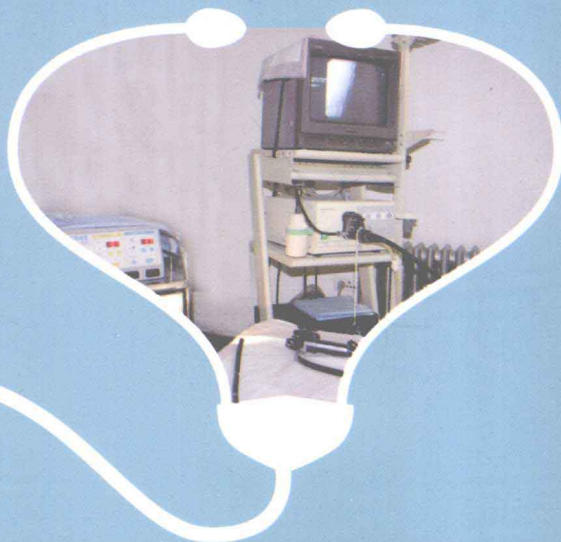


普通高等教育“十二五”规划教材
全国高等院校规划教材



医疗信息系统

主编 康晓东 张建春 胡春红

清华大学出版社

“十二五”普通高等教育
国家级规划教材

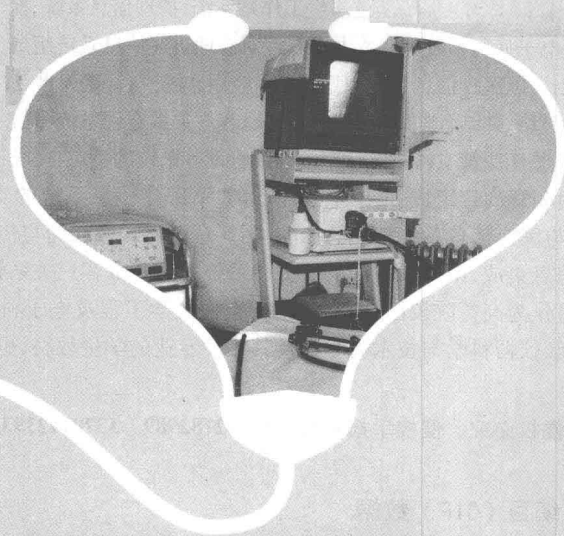


医疗信息系统

主 编 王 颖 副主编 王 颖 王 颖

清华大学出版社

普通高等教育“十二五”规划教材
全国高等院校规划教材



医疗信息系统



主编 康晓东 张建春 胡春红

清华大学出版社
北京

内 容 简 介

广义上的医疗信息系统包括医生工作站系统、护理信息系统、检验信息系统、放射信息系统、医疗影像存储与传输系统以及电子病历系统等。本书在有针对性地讲述信息系统组成及开发理论的基础上,立足国内外医疗信息系统的发展现状,以医疗信息系统规划、设计、应用、维护与管理为重点,以医学数字成像、美国卫生信息传输标准及电子病历等为脉络,并结合临床决策支持系统、客户关系管理系统等信息技术规范,重点讲解了医学影像传输存储、电子病历应用、临床决策支持及医疗单位辅助管理等系统的特点、集成与信息交换等方面的内容。

本书由8章内容组成:第1章概述、第2章医疗信息系统的支撑技术、第3章医疗信息系统的规划与管理、第4章与医疗信息系统相关的标准、第5章医学影像存储与传输系统及其数据库的开发实现、第6章电子病历及其信息集成、第7章临床决策支持系统和第8章客户关系管理系统辅助医疗管理系统的设计与应用。

本书可作为高等院校医学影像技术、生物医学工程及电子科学与工程类、控制理论与工程类、计算机理论与技术类、仪器科学与技术类专业及其相关专业的学生教材,也可供相关领域与专业的科研人员参考。

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

医疗信息系统/康晓东,张建春,胡春红主编. --北京:清华大学出版社,2011.2

(普通高等教育“十二五”规划教材 全国高等院校规划教材)

ISBN 978-7-302-24437-0

I. ①医… II. ①康…②张…③胡… III. ①医疗保健-管理信息系统-高等学校-教材
IV. ①R197.324

中国版本图书馆CIP数据核字(2010)第257884号

责任编辑:罗 健

责任校对:刘玉霞

责任印制:王秀菊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

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

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185×260 印 张:13.75 字 数:349千字

版 次:2011年2月第1版 印 次:2011年2月第1次印刷

印 数:1~3000

定 价:29.80元

产品编号:036853-01

前 言

广义上的医疗信息系统包括医生工作站系统、护理信息系统、检验信息系统、放射信息系统、医疗影像存储与传输系统以及电子病历系统等。

随着信息技术与临床医学的不断融合与发展,通过信息技术将医疗单位信息集成数字化医疗模式是大势所趋。医疗信息系统的规划、建设、管理与应用方兴未艾。

本教材编者在天津医科大学相关课程教学大纲和讲义的基础上,结合天津医科大学前期取得的科研成果,通过多次到医疗信息系统建设与应用比较深入的医院,特别是国内医疗信息化应用较早的部队医院实地考察,并广泛听取了计算机学会、中国生物医学工程学会和中华医学会等业内专家的意见后,组织有关研究、开发和应用医疗信息系统的专家学者,经过充分调研、酝酿和讨论,共同完成本教材的编写任务。

本教材主要供高等院校医学影像技术、生物医学工程专业学生使用,也可作为电子科学与工程类、控制理论与工程类、计算机理论与技术类、仪器科学与技术类专业及相关专业的学生教材。

本教材在有针对性地讲述信息系统组成及开发理论的基础上,梳理了国内外医疗信息系统的发展脉络,以医疗信息系统规划、设计、应用、维护与管理为重点,以医学数字成像、美国卫生信息传输标准及电子病历等为主线,并结合临床决策支持系统、客户关系管理系统等信息技术规范,分别对医学影像传输存储、电子病历应用、临床决策支持及医疗单位辅助管理等系统的特点、集成与信息交换等方面内容进行了重点讲解。

本教材由8章内容组成:第1章概述、第2章医疗信息系统的支撑技术、第3章医疗信息系统的规划与管理、第4章与医疗信息系统相关的标准、第5章医学影像存储与传输系统及其数据库的开发实现、第6章电子病历及其信息集成、第7章临床决策支持系统和第8章客户关系管理系统辅助医疗管理系统的设计与应用。本教材适宜学时为54~72学时。

编者感谢海内外相关领域学者们的工作,是他们的真知灼见充实了本教材的内容(引用的参考文献若有疏忽漏注处,祈请谅解);还要感谢清华大学出版社的朋友们,感谢他们为本教材的出版贡献了他们宝贵的时间和精力。

最后,本教材是集体努力的结晶。康晓东博士起草了大纲,程岚岚硕士(天津科技大学讲师)撰写了第1章和第5章,郑小慎博士(天津科技大学副教授)撰写了第2章,赵德新博士(天津理工大学副教授)撰写了第3章,乔清理博士(天津医科大学副教授)撰写了第4章,胡春红博士(天津医科大学副教授)撰写了第6章和第8章,张建春硕士(中国人民解放军第一医院主任医师)撰写了第7章。天津医科大学研究生刘玲

玲、王昊和段敬豪参与了文档资料的整理工作。全书由康晓东教授统稿。

诚然,本教材难免有错漏和不足之处,故诚恳地希望各位读者以及从事相关研究工作的专家学者提出宝贵的意见和建议。

编 者
2010 年 12 月

目 录

第1章 概 述	1
第1节 信息与医学信息	1
一、信息及其分类	1
二、信息特征与信息表达	3
三、医学信息	8
第2节 医疗信息系统	10
一、系统和系统工程	10
二、医疗信息系统组成	12
第3节 建设具有中国特色的医疗信息系统	15
一、医疗信息系统建设的主要发展阶段	15
二、美国通过“互操作”大力推进医疗信息化发展的经验	16
三、中国医疗信息化发展状况	18
第2章 医疗信息系统的支撑技术	20
第1节 医疗信息系统的网络与通信技术	20
一、医疗信息系统的网络布局	20
二、网络连接	25
三、网络通信	32
第2节 医疗信息系统的数据库技术	36
一、数据模型与数据库	37
二、数据库新技术	40
第3节 医疗信息系统开发平台与建模	43
一、医疗信息系统开发平台	43
二、医疗信息系统建模	45
第4节 医疗信息系统的数据安全	46
一、网络安全威胁	46
二、医疗信息系统安全	47
第3章 医疗信息系统的规划与管理	49
第1节 医疗信息系统的规划与需求分析	49
一、系统规划	50
二、医疗信息系统开发的需求分析	53

第2节 医疗信息系统的开发与实现	60
一、面向过程的信息系统开发方法	61
二、面向对象的信息系统开发方法	63
三、其他信息系统开发方法	65
四、医疗信息系统的实现	68
第3节 医疗信息系统评价与运行	69
一、医疗信息系统评价	69
二、医疗信息系统测试与维护	71
第4节 医疗信息系统主要子系统	73
一、基于局域网的医疗信息系统	73
二、基于广域网的医疗信息系统	77
第4章 与医疗信息系统相关的标准	82
第1节 标准与标准化组织	82
一、标准	82
二、标准化组织	85
第2节 医疗信息传输与交换标准	86
一、美国卫生信息传输标准	87
二、医学数字成像和通信标准	89
三、医疗信息集成模型 IHE	93
四、ANSI X12 与 IEEE 其他标准	96
第3节 临床数据标准	97
一、统一医学语言系统	97
二、电子病历标准	99
第4节 国内外医疗信息标准化现状及发展趋势	101
第5章 医学影像存储与传输系统及其数据库的开发实现	105
第1节 医疗信息系统中的影像信息	105
一、影像设备与影像信息采集	106
二、影像信息数据及其时间分布	109
三、医学影像传输、显示与存储	111
第2节 PACS 及其存储机制	119
一、PACS 的组成	120
二、PACS 中的存储系统	123
三、PACS 数据库的建立	126
第3节 兼容 DICOM 的 PACS 数据库实现技术	127
一、基于 DICOM 的医学影像数据库数据组织技术	127
二、基于 DICOM 的医学影像数据库数据存储技术	130
三、医学影像数据库数据索引技术	136

第 6 章 电子病历及其信息集成	139
第 1 节 电子病历与电子病历系统	139
一、电子病历及其特点	139
二、电子病历系统	140
第 2 节 电子病历结构化	141
一、电子病历的文档表示	141
二、电子病历的结构化数据生成	147
第 3 节 电子病历管理	150
一、基于 Oracle XML DB 的电子病历临床文档管理	150
二、面向叙述性内容的电子病历检索	153
第 4 节 电子病历信息集成系统	156
一、电子病历信息集成系统结构	156
二、病历文档分类及编码	157
三、系统及其向区域合作医疗的扩展	158
第 5 节 基于 XML 与 HL 7 的电子病历信息交换	159
一、交换平台与交换引擎	159
二、交换信息中心与网关	160
三、用 HL7 DTD 实现 HL 7 消息与 XML 语句的转换	162
第 7 章 临床决策支持系统	164
第 1 节 CDSS 的功能与发展	164
一、CDSS 的分类与功能	164
二、CDSS 的特点	166
三、CDSS 的发展趋势	168
第 2 节 构建 CDSS 的方法	170
一、贝叶斯概率方法	170
二、决策树方法	171
三、神经网络方法	176
四、遗传算法与粗糙集方法	178
第 3 节 CDSS 中的推理规则	182
一、CDSS 中的规则推理	182
二、CDSS 案例推理	184
第 4 节 DW + OLAP + DM 医学决策支持技术简介	185
第 8 章 客户关系管理系统辅助医疗管理系统的设计与应用	189
第 1 节 CRM 相关理论	190
一、CRM 产生的背景与特征	190
二、CRM 的体系框架分析	194
三、CRM 数据库分析与技术应用	196

第2节 CRM 辅助医疗管理系统的设计 199

一、医疗辅助管理系统业务流程分析与体系结构设计..... 200

二、系统功能模块设计..... 201

三、系统的安全性设计..... 203

第3节 CRM 辅助医疗管理系统的实现 204

一、主要功能模块实现..... 204

二、CRM 辅助医疗管理系统的测试 207

第1章 概 述

人的眼、耳、鼻、舌、身都可以感知信息。信息自古就有，但受古代社会文明发展水平的制约，人们对信息的重视程度不够。随着社会的进步，人们对信息的依赖程度越来越高，信息科学、材料科学和能源科学一起被誉为当代文明的“三大支柱”。

第1节 信息与医学信息

信息的概念十分广泛，信息的定义有上百种。从本质上说，信息是事物自身显示其存在方式和运动状态的属性，是客观存在的事物现象。

一、信息及其分类

（一）信息的内涵

“信息”一词由来已久。国内最早的记录见于公元3世纪的《三国志》，国外最早的记录见于经济学家 Veblen 于1919年出版的《资本的性质》一书。但学术界公认的是1948年由信息论的奠基人 C Shannon 在《通信的数学理论》中首次提出的信息定义——“信息是用来消除不确定性的东西”。

随着社会不断进步与发展，信息理论在20世纪中后期获得空前发展：它涉及数学、通信理论、控制论、计算机科学、人工智能、电子学、自动化技术及物理、生命科学等多个领域，因而也就决定了信息在不同的学科范畴中具有不同的含义和特征。

通信学家认为信息是不定度的描述；数学家认为信息是概率论的发展；物理学家认为信息是熵的理论；哲学家认为信息是认识论的一部分；管理学家认为信息是提供决策的有效数据；情报学家则认为信息是生物及自控系统与外界交换的一切内容。

据不完全统计，目前有关信息的定义已逾百种。这一方面是由于学术界的研究人员因其各自所处的观察与研究角度及侧重点不一样而造成的；另一方面也说明信息概念的确很复杂，它是一个多维、多层次含义的概念。

当今学术界有两种观点能为大多数人所接受：第一种观点认为信息是事物运动的状态和方式，即信息不是事物本身，但反映了事物的特征与特性；不同事物有不同的特征，并在不同的条件下发生变化，这种特征与变化就是信息。第二种观点认为信息就是一组具有意义的事实或数据。

（二）信息与物质和能量的关系

信息与物质、能量是有区别的，同时信息与物质、能量之间又存在着密切的联系。缺少物质的世界是空虚的世界，缺少能量的世界是死寂的世界，缺少信息的世界是混乱的世界。

物质是信息存在的基础。信息是一切物质的基本属性，认知主体对于客观物质世界的反

映都是通过信息来实现的，但信息不是物质，也不是意识，而是物质与意识的中介；信息的产生、表述、存储、传递等都要以物质为基础，但物质具有质量，且遵循质量守恒定律，而信息本身没有质量，也不服从守恒定律；信息对物质有依附性，任何信息都离不开物质，都要以物质作为载体，但信息内容可以共享，其性质与物质载体的变换无关。

能量是信息运动的动力。信息的传递、转换、获取、利用过程都要耗费一定的能量。信息必须与能量结合才具有活力，但信息效用的大小并不由其消耗的能量决定；各种形式的能量或信息在传递过程中都可以互相转换，但能量的传递与转换过程遵循能量守恒定律，而信息在传递与转换过程中并不服从守恒定律；信息的传递与获取离不开能量，能量的驾驭和转换则又需要信息。

物质可为人类提供材料，能量可为人类提供动力，信息可为人类提供知识和智慧。对于一个系统来说，物质使系统具有形体，能量使系统具有活力，信息则使系统具有灵魂。只有三者的有机结合，才能使系统真正发挥其功能，朝着进步的方向演化。

正是信息使得自然界、人类社会和任何一个单位、组织成为有序的整体。表 1-1 表明了信息的重要性。

表 1-1 信息的重要性

成 分	涉及活动	危机的时间
事务单据	作业	几小时至几天
作业报告	作业控制	几天至几周
计划和控制报告	管理控制	几周至几个月
长期发展报告	战略计划	几个月至几年

(三) 信息分类

如果用不同的标准对信息进行分类，可以把信息划分为以下几种类型。

1. 按照信息的发生领域，可将信息划分为物理信息、生物信息和社会信息。

(1) 物理信息：指无生命世界的信息。无生命的世界每时每刻都在散发着大量的信息（形形色色的气候变化、地球运动、天体演化等）。只是由于条件的限制，人们对于这类信息现象的认识还远远不够。

(2) 生物信息：指生命世界的信息。生物信息中的遗传信息就是生命进化的重要原因。没有信息，就没有丰富多彩的生物界，更不会出现人类社会。

(3) 社会信息：指社会上人与人之间交流的信息，包括一切人类社会运动变化状态的描述。按照其活动领域，社会信息又可分为科技信息、医学信息、经济信息、政治信息、文化信息等。社会信息是人类社会活动的重要资源，也是社会大系统的一类构成要素和演化动力。

2. 按照信息的表现形式，可将信息划分为消息、资料 and 知识。

(1) 消息是关于客观事物发展变化情况的最新报道。消息反映的是事物当前的动态的信息，因此生存期短暂，有较强的时间性，主要用于了解情况和行为决策。

(2) 资料是客观事物的静态描述与社会现象的原始记录。资料反映的是客观真实的记载，因此生存期长久，有较强的累积性，主要用作论证的依据。

(3) 知识是人类社会实践经验的总结，是人类发现、发明与创造的成果。知识反映的

是人类对客观事物的普遍认识和科学评价,因此对人类社会活动有重要的意义。人们通过学习掌握知识,可以增长创造才能,提高决策水平,更有效地开展各项社会活动。

3. 按照性质的不同,可将信息划分为语法信息、语义信息和语用信息。

从主体对信息的认识层次上看,由于主体有感受力,能够感知事物运动状态及其变化方式的外在形式,由此获得的信息称为语法信息;由于主体有理解力,能够领会事物运动状态及其变化方式的逻辑含义,由此获得的信息称为语义信息;又由于主体具有明确的目的性,能够判断事物运动状态及其变化方式的效用,由此获得的信息称为语用信息。语法信息、语义信息和语用信息三位一体的综合,构成了认识论层次上的全部信息(全信息)。

(1) 语法信息是信息认识过程的第一个层次。它只反映事物的存在方式和运动状态,而不考虑信息的内涵。语法信息只是客观事物形式上的单纯描述,只表现事物的现象而不深入揭示事物发展变化的内涵及其意义。这一层次涉及符号的数目、信源(信息的发生方)的统计性质、编码系统、信道(媒介)容量等,主要研究信道传递信息的能力,设计合适的编码系统,以高度的可靠性快速有效地传递数据。

(2) 语义信息是信息认识过程的第二个层次。它是指认识主体所感知或所表述的事物的存在方式和运动状态的逻辑含义;语义信息不仅反映事物运动变化的状态,而且还要揭示事物运动变化的意义。从信源发出的数则消息,如果只是从通信符号的统计数量来看,其信息量可能相等,但信息量相等的消息其意义却可以是完全不同的。在信息检索中就要考虑到信息的语义问题。

(3) 语用信息是信息认识过程的最高层次。它是指认识主体所感知或所表述的事物存在方式和运动状态,相对于某种目的所具有的效用。语用信息就是指信源所发出的信息被信宿(信息的接收方)接收后将产生的效果和作用。同语义信息相比,它对信宿的依赖性更强,而且与信息传递时间、地点、环境条件等有着密切的关系。

另外,按照信息载体的性质,可将信息分为语音信息、图像信息、文字信息、电磁信息和光电信息等。按照携带信息的信号的性质,可将信息分为连续信息、离散信息和半连续信息等。

按照事物运动的方式,还可将信息进一步划分为概率信息、偶发信息、明确信息和模糊信息等,见图1-1。本书所涉及的信息主要指概率信息。

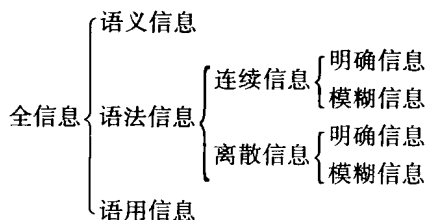


图 1-1 不同性质的信息

二、信息特征与信息表达

信息的特征指信息区别于其他事物的本质属性。信息的基本属性是可传递性。信息传递指通过一定媒介使信息从时间或空间的一点向另一点移动的行为。信息交流就是指各个个体借助于共同的符号系统所进行的信息传递、交换与分享。

（一）信息特征

1. 普遍性 信息是事物运动的状态和方式。只要有事物存在，只要有事物的运动，就会有其运动的状态和方式，就存在着信息。信息与物质、能量一起，构成了客观世界的三大要素。

2. 表征性 信息不是客观事物本身，而只是事物运动状态和存在方式的表征。一切事物都会产生信息，信息就是表征所有事物属性、状态、内在联系与相互作用的一种普遍形式。

3. 动态性 客观事物本身都在不停地运动变化，信息也在不断发展更新。特别是从语用信息的观点来看，事物运动状态及方式的效用是会随时间的推移而改变的。因此，在获取与利用信息时必须树立时效观念，不能一劳永逸。

4. 相对性 客观上信息是无限的，但相对于认知主体来说，人们实际获得的信息（实得信息）总是有限的。并且，由于不同主体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（语法信息、语义信息和语用信息）肯定各不相同，即实得信息量是因人而异的。

5. 依存性 信息本身是看不见、摸不着的，它必须依附于一定的物质形式（如声波、电磁波、纸张、化学材料、磁性材料等）之上，不可能脱离物质单独存在。把这些以承载信息为主要任务的物质形式称为信息的载体。信息没有语言、文字、图像、符号等记录手段便不能表述，没有物质载体便不能存储和传播，但其内容并不因记录手段或物质载体的改变而发生变化。

6. 可传递性 信息可以通过多种渠道、采用多种方式进行传递，一般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递。信息传递要借助于一定的物质载体，因此，实现信息传递功能的载体又称为信息媒介。

一个完整的信息传递过程必须具备信源、信宿、信道和信息 4 个要素。

7. 可干扰性 信息是通过信道进行传递的。信道既是通信系统不可缺少的组成部分，同时又对信息传递有干扰和阻碍作用。

一般把任何不属于信源原意而加之于其信号上的附加物都称为信息干扰。

8. 可加工性 信息可以被分析或综合，可以被扩充或浓缩，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。如果在信息加工过程中没有任何信息量的增加或损失，并且信息内容保持不变，那么就意味着这个信息加工过程是可逆的，反之则是不可逆的（实际上信息加工都是不可逆的过程）。

9. 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占有和共同享用，也即信息在传递过程中不但可以被信源和信宿共同拥有，而且还可以被众多的信宿同时接收利用。物质交换遵循易物交换原则，失去一物才能得到一物；信息交换的双方不仅不会失去原有信息，而且还会增加新的信息。

（二）信息表达

信息交流过程涉及 4 个基本要素：信息的发出者（信源）、信息的接收者（信宿）、交流手段（媒介）和交流对象（信息）。从本质上说，当信息交流双方凭借一定的媒介参加交流活动时，就形成了一种交流关系，在这种关系中，共享的是表述信息的符号，而未必是信息本身。因此，信息交流也可以说是符号的交流。

1. 符号 信息是事物的表征,但表征事物的信息本身是不能直接为人们所认识的。实际上,它是通过代表事物的代码——符号来传达的。

符号必须能够表达意义。没有意义,就没有符号;没有符号,意义也无从表现。当然,符号并不是完美无缺的表达工具^①。一切符号在形式上都带有随意性。

符号形式与符号内容在逻辑上是互为依存的关系,但两者在信息交流中并不一定总是对称的。符号形式和符号内容之间相互依存关系的对称运动只限于假设二者以“理想的”代码为依据的场合,例如 Morse 电码对各种符号都十分明确地规定了其符号形式和符号内容,并使二者之间具有排他性的一一对应关系。

符号的类型是多种多样的。人类在生活和生产劳动过程中,为了方便表达,为了在不同情况下满足不同的需要,创造了许多符号系统。人类使用的符号系统大致可以分为两大类:一是语言符号系统,人们通过发音、书写、键盘输入等口头的或字符的形式进行交流。其中口语是有声的符号,文字是无声的符号;二是非语言符号系统,人们可以运用面部表情、手势或其他身体动作,时间、空间距离,色彩、服饰、气味,声响、图形、图像和器物,以及音乐、舞蹈、美术作品等艺术形式进行交流。它也可以分为有声的和无声的两类^②。

2. 编码 编码是把一定的意义、信息转换为代码的过程。所谓“码”,就是按照一定规则排列起来的符号或信号序列。通过编码,人类的思想才得以交流,信息才能从一个系统传递到另一个系统。编码的功能主要体现在三个方面:一是可以简化信息传递的形式,以提高信息传递的效率和准确度;二是可以对信息单元的识别提供一个简单、清晰的代号,以便于信息的存储与检索;三是可以显示信息单元的重要意义,以协助信息的选择和操作。

通过与编码相反的过程可以把这些代码还原成编码前的信息形式。这个过程就称为解码或译码。两个系统必须具有共同的编码和译码机制才能进行信息交流。

编码最初是指把文字变换成由点、划和间隔组成的电码(如著名的 Morse 电码就是其中的一种)。后来,人们把编码的概念推广到整个信息表述与传达过程中。人们几乎可以运用声、光、电等世间一切物质运动形式对信息进行编码,以形成信息交流系统。对人类的信息交流系统来说,存在着两个层次的编码活动:人们用语言、文字及其他辅助性的非语言文字符号来表达一定的思想内容的过程是初始编码。通过初始编码,可以建立需要表达的思想内容与音素、字母或其他事物之间的确定联系,即把思想内容用一段话、一段文字或其他符号表现出来。为了有效地进行信息交流,人们还建立和发展了多种多样的通信系统。在通信系统中,通过某种机械装置(编码器),把信息变换成适合在信道上传输的信号的过程就是二次编码。若将信号理解为对于时间的物理变量,那么二次编码实际上就是把一定形式的符号变换为另一种形式的信号序列,二次编码又可分为信源编码和信道编码两种方式,其目的都是把符号变换成便于在信道中传递的形式。

编码的基本原则是:① 编码的方法必须是合理的,能够适合使用者和信息处理的需要;② 编码时要预留足够的位置,以适应需要的发展变化,即具有可扩展性;③ 每一编码都能代表一个确定的信息内容,每一信息都有一个确定的编码表示;④ 编码的结构必须易为使

① 任何一套符号都不能把一个人的全部感觉和内部的所有活动表达出来,虽然他可能把全部身心状态投入到编制符号的任务之中。

② 正是由于人们在进行信息交流时大多同时采用几种符号系统,人类社会的信息交流活动才会如此丰富多彩。这同时也增加了人们理解与沟通的难度。

用者了解和掌握；⑤ 必须建立和不断完善编码标准化体系，以避免混乱和误解。

(1) 初始编码——普通语义学的解释：初始编码是人们在一定的文化历史背景下形成的，它只在特定的文化群体中有意义。一旦越过文化边界与另一文化交流，就会出现种种麻烦。

(2) 二次编码——通信工程问题：通信系统的基本问题是要解决信息交流的有效性与可靠性这两个方面的问题，即以最大的速率传递信息，而且要保证在干扰存在的条件下，能够准确地和最佳地再现信息，这就涉及整个通信过程，包括信源、信宿、信道以及编码问题。

信源问题：信源问题的核心是信源包含的信息到底有多少？这要做定量描述，用信源熵 $H(x)$ 来测度。

信宿问题：就是接收者能收到多少或提取多少由信源发来的信息量 $I(p)$ 。这里不涉及信息的意义以及对收信人是否有用的问题。

信道问题：主要是信道最多能传送或存储多少信息量，即所谓信道容量的问题。从通信的有效性出发，它要求以最大的通信速度传送最大的信息量。

信道在单位时间内传递的信息量称为传信率。在一定的信道中，它随信源性质和编码方式而变化。信道在单位时间里所能传递的最大信息量称为信道容量 C 。

为提高通信系统的有效性与可靠性，在二次编码中所使用的编码规则应满足下列条件：

(1) 保证在编码时所使用的代码序列中码元（即所用的代码组合）尽可能少。码元少，则信道一定时，信道容量就大，从而在这种信道容量下，可提高信息传输的有效性。凡是能载荷一定的信息量，且码元的平均长度最短，可分离的变长码的码元集合称为最佳码。为此必须使出现概率大的信息符号编以短字长的码元，对于概率小的符号编以长字长的码元，则平均字长一定最短。由于一般信源都存在较大的冗余度，所以信源编码的主要任务是保证适当的编码，借助减小冗余度的办法来降低信息符号所需的平均码元数，提高编码效率。

对于给定的信道，如果传信率小于信道容量，又没有噪声干扰，就可以找到一种编码方法使信息准确无误地传递。反之，如果传信率大于信道容量，那么必然会产生差错。采用不同的编码方式可以有不同的传信率，因此可通过选择传信率最高的编码方案来最有效地利用信道。

(2) 要求通过编码后，能从编码序列中无失真（或限定失真大小）地恢复原始符号序列的概率为最大，其目的是为了无失真或失真较小地传输信息，以提高信息传输的可靠性。信息传递一般都有干扰存在，为了增加通信的可靠性，必须采用具有一定抗干扰能力的编码。而通信系统的有效性与可靠性是相互矛盾的，Shannon 提出的有噪声信道编码定理（即 Shannon 第二定理^①）巧妙地使这个矛盾得到了统一。

现实的信息系统是由许多信道组成的，不管在哪一信道，当传信率大于信道容量时，就会出现差错，减少有效信息量。因此，多信道信息系统有一个相互匹配的问题。

（三）信息量

信息是一种重要的资源，为便于管理和处理就需要对其进行科学的度量。但由于信息本身具有虚拟性，这决定了信息的度量要比物质和能源的度量困难得多。

^① Shannon 第二定理说明，在有噪声信道中，采用纠错码等方法，用有冗余度的方式来传递信息，可以降低产生误差的概率。

1. 信息量的概念 信息量可以表示事物不确定性的多少。对于要决策的事情，常常存在多种可能性。为减少不确定性，就需要获取信息。随着信息的获得，由多种可能性逐渐变为少数几种可能性，以便尽可能获得准确的信息，为最终决策提供最准确的依据。

实际上，在做决策时之所以犹豫不决，就是因为事情发展存在很多不确定性，也就是决策的结果存在很多种可能性。越是不确定的事情越是难以决策，也就需要获得更多的信息来减少这种不确定性。这就是信息对决策的决定作用。因此，可以用这种不确定性的多少来表示信息的量。正是基于这一点，Shannon 把信息定义为“事物不确定性的减小”，也就是说信息作用的本质是减少事务的不确定性。因此，不确定性的多少就可以作为信息量的一种度量方法。信息的这种不确定性也称为信息熵。

熵的概念来源于热力学的研究。早在 1850 年，R J E Clausius 就发现，对于一个热力系统，虽然能量是守恒的，但能做功的能量并不守恒。他将这种无效热能称为熵，并发明了一个新的词汇表示熵（entropy）。1865 年，Clausius 提出，对一个孤立的系统，其熵只能增加。这个原理就是著名的热力学第二定律，也被称为熵增原理。

后来科学家发现，热力学第二定律不仅对热力学系统有效，并且可以在其他各种系统中发现熵的影响^①。这项研究结果震动了整个科学界。

当一件事情存在很多种可能性时，所表现出的也是一种混乱、无序的状态。随着信息的不断获取，情况逐渐清晰，逐渐变为有序状态。因此，Shannon 用熵来表示事物的不确定性，即信息量的多少。N Wiener 则认为可以用熵来定义信息：“信息即负熵。”

2. 信息量的计算 信息量的计算公式如式（1-1）所示：

$$I = \log \frac{\text{后验概率}}{\text{先验概率}} \quad (1-1)$$

式（1-1）中，后验概率表示事情发生后的概率，这时结果是确定的，因此后验概率为 1。先验概率是在计算信息量时，该事情估计的可能性的多少和概率。因现在数据处理都是在计算机中以二进制进行，所以在计算对数时一般是以 2 为底。这时计算出的信息量的值实际为二进制位（bit，b）的值。

若有 n 种可能不等概率，则可得出一般情况的信息量计算公式为

$$I = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i} = - \sum_{i=1}^n P_i \log_2 P_i \quad (1-2)$$

式（1-2）中， i 表示 $1 \sim n$ 可能性的数目； P_i 表示第 i 种可能出现的概率。

3. 信息的冗余度 信息的冗余（redundancy）指用于表达信息的量和实际信息量间的差值。冗余是信息存储管理和信息系统设计中的一个重要概念。在信息系统中，一方面为了减少信息的存储容量和避免信息的不一致性，要尽可能减少信息的冗余；另一方面，有些情况下信息的冗余（如为了检验信息的传输是否正确所加入的冗余位等）又是必需的，这时冗余的信息就不是多余和无用的信息了。合理地处理数据的冗余也是信息系统设计者要考虑的问题之一。

信息的冗余度的计算公式如式（1-3）所示：

$$I = 1 - \frac{I_n}{I_m} \quad (1-3)$$

^① 诺贝尔奖得主 I Prigogine 说：“我们科学遗产中包括两个至今未得到答案的基本问题，无序与有序的关系以及什么是熵。” A Einstein 甚至断言：“熵理论，对于整个科学来说是第一法则”。足见熵理论在科学发展史中的重要影响。