

XINXISUZHIYUXINXIJIANSUO

信息素质与信息检索

贺伟 著



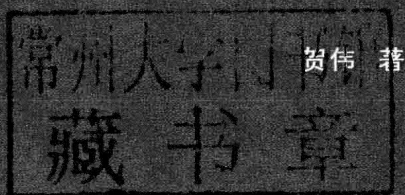
中国出版集团





现代出版社

XINXISUZHIYUXINXIJIANSUO

信息素质与信息检索



 中国出版集团
 现代出版社

图书在版编目(CIP)数据

信息素质与信息检索 / 贺伟著. — 北京: 现代出版社, 2014.5

ISBN 978-7-5143-2699-4

I. ①信… II. ①贺… III. ①情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2014)第 066176 号

信息素质与信息检索

| | |
|------|-----------------------------------|
| 作 者 | 贺 伟 |
| 责任编辑 | 陈世忠 |
| 出版发行 | 现代出版社 |
| 地 址 | 北京市安定门外安华里 504 号 |
| 邮政编码 | 100011 |
| 电 话 | 010-64267325 010-64245264(兼传真) |
| 网 址 | www.1980xd.com |
| 电子信箱 | xiandai@cnpita.com.cn |
| 印 刷 | 三河市宏顺兴印刷有限公司 |
| 开 本 | 880 × 1230 1/32 |
| 印 张 | 12.5 |
| 版 次 | 2014 年 4 月第 1 版 2014 年 4 月第 1 次印刷 |
| 书 号 | ISBN 978-7-5143-2699-4 |
| 定 价 | 36.00 元 |

版权所有,翻印必究;未经许可,不得转载

前 言

面对浩瀚的网络数字信息环境,我们必须充分地了解现代信息技术及信息检索,了解信息资源及时空分布。信息越来越成为社会领域中最活跃、最为决定性因素之一。以信息获取能力、信息利用能力和信息甄别能力为主要内涵的信息素质是信息社会中人的综合素质的重要组成部分,信息素质已成为每个社会成员的基本生存能力,更是学习型社会必备的基本素质。信息素质主要有两个层面,一是信息知识能力,它标志信息专业知识的水平;二是信息的认识和意识,它表明信息认识水平,主要包括信息收集、整序、利用和评价方面的素质。信息素质既是一种能力素质,更是一种基础素质。信息素质作为一种高级的认知技能,同批判性思维、问题解决的能力一起构成了学生进行知识创新和学会如何学习的基础。

现代社会受教育者不应是一个只会吸纳大量事实信息的人,而是一个知道如何检索、评价和应用所需信息的人。具有信息素质的人,应该“能知道何时需要信息,并拥有确定、评价和有效利用所需信息的能力,从根本意义上来说,具有信息素质的人是那些知道如何进行学习的人。他们知道如何进行学习,是因为他们知道信息是如何组织的,如何寻找信息,并如何利用信息。他们能为终生学习做好准备,因为他们总能找到为做出决策所需的信息。了解自己的信息需求;承认准确和完整的信息是指定明智决策的基础;能在信息需求的基础上系统阐述问题,具有

识别潜在信息源的能力;能指定成功的检索策略;能检索信息源,包括能利用计算机为基础的信息技术或其他技术;具有将新信息结合到现存的知识体系中的能力;能采用批判性思维利用信息并解决问题。”

本书内容结构以信息素质——信息技术——信息检索为主轴,较系统地介绍了信息素质、信息技术及信息检索的基础知识,见解新颖、内容翔实,普及性、实用性是该书的特色。同时,本书作为一本具有较强可读性、可操作性的信息检索图书,对有意了解信息检索这门学科的读者,也是一部大有裨益的好书。

本书共分12章,从信息素质教育的角度,分别对涉及信息素质教育的信息资源、信息检索工具、数字图书馆应用以及信息利用、评价等问题进行了较深入的探讨。

本书在写作过程中,参考和借鉴了许多前人研究成果,在此,向有关作者一并表示谢意。由于水平有限,书中有不足或者错误的地方,敬请广大读者提出批评和指正。

贺伟

2014年3月于怀化医学高等专科学校

目 录

| | |
|--------------------------|-------|
| 1 绪论 | (1) |
| 1.1 信息与知识 | (1) |
| 1.2 信息源 | (17) |
| 1.3 信息检索的意义 | (21) |
| 2 信息素质的培养教育 | (24) |
| 2.1.信息素质教育 | (24) |
| 2.2 信息意识 | (28) |
| 3 信息知识、能力、道德的培养教育 | (39) |
| 3.1 信息知识的培养教育概念和内涵 | (39) |
| 3.2 信息能力的教育培养 | (43) |
| 3.3 网络信息伦理道德与素养 | (68) |
| 4 信息检索的基本原理 | (83) |
| 4.1 信息检索工具 | (83) |
| 4.2 信息检索语言 | (90) |
| 4.3 信息检索与检索系统 | (96) |
| 4.4 信息检索原理 | (99) |
| 4.5 信息检索方法和步骤 | (102) |
| 5 传统文献信息检索 | (123) |
| 5.1 传统文献信息源 | (123) |
| 5.2 传统文献信息检索工具及其类型 | (128) |
| 5.3 传统文献信息检索工具编排方法 | (176) |
| 6 各种信息工具的检索 | (185) |

| | |
|-----------------------------------|-------|
| 6.1 图书检索 | (185) |
| 6.2 报刊工具书检索 | (197) |
| 6.3 国内特种文献检索 | (212) |
| 7 计算机信息检索 | (245) |
| 7.1 计算机检索概述 | (245) |
| 7.2 联机检索 | (262) |
| 7.3 光盘数据库检索 | (270) |
| 8 互联网络检索 | (303) |
| 8.1 INTERNET 概述 | (303) |
| 8.2 INTERNET 检索工具概述 | (310) |
| 8.3 网络信息资源的检索步骤 | (325) |
| 9 外文期刊全文信息检索 | (328) |
| 9.1 ELSEVIER SCIENCE 全文期刊检索 | (328) |
| 9.2 SPRINGER LINK 数据库 | (331) |
| 9.3 KLUWER 电子期刊检索 | (335) |
| 10 数字图书馆 | (341) |
| 10.1 数字图书馆及其相关概念 | (341) |
| 10.2 数字图书馆的功能及相关技术 | (355) |
| 10.3 数字图书馆的信息服务 | (358) |
| 11 信息的加工整理与分析 | (361) |
| 11.1 信息搜集与整理 | (361) |
| 11.2 信息分析的一般步骤与方法 | (371) |
| 11.3 信息分析研究成果的撰写 | (375) |
| 12 信息检索效果及其评价 | (379) |
| 12.1 信息检索效果 | (379) |
| 12.2 网络信息检索效果评价 | (385) |
| 参考文献 | (391) |

1 绪论

1.1 信息与知识

1.1.1 信息

自人类社会进入 21 世纪以来,信息成为了最常用的语汇,信息的价值日益被重视,信息、材料、能源被看作是维持社会发展的三个基本要素。随着科学技术的发展和社会的进步,“信息经济”、“知识经济”也成为越来越热的话题。因此有必要对信息与知识的本质及其价值进行一些初步的探讨,以期达到科学利用的目的。

1.关于信息的基本概念

信息一词的基本含义是指消息、音讯,在信息论中,信息是指用符号传送的报道,报道的内容是接受符号者预先不知道的。随着时间的推移,信息的基本内涵在不断变化,人们有关信息的概念也在不断的发展变化。同时,由于信息一词具有广泛的社会意义和很大的包容性,人们对它的理解和解释也不尽相同。迄今为止,人们对信息的定义尚无一个统一的结论。为了最大限度地概括信息的本质含义,学术界普遍主张从哲学的角度去认识信息,认为“信息是物质存在方式和运动状态所蕴含的间接存在物的标志”;“信息是对客观世界中各种事物的变化和特征的反映”;“信息是物质的一种属性,是物质存在的方式和运动的规律与特点”;“信息是反映出来的物质属性”等。尽管关于信息定义的表达

述不尽相同,但大多认为信息是客观世界的具体存在,是客观事物特性的表征,是各种事物的变化规律及其特征的反映。从这个意义来说信息的存在历史无疑是久远的,尤其是一些自然信息,其存在历史甚至远远早于人类社会的历史。尽管信息产生和存在的历史久远,但它作为一个科学概念被使用,则是近代以来的事情,而且人们更多的是从学习、工作与生活的实际需要出发,去感知信息、解释信息、界定信息。信息作为一种可供利用的资源,一般来说具有客观性、相对性、依附性、传递性、可加工性、共享性和时效性等特征。

2.信息的本质

对信息的直观而又通俗的理解就是“消息”,它总是伴随着一种传递的过程。在信息论中,发出信息的一方称为信源,接受信息的一方称为信宿。信息从信源到信宿的传递必然借助于某种物理手段,称之为信道。信源产生某种物理量的变化,称之为信号,所要传递的信息就包含在这种信号中。通过信道,信宿感受到一种变化的物理量,就可以按照既定的规则破译出其中包含的信息。

信号本质上是一种变化的物理量,它反映的是信源本身的一种运动,一个静止不动的物体是没有资格作为信源的。从另一个角度来看,基于事物之间普遍联系的原则,宇宙中并不存在孤立的物体,因此任何运动的物体都会发出信号,这也是客观世界的可认识性的根源。

然而,信号并不等同于信息。信息论的奠基人香农(Shannon)把信息量定义为所消除的不确定度的大小,控制论的创始人维纳(Wiener)则提出信息量的基本单位就是“对具有相等概率的二中择一事件作选择时所传递出去的信息”。这个单位也就是现在常用的比特(bit),例如,“去工作”和“去休息”这两个行为是二种择一行为,假定二者发生的概率相同,那么当我们被告

知“去工作”时,我们就获得了 1bit 的信息。如果我们已经确信“明天下雨”这件事,那么即使再被告知关于明天是否下雨的任何信息,所获得的信息量也只能是 0。

可以看出,接收者从信号中获得的信息量的大小不仅仅是由信息的发出者——信源所决定的,而且与信息的接收者对相关事件的确定程度也有关。信息的接收者对信号所描述的事件越确定,能够从中获得的信息量也就越少。

香农所定义的信息量只是基于信号本身的一种统计属性,即信号中各个符号的出现概率和它们之间的相关程度,与信息的接收者没有任何关系。然而不要忘记,这里的出现概率和相关程度本身就是对即将到来的信号的一种认识,而非信号接收完毕后再对其进行统计的结果。这种认识只能是来源于接收者本身的所拥有的知识和智慧。例如,将一枚硬币抛掷出去,可能得到正面,也可能得到反面。由于在我们的知识中已经知道正反面出现的概率是相等的,所以每得到一个结果,我们就获得了 1bit 的信息量。实际操作中,常常将已经接收到的信号的统计结果应用到全部信号上,这是一种从局部到整体的归纳推理,因而不能保证是完全正确的。例如,在莫尔斯(Morse)电报码中,s 和 e 等字母的编码最短,这是因为它们在现有的英文典籍中出现的频率最高,但显然不能认为所有的电文都是如此。

知识掌握的多少对不同的信息接收者来说,是有差异的,因为从一段信号中得到信息量,不仅要求了解信号本身统计属性,而且要理解信号本身的含义,同样一个符号 0 或者 1,它可能代表“去工作”和“去休息”,也可能代表“禁止”和“允许”等,如果信息接收者对此一无所知,则会陷入无所适从的境地,也就得不到任何信息量。另外的例子是信息的加密,接收方如果无法进行有效的解密也就不可能获得足够的信息。

然而,信息论所描述的信息量确实是信号本身的一种量度,

而不在于它是怎样被接收的。从一段相同的信号中,一个拙劣的接收者可能获得比一个高明的接收者少得多的信息量。信息论认为信号包含的信息量是一个定值,接收者之所以不能完全获得其中所包含的信息乃是由于它没有足够的鉴别其中所包含的信息。事实上,这种定义就是假定了存在一个“超级接收者”,它比任何其他接收者都要高明。它能够完美的从一段信号中破译出所有包含的信息。没有一个接收者能够从同样一段信号中得到比“超级接收者”还多的信息。在这种定义下,信号所包含的信息量就是它被“超级接收者”所能识别的信息量。

从这种定义出发,不难得出,任何信息的处理和再传输都只能减少信息,而不能使信息有所增加,因为凡是经过处理或再传输获得的信息,“超级接收者”自然也能够得到,因而这些信息应被视为包含于原信息本身中。

香农的信息量定义对狭义的信号概念从数学上给出了一个可计算的性质,具有很高的理论价值,由此衍生的整个信息论在通信、计算机等领域发挥着深远的影响。但需要指出的是,该定义针对的是随机变量产生的信号,对于有一定规律的信号是不完全适用的。因此,Kolmogorov 提出了算法信息量的定义,通俗地讲,一段信号所包含的信息量的大小由产生这段信号的最短的计算机程序的长度来决定,当然,这里的计算机也不是任意一台现实中的计算机,而应该是一台“理想”的图灵(Turing)机。这种定义虽然使信息量变得难以计算,但却更加揭示了信息的本质含义。如果一段长长的信号能够由一个很短的程序产生,那么它所包含的信息量就不会比这个很短的程序更多。最好的例子是计算机软件中常用的伪随机码,每个伪随机码都是从一个被称为种子的编码通过固定的算法计算出来的,而且通常当前产生的伪随机码就作为下一个伪随机码的种子。因此,只要知道了第一个种子和产生伪随机码的算法,以后产生的所有伪随机

码序列也就确定了, 尽管这些序列中 1 和 0 出现的概率近乎相等, 也不会再提供 1bit 的信息量。从这个意义出发, 所有的演绎推理都不能产生额外的信息量, 因为推理的结果是从原有的条件——原始信息经过逻辑推理这种固有的“程序”得到的, 推理结果提供的信息本来就是包含在原始信息中的。

以上只是从“消息”、“信号”的角度对信息所作的理解。在现代科学中, 信息一词的含义早已超越了这种直观的理解。早在信息论诞生之前, 玻尔兹曼(Boltzmann)在统计物理中就提出了一个用微观粒子状态数表示的熵的公式, 这一公式与香农的信息定义形式上只差一个负号。事实上, 香农在其 1948 年发表的信息论奠基性论文《通信的数学理论》中就是用“熵”这个词来称呼本文中所说的信息。这种相似性绝不是偶然的。信息即是负熵, 得到信息, 则自身的熵减少; 失去信息, 则自身的熵增加。当信息与熵联系起来的时候, 它就成为了一个类似于质量和能量的可以度量的实实在在的物质属性, 亦即一种客观存在。

这种概念在耗散结构的理论中得到了应用。根据热力学第二定律, 任何孤立系统的熵都是随时间增加的, 这样最终会达到一个均匀的“热寂”状态。我们知道, 熵是度量系统无序程度的物理量。这就是说, 一个孤立的系统最终会成为一个完全无序的系统。而耗散结构不是孤立系统, 它能从周围环境中有选择的吸收信息, 从而减少自身的熵, 向着越来越有序的方向发展。生物体就是一个耗散结构, 因此能够不断的成长直至衰亡。

人类对自然界的认识过程就是一个不断的接收来自客观世界的信息的过程。从信息论的角度来看, 物体发出的原始信号包含的信息量无疑是最丰富的。任何观测手段如显微镜、望远镜等都相当于对原始信号的一种处理, 因而不可能增加关于被观测物体的更多的信息量, 相反通常还会对原始的信息有所损坏。那为什么人类还要使用各种各样的观测工具呢? 这是因为, 人的肉

眼是一个相当拙劣的信息接收者，通过有效的工具从受损的信号中得到的信息量往往比直接用肉眼从原始信号中得到的信息量还要丰富。各种各样越来越先进的观测工具，成为人自身感官的一种外延，使人类能够从自然界中挖掘到越来越多的信息，有力地促进了人类对客观世界的认识。从信息的更本质的意义上看来，人类乃至人类社会不断的趋于有序化，向更高的组织形态发展，皆是因为从周围环境中不断摄取信息的缘故。

3.信息的类型、特点与功能

(1)信息的类型

信息上作用于社会生活的每一个领域的。信息活动是人们进行一切社会活动时必然伴随的活动。它既纵贯人类社会发展的整个过程，又在每一个历史时期内横跨当时社会生活的每一个领域。因此，信息是复杂多样、丰富多彩的。

①按信息的内容特点分为两种：事实信息和分析信息。

事实信息。事实信息是表达客观现实的信息。这类信息提供确切的事实或数据，回答这样一些问题，如：长江的源头是在什么地方？全国年生产总值有多少？等等。回答这些问题通常只要一个简短的答案，无须说明。这些信息常出现在辞典、百科全书、年鉴、手册、名录等工具书中。

分析信息。分析信息是指对事物、事件等进行说明、分析、解释的信息。分析信息研究行为、思想和事件之间的相互关系、内涵以及背景、原因等。在研究分析某一问题时，往往需要多个信息来源。分析信息通常可在图书、期刊等中找到。

②按信息的表达形式分四种：文字信息、声像信息、实物信息、机读信息。

文字信息。用文字来表达其内容的信息资料，如各种图书、期刊、报纸等。

声像信息。通过声频信号负载与传递的信息，如各种广播、

电视、电影、图片等。

实物信息。通过实物来传递的信息,如产品样品、历史遗迹等。

机读信息。通过计算机阅读的信息,如计算机磁带、磁盘、光盘上存储的信息等。

③按信息的应用范围分:白色信息、灰色信息、黑色信息。

白色信息。公开出版发行、流通和传递,并人人均可使用的信息,如公开发行的图书、报纸、光盘、数据库等。

灰色信息。非公开出版发行、流通和传递,从正常途径难以获得的信息,如内部资料、技术报告、会议资料等。此类信息出版量少、发行渠道复杂、流通范围所限,不易收集。

黑色信息。指人们未破译或未被识别的信息,也指处于保密状态的信息,如考古发现的古代文字、未解密的政府文件、内部档案、个人日记等。

④按信息的载体形式分:印刷型、缩微型、视听型、电子型。

载体是信息存在的物质形态。随着信息记录和存取技术的发展,信息的物质形态呈现多样化。

印刷型。印刷型亦称纸介质型,它是以纸张为载体,以手写、石印、刻印、胶印和影印等为手段形成的文献。这是传统文献的基本形式,其优点是方便阅读,流传广泛,价格便宜;缺点是存储信息密度低,体积大。

缩微型。缩微型是以感光材料为载体,以缩微照相为手段形成的文献,即利用光学技术将文字、图形、影像等信息按比例缩小产生的文献。目前最常见的有缩微胶卷和缩微胶片两种产品。用户可通过缩微阅读机阅读,或通过缩微阅读复印机阅读,也可将其复印成纸质文献。这类文献的优点是信息存储密度高,体积小,容易保存和流通;缺点是阅读时必须借助缩微阅读机等设备,保存条件高,不易普及。

视听型。视听型亦称声像型,是以感光材料或磁性材料为载体,通过特定的设备,使用声、光、磁、电等技术记录声音和图像,并通过视听设备存储和播放信息的文献形式,如录像带、录音带、唱片、影视片等。其优点是图文声像并茂,直观,易于接受;缺点是收看、收听必须借助一定的设备。

电子型。电子型亦称机读型,是以磁性材料为存储介质,以打字、穿孔或光学字符识别装置为记录手段,通过编码和程序设计,把信息输入计算机,存储在磁盘、磁带和光盘上,转换成数字语言和机器语言,阅读时再通过计算机将其内容按要求输出的文献形式。这类电子出版物包括电子图书、电子期刊、计算机文档、联机信息库和光盘数据库等。其优点是信息存储密度高,存取速度快,信息量大,交互性强,信息更新快,保存时间长;缺点是阅读时必须借助计算机等设备及较好的网络条件。

⑤按信息加工的深度分:零次信息、一次信息、二次信息、三次信息。

零次信息。零次信息是指未经记录、未公开、无法通过载体在较大范围内传播的信息,主要以口语或实物方式进行传递,直接作用于人的感觉器官,信息的载体是人脑或实物本身。它包括语言信息和实物信息,如会议口头交流、私人信件、未发表的演讲稿等。前者具有方便及时、针对性强、反馈快等特点;后者的特点是真实、直观。其不足是难于管理和积累。零次信息是人们获取知识信息不可忽视的信息源,不仅在内容上具有一定的价值,而且可有效弥补公开信息传播耗时的不足。

一次信息。一次信息是指第一次报道、第一次书写、第一次出版的信息。一次信息是人们在社会实践和科学研究中根据科研成果、发明创造而撰写的文献。一次信息的文献形式是一次文献,也称第一手资料。它的载体形式有图书、期刊论文、会议论文、科技报告、专利文献、政府出版物、标准文献和学位论文等

一次信息记录的信息内容具体、详尽,具有新颖性、创造性、系统性等特点,有直接参考、借鉴和使用的价值,是信息检索和利用的主要对象。其不足是出版分散,缺乏系统性,未经过科学组织,呈无序状态而难以被系统获得和掌握,要寻找有关信息往往需要借助二次信息来完成。

二次信息。二次信息是通过分散、无序的一次信息进行加工、整理、浓缩而形成的信息。它以特定的方法汇集某一范围内的信息,用科学的方法进行加工整理,以简练的语言,以不同的深度揭示一次信息的内外特征,并提供多途径检索,将分散、无序的大量一次信息转变为有序、便于管理的系统,从而有利于人们有效利用一次信息,成为查找一次信息的工具,如各种目录、文摘、题录、索引等。二次信息具有浓缩性、汇集性、检索性等特点,有对一次信息进行报道和指引的作用。

三次信息。三次信息是在大量利用一次信息和二次信息的基础上进行汇集、综合、分析而形成的信息。人们根据某种需要和目的,利用二次信息为手段,将大量的一次信息加以全面系统的再度选择、分析和综合,编写成为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。它的文献形式是三次文献,可分为综述研究类和参考工具类两种类型。前者包括综述、述评、进展报告等;后者包括年鉴、百科全书、手册、指南,以及书目之书目、文献指南等。三次信息具有综合性高、针对性强、系统性好、信息面广等特点,有较高的实际使用价值,能直接参考、借鉴和利用。

综上所述,从零次信息到一次信息、二次信息、三次信息,是一个从分散到集中,从无序到有序,对信息进行不同层次加工的过程。零次信息是一次信息的素材,一次信息是二次信息、三次信息的来源和基础,二次信息、三次信息是对一次信息进行整理、组织、加工、综合而形成的。

(2)信息的特点:

客观性、时效性、相对性、依存性、可传递性、可加工性、共享性。

客观性。指信息是客观存在的,不以人们的意志为转移的。它是现实世界中各种事物运动与状态的反映,它可以被人们所感知、处理、存储和使用。它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。

时效性。客观事物处于不停的运动变化之中,反映客观事物的信息也在不断发展更新。因此,获取和利用信息时必须树立时效观念。

相对性。客观上信息是无限的,但对信息用户来说,实际可获得的信息总是有限的。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也各不相同,数量和质量上都会有所差异。

依存性。即信息本身必须依附于一定的物质形式(如纸张、声波、电波、磁性材料等),不可能脱离物质载体而独立存在。没有语言、文字、图像、符号等记录手段信息便不能表述,没有物质载体信息便不能存储和传播,但其内容不会因记录手段和依附载体的变化而改变。

可传递性。任何信息只有经过从信源出发,经过信息载体才能被信宿接收并进行处理和运用。信息传递跨越时空,一个完整的信息传递过程包括信源、信道、信宿和信息四个基本要素。

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。

共享性。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。一条信息被用户吸收和利用并不影响信息的本身和被其他用户的反复使用,各用户分享的份额也不会