

信息检索

XINXIJIANSUO

刘俊熙 王立义/编著



<http://www.altavista.com>

北京图书馆出版社

信 息 检 索

刘俊熙 王立义 编著

北京图书馆出版社

图书在版编目(CIP)数据

信息检索/刘俊熙,王立义编著. —北京:北京图书馆出版社,2002.10

ISBN 7-5013-1953-7

I. 信… II. ①刘…②王… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2002)第 067438 号

书名 信息检索

著者 刘俊熙 王立义 编著

出版 北京图书馆出版社 (100034 北京西城区文津街 7 号)

发行 (010)66126153 传真(010)66174391

E-mail Btsfxb@publicf.nlc.gov.cn

Website www.nlcpress.com

经销 新华书店

印刷 北京华正印刷厂

开本 787×1092 毫米 1/16

印张 16

版次 2002 年 10 月第 1 版 2002 年 10 月第 1 次印刷

字数 300(千字)

书号 ISBN 7-5013-1953-7/G·524

定价 25.00 元

序 言

“知识的一半就是知道在哪里去寻求它。”这是雕刻在美国一所大学的大门上的一句名言。俄罗斯学者布留索夫(Брюсов Влери́й Яковле́вич)在《论目录学对于科学的意义》一文中写道:“有人说,学问与其说是知识的储蓄,倒不如说是善于在书海中找到知识的本领。”

今天,人类社会正处在一个知识经济、科技创新的网络时代。已经突破和将要突破的信息传播和交流技术,使得知识、信息加速更新和快速传递,给我们所处的社会带来了超乎想象的变化。面对浩瀚的、飞速更新的信息,如何有效地开发、科学地存储和组织、合理而充分地利用有价值的最新信息,以提高个人、企业和国家的竞争能力,是现代科研人员必须充分意识和亟待解决的重大和现实的课题。

当今,在网络化环境下,存储技术、计算机技术和通信技术的结合,大大缩短了时间和空间的距离,为存储、组织和利用信息提供了物质的可能。同时也使得信息检索理论和方法发生了质的变化,产生越来越多的新问题需要信息检索领域的专业人员予以重视和解决。

为此,各学科的专业人员都应该在培养信息意识的基础上,掌握建立各类信息数据库和开发计算机检索系统的技术,熟悉获取信息的途径和技能,学会正确设计检索策略及准确分析信息的能力,完善自己的知识结构,服务于各自的工作领域已成为参与社会竞争的重要能力。

囿于作者的学识和水平,以及成书与信息检索技术发展的相对滞后性,书中疏漏、错误在所难免,敬请读者赐教。

刘俊熙

王立义

2002年08月

目 录

第一章 信息概述	(1)
1 信息引论	(1)
2 知识、情报和文献	(6)
3 信息交流	(10)
第二章 信息系统的组织和存储	(18)
1 信息系统概述	(18)
2 信息系统的组织	(20)
3 信息系统的存储	(25)
4 计算机存储概述	(26)
第三章 信息检索语言	(29)
1 检索语言概述	(29)
2 检索语言的原理	(29)
3 检索语言基础知识	(30)
4 检索语言的作用	(32)
5 检索语言的类型	(33)
6 信息网络化对传统规范化语言的冲击	(38)
7 在网络环境下检索语言的发展趋势	(39)
8 关于检索词	(40)
第四章 信息检索原理	(41)
1 信息检索的意义	(41)
2 信息检索的原理	(42)
3 信息检索的类型	(42)
4 信息检索方法	(43)
5 信息检索的具体步骤	(45)
6 检索效果的分析及评估	(51)
7 关于信息检索的几个问题	(53)
8 检索策略的优选	(55)
第五章 综合信息源及检索示例	(58)
1 百科全书及检索示例	(58)
2 类书及检索示例	(60)
3 政书及检索示例	(62)
4 年鉴及检索示例	(64)
5 手册及检索示例	(66)
6 名录及检索示例	(67)
7 表谱和图录及检索示例	(68)

第六章 书目、论文信息系统及检索示例	(71)
1 书目信息系统.....	(71)
2 论文信息系统.....	(73)
3 社科各类信息检索示例.....	(76)
第七章 计算机信息检索	(84)
1 计算机信息检索的发展历史.....	(84)
2 计算机信息检索的必备条件.....	(85)
3 计算机信息检索系统的构成.....	(85)
4 计算机信息检索的类型.....	(86)
5 机检与手检的比较.....	(87)
6 信息处理和信息技术.....	(87)
第八章 数据库、光盘数据库和文档构成	(93)
1 数据库.....	(93)
2 文档结构及检索原理.....	(95)
3 数据库的检索原理.....	(98)
4 数据库分类.....	(98)
5 光盘数据库	(100)
6 《中国学术期刊(光盘版)》数据库	(104)
7 《中国人民大学书报资料中心全文数据库》	(113)
8 《中国人民大学书报资料中心索引光盘》	(116)
9 《中文科技期刊数据库》	(118)
10 《中文社科报刊篇名数据库》.....	(122)
第九章 信息检索的技术实现	(128)
1 基本检索技术	(128)
2 高级检索功能	(131)
3 相关的检索技术	(132)
4 智能检索技术	(134)
5 网络信息检索策略	(135)
6 信息检索技术发展趋势	(139)
第十章 Internet 资源网络信息检索	(142)
1 网络的基础概念	(142)
2 Internet 概况	(143)
3 Internet 基本知识	(145)
4 Internet 主要服务功能	(147)
5 Internet 连接	(149)
6 WWW 信息检索:联机书目检索(OPAC)	(150)
7 WWW 信息检索:搜索引擎	(152)
8 WWW 信息检索:FirstSearch	(167)
9 WWW 信息检索:Dialog	(173)

10 网络信息检索发展展望·····	(180)
第十一章 计算机信息检索示例·····	(183)
1 《中文社科报刊篇名数据库》检索示例 ·····	(183)
2 《中国人民大学书报资料中心数据库》检索示例 ·····	(184)
3 《中国学术期刊数据库》检索示例 ·····	(186)
4 《中文科技期刊数据库》检索示例 ·····	(190)
5 联机书目检索(OPAC)检索示例:清华大学 ·····	(191)
6 联机书目检索(OPAC)检索示例:哈佛大学商学院 ·····	(193)
7 搜索引擎检索示例:Excite ·····	(194)
8 搜索引擎检索示例:Yahoo! 转 Google、HotBot ·····	(196)
9 多元搜索引擎检索示例:ASK、Metacrawler ·····	(199)
第十二章 社科信息研究及成果的表述·····	(203)
1 学术成果的表述形式概论 ·····	(203)
2 学术论文概述 ·····	(203)
3 学术论文的选题 ·····	(204)
4 学术论文的撰写 ·····	(207)
5 学术论文的语言运用 ·····	(213)
6 学术论文的基本格式 ·····	(214)
7 学术论文的讲解和答辩 ·····	(215)
8 学术论文的发表 ·····	(216)
附录 1:国内外分类法演变和发展 ·····	(217)
附录 2:有关词表简介 ·····	(219)
附录 3:First Search 数据库 ·····	(221)
附录 4:DIALOG 系统检索指令格式、功能及举例 ·····	(224)
附录 5:DIALOG 系统综合数据库 ·····	(227)
附录 6:中文社科 Internet 网站信息资源 ·····	(229)
主要参考文献·····	(231)

第一章 信息概述

1 信息引论

人类生活的三个基本要素是材料、能源和信息,由这三个基本要素分别形成材料科学、能源科学和信息科学,被人们确认为现代科学技术的三大支柱。它们的发展大大地推动了人类文明进步的进程。

1.1 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。据说上古时期,诺亚的方舟在洪水中飘荡许久,当放出的飞鸽衔回一束橄榄枝,意味着带回了洪水已退的信息。那么李太白的诗“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。”给我们带来的显然就是庐山瀑布的信息;苏东坡的词“大江东去,浪淘尽、千古风流人物。……”给我们传递的是赤壁怀古的信息。

现代人更是如此,在某种意义上,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。而其中信息处理其实是我们每一个人日常生活中最平常最普通的活动了,并无什么稀奇的地方。正如十九世纪的印度哲学家辩喜所说:“世界上最伟大的东西是最简单的东西,它和你自己存在一样简单”。信息的传递速度可达光速极限,传播范围可及星际空间。自古以来,人们随时都在自觉地接收、传递、存储和利用信息。

然而,在商业活动中、在间谍活动中,人们竟把信息吹捧得如此神乎其神!我们把信息与间谍卫星联系在一起,高速运转,或者与“詹姆斯·邦德(007)”联系在一起。似乎很难把家庭购物和升学择业看作是信息处理的一个组成部分。其实信息与你我如此贴近,犹如日常生活中你每天都会面对的衣服、菜肴和文件一样简单。

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。使我们坐在家里也能知道几分钟之前发生在纽约某个董事会上的事情,或者发生在拉美某个政府部门的事情。

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

在国外,1928年,奈奎斯特(Nyquist)和哈特莱(R. V. Hartley)在一篇题为《信息传输》的论文中,把“信息”理解为选择通信符号的方式,并指出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。

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,他在著名论文《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。信息传递、交流的目的就是要消除信息接受者对于发出信息可能会发送出哪些消息的不确定性。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。在接收者看来,信息必须是事先不知道其内容的新消息。可见,申农的信息定义是从信息在通信过程中而作用角度提出的。

几乎同时,美国科学家维纳(N. Wiener)发表了《时间序列的内插、外推和平滑化》一文和《控制论》一书中表明,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指导衣食住行。商品购买者需要比较不同商品价格和质量的信总;足球教练员需要掌握最新的战术、球员状况、对手的伤病员等信息。每一个组织的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。每一个国家的决策部门同样需要利用内外信息作润滑剂,以使国家机器正常运转。无论是科学技术的研究,还是生产力的发展,都要利用蕴含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。

1.2 信息的特征

所谓信息的特征,就是指信息区别于事物的本质属性。信息的基本特征是:

(1) 客观性

信息的客观性是指信息是客观存在的,不是虚无飘渺可以随意想像和“创造”的事物,其存在也是不以人的意志为条件的。信息是现实世界中各种事物运动与状态的反映,它可以被人所感知、处理、存储、传递和使用。因此,信息的客观性还表现在反映客观世界变化的信息包含在各种物质之中。

(2) 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。事物运动状态及方式的效用是会随时间的推移而改变的。因此获取与利用信息时必须树立时效观念,不能一劳永逸。例如在光盘数据库的检索中,数据的更新往往是以月为单位的,网络数据库的数据更新有时甚至是以周、日为单位的。

(3) 相对性

客观上信息是无限的,但相对于信息用户来说,人们实际获得的信息(实得信息)总是有限的。并且由于不同的信息用户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各不相同,即实得信息量是因人而异的。例如当用户的检索课题为“知识产权”,那么对于信息量庞大的数据库来说,只有符合“知识产权”这一主题的信息才是符合用户需要的。那么这部分的信息量对于整体信息来说是相对有限的。

(4) 依存性

信息本身必须依附于一定的物质形式(如声波、电磁波、纸张、化学材料、磁性材料等等)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手段或物质载体的改变而发生而发生变化。例如关于房产的信息刊登在报刊上、发布在电视节目中,存储在光盘数据库中其信息内容和价值是同样的。

(5) 可传递、亦可干扰

是指任何信息只有从信源出发,经过信息载体的传递才能被信宿接收并进行处理和运用。信息的传递性不仅表现在跨越时空的传递方面,而且还表现在通过语言、文字、图像、电磁波、磁盘等不同的载体形式以及电子计算机、人际交流、文献交流和大众传媒等手段方面。信息跨越时空的传递特性是实现信息资源共享的基础。

一个完整的信息传递过程必须具备信源(信息发出方)、信宿(信息的接收方)、信道(媒介)和信息四个基本要素。其中信道同时对信息传递有干扰和阻碍作用。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引等等干扰,会给信息用户的分析和判断产生或多或少的干扰和阻碍。

(6) 可加工性

信息可以被分析或综合,扩充或浓缩,也就是说人们对信息进行加工处理。所谓信息加工,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。例如文摘可以看作是信息的分析和浓缩的结果。如果在信息加工过程中没有人和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,反之则是不可逆的。实际上信息加工都是不可逆的过程,因为在信息加工过程中信息量的增加和减少是不可避免的。

(7) 共享性

共享性主要是指同一内容的信息可以在同一时间、或不同时间里被多个信息用户使用。一条信息被用户的吸收和利用并不影响数据库内存储和被其他用户反复使用,各用户分享的信息份额也不因为分享人的多少而受影响,信息照样广泛地传播扩散,供全体接收者共享。正如萧伯纳所举的“苹果与思想”的例子,苹果交换以后交换双方仍然各自仅有一个苹果,但是信息交换之后双方都有了两种信息。因此信息的共享性是推动社会交流的主要原动力。

1.3 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、制订方案、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

人类的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。

1.4 信息的类型

综上所述,信息是作用于社会生活每一个领域的。信息活动是人们进行一切社会活动时必然伴随的活动。因为它既纵贯人类社会发展的整个历史过程,又在每个历史时期内横跨当时社会生活的每一个领域。

(1) 按信息的表达形式划分

文字信息:指用文字来表达其内容的信息资料,如各种图书、期刊、胶卷、胶片等;

声像信息:指通过声频信号负载和传递的信息,如通过图片、电影、录音、广播、讲演等。

实物信息:指通过实物来传递的信息,如样品、样机等。

机读信息:指通过计算机阅读的信息,如计算机磁带、磁盘上存储的信息。

(2) 按信息的应用范围划分

白色信息:指公开发行、流通和传递的信息,其蕴涵的信息人人均可使用。包括图书、报刊、缩微胶卷、光盘、数据库等。例如《人民日报》上刊登的信息。

灰色信息:指非公开发行、流通和传递,从正常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。如内部的刊物、技术报告、会议资料等。

黑色信息:指人们未破译或未被辨识的,也指出于保密状态的信息。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、私人信件等。

(3) 按信息加工的层次划分

零次信息:指未经记录、未公开、本身无法通过载体在较大范围内传播的信息。如口头传播的或实物展示的信息、会议口头交流、私人通信、有用的手稿、未发表的演讲稿等。

零次信息是人们获取知识信息不可忽视的信息来源。它不仅在内容上具有一定的价值,而且它有效地弥补了公开信息传播费时的不足,其新颖性更为社会各方面所关注。但由于其传播的范围极为有限、搜集困难,核实验证困难,贮存、保管困难,因而难以把它列为信息检索的对象。获得这方面的信息,很大程度取决于信息用户的信息意识。

一次信息:指以信息编制者的研究成果为依据而创作(撰写)的,未经情报加工的原始信息。习惯上也称原始信息。如图书、期刊论文、科研报告、会议论文、学位论文等。所记录的信息一般比较具体和详尽,有很高的直接参考和借鉴使用价值,是信息检索的主要对象和最终目标。但数量较大,出版分散,缺乏系统性,未经过科学的组织,呈无序状态并难以系统地获得和全面地掌握是它的最大弱点。

二次信息:指对一次信息进行加工整理,使之简化(如著录信息特征,摘录信息内容要点等),或分类编辑之后所得的产物,或为了便于检索和利用一次信息而编辑出版的产物。它以特定的方法汇集某一范围内的信息,用科学的方法加工整理,以简练的语言,不同的深度揭示一次信息的外部特征和内容特征,并提供多检索途径,将分散、无序的大量一次信息转变成有序的、便于管理的系统,从而便于人们有效利用一次信息,因而成为查找一次信息的工具,如各种书目、题录、索引、文摘等检索工具书。我们学习信息的检索方法,主要是学习和掌握二次信息的使用方法。

三次信息:指根据某种需要和目的,利用二次信息作为手段,将大量的一次信息加以全面系统的再度选择、分析和综合,编写成具有专指性内容和使用目的更为明确、效果更为直接的三次信息。例如专题评述、动态综述、百科全书、年鉴、数据手册等等。三次信息的整理加工过程,是一个对一次信息所提供的内容进行综合分析、提炼压缩和研究编写的思维创作过程。另外,三次信息一般附有大量参考信息,也是查找一次信息的重要途径。

从一次、二次到三次信息形式的变化反映了信息的集中和有序化的过程。

将信息划分为以上几种层次和类型,只是一种近似的方法,各次信息的界限有时并不十分明显和严格。我国许多古典目录书,具有信息检索工具的功能,自然应划分为二次信息。但是,它们体现了我国目录学的优良传统,可起到“辨章学术,考镜源流”的作用,则亦可归入一次信息或三次信息。例如,《四库全书总目提要》中的“总叙”及“类叙”,可视为中国数千年的文化简史,这方面的内容就很难说仅属于二次信息的范畴。又如,从总体看,“年鉴”可归入三次信息,但各种年鉴中收列的许多专论则是一次信息,其中的“论文选目”、“新书要目”则是二次信息。因此,在检索和利用报刊信息时,既要在一、二、三次信息功能的角度去充分认识各次信息在科研工作中的作用,但又不可十分拘泥。只有这样,才能准确、全面地利用各种信息。

1.5 信息源

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。

(1) 个人信息源

个人通过语言这一信息交流符号不断地创造和传播各种最新信息,所以参与社会信息交流的每个人都是一个独立的信息源。主要的获取方式是口头交流。其特点是及时、新颖、主观随意和瞬时性。

(2) 实物信息源

一切物质实体蕴涵着的丰富信息均可视为实物信息,它给人们提供了充分认识事物的物质条件。其特点是直观、真实、隐蔽(需要观察和分析)和零散(散乱没有规律,很难加工)的。

(3) 文献信息源

这是指用一定的记录手段将系统化的信息内容存储在各类载体上而形成的一类信息源。在本章下一节 2.3 中有详细阐述。

(4)数据库信息

所谓数据库,就是在一定的计算机软硬件技术支持下,按照一定方式和结构组织起来的,具有最小冗余和较高独立性的大量相关数据的集合。详细内容参见第八章。

(5)组织机构信息源

组织机构是社会的细胞,需要不断地与外界交流信息来发挥其最佳的控制功能。因此组织机构既是社会信息的集散地,也是发布各种专业信息的主要源泉。其特点是权威性、垄断性等。

2 知识、情报和文献

2.1 关于知识

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,并且仅限定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。

2.2 情报

情报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。情报是特定的知识,是知识的一部分。

2.3 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。它只有当以一定的形式,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世界认识的结晶,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。

新版《辞海》对文献的定义作了解释:“记录有知识的一切载体的统称,即用文字、图像、符号、声频、视频等手段以记录人类知识的各种载体(如纸张、胶片、磁带、磁盘、光盘等)。”

根据这一解释我们可以看到“文献”一词涵义已经得到深化:古代甲骨文记录以龟甲、兽骨作为材料是文献;有文字和图形的碑刻、竹简和帛书是文献;有铭文的青铜器是文献;现今的机读资料、电子出版物、缩微制品等等都是文献。尽管文献的载体材料、记录手段在不断演进,但是构成文献的三大要素依然故我,即被记录的知识内容、承载知识内容的载体和记录知识内容的手段。

(1) 文献的形态

早期形态:古陶文、甲骨文、金文、石鼓文、版牍、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,纸张彻底取代了简牍,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。

缩微媒体还可以作为计算机数据的存取载体,分别有输入胶片 CIM(Computer Input Microfilm)和输出胶片 COM(Computer Output Microfilm)。另外,一些公司,如佳能和美能达等已经推出了能方便地检索、复制和传送缩微图像的设备。这种设备能把图像扫描到计算机中,而这些数字化的图像便能被传送到异地的激光打印机、传真机或其他信息工作站上。缩微技术加上电子计算机技术极大增强了缩微媒体的功效。

美国 UMI 公司(University Microfilm International)是世界上最大的一个缩微公司,可提供 2.7 万多种杂志和报纸的缩微胶卷。目前我国已有《四库全书》、《古今图书集成》、《大清实录》等被缩制成胶卷,可资利用。

声像文献:又称视听资料或声像资料。它以感光材料和磁性材料为记录介质,使用特定的设备,用声、光、磁、电等技术将信息表现为声音、图像、影视和动画等形式,给人以直观、形象的感受。它包括唱片、录音带、幻灯片、电影电视片、录像带、激光唱盘、多媒体学习工具等。近十

年来又推出了高密度存贮的唱盘和视盘。这类文献存贮密度高,内容直观真切,在帮助人们观察罕见的自然现象、瞬间的物理化学过程和探索物质结构时能起到文字型文献起不到的独特作用。其表现力强,形象生动,易于理解、接受,传播效果好。这类文献在整个文献中所占的比重正日益增大。

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 600M。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。

近年来,多媒体(multi-media)技术引人注目。多媒体是多种媒体的综合与发展。它是一种数字化的视听媒体,采用超文本(hypertext)或超媒体(hypermedia)方式,除文字外,还包括图片、动画、音乐、语言等信息,在内容表述上具有多样性与直观性,并具有人机交互的友好界面。因此,多媒体电子型的,也是声像型的,它在科技、教育、出版和新闻等领域正在得到日益广泛的应用。

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。

(2) 信息系统中文献类型的识辨

检索系统著录的文献来源款目项,一般对摘录的文献类型不加明显区分,需要专业人员自己辨识。各种类型文献一般都具有一定的著录内容,只要细心观察,就不难识别。下面举例说明主要文献类型的外文著录特征。

1) 图书著录内容

Digital Filters and Their Applications^①. Academic Press^②, London, England^③, 1978^④, 393p^⑤. (0-12-159250-2) ^⑥

注释:① 书名;② 出版社;③ 出版地;④ 出版年;⑤ 图书总页码;⑥ 国际标准书号(0 表示英语地区;12 表示出版社代码;159250 表示该书书号;2 表示该书计算机检验号)。

2) 期刊著录内容

Appl. Solar Energy^①, V. 15^②, No. 6^③, 1979^④, P. 34-35^⑤, ISSN 3645-8102^⑥

注释:① 刊名(大都采用缩写方式);② 卷号;③ 期号;④ 年份;⑤ 该刊的页码;⑥ 国际标准连续出版

物号。

3) 会议文献著录内容

Proceedings of the Society of Photo-Optical Instrumentation Engineers^①, V. 169^②, 1979^③, P. 42-47^④

注释:①会议名称(常用的会议特征有 Conference, Congress, Symposium, Convention, Workshop 等,常用的主办会议的机构特征有 Society, Association 等);②会议文献的卷期;③会议年代;④会议文献的页码。

另外,会议文献的识辨特征有:如会前出版物的 Paper,会后出版物的 Proceedings 等。

4) 学位论文著录内容

ENGINEERING, ELECTRONICS AND ELECTRICAL^①

Neural network-based detection and tracking of maneuvering targets in clutter for radar applications^②. Amoozegar, Seyed Fared^③, Ph. D. ^④ The University of Arizona^⑤, 1994^⑥, 276pp^⑦. Director: Malur K. Sundareshan^⑧

Order Number DA9502624^⑨

Until the recent……^⑩

注释:①类目名称;②论文题名;③著者姓名;④学位级别;⑤授予学位单位;⑥授予学位时间;⑦原文页码;⑧指导教师;⑨UMI 原文订购号;⑩文摘。

另外,学位论文的识辨特征有:①篇名一般均著录有学位和学位论文的名称,如 Ph. D. Dissertation 即哲学博士学位论文,若硕士论文则为 Master Thesis。

2.4 信息、知识、情报和文献的关系

综上所述,世界是物质的,物质的运动便产生了信息;各种信息经过人们进行系统化的加工处理,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。其间的关系如图 1-1 所示:

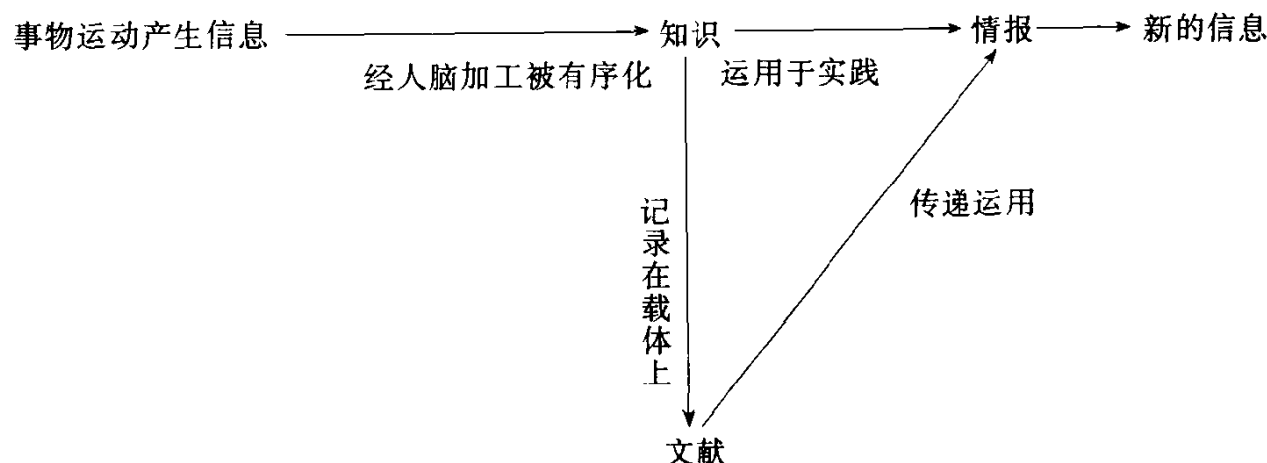


图 1-1 信息、知识、文献和情报之间的关系

3 信息交流

3.1 信息交流

信息交流是人类社会中亘古及今、渗透人类一切社会活动的现象。无论是在人类以渔猎为生、栖身山洞的远古时代,还是在信息产业迅速发展的今天,尽管信息交流的方式和手段不断地随着社会的发展和科技的提高而变化着,但人类从未停止过信息交流的活动。正是由于这些创造性的信息交流活动,人类文明得以形成并代代相传,文明社会的历史得以不断延伸和发展。同时,信息交流也是联结关系纵横的社会网络中各种关系的纽带,是社会各部门协调的机制,是凝聚社会的一股无形力量。人与人之间必须通过相互间的信息交流,思想感情交流,达到相互了解的境界,才能实现有效合作。换言之,人们必须进行信息交流活动,并依赖于这种信息交流维持群体间的各种社会关系,同时维持社会群体的存在与发展。

人类的行为千差万别,其行为的背后都隐藏着不同的信息运动规律,并受信息流的控制和引导。信息流贯穿人类一切行为的始终,广义地说,人类的行为都是信息行为的一种表现形式。

信息交流就是指各个个体借助于共同的符号系统所进行的信息传递、交换与分享,其本质是一种知识交流活动。而知识交流的构成表现在个人知识与社会知识的相互转换运动过程中,它又具有社会性,使信息交流成为个人发展和社会进步的必要条件。

信息交流的社会功能:促进了人类作为个人与社会群体间的沟通、维持了社会系统的动态平衡、建立起共同的社会规范与社会交流机制。信息交流还对社会变化产生影响,成为社会变化的重要因素,是实现社会控制的一个重要的自我组织系统。

信息交流过程涉及四个基本要素:信息的发出者(信源)、信息的接收者(信宿)、交流手段(媒介)和交流对象(信息)。

信息社会的最基本特征是信息交流的全球化。它改变了信息交流的时间、距离、空间,改变了传统信息交流方式和信息流动的速度。越境信息流(Transborder Data Flow,简称 TDF)是指点对点、点对面的跨越国家政治疆界的数字化电子数据和信息的传输。在信息社会中,跨越国家政治疆界流通的数据、知识、信息均属于 TDF 的范畴,其本质就是信息在世界范围内的传递和交流。

3.2 信息交流模式

(1) 拉斯韦尔模式

1948 年美国政治学家拉斯韦尔(Harold D. Lasswell)提出一个著名的命题:“描述传播行为的一个方便的方法,是回答下列五个问题:谁?说了什么?通过什么渠道?对谁?产生什么效果?”此后,这句话被称为“拉斯韦尔公式”而被人们所广为引用,成为传播学研究的基本框架。该模式的缺陷是,它只是单向流动的线性模式,过高地估计了传播的效果,忽略了反馈的作用,而反馈在信息交流中的作用是很显著的。

(2) 申农—韦弗模式

1949 年,美国贝尔电话实验室的申农(Shannon)及其合作者韦弗(Weaver)提出了通信系