

WENXIAN XINXI JIANSUO YU LIYONG JIAOCHENG
KEJI WENXIAN SHUJUKU YU
OA KEJI ZIYUAN DE FAZHAN LIYONG

文献信息检索 与利用教程

——科技文献数据库与
OA科技资源的发展利用

李霞 胡文忠 王文棣 主编



北京邮电大学出版社
www.buptpress.com

文献信息检索与利用教程

——科技文献数据库与 OA 科技资源的发展利用

主 编 李 霞 胡文忠 王文祿
参 编 胡雯君 巩文峰

北京邮电大学出版社

· 北 京 ·

内 容 简 介

本教材概述了文献信息检索的相关概念、基本原理、方法及基本技术;重点介绍了几种重要的中文文献数据库和外文文献数据库及使用技巧,详细论述了免费网络科技信息资源(OA 资源)的发展状况及其获取途径;同时增加了学术论文写作和学术论文投稿方面的知识介绍。

图书在版编目(CIP)数据

文献信息检索与利用教程——科技文献数据库与 OA 科技资源的发展利用/李霞,胡文忠,王文棣主编. —北京:北京邮电大学出版社,2010.3

ISBN 978-7-5635-2233-0

I. ①文… II. ①李…②胡…③王… III. ①科技情报—情报检索—教材 IV. ①G252.2

中国版本图书馆 CIP 数据核字(2010)第 015828 号

书 名: 文献信息检索与利用教程——科技文献数据库与 OA 科技资源的发展利用

主 编: 李 霞 胡文忠 王文棣

责任编辑: 崔 珞

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号(邮编:100876)

发 行 部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 北京忠信诚胶印厂

开 本: 787 mm×1 092 mm 1/16

印 张: 17.25

字 数: 405 千字

印 数: 1—3 000 册

版 次: 2010 年 3 月第 1 版 2010 年 3 月第 1 次印刷

ISBN 978-7-5635-2233-0

定 价: 32.00 元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

序

文献信息检索是获取知识信息的捷径。掌握文献信息检索的方法与技能,可以帮助人们快捷、准确、全面地获取所需知识,最大限度地节省查找时间,使文献信息得以充分的利用;文献信息检索是科学研究的向导。要进行有价值的科学研究,必须依赖文献检索,全面获取相关文献信息,及时了解各学科领域出现的新问题、新观点,以确定自己的研究起点和研究目标;掌握文献信息检索技能还是终身教育的基础。科技的迅速发展使知识的总量呈指数增长,知识的陈旧速度也明显加快,这就要求人们要具有终身学习的能力,必须具备一定的寻找、获取有效信息的能力,这种能力在很大程度上决定于信息需求者对新知识的敏感力和接受力,因此具备文献信息检索的技能是现代人终身学习的需要,也是一个科学工作者知识创新的基础。

当代文献具有数量多、增长快,出版形式多、文种多,文献异常分散,文献失效加快等特点。同时,信息技术、计算机技术、网络资源的发展也给文献信息检索带来了空前的变化。要适应这些变化,更好地利用文献信息资料,发挥文献信息的作用和功能,就必须掌握文献信息、文献检索的基本理论和基本知识,从而为全面掌握文献检索的技能打好基础。

由李霞、胡文忠、王文棣主编的《文献信息检索与利用教程——科技文献数据库与OA科技资源的发展利用》一书,内容包括了文献信息检索的概念范畴及检索原理、主要中文数据库、主要外文数据库、免费网络科技信息资源、学术论文的撰写与投稿、核心期刊及其评价以及根据检索目的而专设的实验实习内容七大部分,本书学科体系完整,特点鲜明,直观生动,既体现了作者扎实、深厚的专业功底,又反映了严谨、求实、创新的学风。我们期待着该专著能够在信息浩瀚的时代对读者信息搜寻和获取能力的提高有所帮助。

刘学录

前言

文献信息检索(Information Retrieval)的含义有广义和狭义之分。广义的文献信息检索是指将信息按一定的方式组织和存储起来,由信息用户根据需求找出有关信息的过程,所以它的全称又叫“信息的存储与检索”(Information Storage and Retrieval)。狭义的文献信息检索则仅指该过程的后半部分,即信息用户从信息集合中找出所需要的信息的过程,相当于人们通常所说的信息查寻(Information Search)。无论是狭义的信息查寻,还是广义的文献信息检索,共同的基础都是对信息的收集、处理、分析和利用的能力。

随着科学技术的进步和社会经济的发展,文献检索也在不断地发展。这种发展其主要特征表现在:一方面,文献数量在急剧增长,文献类型在不断多样化,文献的存储方式日新月异;另一方面,文献的利用者在很多时候也是文献的创造者。所以,文献信息检索技能在科学研究,乃至日常生产生活中的应用越来越广泛,作用日益突出,尤其是在学术交流中,离开了文献检索几乎寸步难行。

计算机信息检索,是指以计算机技术为手段,通过光盘和联机等现代检索方式进行信息检索的方法。与手工检索一样,计算机信息检索能力已经成为科技人员的一项基本功,这一能力的训练和培养对科技人员适应社会发展,做好科学研究、技术推广和服务社会等工作极其重要。一个善于从电子信息系统中获取信息的科研人员,必定比不具备这一能力的人有更多的成功机会。国际社会已经将互联网络检索专家作为未来十大热门职业之一。

为了适应文献检索的这些新特点和新变化,本书的重点为网络信息资源的检索,与同类专著比较,新增了网络免费资源的检索,即开放索取(OA)网络资源的介绍,学术论文的撰写与交流的基本知识,以及学术期刊评价及核心期刊的相关知识。本书大量利用文献数据库检索页面截图介绍的方法,直观生动;同时增加了针对性的以检索目的不同的典型案例检索的实践内容。本书是一本深入浅出的实用性和理论性并重的专业论著,可以作为非图书情报专业本科生、研究者的学习材料,也可以作为科技人员的检索工具书。

全书由七部分组成,其中第一章、第三章、第五章及附录由李霞副研究馆员撰写;第二章、第六章第一节由胡文忠副编审撰写;第四章第一、三节,第六章第二、三节由胡雯君编辑撰写;第四章第二、四节由王文棣副教授和巩文峰讲师撰写。由于编者水平有限,疏漏之处在所难免,恳请读者批评指正。

作者

目 录

第一章 文献信息检索概论	1
第一节 基本概念	1
第二节 文献信息检索概述	7
第二章 中文文献数据库	20
第一节 中文文献数据库概述	20
第二节 中国知网 CNKI	21
第三节 万方数据库	66
第三章 外文文献数据库	96
第一节 外文数据库概述	96
第二节 SpringerLink	105
第三节 Ovid	112
第四章 免费网络科技信息资源	125
第一节 Open Access	125
第二节 搜索引擎	146
第三节 在线字词典	160
第四节 特种文献	193
第五章 学术论文的撰写与投稿	218
第一节 学术论文的概念与特点	218
第二节 学术论文类型与写作格式	220
第三节 学术论文的投稿和发表	232
第六章 核心期刊及期刊评价	235
第一节 核心期刊的概念	235

第二节	期刊影响因子及其他评价指标.....	236
第三节	国内外主要核心期刊及文献评价系统.....	238
实习报告.....		238
实习一	以了解科研动态为目的的检索实习.....	243
实习二	以撰写综述为目的的检索实习.....	249
实习三	以查新论证为目的的检索实习.....	253
实习四	以查找专利文献为目的的检索实习.....	259
实习五	以传统图书馆利用为目的的检索实习.....	263
实习六	以编辑采集电子文献信息为目的的检索实习	265
参考文献.....		266

第一章 文献信息检索概论

第一节 基本概念

一、文献

(一) 文献的概念

“文献”一词最早见于《论语·八佾》，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。《虞夏书·益稷》也有相关的引证说明“文献”一词的原意是指典籍与宿贤。

宋代马端临《文献通考》中将文与献作为叙事与论事的依据：“文”是经、史历代会要及百家传记之书；“献”是臣僚奏疏、诸儒之评论、名流之燕谈、稗官之记录。在他的影响之下，宋代之后对文献的认识仅限于一般的文字记载，将不能表达为文字记载的东西，不称为文献。

狭义地讲，文献是图书、期刊等各种出版物的总和，但随着时代的发展，文献的内涵和外延都远远超出了其原来的含义，现代的文献更多地具有信息情报的特征。因此，广义地讲，文献是用文字、图形、符号、声频、视频等技术手段记录人类知识信息的一种载体，它是固化在一定物质载体上的知识。因此，GB/T4894—1985 中将文献定义为：记录知识（信息）的一切载体。

(二) 现代文献的构成要素及功能

1. 文献的构成要素

文献是记录、积累、传播和继承知识的最有效手段，是人类社会活动中获取情报的最基本、最主要的来源，也是交流传播情报的最基本手段。从文献的概念和属性可以看出：文献既不是知识本身，也不是记录知识的物质载体本身，而是由两者构成的以知识为内容，载体为外部形式的组合体。因此，知识性的内容、文献信息的记录符号系统、文献的记录方式以及文献的载体形式就成为文献构成的 4 大要素。

知识内容是指文献中所记载的人类的思想意识、知识信息等，是文献存在的根本，也是文献的实质所在。文献是人类思想和智慧的结晶，人们生产、传递、保存和利用文献的主要目的，都是为了获得和利用其中所记载的知识内容。如果离开了知识内容，文献就失去了存在的基础，所有有关文献的活动也都失去了意义，因此，知识内容是文献最基本的



组成要素。

信息符号是信息内容的标记和表达手段。现代文献的信息符号主要是指语言文字、声频、视频和编码等。语言文字的产生标志着人类文明的进步和发展,是人们交流最通用的信息符号,也是传递知识信息的主要书写形式。声频和视频能够反映较直观的信息内容,易被人们直接识别和理解。编码符号是经过人工编排处理和加工转换过的字符,一般用于计算机软件编程,是人和计算机都能识别处理的信息符号,它使信息内容的加工存储、传递检索的效率显著提高。

载体材料是指知识内容所依附的物质媒介。世界上的物质材料很多,它们并不是都能记录和传播知识信息的,只有具备记录和传播知识信息的物质材料才能被称为载体材料。载体材料大致可分为3种:一是记忆有知识信息的人脑,在人脑中的知识信息通过人们相互间的语言交流或言传身教而被传播;二是承载有知识信息的实物,依附于古代文物、样品和建筑等实物之中的知识信息,必须通过人们对实物的进一步分析研究,才能被认识和吸收;三是通过一定的方式专门记录知识信息的物质材料,如古代常用的载体有甲骨、金属、石、竹简、帛、纸张,现代的有胶片、磁带、优盘等。应当注意的是,只有将知识信息以一定的方式记录在专门的载体上,才能称为文献,而前两种载体——人脑和实物,一般不称为文献。

记录手段是指将标识知识信息的符号记录在载体材料上的方法。文献的记录手段是随着科技的发展而改变的,不同的历史时期有不同的记录方式出现。文献最早的记录手段是刻划、手写,后来有了印刷、打字、拍摄、录制和复印,随着计算机的发明和使用,又有了复制、刻录和计算机输入等方法,但从古至今,印刷物一直是文献记录的主要手段。

2. 文献的功能

文献具有多种功能,但根本功能是它的存储功能和传递功能。人类在改造世界的实践中所获得的知识成果,需要通过文献记录保存下来,所以说文献是知识的宝库,它是人类世世代代留传下来知识的结晶,是人类文明发展的见证。如果没有文献,人类的知识就不能够被集中、延续和继承,人类文明就难以继续向前发展。在人类历史上,由于文献的缺失,使许多古代文明成为今天失落的文明,如“玛雅文明”和“西夏文明”的消逝都是活生生的例子。此外,文献在人们日常生活中主要作为知识信息的传递工具而存在的,它的传递功能有纵、横两个方向。纵的方向,从古及今,在时间维度,文献为人类知识的继承提供了条件,人们通过阅读文献就可以了解前人的思想感情和实践经验,使今人可以不必重复前人已经做过的事情。横的方向,文献打破了空间地域的界限,使人类知识得以广泛地传播和融合,成为联系世界和沟通全人类思想的纽带。总之,文献的存储功能和传递功能是相辅相成的,正是借助于这两种功能,文献才能够跨越空间和时间,成为人类文明进程中不可或缺的知识财富。

(三) 文献的分类

随着信息时代的发展,现代文献的数量正在以几何倍数增长,载体形式五花八门,因此,从不同的应用标准和角度,对文献的划分也不同。目前,人们对于文献的划分主要有以下几种类型。



1. 根据载体划分

根据载体形式划分可分为:印刷型、缩微型、机读型和视听型文献。

印刷型文献是一种传统的文献形式,也是文献的最基本的载体形式。它主要包括铅印、油印、胶印、石印等各种资料。这种文献形式之所以能从古代沿用至今是因为它有便于阅读和流传的优点。

缩微型文献是以感光材料为载体的文献,又可分为缩微胶卷和缩微平片,优点是体积小、成本低、便于保存、转移和传递,但阅读时须用阅读器。由于此类文献存在阅读不便的缺点,至今也未成为文献的主要形式。

机读型文献是一种新形式的载体,它主要通过编码和程序设计,把文献变成符号和机器语言,输入计算机,存储在磁带或磁盘上,阅读时再由计算机输出。此类文献能存储大量情报,可按任何形式组织这些情报,并能以极快的速度从中取出所需的情报。近年来出现的电子图书即属于这种类型。

视听型文献是以声音和图像形式记录在载体上的文献,如唱片、录音带、录像带、科技电影、幻灯片等。

2. 根据出版形式及内容划分

根据出版形式及内容可分为:图书、连续性出版物、特种文献。

(1) 图书:指凡篇幅达到 48 页以上并构成一个书目单元的文献称为图书。

(2) 连续性出版物:包含期刊、报纸、年度出版物等。

(3) 特种文献:包括科技报告、会议文献、专利文献、学位论文、标准文献、政府出版物、档案资料、产品资料等。

3. 根据文献的加工层次

根据文献的加工层次可分为:零次文献、一次文献、二次文献、三次文献等。

(1) 零次文献。没有经过加工的用于非正式交流的零散文献,具有零散性,主要特征是没有经过正式公开发表的文字性材料。如实验记录、上课笔记、文章手稿、书信等文字材料。

(2) 一次文献。凡以作者本人的生产和科研工作成果为依据而创作的原始文献,称为一次文献。判断文献是否是一次文献,是根据文献的内容而不是它的载体形式。只要是作者根据自己的科研成果而发表的原始创作,不论是手稿、印刷品、缩微品、声像制品、机读资料等,都是一次文献。

(3) 二次文献。二次文献是对一次文献进行加工整理和提炼压缩的产物。为了便于读者全面了解和准确查找所需的一次文献,将分散的原始文献按照一定的体例排列、加工整理,描述其形式特征,揭示其内容要点,以便广泛、系统、完整地报道某一学科、专题领域的文献资料。如各种书目、索引、文摘、题录、简介等,二次文献的主要功能是检索与通报一次文献。

(4) 三次文献。三次文献是对原始文献群的内容进行系统分析、综合评述而编写的文献资料,是高度浓缩加工的再生科研文献,它包括综述研究和参考工具两大类型。

从零次文献、一次文献、二次文献到三次文献,是一个由分散到集中,由无序到有序,由博而精的对知识信息进行不同层次加工的过程。它们所含信息的质和量是不同的,对于改善人们的知识结构所起到的作用也不同。零次文献和一次文献是最基本的信息源,



是文献信息检索和利用的主要对象;二次文献是一次文献的集中提炼和有序化,它是文献信息检索的工具;三次文献是把分散的零次文献、一次文献、二次文献,按照专题或知识的门类进行综合分析加工而成的成果,是高度浓缩的文献信息,它既是文献信息检索和利用的对象,又可作为检索文献信息的工具。

二、情报

(一) 情报的概念

情报一词渊源于拉丁文,原意为消息传递。英文中的 information 一词,其基本含义也是指经过传递或交流的原先不知道的信息。我国古代兵书中虽然没有出现“情报”一词,但《孙子兵法》中的“先知”、《兵经》中的“知”就是指情报,具有情报的含义。从情报字面了解,“情”即情况、消息、知识等,“报”即报道、传递、交流等,因此可将情报定义为:可被传递的有价值的信息。

(二) 情报的属性

从情报的定义可以总结出,情报应该具有三方面的属性,即信息性、价值性和传递性。

情报的信息性,是指情报必定是信息,不是信息的情报是不存在的。情报具有信息所包含的一切属性。情报是信息的一部分,即信息范围大,信息包含情报;情报范围小,情报只是信息的一部分。因此,可以说凡情报都是信息,但不可以说凡信息都是情报,只能说有的信息是情报。情报是能够用于进行思想交流的社会信息,是人们在社会实践中认识客观世界规律性所获得的知识。概言之,情报是针对特定的对象需求而搜集、加工、传递,并在实践中具有参考、借鉴和继承作用的新知识。

情报的价值性,是指情报往往可以用来作为决策和行动的依据,肯定其具有的或大或小的用处,要判断情报的价值大小,就要从情报的本身的真实性,情报对所需人群的针对性,以及情报的新颖性、时效性和完整性 5 方面综合进行评价。

情报的传递性,是指情报一定是经过传递或交流的,未经传递或交流的信息都不是或还不是情报。

三、信息

(一) 信息的概念

“信息”一词是汉语固有的,而非舶来词汇。南唐李中《暮春怀故人》诗:“梦断美人沉信息,目穿长路倚楼臺”,这里的“沉消息”即杳无消息之意。宋陈亮《梅花》诗:“欲传春信息,不怕雪埋藏。”《水浒传》第四十四回:“宋江大喜,说道:只有贤弟去得快,旬日便知信息。”等例句中,信息都具有“消息”、“音信”等含义。《汉典》中将信息解释为:“1)音信消息。2)现代科学指事物发出的消息、指令、数据、符号等所包含的内容。人通过获得、识别自然界和社会的不同信息来区别不同事物,得以认识和改造世界。在一切通信和控制系统中,信息是一种普遍联系的形式”。

从信息的发展历史看,信息的定义已有 100 多年的历史,目前有关信息的定义有过多



过泛之势,但大多都将解释当成定义。现在通常认为信息是指传递的消息,信息广泛存在于自然界、生物界和人类社会。信息是多种多样、多方面、多层次的,具有自己的结构和成分。由于从不同的角度人们对信息的认识不同,所以对于信息至今仍然没有统一的定义。在众多定义中,最经典和最有代表性的有:“信息是事物及其属性标识的集合”;“信息是物质、能量、信息及其属性的标示”;“信息是确定性的增加,即肯定性的确认”。这些基于信息学的经典定义,虽然都从信息的本质与属性方面对信息加以了定义,但这些定义都有些抽象和难懂。因此,有人从信息检索领域及其应用的角度对信息作出了以下相对而言比较通俗的定义:信息是对客观世界中各种事物的变化和特征的反映,是客观事物之间相互作用和联系的表现,是客观事物经过感知或认识后的再现。

(二) 信息的属性

理解信息的概念,仅从信息学专业抽象的定义理解还是比较困难的,而从信息的属性方面归纳理解,信息的概念就会变得易懂和熟悉。信息具有多种属性,有的学者提出的属性多达12种之多,如抽象性、积累性、可变性、流动性、无量纲性、共享性、可重复性、可传递性、可复制性、可分割性、可压缩性、可存储性等。刘廷元在其专著《信息检索教程》中就将信息的本质属性概括为:反映性、无量纲性、不守恒性和可共享性。他认为:“在人们提出的这些属性中,有四个属性是最基本的,它是信息不同于物质和能量的根本区别,即物质和能量是本原的、有量纲的、守恒的、独占的,而信息则是反映性的、无量纲的、不守恒的和可共享的。”虽然对于信息的属性有诸多的论述,但人们可以感知和容易理解的主要属性可以归纳为以下几点。

(1) 客观性。信息是客观存在的,它源于客观存在的物质及其运动。信息的客观性还表现在它是以物质的客观存在为前提,即使是主观信息,如决策、判断、计划、指令等,也具有其客观的背景。

(2) 传递性。信息的产生同信息的传递是联系在一起的,信息只有在传递过程中才能发挥作用。信息的传递过程是一个重复使用的流通过程,在这一过程中,信息的占有者不会因传递信息而失去信息,一般说来,也不会因多次使用而改变信息的自身价值。

(3) 依附性。信息的存在必须有一定的物质载体,信息依靠载体进行交流和传递。印刷物、声音、光电波,以及胶片、磁带、光盘、人等生物体无一不是信息的载体。

(4) 可塑性。信息可以加工处理,可以压缩、扩充和叠加,也可以变换形态。在流通和使用过程中,经过综合、分析、再加工,原始信息可以变成二次信息和三次信息;原有的信息价值也可以实现增值;为了有效地交流和传递,借助于先进的信息技术,文本、图像、数字、语音等各种形态的信息均可实现互相转换。

(5) 时效性。现代社会中,信息的使用周期正在迅速缩短,信息的价值实现取决于及时地把握和运用信息。信息是活跃的,不断变化的,及时地获取有效的信息将获得信息的最佳价值,如时效性很强的天气预报、经济信息、交易信息、科学信息等。不能及时地使用最新信息,信息的价值就会随其滞后使用的时差而贬值。

(6) 共享性。信息的共享性主要表现在同一内容的信息可以在同一时间由两个或两



个以上的使用者使用,而信息的提供者并不失去所提供的信息内容和信息量,它是信息资源的广泛提供与利用的基础。

四、知识

知识到底是什么,目前仍然有争议。我国对知识的定义一般是从哲学角度作出的,如在,《博弈圣经》中知识的描述是“把识别万物实体与性质的是与不是,定义为知识”。《中国大百科全书·教育》中“知识”条目是这样表述的:“所谓知识,就它反映的内容而言,是客观事物的属性与联系的反映,是客观世界在人脑中的主观映像。就它的反映活动形式而言,有时表现为主体对事物的感性知觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识”。从这一定义中可以看出,知识是主客体相互统一的产物。它来源于外部世界,所以知识是客观的;但是知识本身并不是客观现实,而是事物的特征与联系在人脑中的反映,是客观事物的一种主观表征,知识是在主客体相互作用的基础上,通过人脑的反映活动而产生的。

根据当代著名科学哲学家卡尔·波普尔的“三个世界”理论,知识可以分成两种,即主观知识和客观知识。主观知识主要是指人们头脑中的经验、观点、思想。在文字产生以前,知识主要以主观知识的形式出现,但随着信息载体发展,主观知识实现主体外存储,从而产生了通过文字、图像等记录于载体的客观知识,它是人类的一个重大进步。主观知识可以随着人的死亡而消失,而客观知识可以代代继承,实现跨越时空的交流、传递与开发利用,因而客观知识的作用和意义不可低估。人类文明的进步,主要是客观科学知识的作用,它也是当代社会信息资源的主体结构。

五、文献、情报、信息及知识的关系

信息是自然界、人类社会以及人类思维活动中普遍存在的。不同的事物有不同的特征,这些特征会通过一定的物质形式(如声波、电磁波、图像等)给人带来某种信息。哲学家们认为信息是物质的一种普通属性,是物质存在的方式和运动的规律与特点。

知识是人类社会实践经验的总结,是人的主观世界对于客观世界的概括和如实反映。知识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,是人的大脑通过思维重新组合的系统化的信息集合。因此,人类通过信息感知世界、认识世界和改造世界,而且要根据所获得的信息组成知识。可见知识是信息的一部分,是一种特定的人类信息。人类的认识,并非客观世界的简单复制和临摹,人类在接受信息时是有区别和有选择的。人类生活环境中普遍存在的信息,这些信息只有经过人脑的接收、选择、处理才能组合成新的信息系统(系统化的信息),即新的知识。新的知识首先发生并存在于人脑中,这就是主观知识。但为了适应人类的交往,知识就应具备为感官所能接受的形式——即借助语言符号来加以表现。一旦将人们头脑中的认识结果通过某种物质载体(纸张、胶片、录音带、磁带等)记录下来,就变成了可以传递的客观知识。随着人类认识的深入发展,这种客观知识已逐步形成为较完整的知识体系——科学体系,这是人类认识客观事物的过程,也是人脑对外界事物传来的信息的加工过程,而认识的飞跃结果即为知识,知识是一



种观念形态的东西,只有人的大脑才能产生它、识别它和利用它。

人的知识是在后天的实践中形成的,知识源于信息,但信息不等于知识。知识是信息通过大脑所产生的结果。换言之,人们认识客观事物的过程就是人脑对外界事物传来的信息进行加工的过程,而认识的飞跃的结果即为知识。也就是说信息被有选择地收入到人们的思维系统,经过人们大脑的储存、甄别、加工、处理、转换等形式而形成知识。

知识与信息的关系是以认识主体为中介的相互转化关系:一方面,信息作为事物的属性、关系和意义的表征是知识得以形成的中介和原料,信息经过主体的生理结构和认知结构的加工处理可以转化为知识;另一方面,知识也可以通过一定手段和途径转化为信息,以信息作为自己的存在形式和传播方式,并以信息的方式发挥作用。在信息时代,知识只有转化成为信息产品才具有价值。

情报是针对特定的对象需求而搜集、加工、传递,并在实践中具有参考、借鉴和继承作用的新知识。

因此从逻辑关系上讲,知识、情报、文献的本质都是信息,它们都是信息的一部分。知识属于信息这一大范畴,是经过人脑思维整合的信息;情报是知识的一部分,知识中一部分被传递着用于特定的目的时就变成了情报;文献是记录有知识的一切载体,是知识中的一部分(客观知识)被人类用特定的载体记载下来而形成的。这几个概念之间的关系如图 1-1 所示。

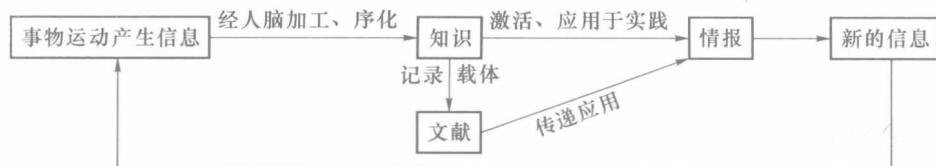


图 1-1

第二节 文献信息检索概述

一、文献信息检索的概念

在信息社会中,信息的不断产生和膨胀,使信息资源不断丰富、庞大。信息资源的共享使人们找到了避免重复劳动、吸取和利用他人研究成果的捷径。面对这日益庞大的可共享的信息资源,人们必须通过一种科学的方法从中获取自己所需的那部分信息,这就是文献信息检索。文献信息检索(Document Retrieval)是在科学研究、撰写论文时所必需的一种获取资料的手段,其概念有广义和狭义之分。

广义的文献信息检索包括文献信息的存储和检索两个过程(Storage and Retrieval)。存储过程是将大量无序的文献信息集中起来,根据文献信息源的外表特征和内容特征,进行加工、整理、分类、浓缩、标引等处理,使其系统化、有序化,并按一定的技术要求建成一



个具有检索功能的数据库或检索系统,供人们检索和利用;而检索过程是指运用编制好的检索工具或检索系统,查找出满足用户要求的特定信息,如图 1-2 所示。

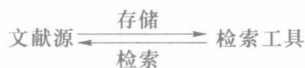


图 1-2

存储过程需要对文献信息进行加工,这种加工处理过程就是对文献特征进行描述的过程,即描述文献信息的内部特征和外部特征,并将这些特征加以标识,再按一定规律组织储存起来,构成完整检索系统的过程。在这里标识就是将从文献中提取的特征进行标注、表达,即用语词或符号来表征文献信息的内部(内容)特征和外部(形式)特征。

狭义的信息检索是指广义信息检索的下半流程,即指依据一定的方法,从已经组织好的大量有关文献信息集合中,查找并获取特定相关文献信息的过程。因此不难看出,狭义的信息检索其实是针对信息用户而言的,这也是通常意义上的文献信息检索过程。

简单地讲,检索的本质就是信息用户的需求和信息集合的比较与选择,即匹配的过程。它是将检索提问标识与存储在数据库中的标引标识进行比较,两者一致或者信息标引的标识包含着检索提问标识,则具有该标识的信息就从数据库中输出,输出的信息就是检索命中的信息。

二、文献信息检索的类型及检索工具

(一) 文献信息检索的类型

1. 按存储和检索的对象划分

(1) 文献信息检索(Document Information Retrieval)

凡是利用目录、文摘或索引等二次信息查找有关信息及这些信息的出处和收藏单位等,都属于文献信息检索,它是将存储于数据库中的关于某一主题文献的线索查找出来的检索,它通常通过目录、索引、文摘等二次文献信息为工具,以原始文献的出处为检索目的,向用户提供有关原文献的信息。文献信息检索是利用检索工具查出相关文献的过程。检索系统不直接解答用户提出的问题,而是提供与之相关的文献名称及出处,供用户筛选使用。

(2) 数据信息检索(Data Information Retrieval)

凡是利用参考工具书、数据库等检索工具检索包含在文献中的某一数据、参数、公式或化学分子式等的检索过程就称为数据信息检索。它是将经过选择、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。数据信息检索的对象是数值性数据,即具有数量性质并以数位形式表示的量化信息。这些数值型数据各种各样,包括物理性能数据、化学物质数据、环境监测数据、各种统计数据、自然资源数据、经济和社会数据等。数据检索不仅能查出数据,而且能提供一定的运算、推导能力。用户可通过使用经过选择、整理、鉴定的数值数据库检索获得一些量化信息,从而为定量分析提供依据。数据信息检索是一种确定性的检索,它在科学研究、工程计算、质量控制、决策管理等方面发挥着重要作用。

(3) 事实信息检索(Fact Information Retrieval)



凡是利用百科全书、数据库等检索工具从存储事实的信息系统中查找出特定事实的过程称为事实信息检索,它是将存储于数据库中的关于某一事件发生的时间、地点、经过等信息查找出来的检索。它既包含数值数据的检索、运算、推导,也包括事实信息、概念等的检索、比较、逻辑判断。事实信息检索的对象是已经存在的各种事实的有关资料,它是数值信息和系统数据信息的混合。一般先从系统中检索出所需信息,然后再加以逻辑推理才能给出结论。

2. 按使用的工具和手段划分

(1) 手工检索(Manual Retrieval)

手工检索是指通过手工的方式进行信息的存储和检索的过程,主要使用的是目录、索引、文摘和各类工具书。优点是灵活性高、费用低、准确性好。缺点是速度慢、受时空限制因素多、途径少、更新周期慢。但对于一些内容专深、回溯期长的研究课题,手工检索依旧是必要的。

(2) 计算机检索(Machine Retrieval)

计算机检索是现代文献信息检索的主要检索方式,是在利用计算机和通讯设备的基础上,通过对各种检索系统(数据库)的检索,查找所需文献信息的方法。检索过程由人机协同完成,具体检索指令由人通过输入计算机而发出,繁复的检索匹配过程则由计算机完成,这样大大减轻了以往人工检索匹配的繁重工作,提高了检索效率。计算机检索与手工检索相比,其本质没有发生根本的改变,变化的只是信息的载体形式、检索手段、存储方式和匹配方法。

3. 系统中信息的组织方式划分

(1) 全文检索(Full Text Retrieval)

全文检索技术是20世纪50年代末产生的信息检索技术。全文检索系统是为人们提供了一条获取文献原文而非文献线索的途径。全文检索是可对存储于数据库中的整本书、整篇文章中的任意内容信息进行的检索。用户可以根据自己的需求从中获得有关的章、节、段、句、词等信息,并且还可以进行各种字词的频率统计和内容分析。

(2) 超文本检索(Hyper Text Retrieval)和超媒体检索(Hyper Media Retrieval)

超文本和超媒体检索是对存储的文本、图像、声音等多媒体信息进行的检索。传统的文本都是线性的,用户必须按顺序阅读。超文本和超媒体却与此不同,它是一种非线性的网状结构,强调中心节点之间的语义联结,用户可沿着交叉链选择阅读自己感兴趣的部分。早期的超文本以文字为主,随着多媒体技术的发展,图像、音频、视频等各种信息大量链接,形成了一种网状的动态信息网,信息的存储结构从单维发展到多维、有向的链接,存储空间范围更大。

(二) 检索工具

检索的迅速、准确是对文献检索工具的基本要求。检索工具所覆盖的某一方面、某一领域、某一时间段信息资源的全面与否以及检索手段的快捷、科学与否是判断该检索工具好坏的重要标准。

信息检索的目的就是获取有价值的情报信息。信息检索工具可以分为传统纸质检索



工具和电子信息检索工具两大类。纸质检索工具是经过一系列的判断、选择、标引、分类、组织、阐释等处理后制作而成的。它是在一次文献的基础上,经过加工整理、编辑而成的二次文献或三次文献。

1. 传统纸质检索工具

纸质检索工具又分为目录、题录、文摘和索引 4 种。

(1) 目录。目录是对图书、期刊或单独成册的出版物外表特征的著录。包括分类目录、书名目录、著者目录、主题目录。目录的种类很多,如国家书目、联合目录、馆藏目录等,网络上的目录型搜索引擎等。

(2) 题录。题录是对单篇文献外表特征的揭示和报道。著录项目一般有篇名、著者、著者单位、文献来源、语种等。由于著录项目比较简单,因此收录范围广,报道速度快,是用来查找最新文献资料的重要工具。

(3) 索引。将书刊资料所刊登的文章的题目、作者、出处及所论及的主题,按一定的原则和方法编排而成。

(4) 文摘。在索引的基础上对原始文献用简明、扼要、准确的文字所作的摘录,让读者了解文献的基本内容。文摘是二次文献的核心,文摘分为指示性摘要和报道性摘要(此两种文摘的区别请看第五章)。

2. 电子信息检索工具

(1) 电子信息资源网站。如国内的万方数据资源系统、中国期刊网、重庆维普、超星数字图书馆等。

(2) 光盘数据库。主要包括文献型数据库、数值型数据库、事实型数据库以及术语型数据库。

(3) 搜索引擎。广义的搜索引擎泛指网络(尤其是万维网)上提供信息检索服务的工具或系统,即在因特网上或通过因特网响应用户的搜索请求,返回相应查询结果的信息技术和系统。狭义的搜索引擎主要指利用网络自动搜索软件或人工方式,对万维网信息资源进行采集、分析与标引,并将索引信息组织成数据库,以网站形式为网络用户提供检索服务的一类信息服务系统。

(4) 电子词典。电子词典主要分为掌上词典、个人电脑词典和在线词典。我国的电子词典自 20 世纪 80 年代末诞生以来得到了迅速发展并且显示出强大的生命力。掌上词典的主要用户是大中学生。以金山词霸为代表的个人电脑词典是目前市场上电脑词典的主流,而在线词典则是目前发展势头很快的一种在线查阅的工具。

三、文献信息检索的方法、途径、步骤和语言

(一) 文献检索的方法

1. 顺查法

顺查法是一种“由远及近”的检索方法。就是从过去某一时间往现在按时间顺序逐年地检索。如果将所查文献信息按时间顺序加以分析,就可以看出研究对象的发展过程及