

信息检索

XINXI JIANSUO

冯涛◎编著



知识产权出版社

全国百佳图书出版单位

信息检索

冯 涛 编 著



知识产权出版社

全国百佳图书出版单位

图书在版编目(CIP)数据

信息检索/冯涛编著. —北京:知识产权出版社,2015.7

ISBN 978-7-5130-3596-5

I. ①信… II. ①冯… III. ①情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2015)第 145361 号

责任编辑:王 辉

责任出版:孙婷婷

信息检索

冯涛 编著

出版发行:知识产权出版社有限责任公司

电 话:010-82004826

社 址:北京市海淀区马甸南村1号

责编电话:010-82000860 转 8381

发行电话:010-82000860 转 8101/8029

印 刷:北京中献拓方科技发展有限公司

开 本:720 mm × 1000 mm 1/16

版 次:2015 年 7 月第 1 版

字 数:250 千字

ISBN 978-7-5130-3596-5

网 址:<http://www.ipph.cn>

<http://www.laichushu.com>

邮 编:100088

责编邮箱:waughui@cnipr.com

发行传真:010-82000893/82003279

经 销:各大网上书店、新华书店及相关专业书店

印 张:12

印 次:2015 年 7 月第 1 次印刷

定 价:38.00 元

出版版权专有 侵权必究

如有印装质量问题,本社负责调换。

目 录

第一章 信息资源概述	1
第一节 信息引论	1
一、信息的概念	1
二、信息的表现形式和特征	3
三、信息的分类	4
四、信息源	7
五、信息的功能	8
六、数据、信息、知识、情报、文献	9
第二节 信息组织	12
一、信息组织概念	12
二、信息组织内容	13
三、信息组织原则	13
四、信息组织要求	14
五、信息组织目的	15
六、信息组织的基本方法	15
第三节 信息素养	18
一、信息环境	18
二、信息素养概念	19
三、信息素养内涵	20
四、信息素养教育	22
第二章 信息检索	28
第一节 信息检索概述	28
一、信息检索的起源与发展	28
二、信息检索概念	29

第二节 信息检索的意义与作用	31
一、继承和借鉴前人的成果,避免重复劳动	32
二、节省查找文献的宝贵时间,提高科研效率	32
三、获取新知识的捷径	33
四、决策的前提	33
第三节 信息检索原理	34
第四节 信息检索语言	35
一、信息检索语言概念	35
二、信息检索语言种类	36
第五节 信息检索方法	43
一、直接法	43
二、追溯法	43
三、综合法	44
第六节 信息检索技术	44
一、布尔逻辑检索	44
二、截词检索	46
三、位置检索	48
四、字段限定检索	49
五、加权检索	50
六、模糊检索	50
第七节 信息检索步骤	51
一、分析研究课题	51
二、选择检索工具(系统)	52
三、确定检索途径	52
四、确定检索方法、调整检索策略	53
五、获取检索结果并整理	54
第八节 检索效果的评价	55
一、查全率和查准率	56
二、影响查全率和查准率的主要因素	56
三、查全率和查准率的关系	57
四、查全率和查准率的合理运用	58

第三章 书目信息检索	59
第一节 书目相关知识	59
一、书目的概念	59
二、书目的类型	59
三、书目的内容	60
四、OPAC	65
五、书目的作用	68
第二节 国家图书馆书目检索	69
一、国家图书馆简介	69
二、国家图书馆馆藏目录检索系统	70
第三节 CALIS 联合目录检索	78
一、CALIS 简介	78
二、CALIS 联合目录检索系统	79
第四节 贵州师范学院图书馆书目检索系统	84
一、贵州师范学院图书馆简介	84
二、贵州师范学院图书馆书目检索系统	84
第五节 馆藏排架检索	93
一、索书号	93
二、藏书排架	94
第四章 数字资源库检索	96
第一节 中国知网检索	96
一、中国知网简介	96
二、中国知网检索介绍	98
第二节 维普网检索	117
一、维普简介	117
二、维普网检索介绍	120
第三节 读秀学术搜索	149
一、读秀学术搜索简介	149
二、读秀学术搜索检索介绍	150

第四节 EBSCOhost 全文数据库检索	158
一、EBSCO 简介	158
二、EBSCO 检索介绍	159
附录一 中图法简表(第五版)	166
附录二 中华人民共和国国家标准	171
主要参考文献	183

第一章 信息资源概述

第一节 信息引论

一、信息的概念

自 20 世纪出现信息的概念后,信息不仅得到了更加深入的研究,还得到了广泛的重视和利用。在 21 世纪,不掌握信息手段已完全与现代化社会格格不入。人们生活中随时都在自觉或不自觉地接收、传递、存储和利用信息。每个人都需要相关信息来指导衣食住行。购买衣服时,我们会比较不同款式的衣服,对比款式信息和价格信息。选择食品时,会了解生产日期、重量、材质、价格等信息。购买住房或旅游住宿时,我们会详细了解面积、价格、地段、配套、交通等相关信息。出行时,会了解天气信息,根据相关信息来确定自己衣服的厚薄,是否携带伞具。与对手谈判,要全面了解对方的背景信息。单位发工资,要综合员工的工龄、学历、职务、考勤、业绩等众多信息,国家出台新政,也必须在广泛调研并征求民众代表的信息基础上进行。因此小到个人,大到国家,信息无处不在,无时不有,无人不用。信息已经成为个人、部门、国家发展的最重要的战略资源之一。信息作为一种隐形的资源,在社会经济发展、日常生活、科学研究、工作学习、国策法规中发挥着日趋重要的作用。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。

信息是有价值的,就像不能没有空气和水一样,人类也离不开信息。因此人们常说,物质、能量和信息是构成世界的三大要素。所以说,信息的传播是极其重要与有效的。在我国,“信息”一词有着很悠久的历史,早在两千多年前的西汉,即有“信”字的出现,常可作“消息”来理解。“信息”连在一

起,最早见于文献《三国志》中的“正数欲来,信息甚大”。再以后,信息作为一个成词,出现更加频繁,其含义均为“音信、消息”。^①

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·香农(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”,即信息是“用来消除未来的某种不定性的东西”。该论文成为信息论诞生的标志。^②控制论的创始人之一维纳认为,“信息是人们在适应外部世界并且使之反作用于世界的过程中,同世界进行交换内容的名称”。有观点认为,信息以物质介质为载体,传递和反映世界各种事物存在方式运动作态的表征。也有人认为,信息是一种被加工为特定形式的数据。还有如“信息是事物相互作用的表现形式”“信息是人与外界相互作用过程中所交换的内容的名称”“信息是使概率分布发生变动的东西”等定义。

由于人们研究信息的角度与目的的差异,对其理解也是见仁见智,说法不一。信息的概念也是十分广泛的,至今还没有对其形成一个完全统一的认识。

数学家认为,信息是概率论的发展。

物理学家认为,信息是以熵作为度量的内容。

哲学家从产生信息的客体来定义,认为信息是事物本质、特征和运动规律的反映。

心理学家认为,信息是存在于人们意识之外的东西,它存在于自然界、印刷品、硬盘,以及空气之中。

经济学家认为,信息是与物质和能量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。

图书情报学家认为,信息是读者通过阅读或其他认知方法处理而获得的事物或记录。

计算机专家认为,信息是经过收集、记录、处理,以能够检索的形式存储的事实或数据。

综合各个学科对信息内涵的深层理解,《中国大百科全书》中对信息的

① 孙福强. 网络信息检索[M]. 北京:北京理工大学出版社,2014:1.

② 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:1-2.

定义为:狭义上讲,信息是指用来消除不确定性的东西。广义上讲,它包含本体论和认识论两个层次。从本体论意义上说,信息泛指一切事物(物质的、精神的)运动的状态和运动的方式,包括事物内部结构的状态和方式,以及外部联系的状态和方式;从认识论意义上说,信息是关于事物运动状态和运动方式的反映。^①

信息的定义之所以呈现多样化,主要原因有三:①信息本身的复杂性。它是一个多元化、多层次、多功能的综合物。②信息科学是一门新兴学科。它的许多分支学科仍在随着社会、经济和科学技术的发展而发展,其内涵和外延还不很确切。③人们出于不同的研究和使用目的,从不同的角度或层次出发,对信息概念就会做出不同的解释。^②

目前,关于信息比较统一和科学的定义是: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的特征性内容的总称。^③人通过获得、识别自然界和社会的不同信息来区别不同事物,得以认识和改造世界。在一切通信和控制系统中,信息是一种普遍联系的形式。它是事物存在的方式、形态和运动规律的表征,是事物具有的一种普遍属性,它与事物同在,存在于整个自然界和人类社会。

二、信息的表现形式和特征

(1) 表现形式

信息一般有4种形态:数据、文本、声音、图像。这4种形态可以相互转化。信息是通过载体表现出来的,我们听到的声音包含信息,看到的景象包含信息,读到的文字包含信息。信息需要通过声音、图像或者文字等表现出来,所以说信息离不开载体。而纸张、胶片、磁带、硬盘、手机、电脑甚至人的大脑等,都是承载信息的媒介。

(2) 特征

信息本身看不见,摸不着,无色、无味、没有形状、没有大小、没有质量

① 孙福强. 网络信息检索[M]. 北京:北京理工大学出版社,2014:2.

② 肖亚明,尹志清,王涛. 信息检索与利用[M]. 天津:天津大学出版社,2009:8.

③ 肖亚明,尹志清,王涛. 信息检索与利用[M]. 天津:天津大学出版社,2009:8.

……它是非常抽象的东西,但它又处处存在,护之塞耳,视之濡目。它既区别于物质和能量,又与物质和能量有相互依赖的关系。^① 综合起来,信息有以下主要特征。

可量度。信息可采用某种度量单位进行度量,并进行信息编码。如现代计算机使用的二进制。

可识别。信息可采用直观识别、比较识别和间接识别等多种方式来把握。

可转换。信息可以从一种形态转换为另一种形态。如自然信息可转换为语言、文字和图像等形态,也可转换为电磁波信号和计算机代码。

可存储。信息可以存储。大脑就是一个天然信息存储器。人类发明的文字、摄影、录音、录像,以及计算机存储器等都可以进行信息存储。

可处理。人脑就是最佳的信息处理器。人脑的思维功能可以进行决策、设计、研究、写作、改进、发明、创造等多种信息处理活动。计算机也具有信息处理功能。

可传递。信息的传递是与物质和能量的传递同时进行的。语言、表情、动作、报刊、书籍、广播、电视、电话等是人类常用的信息传递方式。

可再生。信息经过处理后,可以以其他形式再生。如自然信息经过人工处理后,可用语言或图形等方式在生成信息。输入计算机的各种数据文字等信息,可用显示、打印、绘图等方式在生成信息。

可压缩。信息可以进行压缩,可以用不同信息量来描述同一事物。人们常常用尽可能少的信息量描述一件事物的主要特征。

可利用。信息具有一定的实效性和可利用性。

可共享。信息具有扩散性,因此可共享。

三、信息的分类

以加工的先后或加工深度划分如下:

文献信息资源以信息加工深度划分,可分为零次文献信息、一次文献信息、二次文献信息、三次文献信息和高次文献信息。

^① 马林山. 信息检索与利用[M]. 合肥:安徽科学技术出版社,2013:4.

(1) 零次信息资源

零次信息资源是指未经记录、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、网络聊天、发言稿、会议记录、实验记录、设计草稿等。具有信息内容新颖、不成熟、不定型的特点,而且没有公开交流,难以获得。

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。^①

(2) 一次信息资源

一次信息资源是指未经加工的原始信息。一般是以作者本人的研究工作或研制成果为依据而撰写,且已公开发行进入社会流通使用的专著、期刊论文、学位论文、专利说明书、科技报告等,习惯上也称原始信息。一次信息资源包含了新观点、新发明、新技术、新成果,提供了新的知识信息,是创造性劳动的结晶;所记录的信息一般比较具体和详尽,具有创造性的特点,有直接参考、借鉴和使用的价值,是人们检索和利用的主要对象。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。如图书、期刊论文、科研报告、会议论文、学位论文等。

(3) 二次信息资源

二次信息资源是对一次信息进行整理、分类编辑或加工处理后得到的信息。即把大量的、分散的、无序的一次信息资源收集起来,按照一定的方法进行整理、加工,使之系统化而形成的各种目录、题录、索引和文摘等检索工具书。

二次信息资源仅是对一次信息资源进行系统化的压缩,无新的知识信息产生,具有汇集性、检索性的特点,易于存储、检索、传递和使用。它的重要性在于提供了一次信息的线索,是打开一次信息知识库的钥匙,可节省人们查找知识信息的时间,有较高的使用价值。

当今网络信息成为信息分析者的重要信息来源。如新浪、搜狐、雅虎中

^① 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:4.

国等很多大型网站都提供大量信息及索引服务,很多信息都是经过处理与加工排序后的信息,对相同主题提供了相关的链接。

(4) 三次信息资源

三次文献信息资源是根据一定的目的和需求,在大量利用一、二次信息资源的基础上,对有关知识信息进行组织、分析、提炼、重组而生成的再生信息资源。如专题报告、技术书、参考工具书、综述、年鉴等都属三次信息的范畴。

三次信息资源具有综合性高、针对性强、系统性好、知识信息面广的特点,有较高的实际使用价值,能直接提供参考、借鉴和利用。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。

(5) 高次信息资源

高次信息资源是在对大量一、二、三次信息资源中的知识信息进行综合、分析、提炼、重组的基础上,加入了作者本人的知识和智慧,使原有的知识信息增值,生成比原有知识品位更高的知识信息新产品。如专题述评、可行性分析论证报告、信息分析研究报告等,具有参考性强、实用价值高、社会效益和经济效益显著的特点。

从零次、一次、高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级次,只是一种粗略的方法,好比人的性格一样,开朗型的人也有忧郁的一面。各级次信息的界限有时并不十分明显和严格。比如从总体看,“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,既要从一、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,但又不可过分去看它究竟是何级次信息,^①只要能为己所用,都可广征博引。只有这样,才能准确、全面地利用各种信息。

以载体材料、存储技术和传递方式划分如下:

(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信息记录存储在纸

^① 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:5.

张上而形成。它既是信息资源的传统形式也是现代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) 非印刷型

非印刷型大致有交流型、缩微型、电子型。

交流型。交流型主要是以人与人之间、人与动物之间、人与自然之间、人与实物之间的信息交流。这部分信息量非常大,随时都在产生,最原始最自然,内容很新颖,但也相当随意不成熟、不稳定,在经多次传播后极大增加人为成分而加大其不确定性。比如甲说的“我很喜欢吃广菜”经不同人多次交流后很可能面目全非地变成“甲很喜欢吃广州的月饼”传播到乙的耳朵中。

缩微型。以磁性或感光材料为载体,采用磁录技术和光录技术将声音和图像信息记录存储在磁性或光学材料上,如唱片、录音录像带、电影胶卷、缩微胶卷、缩微卡片等。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。但是,其缺点是必须借助缩微阅读设备才能阅读,需要一定经济投资。

电子型。目前,电子信息已成为人们办公、经商、交流非常重要的渠道,渗透在各个领域发挥着重要作用。相应的电子设备有 MP3、电视、电脑、手机等。主要特点是存储信息密度高,读取速度快、网络化程度高、远距离传输、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,使人类知识信息的共享能得到最大限度的实现。但须借助于一定的设备才能阅读。

四、信息源

信息源是用户获取信息的来源。联合国教科文组织《UNFSCC》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。”^①

按信息源产生的时间顺序来划分如下。

(1) 先导信息源:是指产生于社会活动之前的信息源。如天气预报、股市预测等。

① 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:5.

(2) 即时信息源: 是指在社会活动中产生的信息源, 如工作记录、实验报告等。

(3) 滞后信息源: 是指在社会活动后产生的信息源, 如报刊、图书等。

按信息源传播形式来划分如下。

(1) 口传信息源: 存在于人脑的记忆中, 人们通过交流、讨论、报告会的方式交流传播。优点是获取速度快、及时、新颖; 缺点是主观随意性强、容易篡改, 其真实性与科学性需进行分析鉴别。

(2) 文献信息源: 凡是以文献作为载体形式的信息源均可称为文献信息源。它存在于各类型文献中(包括印刷型信息源和电子信息源等), 人们可以通过阅读、视听学习等方式交流传播。文献信息源是比较正规的信息源, 是人类社会所特有的人工信息源。由于文献在保存和传递信息方面具有重要的作用, 可以说文献信息源是人们获取信息最基本、最主要的来源。文献信息源具有的特殊优势是: ①能够使信息得到长久的保存和广泛的传播; ②所载信息固定而明确, 便于进行多方面的加工和利用; ③是确认人类活动、进行社会规范的工具。文献信息源的主要缺陷是在传递和交流上不够灵活、生动。^①

(3) 实物信息源: 存在于自然界和人工制品中, 人们可通过实践、实验、采集、参观等方式交流传播。优点是生动直观、选择性和针对性强; 缺点是不够灵活、零散和无规律, 有时要受场地和季节限制。

五、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

^① 肖亚明, 尹志清, 王涛. 信息检索与利用[M]. 天津: 天津大学出版社, 2009: 15.

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链接和融合。

(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

六、数据、信息、知识、情报、文献

数据、信息、知识、情报、文献的概念和关系。

(1) 数据

数据是指经整理的,可判读的描述事物特征的符号序列,它记录或载荷着信息。数据具有数值属性和物理属性。各种数字、字母、符号的组合、语

^① 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:3-4.

言、声音、图形、图像等都可以是数据,数据本身表达不出任何意义,它经过加工后可成为信息。^①

(2) 信息

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的特征性内容的总称。^②人通过获得、识别自然界和社会的不同信息来区别不同事物,得以认识和改造世界。在一切通信和控制系统中,信息是一种普遍联系的形式。它是事物存在的方式、形态和运动规律的表征,是事物具有的一种普遍属性,它与事物同在,存在于整个自然界和人类社会。

(3) 知识

知识是人类社会实践经验和认识的总结,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加工、重新组合的系统化信息的集合。因此,人类不仅要通过信息感知世界,认识和改造世界,而且要将所获得的部分信息升华为知识。^③也就是人们在认识和改造世界的过程中,对信息认知的那部分内容就是知识,可见知识是信息的一部分。知识仅存在于人类社会。

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。

(4) 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知识。^④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。情报是被传递着有特定效用的知识,它是为一定目的,具有一定时效,经过发送由载体传递,能为感觉器官或感觉系统接收的情况的报导。只有具备了有准备的头脑,包括知识基础和意识准备,信息才有可能

① 孙福强. 网络信息检索[M]. 北京:北京理工大学出版社,2014:2.

② 肖亚明,尹志清,王涛. 信息检索与利用[M]. 天津:天津大学出版社,2009:8.

③ 张林龙. 实用信息检索[M]. 上海:上海中医药大学出版社,2004:2.

④ 刘俊熙,王立义. 信息检索[M]. 北京:北京图书馆出版社,2002:6.