



MIANXIANG WANGLUO XINXI SHUJUKU YU SOUSUO YINQING

NET

面向网络信息:

INFORMATION

刘延章 主编

数据库与搜索引擎

西北工业大学出版社

面向网络信息：数据库与搜索引擎

主 编	刘延章		
副主编	马 红	方 明	吕志梅
	李 志	秦鸿霞	董全中
	谢琳惠	雷金民	
编 者	王爱丽	孙爱玲	乔 青
	杨 琳	张昕宇	

西北工业大学出版社

【内容简介】 本书旨在为信息工作者和高校师生阐述网络信息的基本知识、组织原理和搜索方法。其内容可概括为三部分：第一部分网络信息，包括网络信息的概念、功能、评价、组织等，重点阐述网络信息的分类组织原理和主题组织原理；第二部分数据库，包括面向网络信息数据库的概念、结构、技术、功能等，重点阐述大型数据库的体系结构和搜索功能；第三部分搜索引擎，包括面向网络信息搜索引擎的概念、原理、评价、功能等，重点阐述综合性搜索引擎和专业性搜索引擎的搜索功能。全书理论系统，方法实用，重点突出，信息量大。该书既可供信息工作者和网络用户参考，亦可供高校师生学习之用。

图书在版编目 (CIP) 数据

面向网络信息：数据库与搜索引擎/刘延章主编. —西安：西北工业大学出版社，2007.4
ISBN 978-7-5612-2201-0

I. 面… II. 刘… III. 计算机网络—情报检索 IV. G354.4

中国版本图书馆 CIP 数据核字 (2007) 第 040177 号

出版发行：西北工业大学出版社

通信地址：西安市友谊西路 127 号 邮编：710072

电 话：(029) 88493844 88491757

网 址：www.nwpup.com

印 刷 者：陕西宝石兰印务有限责任公司

开 本：787 mm×1092 mm 1/16

印 张：27.625

字 数：856 千字

版 次：2007 年 4 月第 1 版 2007 年 4 月第 1 次印刷

定 价：36.00 元

前 言

在当今社会里，网络信息已成为人们工作、学习和生活中不可或缺的重要组成部分，正所谓“不可一日无此君”。那么，网络信息是如何产生和发展的？浩如烟海的网络信息又是如何组织起来的？面向网络信息的数据库，其功能结构是什么？人们从哪些途径来获取数据库中的网络信息？面向网络信息的搜索引擎，其工作原理是什么？人们如何利用它来搜索网络信息？这正是本书所试图回答的主要问题。

本书是在编者多年来学习、研究和使用的网络信息、数据库和搜索引擎的基础上撰写而成的，其特点主要是：①以网络信息贯穿始终。本书共分三大部分，即网络信息、数据库、搜索引擎。但不管哪一部分，都是以网络信息的组织方式和获取途径为重点的。②以理论指导实践。在上述三大部分中，我们首先是阐述理论。在阐述理论的时候，力求其系统性和完整性。其次是描述实践。在描述实践的时候，力求其实用性和可操作性。③贯彻三个面向，即：面向广大高校师生，面向众多中文网站，面向当代高新技术。④信息量大。本书尽可能多得列举大型数据库并阐述其搜索功能，尽可能多得提供重要网站及其搜索引擎，尽可能多得提供参考文献。

全书共分三编十章七十七节，其撰写人员分工如下：马红负责撰写第一章、第七章的第十节至第十二节；方明负责撰写第二章、第七章的第十三节至第十五节、第九章的第六节；吕志梅负责撰写第三章、第四章的第三节、第九章的第一节至第四节；李志负责撰写第四章的第一节至第二节；秦鸿霞负责撰写第五章及第九章的第五节及第七节至第十二节；董全中负责撰写第六章、第九章的第十四节至第十六节；谢琳惠负责撰写第七章的第一节至第九节；雷金民负责撰写第八章、第九章的第十七节至第二十节；王爱丽负责撰写第十章的第一节至第四节；孙爱玲负责撰写第十章的第六节至第九节；乔青负责撰写第十章的第十节、第十三节至第十五节；杨琳负责撰写第九章的第十三节、第十章的第五节及第十一节至第十二节；张昕宇负责撰写第十章的第十六节。刘延章教授负责制定全书大纲、全书统稿、定稿及校对工作。

本书也是编者在借鉴当代学者研究成果的基础上编写而成的。在编写的过程中，参阅了大量的相关著作、文章和网站，多数作为参考文献列在本书有关章节之后，但仍有一些限于篇幅未能列出，在此还请有关权利者谅解！另外，本书从2005年5月20日开笔，到2006年4月20日封笔，历时近一年矣。编者在2006年底校对打印稿时发现：有些网站已经改版了，甚至销声匿迹了；而新的网站却是“小荷才露尖尖角”。只是本书难以反映这种吐故纳新的现状，还望读者见谅。由于编者水平所限，书中一些分析、归纳和评价不一定准确和恰当，敬请广大读者批评指正。

编 者

2007年1月

目 录

第一编 网络信息	1
第一章 网络信息概论	1
第一节 网络	1
第二节 信息	6
第三节 网络信息	13
参考文献	17
第二章 网站信息组织	18
第一节 网络信息组织方式	18
第二节 网站建设与网页设计	23
第三节 网站分析与标引	26
第四节 网站描述与排序	30
第五节 网站评价标准体系	33
参考文献	36
第三章 网站分类体系评价	37
第一节 五大中文网站分类体系评价	37
第二节 中文网站分类体系整体评价	51
参考文献	55
第四章 网络信息分类组织法	56
第一节 网络信息分类表的设计与编制	56
第二节 网络信息分类表的标准化	89
第三节 网络信息分类表的发展趋势	92
参考文献	95
第五章 网络信息主题组织法	96
第一节 主题组织法及其类型	96
第二节 叙词表的结构体系及其应用	99
第三节 《中国分类主题词表》简介	105
第四节 关键词组织法的原理与应用	108
参考文献	113

第二编 数据库	114
第六章 数据库概论	114
第一节 数据库与数据库系统	114
第二节 数据库类型及其简介	122
第三节 数据库技术的产生及其发展	130
第四节 我国的数据库产业	134
参考文献	137
第七章 大型数据库系统	138
第一节 CNKI 工程及其检索功能	138
第二节 CALIS 及其检索系统	143
第三节 万方数据资源系统及其检索功能	148
第四节 VIP 系统及其检索功能	156
第五节 全国报刊索引数据库及其检索功能	161
第六节 中国人民大学复印报刊资料数据库及其检索功能	164
第七节 中国数字图书馆及其检索功能	167
第八节 超星数字图书馆及其检索功能	170
第九节 书生之家数字图书馆及其检索功能	172
第十节 DIALOG 系统及其检索功能	173
第十一节 OCLC 系统及其检索功能	183
第十二节 Web of Knowledge 及其检索功能	188
第十三节 OVID 系统及其检索功能	192
第十四节 OPAC 系统及其检索功能	194
第十五节 其他大型数据库系统	197
参考文献	204
第三编 搜索引擎	206
第八章 搜索引擎概论	206
第一节 搜索引擎的产生及其发展	206
第二节 搜索引擎的工作原理	209
第三节 搜索引擎的基本类型	213
第四节 搜索引擎的缺陷及其完善	219
第五节 搜索引擎的评价体系	222
参考文献	225

第九章 综合性搜索引擎.....	226
第一节 搜索引擎的常用方法.....	226
第二节 雅虎网站及其搜索引擎.....	232
第三节 搜狐网站及其搜索引擎.....	235
第四节 新浪网站及其搜索引擎.....	238
第五节 网易网站及其搜索引擎.....	240
第六节 百度网站及其搜索引擎.....	243
第七节 Google 网站及其搜索引擎	246
第八节 中搜网站及其搜索引擎.....	250
第九节 3721 网站及其搜索引擎	254
第十节 腾讯网站及其搜索引擎.....	256
第十一节 TOM 网站及其搜索引擎	257
第十二节 21CN 网站及其搜索引擎	260
第十三节 万纬搜索及其搜索引擎.....	262
第十四节 Excite 搜索及其搜索引擎.....	264
第十五节 Infoseek 搜索及其搜索引擎.....	266
第十六节 ATW 搜索及其搜索引擎	268
第十七节 Inktomi 搜索及其搜索引擎	270
第十八节 Vivisimo 搜索及其搜索引擎	272
第十九节 Openfind 搜索及其搜索引擎	274
第二十节 其他综合性搜索引擎.....	276
参考文献.....	282
第十章 专业性搜索引擎.....	285
第一节 音乐影视网站及其搜索引擎.....	285
第二节 旅游交通网站及其搜索引擎.....	291
第三节 军事体育网站及其搜索引擎.....	297
第四节 卫生保健网站及其搜索引擎.....	304
第五节 文化事业网站及其搜索引擎.....	312
第六节 文学艺术网站及其搜索引擎.....	320
第七节 教育人才网站及其搜索引擎.....	325
第八节 经济金融网站及其搜索引擎.....	332
第九节 政府法律网站及其搜索引擎.....	344
第十节 历史地理网站及其搜索引擎.....	349
第十一节 哲学宗教网站及其搜索引擎.....	354
第十二节 科学技术网站及其搜索引擎.....	364

第十三节	电脑网络网站及其搜索引擎.....	373
第十四节	农业环保网站及其搜索引擎.....	379
第十五节	新闻媒体网站及其搜索引擎.....	387
第十六节	生活服务网站及其搜索引擎.....	395
附录.....		428
附录一	中文编目著录格式（缩写）.....	428
附录二	中国机读目录格式简介.....	428
附录三	MARC 网络资源编目	432

第一编 网络信息

第一章 网络信息概论

第一节 网 络

一、网络的含义

网络 (Network) 是指通过若干个节点连接而成的系统。人们通常所说的网络, 多指计算机网络。计算机网络 (Computer Network) 是指用通信链路把分布各地的计算机连接起来, 按照特定的协议标准进行数据通信, 以实现资源共享的信息系统。由此可见, 计算机网络是由三大基本要素组成的: 一是通信介质, 如电缆、光纤、双绞线等有线介质, 又如微波、红外线、激光等无线介质。二是计算机。计算机硬件有中央处理器 (由运算器和控制器构成)、存储器 (如光盘、硬盘、软盘等)、输入设备 (如键盘、鼠标、光笔、触摸屏、扫描仪等)、输出设备 (如显示器、打印机等)。计算机软件可分为两大类: 系统软件和应用软件。前者包括系统管理程序、系统支持程序和系统开发程序; 后者包括通用应用程序和专业应用程序。三是网络协议。计算机网络协议主要是指 TCP/IP 协议。从字面上看 TCP/IP 包括两个协议, 即传输控制协议 (TCP) 和国际协议 (IP), 实际上 TCP/IP 是一组协议, 是 Internet 的协议族, 包括上百个协议, 如远程登录、文件传输、电子邮件等协议。

二、网络的产生和发展

1. 网络产生和发展的 3 个阶段

(1) 面向终端的计算机网络。早在 20 世纪 50 年代, 美国国防部就建立了半自动地面防空系统。该系统是以单个计算机为中心的远程联机系统, 是把计算机技术和通信技术相结合, 构成面向终端的计算机网络。

(2) 计算机-计算机网络。20 世纪 60 年代, 美国国防部就支持美洲电脑公司开展“高等研究项目局网络” (ARPANET, 又称阿帕网) 的研究计划, 并于 1974 年开始运行。ARPANET 的研究成功并运行, 标志着计算机-计算机互联系统的开端。

(3) 开放式标准化网络。20 世纪 80 年代, 计算机网络专家开始注意计算机-计算机通信的规范问题。1984 年, 国际标准化组织 ISO 就颁布了一个称之为“开放系统互连基本参考模型”的国际标准 ISO7498, 简称 ISO 参考模型或 ISO/RM, 后来被 TCP/IP 协议所代替。TCP/IP 协议成为国际互联网的基础, 成为开放式标准化网络产生和发展的基石, 也为计算机网络的发展开辟了无限广阔的前景。

2. 阿帕网

阿帕网 (ARPANET) 是因特网 (国际互联网, Internet) 的基础。20 世纪 60 年代初, 美国兰德公司在空军支持下, 研究通过蜘蛛网状的分布式网络进行通信, 1965 年正式提出为空军建设该系统。但因故搁置。同年, T. 麦瑞尔的美洲电脑公司向高等研究项目局 (ARPA) 提交了一份研究计划, 计划中提到要在马萨诸塞州和加州之间进行不同电脑的链接实践。1967 年 10 月, 该公司正式提出了建立 ARPANET 计划。1972 年 10 月, 在国际电脑通信大会上成功地演示了美国各地 40 多台电脑的互联。1974 年, 由于 ARPANET 的规模扩大, 出现了兼容问题。于是该公司与高校及有关公司的研究机构签定合同, 共同研究出了网络控制程序 (NCP)。同年 5 月, 已基本成熟的 ARPANET 移交给了美国国防通信局正式运行。

3. 因特网

因特网是在阿帕网的基础上发展起来的。1972 年 10 月, 在美国华盛顿召开的首次国际电脑通信会议上, 决定成立一个国际互联网工作组, 研究并建立计算机之间进行通信的标准规范, 于是 TCP/IP 协议便应运而生。1983 年 1 月, 美国国防通信局将 ARPANET 的核心协议 NCP 改为 TCP/IP 协议。这标志着因特网已经诞生了。1986 年美国国家科学基金会 (NSF) 建立了大学之间的互联骨干网络 (NSFNET), 这是因特网由军用转人民用的重要一步。1994 年 NSFNET 转为商业运营, 这就是人们所共知的因特网的前身。1994 年 4 月中国通过登记注册, 正式与因特网联网, 成为第 71 个网员国。目前全世界已有 160 多个国家入网。

三、网络的层次和分类

1. 局域网

局域网 (LAN) 是局限于某个地域范围内的互联网。通常是把校园、大楼、实验室、办公室等一定范围内的各种计算机、终端与外部设备相互连接而形成网络。局域网被广泛地应用到连接企业或机构不同办公室的个人电脑、工作站、打印机等设备, 实现设备共享和数据交换, 因此, 局域网技术是计算机网络中发展最快的一门技术。

2. 城域网

城域网 (MAN) 是连接城市之内各单位之间的互联网, 它比局域网范围要广。如果说局域网是连接一个校园之内 (通称校园网) 的话, 那么, 城域网则是把若干个校园网连接起来的网中网。城域网可以把几十公里范围之内的学校、企业、机关、公司等众多局域网互联起来, 借以实现众多用户之间的数据、图形、语言、视频等多种信息的传输功能和交换功能。

3. 广域网

广域网 (WAN) 也称远程网, 它是连接不同地区的互联网, 它比城域网更广泛。广域网可以把相距几千公里的一个国家或几个国家, 甚至几个洲的局域网和城域网连接起来, 形成全国性或国际性的远程网络。广域网的通信子网可以利用公用分组交换网、卫星通信网和无线分级交换网等, 把分布在不同地区的计算机系统连接起来, 以便实现数据的传输和交换。

4. 国际互联网

国际互联网 (Internet) 通称因特网, 它是连接全球范围内所有网络的网络。其中包括小规模局域网、中等规模的城域网以及大规模的广域网。因特网可以通过普通电话线、高速率专用线、卫星、微波和光缆等线路, 把不同国家的大学、机关、公司、企业、科研部门、政府部门、军事部门等的网络连接起来, 让全人类共享信息资源。

如果说局域网和城域网是把计算机连接成为一个可以相互通信的系统的話，那么，广域网和因特网则是把局域网和城域网连接成为一个可以相互通信的系统。

四、网络的主要功能及其表征

1. 电子邮件

电子邮件 (Electronic Mail, 简称 E-mail) 是通过计算机网络发送信件和传递信息的一种手段。网络用户可以给任何一个拥有邮件地址的人发送电子邮件。

2. 文件传输协议

文件传输协议 (File Transfer Protocol, 简称 FTP) 是计算机网络上主机之间传送和复制文件的一种服务协议。这些文件可以是文字、表格、计算机程序、图形、声音、影像等。

3. 远程登录

远程登录 (Telnet) 是由本地主机通过网络登录到远程的另一台主机上，一旦登录成功，自己的计算机就可以成为远端主机的一部终端机，可以享受该主机的所有功能和资源，但用户必须先拥有该主机的账号才能合法登录。

4. 万维网

万维网 (World Wide Web, 简称 WWW 或 Web) 是以超文本技术为基础，用面向文件的阅览方式，提供一定格式文本信息和图形信息的服务系统。目前的万维网不但可以实现文本之间的连接检索，还可以提供声音、图像、动画等多媒体信息的连接检索。

5. 电子公告牌

电子公告牌 (BBS) 是用户之间进行信息交流的园地。它的主要功能包括：①提供给用户阅览文章和信息；②用户可在站点内发表文章或信息；③用户可对站点内他人的文章进行评论；④用户可与同一站点内的其他用户实时对话、发电子邮件；⑤提供免费的文件传输、软件获取等。

6. 网络新闻

网络新闻 (Net News) 是利用计算机网络在大范围内向众多用户快速传递新闻的一种方式。在网络新闻中，用户被组织在“新闻组”的专题下参与讨论。每一则信息称之为篇文章，每一篇文章采用电子邮件方式发给网络新闻组，每篇发往网络新闻的文章被放在一个或几个相关的新闻组中，用户可在客房终端利用新闻阅读程序，组织并浏览这些文章。若遇到问题或难题，也可以向有关新闻组发求助函。

7. 网络聊天

网络聊天 (International Relay Chat, 简称 IRC) 是一种国际性的多人聊天系统。网络聊天是指使用特定的客户端软件连接到 IRC 服务器，然后以客户端→服务器→客户端的方式，使双方用户能够交换信息的系统。服务器在信息传递中间充当了一个中继信息的作用，因此，IRC 的中文名称有时又称之为“网络中继聊天”。它具有保密性强、屏幕刷新快、便于管理等优点。

8. 网络电话

网络电话 (Voice Over Internet Protocol, 简称 VOIP)，就是将语音数据通过打包压缩并通过 IP 网络进行传输的一种语音传输技术。IP 网络电话就是用户用普通电话拨打接入号 17909 进入服务器或叫网关服务器，用户的语音数据通过打包通过网关进入 IP 网络到达目的地，再通过网关进入目的地的号码。

五、网络的应用举例

1. 电子政务

电子政务 (Electronic Government, 简写 EG) 是指面向政府机关的优化工作流程和扩展服务功能的基于因特网技术的综合信息系统。其目的在于以网络技术为手段, 优化和扩展政府机构的业务模式、管理模式和服务方式, 取代手工操作, 借以提高政府各部门的工作质量和效率。电子政务主要有三种形式: 一是内容发布, 即通过政府网站发布有关信息; 二是信息和数据交换, 即政府与公民进行网上交流; 三是业务工作的高度集成, 即办公实现自动化。

2. 电子商务

电子商务 (Electronic Commerce, 简写 EC) 是指以网络技术为平台, 通过电子通信方式, 进行商贸活动的系统。电子商务发展到今天, 已经渗透到企业的各个方面, 如原料和设备的网上采购、产品网上销售及网上售后服务等, 再如与客户网上洽谈、订货、在线付 (收) 款、开电子发票、电子报关、电子纳税等。当然还涉及政府机关 (如工商、税务、海关等)、金融机构、认证机构、配送中心等, 因此说, 电子商务是一个庞大的网络工程。电子商务主要有三种类型: 一是企业内部的电子商务; 二是企业之间的电子商务; 三是企业与消费者之间的电子商务。

3. 数字图书馆

数字图书馆 (Digital Library) 就是指运行在高速宽带网络上的分布式超大规模的可跨库检索的海量数字化信息资源库群。由此可见, 数字图书馆是以数字化的信息为基础, 以分布式海量信息资源库群为支撑, 以智能检索技术为手段, 以电子商务为管理方式, 以高速宽带网络为传输通道, 将文本、图像、语音、音响、影像、影视、软件及科学数据等多媒体信息进行生产、加工、存储、检索、传递、保护、利用、归档、剔除等。数字图书馆的特点是收藏数字化、操作电脑化、传递网络化、信息存储自由化、资源共享化和结构链接化。美国 1993 年 9 月首次提出 “数字图书馆倡议”, 并于 1994 年 6 月召开了 “第一届数字图书馆理论与实践研究” 的国际会议。

4. 办公自动化

办公自动化 (Office Automation, 简称 OA) 是指在系统科学和行为科学的指导下, 通过计算机技术、通信技术等先进技术与先进办公设备的有机结合, 从而高效地处理办公信息, 进而提高办公质量和科学决策水平的综合性技术。办公自动化系统可分为事务型办公系统、管理型办公系统和决策型办公系统。办公自动化具有以下特点: 一是集成化, 二是智能化, 三是多媒体化, 四是电子数据交换。办公自动化是由人、硬件和软件共同组成的一个庞大系统。人是指办事员、网络工程师、设备维护人员等; 硬件是指计算机、复印机、速印机、电话机、传真机、光盘机及网络设备等; 软件是指网络操作系统、文字处理软件、专项工作程序软件等。

5. 远程教育

远程教育 (Distance Education, 简称 DE) 是应用信息网络技术和多媒体技术, 在数字化的环境下, 师生进行教与学的一种交互式教育方式。远程教育涵盖了信息源、网络传输、多媒体终端和网络教学管理四部分。远程教育模式主要有两种: 一是以群体为基础的远程教育; 二是以个别为基础的远程教育。远程教育体系以多媒体网络传递文、图、声、像在的各种教学信息形成交互式的教学模式, 具有可视远程授课、点播、同步课业辅导、远程讨论交流、咨询答疑等功能。由此可见, 远程教育的突出特点是教育形式不受时空限制; 教育对象不受校园限制; 教育资源不受单位限制; 学生学习不受教师限制。总之, 远程教育具有开放性、延伸性、灵活性、自由性、普及性等特点。

6. “三上网工程”

“三上网工程”是由信息产业部支持的、中国电信发起的“政府上网工程”、“企业上网工程”和“家庭上网工程”的统称。“政府上网工程”已于1999年启动。目前已建立了3200多个政府网站,70%以上的政府部门在网上设立了办事窗口。“企业上网工程”于2000年7月启动。目前已有100多万家企业上网,80%以上的企业在网上建立了自己的网站或网页。“家庭上网工程”于2001年12月启动。目前至少有1亿多家庭在网上进行在线游戏、视频点播、网上购物、可视电话、浏览新闻、查阅资料、个人理财、金融投资等。

六、图书馆网络化

1. 图书馆网络化的含义

图书馆网络(Library Network)是指利用通信技术和计算机网络结构,按照一定的协议,把异地的图书馆都连接起来,从而实现信息资源共建共享的系统。图书馆网络化是指以大量的数字化信息资源为基础,以高速的数字通信网络为平台,采用分布式管理,满足面向对象查询需要的过程和程度。由此可见,图书馆的网络化应具备四大要素:一是信息资源数字化。不管是馆藏书籍、期刊、录音带、录像带、甚至古籍善本、字画、X光片、缩微制品等,都必须数字化。这是图书馆网络化的物质基础。二是信息存取网络化。对图书馆内部来说,是由高速主干网连接数台服务器和 workstation 而建成的局域网;对图书馆外部而言,是通过数台广域网服务器进入国家级的主干网和因特网。这是图书馆网络化的必由之路。三是采用分布式管理。是指通过一定的协议把异地的图书馆互联起来,就像因特网通过TCP/IP协议把各地网站互联起来一样。四是用户服务信息化。如远程查询、在线问答、电子邮件服务、馆际互借服务、在线推送服务(即根据一定的协议,从信息源中选择信息,并通过一定的方式,定期地把信息传递给用户的一种服务方式)等。图书馆利用信息化的手段,向用户提供原始数据和信息。这是图书馆网络化的目的。

2. 图书馆网络化的发展

从国外来看,图书馆的网络化大体上经历了四个阶段:

第一个阶段是20世纪60年代,为单机操作阶段。当时美国及西方一些国家的计算机开始在图书馆中得到应用。当时只是把文献的加工处理交给计算机,如把文献信息输入计算机,然后再打印成卡片,以代替手工劳动。

第二个阶段是20世纪六七十年代,为局域网阶段。当时美国及西方一些发达国家开发出了图书馆集成管理系统,建立了图书馆的局域网,实现了图书馆内部的计算机管理。

第三个阶段是20世纪七八十年代,为广域网阶段。当时美国等西方一些发达国家的高校图书馆开始实现联机编目,并建立了联机书目信息数据库和专题数据库,组建了广域网,开始了联机检索和初步的资源共享。

第四个阶段是20世纪末至21世纪初期,为因特网阶段。20世纪90年代因特网得到了快速发展,一些图书馆纷纷在因特网上建立网站,探索信息服务的新方法,于是数字图书馆便应运而生。

从国内来看,图书馆的网络化大体上也是经过了四个阶段:

第一个阶段是20世纪七八十年代初期,为起步实验期。当时的中国科学院计算机研究所和中国科学技术情报研究所分别对文献进行了输入和输出实验,实现了借用计算机操作来代替手工操作。

第二个阶段是20世纪80年代中期,为开发应用期。当时的中国科学院计算机研究所已开发出各类型单项和多项的计算机管理系统,并开始在当时中国科学院图书馆应用。

第三个阶段是20世纪80年代末期到90年代初期,为集成系统发展期。当时我国很多高等院校的图书

馆,纷纷要求实现各个业务环节的计算机管理,于是一些大学图书馆(如北京大学、北京邮电大学、深圳大学、西安交通大学、南京大学等)开发研制出了图书馆集成管理软件,一些软件开发公司(如北京文津、北京丹诚、大连博菲特、重庆图腾等)抓住这一商机,也纷纷开发出图书馆集成管理软件。

第四个阶段是20世纪90年代中期以后,为网络化开始时期。1995年,北京大学信息管理系建立了我国第一个电子阅览室。同年,清华大学与IBM公司中国研究中心合作推出了“清华大学数字图书馆系统”建设计划。1996年,国家图书馆提出了“中国试验型数字图书馆”项目,并由国家图书馆牵头,上海图书馆、辽宁省图书馆、南京图书馆、广东中山图书馆、深圳图书馆、广西桂林图书馆等七馆合作项目开始启动。同年,中国科学院文献情报中心开始实施“中国科学院百所联网工程”,并向四个地区和100多个中国科学院系统的研究所提供书目信息查询服务。1997年,上海交通大学开始建设数字图书馆。同年,上海图书馆的数字化工程开始启动。1998年,中国数字图书馆工程项目开始启动。1999年,中关村科技园数字图书馆建设拉开序幕。2000年,北京大学图书馆成立了数字图书馆研究所。2001年,江苏省数字图书馆开始实施。2002年以后,各省馆、各高校馆、各地、市馆都投入了大量的资金和人力来建设数字图书馆。

第二节 信 息

一、信息的含义

从广义讲,信息(Information)就是对自然和社会所产生出来的各种事物运动状态及其变化的反映;从狭义讲,信息就是人们在各种活动过程中所产生和接受的各种消息。关于信息概念的表述据说有不下30种,但都不是公认的定义。现择其要者归纳如下:

- (1) 信息是指事物运动状态及其改变方式;
- (2) 信息是万物赖以生存和进化的音信和信号;
- (3) 信息是事物属性的表征;
- (4) 信息是人们所需要的知识、消息、音信、情报、情况、指令等;
- (5) 信息是具有新知识、新内容的消息;
- (6) 信息是物质和精神存在状态的反映;
- (7) 信息是产生于物质又依附于物质的一种信号;
- (8) 信息是可以识别、存储、传播的一切自然与社会现象的关系和特性;
- (9) 信息是事物之间差异的再现形式;
- (10) 信息就是人们在生产和生活等活动中通过各种方式了解到的消息。

《中国大百科全书》中对信息的解释如下:信息就是关于事物运动的状态和规律的表征,也是关于事物运动的知识。信息就是用符号、信号或消息所包含的内容,来消除对客观事物认识的不确定性。我国国家标准《情报与文献工作词汇基本术语》(GB4894—1985)对信息含义作如下描述:信息是物质存在的一种方式、形态或运动状态,是事物的一种普遍属性,一般指数据、消息中所包含的意义,可以使消息中所描述事件的不确定性减少。信息论的创始人C. E. 香农在其著作《通信的数字理论》中对信息概念作如下描述:信息是用于消除随机不确定性的东西。控制论的创始人N. 维纳在其著作《控制论与社会》中认为:信息是人们在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交换的内容的名称。美国的《韦氏大词典》对信息的解释是,信息是通信的事实,是在观察中得到的数据、新闻和认识。

不管如何定义信息,但信息都必须具备四大要素,即信源、信道、信宿和载体。发出信息的物体称之为信源,传播信息的通道称之为信道,接受信息的物体称之为信宿,承载信息的物体称之为信息载体。

二、信息的属性

信息的属性是由信息的本质所决定的。由上述的含义可知,信息的本质具有多样性,因此,信息的属性也具有多样性。现择其要者列举如下:

1. 信息的普遍属性

信息是存在于自然界和人类社会中的无形财富,它是和材料、能源并存的社会发展的三大支柱之一。可以说物质不灭定律也适合于信息,即信息存在着不灭定律,物质取之不尽,信息也用之不竭。信息无处不在,也无时不有。

2. 信息的存储属性

信息的存储形式很多,最基本的形式是大脑存储。最现代化的方式是计算机存储,可以是文字、图表、符号、声频、视频等。信息存储的方式将随着科技的发展而会更加先进。

3. 信息的传递属性

信息是通过传递来体现其自身价值的。信息传递的路径很多,可以通过口头、电话、动作、表情、气味、颜色等直接传递,也可以通过文字、图像、声音等间接传递。由于网络技术和通信技术的发展,信息从信源到信宿的传递时间越来越短,对人类社会的影响也越来越大。

4. 信息的共享属性

信息从信息源产生之后,通过信道可以传递给多个信宿,也就是说,拥有不同的信宿可以得到相同的信息。这就是信息共享的原理所在。在信息化的今天,很多信息(如天气、交通、新闻等)都是可以共享的,而且不会因多次使用而改变信息的自身价值。

5. 信息的时效属性

事物是发展变化的,信息也是如此。信息的价值在于利用,再好的信息如果错过时效,也无价值可言。也就是人们常说的“机不可失,时不再来”。如股市行情,瞬息万变,把握不好就会损失惨重。

6. 信息的失真属性

信息在产生、加工、整理、传递等过程中,往往会受到主观和客观因素的影响,使信息失真,成为虚假信息或垃圾信息,如果接受者信以为真,有时会造成极大的损失。如股市分析师由于受某种利益的驱动,传播虚假信息,造成股民血本无归的现象屡见不鲜。

7. 信息的再生属性

当人们收到某种形式的信息之后,经过加工处理,还可再生成其他形式的信息。或者从一种信息载体再生成另一种信息载体。原始信息经过综合、分析、加工等,可以转化成二次信息和三次信息,原有的信息价值也可以实现增值。如网络信息,经过加工处理之后,可再生成文献、录像带、录音带等载体的信息。

8. 信息的感应属性

信息可以通过人的感觉器官去感知它,也可以通过各种传感器、触摸屏去识别它。当计算机具有人工智能后,也可以通过声音、颜色、气味、图像等来感知和识别外部各种载体信息。

9. 信息的依附属性

信息不可能独立存在,总是依附于一定的载体,即信息的承担者。声音、语言、文字、颜色、图像、符号、电、磁、光、生物等都是信息信号;纸张、胶片、光盘、磁带、人的大脑、计算机等都是

信息载体。

三、信息的类型

由于划分的标准不同,信息的类型也就不一样。

(1) 按照信息发生的领域来划分,可分为自然信息和社会信息两大类。前者包括天体信息、物质信息、生物信息等;后者包括人际信息、活动信息、精神信息等。

(2) 按照信息的学科属性来划分,可分为社会科学信息和自然科学信息两大类。前者包括政治信息、经济信息、文化信息、生活信息等;后者包括科技信息、农业信息、工业信息、环保信息等。

(3) 按照信息的加工程度来划分,可分为一次信息、二次信息、三次信息。前者是指未加工的或粗加工的原始信息;中者是指对一次信息进行加工之后而产生的信息;后者是对二次信息加工整理之后而产生的信息。

(4) 按照信息的载体来划分,可分为大脑信息、文献信息、实物信息、电子信息、光纤信息、网络信息等。

(5) 按照信息的传播范围来划分,可分为公开信息、非公开信息、半公开信息、绝密信息、机密信息、解密信息等。

(6) 按照信息所采用的符号来划分,可分为语言信息、文字信息、图画信息、声像信息、电磁信息、颜色信息、光泽信息、代码信息等。

(7) 按照信息的内容来划分,可分为客观信息和主观信息。前者是指对事物进行客观描述的信息内容;后者是指对事物进行分析后表述主观意见的信息内容。

(8) 按照信息传播的渠道来划分,可分为口传信息、实物传播信息、文献传播信息、电子传播信息等。

(9) 按照信息的作用来划分,可分为有效信息、无效信息、垃圾信息等。

(10) 按照信息的真伪来划分,可分为真实信息、虚假信息等。

四、信息化社会

1. 信息化的含义

信息化(Informationization)就是在信息基础结构上综合各种要素所展开的信息应用过程和程度。这里说的信息基础结构包含两大基础设施:一是通信基础设施(即电信网络);二是信息基础设施(即信息网络)。这里说的各种要素如下:

(1) 人才要素,是指管理、开发、利用各种设施和信息的各种人才。这是实现信息化的动力。

(2) 技术要素,是指通信技术、网络技术、数字化技术等。这是实现信息化的催化剂。

(3) 产业要素,是指通信业、计算机业、软件业、数据库业、传媒业、电子制造业、信息加工业等。这是实现信息化的支柱。

(4) 环境要素,是指标准、安全、政策等法律法规。这是实现信息化的保障。

这里所说的信息应用,是指信息技术、信息资源和信息产品的开发和利用,涉及到国民经济和社会发展的所有领域。

2. 信息高速公路

信息高速公路(Information High Way,简称IHW)实质上是一个庞大的交互式的多媒体高速计算机通信网络系统。它是国家信息基础设施(National Information Infrastructure,简称NII)的一种形象比喻。它是以现代计算机网络与通信技术为基础、以光导纤维通信和卫星通信为前提、以数据库为信息源、双向大容量、高速度的电子数据的传输系统。由此可见,信息高速公路建设是一个庞大的

系统工程。

早在 1991 年, 戈尔 (时任美国参议员) 就提出“国家高性能计算机技术法案”, 建设高宽带的国家研究和教育网络。1992 年, 克林顿和戈尔 (时任克林顿政府副总统) 又明确提出建设信息高速公路, 并作为竞选总统的一个纲领。1993 年 1 月, 克林顿授权成立了“信息基础设施特别小组”。1993 年 4 月, 美国以通信公司 AT&T 为首的 14 家公司, 联合向国会提出建设信息高速公路的实施方案。1993 年 9 月 15 日, 戈尔 (时任美国副总统) 和布朗 (时任商业部长) 宣布美国要建设为期 20 年、耗资 4 000 亿美元、惠及美国人民的信息高速公路计划。1993 年 11 月 15 日, 克林顿政府正式批准了“国家信息基础结构: 行动计划”。美国政府通过一项拨款法案, 规定在 5 年内, 每年拨款 10~20 亿美元, 用于推进 NII 工程的实施。1993 年 12 月, 欧共体提出了建立欧洲信息高速公路计划。1994 年 2 月 23 日, 法国总统召开内阁会议, 会上提出了法国建设信息高速公路计划。1994 年 3 月 21 日, 戈尔又在阿根廷举行的第一次世界信息发展会议上, 提出了建设“全球信息基础结构 (GII)” 的设想。韩国政府于 1994 年 3 月 23 日正式确定建设韩国信息高速公路综合计划。1994 年 8 月, 日本内阁作出决定, 成立以内阁总理大臣为首的“高度信息化社会推进本部”, 同年发表了日本将在 2010 年建成信息高速公路的意见书。新加坡早在 1980 年就提出建设电子基础设施的计算机化社会。俄罗斯、加拿大、德国等也都先后提出本国建设信息高速公路的计划。1995 年 2 月, 西方 7 国和欧盟在比利时召开了“信息社会圆桌会议”, 确定了发展全球信息化社会的行动准则。1996 年 7 月在南非召开了“信息社会与发展” 世界大会, 共有 50 多个国家 100 多名代表参加。2000 年 7 月在日本冲绳召开了八国集团首脑会议, 通过了“全球信息社会冲绳宪章”。2000 年 8 月 21 日在北京召开了第十六届计算机大会, 大会主旨是新世纪的信息管理。

五、中国的信息化建设

1. 信息网络建设

我国信息化建设真正起步是 1993 年 12 月成立的以邹家华为主席、胡启立为副主席、中央 24 个部委参加的“国家经济信息化联席会议”。在这次会议上提出了“三金工程” (即金桥工程、金关工程、金卡工程), 后来发展成为“金”字系列工程, 从此才真正拉开了我国经济信息化建设的序幕。为便于了解, 现把“金”字系列工程中最主要的工程列入表 1.1。

表 1.1 “金”字系列工程一览表

序号	简称	工程全称
1	金桥	国家公用经济信息通信网
2	金关	国家对外经济贸易信息网
3	金卡	自动化支付系统及电子货币工程
4	金税	税务征收管理服务信息系统
5	金智	教育科研计算机网与人才工程
6	金企	国家企业生产与流通信息系统工程
7	金农	农业综合管理及信息服务系统
8	金卫	国家医疗卫生信息网络工程