

A decorative graphic on the left side of the cover consists of a grid of colored squares. The top four rows contain squares with warm, abstract colors like red, orange, and yellow. The bottom three rows contain squares with cooler, more naturalistic colors like green and brown. The squares are arranged in a grid that is 7 squares wide and 7 squares high, though some squares are missing, creating an irregular pattern.

全国普通高等院校
信息管理与信息系统专业规划教材

信息采集学

夏南强 主编

殷克涛 谢乘风 副主编

清华大学出版社

全国普通高等院校
信息管理与信息系统专业规划教材

信息采集学

夏南强 主编
殷克涛 谢乘风 副主编

清华大学出版社
北京

内 容 简 介

本书是国内第一部信息采集学教材,也是第一部全面地论述信息采集知识的著作。全书在阐述信息采集基本理论的基础上,对纸质信息资源、数字化信息资源、人类社会信息资源的采集工具、方法和途径,信息服务机构所拥有的信息资源和信息服务方式与如何借助它们实施信息采集,以及信息采集相关的法律问题等做了系统而又全面的介绍和论述。本书内容丰富,结构严谨,知识新颖,自成体系。

本书既可作为高等院校信息管理与信息系统、信息资源管理、新闻学、社会学、政治学、图书馆学等专业的基础课教材,也可以供企业信息高层主管、咨询服务机构人员、网络从业人员和相关领域的科学研究工作者参考使用。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

信息采集学 / 夏南强主编. --北京:清华大学出版社, 2012.5

(全国普通高等院校信息管理与信息系统专业规划教材)

ISBN 978-7-302-28033-0

I. ①信… II. ①夏… III. ①信息学—高等学校—教材 IV. ①G201

中国版本图书馆 CIP 数据核字(2012)第 025880 号

责任编辑:白立军 战晓雷

封面设计:常雪影

责任校对:时翠兰

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京市清华园胶印厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:13.5

字 数:318 千字

版 次:2012 年 5 月第 1 版

印 次:2012 年 5 月第 1 次印刷

印 数:1~3000

定 价:25.00 元

产品编号:037112-01

本书编写人员

主编

夏南强

副主编

殷克涛 谢乘风

撰稿人

夏南强 殷克涛 赵富红 谢乘风

张 煊 王慧杰 胡红翠 喻 琳

郭力维 肖祥云 王 暘 陈 明

前言

人类已经进入信息社会,信息无所不在,无所不有,充斥于宇宙,萌生于万物。许多信息能够被人类所认知,但也有许多信息尚不能为人类所认识。本书所讨论的信息采集,显然是针对人类已经认知的信息。

人类认知的信息,也是存在于人类社会的信息。其主要表现形式是文字、语言、声音、形态、图像和符号。它们或借助于一定的载体而呈现,或在人与人的交往中起即时思想沟通、情感沟通的作用。在现代技术条件下,前者能够长久保存流传;后者则大都随时湮灭。仅此而言,我们所说的信息采集,也不可能是将人类社会存在的所有信息都视为采集对象的。信息采集的对象只能是信息资源,即经过人类主观判断处理后,能够被利用来创造物质财富或精神财富,并且能够被人类所认识、了解、传播或传输的文字、声音、图像、符号、形态、气味、数据和实物。但为了叙述的简便,并为了兼顾社会上约定俗成的见解,本书还是将信息资源的采集统称为信息采集。

人类社会中流行的信息,在当代有三大主要类别。即以纸质为载体的文献信息,以网络为载体的数字化信息,以及蕴藏于人的大脑的事实、观点、见解和知识等方面的信息。前两类信息是显性信息,后一类信息则是隐性信息,尚须通过调查采访发掘。本书所讨论的信息采集即以此三大类别信息为对象。

毋庸赘言,在当今社会,只要是正常人的日常生活、学习和工作,都离不开信息采集。这种信息采集,有时是无意识地“悄悄”进行的,有时则是有意识地专门施行的。社会发展越迅速,信息采集对于人的发展、企业的发展和国家的发展就越重要。遗憾的是,迄今国内尚无一部专门论述信息采集知识的著作,也没有一部大学专业教材。

本书承乏而编,在介绍信息采集基本理论和基本知识的基础上,对纸质信息资源、数字化信息资源和人际社会信息资源的采集工具、采集方法和采集途径,当代信息服务机构所拥有的信息资源和信息服务方式与如何借助其实施信息采集,以及信息采集相关的法律问题等作了系统而全面的论述,期望对读者从事信息采集和利用工作能有所帮助。由于参考资料比较缺乏,编写时间匆促,我们虽然尽了最大的努力,仍难免存在问题和不足之处。为了以后进一步完善,希望广大读者和专家提供宝贵意见。

本书由夏南强提出编写大纲,殷克涛、谢乘风参与了大纲的修改讨论。硕士研究生王慧杰、胡红翠、喻琳、郭力维、肖祥云、王暘、陈明编写了部分章节初稿。全书由殷克涛协助夏南强完成文字的修改和整理工作,最后由夏南强统编、定稿。

本书在写作过程中参考了大量的研究文献。为阐述信息服务机构的信息资源及其服务,也引述了一些信息服务机构网站上的介绍文字。限于篇幅,未能一一注明。此外,张大斌教授审阅了第7章书稿,并提出了一些修改意见和建议,谨此一并表示深深的谢忱。

夏南强

2011年11月18日

· III ·

目 录

第 1 章 绪论	1
1.1 信息	1
1.1.1 信息的概念	1
1.1.2 信息的特征	3
1.1.3 信息的分类	4
1.2 信息资源	5
1.2.1 资源与信息资源	5
1.2.2 信息与信息资源的关系	8
1.2.3 信息资源的类型	8
1.2.4 信息资源的特点	9
1.2.5 信息资源的作用	10
第 2 章 信息采集基本理论	13
2.1 信息采集的内涵	13
2.1.1 信息采集的实质	13
2.1.2 信息采集的对象	15
2.1.3 信息采集的工作流程	15
2.1.4 信息采集在信息服务工作中的地位	18
2.2 信息采集的功能	19
2.2.1 信息采集使信息资源“脱胎”于信息	19
2.2.2 信息采集为科学研究提供参考资料	19
2.2.3 信息采集为科学决策提供支撑	20
2.2.4 信息采集为企业经营管理提供依据	20
2.2.5 信息采集为社会生活提供便利	21
2.3 信息采集的原则	22
2.3.1 针对性原则	22
2.3.2 真实性原则	23
2.3.3 系统性原则	23
2.3.4 全面性原则	23
2.3.5 及时性原则	24
2.3.6 预见性原则	24
2.3.7 科学性原则	24
2.3.8 计划性原则	24
2.3.9 经济性原则	25

第3章 信息采集的途径与过程	26
3.1 信息采集的途径	26
3.1.1 自然观察描写途径	26
3.1.2 社会调查走访途径	28
3.1.3 文献检索查阅途径	28
3.1.4 网络信息共享途径	31
3.2 信息采集的过程	33
3.2.1 信息需求分析	33
3.2.2 信息甄别	33
3.2.3 信息选择	35
3.3 信息采集业务分析	36
3.3.1 信息采集过程中分析	36
3.3.2 信息采集结果分析	37
第4章 信息采集策略	38
4.1 信息全采集策略	38
4.2 选择性信息采集策略	40
4.3 “呈缴本制度”式的信息采集策略	42
4.4 委托式信息采集策略	43
4.5 分布式信息采集策略	45
4.6 信息采集策略的选择	47
第5章 信息采集方法	48
5.1 观察法	48
5.1.1 观察法的特点	49
5.1.2 观察法的类型	49
5.1.3 观察法的运用	50
5.2 社会调查法	51
5.2.1 访问调查法	51
5.2.2 问卷调查法	53
5.2.3 网络调查法	56
5.3 信息采购和交换法	59
5.3.1 信息采购法	59
5.3.2 信息交换法	60
5.4 信息索取与捐赠法	61
5.5 信息检索法	62
5.6 咨询法	63
5.6.1 咨询法的特点	64
5.6.2 咨询法采集信息的方式	64
5.6.3 咨询机构所提供的咨询服务	65

5.7	信息征集法	66
5.7.1	征集法的分类	67
5.7.2	利用征集法需要注意的问题	67
第6章	社会信息的采集	68
6.1	社会信息及其类型	68
6.1.1	社会信息的含义	68
6.1.2	社会信息的类型	69
6.2	社会信息的特点	70
6.2.1	社会性	70
6.2.2	时代性	70
6.2.3	复杂性	70
6.2.4	地域性	71
6.2.5	民族性	71
6.3	人际信息的采集	71
6.3.1	人际信息采集的内容	72
6.3.2	人际信息采集的方式	73
6.3.3	人际信息采集的过程	74
6.4	组织机构信息的采集	75
6.4.1	公共服务机构信息的采集	76
6.4.2	私人服务机构和社会团体信息的采集	77
6.4.3	企业信息的采集	78
6.5	社会环境信息的采集	81
6.5.1	社会环境信息采集的内容	82
6.5.2	社会环境信息采集的途径	82
第7章	纸质文献信息的采集	84
7.1	纸质文献的类型	85
7.1.1	按文献载体分类	85
7.1.2	按文献加工深度分类	85
7.1.3	按文献编辑出版形式分类	86
7.2	纸质文献的特点	89
7.2.1	数量巨大	89
7.2.2	知识蕴藏丰富	90
7.2.3	携带和阅读方便	90
7.2.4	以纸质为载体的艺术作品的不可替代性	90
7.2.5	某些纸质文献具有文物收藏价值	90
7.3	纸质文献采集需要具备的知识素养	91
7.4	纸质文献的采集途径	93
7.4.1	工具书	93

7.4.2	集成性纸质文献汇编	100
7.4.3	利用纸质文献汇编采集信息要注意的问题	103
7.5	纸质文献的数字化及其采集	106
第8章	网络信息的采集	109
8.1	网络信息资源的类型和特点	109
8.1.1	网络信息资源的类型	109
8.1.2	网络信息资源的特点	111
8.2	网络信息采集途径	112
8.2.1	基于搜索引擎途径	113
8.2.2	基于联机数据库途径	116
8.2.3	基于信息门户途径	120
8.2.4	基于社交网络服务途径	122
8.2.5	基于维基途径	124
8.2.6	基于博客和社区途径	125
8.2.7	基于RSS采集途径	125
8.2.8	其他采集途径	126
8.3	网络信息采集相关技术	127
8.3.1	网页采集技术	127
8.3.2	信息检索技术	129
8.3.3	信息抽取技术	132
8.3.4	数据挖掘技术	135
8.3.5	信息过滤技术	141
第9章	借助于信息服务机构的信息采集	144
9.1	大众图书馆系统的信息服务	144
9.1.1	大众图书馆及其类型	144
9.1.2	大众图书馆的文献信息服务	146
9.1.3	大众图书馆的网络信息服务	148
9.1.4	借助大众图书馆的信息采集	150
9.2	中国科学院国家科学图书馆的信息服务	151
9.2.1	中国科学院国家科学图书馆的馆藏特色	152
9.2.2	中国科学院国家科学图书馆的服务	152
9.2.3	借助中国科学院国家科学图书馆的信息采集	153
9.3	中国经济信息系统的信息服务	154
9.3.1	中国经济信息系统	154
9.3.2	业务系统信息服务内容	155
9.3.3	中国经济信息网	159
9.3.4	借助中国国家经济信息系统采集信息	160
9.4	新华社的信息产品服务	161

9.4.1	新华社的信息资源与服务	161
9.4.2	新华社的信息产品	162
9.4.3	借助新华社信息产品及服务采集信息	166
9.5	上海行业情报发展联盟	167
9.5.1	上海行业情报发展联盟简介	167
9.5.2	行业服务系统架构	168
9.5.3	行业情报服务类型	169
9.5.4	借助上海行业情报发展联盟采集信息	170
9.6	科学数据共享工程	171
9.6.1	科学数据共享工程简介	171
9.6.2	科学数据共享工程体系结构	171
9.6.3	科学数据共享标准	173
9.6.4	科学数据采集	174
9.7	全国文化信息资源共享工程	174
9.7.1	全国文化信息资源共享工程简介	174
9.7.2	全国文化信息资源共享工程框架体系	175
9.7.3	利用共享工程对文化信息资源的采集	175
9.8	信息咨询公司的信息服务	176
9.8.1	咨询产业的层次	177
9.8.2	信息咨询公司的服务	178
9.8.3	借助信息咨询公司的信息采集	179
9.9	国外著名服务机构的信息服务	179
9.9.1	美国国会图书馆	180
9.9.2	大不列颠图书馆文献提供中心(BLDSC)	180
9.9.3	联机计算机图书馆中心(OCLC)	180
9.9.4	文献快递中心(UnCover)	180
9.9.5	德国 Subito 在线馆际互借系统	181
9.9.6	加拿大科学技术资料中心(CISTI)	181
第 10 章	信息采集中的法律问题	182
10.1	信息采集的基本法律问题	182
10.1.1	采集信息的合法性问题	182
10.1.2	信息采集者的限制问题	183
10.1.3	采集信息的许可问题	184
10.1.4	采集信息的隐私问题	185
10.1.5	采集信息的国际问题	186
10.2	信息采集的法律管制	186
10.2.1	信息采集中的知识产权法律管制	187
10.2.2	信息采集中的隐私权法律管制	188

10.2.3	公共信息采集权的法律管制·····	189
10.2.4	特殊信息特许权的法律管制·····	190
10.3	信息采集中的权利与义务·····	191
10.3.1	信息采集者的权利与义务·····	191
10.3.2	信息发送者的权利与义务·····	193
10.3.3	信息服务平台的权利与义务·····	195
10.4	信息采集中的法律纠纷·····	197
10.4.1	信息采集活动的侵权行为·····	197
10.4.2	侵权行为的法律责任·····	198
10.4.3	调解与仲裁·····	200
参考文献 ·····		201

第1章 绪 论

社会信息化、信息社会化是人类社会发展的必然趋势。作为信息社会的主导因素,信息在当今社会备受关注。信息的生产、加工和利用一直主导着人们的生活。随着信息开发和利用的深入,信息资源作为人类社会战略资源的重要性日益彰显。信息资源采集、开发和利用的水平成为衡量社会发展程度的重要标志。现代社会,各国政府、科研机构和企业纷纷致力于信息资源的开发利用,以提高工作和服务效率,多出高质量的科研成果,增加经济效益和社会效益。信息环境的改变使人们意识到,只有认识信息,掌握信息的特征,才能合理、高效地开发信息资源。而明确信息资源的重要性,对信息资源合理采集、组织和利用,造福于社会,造福于人民,也是社会信息化与信息社会化发展过程中不可或缺的重要环节。

1.1 信 息

“信息”一词,在我国早就见于可以考据的古代文献中。如:见于一千多年前崔备的《清溪路中寄诸公》诗:“野笋资公膳,山花慰客心。别来无信息,可谓井瓶沉。”;南唐李中《暮春怀故人》诗:“梦断美人沉信息,目穿长路倚楼台。”在古代,信息都是“音信、消息”的意思。现在,“信息”作为我们耳熟能详的语词,已广泛出现在现代社会生活中。信息科学、信息管理、信息技术、信息化、信息经济和信息社会等与“信息”相关的用语随处可见。“信息”已经被赋予了新的含义。

1.1.1 信息的概念

关于信息的概念,目前的说法很多,至今还没有形成统一的观点。信息研究领域的专家、学者都是各持己见,争辩不休。仔细分析各方的论点,发现造成不同见解的主要根源在于研究信息的角度和目的不同。有关“信息”概念的解释可以归纳为以下几类。

第一类是从“事物”着眼来定义。典型的代表有 G. 克劳茨以及我国的钟义信、岳剑波等。G. 克劳茨认为信息是“物质成分与意识成分按着完全特殊的方式融合起来的产物”,它是一种“客观而不实在的东西”。钟义信将信息理解为“信息是事物存在的方式或运动的状态,以及这种方式/状态的直接或间接的表述。”岳剑波指出,“所谓信息,并非指定事物本身,而是指用来表现事物特征的一种普遍形式。”这类信息的定义借用哲学语词,试图抓住信息的本质。但是,概念的定义过于空泛,不易理解,忽略了信息的多样性,在进行信息研究时不好把握。故钟义信后来又提出了狭义的概念:“信息就是消息,即通信,包括面对面的通信和远距离通信的内容。”

第二类是从信息的可交流、交换性着眼来定义。这类定义以美国的维纳为代表。美国数学家、控制论的奠基人诺伯特·维纳在他的《控制论——动物和机器中的通讯与控制问题》中认为,信息是“我们在适应外部世界,控制外部世界的过程中同外部世界交换的内容的

名称”。这类定义着眼于信息的可交流、交换性,可以说,为信息利用开了先河。但是,单从这些特征来定义比较片面,难以把握信息的整体内涵。

第三类注重信息的表现形式,从信息的可记录性来定义。这类观点以史密斯、谢拉为代表。在信息资源管理领域颇有建树的美国学者史密斯认为信息是“事物或记录”,许多事物都可以是信息,文本固然是信息,图片、录音磁带、博物馆陈列品、自然物体、试验和事件也是信息。谢拉(J. Shera)从生物学家与图书馆员的眼光来定义,认为信息都表示一种“事实”,它是我们通过感觉所获得的刺激,它可以是一个独立的事实,也可以是一组事实,但它仍然是一个单位,是一个思维单位。这种定义从单一的形式来看待信息,不注重信息的内容,使信息表现形式与内容脱节,在认识上缺乏完整性,也缺乏以发展的眼光来看待新形式问题。致使人们只能从表面认识事物,很容易在信息的采集和开发上陷于困境。

第四类是从信息的内容和意义特性着眼来定义。麦凯(D. McKay)认为,信息是有助于表述的,当“我们所知道的内容”发生变化时,我们就接受了信息。信息是能够对一个表述或事件状态进行逻辑转换或强化的内容。表述可以是明晰的(如一幅地图或一剂药方),也可以隐含在接受者的定向行为中。我国的 GB/T70722002 标准(即《信息分类和编码的基本原则与方法》)将信息定义为“一切有含义的具体或抽象事物或概念的真相及相关陈述,通过数据、消息及其进一步细节表达出来”。这类观点抓住信息的可识别性,以及信息的有意义特性,作为解释信息的出发点,从内容与意义来看待信息,能够抓住信息所表达的含义。但是这类定义忽略了信息的主体与外在形式,使信息的含义变得难以琢磨。

第五类从信息的可选择、可获取角度出发来定义。米勒(G. Miller)认为信息是我们进行选择时所需要的某种东西。不论它是什么内容,我们所需要的信息的量取决于选择时的复杂程度。当我们面临许多互相平等的选择,并作出某件事能够发生的选择时,我们所需要的信息量就比在两种选择方案中选择一种方案时所需要的信息量多。《辞海》(1999 年版)对信息的解释是:“信息是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,通常需要通过处理和分析来提取。”这类定义显然比较狭隘,没有从哲学的高度来认识事物的本质。众所周知,信息不一定可供人类选择、获取,如动物、植物发出的信息。蚂蚁利用味觉和嗅觉获取信息,彼此进行联系,识别同窝伙伴;野猪在平时总是把尾巴转来转去,但一旦觉察到有危险时,就会扬起尾巴,在尾尖上打个小卷给同伴以报警信息;蜜蜂在发现蜜源以后,就会用特别的“舞蹈”方式(圆形舞和 8 字形摆尾舞)发出信息,向同伴通报蜜源的远近和方向。日本京都大学的科学家研究甚至发现,利马豆植株会产生特殊的化学物质,从而发出危险信息保护自己并警告邻居们害虫蛛螨正在迫近。动物乃至植物之间通过气味、行为或声音等表达出各种信息用于它们之间传递,自然生物间充斥着各种信息。

第六类从人对事物认知的角度出发定义信息,认为信息是知识。持这类观点的学者有布鲁克斯、米哈伊洛夫等。英国的布鲁克斯(B. C. Brooks)认为,“信息是使人原有的知识结构发生变化的那一部分知识。”前苏联信息(情报)专家米哈伊洛夫将信息(情报)视作为存储、传播和交换的对象的知识。《牛津词典》对信息的解释是“信息就是谈论的事物、新闻和知识”。国际标准组织 ISO 对信息的解释是“对人有用的,能够影响人们行为的数据,而在日常用语中,信息仍然是消息、情报、情况、资料、知识等的总称”。实际上我国许多学者都认为,信息包含知识。从这个意义上讲,将信息简单地认同为知识实际上是片面地理解和阐释

了信息。

信息是事物存在的方式或运动状态的表现。随着科学技术的进步,人们对人类社会和自然界的认识越来越深入,社会信息不仅包括了人与人之间的信息交流,还包括了人与机器之间、机器与机器之间的指令转换,以及其他动物界、植物界气息、信号的发出,等等。在研究、理解信息时,应重点强调信息的主体与表现方式。武汉大学严怡民的定义综合了前面几种观点,突出信息的主体,认为“信息是生物以及具有自动控制系统的机器通过感觉器官或相应设备与外界交换的一切内容”。这种观点比上述几类观点前进了一大步,但定义比较笼统,不太清晰、明确,也没有顾及信息的表现形式。而且信息的表现,也并非为了与外界交换。

我们认为,信息是生物以及具有自动控制系统的机器通过感觉器官、细胞组织或者具有类似功能的设备部件所发出或表现出来的文字、语言、声音、形态、图像、符号或气味等。这样定义,既抓住了信息的本质特征,也把握住了信息的表现形式,有利于对信息的认识、采集、分析、获取和利用。

1.1.2 信息的特征

维纳说过:“信息就是信息,不是物质也不是能量。不承认这一点的唯物论,在今天就不能存在下去。”这说明信息作为独立的概念具有区别于其他事物的特性。信息不是具体的物质,无法利用工具测量。信息既有与物质相同之处,但更有其独特的地方。信息的特性主要表现在以下几个方面。

1. 普遍性

按照广义的理解,信息是事物的存在方式或运动状态的体现。作为事物的呈现方式——物质,是具有普遍性的,是无处不在的,因此,信息具有物质的特性,具有普遍性。在我们生活的世界中,无论空中、水底还是陆地,生物无处不在。这些生物与人类共存,组成多姿多彩的世界。在这世界上,各种有形的、无形的信息充斥其间。有些信息为人类所认识、察觉、体会、感觉;有些则尚不为人类所认识、感知和了解。人类通过其发展,借助技术和工具,在逐渐拓展其信息的视野。尽管存在许多难以认知的信息,但可以肯定的是,信息是普遍存在的。

2. 动态性与时效性

事物是在不断变化发展的,信息作为事物的反映也必然随之变化发展。因此,信息具有动态性,会随着事物的变化而展现出不同的内容。社会的不断变化,不断进步,改变着人们对信息的认识,也提醒我们要用发展的眼光来看待信息,定义信息,利用信息。当今社会,许多信息能够产生巨大的经济效益。特别是重大的商业信息,往往能够带来巨大的财富。一个能够产生经济效益的信息,其使用价值会随着时间的流逝而衰减。从产生到被利用,时间越短,其利用价值越大,反之就越小。这也就反映出信息具有时效性的特点。

3. 可识别性、可传递性和可共享性

信息作为人类社会交流的内容通常以文字、图像、声音、符号、数据和形态等形式展现出来,为人类所识别。这种可识别的信息经过人类智慧的加工处理,以简单易懂的方式为人们所理解。同时,信息这种能够被人类用感觉器官和科学仪器等方式来获取、整理和认知的特性,也进一步促进了信息的交流,实现了信息在人与人、人与物、物与物之间的传递。这种传递最大的影响是人类文明的发展、科技的进步。可传递性使同一信息可以在不同时间、不同空间被多个个体所共有,而且还能够被无限地复制、传递,甚至供人类共享。

4. 可扩充性与可压缩性

随着时间的推移,事物在不断变化,信息也可以不断被扩充,以应对事物的相应反映。同时,从系统论的观点出发,整体大于部分,大量的信息聚集在一起,经过加工整理,可以表达更为丰富的内容。信息的可扩充性为信息汇集、组织、分类和利用奠定了基础。现代社会的很多联机数据库是不断在进行扩充的,随着信息的与日俱增,其扩充的进度非常惊人。技术的进步使得人类还可将聚集的信息浓缩,存储在不同的载体上,形成多样的信息种类。2010年东京日本电子高新技术博览会(CEATEC)上,东京电气化学公司(TDK)展示了一款可存储1TB内容的光盘原型,该技术让一张光盘拥有16个存储数据的层面,每层可存下32GB信息。信息的这种可压缩性为人们利用信息提供了便利。

正是具有这些非常重要的特征,信息在人们生活中的地位才显得更加突出,所起到的作用也就更大。早在1944年,罗斯福总统就在给科学研究与发展局(R&D)局长V. 布什博士的信中指出,在战争期间,“由R&D发起,大学与私营企业中数千名科学家所发展的信息、技术与研究经验,在和平日子里可为改善国民卫生,建立能够带来就业机会的新企业,改进国家生活标准上做好准备。”可见,开展信息的采集和利用活动,对整个国民经济发展起着至关重要的作用。

1.1.3 信息的分类

信息广泛分布于自然界与人类社会。依照不同标准可以把信息区分为不同的类型,各类型之间相互交叉,并无明确界限。

1. 按产生的来源分

按信息产生的来源,可以将信息分为自然信息和社会信息。自然信息是自然物及自然现象的存在方式和状态的自身显示,包括自然界中人类以外的各种生物信息、非生物物质存在与运动信息(如地震)、非生物物质与生物物质之间的作用信息等。它不依赖于人的存在而存在。社会信息是人类对自然、人类社会现象及人类思维的反映。它以人类的活动为出发点,是人类在社会活动中所认识、产生和传播的信息。其作用对象主要是人,并能被人类所认识或利用。

2. 按认识论层次分

按认识论层次,可以将信息分为语法信息、语义信息和语用信息。我们通常把与事物运动状态及变化方式的因素相联系的认识论信息称为语法信息,把与含义因素相联系的认识论信息称为语义信息,把与效用相联系的认识论信息称为语用信息。这3个层次依次体现了人们对事物认识的程度,也是人们认识事物、理解事物的过程。

3. 按信息所体现或表述的内涵分

按信息所体现或表述的内涵,可以将信息分为政治信息、经济信息、军事信息、法律信息、艺术信息等。政治信息是指人类政治活动中形成的信息总和,主要包括政治制度、意识形态、国内外政治态势、国家方针政策等信息。经济信息是指在经济活动中形成的信息总和,随经济活动而产生和发展,主要包括国家经济政策、社会生产力发展、国民经济水平、经济系统比例与结构、新技术开发与应用、金融、市场、生活需求等方面的信息。军事信息是指在军事活动中所需要的信息总和,比如国家军备力量评估方面的信息、战争信息、武器信息等。法律信息主要由法律制度、法律体系、立法、司法等各种法律法规构成,还包括法律执行信息、社会治安信息等。艺术信息主要是社会个体和群体在发展过程形成的与艺术相关的信息,如书法、美术、戏曲、电影等方面的信息。

4. 按信息的运动状态分

按信息的运动状态,可以将信息划分为动态信息和静态信息。动态信息指时间性强,处于不断变化的信息,如市场信息、交易所价格指数、金融信息、战场军事信息、时事新闻等。静态信息则指被记载并存储在一定载体上的信息,如纸质文献信息、磁盘文献信息、光盘数据库信息等。

1.2 信息资源

随着人们对信息认识的加深,人类从事信息活动的范围越来越大,信息产生的效益也越来越明显。信息推动社会进步的作用以及现代信息技术的发展使信息采集和利用的领域正逐步扩散到全世界,人类社会也由此进入信息社会。在信息社会中,信息的地位越来越重要。特别是现代科技的发展促使信息成为推动社会发展的重要因素,信息资源也成为社会重要的资源。将信息作为资源来采集、存储、开发与利用也成为现代社会工作中最重要的领域之一。

1.2.1 资源与信息资源

1. 资源的定义

顾名思义,“资源”之所以能成为资本的来源,首先,它可以代表财富的来源。如《辞海》(1999年版)对“资源”的解释是:“资财的来源,一般指的是天然的财源”。《现代汉语词典》

(2005年版)的解释是：“生产资料或生活资料的天然来源”。西方学者兼经济学家阿·兰德尔(A. Randall)在《资源经济学》中提出，“资源是有人发现的有用途和有价值的东西”，并将资源分为两个范畴：一是自然界赋予的自然资源；二是人类社会中人的劳动创造的各种资源。其次，资源具有可用性。也就是可以以某种形式为人类所开发利用。资源来源于英语的 resource 一词，英语的 resource 则源自法语 resourdre，在法语中为“再生(兴起)”之意。根据《牛津英语大词典》(第2版)等释义，资源的义项有6项：

- ① 满足需求和匮乏的手段，在必需时可以提取的原料与存储；
- ② 援助或帮助的可能性；
- ③ 可以从困境和紧急状态下解救出来的方法、权宜之计、办法、谋略；
- ④ 放松与娱乐的手段；
- ⑤ 采取手段达到目的或应付困难的能力；
- ⑥ 定语和搭配，如资源配置、资源密集、资源贫困、资源中心等。

从以上释义我们可知，resource 一词既可以指物质的自然资源，也可以指非物质的精神和智力方面，或者兼指两者。再次，资源是一定的客观存在。这种客观存在一般为人类所熟知。马费成在《信息资源管理》中说：“资源是指自然界和人类社会中一种可以用以创造物质财富和精神财富的具有一定量的积累的客观存在形态，例如：土地资源、森林资源、海洋资源、人力资源等。资源一般可以分为经济资源和非经济资源两大类。”

上述有关“资源”的定义，除了马费成的定义比较符合现代社会有关“资源”的客观实质之外，其余定义大都忽视了“人”和“信息”这两个很重要的因素。但马氏的定义中“具有一定量的积累的”这个定语以及“形态”这个词语使用欠妥。在当今社会，不仅“物”是资源，可以用来创造财富和价值，“人”和“信息”也是资源，甚至是更重要的资源。在政治、经济、军事等方面的重大决策中，“信息”作为资源，从某种意义上说，比“物质”资源更为重要；在科技强国、创造发明中，一流科学家也可视作“资源”，而且肯定比“物质资源”更为重要的资源，是各国争相“抢夺”的对象。因此，可以从高度抽象的角度把“资源”定义为：资源是一切可被人类开发和利用的客观存在。

2. 信息资源的定义

在信息没有充分受到重视之前，作为资源，物质、能源一直是人类争夺的对象，有时甚至引发战争。矿山开采金属原料、地底和海洋石油开发等活动都是人类追求物质资源的表现。随着科学技术的发展和信息社会的到来，信息的重要性愈来愈被人们所认识，信息推动社会进步的作用不言而喻。信息已经成为一种人类赖以生存和发展的战略资源。

随着信息量的增长和信息技术的进步，人们对信息资源有了新的认识。但是，仁者见仁，智者见智。国内外专家学者对信息资源这一概念尚未达成统一的观点。现将有代表性的见解列举如下。

美国著名信息管理专家霍顿(F. W. Horton)从政府文书管理的角度认为：“信息资源有两层意思：一是当资源一词为单数时，信息资源是指某种内容的来源，即包含在文件和公文中的信息内容；二是当资源一词为复数时，信息资源指支持工具，包括供给、设备、环境、人员、资金、环境等。”