

普通高等教育“十一五”立项教材

信息检索导论

主 编 付兴奎

副主编 杨晓光 王春梅 李阿丽

编 委 (按姓氏笔画为序)

王靖会 王春梅 王春梅 付兴奎

安秀敏 陈文勇 张宏亮 张立娜

李亚波 杨晓光 李阿丽 林 娜

吉林大学出版社

图书在版编目 (CIP) 数据

信息检索导论/付兴奎主编. —长春: 吉林大学出版社,
2008. 6

ISBN 978-7-5601-3074-3

I. 信… II. 付… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字 (2008) 第 085558 号

书 名: 信息检索导论

作 者: 付兴奎 主编

责任编辑、责任校对: 孟亚黎

吉林大学出版社出版、发行

开本: 787×1092 毫米 1/16

印张: 15.875 字数: 292 千字

ISBN 978-7-5601-3074-3

封面设计: 孙 群

吉林科华印刷厂印刷

2008 年 6 月第 1 版

2008 年 6 月第 1 次印刷

定价: 26.00 元

版权所有 翻印必究

社址: 长春市明德路 421 号 邮编: 130021

发行部电话: 0431-88499826

网址: <http://www.jlup.com.cn>

E-mail: jlup@mail.jlu.edu.cn



文献检索与利用课开设初期,为了满足本校教学需要,馆内的几位同志曾经编写了本教材。仓促上阵,不尽人意之处可想而知。不久,开始使用系列教材,内容、结构与印刷质量等都有所改善。但基本上仍是讲手检,以介绍常用书刊为主线。由于没有出版保证,实际使用时间不长。接着再次自编或借用其他教材,因为缺乏稳定性,既不利于提高教学质量,也带来不少麻烦。上个世纪末年,曾经参与过几校合编教材,虽然彼此合作得很好,但在持续利用上也有其难处。如今还得自编。这种自编→(大)统一→自编→(小)统一→自编的过程,其他开课院校可能同样或部分地经历过。尽管将来仍有可能实行某种形式的统一,当前以自编满足教学需要,乃是必要的可行措施。说“自编”,是从由本馆人员编写,主要供本校教学使用角度而言。其实,经过几个阶段的磨练,已不是前两次的简单重复,而是在诸多有利因素构成的螺旋形上升形势下的一次新创作。既然公开出版,自也会扩大社会交流,有被他校参考利用的可能。那么,这也可以说是为将来的统一注入一点力量。

这本书和前述几个阶段使用的教材比较起来,最突出之处是机检理念贯彻全书,直接阐释机检的就占了12章中的7章。在这网络信息迅速发展与普及的时代,让大学生能够适应新形势,能够充分利用日新的检索条件,这是完全必要的。但是,当今毕竟处在由手检走向机检的过渡阶段,在强调机检的同时,适当照顾手检,从实际情况出发,让大学生也能驾轻就熟,用两章篇幅介绍印刷本中外文检索刊物,也是完全必要的。着眼于社会现实、关注大学生当今和未来一段时间里的需求,予他们能在不同环境下“渔”的知识和技能,乃是文献检索课当今的立足点。

高等农业院校同志们编写的检索书籍,具有农业方面特色是很自然的事。不过把这本书和前述几个阶段教材比较起来,这一特色已有所淡化,这是因为机检占了重要地位,着重阐述具有通论性质的机检方法,摆脱了对与农业相关的检索刊物的详细剖析,从而省去了大量的农业文献示例。实际上,农业科学有其广阔的领域,每一学习本课的大学生,只与其中的个别部分相关,书中专业示例能与他们所学专业对应的可能性很少,关键在于授课教师当堂联系实际地举例。本书已较侧重于通论方法的阐述是可取的。通用化也许是文献检索课教材的发展趋势。

书中以两章篇幅讲述信息利用和论文写作,颇有见地。检索的目的是利用信息,尽管信息检索是信息利用中的重要部分,但毕竟只是个部分。大学生应当同时了解其他部分,识其全貌,掌握全过程。强调课程名称和内容相应而排斥利用部分,既是对信息利用的割裂,也脱离了大学生需求的实际。一些调查表明,目前高中毕业生对信息利用仍不甚了了,于论文写作也不太熟悉。大学阶段在这方面予他们以培训,对他们学习有益,于将来工作也有益。自然,开有写作课的院系没有必要重复。

信息检索有其理论,对非图书情报专业大学生说来,较为重要的是其方法部分,尤其是实用技巧。文献检索课教材就应该侧重于这方面,本书对此也处理得较好。这是轻视理论,非也,此乃因人制宜。信息检索有其学术水平,有其世界与国内先进水平,有其国内其他高校的经验,这些当然应是编写这类书时需要考虑的,但更重要的是所面对主要读者的水平。过与不及,皆非所宜,相应的越好越恰当。因为文献检索课是为现实的大学生服务的,绝对不能脱离实际。

信息检索涉及计算机、数据库、用户等诸多动态因素与新技术,发展快、变数多,能尽快地反映相关方面新事物才是好书。希望同志们紧跟时代步伐,监控相关方面新进展,为修订充实再版做准备!

江乃武

2004年7月

目 录

第一章 信息检索概述	(1)
第一节 信息的概念与特征	(1)
第二节 信息的功能	(4)
第三节 信息源的类型和特点	(5)
第四节 知识、信息和文献	(10)
第五节 信息检索	(11)
第二章 计算机信息检索	(18)
第一节 计算机信息检索发展概况	(18)
第二节 计算机系统与网络知识	(20)
第三节 计算机信息管理与数据库知识	(24)
第四节 机检系统的组成与类型	(28)
第五节 计算机信息检索原理	(32)
第三章 常用中文文献光盘数据库	(41)
第一节 中国期刊全文数据库(CJFD)	(41)
第二节 中国人民大学书报资料中心复印报刊资料数据库	(48)
第三节 中文科技期刊数据库	(50)
第四节 全国报刊索引光盘数据库	(54)
第五节 万方数据库系列	(57)
第四章 常用西文光盘数据库检索	(63)
第一节 AGRIS(国际农业科技信息系统)	(63)
第二节 AGRICOLA(美国国家农业图书馆书目数据库)	(72)
第三节 CAB(国际农业生物科学中心数据库)	(73)
第四节 BA on CD (美国生物学文摘光盘数据库)	(75)
第五节 CA(美国化学文摘数据库)	(77)
第六节 EI(工程索引数据库)	(84)
第五章 网络信息资源检索	(97)
第一节 网络检索基础	(97)
第二节 Internet 提供的服务	(102)
第三节 Web 信息检索工具	(106)
第四节 网络信息检索发展展望	(111)
第六章 搜索引擎概述	(114)
第一节 概述	(114)
第二节 综合性搜索引擎	(117)
第三节 著名搜索引擎的使用	(120)
第四节 专业型搜索引擎	(125)

第七章 网络信息资源综合利用	(132)
第一节 Internet 网上信息资源开发与利用的对策	(132)
第二节 查找电子期刊	(134)
第三节 查找免费专利资源	(139)
第四节 查找标准信息	(147)
第五节 查找会议文献信息	(149)
第六节 查找科技报告	(150)
第七节 查找学位论文	(154)
第八节 公司及产品的中文信息	(155)
第九节 网上词典和物性数据	(157)
第十节 在 Internet 上寻人	(158)
第十一节 网上搜索小技巧	(160)
第八章 中文检索工具介绍	(165)
第一节 《全国报刊索引》	(165)
第二节 《中国农业文摘》	(167)
第三节 《中国生物学文摘》	(169)
第四节 《中国药学文摘》	(171)
第五节 《中国学术期刊文摘》	(173)
第六节 《中国专利公报》及《中国专利索引》	(174)
第七节 其他中文检索工具	(179)
第九章 外文检索工具介绍	(183)
第一节 美国《化学文摘》	(183)
第二节 美国《工程索引》	(193)
第三节 国际农业和生物科学中心文摘	(196)
第四节 美国《生物学文摘》	(201)
第五节 《食品科学与技术文摘》	(205)
第六节 《国际药学文摘》	(207)
第七节 《德温特生物技术文摘》	(209)
第十章 科技信息服务机构	(213)
第一节 科技信息服务机构概述	(213)
第二节 科技文献信息服务机构简介	(214)
第十一章 信息研究与利用	(219)
第一节 信息调查与资料的收集、鉴别与整理	(219)
第二节 信息分析方法	(231)
第三节 信息调研报告	(232)
第十二章 科技论文写作	(236)
第一节 科技论文概述	(236)
第二节 科技论文的选题	(237)
第三节 科技论文的撰写	(241)
第四节 科技论文的基本格式	(243)

第一章

信息检索概述



第一节 信息的概念与特征

一、信息的概念

信息的概念是十分广泛的。世间万物的运动,人间万象的更迭,都离不开信息的作用。李白的诗“日照香炉生紫烟,遥看瀑布挂前川,飞流直下三千尺,疑是银河落九天。”给我们带来了庐山瀑布的信息;苏东坡的词“大江东去,浪滔尽,千古风流人物。”给我们传递的是赤壁怀古的信息。

信息的概念是十分普遍的。客观世界中存在着各种各样的信息,生命的进化也需要信息,人类的生活更是需要信息。没有信息,千变万化的事物之间就没有了联系,也就没有大千世界的统一。

在某种意义上说,不掌握信息手段就不成其为人类了。从购物到旅游,从升学到择业,直至收购一家跨国公司,生活就是对信息的收集、分析和在此基础上的行动。而信息处理其实就是我们每一个人日常生活中最常见的活动了。值得注意的是,现在的信息交流和传递的载体、方式等发生了巨大的变化。使我们坐在家就能知道几分钟前发生在世界各地的事情。

信息作为科学的概念,首先是在信息论中得以专门研究的。信息论是一门年轻的科学,关于信息论的研究工作可以说是从通信工程研究开始的。1928年,哈特莱(R. V. L. Hartley)在《贝尔系统技术杂志》上发表了一篇为《信息传输》的论文。在这篇论文中,哈特莱把“信息”理解为选择通信符号的方式。例如,假定他在符号表中选择了这样一些符号:“I am well.”他就发出了“我平安”的信息;如果他选择了“I am sick.”这些符号,他就发出了“我病了”的信息。他还注意到,不管符号所代表的意义是什么,只要从符号表中选择的符号数目一定,发信者所能发出的信息的数量就被限定了。哈特莱的思想和研究成果,为信息论的创立奠定了基础。

我国汉语中很早就有“信息”这个词。早在一千多年前,唐朝诗人立中在《碧云集·暮春怀故人》一诗中就留下了“蒙断美人沉信息,目穿长路倚楼台”的佳句。当时,“信息”指的是音信、消息。

所谓信息,并非指事物本身,而是指用来表现事物特征的一种普遍形式。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/状态的直接或间接的表述。”从本质上说,信息是事物自身显示其存在方式和运动状态的属性,是客观存在的事物现象。但是,信息与认知主体又有着密切的关系,他必须通过主体认知才能被反映和揭示。这表明,信息是一种比运动、时间、空间等概念更高级的哲学范畴,是一个复杂的、多层次的概念。

实际上,信息的概念是有层次的。在信息概念的诸多层次中,最重要的是两个层次:一个是没有任何约束条件的本体论层次,另一个是受主体约束的认识论层次。从本体论层次上来考察,信息是一种客观存在的现象,是事物的运动状态及其变化方式。世间一切事物都在不停运动,因此,都在不断地产生着本体意义上的信息。站在主体的立场来考察信息概念,就会引出认识论层次上的信息定义:信息就是主体所感知或所表述的事物运动状态及其变化方式,是反映出来的客观事物的属性。

信息作为一门严密的科学,主要应归功于贝尔实验室的美国科学家克劳德·申农(Claude E. Shannon)1948年在他的著名的论文《通信的数学理论》中把信息解释为“两次不定性之差”,即通信的意义在于消除某种不定性。该论文成为信息论诞生的标志。申农认为,信息的多少意味着消除了的不确定性的多少。所谓不确定性,就是对客观事物的不了解、不肯定。通信的直接目的就是要消除接收端(信宿)对于发出端(信源)可能回发出那些消息的不确定性。因此,信息被看作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。比如人们收听广播,听到了一些新闻,也就是接收到了一些消息。这些消息的内容可能是已经知道的,也可能是还不知道的。事先已经知道的消息不是信息,因为人们不能从中获得新内容或新知识以消除不确定性。在接受者看来,信息必须是事先不知道其内容的新消息。由此可见,申农的信息定义是从信息在通信过程中的作用角度提出来的。

几乎与申农同时,维纳(N. Wiener)也发表了《控制论——或关于在动物和机器中控制和通信的科学》一文,标志着控制论这门新兴学科的诞生。维纳认为,信息就是我们适应外部世界,并把这种适应反作用于外部世界的过程中同外部世界进行相互联系、相互作用、相互交换的一种内容。维纳的理论为我们提供了一条深入揭示信息本质的正确途径。

从此,“信息”的概念才被广泛应用。世界上每个人都需要有关信息来知道衣食住行。购置商品需要商品价格和质量方面的信息;体育比赛需要掌握最新的战术、运动员状况、对方的运动员身体状况的信息。每一个组织管理机构要想使该组织获得成功,取决于是否有效地组织了各分支部门之间的信息交流。每一个决策部门都需要利用国内外信息,以确保决策的科学性和准确性。无论是科学技术研究,还是生产力的发展,都要利用蕴涵了新知识和新内容的信息,避免重复无谓的劳动,才跟得上时代的步伐。

二、信息的特征

所谓信息的特征,就是指信息区别于其他事物的本质属性。

信息的基本特征是:

1. 普遍性

信息是事物运动的状态和方式,只要有事物的运动,就会有其运动的状态和方式,就存在着信息。无论在自然界、人类社会,还是在人类思维领域,绝对的“真空”是不存在的,绝对不运动的事物也是没有的。因此,信息是普遍存在着的。信息与物质、能量一起,构成了客观事物的三大要素。

2. 表征性

信息不是客观事物本身,而只是事物运动状态和存在方式的表征。一切事物都会产生信息,信息就是表征所有事物属性、状态、内在联系与相互作用的一种普遍形式。宇宙时空中的事物是无限的,表征事物的信息现象也是无限的。

3. 动态性

客观事物本身都在不停地运动变化,信息也在不断发展更新。特别是从语用信息的观点来看,事物运动状态及方式的效用是会随时间的推移而改变的。因此,在获取与利用信息时必须树立时效观念,不能一劳永逸。

4. 相对性

客观上信息是无限的,但相对于认识主体来说,人们实际获得的信息总是有限的。并且,由于不同主体有着不同的感受能力、不同的理解能力和不同的目的性,因此,从同一事物中获取的信息肯定是不同的,也就是说,实得的信息量是因人而异的。

5. 依存性

信息本身是看不见、摸不着的,他必须依附于一定的物质形式之上,不可能脱离物质单独存在。我们把这些以承载信息为主要任务的物质形式称为信息的载体。信息没有语言、文字、图像、符号等记录手段便不能表述。没有物质载体便不能存储和传播,但其内容并不因记录或物质载体的改变而发生变化。

6. 可传递性

信息可以通过多种渠道、采用多种方式进行传递,我们把信息从时间或空间上的某一点向其他点移动的过程称之为信息传递。信息传递要借助于一定的物质载体,因此,实现信息传递功能的载体又称为信息媒介。一个完整的信息传递过程必须具备信源(信息的发出方)、信宿(信息的接受方)、信道(媒介)和信息四个基本要素。

7. 可干扰性

信息是通过信道进行传递的。信道既是通信系统不可缺少的组成部分,同时又对信息传递有干扰和阻碍作用。我们把任何不属于信源加之于其信号上的附加物都称之为信息干扰。例如索引数据库提供的线索型信息,不少信息受到标题、作者、主观的分类和主题的标志等等干扰,会给信息用户的分析和判断产生或多或少的干扰和阻碍。

8. 可加工性

信息可以被分析或综合,扩充或浓缩,也就是人们所说的对信息进行加工和处理。所谓信息加工,就是把信息从一种形式变换成另一种形式,同时在这个过程中保持一定的信息量。如果在信息加工过程中没有任何信息量的增加或损失,并且信息内容保持不变,那么就意味着这个信息加工过程是可逆的,反之就是不可逆的。实际上信息加工都是不可逆的

过程。

9. 可共享性

信息区别于物质的一个重要特征就是它可以被共同占有,共同享用,也就是说,信息在传递过程中不但可以被信源和信宿共同拥有,而且还可以被众多的信宿同时接收使用。物质交换遵循易物交换原则,失去一物才能得到一物;信息交换的双方不仅不会失去原有的信息,而且还会增加原有的信息;信息还可以广泛地传播扩散,供众多用户共享。

第二节 信息的功能

1. 信息是感知世界的中介

信息是介于物质世界和精神世界之间过渡状态的东西,是人们用来认识事物、感知世界的不可缺少的中间环节。它贯穿于认知活动的始终,认知过程本身就是一个以信息为中介的信息运动过程。人类认识世界和改造世界的过程,是一个不断从客观世界获得信息,并对信息进行加工处理,形成新的认知结构,然后通过实践活动反作用于客观世界的过程。

2. 信息是管理决策的依据

管理决策是一个动态过程,其程序一般包括发现问题、确定目标、制定方案、评估选优、实施决策、追踪反馈等环节。决策需要综合众多因素,但决定性的因素是取决于对客观实际的了解,对未来形势及后果的正确判断,而这些都需要依赖于全面、及时和准确的信息分析研究。信息活动贯穿于科学决策的全过程,并渗透到决策过程的每一个环节。

随着社会的发展,信息对于管理决策的作用日趋重要,这是因为社会的庞大和复杂对信息的依赖越来越多,而管理决策的正确与否,将直接影响整个社会各个系统。

3. 信息是科学研究的必要条件

人类知识的继承性和共享性使得任何一项科学研究都必须借鉴前人的研究成果和依靠同时代其他人的帮助。这就是说,科研工作需要在时间上和空间上的信息传递。

从另一方面来说,世界本来是一个统一的整体,人们为了研究的方便,人为地把统一客观世界划分成若干个学科领域。如今这种分割阻碍了科学的整体化的发展,也不利于各门具体学科的纵深发展。于是,便出现了科学的“微分化”和“积分化”的趋势,产生了一批交叉学科、边缘学科。多学科的知识协作和发展需要信息的联接和融合。

4. 信息是社会发展的资源

人类在使用物质资源和能量资源的基础上,开始重视生产、处理、传递和利用信息的能力,信息资源与物质资源、能量资源一起,共同构成了现代人类社会资源体系的三大支柱。物质作为材料,能量作为动力,信息作为知识和智慧,正如一个人的体质、体力和智力,只有三者健全发展的人,才是一个真正健康的人。信息资源是人类借以对其他资源进行有效管理的工具,它在推动社会经济发展、促进人类社会进步等方面正发挥着日益重要的作用。

第三节 信息源的类型和特点

信息源是用户获取信息的来源。联合国教科文组织(UNESCO)把信息源定义为:“个人为满足其信息需要而获得信息的来源。”这显然是从信息使用者角度来说的。从绝对意义上看,只有信息产生的“源头”,才能称作信息源。

一、以载体材料、存储技术和传递方式划分信息类型

文献信息源以载体材料、存储技术和传递方式为依据可分为印刷型、缩微型、声像型和电子型。

1. 印刷型

以纸质材料为载体,采用各种印刷术把文字或图像记录存储在纸张上而形成。它既是文献信息资源的传统形式,也是现代文献信息资源的主要形式之一。主要特点是便于阅读和流通,但因载体材料所存储的信息密度低,占据空间大,难以实现加工利用的自动化。

2. 缩微型

以感光材料为载体,采用光学缩微技术将文字或图像记录存储在感光材料上,有缩微平片、缩微胶卷和缩微卡片之分。主要特点有:存储密度高、体积小、重量轻、便于收藏;生产迅速,成本低廉;须借助缩微阅读机才能阅读,设备投资较大。现在可以通过计算机缩微输入机(CIM)把缩微品上的信息转换成数字信息存储在计算机中,使缩微品转换为磁带备用,也可以通过计算机输出缩微机(COM)把来自计算机中的信息转换成光信号,摄录在缩微平片或胶卷上,摄录速度可达每秒 12 万字符,大大缩短了缩微型信息资源的制作周期。

3. 声像型

以磁性和光学材料为载体,采用磁录技术和光录技术将声音和图像记录存储在磁性或光学材料上,主要包括唱片、录音录像带、电影胶卷、幻灯片等。主要特点是存储信息密度高,用有声语言和图像传递信息,内容直观,表达力强,易被接受和理解,但须借助于一定的设备才能阅读。

4. 电子型

按其载体材料、存储技术和传递方式,主要有联机型、光盘型和网络型之分。联机型以磁性材料为载体,采用计算机技术和磁性存储技术,把文字或图像信息记录在磁带、磁盘、磁鼓等载体上,使用计算机及其通信网络,通过程序控制将存入的有关信息读取出来。光盘型以特殊光敏材料制成的光盘为载体,将文字、声音、图像等信息采用激光技术、计算机技术刻录在光盘盘面上,使用计算机和光盘驱动器,将有关的信息读取出来。网络型是利用 Internet 中的各种网络数据库读取有关信息。电子型信息资源具有存储信息密度高,读取速度快,易于网络化和网络化程度高、高速度、远距离传输信息的特点,使人类知识信息的共享能得到最大限度的实现。

二、以文献的加工深度划分信息类型

科技文献按照对其加工层次的不同可以分为一次文献、二次文献和三次文献。这主要

是根据文献内容信息含量有无变化而划分的。

1. 一次文献

凡未通过翻译、缩编或被第三者评价等手段所过滤的原始文献,不论撰写时是否参考或引用别人的资料,不论载体和出版类型,不论是否正式出版,均为一次文献。其特点是含有前所未有的发明创造,或者是一些新的见解与理论。一次文献包括期刊论文、专利文献、会议文献、学位论文和科技报告等。一次文献是对知识的第一次加工,是信息的基础,也叫信息源。

2. 二次文献

二次文献是为了一定的目的,将分散的无组织的一次文献经过收集选择、加工整理、简化组织工作,形成检索工具。如著录文献特征(篇名、著者、分类号、关键词、文献出处等),摘录内容要点经重新合理编排成为系统的文献,以便查找与利用,如书目、索引、文摘等,即所谓的检索工具。二次文献通常是由图书信息机构组织人力、物力编辑出版的,起着汇集信息、提炼信息和提供信息、处理信息和信息线索的作用。它是查找一次文献的线索和桥梁。

3. 三次文献

三次文献是利用二次文献并在其指导下,对大量一次文献的内容进行综合分析、研究加工后编写成的文献,是对知识的第三次加工,属于这类文献的有年鉴、进展、述评、综述、手册、指南、专著等。三次文献称之为高级信息产物,它是在充分研究已经发表文献的基础上,对某一专题已经取得的成果、进展或述评深入探讨,或预测发展趋势,因此反映的信息量大,系统性强,对了解某专题领域的研究水平与动态具有指导意义。篇末附有该专题领域的大量参考文献,为读者提供了该专题领域的主要参考文献的线索,因而又具有文献检索工具的功能。

从一次文献到二次文献、三次文献的过程是一个由博而精,由分散到集中,由无组织到有系统的过程。从文献检索来说,一次文献是检索的对象,二次文献是检索的手段与工具;从时间来说,一次文献的发表早于二、三次文献;从准确性来说,一次文献比二、三次文献提供的数据准确性强。三者功能不同,可根据不同需要分别选用。

三、以撰写的目的和文体划分信息类型

文献信息源以撰写的目的和文体划分,主要可分为著作、学术论文、专利说明书、科技报告、技术标准、科技档案、产品资料等。

按照我国 1983 年制定的国家标准(GB3469—83),文献类型被分为 26 类。除以文字记录、印刷出版的文献外,还有光盘、缩微胶卷、缩微平片、录音带、录像带、唱片、电影片、幻灯片等。目前最大量的还是印刷型文献,通常被分成十大类,号称“十大信息源”。

1. 科技图书

科技图书,是对科技知识、科技成果、经验总结等进行系统全面的概括与论述,并以书本(按印刷业要求每本不少于 50 页,少于者称为小册子)形式出版发行。由于经过著者的选择、核对、鉴别和系统组织,因而每本书的内容比较成熟、可靠和完整,可归为三次文献。它是传播科技知识、教育和培养科技人才的主要手段,是综合、积累和传送科技知识的一种重要信息源。但由于科技图书的知识积累和编辑出版的时间较长,其传递信息的速度较慢,新颖性较差,一般适用于教育培训。尽管如此,它仍然是一种重要的科技信息的载体,一种重

要的科技信息源。

2. 科技期刊

科技期刊指定期或不定期出版的连续性出版物,出版周期最长不超过一年。它采用固定的名称和统一的版式外形,有卷期号或年月顺序号,每期内容不重复。

科技期刊的特点是数量大、品种多、内容新颖、学科广、信息多、文种多、报道迅速、反映及时。它是获取科技信息的主要来源。特别是一些重点期刊或称核心期刊,由于其发表专业文献密度大、学术性强,报道的研究成果较新、质量较高,能及时反映学科发展的最新动向和科学研究的先进水平,因此使用寿命、利用频率和效益都较高,成为科学交流的主要工具。有人测量过大约有 65% 的信息来自科技期刊,它是重要的科技信息源。

3. 科技报告

科技报告也称研究报告或科技总结报告。它是有关某项研究工作的情况或成果의正式报告,或是研究过程中阶段进展情况的实际记录。科技报告分阶段性报告和终结报告。总的特点是每份报告单独成册,有机构名称,有统一编号,内容新颖专深,叙述详尽,其内容常经主管部门审定,比较成熟可靠。但科技报告中的阶段报告,审查不严格,出版发行又不规则,且具有一定保密性,因而流通受限制。科技报告作为重要的信息来源,在 20 世纪得到很大发展,尤其是第二次世界大战期间各国军事研究大为发展,科技报告相得益彰。由于科技报告在一定程度上反映一个研究机构乃至一个国家的科研能力和水平,所以美、英、法、德等国,每年都发行许多科技报告,其中最著名的是美国二次大战后出版的四大报告:即美国军事国防部门的 AD 报告、美国商务出版局的 PB 报告、美国能源部的 DOE 报告和美国宇航部门的 NASA 报告。这四大报告连续发行了 40 年,量大面广,内容涉及科学技术的许多领域,如:数、理、化、天、地、生、农、医、工程、宇航、军工、能源、交通运输、环境保护等,参考价值较高。

科技报告根据其内容、密级、体裁等特征,可分为以下四种类型:

- (1)按科技报告的内容可分为基础理论研究报告和工程技术报告等。
- (2)按科技报告的形式可分为报告书、技术札记、备忘录、论文、通报、技术译文等。
- (3)按科技报告所反映的研究进度可分为初步报告、进展报告、中间报告、终结报告等。
- (4)按报告的流通范围可分为绝密报告、机密报告、秘密报告、非密限制发行报告、非密报告、解密报告等。

科技报告总的来说,保密的多、公开的少,有较大的情报价值,对科研工作可起到直接的借鉴作用。

4. 科技会议文献

科技会议是科技交流的一条重要渠道。按照会议的规模,科技会议分为基层会议、全国会议和国际性会议。而在这些会议上宣读的论文、报告或交流的书面论文都被称为科技会议文献。它的主要特点是反映新成果较快,质量较高,专业性强,往往代表着某一学科或专业领域的最新学术研究成果,基本上反映该学科或专业当时的学术水平、动态和发展趋势。会议文献历来是重要的科技信息来源之一。

世界上每年举行上万次的学术会议来交流科研成果,讨论重大的学术问题。世界会议信息中心对会议文献与期刊论文进行对比分析后认为,会议文献所反映的信息比期刊平均

早一年,其出版形式灵活,新兴学科的专业文献集中,许多会议文献不在期刊上发表。

5. 专利文献

专利文献是指在建立了专利制度的国家中,发明人向本国或外国专利局提出申请保护某项发明时呈交的一份详细技术说明书,经审查、批准、出版后就成为通常所说的专利说明书。广义地讲专利文献包括专利公报,专利法律及有关文件和专利检索工具书刊。

6. 标准文献

标准文献是将技术标准以文字形式记载在一定载体上的一种技术文献。它是由技术标准及其具有标准性质的类似文件所组成的一种特定形式的技术文献体系。

7. 政府出版物

政府出版物是指各国政府所属各部门发表、出版的文件。分为行政文件(如议会讨论记录、听证记录,政府及各部法令、政策、外交文书,调查统计资料等)和科技文献(如科技报告、标准规格、地质水文航线图等)两大类。其中科技文献占整个政府出版物的 30% ~ 40%。显然,政府出版物和科技报告等其他类型的科技文献有一定重复,所包括的学科和专业很广。政府出版物的特点是具有正式性和权威性。根据其性质,分为公开资料、内部资料、机密资料(秘密、机密、绝密 3 个密级)等 3 种。

政府出版物集中反映了各国政府各部门对有关工作的观点、方针、政策,对了解某一国家的科学技术和经济状况及政策,具有一定的参考价值。加之其售价较低廉,有些是免费供应的,又是官方文献,所以其可靠性强,深受文献信息单位的重视。

目前,很多国家都设有专门机构(如美国政府出版局,英国皇家出版局等)负责办理政府出版物的发行工作。据不完全统计,美、英、法、日等国的政府出版物每年多至几万种,并且还在逐年增加。

世界上公开出版发行、数量最大、内容最丰富的是美国政府出版物(U. S. Government publications),它是联邦政府及其直属机构的出版物。其出版者是美国国会、联邦法院、国防部、内政部、财政部等,以及直属联邦政府各委员会、局、署、院等几百个单位。其学科范围包括法律、经济、贸易、军事、科学、工程技术、农业、医药卫生等。其内容十分庞杂,包括法律文件、法庭审判记录、委员会报告、统计报告、调查报告、听证记录、章程、法规,其中还有一部分 AD、PB 报告、专利、会议文献等。出版形式有小册子、图书、期刊、图表等。美国政府出版局文献管理处是美国政府出版物的发行机构。

8. 学位论文

学位论文是研究单位和高等院校本科毕业生、硕士和博士研究生申请学位时撰写的论文(英国通称为 Thesis,美国称为 Dissertation)。学位论文是经审查的原始研究成果,具有内容专一、阐述详细、有独到见解、参考文献比较系统等特点。

学位论文一般分为两种类型:一种是调查研究性的,著者根据自己或他人的大量的资料或数据,进行科学的分析研究核实,从而对某一课题作出概括性的总结,提出本人的独特见解。另一种是理论的研究和探讨,这种论题一般也需占有大量的资料,根据前人已想出的论点或结论,加上自己的研究,提出自己的进一步的论点和新的探讨。

各级学位论文都是在有专长的导师指导下研究成功后撰写的,并经过答辩和评审,具有一定质量。论文的研究主题,大多数是某学科的新课题,具有一定独创性。论文对问题的探

讨比较专深,特别是博士论文,具有较高的学术水平。此外,论文叙述也比较系统和详细,因此,学位论文是有一定的信息参考价值的。学位论文通常保存在授予该学位的大学图书馆里,当需要索取和复制时,要办理一定的手续。例如要得到学位论文作者的同意等。

由于学位论文公开发表的机会很少,除少数印成单行本或在期刊上、学术会议上发表外,一般都不出版发行,因而份数有限,馆藏分散、搜集困难、没有得到充分利用。

9. 产品样本

产品样本一般是指厂商为了推销其产品而印发的商业宣传品,这些大多是对定型产品的性能、构造原理、用途、使用方法、操作规程、产品规格等所作的具体说明,并有较多的外观照片和结构图。这类文献在技术上比较成熟,数据比较可靠,直观性强,有参考价值。但此类文献的使用寿命,随产品更新周期的缩短而降低。

产品样本的种类有:单项产品样本、产品说明书、产品目录、企业产品一览、产品样本集、产品数据手册、厂商企业介绍、厂刊外贸刊物等等。

产品样本是外贸或引进技术的主要参考材料,是工程和产品设计人员的重要信息源,是使用、维修国外产品的指南。产品样本在这方面所提供的信息是其他文献所无法比拟的。

10. 科技档案

科技档案又称技术档案,它是指科研生产活动中形成的,有具体事物对象的技术文件,是图纸、图表、照片和原始记录等的总称。详细内容包括任务书、技术指标、审批文件、研究计划、方案大纲、技术措施、调查材料、设计资料、试验和工艺记录等。科技档案包括科技活动中准备、试验、总结各阶段形式的全部技术文件,往往成套收藏。它不仅反映科技活动过去的历史,更能如实、准确地反映科技发展的现状。因此,它可能重复实践,或经修改后再实践,并且它还是追究技术责任的依据。它是生产建设与科学技术研究工作中积累经验、吸取教训和提高质量的重要文献,具有较大的利用价值。

科技档案的特点有:

(1)科技档案比较真实、详细、准确、可靠。

(2)大量未能发表的科技文献均存贮于科技档案中。一个单位的科技成果报告,因多方面原因,不可能全部发表,尤其是尖端技术以及对国民经济发展有重大影响的技术项目,也因保密而不能发表,所以能够发表的科技文献只是全部科技文献中的一部分,而其余大量科技文献却贮存于科技档案中。

(3)科技档案具有长久保存价值,是科技贮备最完善、最可靠的形式。

(4)科技档案保密性较强,一般不对外开放,所以查找外单位的科技档案较困难。

科技档案种类繁多,成分复杂,大体上可归纳为:工程设计技术档案,基本建设档案,生产技术档案,设备档案,自然科学研究技术档案,专门性的科技档案等6种。

根据科技档案的归类的特点,一般有:按工程项目分类,按产品型号分类,按科研专题分类,按专业分类,按领域或时间分类5种方法进行归类管理。

科技档案是在科研生产中形成的,是技术和生产活动的真实记录和经验总结。在生产技术活动中,有完备的技术档案,生产技术工作就可以有秩序地进行,没有技术档案,或者虽有技术档案,但管理不完善,科技档案不完整、不准确,生产技术工作就失去了依据,使生产技术工作受到损失。

除上述十大科技信息源外,还有报纸、科技电影、卫星资料、电子地图、广播电视等大众传媒,这些文献信息都具有特殊的参考价值。尤其在现代信息网络环境下,网络上每时每刻都传输着大量的数字化电子化的各类信息,是极其丰富的信息资源。

第四节 知识、信息和文献

一、知识

知识是人类社会实践的总结,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,人类不仅要通过信息感知世界、认识和改造世界,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。

在知识经济社会中,知识和知识产品已成为经济发展的第一推动力,成为社会经济发展的一个先决条件。现在世界竞争的核心归根到底是智力和知识的竞争,谁拥有了知识和信息,谁就能够在竞争中抢占高新技术的制高点和最前沿。因此,知识在经济发展中的互动作用愈来愈明显。

知识从不同的角度一般可以划分为下列四类:

知事(know-what),即关于事实的知识,例如,上海第三产业的就业人数有多少,《中美农业贸易协议》签署的时间等,在日常生活和工作中人们常常会使用到这些知识。

知因(know-why),即关于科学原理和自然科学方面的知识。这类知识是国家技术进步和工商企业发展产品及工艺的基础。科学院、大学和各类研究机构是生产这类知识的源泉。

技能(know-how),即做事的技巧和能力。技能往往是属于单个的企业或个人所拥有的知识,并且仅限于其自身范围内使用而不得向外传播的知识,秘诀和窍门就属于这一类知识。

知人(know-who),即关于谁知道什么和谁知道如何做什么的信息。了解这类知识可以与有关专家建立联系从而有效地利用他们的知识。尤其在知识经济条件下,运用知人这一类知识对于科学管理和组织来说,是非常重要和有效的手段。

二、信息

信息是指那些被人们用来解决特定问题所需要的、经过激活过程活化了的的知识。这里的激活过程,就是指对文献(即知识)进行加工整理,使之有序化、系统化。信息是特定的知识,是知识的一部分。

三、文献

人类的信息、知识的存在形式基本上有三种:①存在于人脑的记忆中,它是属于人们主观精神世界的东西。它只有以一定的形式,通过一定的载体表达时,才能为其他人所感知。②存在于实物中,如古文物、样品、样机、物品等。人们可以通过研究实物而获得某种知识;③用文字、图形、代码、符号、声频、视频等技术手段记录在一定的载体上,例如刻在甲骨上、印在纸上、摄在感光胶片上、录在磁性载体上,这就是文献。

文献记录和反映着社会发展、科学技术的成就及水平,汇集着世世代代的广大人民群众对自然界认识的结晶,记载着无数成功或失败的经验教训,反映着人类的文明程度,是人类进步的重要基础。

文献的形态:

早期形态:古陶文、甲骨文、金文、石鼓文、木板、简册、帛文。

印刷文献:从西汉开始,我国发明了比帛更为优越的文献载体——纸。东汉蔡伦造出了廉价而优质的纸张(定名蔡侯纸)以后,纸质文献便逐步盛行起来。自东晋末年起,纸张彻底取代了竹简、丝帛,成为文献的主要载体,也是至今传播知识信息的主要形式。主要的类型有:图书、报刊、会议文献、学位论文、政府出版物、档案、统计资料等。

缩微文献:指把知识记录在缩微胶卷或平片上形成的文献。它采用感光材料为存贮介质,通过光学摄影技术把文献的体积缩小,固化到感光材料上而产生的一种文献形式。随着激光和全息照像技术的应用,又出现了超级缩微胶片和特级缩微胶片。一张全息胶片可存贮 20 万页文献。由于它具有体积小、容量大、便于保存和转移、成本低等优点,为珍贵文献的存贮和保藏提供了可靠的条件。

电子文献:全称计算机可读型文献。是以磁性材料为存贮介质,以穿孔、打字或光学字符识别装置为记录手段,通过计算机对电子格式的信息进行存取和处理而产生的文献。它将文献信息通过编码和程序设计使文字和图像转换为数学语言和机器语言输入计算机,存贮在磁带、磁盘或磁鼓上,阅读时又由计算机将其转换为文字或图像,显示在终端屏幕上。它们不仅有很高的信息存储密度,还有很高的信息存取速度,并具有电子加工、出版和传递功能。比如,一张普通的光盘 CD-ROM 的信息存储量可达 600 兆,这些电子出版物包括电子图书、电子期刊、电子新闻、各种联机信息库、光盘数据库产品、软盘和磁带等产品,以及电传视讯和电传文本,还包括电子邮件,等等。

电子文献的产生,被认为是人类在知识生产和交流方面的第四次革命。曾有过的三次革命是:几十万年前语言的出现,开始了人类的思想交流;几万年前文字的产生,开始有了文字的记载;近千年前的活字印刷术的发明,使人类的知识可以广泛地传播开来,推动社会文化和科技的发展。电子出版物的诞生开始了人类历史上最快速、高效的知识生产和传播。人们通过计算机阅读、编辑、出版、检索和获取信息,通过网络远程访问计算中心各种类型的数据库资源。电子出版业的迅猛发展必将极大地加速社会信息化的进程。

各类型文献中,印刷型文献是最基本的,电子型文献是发展方向。电子出版物的出现是社会信息化的一个里程碑,它有着广阔的发展前景。它的产生并不意味着对其他信息媒体的完全取代,各种媒体的产生和存在有其特定的环境条件和需要。因此,相互间将在相当长时间内共存,相互补充、渗透,发挥各自的综合的优势,共同促进信息的繁荣与人类的文明。

第五节 信息检索

一、文献检索的概念

广义的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要找