

高等学校教材

# 信息技术基础

主编 鄂大伟 庄鸿棉

参编(以姓氏笔划为序)

王兆明 王家聚 叶文来 庄鸿棉

张仪华 郑嘉琳 贾红伟 鄂大伟

黄凯明 谢绵陞

高等教育出版社

## 内 容 提 要

本书的编写是对教育部确定的高等学校计算机基础教学三个层次教学体系要求的计算机文化基础层次的探讨,教学内容定位在计算机文化基础教学的层次上。

本书以信息的获取、传输、存储、处理、加工、发布和应用为主线,全面介绍信息技术涉及的各种技术——微电子技术、通信技术、计算机技术和网络技术。主要包括:信息与信息技术,信息源与信息采集方法,信息处理工具——计算机系统,信息的存储与压缩,信息数据结构、组织和管理,信息的传输,网络与通信技术,信息的检索与利用,信息系统的设计、开发与应用,信息的发布——多媒体与网络,信息安全与管理。

本书的一个显著特点是,全面介绍信息技术的相关概念与知识,内容丰富,图文并茂,知识性和可读性较强,适合高等学校非计算机专业学生使用,对于希望了解信息技术的人员来说,本书也是一本较好的参考书。

本书还有配套的上机指导用书、光盘课件和授课演示文档。

# 序 言

自文明肇始以来,人类就生活在信息的海洋中。人类社会的生存和发展,时刻都离不开接收信息、传递信息、存储信息和利用信息。原始人的“结绳记事”,古人的“烽火报警”无疑是信息的表示、存储和传送的方法之一。如果说文字的出现、印刷术的发明,是人类文明发展进程中的两个重要里程碑,那么,以计算机技术为代表的信息技术的发展和普及,将成为人类文明发展史上的第三个里程碑。

以计算机技术、网络与通信技术和微电子技术为代表的现代信息技术,正在改变人们传统的生活、学习和工作方式,同时也影响着教育的内容与方法。信息技术教育成为全世界教育课程改革的热点研究课题。作为序言,这里想谈谈编写本书的动机与目的。

信息技术教育是指学习、利用信息技术,培养信息素养,促进学与教优化的理论与实践。信息技术教育的本质是利用信息技术培养信息素养。信息素养是指人所具有的对信息进行获取、识别、加工、利用、评价和管理的知识、能力等各方面基本品质的总和。

为了迎接 21 世纪的挑战,我国已决定中小学在今后几年里基本普及信息技术教育,国家教育部门已经将信息技术课纳入中小学必修课程,把学生培养成有信息素质的终身学习者。这是我国面向 21 世纪国际竞争、提高全民素质、培养具有信息素质和创新素质的新型人才的一项重要举措。随着中小学信息技术教育的实施,中小學生将掌握计算机和网络的基本知识技能,学生的信息素养将会大幅度的提高。目前,经过基本信息技术教育的高中毕业生将陆续升入大学,他们对计算机基本操作技能的掌握程度会越来越好,这是不争的事实。

从学生受教育的系统性和整体性来看,大学中的信息技术教育应充分考虑学生在中小学时期所打下的信息技术的基础。在规划大学信息技术教育的课程时,应首先了解中小学信息技术课程的开设情况,并结合学生对信息技术教育内容的掌握程度,有的放矢地进行教学,以避免教育的重复、资源的浪费。使那些在中小学接受了信息技术教育的学生,在步入大学殿堂后,能够受到更高层次的、与培养目标相适应的信息技术教育。

在由原先大学作为计算机教育的起点开始过渡到以中小学作为普及信息技术教育的起点的背景下,大学信息技术教育的内容是什么,如何体现大学信息技术教育的特点,如何与中小学衔接,如何紧跟迅速发展的信息技术,构建适合我国国情的大学信息技术课程与教材体系;如何进行信息技术教育与其他学科教育的整合以及课堂教学和实践环节的整合,这些都是高校计算机基础教学课程改革需要认真思考的问题。

当然,对以上问题的认识,人们有着不同的观点。我们以为:大学信息技术教育其实质是对信息技术教育根本目的的认识,对学习目标的理解,对课程性质的定位。信息技术教育的根本目的是培养创新人才,不是培养机器的操作者。从更深层的意义上讲,信息技术教育是一种素质教育,不是以某种技能掌握为目标的技能培训。不能以计算机基础知识与操作技能的学习与掌握来替代信息技术的学习与掌握,只能将计算机作为信息获取、分析、处理的工具进行学习。对计算机技术的学习应从信息技术的高度、信息技术的角度进行学习,而不是脱离这个前提进行学习。

基于以上认识,我们早在两年前就开始进行大学信息技术教学改革的尝试。当时的初衷

是准备给大学生开一门类似于“信息技术基础”的课程。可是寻遍书店也未找到合适的教材。因为目前关于大学计算机文化基础的教材比比皆是,而简明扼要介绍信息技术的教材甚鲜,即使有些挂以“信息技术”之名的大学教材仍未脱离“计算机文化基础”的窠臼。这才使我们萌生了编写一本全面介绍信息技术的相关概念与知识,内容丰富,可读性强的教材,通过教材与内容的改革来贯彻我们对大学信息技术教学改革的思想与理念。我们从2001年上半年起开始策划,编制讲义,并经过在教学中的试用,再经修改,历时两年多。现在,本书终于与读者见面了,我们感到十分欣慰。

本书冠以“信息技术基础”之名,意即导明要义,窥其整体,论述梗概,阐其原理,而不涉及所述对象使用或操作的繁文缛节。本教材内容以信息的获取、存储、处理、传输、组织、管理、发布和利用为主线,从当代信息科学的最新成果中汲取知识,力求与信息技术的相关学科相互融合,使课程内容超越原“计算机文化基础”的局限,最大限度地回归和体现“信息技术”的“整体”面目,成为培养大学生信息能力、信息素养和创新意识的载体。这正是我们所期待的。

在本书的编写过程中,我们力求能在结构和内容上有所创新,并有较鲜明的特色。一是取材新颖,许多内容取自于国内外最新的资料和 Internet 资源,以反应信息技术当前的发展;其次,为使学生了解信息技术的全貌,开拓视野,对在信息技术发展史上出现的重要事件、人物或概念以提示方式介绍,旨在增加知识性;第三,全书包含的信息量较大,深浅程度不一,目的是让具有不同需求的读者都有收益。在教学中可根据教学对象的专业背景和需求对内容加以调整,使“深者得其深,浅者得其浅”。

这套教材的另一个特色是克服了传统书本媒体的局限性,提供了供教师教学使用的电子教案和供学生自学使用的网络版课程。网络课程可以在教师教学计划的指导之下,实现任何时间、任何地点的学习。网络课程的设计是对教材内容的进一步补充和说明。学习者还可以通过网络教材提供的网络链接查阅到众多相关的学习资源,实现现实世界与课堂教学之间的交互功能。

在信息化社会的今天,信息素养已成为科学素养的重要构成部分。迅速地筛选和获取信息、准确地鉴别信息、创造性地加工和处理信息,将是所有社会成员应具备的、终生有用的基础能力之一。在本课程的学习中,我们希望能够营造一种信息素质教育的环境。学生应能关注学习过程,积极参与自主的学习活动,提出与自己学过的学科内容有关的问题,使用环境、资源、工具等进行学习,从而实现教学相长。“博学而笃志,切问而近思”,学则固矣。

参加本教材编写的教师有(按章节顺序排列):庄鸿樾(第1章)、王兆明(第2章)、鄂大伟(第3、4章)、谢绵陞(第5章)、黄凯明(第6章)、王家聚(第7章)、张仪华(第8章)、叶文来(第9章)、郑嘉琳、贾红伟(第10章)。全书由鄂大伟教授和庄鸿棉教授策划和统稿。

在本书的编写过程中,得到许多兄弟院校老师的积极关注,并提出了许多有价值的意见。高等教育出版社对我们的教学改革与探索给予了大力支持。此外,在教材编写过程中摘录或参考了许多著作和国内外网站内容,在此一并深致谢忱。

在结束这篇似乎有些冗长的序言之前,最后我们想说的是:信息技术的发展一日千里,相关的每个学科都变得越来越精微和深奥。囿于作者的水平及篇幅所限,俾本书内容难以准确反映和把握信息的整体与全貌,疏漏、欠妥、悖谬之处,恳请读者指正。以期冀收博见,嘉惠来学。

作 者

2003年4月

# 目 录

第 1 章 信息与信息技术 .....	(1)	本章小结 .....	(30)
1.1 探索信息的真谛 .....	(1)	思考与练习 .....	(30)
1.1.1 什么是信息 .....	(1)	参考文献 .....	(31)
1.1.2 从信息论到信息科学 .....	(3)	第 2 章 信息源与信息采集方法 .....	(33)
1.1.3 香农对信息的定义 .....	(3)	2.1 自然界的三种资源 .....	(33)
1.1.4 信息的度量 .....	(4)	2.1.1 物质、能量和信息之间的关系 ...	(34)
1.1.5 香农信息论的局限性 .....	(6)	2.1.2 信息的来源 .....	(35)
1.1.6 数据、消息、信号与信息 .....	(7)	2.1.3 数据与信息的再认识 .....	(36)
1.1.7 信息的基本特性 .....	(7)	2.2 基于人工系统的信息采集方法 .....	(38)
1.1.8 信息的基本作用 .....	(9)	2.2.1 直接观察法 .....	(39)
1.2 信息科学与信息技术 .....	(11)	2.2.2 社会调查法的调查对象范围 和手段 .....	(39)
1.2.1 科学的定义 .....	(11)	2.2.3 查阅资料 .....	(40)
1.2.2 信息科学 .....	(11)	2.3 基于计算机系统的信息采集方法 .....	(41)
1.2.3 信息技术 .....	(12)	2.3.1 信息的数字化过程 .....	(41)
1.2.4 信息技术的核心 .....	(12)	2.3.2 语音信息的采集与数字化 .....	(42)
1.2.5 信息技术的发展历史 .....	(15)	2.4 图像信息的采集与处理 .....	(46)
1.3 信息化与信息社会 .....	(15)	2.4.1 色彩与色彩模型 .....	(46)
1.3.1 信息化社会的标志 .....	(15)	2.4.2 图像的数字化处理过程 .....	(49)
1.3.2 构建信息化社会 的信息高速公路 .....	(16)	2.4.3 图形和图像的概念区别 .....	(51)
1.3.3 知识经济成为信息化社会 主要标志之一 .....	(18)	2.5 视频信息的采集及处理 .....	(54)
1.3.4 社会信息化的基础 是企业信息化 .....	(18)	2.5.1 视频基础 .....	(54)
1.4 信息产业与信息人才 .....	(19)	2.5.2 YUV 颜色空间 .....	(55)
1.4.1 信息产业的概念 .....	(19)	2.5.3 视频图像的数字化处理过程 .....	(56)
1.4.2 信息产业的结构 .....	(20)	2.5.4 视频卡的组成及其主要功能 .....	(57)
1.4.3 信息产业的特点 .....	(23)	2.5.5 视频压缩技术 .....	(58)
1.4.4 信息产业对人才的需求 .....	(25)	本章小结 .....	(58)
1.4.5 信息人才的职业划分 与人才培养 .....	(26)	思考与练习 .....	(59)
1.5 关于大学信息技术教育的思考 .....	(27)	参考文献 .....	(59)
1.5.1 信息能力与信息素养 .....	(27)	第 3 章 信息处理工具 ——计算机系统 .....	(61)
1.5.2 信息技术教学的目标 .....	(28)	3.1 计算机的过去、现在与未来 .....	(61)
1.5.3 大学信息技术教育 包括的内容 .....	(28)	3.1.1 计算模型与图灵机 .....	(61)
1.5.4 对信息技术课程性质的认识 .....	(29)	3.1.2 存储程序式计算机的基本结构 与工作原理 .....	(62)
		3.1.3 现代计算机的“史前” 时代( - 1946) .....	(63)

3.1.4	第一台现代电子数字计算机的诞生 .....	(64)	4.3.2	图像数据压缩的可能性 .....	(106)
3.1.5	现代计算机发展的四个阶段 .....	(65)	4.3.3	压缩编码方法的分类与评价 ...	(107)
3.1.6	计算机有智能吗? ——关于第五代计算机 .....	(68)	4.3.4	数据压缩编码技术的发展 .....	(107)
3.1.7	巨型计算机与矢量计算 .....	(70)	4.3.5	图像压缩编码方法 .....	(108)
3.2	计算机系统 .....	(71)	4.4	静态图像压缩标准——JPEG .....	(112)
3.2.1	计算机系统的组成 .....	(71)	4.4.1	离散余弦变换 .....	(112)
3.2.2	微型计算机系统 .....	(72)	4.4.2	量化 .....	(114)
3.2.3	微型计算机的总线及标准 .....	(72)	4.4.3	编码 .....	(115)
3.2.4	计算机与外部设备的接口及标准 .....	(73)	4.4.4	新一代静态图像压缩标准——JPEG2000 .....	(115)
3.3	微型计算机的硬件系统 .....	(74)	4.5	视频图像压缩标准 MPEG .....	(116)
3.3.1	主机 .....	(74)	4.5.1	什么是视频 .....	(116)
3.3.2	外存储器 .....	(76)	4.5.2	MPEG 家族 .....	(117)
3.3.3	输入设备 .....	(78)	4.5.3	MPEG 压缩过程 .....	(119)
3.3.4	输出设备 .....	(81)	本章小结 .....	(122)	
3.4	计算机软件系统 .....	(82)	思考与练习 .....	(124)	
3.4.1	软件的定义及其性质 .....	(82)	参考文献 .....	(125)	
3.4.2	软件危机与软件工程 .....	(86)	第 5 章 信息数据的结构、 组织和管理 .....	(126)	
3.4.3	软件技术的历史回顾 与发展 .....	(88)	5.1	为什么要进行数据组织 .....	(126)
3.5	程序设计方法 .....	(91)	5.2	数据结构 .....	(127)
3.5.1	计算机程序的性质 .....	(91)	5.2.1	数据结构要解决什么问题 .....	(127)
3.5.2	什么是算法 .....	(91)	5.2.2	线性表 .....	(130)
3.5.3	程序、数据结构、算法及其关系 .....	(92)	5.2.3	树形结构(层次结构) .....	(132)
3.5.4	结构化程序设计 .....	(93)	5.2.4	图 .....	(134)
3.5.5	程序设计的步骤 .....	(94)	5.3	关系数据库 .....	(136)
3.5.6	从面向过程(OP)到面向对象(OO) .....	(95)	5.3.1	数据处理技术的发展 .....	(136)
3.5.7	第四代语言 .....	(95)	5.3.2	要会使用数据库,我们应掌握些什么 .....	(137)
本章小结 .....	(96)		5.3.3	数据模型 .....	(138)
思考与练习 .....	(97)		5.3.4	关系数据库标准语言——SQL .....	(144)
参考文献 .....	(97)		5.4	数据库设计 .....	(148)
第 4 章 信息的存储与压缩 .....	(99)		5.4.1	数据库设计的全过程 .....	(148)
4.1	信息存储技术的三次革命 .....	(99)	5.4.2	关系数据库的模式设计理论 ...	(150)
4.2	计算机存储系统 .....	(100)	5.5	数据库的保护 .....	(152)
4.2.1	信息存储技术的金字塔结构 ...	(100)	5.6	数据库技术的新发展 .....	(154)
4.2.2	磁存储技术 .....	(101)	5.6.1	数据仓库 .....	(154)
4.2.3	光存储技术 .....	(102)	5.6.2	空间数据库(Spatial Data Base) ...	(156)
4.3	图像数据压缩与编码技术 .....	(105)	5.6.3	工程数据库 .....	(156)
4.3.1	熵与信息压缩 .....	(105)	本章小结 .....	(157)	
			思考与练习 .....	(157)	
			参考文献 .....	(158)	

第 6 章 信息的传输 网络 与通信技术 .....	( 159 )	8.1.2 系统模型及其建模 .....	( 235 )
6.1 通信系统模型 .....	( 159 )	8.2 信息系统的分类 .....	( 240 )
6.1.1 传输介质及其工作原理 .....	( 160 )	8.2.1 事务处理系统 .....	( 240 )
6.1.2 传输速度 .....	( 167 )	8.2.2 管理信息系统 .....	( 241 )
6.1.3 传输模式 .....	( 168 )	8.2.3 决策支持系统 .....	( 241 )
6.1.4 基带传输与宽带传输 .....	( 170 )	8.2.4 人工智能与专家系统 .....	( 241 )
6.2 计算机网络 .....	( 171 )	8.3 信息系统的开发 .....	( 242 )
6.2.1 计算机网络的分类 .....	( 172 )	8.3.1 信息系统开发的一般过程 .....	( 243 )
6.2.2 计算机网络的结构组成 .....	( 175 )	8.3.2 生命周期法 .....	( 243 )
6.2.3 通信协议 .....	( 176 )	8.3.3 原型法 .....	( 244 )
6.2.4 网络服务 .....	( 184 )	8.3.4 面向对象的开发方法 .....	( 245 )
6.2.5 网络硬件 .....	( 191 )	8.3.5 CASE 方法与工具 .....	( 250 )
6.2.6 宽带网络实现技术 .....	( 198 )	8.4 电子商务 .....	( 251 )
本章小结 .....	( 202 )	8.5 EDI、MRP II 与 ERP .....	( 253 )
思考与练习 .....	( 203 )	8.5.1 EDI .....	( 253 )
参考文献 .....	( 203 )	8.5.2 MRP II .....	( 254 )
第 7 章 信息的检索与利用 .....	( 205 )	8.5.3 ERP .....	( 254 )
7.1 信息检索的原理及一般步骤 .....	( 205 )	本章小结 .....	( 255 )
7.1.1 信息检索 .....	( 205 )	思考与练习 .....	( 255 )
7.1.2 信息检索的途径 .....	( 206 )	参考文献 .....	( 256 )
7.1.3 信息检索的一般步骤 .....	( 207 )	第 9 章 信息的发布 .....	( 257 )
7.2 信息检索的方法 .....	( 209 )	9.1 信息发布的演变 .....	( 257 )
7.2.1 手工信息检索 .....	( 209 )	9.1.1 多媒体 .....	( 257 )
7.2.2 计算机信息检索 .....	( 209 )	9.1.2 多媒体应用系统 .....	( 258 )
7.3 信息检索工具介绍 .....	( 215 )	9.1.3 超文本( Hypertext ) .....	( 260 )
7.3.1 国外著名检索工具 .....	( 215 )	9.1.4 超链接( Hyperlink ) .....	( 261 )
7.3.2 网络信息检索工具 .....	( 218 )	9.1.5 超媒体( Hypermedia ) .....	( 261 )
7.4 信息检索的运用 .....	( 222 )	9.1.6 多媒体与因特网 .....	( 263 )
7.4.1 中文搜索引擎的使用 .....	( 222 )	9.2 因特网 .....	( 263 )
7.4.2 常用搜索引擎介绍 .....	( 223 )	9.2.1 因特网的过去和现在 .....	( 264 )
7.4.3 网络信息挖掘 .....	( 225 )	9.2.2 因特网的工作原理 .....	( 265 )
7.5 信息应用 .....	( 227 )	9.2.3 因特网在线服务 .....	( 269 )
7.5.1 网络环境下的信息资源利用 .....	( 228 )	9.2.4 新兴的网络服务 .....	( 272 )
7.5.2 数字图书馆 .....	( 229 )	9.2.5 WWW 浏览器 .....	( 275 )
本章小结 .....	( 230 )	9.3 网络信息发布技术 .....	( 276 )
思考与练习 .....	( 231 )	9.3.1 标记语言的演化 .....	( 276 )
参考文献 .....	( 231 )	9.3.2 HTML .....	( 277 )
第 8 章 信息系统的设计、 开发与应用 .....	( 233 )	9.3.3 网页制作三剑客 .....	( 279 )
8.1 信息系统概述 .....	( 233 )	9.3.4 Web 服务器 .....	( 282 )
8.1.1 信息系统的定义与结构 .....	( 234 )	9.3.5 网页发布 .....	( 283 )
		本章小结 .....	( 284 )
		思考与练习 .....	( 284 )
		参考文献 .....	( 285 )

第 10 章 信息安全与管理 .....	( 286 )	10.2.5 数字签名 .....	( 306 )
10.1 计算机犯罪 .....	( 286 )	10.2.6 对上网用户的几点建议 .....	( 306 )
10.1.1 计算机犯罪 .....	( 286 )	10.3 信息政策与法规 .....	( 307 )
10.1.2 计算机病毒 .....	( 289 )	10.3.1 信息道德 .....	( 307 )
10.1.3 黑客 .....	( 294 )	10.3.2 国外的信息政策与法规 .....	( 308 )
10.2 保障信息安全的措施 .....	( 295 )	10.3.3 我国的信息政策与法规 .....	( 309 )
10.2.1 信息安全 .....	( 295 )	本章小结 .....	( 310 )
10.2.2 防火墙 .....	( 299 )	思考与练习 .....	( 310 )
10.2.3 安全检查(身份认证) .....	( 302 )	参考文献 .....	( 311 )
10.2.4 加密 .....	( 303 )		

# 1

## 信息与信息技术

信息犹如空气一样普遍存在于人类社会时空之中。它作为一种客观存在,从远古直到当今的文明社会,一直都在积极发挥着人类意识到或没意识到的重大作用。“信息”是本书中最基本的概念,并作为一条主线贯穿始终,所以首先需要弄清:什么是信息?它的实质是什么?它有什么特征?它怎样度量?对这些问题的透彻理解,是收集、处理和利用信息的前提,让我们就从这里开始探索信息资源的宝库,迈向信息科学的大门。

### 1.1 探索信息的真谛

#### 1.1.1 什么是信息

信息一词来源于拉丁文“Information”,且在英语、法语、德语、西班牙语中同字,以及俄语、南斯拉夫语中同音,表明了它在世界范围内使用的广泛性。“信息”一词在我国有着很悠久的历史,早在两千多年前的西汉时期,“信”字就出现了。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。



就一般意义而言,信息可以理解成消息、情报、知识、见闻、通知、报告、事实、数据等。但真正被作为一个科学概念探讨,则是上个世纪20年代的事,而被作为科学为人们普遍认识和利用则是近几十年的事情。

对于什么叫信息,迄今说法不一;“信息”使用的广泛性使得我们难以给它下一个确切的定义。专家、学者从不同的角度为信息下的定义达十几种之多。下面所叙述的几种定义是人们从不同角度对信息的理解。

最早对信息进行科学定义的是哈特莱(Ralph V. L. Hartley)。他在1928年发表的《信息传输》<sup>①</sup>一文中,首先提出“信息”这一概念。他认为,发信者所发出的信息,就是他在通信符号表中选择符号的具体方式,并主张用所选择的自由度来度量信息。例如,假定某人在符号中选择了这样一些符号:“I am well.”就发出了“我平安”的信息,如果他选择了“I am sick.”这些符号,就发出了“我病了”的信息。不管符号所代表的意义是什么,只要从符号集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

<sup>①</sup> Ralph V. L. Hartley. Transmission of Information. Bell System Tech. Journal, vol. 7, 1928

奠定了基础。

1948年,信息论创始人、美国科学家香农(C. E. Shannon)从研究通信理论出发,第一次用数学方法定义“信息就是不确定性的消除量”,认为信息具有使不确定性减少的能力,信息量就是不确定性减少的程度。所谓不确定性,就是对客观事物的不了解、不肯定。因此,信息被看做是用以消除信宿(信息的接收者)对于信源(信息的发出者)发出哪些消息的不确定性。他还用概率统计的数学方法来度量不确定性被消除的量的大小。

几乎是在同时,控制论创始人之一、美国科学家维纳(N. Wiener)在1948年发表的名著《控制论——动物和机器中的通信与控制问题》<sup>②</sup>一书中曾经指出“信息就是信息,不是物质,也不是能量。”后来,维纳在《人有人的用处——控制论与社会》<sup>③</sup>一书中写道:“信息是在人们适应外部世界,并且使这种适应反作用于外部世界的过程中,同外部世界进行互相交换的内容的名称”;“要有效地生活,就必须有足够的信息。”在这里,维纳把人们与外界环境交换信息的过程看成是一种广义的通信过程,试图从信息自身具有的内容属性给信息下定义。这两本著作标志着控制论这门新兴学科的兴起。

关于信息的定义,有人提出用变异量来度量,认为“信息就是差异”。持这种观点的典型代表是意大利学者朗格(G. Longe)。他提出:“信息是反映事物的形式、关系和差别的东西。信息是包含于客体间的差别中,而不是在客体本身中。”按照这种观点,自然界和人类社会普遍存在可传递的差异性。差异越大,信息量就越大,没有差异就没有信息,不可传递的东西也不是信息。所谓信息量就是对事物差异度的量度或测度。

我国信息论学者钟义信教授认为:“信息是‘事物运动状态和方式,也就是事物内部结构和外部联系的状态和方式。’<sup>④</sup>

《辞源》中将信息定义为:“信息就是收信者事先所不知道的报导。《韦氏词典》(美国)对信息的描述是:“信息是用以通信的事实,是在观察中得到的数据、新闻和知识。”

对于信息的含义,至今仍是众说纷纭,莫衷一是,人们出于不同的研究目的,从不同的角度出发,对信息作用的不同理解和解释而对信息做出了定义。各种信息定义都反映了信息的某些特征。这样,难免就会产生差异性、多样化。

随着时间的推移,时代将赋予信息新的含义。现代“信息”的概念,已经与半导体技术、微电子技术、计算机技术、通信技术、网络技术、多媒体技术、信息服务业、信息产业、信息经济、信息化社会、信息管理、信息论等含义紧密地联系在一起。但信息的本质是什么?这仍然是需要进一步探讨的问题。



Ralph V. L. Hartley



N. Wiener

② N. Wiener. 控制论——动物和机器中的通信与控制问题. 北京: 科学出版社, 1963年

③ N. Wiener. 人有人的用处——控制论与社会. 北京: 商务印书馆, 1978年

④ 钟义信. 信息科学原理. 北京: 北京邮电大学出版社, 1996年

### 1.1.2 从信息论到信息科学

自 20 世纪初以来,特别是 20 世纪 40 年代,通信技术的迅速发展,迫切需要解决一系列信息理论问题,例如如何从接收的信号中滤除各种噪声,怎样解决火炮自动控制系统跟踪目标问题等。这就促使科学家在各自研究领域对信息问题进行认真的研究,以便揭示通信过程的规律和重要概念的本质。

信息论作为一门严密的科学,主要应归功于美国应用数学家香农(C. E. Shannon)。1948 年,香农在《贝尔系统技术杂志》上发表重要论文《通信的数学理论》。1949 年,香农又发表另一重要论文《在噪声中的通信》。在这些论文里,香农提出了通信系统模型、度量信息的数学公式以及编码定理和其他一些技术性问题的解决方案。香农的研究成果标志着信息论(Information Theory)的诞生。由于香农提出的信息论是关于通信技术的理论,它是以数学方法研究通信技术中关于信息的传输和变换规律的一门科学。所以,人们又将其称为狭义信息论,或经典信息论。



Shannon(1916—2001)

信息论发展的第二个阶段是一般信息论。这种信息论虽然主要还是研究通信问题,但是新增加了噪声理论、信号的滤波与检测、信号的编码与译码、信号的调制与解调以及信息的处理等问题。通信的目的是要使接收者获得可靠的信息,以便做出正确的判断与决策。为此,一般信息论特别关心信号被噪声干扰时的处理问题。

信息论发展的第三个阶段是广义信息论。它是随着现代科学技术的纵横交错的发展而逐渐形成的。一般地说,在对信息的研究中,仅考虑其形式的方面而不考虑其内容和用途,即是狭义信息。如果考虑信息的语义和有效性问题,则是广义信息。广义信息论,则超出了通信技术的范围来研究信息问题,它以各种系统、各门科学中的信息为对象,广泛地研究信息的本质和特点,以及信息的获取、计量、传输、储存、处理、控制和利用的一般规律,从而拓宽了信息论的研究方向,使得人类对信息现象的认识与揭示不断丰富和完善。显然,广义信息论包括了狭义信息论和一般信息论的内容,但其研究范围却比通信领域广泛得多,是狭义信息论和一般信息论在各个领域的应用和推广,因此,它的规律也更一般化,适用于各个领域。所以广义信息论是一门横断学科,即信息科学。

### 1.1.3 香农对信息的定义

香农在他发表的著名论文《通信的数学理论》中,从研究通信系统传输的实质出发,对信息下了科学的定义,并进行了定性和定量的描述。

香农认为:信息是有秩序的量度,是人们对事物了解的不确定性的消除或减少。信息是对组织程度的一种测度,信息能使物质系统有序性增强,减少破坏、混乱和噪音。

香农提出:信息的传播过程是“信源(信息的发送者)把要提供的信息经过“信道”传递给“信宿(信息的接收者),信宿接收这些经过“译码(即解释符号)的信息符号的过程。并由此建立了通信系统模型。

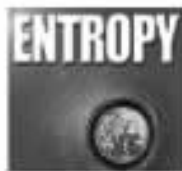
什么是信道呢?信道是在物理线路上划分的逻辑通道。由于物理上的限制,信道都只有有限的带宽,而且存在噪声,因此信道能够传递的最大数据速率是受信道带宽制约的。对于这个问题,奈奎斯特(H. Nyquist)和香农先后展开了研究。由此,香农推出了受噪声(所谓噪声是指“外加于信号之上,而非属信息源本身的信号”)干扰的信道情况下传输速率与信噪比(信号功率与噪声功率之比)之间的关系,指出了用降低传输速率来换取高保真通信的可能性。该公式已广泛用于有噪声情况下的信道最大传输速率的计算。

在香农确定信息量名称时,将热力学中“熵”的概念应用到信息领域,理由是不定性函数在统计力学中已经用在熵下面了。一个系统的熵就是它的无组织程度的度量。而一个系统中的信息量是它的组织化程度的度量,这说明信息与熵恰好是一个相反的量,信息是负熵,所以在信息熵的公式中有负号,它表示系统获得信息后无序状态的减少或消除,即消除不定性的多少。

由于熵表达了事物所含的信息量,我们不可能用少于熵的比特数来确切表达这一事物。所以这一概念已成为所有无损压缩的标准和极限。同时,它也是导出无损压缩算法达到或接近“熵”的编码的源泉。

## 什么是熵

“熵”的概念起源于热力学,是度量分子不规则热运动的物理量,即“不确定性”在热力学里用“熵”来度量。熵表示系统的无组织或混乱程度,熵愈大意味着该系统愈混乱无序,熵愈小表明该系统的组织程度愈高。香农的伟大贡献在于,将热力学中“熵”的概念应用到信息领域,并利用概率分布的理论给出信息量——熵的概念,即信息是以它对事物不确定性的减少或消除来度量。



### 1.1.4 信息的度量

我们从一则消息中获得了信息,那么我们获得的信息有多少呢?量度信息多少的测度就是信息量。信息的度量反映了人们对信息的定量认识。

根据香农的有关信息的定义,信息如何测度呢?显然,信息量与不确定性消除程度有关。消除多少不确定性,就获得多少信息量。用数学的语言来讲,不确定就是随机性。不确定性的多少可以直观地看成是事先猜测某随机事件是否发生的难易程度。

按照香农的定义,信息量的度量公式可描述为:

$$I = \log \frac{\text{后验概率}}{\text{先验概率}} \quad (1.1)$$

它实际上就是后验概率和先验概率之比。在信道无干扰情况下,由信源传来的消息告诉信宿某事件已经发生,则某事件必然发生了。按照概率的定义,对于确定发生的必然事件,其概率为1。这时公式(1.1)的后验概率为1。

假设  $P_i$  是第  $i$  个消息出现的先验概率,则第  $i$  个消息的信息量为:

$$I_i = \log \frac{1}{P_i} = -\log P_i \quad (1.2)$$

第  $i$  个消息可能有  $n$  种状态,那么输出这个消息的总信息量的期望值(即平均信息量)为:

$$H = - \sum_{i=1}^n P_i \log P_i \quad (1.3)$$

在以上公式中,对数的底数从理论上而言可以取任何数。当底数为 2 时,信息的计量单位为比特(bit),即二进制单位。

以上就是香农关于信息的度量,通常也称为概率信息。它是一个科学的定义,有明确的数学模型和定量计算。

香农公式与日常用语中的信息的含义是一致的。例如,设某一事件是在预料中一定会发生的必然事件,如果该事件果然发生了,收信者将不会得到任何信息。因为根据公式(1.1),这时  $P_i = 1$ ,所以:

$$I_i = \log \frac{1}{P_i} = -\log 1 = 0$$

反之,如果某事件发生的概率很小,即猜测它是否发生的不确定性很大,一旦这个事件发生了,收信者就会觉得很意外的惊讶,感到获得信息量很大。根据公式(1.1),因为  $P_i \ll 1$ (表示概率远小于 1),故:

$$I_i = \log \frac{1}{P_i} \gg 1$$

只有当事件发生与不发生的概率相同时,其信息量才达到最大(考虑一下,为什么?)。

香农的信息度量公式排除了对信息主观上的含意。根据上述公式,同样一个消息对任何一个收信者来说,所得到的信息量都是一样的。

## 香农——现代信息论的创始人

香农(C. E. Shannon)是现代信息论的创始人,也是电子计算机理论的重要奠基人之一。现代信息论的出现,对现代通信技术和电子计算机的设计产生了巨大的影响。如果没有信息论,现代的电子计算机是无法研制成功的。香农在美国密歇根大学和麻省理工学院学习时,修过布尔代数课,并在布尔的指导下使用微分析仪对继电器电路进行分析。他认为这些电路的设计可用符号逻辑来实现,并意识到分析继电器的有效数学工具正是布尔代数。1938年,香农发表了著名的论文《继电器和开关电路的符号分析》,首次运用布尔代数进行开关电路分析,并证明布尔代数的逻辑运算可以通过继电器电路来实现,明确地给出了实现加、减、乘、除等运算的电子电路的设计方法。香农在贝尔实验室工作中进一步证明,可以采用能实现布尔代数运算的继电器或电子元件来制造计算机。香农的理论还为计算机具有逻辑功能奠定了基础,从而使计算机既能用于数值计算,又具有各种非数值应用功能,使得以后的计算机几乎在任何领域中都得到了广泛的应用。



正在思考的 Shannon

1948年香农发表的长达数十页的《通信的数学理论》(The Mathematical Theory of Communication)成了信息论正式诞生的里程碑。在他的通信数学模型中,信息的度量问题被清楚地提了出来,得出了著名的计算信息熵的公式。

2001年2月24日,当代最伟大的数学家和贝尔实验室最杰出的科学家之一、84岁的香农博士逝世。

### 1.1.5 香农信息论的局限性

从信息和信息论的提出到现在已经有半个多世纪。早期的信息论是关于研究信息的本质和传输规律方面的,重点是信息是什么、信息量计算、概率、信道、编码方面的内容,比特(Bit)也就是从那个时候开始提出来的。这些理论是香农在解决通信问题时提出的。这个方法成功地解决了通信过程中有关信息量的各种问题。

信息论创立后,信息概念广泛渗透到物理学、化学、生物学、心理学、医学、经济学、语言学以及电子技术等各个领域。随着信息概念的广泛应用,香农信息论的局限性逐渐暴露出来。其实,香农作为一个通信科学家,所定义的仅仅是通信模型中的信息,是对信息定义的狭义解释。可是若把这一信息量的计算方法引入其他领域时,就会出现极大的局限性。

首先,香农对信息的定义的出发点是假定事物状态可以用一个以经典集合论为基础的概率模型来描述。然而实际存在的某些事物的运动状态要寻找一个合适的概率模型往往是非常困难的。对某些情况来讲,是否存在这样一个模型还值得探讨。

其次,这个定义和度量考虑收信者的主观特性和主观意义,也撇开了信息的具体含义、具体用途、重要程度和引起后果等因素。这与实际情况不完全一致。

以上问题的出现,主要是由于狭义信息论没有解决信息的语义问题和有效性问题。语义信息是指当认知主体在获得信息时,不仅要知道“是什么形式”,而且还要理解“是什么意思”。也就是说,信息的具体含义是什么。香农定义并没有解决语义信息的度量问题。这样,它的适用范围就受到严重的限制。

捍卫一种理论的最好方法就是发展它。20世纪60年代以来,有人提出了信息语义问题,并进行了一些定量研究。如1964年卡尔纳普(R. Carnap)等人提出了“语义信息”;1971年高艾斯(S. Guisasu)等人提出“有效信息”的概念;1974年哥廷格尔(H. Gottinger)提出“无概率(主观)信息”等概念。此外,还有人提出“模糊信息”的概念。1972年,德路卡(A. Deluca)等人建立了在模糊集合上的非概率的模糊熵定义,试图创立“模糊信息论”。但是,人类在关于语义信息的定性分析、定量分析和结构分析方面,长期以来并未能获得实质性的重大突破,因此至今仍无法很好地解决信息语义测度问题。

我们相信,随着人们对信息这一概念的不断深入研究,将会得出更合理、更确切的信息的定义和测度,达到彻底揭示信息的本质,全面和准确地把握和利用信息。



## 香农谈狭义信息论

香农理论诞生不久,许多人把它用到日常信息交流场合,于是漏洞百出。为此,香农要求人们对他的理论的应用不要超出随机统计领域。其实,曾和香农合著《通信的数学理论》一书的 W. Weaver 在一篇论文中就提出通信的三个水平:水平 A——通信的技术问题,如香农理论研究的;水平 B——考虑到语义问题;水平 C——考虑到效用或价值问题。

### 1.1.6 数据、消息、信号与信息

在日常生活中,人们并不刻意区分数据、消息、信号之间的区别,因为它们本身与信息有着非常天然的和紧密的联系。但是,从信息科学的角度来看,信息的含义则更为深刻和广泛,它是不能等同于数据、消息、信号的。

#### 1. 数据

数据是对客观实体的一种描述形式,是信息的载体。信息和数据的区别可以理解为:数据是未加工的信息,而信息是数据经过加工以后的能为某个目的使用的数据,信息是数据的内容或诠释。将数据加工为信息的过程称为信息加工或处理。数据可分为模拟数据和数字数据两种形式。模拟数据是在某个区间内连续的值,例如声音和视频,其强度是连续改变的波形;温度和压力,也都是连续值。数字数据是离散值,例如大多数用传感器收集的数据是非连续的值。

#### 2. 消息

人们常常错误地把信息等同于消息,认为得到了消息,就是得到了信息。

“消息”是英文 message 的中译。信息论的先驱哈特莱 1928 年在《信息传输》这篇论文中,曾经阐述过“消息”和“信息”的关系和差异。他认为信息是包含在消息中的抽象量,消息是具体的,其中蕴含着信息。

按照香农理论,在通信过程中,信息总是经过编码(符号化)成为消息以后,才能经由媒介传播的,而信息的接收者收到信息后,总是要经过译码(解读)才获取其中的信息的。在这一过程中,不管接收者的解读能力如何,不管他是否确实理解了其中的内容,不管其中的内含是否确实消除了接收者的不确定性,消息依然是消息,消息包含的依然是信息,这种客观存在是不会因接收者的状况而改变的。

#### 3. 信号

既然信息不同于消息,当然也不同于信号。

在各种实际通信系统中,为了克服时间或空间的限制而进行通信,必须对消息进行加工处理。把消息变换成适合信道传输的物理量,这种物理量称为信号。信号携带着消息,它是消息的运载工具。

信号是数据的电磁或光脉冲编码。信号可以分为模拟信号和数字信号。模拟信号是一种随时间而连续变化的信号。数字信号则是在时间上的一种离散信号。

### 1.1.7 信息的基本特性

信息是客观事物运动状态和存在方式的反映。所谓信息的特性,就是指信息区别于其他

事物的本质属性。信息的基本特征有：

### 1. 信息的普遍性、无限性和客观性

世界是物质的,物质是运动的,运动的物质既产生也携带信息。无论是自然界还是人类社会,对客观物质世界间接和概括反映的人类思维都处于永恒的运动之中,因而信息是普遍存在的。由于宇宙空间的事物是无限丰富的,所以它们所产生的信息也必然是无限的。诚然,由于人类在一定历史阶段认识领域的有限性,在此阶段获得的信息也只能是有限的,但并不能由此否认信息资源的无限性。

同时,普遍存在着的信息又是客观的。客观世界的一切事物都在不断地运动变化着,并表现出不同的特征和差异。这些特征变化就是客观实在,并通过各种各样的信息反映出来。从有人类存在以前直至今天,人类及人类以外的各种生物就利用着客观存在的大自然中无穷无尽的信息资源。信息的客观性还表现为它是物质的客观存在为前提的,即使是主观信息,如决策、判断、指令、计划等,也有它的客观实际背景,并受客观实践的检验。因此,信息必须真实、准确,必须如实地反映客观实际。

### 2. 信息的可共享性

信息区别于物质的一个重要特征是它可以被共同分享和占有。信息的共享性有两层含义:一是信息交换的双方,即传播者和接受者都可以享有被交换的同一信息;二是信息在交换或交流过程中,可以同时为众多的接受者所接收和利用。

信息的分享不仅不会失去原有信息,而且还可以广泛地传播与扩散,供全体接收者所共享。与其相反,物质的交换遵循易物交换原则,你的所得,必为我之所失,其结果是零和的。



### 3. 信息的可存储性

任何信息都是以某种物质的特定的运动形式表现出来的,它不能独立存在于某种物质之外,必须依附于物质载体而存在,需要物质承载者。也就是说,信息能够以一定的方式存储在某种物质载体之中。信息在时间上的传递通常被称为信息的存储。人们存储信息的目的在于利用信息。实际上,人类文明就是这样传承下来的。

人类除运用大脑进行信息存储外,一般要运用语言、文字、图像、符号等记载信息,并通过声波、光波、电波等信息媒体进行传递。如果要使信息长期保存下来,还必须采用纸张、胶卷、磁带、磁盘等实物作为它的载体加以存储。没有物质载体,信息就不能存储和传播,但其内容并不因记录手段或物质载体的改变而发生变化。

### 4. 信息的可传输性

人们要获取信息必须依赖于信息的传输。把信息从时间或空间上的某一点向其他点移动的过程称为信息传输。一个完整的信息传输过程必须具备信源、信宿、信道和信息四个基本要素。

信息可以通过多种渠道、采用多种方式进行传输。人与人之间信息传输一般依赖语言、文字、表情、动作,社会信息的传输则通过报纸、杂志、文件等。随着现代通信技术的发展,信息可以通过电话、电报、广播、通信卫星等通信手段进行传输。信息的传输成本远远低于物质和能

源的传输。

### 5. 信息的可扩散性

信息富于渗透力,力图冲破保密的非自然束缚。信息好像热源,它总是力图向温度低的地方扩散。信息的浓度越大,信息源和接收者的梯度越大,信息的扩散力度越强。在日常生活中,越是离奇的消息,越是爆炸性的新闻,它就传播的越快,扩散的范围越大,这正说明了信息扩散的威力。

### 6. 信息的可转换性

信息的传递是同物质和能量的传递相关的,其传递过程中必将伴有一定的物质及其运动的传递或变换、能量的传递或能量形式的变换。

此外,信息在变换载体时的不变性,使得信息可以方便地从一种形态转换为另一种形态。如信息可以转换为语言、文字、数据、图像等形式,也可转换为计算机代码、电磁光信号等。由此可知,信息对于载体的可选择性使得如今的信息传递不仅可以在传播方式上加以选择,而且在传递时间和空间上提供了极大的方便,并使得人类开发和利用信息资源的各项技术的实现成为可能。

### 7. 信息的可度量性

信息也是可以度量的。信息论的创始人香农,舍去事件发生的时间、地点、内容以及人的情感等因素,只考虑事件发生的状态数目及每种状态发生的可能性大小,给出用以度量信息的熵函数,这是信息度量最基本的一种方法。虽然这种信息度量的方法有其局限性及其适用的条件和环境,但信息论、信息科学的发展也正是以此为起点,并在发展过程中不断改进和创新信息的度量方法,扩大大量信息的范围和层次。

### 8. 信息的可压缩性

信息可以进行浓缩、集中、概括以及综合,而不至丢失信息的本质。信息压缩在实际中很有必要。因为我们没有能力收集一个事物的全部信息,也没有能力和必要存储所有的信息,这叫信息的不完全性。只有正确地舍弃冗余信息,才能正确地使用信息。

## 1.1.8 信息的基本作用

信息作为一种客观存在,它一直都在积极地发挥着人类意识或没有意识到的重大作用。科学技术在近两个世纪所取得的空前进步,使人们终于认识到,信息是与物质和能源可以相提并论的用以维系人类社会存在及发展的三大要素之一。因此,只有科学地了解和认识信息的基本作用,才能更好地把握信息,进而才能使信息更好地为科学技术、经济和社会发展服务。

概括起来,信息的基本作用主要体现在以下几个方面。

#### 1. 信息是人类认识客观世界及其发展规律的基础

信息的基本功能主要表现为信息的认识功能。信息是客观事物及其运动状态的反映,是揭示客观事物发展规律的重要途径。客观世界里到处充满着各种形式和内容的信息,人类的认识器官,包括感觉器官和思维器官,对各种渠道的信息进行接收,并通过思维器官将已收集到的大量信息进行鉴别、筛选、归纳、提炼、存储而形成不同层次的感性认识和理性认识。在这一认识过程中,人类是认识论的主体,信息是认识论的客体。