



信息检索

Information Retrieval

主 编：马桂琴



兰州大学出版社

LANZHOU UNIVERSITY PRESS

信 息 检 索

Information Retrieval

主 编 马桂琴

副主编 王亚炜 李惠琴

水 静 常雪琴

兰州大学出版社
LANZHOU UNIVERSITY PRESS

图书在版编目(CIP)数据

信息检索 / 马桂琴主编. —兰州: 兰州大学出版社,
2006. 6

ISBN 7-311-02804-3

I. 信... II. 马... III. 情报检索-高等学校-教材 IV. G 252.7

中国版本图书馆 CIP 数据核字(2006)第 067074 号

信 息 检 索

主 编 马桂琴

副主编 王亚炜 李惠琴 水静 常雪琴

兰州大学出版社出版发行

兰州市天水南路 222 号 电话: 8912613 邮编: 730000

E-mail: press@onbook.com.cn

<http://www.onbook.com.cn>

兰州奥林印刷有限责任公司印刷

开本: 787 × 1092 1/16 印张: 20.75

2006 年 6 月第 1 版 2006 年 6 月第 1 次印刷

字数: 500 千字 印数: 1 ~ 1000 册

ISBN 7 - 311 - 02804 - 3/G · 972 定价: 32.00 元

内 容 提 要

面对海量的信息，如何准确、及时、有效地获取自身要求的信息是我们每个人面临的挑战，信息检索对培养我们的信息意识以及提高我们的信息技能都具有极大的帮助。信息检索不仅是信息管理与信息系统专业学生的必修课程，也是一门面向多学科专业的公共课程，同时也是广大信息需求者的必修课程。本书针对这一需求，从信息检索的存储开始，详细讲解信息检索的基本理论知识，并从手工到计算机再到因特网讲解具体的信息查找方法，同时在每章开始说明本章的组织思路，在每章后设计了信息检索思考题。

本书既可作为信息管理与信息系统、图书馆学、情报学的专业教材，也可作为高等学校信息检索类公共课教材以及供其他的信息检索需求者参考。

前 言

这几年来,全国各地有许多版本的信息检索教材,以适应目前信息检索教学以及信息检索需求方面的要求,当我们面对这么多的优秀教材时,开始考虑应该怎样才可以突出自己的特色,且可以比较全面的涵盖信息检索的全貌,所以在本书的撰写过程中,对于本教材体系进行了认真的考虑并结合编者多年的检索课教学,形成了以下的思想:

第一,理论与实践相结合。作为信息检索来讲,在教材编写以及教学过程中基本有两种方法:第一种方法,注重理论与实践;第二种,直接讲解或者介绍检索工具和系统。作为后者,编者认为仅仅是“授之于鱼”,此法虽然效果明显、快捷,但是短效。作为信息检索来讲,应该是一门方法课,它的独特之处在于有强大且内容丰富的理论基础,而且作为信息检索来讲,包括信息的存储和信息的检索两个方面,只有掌握了信息的存储过程,才可以知其然然后知其所以然,而信息的存储涉及到的就是信息检索大量的理论知识。所以笔者认为,即使不是专业的信息检索人员,在掌握信息技能的过程中,不应该急功近利,只有在扎实的理论指导下,面对信息技术日益发展的今天,才不至于只知其一,不知其二。

第二,计算机与手工相结合。信息检索已经发生了很大的变化,从以前的手工发展到光盘、联机到今天的因特网,许多的检索工具不仅具有传统载体的服务形式,更多的是提供网络的服务,而网络检索也因为其快捷和便利赢得了人们广泛的喜爱。但是,手工检索由于其历史悠久,编辑严格,内容系统,提供给我们一种网络暂时无法完成的内容,所以,面对大量的信息检索需求人员,尤其是年轻一代的人员来讲,在检索信息的时候,不能是只想到网络,而应该同时兼顾手工,只有将两者结合起来,才会具有比较好的检索效果。

第三,全面新颖。本教材的统稿思想是涵盖信息检索的两个方面,因此不管是在理论的讲解上,还是在实际的检索上,都力求详实、全面、系统。同时,大量的数据以及实例尽量采用最新的并且是第一手的资料,避免陈旧。

第四,图文并茂。在教材的编写过程中加入了大量的图示以加强解说效果,这样即使有些系统读者无法接触,也可以通过详实的图片了解此系统的概貌。

本书基于以上思想,主要分为三大部分:首先介绍信息检索的基本知识点;然后从信息检索的存储角度讲述相关的理论知识,并力求通过传统信息组织的讲解,引起对网络信息组织所面临的问题的思考;最后讲述信息查找与利用,介绍相关的检索工具、参考工具以及一些重要的检索系统,包括网络检索工具的介绍。

本教材的筹划、内容大纲、编写组织由马桂琴负责。各章节的具体编写分工如

下：水静负责第1章、第5章5~9节；常雪琴负责第2、3章；李惠琴负责第4章、第5章1~4节；王亚炜负责第6、7章；马桂琴负责第8章。

本教材不仅可以作为信息管理与信息系统专业、图书馆学、情报学专业信息检索课程的教材，也可以作为高校信息检索与利用公共课程的教材以及一些信息检索需求者的指南性的工具。

本书在撰写的过程中，得到院领导、家人以及朋友们大力支持和关怀，在此深表谢意！

本书在编写过程中参阅了大量的国内外文献尤其是一些优秀的信息检索教材，不仅是内容，更多的是研究了他们的思想体系，同时也引用了相关的网站以及一些图书馆网站的内容，在此，我们向有关作者和编者表示谢意！

本书虽经编者们的努力，但由于我们水平有限，撰写时间较为仓促，再加上信息检索在资源类型、检索平台方面的变化以及来自日常工作学习的压力，疏漏、不足乃至错误之处在所难免，真诚地期望得到读者和专家的批评指正。

马桂琴
2006年4月

目 录

第一章 信息资源以及信息检索	1
1.1 信息以及信息资源	1
1.2 信息资源	4
1.3 文献信息资源	7
1.4 信息检索以及信息检索原理	11
1.5 信息检索的意义	17
第二章 信息组织	21
2.1 信息组织概念	21
2.2 传统信息组织	22
2.3 网络信息组织	54
2.4 信息检索的过程与方法	65
第三章 信息检索系统评价	68
3.1 信息检索系统的类型及其特点	68
3.2 信息检索系统评价	70
第四章 信息检索技术	80
4.1 信息检索模型	80
4.2 计算机信息检索的技术与方法	85
4.3 信息检索技术的发展	90
第五章 检索工具以及相关网站	94
5.1 检索工具概述	94
5.2 中文三大检索系统介绍	100
5.3 外文重要检索系统介绍	109
5.4 图书、期刊检索工具概述	117
5.5 期刊学术论文检索	143
5.6 专利文献检索	162
5.7 政府出版物检索	168

5.8 标准及标准文献检索	173
5.9 科技报告检索	177
第六章 参考工具以及事实数据的检索	182
6.1 参考工具概述	182
6.2 百科全书	187
6.3 年鉴	198
6.4 名录	204
6.5 辞典	207
6.6 地理资料	213
第七章 光盘以及国际联机检索系统	218
7.1 光盘检索	218
7.2 国际联机检索	226
7.3 世界著名的国际联机检索系统简介	230
7.4 其它国际联机检索系统	237
7.5 我国联机检索系统	239
第八章 因特网信息检索	242
8.1 网络信息资源概述	242
8.2 网络信息检索工具概述	250
8.3 目录搜索引擎	255
8.4 关键词搜索引擎	266
8.5 元搜索引擎	288
8.6 专题信息检索工具	298
附录 1:《中国图书馆分类法》简表	312
附录 2: 中国机读目录格式(CNMARC)	315
参考文献	322

第 1 章 信息资源以及信息检索

[本章思路]

信息检索到底应该解决什么问题? 信息检索将针对什么样的信息资源? 这都是我们学习信息检索的时候首先要深刻认识到论题。在本章中, 我们先从信息入手, 讨论信息资源以及信息检索的相关概念, 以明确我们在这门课程中将要学习的知识内容。

[本章重点]

1. 信息、知识及情报相关概念。
2. 信息检索的定义以及原理。
3. 信息资源的类型以及各种文献信息资源的特点。
4. 信息检索的意义。

1.1 信息以及信息资源

1.1.1 信息、知识、情报

1. 信息 (Information)

信息作为名词术语, 在中国最早出现于唐代诗人李中的诗句中: “梦断美人沉信息, 目穿长路倚楼台。” 但是, 信息作为一个科学概念以及科学研究对象, 却只有四、五十年的历史。然而, 在人类社会的演变和发展过程中, 信息一直在积极地发挥着人类已经意识到或还没有意识到的重要作用。

圣经中记载, 在上古时期, 洪水泛滥, 诺亚自造方舟, 避免了灭顶之灾。方舟在洪水中漂荡许久, 诺亚想知道洪水是否已退, 就放出飞鸽, 待飞鸽衔回一枝橄榄枝, 诺亚根据飞鸽的返回时间及新嫩的绿枝这样的信息推断: 洪水已退, 和平来临了, 人类获救了。这是人类利用信息的最早文字记载。

自古以来, 人们随时都在自觉不自觉地接收、传递、存贮和利用信息。人类的信息活动也从来没有间断过。

20 世纪科学技术的空前进步, 使人们对信息的认识和研究逐步建立在科学的基础上, 形成了以信息为研究对象的理论——信息论, 并迅速发展为一门具有丰富内涵的新学科——信息科学。

1948年,美国科学家申农(C.Shannon)和维纳(N.Weiner)先后发表了《通信的数学理论》和《控制论:关于动物和机器中控制和通信的科学》两篇著作,为信息学的建立奠定了理论基础。

然而,对于信息的定义,人们还没有一致的认识。

申农这样描述信息:“信息是用以消除随机不确定性的东西”,从信息具有减少人们认识的不确定性的功能上概括了信息的特征。

维纳认为,“信息就是信息,不是物质,也不是能量”,把信息看成是与物质、能量具有同等重要性的客观世界三大要素之一。

我国著名的科学家钱学森称信息是“激活了、活化了的的知识”,认为信息“就是为了解决一个特定的问题所需要的知识”,对信息进行了科学的概括。

可见,信息具有极其丰富的涵义,概括起来,信息的概念有三点是需要共同强调的:

●为了一定目的而系统收集起来的、正确的事实所构成的知识,也就是说,经过传递而又为人们所吸收的知识构成信息。

●通过传递而进入人类社会交流体系的运动着的知识构成信息,即一方面信息可以通过一定的物质载体或其它交流形式进行传递,另一方面获取信息必须经过传递。

●信息不仅是传递中的知识,而且必须是有效的知识。凡是人类社会的信息均构成一定的效用,或与接收者原来知识状态吻合而被认可;或有相异,从而发生知识的重构,生成新的知识结构。

因此,可以说,在社会交流过程中,知识变为信息,信息转化为知识,它们总是处在相互作用之中。知识是静态的,以实践为基础,通过抽象思维,对客观事物规律性给予概括。知识信息是人类社会中客观存在着的,以知识形态为主,同时包括数据、新闻、消息等非知识形态的社会信息。知识信息是知识的激活,知识是知识信息的条理化。

其中有一点是可以肯定的,即通常所说的原始数据并不能称之为信息。原始数据和信息的主要区别就在于:一切信息虽然都是由数据组成的,但并非一切数据都能消除不确定性而产生有用的信息。

因此,只有科学地了解和认识信息,才能更好地把握信息,才能使信息更好地为科学技术、经济和社会发展服务。

从客观上看,信息是指一切事物存在方式和运动规律的表征。它是事物的一种普遍属性,只要事物存在,就会有表征其属性的信息。它广泛存在于自然界和人类社会中,信息无处不在。人们正是通过自然界和人类社会中产生的不同信息来区别和认识各种事物的。从主观上看,信息是指人们对世界的认识所形成的各种知识、学问、消息等,它与人类的智能活动密切相关。这里所说的信息主要是指后者,它是人们认识世界和改造世界,取之不尽、用之不竭的宝贵资源。在人类进入信息社会的时代,信息已成为发展科技、经济、文化、教育的重要支柱之一。

信息按表现形式可分为文字信息、图像信息、数值信息、语音信息等类型。各种类型的信息来源于自然界和人类社会中,人们从事各种社会活动如生产、科研、经营、生活等都会产生和运用信息。而对于大多数信息用户而言,文献是其获取信息的

主要来源。

2. 知识 (Knowledge)

知识是人们在社会实践中积累起来的经验,是对客观世界物质形态和运动规律的认识。人们在社会实践中,不断接收客观事物发出的信号,经过人脑的思维加工,逐步认识客观事物的本质,这是一个由表及里、由浅入深、由感性到理性的认识过程。由此可见,知识的产生来源于信息,通过信息传递,并对信息进行加工的结果。

关于知识内涵的阐释,比较经典的有:

《中国大百科全书》:“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象。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。”

国家科技领导小组在1998年《关于知识与国家知识基础设施的研究报告》中的定义:“经过人的思维整理过的信息、数据、形象、意象、价值标准以及社会的其他符号化产物,不仅包括科学技术知识——知识中最重要的部分,还包括人文社会科学的知识、商业活动、日常生活和工作中的经验和知识,人们获取、运用和创造知识的信息,以及面临问题作出的判断和提出解决方法的知识”。

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的集合。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,成为经济发展的一个先决条件。现今世界竞争核心归根结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。

经济合作与发展组织(OECD)1996年在《基于知识的经济》(Knowledge-based Economy)报告中认为知识可以分为四类:

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,中俄《璦琿条约》签署的时间,洗面奶含有哪些养颜的成分等。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。

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,是非常重要的和有效的手段。

3. 情报 (Information)

不管是情报也好,信息也好,其英文单词为同一个,即“information”,实质上是

同一个英文单词在国内有不同译法所致。国内人士一般认为,广义的信息是只要一切可视、可感、可听等借助于各种载体所发出的信号、符号等均为信息;而情报则属狭义的范畴,即情报是被传递的知识。它是针对一定对象的需要传递的,并且是生产实践和科学研究中起继承、借鉴或参考作用的知识。情报是知识的一部分,即被传递的部分,而这种知识又是以前尚未知道的。知识要转化为情报,必须经过传递,并为使用者所接受,才能发挥其使用价值。

情报的特点:

(1) 综合性

不管是何种情报都可以说是经济情报、政治情报、管理情报、科技情报等各种情报交叉、渗透的综合体,其内容反映的是一个综合性的知识领域。

(2) 及时性

由于科学涉及的各种研究主体与整个社会的发展密切相关,各种主体活动又与空间、时间、环境相联系,离开了某一特定的时间、空间和环境,其情报的实用价值就会降低甚至毫无价值。

(3) 系统性

情报所反映的是一个综合的知识内涵,由若干个具有特定内容和相关性质的若干现象,彼此之间都是相互联系、相互作用的,从而构成互相制约的系统。这些关系的综合,就构成了某种内在关联的系统。

(4) 有效性

情报是有用的知识所构成的。情报只有在其被人们接收后,而且其使用价值得到充分体现时本身才有价值。离开了情报的有效性,再多的信息都是废物。

按照不同的应用范围,情报可分为科技情报、社会科学情报、经济情报、政治情报、军事情报等。

4. 信息、知识及情报三者的关系

综上所述,世界是物质的,物质运动产生了信息;各种信息经过人们系统化的加工处理,便转化为知识;知识经过系统化的加工处理转化为情报;情报应用于实践,解决实践中存在的问题,创造出物质财富或精神财富,这时的情报,便转化为生产力,产生新的信息,形成一个无限循环的转化过程。这也表明,信息包含知识,知识包含情报。它们不仅是包含关系,而且可以相互转化。例如,知识在需要被用来解决特定问题时,便转化为情报;情报在不需要利用时,便还原为客观的知识。知识和情报,对于既不能认识又不能理解的人来说,它们都不过是一种信息。

1.2 信息资源

信息是一种重要的资源,已成为人们的共识。中共中央办公厅、国务院办公厅在2004[34]号文件《关于加强信息资源开发利用工作的若干意见》中指出:“信息资源作为生产要素、无形资产和社会财富,与能源、材料资源同等重要,在经济社会资源结构中具有不可替代的地位,已成为经济全球化背景下国际竞争的一个重点。”从国家政策层面上明确了信息资源及其开发利用的战略地位。信息资源涵盖了传统沿用的文献、情报、知识等概念。

信息资源与自然资源、物质资源相比,具有以下几个特点:

1. 具有开发、利用和价值转化性。信息资源可以被重复开发利用,其价值在使用中得到体现,在开发中不断增值。

2. 信息资源的利用具有很强的目标导向。不同的信息在不同的用户中体现不同的价值。

3. 信息资源具有整合性。人们对其检索和利用,不受时间、空间、语言、地域和行业的制约。

4. 信息资源是社会财富,任何人无权全部或永久买下信息的使用权,它是商品,可以被销售、贸易和交换。

5. 具有流动性、可传播性。信息资源借助于各类媒介,比如网络、电视、电话、印刷品、声像、电子信息、数据库等,可以广泛向社会传播。

1.2.1 信息资源的概念

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三点:

(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三部分构成: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。

广义的信息资源概念把计算机及通信基础设施,以至信息技术、人才都包括在内,而狭义的信息资源概念则特指信息的内容。

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载信息资源、实物信息资源、文献信息资源、网络信息资源四种类型。

(1) 体载信息资源

体载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。

(2) 实物信息资源

实物信息资源是指以实物为载体的信息资源。依据实物的人工与天然特性又可将实物信息资源分为以自然物质为载体的天然实物信息资源和以人工实物为载体的人工实物信息资源,例如,产品、样品、样机、模型、雕塑等。

(3) 文献信息资源

指以文字、图形、符号、声频、视频等方式记录在各种载体上的知识和信息资源,是人们目前利用得最多的信息资源。文献信息资源的检索将在本书第五章中详细讲述。

(4) 网络信息资源

网络信息资源是指以网络为纽带链接起来的信息资源和以网络为主要交流、传递、存储的手段与形式的信息资源。网络信息资源通过网络将原本相互独立、分布于世界各地的数据库、信息中心、文献中心等联结在一起,形成一个内容与结构全新的信息整体。

因特网信息资源包罗万象,广泛分布在整个网络之中,没有统一的组织管理机构,也没有统一的目录。但按照其采用的网络传输协议的不同,可将因特网信息资源划分为万维网(World Wide Web,简称WWW或Web)信息资源、Telnet信息资源、FTP信息资源、用户服务组信息资源、Gopher信息资源等。

2. 按信息资源传递的范围划分

信息资源按其传递的范围,可以分为公开信息资源、半公开信息资源和非公开信息资源三种。

(1) 公开信息资源

公开信息资源又称共享信息资源或白色信息资源,指公开发行人、流通和传递的信息资源,其蕴涵的信息人人均可以使用。一般来说,公开信息资源的数量最大,而且能够作为信息商品进入流通领域。

(2) 半公开信息资源

半公开信息资源又称灰色信息资源,指非公开发行人、流通和传递,从常规途径难以获取的信息(内部信息)。这类信息出版量小,发行渠道复杂,流通传递范围有限,不易收集。例如,内部的刊物、技术报告、会议资料等。

(3) 非公开信息资源

非公开信息资源又称黑色信息资源,指人们未破译或未被辨识的,也指处于保密状态的信息。例如,考古发现的古老文字,未解密的政府文件,内部档案,个人日记,私人信件。

3. 按信息资源的加工程度划分

信息资源按加工程度,可以分为一次信息资源、二次信息资源和三次信息资源。在“文献检索”和“情报检索”的历史阶段,曾有一次文献、二次文献、三次文献和一次情报、二次情报、三次情报之称谓。

(1) 零次信息资源

零次信息资源指的是信息的内容尚未经专门机构加工整理就直接作用于人的感觉信息情报,是未经发表和有意识处理的最原始的资料。零次信息主要包括形成一次文献以前的知识信息,即未经记录,未形成文字材料;未公开于社会即未经正式发表的原始的文献,如书信、手稿、记录、笔记等。零次信息一般是通过口头交谈、参观展览、参加报告会等途径获取。

(2) 一次信息资源

一次信息资源也称原始信息,是指人们根据自己的科学实验、生产实践的成果而撰写的文献,一般指公开出版的图书、期刊论文、科技报告、会议论文、学位论文、专利等。无论信息存储于何种载体,只要是原始资料就是一次信息资源。

(3) 二次信息资源

二次信息资源也称检索工具,是指对一次信息加工、整理后形成的各种检索工具,如目录、题录、文摘等,它不对一次信息提供评论,仅提供一次信息的检索线索,也就是说,二次信息是把一次信息从分散、无序状态变成集中、有序化的状态,供人们迅速、准确地查询所需信息

(4) 三次信息资源

三次信息资源是指对一、二次信息按知识门类或专题进行综合加工的产物。如百科全书、年鉴、手册、综述等,主要包括一些参考工具。

1.3 文献信息资源

现代信息资源从载体的角度理解,既包括传统的印刷型的资源,也包括网络在内的基于计算机检索的信息资源,但是对于信息检索来讲,都是文献信息资源的范畴,而文献信息资源也是信息检索的主要对象。因此在本教材中文献信息资源是我们检索的重点,也是讲述的重点。

1.3.1 文献概念

国际标准化组织《文献情报术语国际标准》(ISO/DIS5217)对文献的定义:“文献是在存贮、检索、利用或传递记录信息的过程中,可作为一个单元处理的,在载体内、载体上或依附载体而存贮有信息或数据的载体。”我国颁布的《中华人民共和国国家标准·文献著录总则》(GB3792.1-83)定义:“文献是记录有知识的一切载体。”随着科学技术的发展,知识的记录手段与存贮载体不断推陈出新,使文献概念的外延不断扩大。现代文献囊括了各种信息载体,如传统的书刊、文稿,如今的音像制品、机读

资料、电子出版物等。

由上述可见：文献是由知识、载体和记录三个基本要素所构成。知识是人们在社会实践中积累的经验，是文献的内容；载体是文献的物质实体，如竹简、纸张、胶卷、光盘等；记录是产生文献的手段，如铸刻、书写、印刷、复制、录音、录像等，它是知识、信息与载体的联系方式。

文献的主要作用是存贮、传播、交流知识和信息。几千年来，文献作为信息的载体，一直是人类了解过去、认识现在和预测未来的重要依据，是传播信息的主要手段和途径，所以人们早已习惯于从中获取信息，并将其视为信息的源泉。文献能帮助人们克服时间与空间障碍，记录、贮存和传递人类已有的知识与经验，从而推动人类知识的增加和科技的进步。文献中所记载的知识和信息内容可以供无数人同时使用或先后、异地反复使用，还可以通过复印、转录、缩微、拷贝等手段保持其原来的内容。合理、有效地开发文献资源，能给人类社会带来巨大的社会效益和经济效益。

1.3.2 文献信息资源类型以及各自的特点

表 1-1 文献信息资源类型以及特点

	分 类
依文献载体划分	印刷文献型：以纸张为存储介质，用印刷技术出版。如：书、刊等。 缩微文献型：以感光材料为存储介质，以缩微照像为记录手段。如缩微胶片等。 电子文献型：以磁性材料为存储介质，以键盘输入或光学字符识别输入为记录手段，通过计算机处理产生。如：磁盘、光盘等。 声像文献型：又称视听资料。以磁性材料或感光材料为存储介质，通过特殊的机械装置记录声音信息或图像信息而产生。如：录像带、激光视盘等。
依出版形式划分	图书——知识系统、全面、可靠；期刊——内容新颖、情报信息量大；科技报告；会议文献；专利文献；标准文献；学位论文；政府出版物；产品资料；技术档案；报纸；新闻稿等。

虽然传统的印刷型文献面临着各种新型文献的冲击，但就目前而言，无论是在文献出版或发表的数量上，还是在用户的比例上，印刷型文献依然占据着主导地位。按照不同的出版形式，印刷型文献分为图书、期刊、专利文献、会议文献、学术论文、政府出版物、产品样本、科技报告、标准文献以及报纸等。

(1) 图书 (Book)

图书大多是对已发表的科学技术成果、生产技术知识和经验经过著者的选择、鉴别、核对、组织而成的，论述比较系统，全面可靠，查阅方便（有目次表、索引），但出版周期较长，知识的新颖性不够。图书一般属于三次文献，但有的专著往往包含著者的新观点，或使用新的方法，新的材料，往往具有一次文献的意义。

图书种类较多，包括专著 (Monograph) 和丛书 (Series of monographs)、教科书 (Textbook)、词典 (Dictionary)、手册 (Handbook)、百科全书 (Encyclopedia) 等各种阅读型图书和参考书。从中可以看出，所谓的图书可以分为二种：一种为普通书籍，一

种为工具书。

(2) 期刊 (Periodical)、报纸 (Newspaper)

期刊又称杂志 (Journal, Magazine), 一般是指具有固定题名, 定期或不定期出版的连续出版物。其特点是出版周期短, 报道文献速度快, 内容新颖, 发行及影响面广, 能及时反映科学技术中新成果、新水平、新动向。期刊发表的论文大多数是原始文献, 许多新的成果、新的观点、新的方法往往首先在期刊上刊登。科学技术的研究人员应熟悉本专业有关的期刊。常常阅读期刊可以了解动态, 掌握进展, 开阔思路, 吸收新的成果。期刊论文是文献的主要类型, 是检索工具报道的主要对象。

报纸也是一种连续出版物。对社会科学特别是对广泛的社会研究和企业经营来说, 报纸是非常重要的信息源。

(3) 报告 (Report)

报告是研究人员或企业围绕某一专题从事研究取得成果以后撰写的正式报告, 或者是在研究过程中每一个阶段的进展情况的实际记录。其特点是内容详尽专深。报告的类型有技术报告 (Technical Reports)、札记 (Notes)、论文 (Papers)、备忘录 (Memorandum)、通报 (Bulletin)、可行性报告 (Feasibility Report) 及市场预测报告 (Market Prediction Report) 等等。报告一般单独成册, 有具体的篇名、机构名称和统一的连续编号 (即报告号)。

报告一般划分为: 保密 (Classified)、解密 (Declassified) 及公开 (Unclassified) 等几种密级。保密的报告经过一定时间后往往会转为解密报告。非密资料中, 又分为非密控制发行和非密公开发行。

(4) 会议文献 (Conference Paper)

会议文献是指国际学术会议和各国国内重要学术会议上发表的论文和报告。此类文献一般都要经过学术机构严格的挑选, 代表某学科领域的最新成就, 反映该学科领域的最新水平和发展趋势。所以会议文献是了解国际及各国的科技水平、动态及发展趋势的重要情报来源。

会议的类型很多, 归纳起来可分为国际会议、全国会议、地区性会议三种。会议文献大致可分为会前文献和会后文献两类。会前文献主要指论文预印本 (Preprint) 和论文摘要; 会后文献主要指会议结束后出版的论文汇编——会议录 (Proceedings)。

据统计, 目前世界上每年有上万次学术会议, 发表学术论文数十万篇。会议论文大都有新思想、新观点, 是科学工作者所重视的情报源。

(5) 专利文献 (Patents)

专利是发明人创造发明了某种新技术, 经政府专利局审批后, 即获得一定年限的垄断权。专利文献, 主要是指专利说明书, 分发明专利和实用新型专利两种。这两种说明书是专利申请人向专利局递送的说明其发明创造的文件。在说明书中, 发明人常常论述其发明解决了什么特殊问题, 解决的方法, 对旧有产品的改进及其它用途等; 同时对企业在引进技术和设备, 以及保护企业自身利益的技术有着非常重要的作用。因此, 专利文献已成了情报的一个重要来源。

(6) 学位论文 (Thesis, Dissertation)

学位论文是高等学校、科研机构的学生为获得某种学位, 在进行科学研究后撰写