

绪论

1.1 信息与信息检索

1.1.1 信息

1. 信息的概念

随着科学技术的飞速发展,现代社会已进入信息时代。通信技术、计算机技术和声像技术等现代信息技术的发展及相互融合,拓宽了信息的传递和应用范围,使人们在广阔范围内可随时随地获取和交换信息。信息的概念十分广泛,围绕信息而出现的信息资源、信息技术、信息系统、信息产业、信息化社会和社会信息化等相关术语不胜枚举。可见,信息的观点、概念和方法已经被政治、经济、科技、文化、生产等各个领域所接受和应用。

那么,信息(Information)究竟是什么呢?信息最早见于我国唐代诗人李中《暮春怀故人》中“梦断美人沉信息,目穿长路倚楼台”的诗句,这里的信息就是消息的意思,中国古朴的信息通常理解为信号和消息,随着科技的不断发展和人类的不断进步,信息的概念也有所深化。

至今为止,信息还没有一个统一的定义,不同的作者站在不同的角度就有不同的理解和解释。例如,信息论的创始人香农认为:“信息是能够用来消除不确定性的东西。”控制论的创立者维纳(N. Wiener)着眼于信息应用,对信息的定义是:“信息是人们在适应外部世界并且使这种反作用于外部世界的过程中同外部世界进行交换的内容的名称。”我国《辞海》中将“信息”解释为:“一是指音讯;消息。二是指通信系统传输和处理的对象,泛指消息和信号的具体内容和意义,通常须通过处理和分析来提取。”我国国家标准《情报与文献工作词汇基本术语》中记载:“信息是物质存在的一种方式、形态和运动状态,也是事物的一种普遍属性,一般指数据、消息中所包含的意义可以使消息中所描述事件的不定性减少。”

综合各家定义的合理内核,这里将信息定义为:信息是指应用文字、数据或信号等形式通过一定的传递和处理,来表现各种相互联系的客观事物在运动变化中所具有特征性的内容的总称。不同事物具有不同的存在方式和运动规律,从而构成了各种事物的不同的特征,发出不同的信息。

总之,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征;是客观事物经过感知或认识后的再现。通俗地讲,信息就是通过信号带来的消息,信息普遍存在于自然界、社会界和思维界。信息与物质、能量构成了现代社会的三大资源,成为社会进步的强劲推动力。

现在,人类已走进信息社会的时代,信息作为一种战略性资源,已成为发展科技、生产、经济、文化、教育的重要支柱之一。

人们了解了信息的概念和含义,也要了解什么是信息检索(Information Retrieval)。信息检索原称情报检索,萌芽于图书馆的参考咨询工作,20世纪50年代才固定成专用术语。随着信息爆炸(Information Explosion)和以计算机技术为核心的信息技术的迅速发展,信息检索的概念与类型都在发生新的变化。

有关信息检索的定义,目前国内外也有不同的表述,本书将其理解为: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出有关的信息过程,所以它的全称又叫信息的存储与检索(Information Storage and Retrieval),这是广义的信息检索。狭义的信息检索则仅指该过程的后半部分,即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询(Information Search)。

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。

2. 信息的分类

由于信息涉及的范围广、数量大、特征多,所以信息的分类比较复杂。从不同角度划分,信息主要有以下几种类型。

(1) 自然信息和社会信息。依据信息存在的领域及信息的内容,信息可划分为自然信息和社会信息。自然信息是指存在于人们主观信息之外,来源于自然界,在自然界中传递的信息。社会信息则是来源于人类社会,是人类对于社会现象和社会活动的认识、反映,是在人际间传播的信息及思维信息。

(2) 语言信息和非语言信息。依据信息的符号种类及信息的性质,信息可划分为语言信息和非语言信息。语言信息是指由作为信息沟通最基本、最重要的工具——语言来表现的信息,它由语法信息、语义信息和语用信息构成。人们总是以信息为媒介与事物打交道的,通过语法信息获得语义、语用信息,经过分析比较产生决策信息,再根据决策信息使事物按照人们预期的目标发展。非语言信息则是指表情、体态等非语言显现的信息。

(3) 动态信息和静态信息。依据信息的状态,信息可划分为动态信息和静态信息。动态信息是指时效性极强的瞬间多变的信息,如军情、警情、新闻之类的信息。静态信息则是指呈现稳定、相对静止状态的信息,如历史文献、档案资料之类的信息。

(4) 一次信息、二次信息和三次信息。依据信息的加工处理程度,信息可划分为一次信息、二次信息和三次信息。一次信息是指在生产实践、社会实践和科学实验中产生的新成果、新知识和经验性总结等方面的信息;二次信息是指依据客观需要,按照一定的科学方式,将特定分散的一次信息进行加工整理而形成的信息;三次信息则是指依据二次信息提供的线索并选用一次信息的内容进行综合分析形成的信息。

(5) 单媒体信息和多媒体信息。依据信息的媒体和载体,信息可划分为单媒体信息

和多媒体信息。用来表示信息的声音、图像、文字、数据等物质载体,是信息的表示媒体。单媒体信息是指只有一种媒体的信息。多媒体信息则是指运用多种媒体综合、协调地表示的信息。多媒体信息不同于多媒体技术和多媒体计算机系统的概念,而现代化的多媒体手段可以将多媒体信息带入一个崭新的应用领域。

此外,依据信息的社会属性,信息又可划分为政治信息、经济信息、文化信息、科技信息、学术信息、法律信息、军事信息、市场信息等。

3. 信息的特点

信息按其性质,分为自然信息和社会信息;信息按其载体,可分为空气信息、电子信息、光纤信息和文献信息等。无论哪种类型的信息,都有着以下共同的特点。

(1) 普遍性。信息既不是物质,也不是能量,是客观事物普遍性的表征。信息是无处不在、无时不有的普遍社会现象。只要有事物存在,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。因此,信息是普遍存在着的。

(2) 传递性。信息可以通过多种渠道、采用多种方式进行传递,人们将信息从时间或空间上的某一点向其他点移动的过程称为信息传递。信息传递要借助于一定的物质载体,实现信息传递功能的载体又称为信息媒介。没有物质作为载体,信息也就不能传递和交换。

(3) 共享性。信息作为一种资源,不同个体或群体在同一时间或不同时间可以共同享用。这是信息与物质的显著区别。信息交流与实物交流有本质的区别。实物交流,一方有所得,必使另一方有所失。而信息交流不会因一方拥有而使另一方失去拥有的可能,也不会因使用次数的累加而损耗信息的内容。信息可共享的特点,使信息资源能够发挥最大的效用。

(4) 扩散性。物质是普遍联系和相互作用的,作为表征物质相互联系和相互作用的信息,一方面同时或连续向其他物质系统扩散传播,另一方面也接收其他物质系统扩散的信息。而且,这样的相互扩散是无穷无尽的。

(5) 时效性。信息是对事物存在方式和运动状态的反映,如果不能反映事物的最新变化状态,它的效用就会降低,即信息一经生成,其反映的内容越新,它的价值越大;时间延长,价值随之减小,一旦信息的内容被人们了解了,价值就消失了。信息使用价值还取决于使用者的需求及其对信息的理解、认识和利用的能力,如时效性很强的经济信息、交易信息、天气预报、科学信息等。

(6) 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小,或成为“信息垃圾”。

(7) 可储性。信息可以收集、加工、整理、筛选、归纳、综合,形成如文字、图像、视频、声音以及用各种编码形式记录下来的数据,并可以通过记忆和各种载体来载荷,这就是信息的可储性。

以上列举的是信息的一般特征,随着人们对信息科学研究的不断深入,一定会发现其更深刻的属性。不同类型和领域的信息具有不同的特征。

1.1.2 知识

知识(Knowledge)是人们对客观事物的存在和运动规律的认识,是人类在改造客观世界的实践中积累起来的认识和经验的总和。

《辞海》对知识的解释是人类认识的成果或结晶,包括经验知识和理论知识。其初级形态是经验知识,高级形态是系统的科学理论。美国霍华德大学图书馆副馆长何光国教授认为知识是:①经验累积的记录;②事实组织的系统化;③对事实的理解;④一种理解的行为或状态;⑤人的已知和未知。当代科学辞典认为知识是一种特定的人类信息,是整个信息的一部分。

知识提供某种经过思考的判断和某种实验的结果,是信息经过多次反复以及人们的加工整理而序列化后形成的。信息经过人脑的储存、识别、加工、处理及转换等形式而形成知识。人们不仅能通过信息感知世界、认识世界和改造世界,而且能将获得的信息转变成知识,作为认识和改造世界的武器。人类在接收了社会和自然界的大量信息后,通过实践活动和大脑的思维活动,将这些信息结合实践活动进行分析与综合,形成新的认识,这种经过加工、孕育后的信息就成为知识。或者说,知识是同类信息的深化、积累,是优化的信息的总汇和结晶。从外延来看,知识包含在信息中。

知识也具有实践性、规律性、渗透性及继承性等属性。

1.1.3 情报

情报主要是指“信息”“资讯”“消息”。“情报”一词最早产生于军事领域,是战时关于敌情的报告,以后在不同的历史时期有着不同的含义。但是,无论情报的内容与形式如何变化,其共同之处就是情报是指为一定目的收集和传递的有特定效用的知识。它具有知识性、传递性和效用性3个基本属性。

情报是指被传递的知识或事实,是知识的激活,是运用一定的媒体(载体),越过空间和时间传递给特定用户,解决科研、生产中的具体问题所需要的特定知识和信息。

情报的知识性是指情报的本质是知识,即任何学科领域的、以任何形式出现的情报都具有一定的知识内容。人们通过阅读图书、报刊,通过听广播、看电视、看电影,通过参观、考察、调研等活动,随时都在接收、传递和利用大量的感性和理性的知识。这些知识就包含着人们所需要的情报。因此情报是引起人们原有知识结构发生改变的知识,是激活、活化的知识。

情报的传递性是指知识只有经过传递才能成为情报。正如上面所述,情报是激活、活化的知识,这种使知识由静态进入动态而转变成活化知识的过程就包含着知识的传递,这种传递是通过语言文字、电磁波等各种情报载体来实现的。广播电视,计算机信息处理,激光、光缆、通信卫星,以及信息高速公路等先进技术的出现与发展,极大地扩大了情报的传递范围,加快了情报的传递速度。所以,情报就是流动的知识。

情报的效用性是指流动着的知识并不都是情报,只有那些能满足特定需要的流动的知识才可称为情报。例如,电台每天通过广播传递的大量信息是典型的流动的知识,但对于大多数人来说,这些广播内容只是消息,只有少数人利用广播的内容增加了知识或解决

了问题。这部分被发挥了效用的流动的知识才是情报。因此,情报是对特定的人而言的,对需要它的人而言是情报,对不需要的人而言它只是信息与知识。所以,情报是一种动态的有用的知识。

1.1.4 文献

“文献”一词最早见于《论语·八佾》,南宋朱熹的《四书章句集注》认为:“文,典籍也;献,贤也。”所以这时的“文”是指典籍文章,“献”是指古代先贤的见闻、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。

《辞海》将文献定义为:“原指典籍和宿贤,现指具有历史价值的图书文物资料,如历史文献。亦指与某一学科有关的重要图书资料,如医学文献。”中国国家标准《文献著录总则》(GB/T 3792.1—1983)对文献的定义为:“文献是指记录有知识的一切载体。”《文献情报术语国际标准(草案)》(ISO/DIS 5127)指出:“文献是指在存储、检索、利用或者传递记录信息的过程中,可作为一种单元处理的,在载体内、载体上或者依附载体而存储有信息或数据的载体。”综上所述,本书将文献定义为:文献就是用文字、图形、符号或声像、视频等技术手段记录知识信息的物质载体,或曰固化在物质载体上的知识信息。

文献的构成主要有3个要素:①有一定的知识内容;②有用以保存和记录知识的技术手段;③有记录知识的物质载体。这3个要素是文献的内涵。不同的知识内容、不同的记录方式和不同的物质载体,可形成不同类型的文献,这就是文献概念的外延。

文献是记录知识信息的物质形式,也是借以传递知识信息的工具。由于文献的存在,人类的知识才得以保存和传播,人类的科学技术和文化才得到继承和发展。文献不仅包括书刊等印刷型出版物,还包括古代的甲骨文、竹简、帛书等,以及当今的声像出版物、电子出版物和 Internet 上的信息。

1.1.5 信息、知识、文献及情报之间的关系

信息、知识、文献及情报之间的关系可用图 1-1 表示。

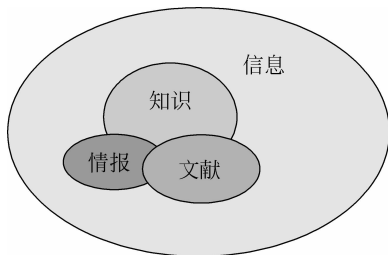


图 1-1 信息、知识、文献及情报之间的关系

简而言之,信息包括知识、文献和情报,知识、文献和情报三者有相交部分但并不重合。文献可以提供信息、知识和情报,但信息、知识和情报获得的方式并不完全是文献。信息是情报和知识的载体,情报是特指的专业信息。知识是信息的内核,是信息中的精华

部分,信息的价值取决于其精华部分的价值。如果说信息是食物,那么知识则是食物中的营养成分。知识依存于信息,信息经过提炼和加工可成为知识;知识组织最终要通过信息组织方式来实现。信息是原料,经过人类的认识活动,成为已知的知识。而被传递、被激活的有用的信息就是情报。

1.2 文献的特点与特征

1.2.1 文献的特点

科学技术不断发展,当今社会已经进入信息时代,这使文献具有了以下特点。

1. 数量急剧增长

随着科学技术和社会生活的飞速发展,人类知识总量在迅速猛增。作为存储、传播知识的载体,文献数量也随着知识量的增加而增加,而且增长速度很快。据统计,目前全世界每年出版各种文献的总量约为 12 000 万册,平均每天出版文献约 32 万册。以我国为例,1949—1954 年,全国共出版图书 21 809 种。2014 年,全国共出版图书 448 431 种,总印数 81.85 亿册(张);出版期刊 9966 种,平均期印数 15 661 万册,总印数 30.95 亿册,定价总金额 249.38 亿元;共出版报纸 1912 种,平均期印数 22 265 万份;出版录音制品 9505 种。

目前,全世界每年出版图书 100 多万种,不到 1 分钟就有 1 本新书问世;期刊 50 多万种;其他文献信息资料 500 多万种;发表科学论文大约 500 万篇,平均每天 14 000 篇左右,每 35 秒就有 1 篇论文发表;每小时出现 20 多项技术发明;每天约有 40 亿个信息单位的信息量向全世界发送,可以说不计其数。表 1-1 所示是 2006—2015 年全国图书总印数。

表 1-1 2006—2015 年全国图书总印数

年 份	总印张/亿册	同比/%	年 份	总印张/亿册	同比/%
2006	64.08	-0.9	2011	77.1	7.52
2007	62.93	-1.79	2012	79.25	2.85
2008	70.26	12.21	2013	83.10	4.87
2009	70.37	-0.36	2014	81.85	-1.51
2010	71.71	1.90	2015	86.62	5.84

文献数量的激增,一方面表明文献信息资源的丰富,但同时也产生了“文献信息污染”,给查找相关文献带来困难。例如,一名化学家每周阅读 40 小时,浏览世界上一年内发表的有关化学方面的论文和著作就要 48 年。因此,就需要人们选择、利用所需的文献,获取所需信息,学会检索和利用相关文献资源。

2. 分布既集中又分散

随着现代科学技术的日益综合与细化,各学科之间的严格界限日趋淡化,学科之间的

相互联系、交叉渗透逐渐增强。这使文献的分布呈现出既集中又分散的现象。1931年,著名文献学家布拉德福(B. C. Bradford)首先揭示了文献集中与分散规律,发现某时期某学科 $1/3$ 的论文刊登在 3.2% 的期刊上;1967年,联合国教科文组织研究了二次文献在期刊上的分布,发现 75% 的文献出现在 10% 的期刊中;1971年,SCI的创始人加菲尔德(E. Garfield)统计了参考文献在期刊上的分布情况,发现 24% 的引文出现在 1.25% 的期刊上;等等。这些研究都表明期刊存在“核心效应”,从而衍生了“核心期刊”(Core Periodicals)的概念。

核心期刊是指那些刊载论文学术质量高、科技信息含量大、刊物学术影响广泛,被认为能反映某种学科(专业)科学研究的前沿动态、学术水平和发展方向的学术期刊。

例如,人们利用英国《科学文摘》统计了“日本光纤通信”方面的论文 322 篇,共发表在 66 种杂志上,有 107 篇文章发表在 3 种杂志上(核心期刊),约有 33% 的文章发表在另 12 种期刊上(相关期刊),还有近 $1/3$ 的文章则发表在另 51 种期刊上(边缘期刊)。文献的这种分布现象是普遍存在的,它提示人们在收集文献信息时,应首先选择本专业的核心期刊,但也不能忽视其他相关期刊。

人们可以使用有关的工具来了解某专业的核心期刊,例如,由北京大学图书馆等单位合编的《中文核心期刊要目总览》(2008 年版)、《国外科学技术核心期刊总览》(2004 年版)、《国外人文社会科学核心期刊总览》(2004 年版)等。

3. 失效加快

目前,随着科学技术的发展,新知识、新理论、新技术、新产品层出不穷,加速了知识的新陈代谢,随之造成文献的新陈代谢,使文献老化加速。据悉,各类文献的平均寿命为图书 10~20 年,科技报告 10 年,学位论文 5~7 年,期刊论文 3~5 年,标准文献 5 年,产品样本 3~5 年。通常用文献的半衰期来描述文献老化的情况。半衰期是指某学科领域目前尚在使用的全部文献中,较新的一半所出版的年限。国外有人统计不同学科文献的半衰期分别为:地理学 16.1 年、地质学 11.8 年、数学 10.5 年、植物学 10 年、化学 8.1 年、生理学 7.2 年、机械工程 5.2 年、化工 4.8 年、物理学 4.6 年、冶金学 3.9 年、生物医学 3 年。

据欧美一些发达国家统计,文献的平均寿命为 5~7 年。这就要求图书界必须加快文献信息的传递速度,才能适应生产技术和科学研究发展的需要。同时,各种类型的文献有不同的有效使用时间。因此,各学科的文獻寿命不同,各国科技发展水平不同,相应的文献寿命也不相同。

4. 内容交叉重复

现代科学技术综合交叉、彼此渗透的特点,导致知识的产生和文献的内容也相互交叉、彼此重复。具体表现为以下几点。

(1) 同一内容的文献以不同文字发表。据统计,当前世界上每年翻译图书约占图书出版总量的 10% ;一些重要的核心期刊被译成多种文字在不同国家出版;同一项发明可以向多个国家申请专利,使专利说明书的内容重复现象严重。据世界知识产权组织统计,世界各国每年公布的专利说明书的重复率达 $65\%\sim 70\%$ 。

(2) 同一内容的文献以不同形式出版。据调查,会议论文有 40% 会后经过整理、修改

发表在期刊上;有相当数量的科技报告后来发表在期刊上,如美国的 NASA 报告有 80%、AD 报告有 60%、美国科学基金会报告有 95%、美国农业部的科技报告有 80%,既出版单行本又发表在期刊上;随着新型载体文献应用的普及,许多文献既出版印刷型的,又有电子版等,如《中国大百科全书》《不列颠百科全书》《工程索引》《科学文摘》《科学引文索引》等既有印刷版,又有电子版。

(3) 许多畅销书内容雷同,选题重复。在激烈的商业竞争中,再版、改版文献数量不断增多,如 2009 年全国共出版图书 30.2 万种。其中,新版图书 16.8 万种,新版图书仅占图书出版总量的 55.6%。

5. 载体及语种增多

随着声、光、电、磁等技术和新材料的广泛应用,新型文献载体不断涌现。传统的纸张型文献已失去了一统天下的局面,多种载体的文献相互依存、相互补充、共同发展已成为发展趋势。新型的非纸张型文献或增大了信息存储密度,延长其保存时间;或加快了信息传递与检索速度,实现了资源共享;或令人闻其声、见其形,使文献家族日臻繁荣。计算机技术、电子技术、远程通信技术、光盘技术、视听技术、网络技术等,构成了信息的现代传播技术。联机检索、交互式图文检索、电子原文传递等现代化信息传播方式已进入实用阶段。信息检索已发展到网络化阶段,人们可以利用分布全球的互联网,多途径、多选择、多层次地检索所需文献。

各国文献所用语种不断增多。据 ISDS(International Serials Data System)1991 年报道,世界上连续出版物使用的语种多达 144 种,常用的有 10 多种,其中英文约占 42%。

1.2.2 文献的特征

文献检索是通过将表示检索要求特征的信息与存储(记录)在文献数据库中的文献特征信息做相符性比较来实现的。检索者的检索要求有两种:①查出具有已知文献外表特征的文献,例如,由作者、书名、ISBN 等途径查找相关文献;②查出具有所需内容特征的文献,例如,根据计算机学科要求编写的一本教材或围绕某一课题收集有关的资料。为了满足检索者这两种检索要求,信息工作者要用特定语言表达分析出的文献的属性或特征,这就是标引。标引人员在编制检索工具(或系统)时,正是按照文献的外表特征和内容特征分别进行标引,以形成满足检索者这两种需求的多种不同的检索途径。

1. 外表特征

外表特征通常包括题目、作者、作者工作单位、专利和科技报告还有专利号或报告号等。这些可以表征一篇特定文献的特征,它可以在文献的封面或扉页,即不打开书本或不看文献的具体内容就可以确定一篇文献的相关信息。根据文献的外表特征,其查准率(Precision Ration)很高。查准率能反映每次从该系统文献库中实际检出的全部文献中有多少是相关的。

也就是说,查准率是指检出的相关文献量与检出文献总量的比率,是衡量信息检索系统检出文献准确度的尺度,即

$$\text{查准率} = \frac{\text{检索出的相关文献量}}{\text{检索出的文献总量}} \times 100\%$$

【例 1-1】 数据库中有 400 000 篇记录,其中有某专题文献 4000 篇相关文献记录,经检索共查出 2000 篇文献记录,查出的文献中只有 1000 篇记录是相关文献,求查准率。

解

$$\text{查准率} = (1000/2000) \times 100\% = 50\%$$

2. 内容特征

假如人们深入文献内容中,则可以发现还可用另外两种方法来表征它。

(1) 一般一篇文献都是论及某一方面的特定问题的,也就是说,与论题相关的词出现的频率较大。以前的研究表明,无论哪一种类型的文献,若对文献中出现的词进行频率统计,会发现所有的词可分为 3 类。

① 文献中出现频率最高的词是冠词、介词和连词等,即其本身没有具体含义的词,如 a、an、the、this、that、or、and、in、on、with 等。

② 绝大部分词在文献中出现的频率较低。

③ 出现频率既不高也不低的词,在文献中约 3~20 个,这些词恰恰是与文献的主题相关度较大的词,称为文献的主题词。主题词即称为文献的主题特征,其与文献的关系如图 1-2 所示。

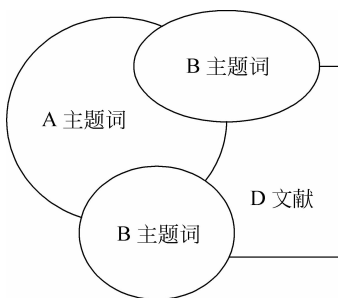


图 1-2 文献主题特征与文献的关系

(2) 一篇文献还可以按照分类方法进行归类,这样,人们可以将某一篇文章归入某一类中。

根据文献的内容特征,其查全率(Recall Ration)很高。查全率是指检出的相关文献量与检索系统中相关文献总量的比率,是衡量信息检索系统检出相关文献能力的尺度。

查全率反映该系统文献库中实有的相关文献量在多大程度上被检索出来,即

$$\text{查全率} = \frac{\text{检出相关文献量}}{\text{文献库内相关文献总量}} \times 100\%$$

【例 1-2】 求例 1-1 的查全率。

解

$$\text{查全率} = (1000/4000) \times 100\% = 25\%$$

文献的外表特征与文献是一一对应的,即一组外表特征只对应一篇唯一的文献,而文献的内容特征与文献却是一种模糊的对应关系,即一篇文献有多个主题词或分类号,一个主题词或分类号也可对应多篇甚至几百篇文献。正因为存在这两种对应关系,才使两种特征的用途大不相同。利用外表特征只能检出很少的文献,有时只用于特定情况下(如已经知道作者名、书名等)。利用内容特征一次能检出一批文献,这对研究者来说是极其有用的。

1.3 文献的类型

文献资源是指人类社会活动中经过加工处理有序化并大量积累后的有用信息的集合,按照按文献加工深度、文献的载体形式和文献的出版形式不同可以分为以下 3 种类型。

1.3.1 按文献的加工深度不同划分

由于文献出版量的急剧增加,再加上文献类型的多样化和分布的离散状态等因素,信息用户要想准确而快捷地从大量无序的原始信息中获取相关信息就如同大海捞针一样困难。为了便于人们利用文献,文献根据加工深度的不同主要分为以下 3 种类型。

1. 一次文献

一次文献(Primary Document)是指作者以本人的研究成果为基本素材而创作或撰写的文献,不管创作时是否参考或引用了他人的著作,也不管该文献以何种物质形式出现,均属一次文献。大部分期刊上发表的文章和在科技会议上发表的论文均属一次文献。一次文献具有如下特点。

(1) 创造性。一次文献是人们根据自己在生产和科学研究中的成果撰写的,报道新成果、新技术、新发明、新创造,是创造性劳动的结晶。例如,专利具有新颖性和创造性,反映了发明创造、技术革新与改进的创造性劳动成果。正由于具有创造性,故一次文献受到人们重视。

(2) 原始性。一次文献是一种原始的创作,一般是作者根据自己所积累的原始素材、原始数据创作而成,故其既有真实可靠的一面,又有特定性和不成熟的一面。

(3) 多样性。一次文献是每个作者的不同成果,故在内容上多样化;另外,在表现形式上也呈现出多样性,有研究报告、学位论文,也有期刊论文、文学读物等。

2. 二次文献

二次文献(Secondary Document)是指文献工作者对一次文献进行加工、提炼和压缩之后所得到的产物,是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。检索工具书和网上搜索引擎是典型的二次文献。二次文献具有如下特点。

(1) 集中性。二次文献集中了某个特定领域的文献,它可以是某个信息部门的所有书刊资料、某个学科领域的文献、某个作者的所有文献等。二次文献是在所集中的某个特定范围的文献基础上,用科学的方法加工整理、组织编排而成的。

(2) 工具性。二次文献可称为工具性的文献,它以特定的方法、简练的语言揭示文献的外部特征和内容特征,并加以科学的编排。它是累积、报道和查找文献线索的一种工具。通常所说的信息检索,即指对于此类文献的有效利用,从中查检到一定的知识信息或某项课题的文献线索。

(3) 系统性。二次文献本身具有自己的系统结构,为了方便利用,一般提供多个检索途径。