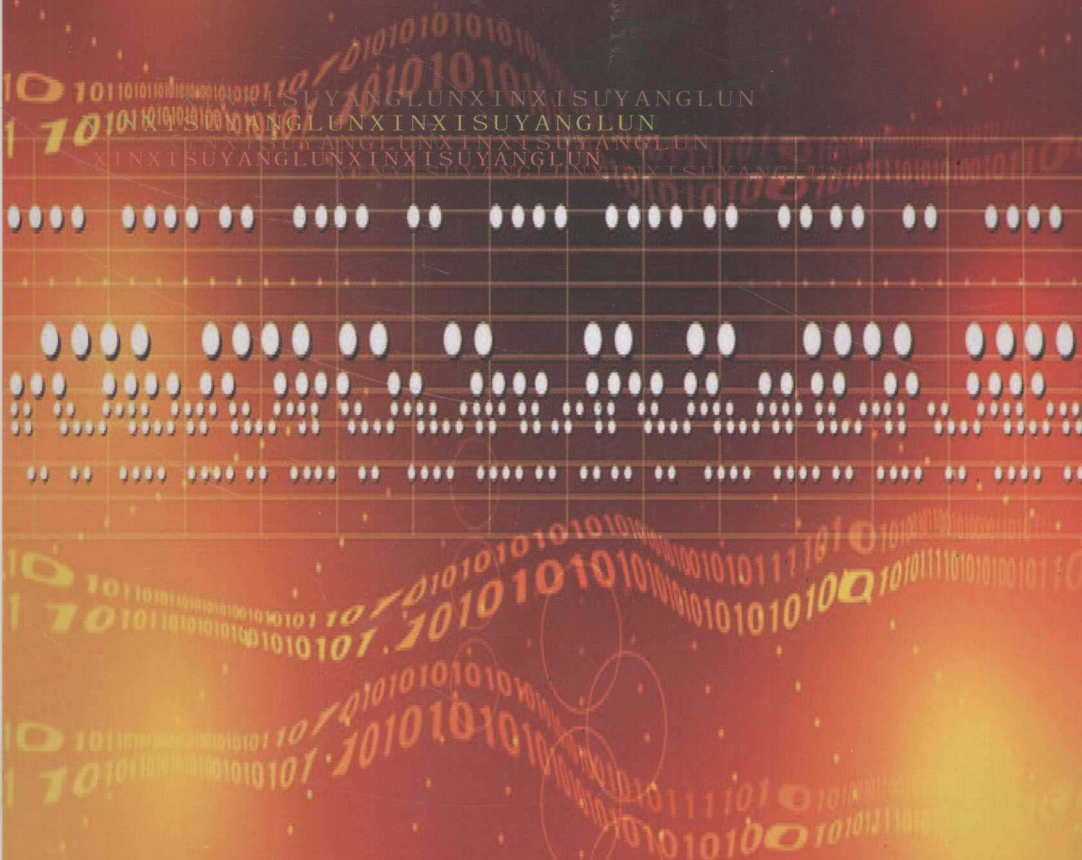


XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN

XINXISUYANGLUNXINXISUYANGLUN

信息素养论

秦殿启 著



南京大学出版社

本书得到盐城工学院学术专著出版基金资助
本书系江苏省高校哲学社会科学研究基金资助项目
“泛在信息社会高校教师信息素养体系的构建”
(项目编号: 2012SJB870009) 的重要成果

信息素养论

秦殿启 著

XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN
XINXISUYANGLUNXINXISUYANGLUN



南京大学出版社

图书在版编目(CIP)数据

信息素养论/秦殿启著. —南京:南京大学出版社,
2012.12

ISBN 978 - 7 - 305 - 10794 - 8

I. ①信… II. ①秦… III. ①信息学—研究
IV. ①G201

中国版本图书馆 CIP 数据核字(2012)第 271725 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

网 址 <http://www.NjupCo.com>

出版人 左 健

书 名 信息素养论

著 者 秦殿启

责任编辑 张黄群 吴 汀 编辑热线 025 - 83596997

照 排 南京紫藤制版印务中心

印 刷 常州市武进第三印刷有限公司

开 本 787×1092 1/16 印张 18.75 字数 480 千

版 次 2012 年 12 月第 1 版 2012 年 12 月第 1 次印刷

ISBN 978 - 7 - 305 - 10794 - 8

定 价 38.00 元

发行热线 025 - 83594756

电子邮箱 Press@NjupCo.com

Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

序

——面向泛在信息社会的公民信息素养

泛在信息社会(Ubiquitous Information Society)是信息技术高度发达,互联网、物联网和RFID有机融合而逐步形成的信息无处不在、信息交流无处不能的信息社会。其特点可概括为信息泛在、技术泛在、服务泛在。在泛在信息社会中,人们的生活方式、学习方式、工作方式以及教育方式将发生巨大变革,信息成为人类的必需品。信息素养(Information Literacy)是“终身学习的核心,它能使人们在整个一生中有效地寻求、评价、利用和创造信息,以便达到其个人的、社会的、职业的和教育的目标,它是数字社会的一种基本人权,能促进所有国家的社会内涵”(联合国教科文组织,《信息社会灯塔:关于信息素养和终身学习的亚历山大宣言》,2005年11月)。因此,提升公民信息素养已经成为一项紧迫而具有战略意义的教育工程。

公民信息素养的研究和培育工作是一个综合工程,是一个没有止境的、系统的、螺旋式上升循环的科学发展工程,可以根据年龄、层次、类别等分步、分级开展,也可以作为一项公益事业有序地推广和普及。高等教育阶段对于开展信息素养教育具有不可替代的战略地位:大学图书馆是科学知识的宝藏、文化信息的海洋,在开展信息素养教育方面具有得天独厚的资源优势,是培育信息素养的摇篮;大学生是接受信息素养教育的最佳受众并将最快成为开展信息素养普及工作的后备力量;大学图书馆有一支从文献检索课教学演进到信息素养教育的专业师资队伍;大学教师有着广博的学科背景和深厚的专业理论,适宜从学科和专业的视阈适时融入信息素养的培育内容。这样就形成了一种合力,这是信息素养教育的极佳途径。

公民信息素养(Citizens Information Literacy)是指公民为适应信息社会而必备的信息方面的知识、能力和文化修养,其本质是公民对信息的驾驭能力和修养,是传统文化素养的延伸和拓展;信息素养是信息社会的必然产物,作为一个公民,提升信息素养是适应泛在信息社会的生存之道。

本书基于上述背景,汲取国内外众多学者的研究精华,试图构建一个信息素养的理论框架,暂且称之为信息素养理论体系。该体系是一个开放式的活的有机体,由三个层面九个要素构成:从知识层面言,信息素养的知识素养包含有信息理论、信息资源和信息整合三个要

素,在信息素养理论体系中处于基础地位;就能力层面讲,信息素养的能力素养包含着信息技能、信息意识和信息创造三个要素,在信息素养理论体系中居于中心位置;依文化层面看,信息素养的文化素养包含了信息网络、信息伦理和信息美学三个要素,在信息素养理论体系中位于最高层,显示了文化的升华作用和传承价值。书中将每个要素分章阐述,每章分为3节,计9章27节。

公民信息素养的培育既需要政府及各类教育培训机构的重视,又需要公民自身主动并积极地参与,这是一个良性互动的过程,整个过程首先建立在对信息素养的价值和意义充分认识的基础之上。其次是前瞻到泛在信息社会信息素养的战略意义,并从提升信息意识开始,从身边的信息需求开始,继而通过主动地、逐步地、全面地吸纳、建立、提升其他信息素养要素,不断适应泛在信息社会对公民的信息素养需要。第三,必须积极提炼公民信息素养的信息美学认识,特别是信息的真善美的审美能力和创造能力。第四,积极利用信息技术新成果,不断提升信息技能。第五,公民信息素养教育必须采取众多学科专家和理论的合力来互动^①。

本书有些概念和观点还需完善,在此抛砖引玉;由于本人学浅才疏,书中错误之处在所难免,敬请读者和专家批评指正!本书可作为高校文献检索课、信息素养课的教材和参考用书,可作为信息素养教育的自学用书,也可供信息素养研究者参考。

^① 秦殿启. 泛在信息社会公民信息素养教育的探索[J]. 图书情报工作网刊, 2012(08): 43-48.

目 录

A 知识篇

第 1 章 信息理论	3
1.1 信息本体理论	3
1.2 专业信息理论	15
1.3 行业信息理论	36
第 2 章 信息资源	47
2.1 理论信息资源	47
2.2 实践信息资源	59
2.3 生活信息资源	75
第 3 章 信息整合	87
3.1 宏观信息整合	87
3.2 中观信息整合	97
3.3 微观信息整合	106

B 能力篇

第 4 章 信息技能	119
4.1 信息分析技能	119
4.2 信息检索技能	131
4.3 信息获取技能	148
第 5 章 信息意识	162
5.1 信息注意力	162
5.2 信息敏锐力	171
5.3 信息洞察力	179
第 6 章 信息创造	186
6.1 提出新方法	186
6.2 阐述新观点	194
6.3 构建新理论	201

C 文化篇

第7章 信息网络	215
7.1 人际网络	215
7.2 社会网络	221
7.3 虚拟网络	230
第8章 信息伦理	239
8.1 信息道德	239
8.2 信息免疫	246
8.3 信息法规	251
第9章 信息美学	263
9.1 信息的真实性	263
9.2 信息的价值性(善)	273
9.3 信息的审美性	282
参考文献	290
跋	293

A 知识篇

第1章 信息理论

第2章 信息资源

第3章 信息整合



第 1 章 信息理论

《老子》第四十五章:大成若缺,其用不弊。大盈若冲,其用不穷。大直若屈,大巧若拙,大辩若讷。躁胜寒,静胜热。清静为天下正。信息理论作为信息素养人的第一知识要素,就应该有一个如老子(老,姓也)所言的虚怀若谷、敦厚朴实、宁静致远的心智和心胸;在信息理论的培育过程中,要博采众长、辩证思考、求真务实。

本章所述信息理论主要指有关信息的重要理论。

信息既包括信息本身,也泛指专业信息、行业信息等内容。

专业信息主要是指和信息密切相关的文献学、图书馆学、情报学、档案学等信息。

行业信息是指各行各业的相关信息,既有显性的,又有隐性的;既有传统的,又有新兴的,是一个数量庞大,资源无法穷尽的信息流。

本章重点探讨信息论、控制论、系统论;文献学理论、图书馆学理论、情报学理论及档案学理论;教育信息资源理论等,深刻理解理论是实践的导师的内涵。

1.1 信息本体理论

1.1.1 信息的几种描述

信息在自然界、人类社会以及人类思维活动中普遍存在。信息最初的表现形式就是信息活动。

1. 《老子》与信息

我国春秋时期的哲学典籍《老子》(亦称《道德经》)中已有“信息”的深刻描述。《老子》第21章有曰:“道之为物,惟恍惟惚。惚兮恍兮,其中有象;恍兮惚兮,其中有物;窈兮冥兮,其中有精;其精甚真,其中有信。”

这几句话可以翻译为:道化生万物的过程是隐约而至的。在隐隐约约之中,有了物的形状;在隐约之中,又有了物的内容。昏暗深沉中,情形渐现。这情形表现(越发)本真,它蕴含着(新生)事物的(数量、体态、色泽、声音、运动规律等等丰富的)信息。可见,《老子》中的信息包含了世界万事万物发生发展的情态本真描述及其运动规律。

2. 英文“信息”(Information)的含义

英文“信息”一词一般用“Information”,该词包括情报、资料、消息、报导、知识、声息、耳目、学问、风声等多层的含义(如图 1-1-1)。

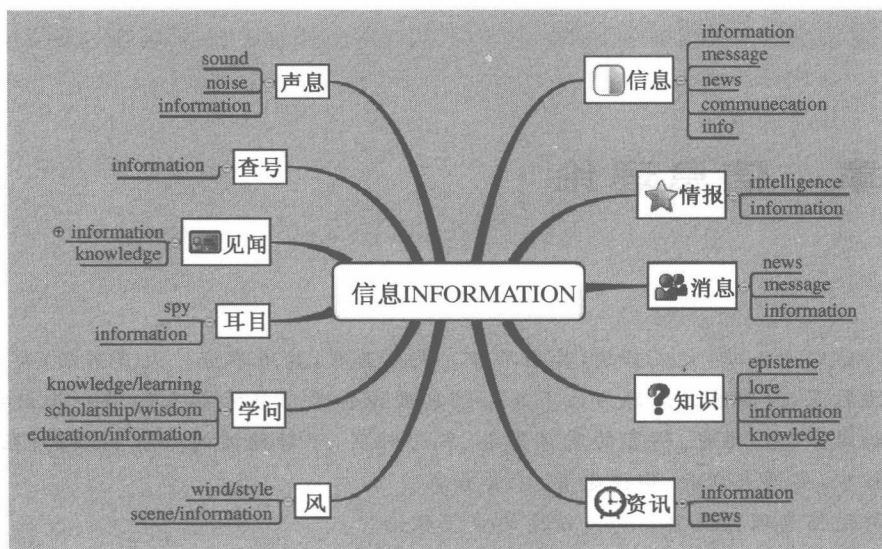


图 1-1-1 英文信息含义图

通常情况下,人们把信息一词看作是消息的同义词,把信息理解为有一定的内容或知识的消息。在科研、教育以及通信领域,信息的含义比消息和情报更为广泛,消息、情报、指令、代码、符号语言、文字等一切有意义内容都是信息。哈特莱把消息、情报、信号、语言定义为载体,承载的内容称为信息。维基百科对信息(Information)的解释如下:

Information, in its most restricted technical sense, is a sequence of symbols that can be interpreted as a message. Information can be recorded as signs, or transmitted as signals. Information is any kind of event that affects the state of a dynamic system. Conceptually, information is the message (utterance or expression) being conveyed. This concept has numerous other meanings in different contexts. Moreover, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, representation, and especially entropy. (From Wikipedia, the free encyclopedia)

维基百科对信息的解释在基本包含了上述的英语信息含义图内容的基础上,它特别强调了信息在通讯系统的含义和熵(Entropy)的引入,这应归功于申农的贡献。

3. 信息作为科学概念的内涵

信息是通过一切媒介传递、表达特定内容的符号与属性,是客观世界一切事物的存在和变化状况与人的认识能力的交换物。这里要特别强调“属性”的概念。

信息是脱离了载体的属性,它用数学或一般符号来表征,是经提升而成的一个新的事物^①。1922年,卡松提出了信息的概念,具体反映在他提出的边带理论,指明信号在调制(编码)与传送过程中与频谱宽度的关系;同年,哈特莱发表《信息传输》的文章,首先提出消息是代码、符号而不是信息内容本身,使信息与消息区分开来,并提出用消息可能数目的对数来度量消息中所含有的信息量,为信息论的创立提供了量化思路。

^① Ralph M Stair, George W Reynolds. 信息系统原理[M]. 张靖,等,译. 北京:机械工业出版社,2000.

1928年,哈特莱强调了“信息是脱离了载体的属性”,或说“信息是脱离了载体的一个新的事物”。此说使得信息与其相应属性有了本质区别,从而为管理者开辟了一个活跃的信息思维空间;同时,也为后来的信息技术开创了一个基本前提^①。比如,黄河干涸,黄河涨水,黄河结冰,黄河改道。黄河是载体,干涸、涨水、结冰、改道是其属性,是脱离了黄河的一个新事物,是有关黄河的水文信息。那么,信息与属性有何关系呢?一个事物的“信息”既已脱离了载体,那它就是具有独立性的新事物。比如,关于干涸的信息,关于涨水的信息,关于结冰的信息,关于改道的信息,已经不是局限于黄河所独具的水文信息。这时,只能说它“代表”了原事物的某项属性(某项特征),而不能说它就是那项属性。既然信息具有独立性,那它就可能具有失真和误差——近似性,但作为属性却不可能。比如,黄河就是黄河,它不是泾河、渭河、淮河,更不是长江、黑龙江;但是,既然信息具有独立性,是新事物,可以有近似性,那就可以表述它、记载它,比如将其表示成符号、语言、文字、声音等。这种表述就是信息的转换,也可转换成便于传递、存储、再提取、再转换、再应用的内容,甚至可作进一步加工、处理,这就是信息技术的内容。可见,一旦意识到信息脱离了原属性的“岗位”,而成为新事物这一理念之后,即为信息论打开了一扇大门,进一步的工作,就只待赋予它数理基础了。这项工作正是申农于1948年完成的^②。

信息作为一个科学概念是在19世纪提出的,申农(Claude Elwood Shannon, 1916—2001年,美国数学家,狭义信息论的创立者)的《通信的数学原理》和维纳(Norbert Wiener, 1894—1964年,美国数学家,控制论的奠基人)的《控制论》,奠定了信息论的基础。

1948年,申农在Bell System Technical Journal上发表了《通信的数学原理》(A Mathematical Theory of Communication)一文,将熵的概念引入信息论中。熵(Entropy)由鲁道夫·克劳修斯(Rudolf Clausius)提出,并应用在热力学中。熵指的是体系的混乱程度,它在控制论、概率论、数论、天体物理、生命科学等领域有重要应用,在不同的学科中也有引申出的更为具体的定义,是各领域十分重要的参量。在信息论中,熵表示的是不确定性的量度。信息论的创始人申农在其著作《通信的数学原理》中提出了建立在概率统计模型上的信息度量。他把信息定义为“用来消除不确定性的东西”。

信息是一种十分广泛的概念,它在自然界、人类社会、宇宙空间以及人类思维活动中普遍存在。不同事物有着不同的特征,这些特征通过一定的物质形式(如声波、电磁波、图像等)给人带来某种信息。例如,人的大脑通过感觉器官所接收到的有关外界及其变化的征候,就是一种信息。因此,信息可以定义为:宇宙空间一切存在物及潜在物的发生、发展、变化、运动、静止等所形成的物象、征兆与人、超人及其他生物具有自动控制功能的系统,通过感觉器官和相应的设备与外界进行交换的一切内容。

4. 波普尔“三个世界”中的信息观

卡尔·波普尔(Karl Raimund Popper, 1902—1994年),英国科学哲学家,出生于奥地利维也纳,父母都是犹太人,第二次世界大战期间移居英国。1934年完成的《科学研究的逻辑》一书标志着西方科学哲学最重要的学派——批判理性主义的形成。由于他在学术上的成就,1965年被英国皇室授予爵士称号。波普尔的对待科学证伪主义态度,使得他在科学哲学史上

① 徐飞,高隆昌.二象对偶空间与管理学二象论:管理科学基础探索[M].北京:科学出版社,2005:111.

② 徐飞,高隆昌.二象对偶空间与管理学二象论:管理科学基础探索[M].北京:科学出版社,2005:112.

具有重要地位。他的“三个世界”理论在多个学科得到诠释,他认为“第一世界”最先存在,而“第二世界”在新的层次上出现,“第三世界”又出现在更高的层次上。

波普尔的“第一世界”是指包括地球在内的全部宇宙自然界,系物理领域,包含了客观物理世界的信息,即本体论意义上的信息,它反映事物运动的状态及其变化的方式。

波普尔的“第二世界”是指人的精神世界,包括所有人的心理活动、心理状态等,系主观现实领域,包含了人类主观精神世界的信息,即主体或认识论意义上的隐性信息,它反映人类能感受的事物运动状态及其变化方式,并处在意识思维状态的潜在信息。

波普尔的“第三世界”又称为“客观知识世界”,包括人类所创造的语言、艺术作品、宗教、科学、技术等,系客观知识领域,包含了客观意义上概念世界的信息,即主体论或认识论意义上的显性信息,它反映人类所表达的事物运动状态及其变化方式,用语言、文字、图像、影视、声音、数据等各种载体来表示,汇成一个实在的自主的信息世界。

5. 现代社会“三大支柱”中的信息观

信息时代是以信息为主体的社会的特称,信息、物质、能源被称为现代社会的三大支柱。信息的重要性日趋显著,在一定程度上超越了物质和能源,成为体现综合国力的重要标志之一。在人类社会运动中,信息被人类应用于物质和能源的研制和开发,物质和能源不断地从生产者手中转运到使用者的工作和生活中,在这个过程中,又产生了一种有关物质、能源运动变化的信息流。有学者结合波普尔的“三个世界”理论,提出了自然物质、信息以及智慧产物的认识。自然物质,有形,独立于信息,作为一元取数,为一,具有分子原子结构,分解后可以在门捷列夫的《化学元素周期表》中找到相应的位置,遵守质能守恒定理,可以进行可控实验;信息,无形,独立于物质,作为一元取数,为一,包含人类(及动植物)的精神、意识、思维;(磁、电等作用力或能量)场、数、规律、印迹(烙印、影响)等,能在不同的物质(介质)中复制、加工、传播;智慧产物,介于自然物质与信息间,为二者的非自然组合体,作为一元取数,时而大于一、时而小于一,如蜂房、金字塔、UFO、因特网等^①。这种认识具有哲理性和启发性。

6. 信息概念的螺旋式发展

查《辞源》(商务印书馆·1989年7月第1版):【信息】指音讯,消息。唐·李中《碧云集》中《暮春怀故人》诗:“馆馆寂寥三月尽,落花重叠盖莓苔。惜春眷恋不忍扫,感物心情无计开。梦断美人沈信息,目穿长路倚楼台。琅玕绣段安可得,流水浮云共不回”,指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。通常须通过处理和分析来提取。信息和物质、能量被称为构成系统的三大要素。信息的量值与信息的随机性有关,如在接收端无法预估消息或信号中所蕴含的内容或意义,即预估的可能性越小,信息量就越大。

两者均将“信息”界定为:音讯、消息、信号的具体内容和意义。哲学家认为,信息是人类认识世界的依据;数学家认为,信息是一种概率;物理学家认为,信息是“熵”(Entropy);通信学认为,信息是“不定度”的描述;图书信息领域的专家认为,信息是可以以各种形式进行传播、记录、出版及发行的观念、事实及论著。

信息是用文字、符号、图案、声音、数据或信号等形式通过一定的载体或技术传递和处理

^① 方关金. 我读波普尔:“三个世界”[EB/OL]. [2012-04-19]. <http://5460.xaonline.com/gy5460/jsp/article/dis-pArt.jsp? dNum=93769&articleType=0>.

来表现各种相互联系的客观事物在运动变化中所具有特征内容的总称。信息是物质的存在方式、形态和运动规律的反映形式及表现特征,是事物所具有的一种普遍属性,它与物质同在,存在于整个自然界、人类社会和宇宙空间。

7. 信息的数学表达

姜春艳等老师用数学表达的熵、条件熵、信息的定义、信息的性质如下^①:

熵的定义

如果任一事物有 N 种情况,每一种情况的概率为 $p_{i,j}=1,2,\dots,N$,有 $\sum_{i=1}^N p_i = 1$,定义熵:

$H = - \sum_{i=1}^N p_i \ln p_i$, 熵 H 可以作为事物不肯定程度的量度,它的值越大,事物不肯定程度越大。

条件熵

如果有两个随机变量 X, Y , 当 Y 取某一确定值 Y_j 时 X 的概率是条件概率,与此对应的熵叫条件熵 $H(X/Y_j)$, 即

$$H\left(\frac{X}{Y_j}\right) = \sum_{i=1}^N - p(X_i/Y_j) \ln p(X_i/Y_j)$$

信息的定义

如有两个随机变量 X 与 Y , 把 X 的熵与 Y 对 X 的条件熵之差定义为 Y 对 X 的信息,信息 $I(X)$ 有

$$I(X) = H(X) - H(X/Y)$$

也可以写成

$$I(X) = H(X) + H(Y) - H(X, Y)$$

信息的性质

① 信息 $I \geq 0$

② 当 Y 与 X 互相独立,信息 $I(X) = I(Y) = 0$

③ $I(X) = I(Y)$, 即 Y 中含有 X 的信息和 X 中含有 Y 的信息相等。

④ 信息可加性:

如有 N 个自变量 X_1, X_2, \dots, X_N 和因变量 Y 之间的信息 $I_{X_1, X_2, \dots, X_N}(Y)$,

可以有: $I_{X_1, X_2, \dots, X_N}(Y) = H(Y) - H(Y/X_1, X_2, \dots, X_N)$

$$= I_{X_1}(Y) + H(Y/X_1) - H(Y/X_1, X_2, \dots, X_N)$$

$$= I_{X_1}(Y) + I_{X_1, X_2, \dots, X_N/X_1}(Y)$$

$$= I_{X_1}(Y) + I_{X_2/X_1}(Y) + I_{X_3/X_1, X_2}(Y) + \dots + I_{X_N/X_1, X_2, \dots, X_{N-1}}(Y)$$

信息的数学表达建立在申农信息论基础之上,对信息的定义和特性进行了数学描述,有助于人们从数学的视角和科学的精神来理解信息、判断信息、运用信息。

8. 信息的基本属性

(1) 普遍性和客观性。世间一切事物都在运动中,都可以看成是具有一定“生命”特征的“活物”,因为这些事物都有一定的运动状态和性状方式的改变,因而一切事物随时都在产生

^① 姜春艳,赵鹏起. 信息理论及其在分类学中的应用[J]. 佳木斯大学学报(自然科学版), 2002(04): 405-408.

信息,即信息的产生源于事物,是事物的普遍属性,是客观存在的,它可以被感知、被处理和存储、被传递和利用。

(2) 依附性和转换性。信息本身是抽象的,它们只有依附在具体的物质载体上,才能体现其客观性和传递性。由于物质载体的多样性决定了其表现形式的多样性,使信息可以从一种载体形态转换成另一种载体形态。同一信息可用多种不同的载体来表现,不同的信息也可以用同一载体来表现。

(3) 传递性与共享性。从信息的客观存在到人们对信息的主观认识,是通过信息的传递来实现的。信息可以为全人类所享用,这种共享性是相对于物质产品的交换而言的,也是信息交流与物质交流的本质区别。

(4) 认知性与创造性。信息可以被人类解读、认识、感知,不同的人对相同的信息的理解和感知程度也是有差异的,被感知和解读的信息如果是零次信息,就可能创造出一次信息;被感知和解读的信息如果是一次信息,就可能创造出新的一次信息或编辑加工成二次信息;被感知和解读的信息如果是一次信息和二次信息,就可能创造出或编辑加工成三次信息。

(5) 选择性与审美性。宇宙之大,时空广袤,物象万千;信息泛在,征兆各异,需求有别;天大,地大,人亦大;人作为宇宙的客体,信息的发生器,不过沧海一粟;人作为信息的主体,主宰万物,掌控世界;人类掌控天下,基于信息,中控信息,高于信息;信息之选择性在于其价值性,在于其真,在于其善;信息之审美性在于其真善美的和谐与韵味;信息具有选择性和审美性是人类利用信息、加工信息和创造信息的根本所在。

1.1.2 信息链相关的几个概念

信息链(Information Chain)由事实(Facts)→数据(Data)→信息(Information)→知识(Knowledge)→“情报”、“智能”(Intelligence)五个链环构成(如图 1-1-2)。在“信息链”中,“信息”的下游是面向物理属性的,上游是面向认知属性的。作为中心链环的“信息”既有物理属性也有认知属性,因此成为“信息链”的代表称谓^①。

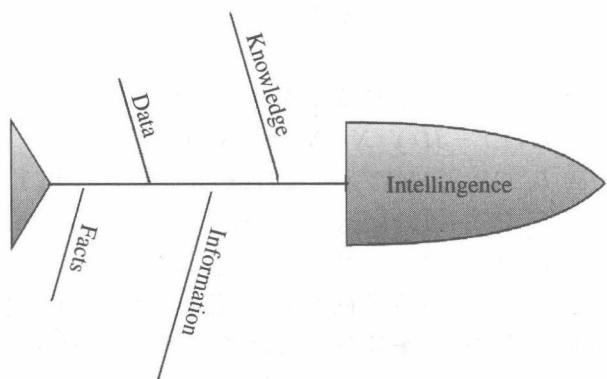


图 1-1-2 信息链示意图

① 百度百科. 信息链[EB/OL]. [2012-05-28]. <http://baike.baidu.com/view/1485174.htm>.

1. 事实(Facts)

事实是有关自然界、人类社会以及整个宇宙一切事物和现象的真实情况,包括人、事、物的本质和属性,事物的发生衍变,事件的前因后果,事态的变化及趋势等等。事实更多地侧重于客观存在的一切事物和现象以及人类社会发生的一切变革和局势情况的真实所在。事实一定层面上也是信息的真实内涵,这和信息美学所审视和最追求的信息的真具有一致性。

2. 数据(Data)

数据是指对事实的数学表示后形成的荷载特定信息的物理符号。数据本来的含义主要指科学实验、检验、统计等所获得的和用于科学研究、技术设计、查证、决策等的数值。信息链中或信息科学中的数据是指荷载或记录信息的按一定规则排列组合的物理符号。可以是数字、文字、图像,也可以是计算机代码。对信息的接收始于对数据的接收,对信息的获取只能通过对数据背景的解读。数据背景是接收者针对特定数据的信息准备,即当接收者了解物理符号序列的规律,并知道每个符号和符号组合的指向性目标或含义时,便可以获得一组数据所荷载的信息。如图书馆学领域的元数据理论,亦即数据转化为信息,可以用公式“数据+背景=信息”表示。

3. 知识(Knowledge)

知识是人们对自然、社会的现象与规律的认识和描述。从本质上说,知识属于认识的范畴。它既不能脱离人类的社会实践,也不能脱离意识和思维。人改造自然和社会的劳动,在实践中直接获得对客观世界的认识,并积累经验,最终转化为知识。意识是人类的机能,它包括认识的感性阶段和理性阶段,从感性认识上升到理性认识,也是一种知识的转化过程。思维指认识的理性阶段,是人脑对客观事物间接的概括和反映:不论是逻辑思维、形象思维还是灵感思维,都要经过一个理性认识过程,经过思考产生思想。

知识有科学知识与非科学知识之分。科学知识是创造性思维的活动的产物,是事物本质和规律性的反映。科学知识要运用科学的方法经过科学检验或科学研究获得。从知识形态看,知识有两类:一是存在于人脑中不能脱离意识的认识、经验、思想和创造性想象,称为“个人知识”;二是借助于脑外载体储存的科学知识、文艺思想及社会思潮等称为“社会知识”。个人知识与社会知识之间可以相互转化。一方面,由于人的记忆能力和有限生命的局限,个人知识经过记录,从人脑载体转移到脑外物质载体,成为社会知识;没有个人知识的转移,没有无数个个人知识的汇集,就不可能有完整的社会知识,就不能实现社会知识的积累与传递,也不能实现个人知识的代代相传与延续。另一方面,个人知识不是靠先天遗传获得的,而是通过后天的教育,通过对社会知识的学习和实践获得的。从这个意义上讲,社会知识已经包含了社会化的知识,它不是脱离社会而孤立存在的。按照知识能否清晰地表述和有效地传递,知识也可以分为显性知识(Explicit Knowledge)和隐性知识(Tacit Knowledge),二者可以互相转换。

毛泽东于1942年2月1日在中共中央党校开学典礼上的演说(见《毛泽东选集》第三卷《整顿党的作风》篇)中指出,什么是知识?自从有阶级的社会存在以来,世界上的知识只有两门,一门叫做生产斗争知识,一门叫做阶级斗争知识。自然科学、社会科学,就是这两门知识的结晶,哲学则是关于自然知识和社会知识的概括和总结。《辞海》中社会科学(Social Sci-

ences)是指以社会现象为研究对象的科学。如政治学、经济学、军事学、法学、教育学、民族学、宗教学、社会学等。它的任务是研究并阐述各种社会现象及其发展规律。人文学科(Humanities)是那些既非自然科学也非社会科学的学科的总和。一般认为人文学科构成一种独特的知识,即关于人类价值和表现的人文主义的学科;人文学科包括如下研究范畴:现代与古典语言、语言学、文学、历史学、哲学、考古学、法学、艺术史、艺术批评、艺术理论、艺术实践,以及具有人文主义内容和运用人文主义方法的其他社会科学。自然科学(Natural Sciences)是研究无机自然界和包括人的生物属性在内的有机自然界的各门科学的总称。自然科学主要涉及天文学、地球科学、生命科学、数学、物理学、化学等领域。

4. 情报(Intelligence)

情报是指被传递的知识或事实,是知识的激活,是运用一定的媒体(载体),越过空间和时间传递给特定用户,解决科研、生产中的具体问题所需要的特定知识和信息。情报应具有3个基本属性:一是知识或信息;二是要经过传递;三是要经过用户使用产生效益。情报不仅取决于情报源,也取决于情报用户。情报对于不能认识和不能理解的人来讲,只不过是一般信息而已。有人估计所谓机密情报,70%以上可以通过公开出版物获得。情报的类型主要有政治情报、经济情报、军事情报、科研情报、技术情报等;也可以分为宏观情报、中观情报、微观情报;或分为战略性情报和战术性情报。

1.1.3 信息论、控制论与系统论

1. 信息论(Information Theory)

(1) 信息论的概念。信息论是关于信息的度量和信息编码、信号处理和分析的科学理论,是运用概率论与数理统计的方法研究信息、信息熵、通信系统、数据传输、密码学、数据压缩等问题的应用数学学科,它主要是研究通讯和控制系统中普遍存在着信息传递的共同规律,以及研究最佳解决信息的获取、度量、变换、储存和传递等问题的基础理论。申农提出“信息就是不确定性的消除”论点,其主旨在于将通信的内容抽象为信息,赋予信息以数学的形式,将通信过程中的信息加工变为严格的数学运算;如果一个信源的实际信息熵小于其占用的编码空间,那么,就可以对信源参数做一种编码压缩的数学变换,使新编码所占用的空间可以压缩到接近甚至等于原始信源的信息熵;这样,就可以用较少的信道容量完成通信而不丢失信息;如信道中有干扰,则增加相应的监督码量来纠错而保证信息的无损传输;最后在接收端,还要做一种译码的数学反变换以完整地恢复原始信号^①。信息论将信息的传递作为一种统计现象来考虑,给出了估算通信信道容量的方法。信息传输和信息压缩是信息论研究中的两大领域。

(2) 信息论研究综述。申农被称为是“信息论之父”。申农在狭义信息论中从通讯角度给出信息的定义,信息是人们对事物了解的不确定性的消除或减少。消除信源发出的不确定度,定义为熵,不定性被消除的量,则用概率统计方法度量。设定信息熵 $H(x)$ 是信源整体的平均不定度,信息 $I(p)$ 是从信宿角度表示收到信息后消除不定性的程度,也就是获得的信息

^① Shannon C E. A mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27: 379 - 423, 623 - 656.