技术档案

技术档案是指在生产建设和科技部门的技术活动中形成的,有一定的工程对象的技术文件的总称。其内容包括:任务书、协议书、技术经济指标和审批文件、研究计划、方案、大纲和技术措施、有关的技术调查材料(原始记录、分析报告)、设计计算、试验项目、方案、数据和报告、设计图纸、工艺卡片以及应入档文件。

8. 产品资料

产品资料是指国内外各厂商为推销产品而印发的商业宣传品,包括产品样本、产品目录、产品说明书、厂商介绍、厂刊或外贸刊物、技术座谈资料等。

9. 标准文献

标准文献是以文件形式出现的标准化工作成果。经过公认的权威当局批准的标准化工作成果,可以采用文件形式或规定基本单位(物理常数)这两种形式固化下来的文件。标准化是为了有关各方的利益,特别是为了达到最佳的经济效果,并适当考虑到使用条件和安全要求,在有关各方的协作下,进行有步骤的特定活动所制定并实施各项规则的过程。

标准文献的特点:制定、审批有一定的程序;适用范围非常明确专一;编排格式、叙述方法严谨统一,措辞准确;技术上具有较充分的可靠性和现实性;对有关各方有约束性,在一定条件下具有某种法律效率;有一定的有效时间,需要随着技术发展而不断修订、补充或废除。

10. 政府出版物

政府出版物是各国政府部门及其所属机构所发表的文件。它的内容广泛,概括起来可分为行政性文件和科技文献两大类。行政文件包括国会记录、司法资料、方针政策、规章制度、决议、指示以及调查统计资料等。科技文献包括各部门的研究报告、技术政策文件等。

二、按照信息加工层次划分

人们在利用传递信息过程中,为了及时报道和揭示信息,对信息进行了不同层次的加工。按加工程度可将信息分为一次文献、二次文献和三次文献。

1. 一次文献

一次文献即以作者本人的生产与科研工作成果为依据而撰写的,并已公开发行进入社会流通使用的原始文献。如专著、学术论文、科技报告、会议论文、专利文献、学位论文等。一次文献的特点是具有学术上的新观点、新发明、新技术、新成果,提供了新的知识信息,是创造性劳动的结晶,有直接参考、借鉴和使用的价值,是人们检索和利用的主要目标。

2. 二次文献

二次文献即将大量的、分散的、无序的一次文献收集起来,按照一定的方法进行整理、浓缩和加工,使之系统化而形成各种目录、索引和文摘。即编制成具有多种检索途径的检索工具。因此,二次文献仅是对一次文献进行系统化的压缩,无新的知识产生,具有汇集性、检索性的特点。它的重要性在于提供了检索一次文献的线索。因此,二次文献又称为检索性文献。

3. 三次文献

三次文献即根据一定的目的和需求,在大量利用一、二次文献的基础上,对有关知识进行综合、分析、提炼、重组而再生的信息资源。如词典、手册、百科全书、年鉴、各种教科书及综述等。因此,三次文献具有综合性高、针对性强、系统性好、知识信息面广的特点。三次文献又称为参考性文献,有较高的使用价值,可直接参考、借鉴和利用。三次文献源于一次文献,又高于一次文献,是一种再创性文献。

从文献的角度看,一次文献是人们检索与利用的主要对象,二次文献是文献信息的检索工具,三次文献是人们考查数据、事实信息的主要信息源。

三、按信息内容划分

1. 按信息的表现形式划分

(1)文献型信息源,即存储语言文字形式信息的各种载体的集合。文献型信息源是目前信息内容最丰富、最可靠的信息,是人们使用最多的信息源。

(2)非文献型信息源,包括电子信息源、数值型信息源、声像型信息源、多媒体信息源和实物及口头信息源。

① 数值型信息源:存储数据形式信息的载体的集合。

② 声像型信息源:存储声音或图像信息的载体的集合,如:磁带、广播、电视。

③ 多媒体信息源:是一种时代发展的产物,它集文字、声音、图像于一体,多以光盘或 Internet 网上资源的形式出现,是目前发展最快、数量最多的一种信息源。

④ 实物及口头信息源:实物信息源是指自然实物和人工实物中所含信息的集合,口头信息源是指在交流、讨论、报告过程中所含的信息集合。

2. 按信息内容的保密程度划分

按信息内容的保密程度划分有公开信息源、半公开信息源和非公开信息源。

3. 按信息内容的应用范围划分

按信息内容的应用范围划分为科技信息源、经济信息源、管理信息源及生活信息源。

4. 按信息传输形式划分

(1)网络信息源:各种网络上的信息集合。

(2)非网络信息源:不用通信设施就能获得的信息集合。

第三节 信息检索概述

一、信息检索的概念

信息检索是指将信息按一定的方式进行加工、整理、组织并存储起来,再根据信息用户的需要找出有关信息的过程。它的全过程又叫信息存储与检索。这是广义的信息检索的含义,主要是对信息工作者而言的。狭义的信息检索则仅指后半部分,即用户根据需要,借助检索工具,从信息集合中找出所需要信息的过程。

信息检索是查找信息的方法和手段,它能使人们在浩如烟海的信息海洋中迅速地、准确地、全面地查找所需要的信息。可以说信息检索对人们的学习、生活和工作等方面都有非常大的作用。

二、信息检索的原理

人类的信息检索行为总是从特定的信息需求而开始,并在特定环境和信息检索系统中完成,这里所说的环境包括产生需求的环境、信息检索系统的运行环境和其他制约因素。特定的检索系统包括完成检索过程所需的一定设施和工具,它可以是图书馆、信息中心或信息经济人,也可以是某种工具书(如文摘索引、目录、资料集、手册、词典等)或机读信息源(如各种机读数据库)。

人类的信息需求千差万别,获取信息的方法也各种各样,但信息检索的基本原理却是相同的,可以把它最本质的部分概括为一句话:对信息集合与需求集合的匹配与选择。

根据信息检索的基本原理,实现信息检索的基本方式可分为传统信息检索和现代信息检索,传统信息检索,简称“手检”;现代信息检索,简称“机检”。

1. 传统信息检索

传统信息检索是检索人员利用手工检索工具手翻、眼看、大脑思维判别、索取原始文献的一种方式。

其优点是:(1)检索条件简单,成本低;(2)在检索过程中可以随时获取反馈信息,及时调整检索策略;(3)可对不同的检索工具同时进行对比,从而提高检索质量;(4)可以参阅检索工具中的附图。

其缺点是:(1)速度慢,效率低,检出的文献款目必须抄录;(2)手工检索工具提供的检索点有限,很难进行多元检索;(3)对于涉及几个概念组合的多主题的文献难于找到。

2. 现代信息检索

现代信息检索是检索人员利用计算机检索系统查找文献的一种检索方式。所谓计算机检索系统包括数据库技术、计算机技术和网络通信技术等。机检可以克服手检的缺点,但机检对设备条件的要求比较高,所需的投资比较大。计算机检索已从单机检索、联机检索发展到今天的网络检索,并向着智能化的方向发展。

三、信息检索研究对象

信息检索作为一门学科,有它自己的研究范围和对象,也有自己的理论、方法和技术。从总体上看,信息检索的研究对象是比较明确具体的,研究范围广泛而边界有些模糊,理论和方

法已经逐渐形成体系。

1. 信息检索的研究范围

信息检索的研究范围包括一切与信息存储与检索有关的系统、过程、理论和方法。一切可供存储和检索利用的信息类型,如文献、数据、事实、知识、声音、图形等,各种信息检索系统及其运行过程,如信息采集、标引、组织、存储、处理、匹配、传送等各种过程中使用的方法,以及在信息检索实践和研究的基础上形成的各种理论和假设,均包括在这个范围内。信息类型侧重于文献,其次是数据和事实。

2. 信息检索研究对象

(1)信息检索理论。它主要包括检索语言与标引理论,信息检索的数学模型,知识表示理论,相关性理论,以及有关的哲学问题。

(2)信息检索系统。它是实现信息检索的物质基础,是现实的研究对象,主要研究它的结构、功能、演变,它的设计开发技术、管理维护技术和评价技术,还研究它与其他信息系统乃至整个外部世界的关系。其中,数据库是信息检索系统的核心部分之一。数据库的建造和维护是一类非常重要的信息技术。数据库的质量直接影响信息检索系统的功能和效率。

(3)计算机信息检索。它涉及到许多计算机设备、软件技术、存储技术革新、检索技术、系统管理和经营知识,市场营销技术等,是一门综合性很强的技术。

(4)检索策略与方法。它是用户从信息检索系统中获取有关信息所必需的。好的检索策略是检索成功的必要前提,计算机的应用为检索策略和方法的发展提供了有力的支持。近30年来,涌现了许多新颖而有效的检索技术和方法,如布尔检索、位置检索、截词检索、加权检索、聚类检索等。人们利用、研究、评价和完善现有的各种检索策略和方法,研究开发新的更有效的策略和方法。

(5)用户研究与培训。用户是信息检索系统的生命。无论是系统的研制开发、管理维护、功能和服务的扩展,还是系统评价,都离不开用户研究工作。用户培训是用户研究的继续,是与用户建立紧密联系和发展新用户的一种非常有效的措施。

此外,还有自动标引、自动分类和自动摘录,以及相关设备等研究领域。

四、信息检索系统

信息检索系统是由一定的设备和信息集合构成,面向一定的用户,具有信息采集、组织、存储、选择和传播等功能。

1. 按照信息存储与检索使用的设备划分

其可分为手工信息检索系统和计算机信息检索系统两种类型。

(1)手工信息检索系统。

手工信息检索系统是在计算机检索系统出现以前进行信息检索的主要检索工具。主要是书本式检索系统,它包括一切以书刊形式提供,为人们查找各种信息或数据的出版物,如文摘杂志、题录或索引刊物、参考工具书等。书本式检索系统编制原理是计算机检索系统产生的基础。

(2)计算机信息检索系统。

计算机信息检索系统主要由计算机硬件和软件系统、数据库、数据通信等设备组成。根据内容不同计算机检索系统分为:联机检索系统、光盘检索系统、网络检索系统。

① 联机检索系统。

国际联机检索就是用户使用终端设备,远距离地从国际联机检索中心迅速而准确地获取数字文献信息资源,使知识信息得到广泛而有效的传播和利用。其实质是数据库和通信的结合。从20世纪60~70年代起,许多国家还先后建立了专门从事计算机检索的机构,如美国的洛克希德公司和系统发展公司,英国的目录检索服务处和图书馆自动化情报服务处与Infoline公司以9及欧洲的ESA等,这些机构都建有大量的数据库联机检索系统,都向全世界联机用户提供电子信息服务。其中,著名的系统有Dialog、ORBIT、ESA-IRS、OCLC、STN等。

② 光盘检索系统。

由于光盘在存储电子信息资源方面具有记录密度高、容量大、成本低、体积小、寿命长、可实现随机存取和检索费用低廉等优点,因此,光盘被广泛用于存储、检索数字信息资源,并产生了一批生产系列光盘的公司,如美国UMI公司和银盘公司等。光盘记载的数字信息资源并不局限于文献信息,还包括各种软件,但可用于检索仍以文献信息为主。

③ 网络检索系统。

Internet是网络通过互联而形成的全球网。它已延伸到地球上几乎每个国家。在Internet网上的所有主机都采用TCP/IP协议连接和通信,使网上各种计算机都遵循该协议所规定的方式进行数据交换,其结果是使得Internet信息资源主要包括电子报刊、电子新闻、电子报告、电子论坛、会议资料、各种软件资料、图像文件、声音文件和电子游戏等。Internet是目前世界上资料最多、门类最全、规模最大的信息库,是人们获取信息的重要来源。

2. 按功能划分

许多信息系统都含有检索的机制和过程。它们存储的数据文件或信息库,在响应用户提问时被处理,然后生成和输出答案。在这个意义上可以把它们看成是广义的检索系统。

(1)文献检索系统,它是狭义的信息检索系统。目前绝大多数信息检索系统都属于此类型。由于文献信息数量巨大,品种复杂繁多,连贯性强,故文献检索系统中的信息存储通常采用结构化的文件结构。检索的目的是要获得书目引文和原始出版物,而不是现实的数据。大型商业性的文献检索系统通常都是专门开发的,检索功能较强,既适用于书目数据和文本信息的存储与检索,也可以用于事实和数值型数据的存取。一方面检索系统主要面向文献资源的管理者和利用者,其用户主要是科研和教学人员。此外,它还能对各种管理和决策活动提供有效的信息支持。

(2)数据库管理系统(Data Base Management System,简称DBMS)是为任何数据库的建立、操作和维护而设计的计算机软件。它最显著的特征是:存储的数据具有确定的结构,即把数据或信息存储在结构化的数据文件中,系统只能使用那些结构完全确定的数据项。这是它与文献检索系统的主要区别。DBMS主要面向数据查询和基层事务管理,其主要用户是行政职员或事务管理员。

(3)自动问答系统(Question Answering System,简称QAS)是一种能直接回答人们提出的具体问题的一种计算机系统,曾被称为事实检索系统,目前已归入专家系统的范畴,问答系统的主要用途是支持或代替专家决策,为知识的采集、存储和利用服务。它采用的知识库技术和自然语言处理技术对文献检索系统和DBMS有潜在的应用价值。问答系统是信息检索系统的高级形式。

(4)管理信息系统(Management Information System,简称MIS)是建立在数学方法(统计理论、计算方法)、管理科学和信息技术的基础上,对各种用于管理的信息进行采集、加工、分析、整理、