



普通高等教育“十一五”国家级规划教材

普通高等教育信息管理类专业规划教材

信息系统 分析与设计

向阳 主编

Information System Analysis and Design



机械工业出版社
CHINA MACHINE PRESS





普通高等教育“十一五”国家级规划教材

普通高等教育计算机专业系列教材

信息系统 分析与设计

第二版



Information System Analysis and Design

机械工业出版社



G202

X281.1

普通高等教育“十一五”国家级规划教材
普通高等教育信息管理类专业规划教材

信息系统分析与设计

主 编	向 阳	
副主编	崔振洪	韩丽川
	赵卫东	
参 编	于长锐	蒋元涛
	赵秀丽	赵英姝
主 审	韩耀军	沈惠璋



机械工业出版社

本书从信息系统基本概念出发,结合当前信息技术发展状况,全面、系统地介绍了信息系统分析与设计的理论、方法和工具。全书共分九章,包括概述、信息系统技术基础、系统规划、系统分析、系统设计、面向对象分析与设计、系统开发与实施、系统评价与运行管理和信息系统新发展。书中主要章节给出了大量案例,章后有思考题、设计题和案例分析,可以帮助读者加深对信息系统基本概念和方法的理解,促进读者对开发工具的掌握。

本书是高等院校信息管理与信息系统本科专业核心课程的教学用书,可作为高等院校信息管理与信息系统及其相关专业学生的教材,也可以供信息系统项目开发人员和项目管理人员参考。本书配有电子教案,需要的教师可到机械工业出版社网站下载。

图书在版编目(CIP)数据

信息系统分析与设计/向阳主编. —北京:机械工业出版社, 2007. 5

普通高等教育“十一五”国家级规划教材. 普通高等教育信息管理类专业规划教材

ISBN 978-7-111-21534-9

I. 信… II. 向… III. ①信息系统—系统分析—高等学校—教材 ②信息系统—系统设计—高等学校—教材 IV. G202

中国版本图书馆CIP数据核字(2007)第073614号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:易敏 版式设计:霍永明 责任校对:刘志文

封面设计:刘科 责任印制:李妍

保定市中画美凯印刷有限公司印刷

2007年7月第1版第1次印刷

169mm×239mm·12.875印张·542千字

标准书号:ISBN 978-7-111-21534-9

定价:32.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

销售服务热线电话:(010)68326294

购书热线电话:(010)88379639 88379641 88379643

编辑热线电话:(010)88379721

封面无防伪标均为盗版

普通高等教育信息管理类专业 规划教材编审委员会

名誉主任：黄梯云 哈尔滨工业大学
主任：李一军 哈尔滨工业大学
副主任：王刊良 西安交通大学
马永开 电子科技大学
姚飞 北京化工大学
卢虎生 内蒙古科技大学
卢才武 西安建筑科技大学
邵军义 青岛理工大学
杨铭 齐齐哈尔大学
邓海平 机械工业出版社

委员（按姓氏笔画排序）：
于俭 杭州科技大学
王成 齐齐哈尔大学
王亮 青岛理工大学
邓晓红 山东建筑大学
田军 西安交通大学
向阳 同济大学
孙军 北京化工大学
李四福 中国地质大学
李志刚 成都理工大学
张宏山 河南科技大学
张喜征 湖南大学
邵培基 电子科技大学
骆正山 西安建筑科技大学
郝戊 内蒙古科技大学
贾红雨 大连海事大学
郭士正 集美大学
翟东升 北京工业大学
蔡建峰 西北工业大学

秘书：易敏 机械工业出版社

前 言

信息系统的开发是集计算机技术、项目管理、通信技术等为一体的信息系统工程。它不但能为企业提供更加准确、及时、全面、详实的数据,而且便于对各种信息进行进一步的加工。另外,它也有利于进一步加强企业管理的科学化、合理化、制度化、规范化。随着互联网的普及应用,飞速发展的网络技术对信息系统的建设和应用也产生了很大的促进作用,其中最具有代表性的发展分支就是以内联网为形式的企业信息系统和以互联网为基础的电子商务系统。由于网络的推进,使得信息系统在信息共享性、实时性、可靠性、快速性等方面有了显著的提升。

本书紧密结合当前网络技术的发展,把网络技术贯穿于系统规划、系统分析、系统设计和系统实施各章节,紧跟技术发展趋势。为了使读者能够对信息系统开发的各个环节有更加感性的认识,我们还引入了当前比较流行的开发和开发辅助工具,如 SQL Server 数据库、Microsoft Visio、ERwin、Microsoft Project 等,读者可以利用这些工具亲自动手进行信息系统的分析与设计。此外,本书还附带了大量案例分析,将抽象的理论知识具体化,避免学生学习本课程有过于空虚的感觉。

本书共分为九章,依次介绍了信息系统概况、信息系统技术基础、信息系统规划、信息系统分析、信息系统设计、面向对象分析与设计、系统开发与实施、系统评价与运行管理、信息系统新发展,内容涵盖了信息系统分析与设计的各个方面。

本书可作为大专院校信息管理及其相关专业本科学生的教材,也可以供软件开发人员和开发项目管理参考。

本书由同济大学向阳教授主编,参加编写的有齐齐哈尔大学崔振洪副教授、上海交通大学韩丽川副教授、山东科技大学赵伟东副教授、上海交通大学于长锐博士后、上海海事大学蒋元涛讲师、山东轻工业学院赵秀丽讲师、齐齐哈尔大学赵英姝讲师。其中,第一章由向阳编写,第二章由向阳和于长锐编写,第三章由韩丽川编写,第四章由崔振洪编写,第五章由赵英姝编写,第六章由赵伟东编写,第七章由于长锐编写,第八章由赵秀丽编写,第九章由蒋元涛编写。

本书的编写过程中,硕士研究生马强、黄蓉、韩婕、王敏、段韧、王连诚、李明田、李雨平、张环宇、万蓬、吕文涛、刘磊、刘昌劲、宋远芳、田刚、赵成、于凤山等同学参与了资料收集、稿件整理和课件制作等工作。在此,对他

们付出的劳动表示感谢。

本书作者制作了配套课件，使用本书作教材的教师可与本书编辑联系索取 (yimin@ mail. machineinfo. gov. cn; yimin9721@ 163. com)。

编 者

目 录

前言

第一章 概述	1
第一节 网络信息时代企业生存与发展的环境	1
第二节 信息系统的概念	4
第三节 信息系统沿革	14
第四节 信息系统开发生命周期	17
第五节 信息系统开发方法与方式	20
思考题	29
习题	29
第二章 信息系统技术基础	30
第一节 信息系统开发环境的建立	30
第二节 数据库技术	37
第三节 系统分析技术	65
第四节 系统设计技术	75
第五节 系统开发中的项目管理技术	83
思考题	91
习题	91
第三章 系统规划	93
第一节 网络环境下企业的信息系统发展战略	93
第二节 信息系统规划的内容、特点与作用	97
第三节 信息系统规划的组织	100
第四节 信息系统规划的方法	102
第五节 信息系统规划报告的撰写与审查	107
第六节 系统规划案例	109
思考题	114
案例分析	114
第四章 系统分析	116
第一节 系统分析概述	116
第二节 可行性分析	122
第三节 业务分析	130

第四节	数据流程分析	143
第五节	数据字典	153
第六节	新系统逻辑模型	157
第七节	系统分析报告	162
第八节	系统分析案例	164
思考题	172
设计题	173
第五章	系统设计	174
第一节	系统设计的原则、特点、任务和步骤	174
第二节	网络化环境下的计算机系统设计	178
第三节	系统功能及其结构设计	181
第四节	系统功能模块及其处理设计	185
第五节	数据库设计	196
第六节	编码设计	214
第七节	输入/输出设计	217
第八节	系统设计报告	223
第九节	系统设计案例	225
思考题	233
设计题	234
第六章	面向对象分析与设计	235
第一节	面向对象方法的形成与发展	235
第二节	面向对象方法的主要概念	237
第三节	面向对象分析与设计工具——统一建模语言 UML	242
第四节	面向对象分析	259
第五节	面向对象设计	278
第六节	面向对象分析与设计案例	289
思考题	300
设计题	300
第七章	系统开发与实施	301
第一节	系统开发的组织工作	301
第二节	系统开发的项目管理	306
第三节	系统开发的文档管理	319
第四节	系统开发监理	324
第五节	程序设计	333
第六节	系统测试	342

第七节 系统切换	348
思考题	351
第八章 系统评价与运行管理	352
第一节 系统评价	352
第二节 系统运行管理	355
思考题	366
第九章 信息系统新发展	368
第一节 企业业务流程重组	368
第二节 ERP、CIMS、SCM 与 CRM	371
第三节 智能信息系统	380
第四节 数据仓库与数据挖掘	387
思考题	397
案例分析	397
参考文献	400

第一章

概述

内容提要

1. 概括性地介绍了网络时代企业生存与发展的内外部环境状况。
2. 详细给出了信息、系统、信息系统的基本概念。
3. 全面介绍了信息系统各发展阶段的特点和内容，以及信息系统的种类。
4. 介绍了信息系统开发生命周期各阶段及其主要内容。
5. 简要介绍了信息系统的开发方法，给出了各种信息系统的开发方式。

第一节 网络信息时代企业生存与发展的环境

20世纪90年代，互联网的迅猛发展使人类社会进入了一个网络处理的时代，全球都开始了一个网络化的发展进程。网络化对信息系统的建设和应用也产生了很大的促进作用，其中最有代表性的发展分支就是以内联网为基础的企业信息系统和以互联网为基础的电子商务系统。正是由于网络的推进作用，使得信息系统的应用范围从传统的只能处理管理问题扩展到了能够处理经营问题，也正是这种网络化的大环境，使得现代企业面临的内外环境产生了很大变化。

一、网络信息时代企业生存与发展的外部环境

(一) 世界经济全球化

当今世界，全球化贸易正在成为世界经济增长的主要动力。头号经济强国美国的经济增长率的提高主要依靠进出口，欧洲和亚洲的发达国家工业经济增长也是如此。对外贸易占美国提供商品和服务的25%之多，像日本和德国这样的国家比例还要更高，而且这个百分比还将继续增长。波音公司在全球有上万个供应商，最远的零件供应甚至要跨越太平洋。跨国公司进行全球贸易当然不是为了体验如今运输发展的给人们带来的便利，而是通过这种世界范围内的采购和销售，利用各地的相对优势得到最低的成本和最大的收益。而这种经济全球化的大环境极大地增加了信息对企业的价值，并为企业提供了生产经营的新机遇。企业可以快速地捕捉到瞬息万

变的有利信息，利用各地商品价格和成本的差异可以获得惊人的收益。为了维持现有的生存地位，为了做大做强，企业就要努力走向全球化，实行全球战略，成为世界产业链中的一个环节。此时，企业不得不开发全球信息系统来跟踪订单、发货和付款，与分销商和供货商联络，一天24小时在不同的环境中经营。简而言之，管理和协调不同环境中的经营活动是企业面临的主要挑战，处在当前经济全球化的环境中，企业信息化的建设已成为企业生存和发展的必经之路。

(二) 知识经济时代的来临

当以信息处理为产品的微软公司的产值已超过美国三大汽车公司产值的总和，当凭借进行全球采购、出色协调管理来提供优质服务的连锁零售商沃尔玛跃居世界500强第一位，当依靠新技术“白手”起家的IT新贵们包揽了福布斯的前几名时，世界经济已经从工业型经济转变成基于知识和信息的服务型经济（A Knowledge and Information-Based Service Economy），即一个以知识为基础的经济时代已经来临，企业生产经营将进入了知识经济时代。

知识经济直接依赖于知识和信息的生产、扩散和应用。软件的发展、网络的产生、虚拟技术的应用，正在使企业资产中无形资产的比例不断增加。知识经济的来临将对我们的生产方式、生活方式、思维方式、管理决策产生重大影响，企业管理将由生产向创新转变，其经济效益将越来越依赖于知识和创新。一个企业要生存和发展，就必须依靠信息系统的支持，用动态的观点来研究面临的新问题。这就是为什么在经济发达国家的许多服务行业（金融、保险和房地产等）甚至将70%以上的投资都用于发展信息技术。

(三) 电子商务日益普及

Internet 在全球迅速普及，使得现代商业具有了不断增长的供货能力、不断增长的客户需求和不断增长的全球竞争三大特征。高速信息获取与处理使得传统的商业模式已经无法跟上现代经济发展。解决传统商务模式弊端的途径就是利用计算机技术、网络通信技术开展电子商务。

中国互联网络信息中心（CNNIC）发布的《第十五次中国互联网络发展状况统计报告》显示，截至2004年底，内地上网用户总数为9400万；《2004年中国B2B电子商务研究报告》显示，截至2004年底，中国电子商务市场规模已经达到了3239亿元人民币。据iResearch预计，到2007年，中国电子商务市场总体规模将会达到17373亿元人民币。这种状态说明，电子商务已成为企业生存和发展的环境中不可或缺的组成部分。

在这种情况下，任何不进行信息化改造的企业，就会被排除在网络平台之外，就无法生存。

(四) 网络信息技术飞速发展，网络信息时代已来临

网络信息技术的发展可以用日新月异这个词来形容，网络信息时代已来临。《第十五次中国互联网络发展状况统计报告》显示，经过十年快速的发展，中国互联网

已经形成规模,互联网应用走向多元化。人们在工作、学习和生活中越来越多地使用互联网,整个社会的运行都搭上了互联网的快车,并打上了互联网的烙印。

网络技术在企业的应用更是快速,Internet本身所具有的开放性、全球性、低成本、高效率的特点,极其适合于现在企业的运作。现在随着信息技术的飞速发展,越来越多的组织、企业、个人都被连接到网络空间,利用Internet组建企业专用的企业内部网(Intranet)和企业之间使用的外联网(Extranet)已势在必行。

二、网络信息时代企业生存与发展的内部环境

现代企业面临的外部环境正在发生巨大的改变,这种改变同时也正促使企业内部环境发生着显著的变化,具体表现在以下几个方面:

(一) 企业决策和监控所依赖的信息空前丰富

企业决策和监控所依赖的信息不仅仅是指企业内部的信息,而且包含大量的企业外部信息。在这个经济全球一体化的时代,这些外部信息包括国内和国际的信息,既包括政策法规、市场行情、客户信息、竞争对手信息、行业信息等企业外部信息,也包括企业生产信息、采购信息、库存信息、销售信息、成本信息、资金信息、人力资源信息等。就市场行情而言,在成千上万的竞争者中,新产品、新价格、新服务、新技术、新热点层出不穷,千变万化。企业能获得的信息空前丰富,用海量来形容一点不为过。面对瞬息万变的国内外环境信息,企业决策与监控所需信息空前丰富。在这种状态下,企业单靠传统手工方式收集和處理信息是行不通的,传统的通信方式也经常难以奏效。只能通过采用先进的通信技术,建立高速的大容量的信息采集、加工、传输、处理和展示系统才能满足需求。

(二) 企业内部业务流程信息化已成为时代要求

在传统经济环境中,企业是围绕物流和资金流来组织生产,以必要的库存支持标准化的批量产品来提高生产效率的。而在信息经济环境中,企业的生产以消费者的需求为导向,只有能快速获取市场需求信息、及时组织生产,并迅速将产品投向市场的企业才能存活下去。企业为了驾驭市场,适应市场的需求,就必须借助电子信息技术建立企业内部快速反应机制,对企业采集的大量信息进行存储、加工、分析和预测,并及时作出正确反应,以指导企业的行为;根据需求订单及现有库存信息,合理组织采购,降低采购与库存成本;缩短产品设计、生产和流通的周期,灵活地进行生产任务和产品结构的调整,不断开发新产品,以适应小批量、多品种的市场需要;充分利用销售信息反馈,改善产品设计,强化信息传播在售后服务中的作用;优化管理,采用新工艺、新方法、新技术,加大技术改造力度,有效地利用生产资源,减少物耗,降低生产成本,建立有效的产品质量控制和工艺流程控制体系。这一切都离不开企业内部业务流程的信息化改造,使企业的信息流能有效、不间断、畅通、迅速地在企业经营活动中发挥作用。

（三）企业工作模式由手工向计算机化转变

据统计,在20世纪90年代,美国在财务会计上90%的工作由计算机完成,物资管理中80%~100%的信息处理由计算机完成,计划管理中则是80%~90%,计算机在很多领域正逐步代替手工劳动。我国虽然起步较晚,但近几年发展较快,特别是微型计算机与网络的普及,使计算机的应用更加广泛地渗透到企业生产经营的各个环节。企业在计划、生产、供应、销售、财务各环节的业务工作中,广泛使用相应的业务信息处理系统,企业已逐渐习惯了计算机代替手工劳动的局面,对信息系统功能的认识也越来越明确。

（四）企业职工计算机网络文化素质不断提高

随着计算机和网络的普及,加之各种操作系统非常友好的图形用户界面,计算机和网络的操作已不再高深莫测,企业职工大都掌握了它们的使用技术。从20世纪90年代以来,国家运用各种政策来推广计算机和网络知识,提高人们对现代信息技术的应用能力。企业职工计算机网络文化素质不断提高,也逐渐感受到现代信息技术给人们生活和工作带来的便捷,对信息系统的开发和运行也持较为欢迎的态度。

第二节 信息系统的概念

现代社会是网络信息社会。随着现代信息技术的快速发展,对信息的处理和利用已深入到人类生产生活的各个方面。人们对信息的处理和利用离不开信息系统,在很多场合都会与信息系统打交道,但要深刻理解信息系统的本质还需要从基本概念开始。

一、信息

（一）信息的定义

信息无处不在,无时不有,无人不用。今天,它已成为使用频率最高的词汇之一。人们对信息的研究和利用越广泛深入,对信息的认识和理解也就越多样化。不同的学科,从不同的角度对信息这个概念有不同的解释。

在经济学家眼里,信息是与物质和能量相并列的客观世界的三大要素之一,是生产力、竞争力和经济成就的关键因素,是社会生产所需要的中心资源。在新闻界,信息被普遍认为是对最新发生的事件的陈述。在图书馆学和情报学领域,信息被认为是被定义的事物(Thing)或记录(Record)。美国学者巴克兰德(M. Buckland)认为,许多事物都可以称为信息,如文本、图片、录音磁带、博物馆陈列品、自然物体、实验、事件等。信息论的创始人申农(C. E. Shannon),从通信系统理论的角度把信息定义为用来减少随机不确定性的东西。还有很多不同的学科都从自身的角度出发给信息予以不同的解释,这里就不一一列举。

据不完全统计,信息的定义有100多种,至今仍没有统一的、能为各界普遍认

同的定义。信息的定义之所以呈现多样化,主要有三方面的原因:第一,信息本身的复杂性,它是一个多元化、多层次、多功能的综合物;第二,信息科学是一门新兴的学科,它的许多分支学科仍在随着社会、经济和科学技术的发展而发展,其内涵和外延不很确切;第三,人们出于不同的研究和使用目的,从不同的角度或层次出发,对信息的概念也就会作出不同的解释。

从信息系统的角度出发,我们最常接触的是信息的处理过程,比如信息的收集、存储、传递、加工、维护、使用等。而这些处理过程中的信息都是借助于数据这个载体来体现。所以要深刻理解信息这个概念,还需要从认识数据这个概念入手。那么,数据到底是什么呢?

数据是记录客观事物的性质、形态、数量特征的抽象符号,例如文字、数字、图形、曲线等。信息是由数据产生的,是对数据加工处理后得到的对客观世界产生影响的数据,是反映具体客观事物特性的一些数据,是进行决策的依据。因此,我们把数据经过加工处理后有特定含义、对决策有价值的结果称为信息。例如财务活动中的数据,经过按财务规则加工处理后,形成有特定含义、供决策者决策参考的财务报表、账册等都是信息。

数据与信息的关系可描述为:数据是信息的载体,数据对决策者的行为没有影响;信息是数据处理的结果,信息影响决策者行为;高层次的数据可能是低层次的信息,低层次的数据只有加工处理后才能成为高层次的信息。

信息的产生过程如图 1-1 所示。

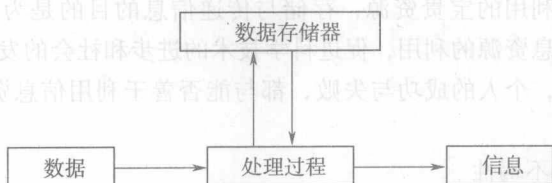


图 1-1 从数据转化为信息的过程

从图 1-1 可以看出,数据是客观事物的一种表现形式,信息是数据经加工处理后的结果,加工处理过程需要将数据从数据存储器中存取。

(二) 信息的特征

信息从不同角度理解,有不同的含义,但信息的特征是具有共性的。信息有如下特征:

1. 存储性

信息是可以存储的。人类不仅通过大脑记忆信息,还利用不同的载体存储信息。所谓载体,就是荷载信息的介质。一切记录信息的介质,如纸张、胶片、磁带、磁盘、光盘等,都是信息载体。正是通过信息的存储,保存了人类的文化遗产,保证了人类文明的延续,使人们在前人智慧的基础上发展和创造新的文明。

2. 传递性

信息是可以传递的。信息的传递有时间和空间传递之分。时间传递是信息通过一定的载体存储,使信息随时间的流逝而传递下去。空间传递是通过一定的方式把信息从一个地方传到另一个地方。由于信息的传递,才出现充满生机和千变万化的世界。人类为了自身的生存与发展,一直在探索和改变信息传递的方式。最初是靠声音和姿势,以后利用火光与符号。语言的出现,是人类传递方式的根本转折。文字的产生和印刷术的发明,使人类传递活动发生了质的变化。电报、电话、广播、电视等电子工具的出现,更使人类的信息传递方式发生了深刻的革命。计算机的使用、光纤纤维的问世、同步卫星的运转、Internet 的普及,使信息传递进入了一个新的时代。

3. 共享性

信息不仅可以传递,而且可以共享。信息与物质不同,不会因为一个人得到了,另一个人就失去了。英国文学家肖伯纳有一个形象的比喻:倘若你有一个苹果,我也有一个苹果,我们彼此交换后仍然各有一个苹果。但是,倘若你有一种思想,我也有一种思想,我们彼此交流这些思想,我们两个人各有两种思想。信息具有的这种特性,为信息的有效利用奠定了基础,极大地缩短了人类认识世界和改造世界的时间,也极大地节省了人力、物力和财力。人们追求信息的目的在于共享,在共享的基础上利用,在利用的基础上创造。

4. 可利用性

信息是一种可利用的宝贵资源,存储与传递信息的目的是为了利用信息资源。人们正是通过对信息资源的利用,促进科学技术的进步和社会的发展。在当今世界,国家的兴盛与衰落,个人的成功与失败,都与能否善于利用信息资源有着密切的联系。

5. 信息资源的不竭性

从整体上说,信息资源不会枯竭。人类生存的一切领域,随时都产生信息。物质世界是无限的,对物质世界的认识也是无限的,因此信息资源也是无限的。从另一个角度来看,人类的创造力是无限的,因此,描述和反映人类所创造出的事物的信息也是无限的。

6. 价值的不确定性

信息的价值体现在信息需要在需要的事情上、需要的时间内和恰当的地点里获得需要的信息,并指导其作出正确的决策。由于人们对信息的需求和理解及判断能力的不同,信息对人们行动的指导作用也是不同的,信息表现出的价值有很大的差别。同时,在特定的条件下,特定的信息对特定需要的用户有强烈的时效性,信息的价值随着时间的改变而改变。信息价值的时效周期,一般分为四个阶段,即升值期、峰值期、减值期和负值期;不同的周期呈现不同的价值,这也是信息的时效性。

三) 信息的度量

人们对信息度量的研究已有很长的历史了。在狭义信息的度量研究方面,已经使通信科学从定性阶段进入定量阶段。然而,对广义信息进行数学描述和度量仍然还有极大的困难。

1. 基于数据量的信息度量

在计算机信息处理工作中,一个通常而简便的信息度量方法,是按反映信息内容的数据所占用计算机存储装置的存储空间大小来衡量信息量的大小。

计算机存储装置的最小存储单位为一位二进制数,称为1bit(比特,位)。常用的存储单元为8位二进制数,即一个字节,记为B。字节之间的换算为:

$$1\text{KB} = 1024\text{B} \quad (\text{KB 为千字节})$$

$$1\text{MB} = 1024\text{KB} \quad (\text{MB 为兆字节})$$

$$1\text{GB} = 1024\text{MB} \quad (\text{GB 为千兆字节})$$

存储于计算机系统中文本文件、数据文件、图像文件、声音文件等,它们含有信息量的大小常用字节数B、KB、GB来度量。存储于计算机之外的信息,如报刊、书籍等,往往直接用发行种类和发行量或者文本的字数来度量。这些也是基于数据量的度量。

2. 基于概率的信息度量

一般来说,人们获取某一事物的数据资料、背景知识等信息后,就对这一事物有了一定程度的认识,就消除了对这一事物在某一方面认识的不确定性。因此,这些数据资料和背景知识等信息中含有信息量的多少可由它们消除对事物认识的不确定性的程度来决定。所以,信息量的大小可定义为信息内容消除人们认识的不确定性的程度,信息量的大小与信息内容消除人们认识的不确定性的程度成正比。

一般来说,可以用概率来表示信息内容消除不确定性的程度,即用概率来表示信息量的大小。例如,一股票投资者有信息:“推荐的100只股票里有一只股票要上涨”,那么这个股票投资者就要在这100只股票的空间里搜索,找到的股票上涨的概率为 $1/100$ 。而另一信息,“推荐的10只股票里有一只股票要上涨”,那么这一股票投资者就要在这10只股票的空间里搜索,找到的股票上涨的概率为 $1/10$ 。显然,后一信息消除股票投资者的不确定性的程度要大于前一信息。

此外,为了度量消除不确定性的程度的大小,需要确定一个合理的度量单位。一旦确定了信息的度量单位,就可以以这个单位为基准去度量实际信息的数量大小了。

在基于概率的信息度量方法中,信息量的单位叫比特(bit,是二进位制数字Binary Digits的缩写)。1bit的信息量是指含有两个独立等概率状态的事件所具有的不确定性能被全部消除所需要的信息。在这种单位制度下,信息量的定义公式可写成:

$$H(x) = -\sum P(X_i) \log_2 P(X_i) \quad i=1, 2, 3, \dots, n$$