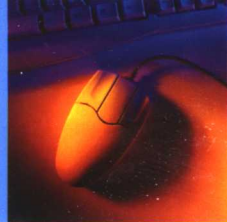




新世纪高职高专教改项目成果教材
Xinshiji Gaozhi Gaozhuan Jiaogai Xiangmu Chengguo Jiaocai



计算机与 信息处理技术

鄂大伟 郭士正 主编



高等教育出版社

新世纪高职高专教改项目成果教材

计算机与信息处理技术

鄂大伟 郭士正 主编

高等教育出版社

内容提要

本书是教育部新世纪高职高专教育人才培养模式和教学内容体系改革与建设项目成果教材,是组织有关教育部高职高专教育专业教学改革试点院校编写的。

全书以计算机技术为主线,以信息处理技术为导引,主要介绍了以计算机技术、通信技术和多媒体技术为代表的现代信息技术,以及如何利用数据库技术和信息系统实现对信息的组织、管理和利用,使读者获得对计算机与信息处理较系统、完整且新颖的认知。

本书适用于高等职业学校、高等专科学校、成人高校、本科院校举办的二级职业技术学院,也可供示范性软件职业技术学院、继续教育学院、民办高校、技能型紧缺人才的培养使用,还可供本科院校、计算机专业人员和爱好者参考使用。

图书在版编目(CIP)数据

计算机与信息处理技术 / 鄂大伟, 郭士正主编. —北京: 高等教育出版社, 2004.7

ISBN 7-04-014764-5

I. 计... II. ①鄂... ②郭... III. 电子计算机—高等学校—教材 IV. TP3

中国版本图书馆 CIP 数据核字 (2004) 第 061878 号

策划编辑 冯 英 责任编辑 张春英 封面设计 王凌波 责任绘图 朱 静
版式设计 胡志萍 责任校对 殷 然 责任印制 孔 源

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-82028899

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京星月印刷厂

开 本 787×1092 1/16
印 张 23.5
字 数 570 000

版 次 2004 年 7 月第 1 版
印 次 2004 年 7 月第 1 次印刷
定 价 29.30 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

出版说明

为认真贯彻《中共中央国务院关于深化教育改革全面推进素质教育的决定》和《面向 21 世纪教育振兴行动计划》，研究高职高专教育跨世纪发展战略和改革措施，整体推进高职高专教学改革，教育部决定组织实施《新世纪高职高专教育人才培养模式和教学内容体系改革与建设项目计划》（教高[2000]3 号，以下简称《计划》）。《计划》的目标是：“经过五年的努力，初步形成适应社会主义现代化建设需要的具有中国特色的高职高专教育人才培养模式和教学内容体系。”《计划》的研究项目涉及高职高专教育的地位、作用、性质、培养目标、培养模式、教学内容与课程体系、教学方法与手段、教学管理等诸多方面，重点是人才培养模式的改革和教学内容体系的改革，先导是教育思想的改革和教育观念的转变。与此同时，为了贯彻落实《教育部关于加强高职高专教育人才培养工作的意见》（教高[2000]2 号）的精神，教育部高等教育司决定从 2000 年起，在全国各省市的高等职业学校、高等专科学校、成人高等学校以及本科院校的职业技术学院（以下简称高职高专院校）中广泛开展专业教学改革试点工作，目标是：在全国高职高专院校中，遴选若干专业点，进行以提高人才培养质量为目的、人才培养模式改革与创新为主题的专业教学改革试点，经过几年的努力，力争在全国建成一批特色鲜明、在国内同类教育中具有带头作用的示范专业，推动高职高专教育的改革与发展。

教育部《计划》和专业试点等新世纪高职高专教改项目工作开展以来，各有关高职高专院校投入了大量的人力、物力和财力，在高职高专教育人才培养目标、人才培养模式以及专业设置、课程改革等方面做了大量的研究、探索和实践，取得了不少成果。为使这些教改项目成果能够得以固化并更好地推广，从而总体上提高高职高专教育人才培养的质量，我们组织了有关高职高专院校进行了多次研讨，并从中遴选出了一批较为成熟的成果，组织编写了一批“新世纪高职高专教改项目成果”教材。这些教材结合教改项目成果，反映了最新的教学改革方向，很值得广大高职高专院校借鉴。

新世纪高职高专教改项目成果教材适用于高等职业学校、高等专科学校、成人高校及本科院校举办的二级职业技术学院、继续教育学院和民办高校使用。

高等教育出版社

2002 年 11 月 30 日

序 言

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重要里程碑。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,正在改变着人们传统的生活、学习和工作方式,同时也影响着教育的内容与方法。

本书以计算机技术为主线,以信息处理技术为导引,主要介绍了以计算机技术、通信技术和多媒体技术为代表的现代信息技术,以及如何利用数据库技术和信息系统实现对信息的组织、管理和利用,使读者获得对计算机与信息处理较系统、完整且新颖的认知。

在本书的编写过程中,我们力求在结构上和内容上有所创新,并有较鲜明的特色。首先,书中许多内容取自于国内外最新的资料和 Internet 上的资源,以反应信息技术当前的发展。其次,全书包含的信息量较大,知识性较强,目的是让具有不同层次、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”。在编写的体例和结构上,符合高职高专教学计划体系的要求和信息技术的教学目标。内容上力求与相关学科相互融合,使教材内容能最大限度地体现计算机技术与信息处理技术的“整体”面貌。

本书由鄂大伟、郭士正主编。参加本教材编写工作的有鄂大伟(第1、8章)、郭士正(第2、3、4、9、12章)、贾红伟(第5章)、黄凯明(第6、7章)、张仪华(第10、11章)。

在本书的编写过程中,得到了高等教育出版社高职高专分社编辑们的大力支持和帮助,没有他们的帮助就没有此书的顺利出版。另外,许多书籍和文章的作者以及网络媒体的提供者们提供的大量信息资源为此书增色不少,在此一并表示谢意,恭疏短语,难竭鄙诚。

计算机与信息处理技术的发展一日千里,相关的各学科都变得越来越精微和深奥。限于我们的水平及篇幅,本书内容仍难以准确反映和把握计算机与信息处理技术的整体与全貌,疏漏、欠妥、悖谬之处难免,恳请读者不吝赐教,以期冀收博见,嘉惠来学。

作 者

2004年2月于厦门集美学校

郑 重 声 明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 58581897/58581896/58581879

传 真：(010) 82086060

E - mail: dd@hep.com.cn

通信地址：北京市西城区德外大街 4 号

高等教育出版社打击盗版办公室

邮 编：100011

购书请拨打电话：(010)64014089 64054601 64054588

目 录

第 1 章 信息与信息技术1	2.2 计算机科学32
1.1 探索信息的真谛1	2.2.1 计算机科学的产生 背景与定义.....32
1.1.1 什么是信息.....1	2.2.2 计算机科学与技术的 层次划分.....33
1.1.2 从信息论到信息科学.....3	2.3 计算机基本工作原理34
1.1.3 香农对信息的定义.....3	2.3.1 信息在计算机内的表示.....34
1.1.4 信息的度量.....4	2.3.2 字符的编码.....36
1.1.5 香农信息论的局限性.....6	2.3.3 计算机逻辑基础.....38
1.1.6 数据、消息、信号与信息.....7	2.3.4 计算机运算基础.....40
1.1.7 信息的基本特性.....7	2.4 计算机系统42
1.1.8 信息的基本作用.....9	2.4.1 计算机系统的组成.....42
1.2 信息科学与信息技术10	2.4.2 微型计算机.....45
1.2.1 信息科学.....10	2.4.3 工作站.....47
1.2.2 信息技术.....11	2.4.4 巨型计算机.....48
1.2.3 信息技术的核心.....13	2.4.5 嵌入式计算机.....50
1.3 信息化与信息社会16	2.5 微电子技术简介52
1.3.1 什么是信息化.....16	2.5.1 微电子技术基础知识.....52
1.3.2 国家信息化指标.....16	2.5.2 数字信号处理器——DSP.....56
1.4 信息产业与信息人才17	2.5.3 现代微电子技术的成就及 发展前景.....58
1.4.1 信息产业的概念.....17	本章小结59
1.4.2 信息产业的特点.....18	思考与练习60
1.4.3 信息产业对人才的需求.....20	参考文献60
1.4.4 IT 人才的职业划分与 人才培养.....20	第 3 章 微型计算机硬件系统62
本章小结22	3.1 微型计算机的硬件系统62
思考与练习22	3.1.1 微型计算机的主板.....62
参考文献23	3.1.2 中央处理器.....64
第 2 章 计算机科学与技术24	3.1.3 存储器.....64
2.1 计算与计算科学24	3.2 计算机的输入与输出设备71
2.1.1 何谓计算.....25	3.2.1 输入输出设备概述.....71
2.1.2 计算的描述.....26	3.2.2 常用输入设备.....71
2.1.3 计算学科的根本问题.....28	
2.1.4 图灵机与计算模型.....28	

3.2.3 常用输出设备	77	保护的历史发展及现状	124
3.3 计算机的总线与接口	81	4.6.2 计算机软件版权保护的 利与弊	125
3.3.1 计算机总线的基本概念	81	4.6.3 计算机软件单独立法 保护的必要性	126
3.3.2 计算机总线标准	83	4.6.4 我国软件著作权保护的建立	126
3.3.3 ISA 总线	83	本章小结	127
3.3.4 PCI 总线	83	思考与练习	127
3.3.5 计算机接口及标准	84	参考文献	128
本章小结	86	第 5 章 多媒体信息处理技术	130
思考与练习	87	5.1 多媒体的概念	130
参考文献	87	5.1.1 媒体的分类	130
第 4 章 计算机软件系统	89	5.1.2 多媒体的定义	131
4.1 计算机软件的概念	89	5.1.3 多媒体计算机系统	131
4.1.1 计算机软件的定义、特点	89	5.1.4 多媒体的主要特征	133
4.1.2 计算机软件的发展阶段	91	5.2 多媒体音频信号处理	134
4.1.3 计算机软件的分类	93	5.2.1 音频信号的形式	134
4.1.4 计算机软件的基本内容	94	5.2.2 模拟音频信号的物理特征	135
4.1.5 操作系统	96	5.2.3 音频的数字化过程	136
4.2 高级语言程序设计	100	5.2.4 声音的采样	136
4.2.1 从机器语言到高级语言 发展的回顾	100	5.2.5 量化	137
4.2.2 语言翻译系统	104	5.2.6 数字音频的编码	138
4.2.3 第 4 代语言简介	105	5.2.7 数字音频的文件格式	139
4.3 结构化程序设计方法	106	5.3 多媒体图形与图像处理	141
4.3.1 程序、数据结构、算法 及其关系	106	5.3.1 计算机图形处理的概念	141
4.3.2 结构化程序的基本 逻辑结构	108	5.3.2 计算机图像处理的概念	142
4.3.3 程序设计的一般步骤	108	5.3.3 图形与图像的区别与联系	143
4.3.4 结构化程序设计方法	109	5.4 图像的数字化过程	144
4.3.5 程序设计的步骤	110	5.4.1 图像数字化概述	144
4.4 面向对象的程序设计方法	111	5.4.2 图像的采样	144
4.5 软件工程	112	5.4.3 图像的量化	145
4.5.1 软件工程与软件过程	112	5.4.4 图像的编码与压缩	146
4.5.2 传统方法学	114	5.4.5 矢量图与位图	146
4.5.3 面向对象方法学	116	5.4.6 图像文件格式	148
4.5.4 软件项目管理	121	5.5 多媒体图像压缩与编码技术	149
4.6 软件著作权保护	124	5.5.1 信息压缩	149
4.6.1 计算机软件版权立法		5.5.2 预测编码	150
		5.5.3 变换编码	151

5.5.4 统计编码	154	7.1.2 域名、域名系统和 Internet	
5.5.5 静态图像压缩标准 JPEG	154	域名空间	226
5.5.6 运动图像压缩标准 MPEG	157	7.2 因特网协议 TCP/IP	228
5.5.7 MPEG 的编码与实现过程	159	7.3 因特网接入技术	233
本章小结	164	7.3.1 拨号连接	233
思考与练习	165	7.3.2 局域网接入	234
参考文献	166	7.3.3 宽带接入	234
第 6 章 计算机网络技术	167	7.3.4 无线接入	235
6.1 通信技术	167	7.3.5 其他连接方式	236
6.1.1 计算机与通信	167	7.4 因特网的服务与资源	237
6.1.2 通信系统模型	168	7.4.1 传统网络服务	237
6.1.3 数据传输的概念与分类	171	7.4.2 新兴起来的网络服务	241
6.1.4 典型的通信接口与标准	176	7.4.3 Internet 的其他应用	244
6.2 计算机网络	180	7.5 网络安全	244
6.2.1 计算机网络的定义	180	7.5.1 网络安全基本术语	244
6.2.2 计算机网络的发展	181	7.5.2 网络系统的安全威胁	244
6.2.3 计算机网络的拓扑结构	183	7.5.3 网络系统的安全性	245
6.3 网络传输媒体	185	7.5.4 网络系统安全结构	245
6.3.1 有线媒体	185	7.5.5 网络应用系统安全体系	246
6.3.2 无线媒体	187	7.5.6 局域网安全解决方案	247
6.4 计算机网络协议	192	7.5.7 广域网安全解决方案	247
6.4.1 协议的概念	192	7.6 网络信息组织与发布	249
6.4.2 标准与标准化组织	193	7.6.1 超文本与超媒体	249
6.4.3 开放系统互连参考模型	194	7.6.2 数据流传输技术	250
6.4.4 网络协议实例分析	197	7.6.3 网页制作与工具	252
6.4.5 网络互连设备	206	7.6.4 流媒体技术	254
6.5 高速宽带网络技术	210	7.6.5 网上音频与视频的实时播放	256
6.5.1 异步传输模式	210	7.7 网站的建设与管理	257
6.5.2 高速以太网	213	7.7.1 网站的建设	257
6.5.3 波分多路复用	216	7.7.2 网站的管理	259
6.5.4 无线局域网	219	本章小结	262
本章小结	219	思考与练习	262
思考与练习	220	参考文献	263
参考文献	221	站点资源	263
站点资源	222	第 8 章 数据库技术基础	265
第 7 章 因特网及其应用	224	8.1 数据管理技术的发展	265
7.1 因特网的基本概念	224	8.1.1 人工管理阶段	265
7.1.1 Internet 和 Intranet	224	8.1.2 文件系统阶段	266

8.1.3 数据库系统阶段	266	9.4.2 参照完整性	284
8.2 数据模型	267	9.4.3 用户定义的完整性	284
8.2.1 由现实世界到数据世界	267	9.5 关系的规范化	285
8.2.2 数据模型	269	9.5.1 为什么要对关系进行规范化	285
8.2.3 层次模型	269	9.5.2 函数依赖的概念	285
8.2.4 网状模型	270	9.5.3 关系的范式	285
8.2.5 关系模型	270	9.5.4 关系规范化小结	287
8.3 数据库系统的组成与模式结构	271	9.6 关系数据库标准语言——SQL	287
8.3.1 数据库系统的组成	271	9.6.1 基本概念	287
8.3.2 数据库系统的三级模式结构	272	9.6.2 数据定义	288
8.3.3 数据库系统研究的范围	273	9.6.3 数据更新	289
8.3.4 学习使用数据库应 掌握些什么	273	9.6.4 查询	290
8.4 数据库技术的发展	274	9.6.5 视图	290
8.4.1 传统数据库技术的发展	274	9.6.6 访问控制	290
8.4.2 新一代数据库技术	274	本章小结	291
本章小结	276	思考与练习	291
思考与练习	277	参考文献	292
参考文献	277	第 10 章 信息系统概述	293
第 9 章 关系数据库	278	10.1 信息系统的概念	293
9.1 关系数据库理论	278	10.1.1 输入、处理、输出与反馈	293
9.1.1 关系模型的特点	278	10.1.2 信息系统的结构	294
9.1.2 关系模型的基本概念及术语	278	10.1.3 基于计算机的信息系统	295
9.1.3 关系所具有的基本性质	279	10.2 系统与建模	296
9.1.4 关系模型支持的三种 基本运算	279	10.2.1 系统组成和概念	297
9.2 关系数据库设计	280	10.2.2 系统性能和标准	297
9.2.1 数据库设计的全过程	280	10.2.3 系统变量和参数	298
9.2.2 需求分析阶段	281	10.2.4 系统建模的描述工具	298
9.2.3 概念结构设计	281	10.3 信息系统的分类	304
9.2.4 逻辑结构设计	281	10.3.1 事务处理系统	304
9.2.5 物理结构设计	281	10.3.2 管理信息系统	304
9.2.6 数据库的实施和维护	282	10.3.3 决策支持系统	305
9.3 数据库的概念模型设计	282	10.3.4 人工智能与专家系统	306
9.3.1 E-R 图方法	282	10.3.5 其他信息系统	307
9.3.2 由 E-R 图导出关系数据模型	282	10.4 电子商务与因特网	308
9.4 关系完整性	284	10.4.1 电子数据交换	308
9.4.1 实体完整性	284	10.4.2 电子商务	309
		10.4.3 企业资源规划	311
		本章小结	312

思考与练习	313	12.3.1 信息安全	338
参考文献	313	12.3.2 网络信息安全的层次结构	339
第 11 章 信息系统的开发模型	314	12.3.3 信息安全的核心技术	340
11.1 信息系统开发的一般过程	314	12.3.4 信息安全的基础和准则	343
11.2 生命周期法	315	12.3.5 协议在信息安全中的 桥梁作用	345
11.3 原型法	316	12.3.6 信息安全的关键	345
11.4 面向对象的开发方法	317	12.3.7 信息安全的保障	346
11.4.1 面向对象技术及其 基本性质	317	12.3.8 信息安全的最终考验	347
11.4.2 面向对象的方法学	319	12.4 防火墙技术	348
11.4.3 软件设计风格	320	12.4.1 防火墙的基本知识	348
11.5 软件能力成熟度模型	322	12.4.2 设置防火墙的目的	349
11.6 CASE 方法与工具	324	12.4.3 防火墙的基本准则	349
11.6.1 CASE 方法的基本概念	324	12.4.4 设置防火墙的作用	350
11.6.2 CASE 工具的分类	325	12.4.5 防火墙的失效之处	350
11.7 选择适当的开发方法	325	12.4.6 防火墙的基本类型	352
本章小结	326	12.4.7 防火墙的身份认证	352
思考与练习	326	12.4.8 防火墙关键技术	353
参考文献	326	12.5 信息安全管理	354
第 12 章 信息系统安全与管理	327	12.5.1 信息安全的分层管理	354
12.1 信息犯罪及其预防	327	12.5.2 信息安全管理标准	355
12.1.1 信息犯罪的基本类型	327	12.5.3 构建信息安全管理体制	356
12.1.2 信息犯罪的主要特征	328	12.6 信息政策与法规	357
12.1.3 计算机犯罪的手段	329	12.6.1 信息道德	358
12.2 计算机病毒	330	12.6.2 国外的信息政策与法规	358
12.2.1 计算机病毒的概念与特点	330	12.6.3 我国的信息政策与法规	359
12.2.2 计算机病毒的传染 方式及症状	331	本章小结	359
12.2.3 黑客	336	思考与练习	360
12.3 信息与网络安全体系的构建	337	参考文献	361
		参考网站	361

第 1 章 信息与信息技术

信息犹如空气一样普遍存在于人类社会的时空之中。也许正是因为我们整天都淹没在信息的海洋中,因而对信息并没有给予太多的关注。信息、物质和能量是组成一切事物的三个基本方面,信息作为一种客观存在,从远古直到当今的文明社会,都一直在积极发挥着人类意识到或没意识到的重大作用。所以我们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

信息的本质究竟是什么?人类始终在不断地追问自己。今天,人类已经跨入信息时代,对于信息的本质,我们到底应当做出怎样的诠释呢?

信息一词来源于拉丁文“information”,且在英语、法语、德语、西班牙语中是同一个字,与俄语、南斯拉夫语中的发音相同,表明了它在世界范围内使用的广泛性。“信息”一词在我国也有着很悠久的历史,早在两千多年前的西汉时期,就出现了“信”字。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。当时的“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。但它真正被作为一个科学概念来探讨,则是 20 世纪 30 年代的事,而作为科学被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,迄今为止说法不一,“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多达十几种。下面所叙述的几种定义是人们从不同角度对信息的理解。

(1) 最早对信息进行科学定义的是哈特莱(Ralph V.L. Hartley)(图 1-1)。他在 1928 年发表的《信息传输》^①一文中,首先提出了“信息”这一概念。他提出消息是代码、符号,而不是信息内容本身,使信息与消息区分开来。他认为发信者所发出的信息,就是他在通信符号表中所选择符号的具体方式,并主张用所选择的自由度来度



图 1-1 Ralph V.L. Hartley
1888—1970

^① V L Hartley *Ralph Transmission of Information In: Bell System Tech. Journal*, 1928, vol. 7

量信息。哈特莱的思想和研究成果为信息论的创立奠定了基础。

(2) 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)所发出消息的不确定性。他还用概率统计的方法,来度量不确定性被消除的量的大小。

(3) 控制论创始人之一,美国科学家维纳(N.Wiener)(图1-2)在1948年发表的名著《控制论——动物和机器中的通讯与控制问题》^①一书中曾经指出“信息就是信息,不是物质,也不是能量。”维纳在这本书的导言中曾明确地指出:“必须发展一个关于信息量的统计理论,在这个理论中,单位信息就是对两种择一的事物做单一选择时所传递出去的信息”。后来,维纳在《人有人的用处——控制论与社会》^②一书中写到:“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定义,这两本著作标志着控制论这门新兴学科的兴起。



图1-2 N.Wiener

(4) 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。

(5) 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。”^③

(6) 我国的权威性工具书《辞源》对信息定义为:“信息就是收信者事先所不知道的报导。”《韦氏字典》(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。人们出于不同的研究目的,从不同的角度出发,对信息的作用有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。信息是一个动态的概念,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。

① Wiener N. 控制论——动物和机器中的通讯与控制问题,北京:科学出版社,1963

② Wiener N. 人有人的用处,北京:商务印书馆,1978

③ 钟义信. 信息科学原理,北京:邮电大学出版社,1996

1.1.2 从信息论到信息科学

从 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中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)(图 1-3)。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的模型、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。



图 1-3 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、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叉的发展而逐渐形成的。一般地说,狭义信息是指在对信息的研究中,仅考虑其形式方面而不考虑其内容和用途。如果考虑了信息的语义和有效性问题,则是广义信息。广义信息论,远远超出了通信技术的范围所研究的信息问题,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。广义信息论的研究与很多学科密切相关,例如,数学、物理学、控制论、计算机科学、逻辑学、心理学、语言学、生物学、仿生学、管理科学等。信息论在各个方面得到了广泛的应用,主要研究以计算机处理为中心的信息处理的基本理论,包括语言和文字的处理、图像的识别、学习理论及其各种其他应用。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,所以广义信息论又称信息科学。

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,对信息做了科学的定义,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,信息能使物质系统的有序性增强,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），及信宿接收这些经过“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，信道能够传递数据的最大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，香农推出了在受到噪声（所谓噪声是指“外加于信号之上，并非属于信息源本身的信号”）干扰的信道的情况下，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的情况下，信道的最大传输速率的计算。

在香农确定信息量名称时，将热力学中的“熵”的概念应用到信息领域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能达到或接近“熵”的编码的源泉。



什么是熵

“熵”（图1-4）的概念起源于热力学，是度量分子不规则热运动的单位，即“不确定性”在热力学里用“熵”来度量。熵表示系统的无组织程度或混乱程度，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，将热力学中的“熵”的概念应用到信息领域中，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的不确定性的减少或消除来度量的。

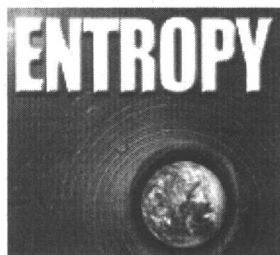


图 1-4 熵

1.1.4 信息的度量

根据香农有关信息的定义，信息源所发出的消息带有不确定性。用数学的语言来讲，不确定性就是随机性。那么如何测度信息呢？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。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可能程度。

信息源发出的消息是随机的，可以用随机变量来表示。如果用符号 X 表示信息源发出的信息，假设事件的基本空间 Ω 包含 m 个元素，即 $\Omega = \{x_1, x_2, \dots, x_m\}$ ，且每一个等可能值的概率为

$$P(X = x_i) = p \quad i = 1, 2, \dots, m$$

那么定义一个随机事件 x 所含的信息量称为 x 的自信息量，即

$$I(x) = \log \frac{1}{p(x)} = -\log p(x)$$

式中 $I(x)$ 代表事件 x 的自信息量, $p(x)$ 为事件 x 出现的概率。

$I(x)$ 只能表示信息源发出的某一特定消息 x 的自信息量, 而对于不同的消息则有不同的自信息量, 所以 $I(x)$ 不足以作为整个信息源的总体信息测度, 因此, 可以定义平均信息量来作为信息总体的测度, 即信息熵。

设 X 为一离散型随机变量, 在集合 $\{x_1, x_2, \dots, x_n\}$ 中取值, 其概率分布为

$$P\{x = x_i\} = p_i \quad i = 1, 2, \dots, n \quad \sum_{i=1}^n p_i = 1$$

定义

$$H(X) = \sum_{i=1}^n p_i \log \frac{1}{p_i} = - \sum_{i=1}^n p_i \log p_i$$

$H(X)$ 被称为离散变量 X 的信息熵。

信息熵是从整个信息源的统计特性来考虑的, 它从平均的意义上来表示信息源的总体信息测度, 它表示信息源 X 在没有发出消息以前, 信宿对信息源 X 存在着平均不确定性。

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公式。在公式中, 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。

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) (图 1-5) 是现代信息论的著名创始人之一, 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。香农在美国密执安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分析仪, 对继电器电路进行分析。他认为这些电路的设计可以用符号逻辑来实现, 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时进一步证明了, 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, 从而使电子计算机既能用于数值计算, 又具有各种非数值应用的功能, 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。

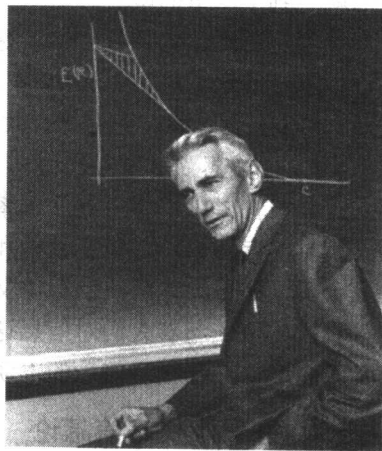


图 1-5 正在思考的 Shannon

1948 年, 香农长达数十页的论文 *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, 信息度量的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、贝尔实验室最杰出的科学家之一,84岁的香农博士不幸去世。

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,早期的信息论是研究信息的本质和传输规律的,重点研究信息是什么,信息量的计算、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,然而为实际存在的某些事物的运动状态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,这个定义和度量要考虑收信者的主观特性和主观意义,同时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”,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了严重的限制。

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.Guiasu)等人提出“有效信息”的概念;1974年哥廷格尔(H.Gottinger)提出的“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,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的场合,于是漏洞百出。为此香农要求人们对它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在其论文中就提出通信的三个水平:水平A——通信的技术问题,如香农理论所研究的;水平B——考虑到语义问题;水平C——考虑到效用或价值问题。