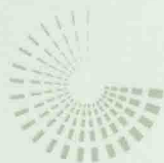


人民日报学术文库

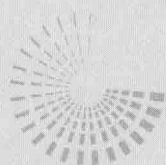


王建行◎著

现代信息检索技术 与应用

XianDai XinXi JianSuo JiShu Yu YingYong

人民日报出版社



王建行◎著

现代信息检索技术与应用

XianDai XinXi JianSuo JiShu Yu YingYong

人民日报出版社

图书在版编目 (CIP) 数据

现代信息检索技术与应用 / 王建行著. —北京:

人民日报出版社, 2014. 12

ISBN 978-7-5115-2980-0

I. ①现… II. ①王… III. ①情报检索

IV. ①G252.7

中国版本图书馆 CIP 数据核字 (2015) 第 011226 号

书 名: 现代信息检索技术与应用

著 者: 王建行

出 版 人: 董 伟

责任编辑: 陈 丹

封面设计: 中联学林

出版发行: 人民日报出版社

社 址: 北京金台西路 2 号

邮政编码: 100733

发行热线: (010) 65369527 65369846 65369509 65369510

邮购热线: (010) 65369530 65363527

编辑热线: (010) 65363105

网 址: www.peopledaily.press.com

经 销: 新华书店

印 刷: 北京天正元印务有限公司

开 本: 710mm × 1000mm 1/16

字 数: 323 千字

印 张: 18

印 次: 2015 年 6 月第 1 版 2015 年 6 月第 1 次印刷

书 号: ISBN 978-7-5115-2980-0

定 价: 54.00 元

前 言

21 世纪在信息时代背景下,信息资源以经成为除物质、能源之外的第三大资源。如何能在浩如烟海的海量信息中快速、准确、有效地检索到所需要信息已成为每一个信息检索者的迫切愿望,也给人们检索和利用信息带来了空前的挑战。满足用户的最大需求不仅要求提高信息检索系统的易用性和信息检索技术,还应充分分析和了解用户的信息检索特点,发现影响用户利用网络信息行为的障碍,有针对性地为用户提供帮助。

在这个高速发展的信息时代,信息就是资源、信息就是财富、信息就是商品、信息就是机会,每个人都渴望及时获得有价值的信息,如果说信息是人类赖以生存、发展的本能,那么信息检索就是每个人必须具备的一项基本技能。作者结合自己多年从事相关课程教学的经验和理论,认为信息检索在信息时代所起作用主要表现为:

1. 信息检索是科学研究的组成部分。科学研究首先是从课题调研掌握资料起步的,信息检索有利于掌握本课题的进展动态,开拓思路,避免无效劳动,把研究水平上升到新的高度。

2. 信息检索是读书治学的基本功。无论是在学习还是在工作期间,都需要进行各种信息检索技术与方法的培训。

3. 信息检索是科学决策的先导。信息化时代的经济管理、政治控制、艺术创造乃至心理状态的演变等,均受到各种社会信息的影响。适时掌握相关信息资源才能实现有效的管理。

目前,信息检索已经发展到网络化和智能化的阶段。信息检索的对象从相对封闭、稳定一致、由独立数据库集中管理的信息内容扩展到开放、动态、更新快、分布广泛、管理松散的网络内容。因此,在未来的时代信息检索必将出现信息智能化、个性化、专业化、多样化的检索引擎。智能搜索引擎是结合了人工智能技术的新一代搜索引擎,它使因特网信息检索从基于关键词检索提高到基于知识或概念

检索,并对知识有一定的理解及处理能力,能够实现分词技术、同义词技术、概念搜索、短语识别及机器翻译等技术。智能检索是基于自然语言的检索形式,机器根据用户所提供的以自然语言表述的检索要求进行分析,而后形成检索策略进行搜索。用户所需要做的仅仅是告诉计算机想做什么,至于怎样实现则无须人工干预,这意味着用户将彻底从烦琐的规则中解脱出来。在检索服务方面,提高检索质量需求最基本的一点就是判定用户是在寻找快速的回应还是精确地检索结果并分析查询中隐含的“意义范围”,即词语在不同领域的含义。

本书对现有的信息检索方法与理论进行了全新的平台构建,通过相关数学模型及算法对信息检索技术的内涵有了详细的分析。全书主要从绪论、信息检索基础理论、应用篇和中、外检索工具介绍四大部分体系全面地对现代信息检索的理论、技术、工具及发展进行了细致性的讲授。

简言之,本书特色主要体现在以下几个方面:

第一,注重信息技术的历史演变,通过纵向的历史分析,阐述了信息检索发展的内在逻辑。有助于读者理解传统信息检索与现代信息检索之间的发展关系。

第二,注重全球化视野,通过横向的比较分析,有利于读者对中、外在信息素养、检索工具等方面知识有一个更全面的认识。

第三,清晰、合理的编排体系。信息检索的编排应遵循从检索的源头到检索的实现以及回归应用的体系构架。从绪论的信息资源、信息素养之后,本书对信息检索的基础理论与技术、应用及检索工具进行了针对性、系统性讲述。

第四,语言简练,表达流畅,逻辑清晰,在理论、技术及实现方面保持一致性。

第五,检索理论、技术及工具符合信息时代的发展趋势。

面对这样一个新技术、新知识不断出现的时代,我们只有不断学习、不断获取新的知识,才能不被社会所淘汰。然而,要快速、有效获得和利用最新的信息,就必须掌握信息检索的相关技能。

虽然作者已经努力使本书尽善尽美,但错误与疏漏之处难免,恳请读者不吝批评指正。感谢人民日报出版社的支持,使本书能够如期完成。

王建行

2015年1月28日

目 录

CONTENTS

第一部分 绪论	1
第一章 信息资源基础	1
第二章 信息素质教育	29
第二部分 信息检索基础理论	38
第三章 信息检索概述	38
第四章 信息检索技术与策略	47
第五章 信息检索语言	54
第六章 信息检索的数学模型	75
第三部分 应用篇	107
第七章 信息检索基础知识	107
第八章 多媒体信息检索技术与方法简介	145
第九章 网络搜索引擎	173
第十章 科技文献检索	186
第四部分 中、外检索工具介绍	214
第十一章 中文检索工具体系	214
第十二章 外文检索工具体系	248
参考文献	280

第一部分 绪论

第一章 信息资源基础

第一节 信息及其相关概念

一、信息的基本概念

信息无时无刻不在,与人类发展历史须臾不可离,从结绳记事到人类登上月球,从日常生活到科学研究,人们都在自觉不自觉地利用信息。信息是普遍存在的,一切信息来源于自然界,来源于人类社会,人们的生产、生活、学习、科研以及社会活动都是信息产生的来源。

信息论创始人美国数学家申农把信息定义为:信息是用来消除不确定性的东西;控制论创始人维纳指出:信息是与外部世界相互作用的过程中所交换的内容的名称;经济学家认为信息是与物质、能源相并列的客观世界的三大要素之一,是为管理和决策提供依据的有效数据;哲学家认为信息是事物本质、特征、运动规律的反映;通信和信息科学家钟义信认为:信息是事物运动的状态与方式,是物质的一般属性;信息资源管理学家和计算机专家认为信息是数据处理的最最终产品,是经过收集、记录、处理、以能检索的形式存储的事实或数据;图书情报学家认为信息是事物或记录,记录包含的信息是读者通过阅读或其他认知方法处理而获取的。

综上所述,我们认为信息是认识主体所感知或所表达的事物运动的状态与方式。

信息的特征主要表现为:

1. 客观真实性。信息是事物存在方式和运动变化的客观反映,客观、真实是

信息最重要的本质特征,是信息的生命所在。

2. 传递性。传递是信息的基本要素和明显特征。信息只有借助于一定的载体(媒介),经过传递才能为人们所感知和接受。没有传递就没有信息,更谈不上信息的效用。

3. 时效性。信息的最大特点是在于它的不确定性,千变万化、稍纵即逝。信息的功能、作用、效益都是随着时间的延续而改变的,这种性能即信息的时效性。时效性是时间与效能的统一性,它既表明信息的时间价值,也表明信息的经济价值。一个信息如果超过了其价值的实用期就会贬值,甚至毫无用处。

4. 有用性(或称目的性)。信息是为人类服务的,它是人类社会的重要资源,人类利用它认识和改造客观世界。

5. 可处理性。这一特征包括多方面内容,如信息的可拓展、可引申、可浓缩等。这一特征使信息得以增值或便于传递、利用。

6. 可共享性。信息与一般物质资源不同,它不属于特定的占有对象,可以为众多的人共同享用。实物转赠之后,就不再属于原主,而信息通过双方交流,两者都有得无失。这一特性通常以信息的多方位传递来实现。

其作用主要体现为:信息和物质、能量成为人类社会资源的三大支柱;信息是主客体的中介,是思维的材料;信息是组织的保证,管理的基础;信息是决策的依据,控制的灵魂。

二、知识的概念

知识的定义在认识论中仍然是一个争论不止的问题。一个经典的定义来自于柏拉图:一条陈述能称得上是知识必须满足三个条件,它一定是被验证过的,正确的,而且被人们相信的^①。知识的概念是哲学认识论领域最为重要的一个概念。个体通过与环境相互作用后获得的信息,在心理学上称之为知识。

Davenport & Prusak 整合了知识的形态、组成元素、主要作用和存储主体等元素,从以下方面描述了知识的定义:(1)知识的形态:知识是一个流动、动态的混合体,随着刺激和学习随时改变、更新;(2)组成元素:包括经验、价值观、情景信息和专业洞察力;(3)主要功能:它提供了一个参考结构来评估和整合新刺激所产生的信息与经验,形成新的结构并可以指导决策和行为;(4)存储主体:它有知者的心智产生并被利用,在组织中知识不仅仅存在于文件与知识库中,更存在于例行的工作、流程、实践与文化中。

^①柏拉图, Theaetetus

世界经合组织(OECD)在1996年的年度报告《以知识为基础的经济》中将知识分为四大类:(1)知道是什么的知识,主要是叙述事实方面的知识;(2)知道为什么的知识,主要是自然原理和规律方面的知识;(3)知道怎么做的知识,主要是指对某些事物的技能和能力;(4)知道是谁的知识(Know - who)涉及谁知道和谁知道如何做某些事的知识。^①

1. 知识分类

按现代认知心理学的理解,知识有广义与狭义之分。广义的知识可以分为两类,即陈述性知识、程序性知识。

(1) 陈述性知识

陈述性知识是描述客观事物的特点及关系的知识,也称为描述性知识。陈述性知识主要包括三种不同水平:符号表征、概念、命题。

符号表征是最简单的陈述性知识。所谓符号表征就指代表一定事物的符号。如学生所学习的英语单词的词形、数学中的数字、物理公式中的符号、化学元素的符号等都是符号表征。

概念是对一类事物本质特征的反映,是较为复杂的陈述性知识。

命题是对事物之间关系陈述,是复杂的陈述性知识。

(2) 程序性知识

程序性知识是一套关于办事的操作步骤的知识,也称操作性知识。这类知识主要用来解决“做什么”和“如何做”的问题,用来进行操作和实践。

策略性知识是一种较为特殊的程序性知识。它是关于认识活动的方法和技巧的知识。如:如何有效记忆,如何明确解决问题的思维方向等。

与哲学不同,认知心理学是从知识的来源、个体知识的产生过程及表征形式等角度对知识进行研究的。如:皮亚杰认为,经验来源于个体与环境的交互作用,这种经验可分为两类:一类是物理经验,它来自外部世界,是个体作用于客体而获得的关于客观事物及其认识;另一类是逻辑——数学经验,它来自主体的动作,是个体理解动作与动作之间相互协调的结果。如儿童通过摆弄物体,获得关于数量守恒的经验,学生通过数学推理获得关于数学原理的认识。皮亚杰对知识的定义是从个体知识的产生过程来表述的。布卢姆在《教育目标分类学》中认为知识是“对具体事物和普遍原理的回忆,对方法和过程的回忆,或者对一种模式、结构或框架的回忆”,这是从知识所包含的内容的角度说的,属于一种现象描述。

知识是不断演进的。从噪音中分拣出来数据,转化为信息,升级为知识,升华

^①经济合作与发展组织(OECD),《以知识为基础的经济》[M],机械工业出版社1997年版。

为智慧。这样一个过程,是信息的管理和分类过程,让信息从庞大无序到分类有序,各取所需。这就是一个知识管理的过程,也是一个让信息价值升华的过程。

反过来,随着信息生产手段的极大丰富,当信息达到一定量的时候,我们发现噪音出现了,信息生产的过程其实也是一个不断衰退的过程,从智慧传播为知识,从知识普及为信息,从信息变为记录的数据,到数据这里已经失去了应用价值,只有记录价值了,从数据衰退到噪音,就已经是纯粹无效的东西了。

2. 知识的相关特征

知识作为一种特殊的信息,它具备了更多的附加特征,也就是说,某一种信息如果越多增加这种特征的烙印,就越接近知识。其特征主要表现为:

(1)隐性特征。知识具备较强的隐蔽性,需要进行归纳、总结、提炼。

(2)行动导向特征。知识能够直接推动人的决策和行为,加速行动过程。

(3)动态特征。知识不断更新和修正。

(4)主观特征。每个人对知识的理解,都会加入自己的主观意愿。

(5)可复制/转移。知识可以被复制和转移,可重复利用。

(6)延展生长特征。知识在应用、交流的过程中,被不断丰富和拓展。

(7)资本特征。知识就是金钱。

(8)倍增特征。知识经过传播不会减少,而会产生倍增效应。一个知识两人分享,就至少有两条。

(9)熟练特征。知识运用越熟练,有效性越高。

(10)情境特征。知识必须在规定的情景下起作用,人类选择知识一般都会进行情境对比。

(11)心智接受特征。知识必须经过人的心智内化,真正理解,才能被准确运用。

(12)结果导向特征。知识不但加速过程,也导向一个可预期的结果。

(13)权力特征。掌握知识的人,即便不在职务高位,也拥有一定的隐性权力。

(14)生命特征。知识是有产生和实效的过程,有生命长短,不是永久有效的。

3. 知识经济

人类进入知识经济的时代意义:知识经济是经济增长直接依赖于知识和信息的生产、传播和使用,以高技术产业为第一产业支柱,以智力资源为首要依托,是可持续发展的经济。按照世界经济合作及发展组织的说法,知识经济就是以现代科学技术为核心的,建立在知识和信息的生产、存储、使用和消费之上的经济。野中郁次郎(Nonaka)认为知识是一种被确认的信念,通过知识持有者和接收者的信念模式和约束来创造、组织和传递,在传递知识的同时也传递着一套文化系统。

知识是从不相关或相关的信息中变化、重构、创造而得到的,其内涵比数据、信息要更广、更深、更丰富。此观点强调知识与背景,以及知识与信息的关系。他在《创造知识的企业》区分两种知识形式:隐性知识(Tacit Knowledge)和显性知识(Explicit Knowledge)。根据这样的划分,前两类知识是显性知识,企业可以通过阅读材料或教材,参加会议和查询数据库获得,对这一类知识可以实现信息化;而后两类知识属于隐性知识,则是更加含蓄的知识,难以量化和信息化,难以通过正式的信息渠道转让。它的特点是:(1)它是一种程序性知识、与行动密切相关;(2)在日常生活中,隐性知识具有有用性,是人们达到价值目标的工具。目标的价值越高,这种知识支持获得目标越直接,这种知识就越有用;(3)隐性知识的获得不能从他人那里获得直接的帮助,它通常靠自己去获得,是不能言传的知识。隐性知识显性化就是知识分享的过程。

三、情报

情报是人们在一定时间内为一定目的而传递的有使用价值的知识或信息^①。

情报的基本属性主要表现为:

(1)知识性。知识是人的主观世界对于客观世界的概括和反映。随着人类社会的发展,每日每时都有新的知识产生,人们通过读书、看报、听广播、看电视、参加会议、参观访问等活动,都可以吸收到有用知识。这些经过传递的有用知识,按广义的说法,就是人们所需要的情报。因此,情报的本质是知识。没有一定的知识内容,就不能成为情报。知识性是情报最主要的属性。

(2)传递性。知识之成为情报,还必须经过传递,知识若不进行传递交流、供人们利用,就不能构成情报。情报的传递性是情报的第二基本属性。

(3)效用性。第三是情报的效用性,人们创造情报、交流传递情报的目的在于充分利用,不断提高效用性。情报的效用性表现为启迪思想、开阔眼界、增进知识、改变人们的知识结构、提高人们的认识能力、帮助人们去认识和改造世界。情报为用户服务,用户需要情报,效用性是衡量情报服务工作好坏的重要标志。

此外,情报还具有社会性、积累性、与载体的不可分割性以及老化等特性。情报属性是情报理论研究的重要课题之一,其研究成果正丰富着情报学的内容。

依据《孙子兵法》,情报等于谍报,可分为“乡间、内间、反间、死间、生间”五种,五种间谍同时都使用起来,使敌人莫敌我用间的规律,这乃是使用间谍神妙莫测的方法,是国君胜敌的法宝。所谓“乡间”,是利用敌国乡人做间谍。所谓“内

^①黄志诚:信息、知识与情报[J],《情报杂志》1996年03期。

间”,是利用敌方官吏做间谍。所谓“反间”,是利用敌方间谍为我所用。所谓“死间”,就是制造假情报,并通过潜入敌营的我方间谍传给敌间(使敌军受骗,一旦真情败露,我方间谍不免被处死)。所谓“生间”,是探知敌人情报后能够生还的人。

按应用范围分类。可分为科学情报、经济情报、技术经济情报、军事情报、政治情报等。

按内容及作用分类。又可分为战略性情报和战术性情报两大类。战略性情报一般是指对解决全局或某一特定领域中(如制定能源政策、城市发展规划等)一些带有方向性、政策性问题所需要的活化了的知识,其中包括科学依据、论证和方案等内容。战略情报的形成需要经过高度的逻辑思维过程并具有较明显的预测性质。战术性情报则是指对解决局部或某一学科领域中的一些具体问题所提供的情报。战略性情报与战术性情报是相互作用、密切关联的,战术性情报是构成战略性情报的基础,战略性情报则可以为战术性情报指明方向。

其研究方法主要表现为:

(1)社会调查法。通过现场调查针对社会现象搜集数据,进行分析。这种方法又可分作直接方法与间接方法两大类,前者主要是用现场观察法,后者又分作访问调查与调查表调查。

(2)文献计量统计方法。对以记录形式进行交流的各个方面进行计量统计,如对于某国、某学科的论文数量、年度增长率、利用状况、对文献中所附参考文献种类、数量等进行计量统计。从中找出变化规律,建立相应数学模型,从定性定量分析中达到掌握变化脉络,进而预测未来可能的变化。

(3)数学分析法。现代数学的许多分支在情报学的研究中都在应用,如在情报检索理论、情报传递的机制、情报采集方案的确定中,概率论、集合论、模糊数学、微分方程、运筹学等均在应用,甚至数论、图论、泛函分析、变分法等,也可以应用。

(4)系统分析与评价方法。对于情报系统各个侧面与总体,可通过引进系统论等方法,进行分析与评价、规划与设计。如情报机构的工作量与处理时间,报道与检索类刊物的时滞,对收藏的文献进行评价,对情报产品进行测试,对情报服务工作进行评价,对索引方式与检索系统进行分析、试验与评价。由于系统论的研究方法众多,对某一具体研究对象来说,就需要比较不同的方法,从中选择最佳方法,以期获得最优的结果。

(5)历史的研究方法。进行历史的研究一般先鉴别一个历史问题,搜集有关史料,形成假说。然后进一步严格搜集与组织史料,认真加以核实,进行分析,得出结论。进行这种研究要力求从经验走向科学,尽管在这一分析中要做到精确是

不容易的。历史研究可以帮助我们了解情报学是如何形成的,促使我们了解过去的事件发生的原因、时间、地点与方式。通过情报学发展过程的剖析,可以加深对于情报学的意义与价值的认识。

情报学的研究方法还很多,诸如德尔斐法、内容分析法、比较分析法、哲学研究法,多综合加以运用,出现空白点分析法、聚类映像法等。

四、文献

文献(literature)是指记录有知识、信息的一切载体。文献的组成要素包括:(1)信息内容—文献的核心。(2)载体材料—承载文献信息符号的物质材料。(3)信息符号—揭示和表达知识和信息的标记符号。记录方式:将包含信息内容的信息符号存储到载体材料上去的方法。^①

1. 文献的类型

按载体类型区分:印刷型、缩微型、声像型、电子型和网络型。

按文献加工层次区分:零次文献。指未经公开发表或交流的文献。如实验记录、草稿、私人日记、笔记、书信、草图等。

一次文献(又称原始文献):是以著者本人的研究或研制成果为依据而创作或撰写的文章,是在科学研究中需重点查找的文献。特点:系统性、创造性、多样性。如期刊论文、专利说明书、会议论文、技术标准、科技报告等。

二次文献(又称检索性文献:如书目、文摘、索引、题录等。)

三次文献(又称参考性文献):如综述、专题述评、学科年度总结、数据手册、百科全书等参考工具书。

零次文献是一次文献的素材,一次文献是二次文献、三次文献最基本的信息源;一次文献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提炼和有序化,二次文献又称为检索工具;三次文献是把零次文献、一次文献、二次文献按照某一特定目的加工而成,是高度浓缩的文献信息。

按版类型区分如下:

(1)图书(一次文献但具有三次文献的性质)特点:内容全、系统,观点成熟,求知解惑之用。周期长,不能及时、迅速地反映最新科研成果。科研利用率14%~19%。检索工具或检索系统:书目、书目数据库。

(2)期刊(最重要的一次文献)期刊:是计划无限期地按期成册出版的连续出版物。ISSN(International Series Standard Number)。特点:刊登多位作者的文章,

^①冯惠玲:《档案文献检索》,高等教育出版社。

速度快,周期短,内容新颖,数量大,品种多。传递科技情报、交流学术思想最基本的手段。科研利用率65%。

检索工具:题录、报刊索引、中国学术期刊数据库等。

(3) 专利文献(集技术、经济、法律于一体的一次文献):广义的专利文献是指一切与专利制度有关的文献,如专利说明书、专利公报、分类表、索引、专利的法律文书等。狭义的专利文献是指专利说明书。专利技术(成果)的特点表现为:独占性、排他性、地域性、时间性。

(4) 标准文献(促进社会产品质量的三次文献):标准文献是按照规定程序编制并经过一个公认的权威机构批准的,供在一定范围内广泛而多次使用,包括一整套在特定活动领域必须执行的规格、定额、规划、要求的技术文件。标准涉及工农业、工程建设、交通运输、对外贸易和文化教育等领域,包括质量、安全、卫生、环境保护、包装储运等多种类型。

标准文献是一种重要的科技出版物。一个国家的标准文献反映着该国的经济政策、技术政策、生产水平、加工工艺水平、标准化水平、自然条件、资源情况等内容,对于全面了解该国的工业发展情况,是一种重要的参考资料。

根据《中华人民共和国标准化法》的规定,我国的标准分为国家标准、行业标准、地方标准和企业标准四级。各级标准的对象、适用范围、内容特性要求和审批权限,由有关法律、法规和规章做出规定。

按约束力分,国家标准、行业标准可分为强制性标准、推荐性标准(/T)和指导性技术文件(/Z)三种;按标准化的对象分,标准可分为技术标准、管理标准和工作标准三大类。

自标准实施之日起,至标准复审重新确认、修订或废止的时间,称为标准的有效期;又称标龄。由于各国情况不同,标准有效期也不同。ISO 标准每5年复审一次,平均标龄为4.92年。我国在国家标准管理办法中规定国家标准实施5年内要进行复审,即国家标准有效期一般为5年。

标准号由颁布机构代号(或说标准代号)、顺序号和颁布年组成。如:GB/T50362-2005《住宅性能评定技术标准》国家标准(GB)行业标准(HB)代号“GB”或“HB”为强制性标准,有“T”的为推荐性标准,有“Z”的为指导性技术文件。需要说明的是,修订的标准期标准号前两项不变,只是颁布年作相应改变。

(5) 产品资料(提供产品技术规格的一次文献)。产品资料是指国内外厂商为推销产品而出版发行的各种商品性宣传品。如公司介绍、产品目录、样本、说明书等。

(6) 会议文献(报道最新科技动向的一次文献):指各类科技会议的资料和出

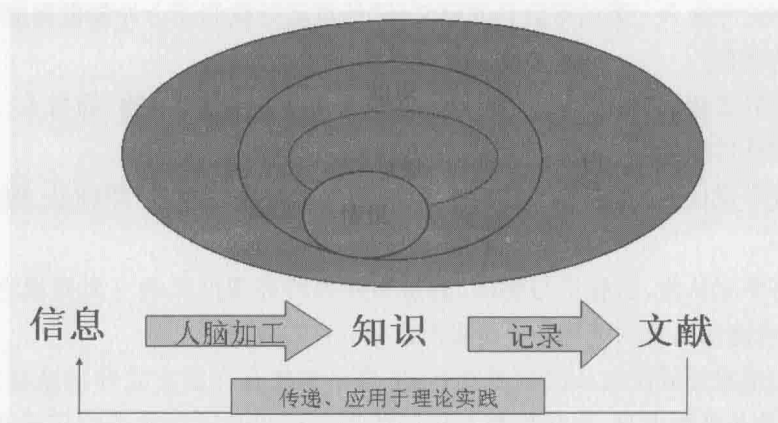
出版物,包括会议前参加会议者预先提交的论文文摘、在会议上宣读或散发的论文、会上讨论的问题、交流的经验 and 情况等经整理编辑加工而成的正式出版物(会议录)。广义的会议文献包括会议论文、会议期间的有关文件、讨论稿、报告、征求意见稿等,而狭义的会议文献仅指会议录上发表的文献。

(7)档案(最忠实的一次文献):社会活动、生产建设和技术工作中所形成的文件的总称,有“第一手材料”、“历史凭证”之美称。由于社会、历史现象不可能重演,因此,对历史上已经形成的文献进行研究、分析、综合已成为社会科学研究的十分重要的手段。

(8)科技报告(既像书又像刊的一次文献):科技报告是关于科研项目或科研活动的正式报告或情况记录,是研究、设计单位或个人以书面形式向提供经费和资助的部门或组织汇报其研究设计或项目进展情况的报告。它以积累、传播和交流为目的,由科技人员按照有关规定和格式撰写,真实而完整地反映科研人员所从事科技活动的内容和经验的特种文献。

五、信息、知识、文献、情报之间的关系

知识是人们对客观事物运动规律的认识,是经过人脑加工处理过的系统化了的信息。知识是人类经验和智慧的总结,知识是信息的内核,知识依存于信息,信息经过提炼和加工可成为知识,知识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。



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的一种载体。文献是记录、积累、传播和继承知识的最有效手段。情报是指传递着有特定效用的知识。因此,情报的三个基本属性是:知识性、传递性和效用性。知识来源于信息,是理性化、优化和系统化了的信息,情报是知识中的一部分,文献是它们的

载体。

第二节 信息资源

信息一词最早出自拉丁语,意思是通知、报道和消息。信息一词也多见于我国古代的诗词中,例如唐诗《暮春怀故人》就有“梦断美人沉信息,目穿长路倚楼台”;宋代女词人李清照《上枢密韩肖胄诗》中也有“不乞隋珠与和璧,只乞乡关新信息”。古人所说的信息,就是指“音信”、“消息”,与今天人们对信息的认识没有本质的不同。

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。信息的广泛应用,导致们对信息的认识和定义上的差别。

哲学家认为:信息是认识世界的一种依据。信息是被其他事物感知的、表征该事物特征的信号内容,即该事物向其他事物传递的信号。

信息论的创始人申农(C. E. Shannon)认为,信息是用来减少随机不确定的东西,即信息是一种不定度。他是从通信系统理论的角度上讲的。

数学家认为信息是一种概念。即信息熵这个概念,它是衡量对某种事物认识的不确定度,即信息的有序度。

控制论专家N·维纳对信息的定义是“信息就是我们适合外部世界进行交换的内容的名称”。

新闻界普遍认为,信息是对事物运动状态的陈述,是物与物、物与人、人与人之间的特征传输。

心理学家认为,信息是存在于意识之外,它存在于自然界、印刷品、硬盘以及空气之中。

经济学家认为,信息是与物质、能量相并列的客观世界的三大要素之一(信息、物质和能源),是为管理和决策提供依据的有效数据。

我们情报领域比较认同的看法是:信息是事物存在的方式和运动状态的表征,信息并非事物本身,而是事物发出的体现它存在和运动状态的信号和消息。因此我们说:信息是客观事物的存在方式与运动规律以及人们对这一切的看法与表述。其功能主要表现在三个方面:一是可深化人们对于世界的科学认识,有助于人们不断地揭示客观世界;二是可以消除人们在认识上的某种不定性。人类认识世界,就是不断地从外界获取信息和加工信息的过程;三是信息可向人们不断

提供知识和情报。

一、信息资源

信息是普遍存在的,但并非所有的信息都是信息资源,信息只有经过人类加工后,可被利用的信息才可称为信息资源。“信息资源”的概念是随着现代信息技术,特别是计算机技术和信息资源管理理论的发展和普及而为人们所接受的。

孟广均等人在《信息资源管理导论》中阐述“信息源不等于信息资源,信息资源是可利用的信息的集合,是高质量、高纯度的信息源。”

卢泰宏和孟广均曾在1992年编译的《信息资源管理专集》中将美国学者对“信息资源”的理解概述为:信息资源=文献信息;信息资源=数据;信息资源=多种媒介和形式的信息(包括文字、图像、声音、印刷品、电子信息、数据库);信息资源=信息活动中各种要素的总称(包括信息、设备、技术和人等)。

由此可以看出信息资源还是一个发展中的概念,是一个具有丰富内涵的术语。我们可以这样理解为:从信息资源所描述的对象来看,信息资源由自然信息资源、机器信息资源、社会信息资源、实物型信息资源组成;从载体和存储方式来看,信息资源由天然型信息资源、智力型信息资源、实物型信息资源和文献型信息资源构成;从信息资源的内容来看,信息资源由政治、法律、军事、经济、管理、科技等信息资源组成;从信息资源的反映面来看,信息资源由宏观信息资源和微观信息资源组成;从信息资源的开发程度来看,信息资源由未开发的信息资源(信息原料)和已开发的信息资源(信息产品)组成。^①

可以说,信息时代更强调信息的收集、整理、加工和使用。信息通过文字符号、声音或图像等形式存储在各种载体上集合在一起就成了信息资源。物资和能源都是不可再生的资源,而信息资源是一种可再生、可反复使用的资源,信息资源是经过人类选取、组织、序化的有用信息的集合。但是同样的信息资源会因使用者的不同而发掘出不同的信息。人们越来越清楚地意识到信息资源日益显著的作用,人类所从事的一切社会活动必须从利用信息资源入手,在未来社会竞争环境中,那些拥有较多信息资源的个人、机构、国家将具有更大的竞争优势。信息资源地位的上升,促进了以计算机和网络为核心的信息技术的进步,先进的信息技术反过来又把信息资源的开发作为重要的应用方面,例如使用现代信息技术可以为人们提供更多、更及时的有用信息。

^① 筭策群、桂学文:《信息经济学通论》1998年版。