

数字图书档案技术

Digital Book and Digital Archives Technology

主编 王爱功 景卫东



时代文艺出版社

数字图书档案技术

Digital Book and Digital Archives Technology

主编 王爱功 景卫东

时代文艺出版社

图书在版编目(CIP)数据

数字图书档案技术/王爱功,景卫东主编.—长春:时代文艺出版社,2006.5

ISBN 7-5387-2128-2

I.数… II.①王…②景… III.①数字技术-应用-图书馆工作②数字技术-应用-档案工作

IV.①G250.7②G270.7

中国版本图书馆 CIP 数据核字(2006)第 046468 号

数字图书档案技术

出 版	时代文艺出版社
地 址	长春市人民大街 4646 号 邮编:130021
电 话	总编办: 0431-5638648 发行科:0431-5677782
网 址	www.shidaichina.com
印 刷	河南日兴印务有限公司
发 行	时代文艺出版社
开 本	787×1092 毫米 1/16
字 数	600 千字
印 张	32
版 次	2006 年 5 月第 1 版
印 次	2006 年 5 月第 1 次印刷
印 数	1500
定 价	48.00 元
版权所有	翻印必究

前 言

本书背景

信息科学是研究和模拟人脑中信息过程的智能科学；信息处理的基本工具是计算机技术，信息传输的基本工具是网络技术；信息的有效利用程度已经成为现代社会发展水平的一种象征或者说是标志。而目前许多图书馆管理、档案管理等相关专业学术著作及高校教材，已经不能适应信息技术的飞速发展，人们获取信息的环境、方式和手段发生了根本的变化，传统的手工信息查询正在逐步演变成数据库信息查询、网络信息查询，网上数字图书馆、数字档案馆将成为新世纪人们获取信息的主渠道。

为了适应信息时代技术更新的需求，培养出具有现代数字图书馆、数字档案馆设计能力的高素质科技人才，同时也满足现阶段各高校的教学需要，为此我们特编写了以数字图书档案七大技术为主且具有浓郁新时代信息的这本技术专著。

本书结构

为了适应信息查询技术飞速发展的需要，本书从整体上跨专业全面详细地介绍了数字图书档案七大技术，即数字图书档案存储技术、数字图书档案传输技术、数字图书档案数据库技术、数字图书档案查询技术、数字图书档案网站技术、数字图书档案安全技术；数字图书档案法规标准与技术，从实用的角度和新技术的角度重点叙述了数字图书档案相关最新技术。

本书读者

本书可作为行政、企事业单位信息管理、图书馆管理、档案管理、科技文献检索等相关领域中、高层管理人员、技术人员知识更新读物，还可供高等院校相关专业师生阅读参考。

本书特点

本书具有两大特点：其一在于本书题材的全面性和新颖性，本书首次将数字图书档案的七大技术融为一体；其二在于本书技术的先进性和实用性，为了使理论和实际相结合，本书不仅加强了基本原理的论述，而且注重实用，编写时力求做到与时俱进，将大量新颖和先进技术，诸如数据仓库与数据挖掘技术、联机分析技术、分布式数据库、B/S 浏览器/服务器技术、分布式系统与群集技术、移动信息检索技术等深入浅出地介绍给广大读者。

本书作者

本书第一章、第二章、第三章第一节由河南省图书馆王爱功编写；第三章第二节、第七节、第八节、由南阳理工学院图书馆李海峰编写；第三章第三节、第四节、第五节、第六节由汝南县图书馆孙冶钢编写；第四章第六节、第七节、第八节、第九节、第十节、第五章第四节由南阳理工学院图书馆姚景芝编写；第四章第一节、第二节、第三节、第四节、第五节、由漯河职业技术学院图书馆张丽萍编写；第五章第一节、第二节、第三节、第五节、第六章由河南省图书馆孔德超编写；第七章和第八章由河南省档案局（馆）景卫东编写。全书由王爱功、景卫东统稿审定。

主 编

2006 年 3 月

目 录

第一章 数字图书档案概论	1
1.1 图书档案基础.....	2
1.1.1 图书基础知识.....	2
1.1.2 档案基础知识.....	8
1.2 图书档案信息网络.....	11
1.2.1 现代图书档案信息网络的设计.....	11
1.2.2 现代图书档案信息管理系统.....	12
1.2.3 决策型图书档案办公系统.....	14
1.2.4 图书档案信息网络的基本功能.....	16
1.2.5 图书档案信息网络的结构与组成.....	16
1.2.6 图书档案信息网络的安全措施.....	18
1.3 数字图书档案馆.....	18
1.3.1 数字图书档案馆的概念.....	18
1.3.2 数字图书档案馆与传统图书馆档案馆的关系.....	19
1.3.3 数字图书档案馆的信息检索服务.....	20
1.4 电子公文归档管理暂行办法.....	23
第二章 数字图书档案存储技术	25
2.1 存储技术概论.....	26
2.1.1 纸存储介质.....	26
2.1.2 缩微胶卷存储介质.....	26
2.1.3 磁介质存储技术.....	27
2.1.4 音频视频存储技术.....	27
2.2 光盘.....	28
2.2.1 光存储介质.....	28

2.2.2 光盘的发展历史	29
2.2.3 光盘的分类	30
2.2.4 光盘的组成材料	31
2.2.5 光盘的组成结构	33
2.2.6 光盘大家族	35
2.3 CD-ROM	36
2.3.1 CD-ROM 标准的制定	36
2.3.2 CD-ROM 的物理格式与逻辑格式	37
2.3.3 CD-ROM 的制作	40
2.3.4 电子出版物	41
2.4 光盘驱动器	42
2.4.1 光盘驱动器的分类	42
2.4.2 光盘驱动器的组成	43
2.4.3 光盘驱动器的读写过程	46
第三章 数字图书档案传输技术	49
3.1 数据传输基础	50
3.1.1 信号及其变换	50
3.1.2 数据传输方式	54
3.1.3 基带与频带	57
3.1.4 数据传输速率	58
3.1.5 数据传输差错控制技术	59
3.2 计算机网络	60
3.2.1 计算机网络的主要功能	60
3.2.2 计算机网络的基本组成	61
3.2.3 计算机网络的拓扑结构	61
3.2.4 计算机网络的分类	62
3.2.5 计算机网络协议标准	63
3.2.6 计算机局域网	64
3.2.7 计算机局域网体系结构及 OSI3 参考模型	67
3.2.8 计算机通信网	75

3.2.9 常用的数字通信网	76
3.2.10 通信网交换技术	78
3.3 接入网技术	85
3.3.1 接入网概述	85
3.3.2 接入网的定义	86
3.3.3 接入网的接口	87
3.3.4 接入网的分类	87
3.3.5 接入网的主要特点	89
3.3.6 接入主网要业务和管理	90
3.4 有线宽带接入网——ADSL 技术应用	92
3.4.1 ADSL 接入技术	92
3.4.2 ADSL 基本原理	93
3.4.3 ADSL 接入模式	94
3.4.4 ADSL 主要特点	95
3.4.5 ADSL 技术应用	97
3.5 有线宽带接入网——HFC 技术应用	98
3.5.1 HFC 接入网技术特点	98
3.5.2 HFC 和 Cable Modem 基本原理	99
3.5.3 HFC 网络应用及解决方案	101
3.5.4 HFC 目前发展状况	106
3.6 互联网 (Internet)	108
3.6.1 初识互联网	108
3.6.2 如何实现与 Internet 的连接	115
3.6.3 文件传输 (FTP) 服务	120
3.6.4 远程登录 (Telnet) 服务	122
3.6.5 电子函件 (E-mail) 服务	124
3.6.6 Internet 现存问题、对策及展望 (Internet II)	127
3.7 移动办公互联网	130
3.7.1 WAP 应用网络模型	130
3.7.2 WAP 网关	131

3.7.3 WAP 应用实例·····	133
3.8 网络新技术应用·····	136
3.8.1 分布式系统与群集技术·····	136
3.8.2 并发系统·····	142
3.8.3 共享存储器多处理机系统·····	144
第四章 数字图书档案数据库技术·····	149
4.1 数据库概论·····	150
4.1.1 数据库基础知识·····	150
4.1.2 数据库的基本功能·····	152
4.1.3 数据库的基本运行环境·····	154
4.1.4 数据库的基本语法规则·····	155
4.2 数据库程序设计基础·····	163
4.2.1 程序建立与运行·····	163
4.2.2 程序中交互赋值语句·····	165
4.2.3 程序中的辅助命令·····	167
4.2.4 结构化程序设计·····	167
4.3 数据库的基本操作·····	170
4.3.1 数据库的建立·····	170
4.3.2 数据库的开关·····	172
4.3.3 数据库的定位·····	172
4.3.4 数据库的显示与打印·····	173
4.3.5 数据库的修改或维护·····	175
4.4 数据库的排序、索引、检索·····	179
4.4.1 数据记录的分类排序·····	179
4.4.2 数据库的索引排序·····	180
4.4.3 数据记录的检索查询·····	181
4.4.4 数据库的统计·····	182
4.5 报表与菜单生成器·····	184
4.5.1 报表生成器·····	184
4.5.2 菜单生成器·····	188

4.6 分布式数据库系统	190
4.6.1 分布式数据库的定义与特点	190
4.6.2 分布式数据库的分类	192
4.6.3 分布式数据库系统的组成	194
4.7 数据仓库技术	195
4.7.1 数据仓库概论	195
4.7.2 数据仓库的基本元素	196
4.7.3 数据仓库模型与设计	197
4.8 联机分析处理	198
4.9 数据挖掘技术	199
4.9.1 数据挖掘的概念和特点	199
4.9.2 数据挖掘的层次空间	199
4.9.3 数据仓库、联机分析及数据挖掘的综合应用	200
4.10 网络数据库查询	201
4.10.1 国内网络数据库	201
4.10.2 国外网络数据库	204
第五章 数字图书档案查询技术	209
5.1 传统图书档案查询技术	210
5.1.1 检索工具	210
5.1.2 档案著录	213
5.2 传统信息检索技术	216
5.2.1 信息检索、文献检索、事实与数据检索	216
5.2.2 信息检索工具	217
5.2.3 信息检索语言	218
5.2.4 国内信息检索工具	221
5.2.5 信息检索方法	222
5.2.6 信息检索技术的发展	225
5.3 现代网络信息查询	226
5.3.1 网络查询概论	226
5.3.2 网络信息资源	229

5.3.3 网络查询工具·····	234
5.4 网络信息查询应用·····	239
5.4.1 文件查询服务软件 Archie·····	239
5.4.2 信息浏览服务软件 Gopher·····	240
5.4.3 广域信息服务器软件 WAIS·····	241
5.4.4 万维网 (WWW) 应用·····	242
5.4.5 网络中文信息搜索引擎·····	248
5.4.6 网上英文搜索引擎·····	258
5.5 联机查询系统 DIALOG·····	265
5.5.1 DIALOG 系统概述·····	265
5.5.2 建立 DIALOG 联机·····	265
5.5.3 DIALOG 的查询特点和查询指令·····	267
5.5.4 DIALOG 的基本查询技术·····	271
第六章 数字图书档案网站技术·····	275
6.1 网站的选择与设计·····	276
6.1.1 网站的三种选择·····	276
6.1.2 网站 Web 网页的设计·····	277
6.1.3 Web 网站的推广·····	278
6.2 服务器分析与选型·····	281
6.2.1 服务器系统分析·····	281
6.2.2 服务器系统设计与选型·····	283
6.3 交换机分析与选型·····	285
6.3.1 以太网二层交换机的基本工作原理·····	285
6.3.2 交换机的选型、连接与品牌·····	286
6.3.3 三层交换机的应用、选型与品牌·····	289
6.3.4 使用 TFS9009E 交换机的图书馆网结构方案实例·····	292
6.4 路由器分析与选型·····	295
6.4.1 路由器的功能特点·····	295
6.4.2 路由器的分类·····	296
6.4.3 路由器的基本工作原理·····	297

6.4.4 路由器的结构	299
6.4.5 路由器与网桥的比较	301
6.4.6 路由器的主要产品品牌、配置和使用	301
第七章 数字图书档案安全技术	305
7.1 数字图书档案安全技术	306
7.1.1 计算机安全	306
7.1.2 计算机病毒	307
7.1.3 计算机网络安全	317
7.1.4 Internet 防火墙 (Fire Wall) 技术与有害信息的过滤	318
7.1.5 数字图书档案安全技术小结	322
7.2 数字图书档案加密技术	324
7.3 数字图书档案知识产权保护	326
7.3.1 知识产权	326
7.3.2 数字图书档案知识产权保护的威胁因素	326
7.3.3 数字图书档案知识产权保护措施	327
第八章 数字图书档案法规标准与技术	330
8.1 新型信息载体材料与数字图书档案技术	331
8.1.1 磁盘	331
8.1.2 130mm 磁盘	332
8.1.3 90mm 软磁盘	332
8.1.4 120mmCD-ROM 光盘	333
8.1.5 130mm 可重写盒式光盘	333
8.2 国家档案局对电子文件应用研究的有关规定	334
8.2.1 参与国家标准的研制	334
8.2.2 制定了一系列适用于公务文件的行业标准	335
8.3 国际研究机构对新型档案载体材料的保护保存要求	335
8.4 新型信息载体材料与档案保护技术的关系	336
8.5 国家标准与档案及相关标准概览	336
8.5.1 国家标准与档案	336
8.5.2 行业标准与档案	337

8.5.3 正在修改和制订的档案标准·····	338
8.6 中华人民共和国档案法·····	339
8.7 中华人民共和国电子签名法·····	342
8.8 中华人民共和国档案法实施办法·····	347
8.9 中华人民共和国国家标准照片档案管理规范·····	351
8.10 纸质档案数字化技术规范·····	366
8.11 档案缩微品制作记录格式和要求·····	371
8.12 档案字迹材料耐久性测试法·····	386
8.13 革命历史档案机读目录软磁盘数据交换格式·····	391
8.14 民国档案目录中心数据采集标准 民国档案机读目录软磁盘数据交换格式外·····	395
8.15 外国组织和个人利用我国档案试行办法·····	399
8.16 河南省档案管理条例·····	400
8.17 河南省公共图书馆管理办法·····	404
8.18 档案管理软件功能要求暂行规定·····	406
8.19 档案工作行业标准目录·····	409
8.20 公务电子邮件归档与管理规则·····	411
本书说明·····	421

第一章

数字图书档案

概论

学
习
内
容

- 图书档案基础知识
- 现代图书档案管理
- 图书档案信息网络
- 数字图书档案馆
- 电子档案管理办法

1.1 图书档案基础

1.1.1 图书基础知识

1. 我国书籍的产生、形成和发展

图书是指用文字、图画或其他符号,在一定的物质材料上记录知识、信息,用以表达、传播、存储思想、文化,并制成卷册的著作,一般泛指书籍、课本、图片。

我国最原始的书,大约出现在公元前 1400 年,已有近四千年历史,其产生和发展大致可分为五个阶段。

原始的书出现在公元前 1400 年前的殷商时代,其书籍载体为甲骨、青铜器、石头,甲骨、青铜器和石头上的“书”,已经具有书籍的记事、传播功能。

最早的书一般指简策、版牍、帛书,简策版牍阶段从春秋到东晋,前后延续 1000 余年,大量通行 900 年,是我国书籍的形成阶段,其标志是写在竹片、木片、木板片上的简策、版牍和写在缣帛上的帛书,载体为竹木片、木版片或丝织的帛。

用经过加工的竹木片作为书写材料的叫“简书”,将写了字的竹木简书用绳编连起来叫“简策”,象形文字写作“册”,即书称册的来历。简策在历史上被称为“简策制度”。穿简成册的细绳子叫编。用经过刮削的木板片作为书写材料的叫“牍”,100 字以下的写在木板上。简策版牍形式的书在我国古代流行约 1700 年。西汉著名史学家、文学家司马迁编撰的我国第一部通史巨著《史记》,全书共 526500 字,以每根简书写 30 字计算,就需要近 2 万根竹简。历史记载,秦始皇亲理朝政,每天批阅文件(简策)的重量就有 50 公斤。在简策版牍出现的同时,一种书写在丝织品上的书——“帛书”出现了。帛书的基本体现方式是文字与图画。

纸写的书一般称为卷轴阶段。东汉蔡伦改进造纸技术后,纸写书逐渐盛行。纸写书就是将文字写在纸上,再按书写顺序把纸粘合起来,前后裱成长卷,有轴可以舒卷,故也称卷轴。古代纸写书从三国末年到五代后汉末年逐渐盛行,后为印本书所替代。到了南北朝,卷轴装的纸写书基本上取代了简书、帛书。迄今为止,考古发现最早的卷轴装的纸写书是晋代陈寿编著的《三国志·吴志》残卷。我国历史上有两部著名的纸写书,一部是成书于明永乐六年(1408 年)的《永乐大典》,全书 22,937 卷,装订成 11,095 册,约 3.7 亿字,当时抄写了正副本两部。另一部是成书于清乾隆四十七年的《四库全书》,全书 79,337 卷,装订成 36300 册,约 7.75 亿字,当时抄写了七部。

印本时代的书——手工业印刷阶段。从唐代到鸦片战争,我国书籍进入了印本时代,印刷术成为书籍的主要生产方式,书籍已定形为册、页,并逐步发展成为线装的形式。其分期大致可分为雕版印书时期、泥活字印书时期、木活字印书时期、铜活字印时期。但由于这个阶段的印刷技术还都是用手工操作的,故称之为手工业印刷阶段。

手工业印刷阶段大致可分为四个时期:

(1) 雕版印书时期

从唐代开始。雕版印刷术,是先将文字或图画用毛笔书写或描绘在一张薄纸上,再将纸反贴在经过加工的雕版板材上,然后再用刻刀刻掉无字、无图部分,使字、图部分凸起,这就是印版。印书时,在印版上刷上墨,印在纸上,即成书页。雕版印刷术的出现,从根本上改变了自商周至唐初近二千年来用手写方式写书的状况。纸的发明,为雕版书技术的产生和发展提供了重要的物质基础,而雕版印书技术的出现又对书籍的大量印刷和广泛传播提供了条件。

(2) 泥活字印书时期

北宋毕升发明了泥活字印刷术,比德国人谷登堡发明金属活字版印刷早了 400 年。

(3) 木活字印书时期

元代农学家王桢于 1298 年所造。

历史上用木活字印刷术的重要著作,清乾隆 56 年程伟元、高鹗合作印制的《红楼梦》120 回本(曹雪芹的《石头记》传世共 80 回,高鹗续写了后 40 回并题名为《红楼梦》),此前流传的都是手抄本。

(4) 铜活字印书时期

1488 年无锡、常州、苏州、南京一带,广泛应用铜活字印书。铜活字印书是以铜字铸成的反文单字,用以排版印刷的书。我国铜活字印成的最有名的书籍,是清雍正六年,用大、小两号铜字印成的《古代图书集成》。

印本时代的书籍装订形式:

从唐代后期起,先后出现了五种装订形式:

(1) 经折装。也叫“折子装”。如:奏折。

(2) 旋风装。也叫旋风叶。

(3) 蝴蝶装。是宋代书籍装帧的主要形式,蝴蝶装的特点是折口两面在展开时是连接为一个完整版面形式,便于对图画的阅读和欣赏。

(4) 包背装,始于元代,盛于明、清,是将书页的正面正折,版心向外,已接近现代书装订形式。

(5) 线装。明末清初盛行,是中国书籍装订最完整的形式,线装与包背装大致相同,区别在于封面和装订,线装书不是用整张封面纸包上,而是采用上下各一张,不包书脊,装订是将书外口戳齐、书脊剪齐后,在书脊边打 4 个眼穿线装订,多卷本常加函套。传统的书籍装订形式发展到线装,已基本上达到完善的境地,现代出版的许多古籍书仍采用线装形式。

现代的书——机械化印刷阶段。现代印刷肇始于 15 世纪中叶,公元 1473 至 1450 年间,德国人谷登堡发明铅合金活字和压力式印刷。从此西方印刷术开始兴起雕刻凹版术、蚀刻凹版技术、间接转印技术及胶印技术相继出现。清代后期,现代印刷技术传入我国,逐渐取代了过去传统的雕版印刷和泥活字、木活字、铜活字的手工印书方式,书籍装订也由线装发展为平装和精装。我国书籍印装进入了近现代印刷技术新时期。到了 20 世纪,现代印刷技术飞跃发展,激光照排、电分制版、胶印印刷、装订联动等新技术的推广和使用,使现代书的印装效果和质量得到了空前的提高。

中国当代的毕昇——王选,中国科学院院士、中国工程院院士。1937 年 2 月 5 日

生于上海，江苏无锡人，2006年2月13日在北京逝世。他所发明的方正汉字激光照排技术，占领国内报业和书刊出版业99%的市场；80%的海外华文报业市场；方正的日文照排技术已占领了日本报刊1/3的市场。汉字激光照排技术真正实现了中国自主知识产权技术的国际化。

现代书的装帧形式可分为：平装、精装、活页装、散页装四种，现代书的装订方式可分为：线订（包括缝纫机订、三眼线订、锁线订）、铁丝订（包括平订、骑马订）、胶粘订（亦称无线装订）、打孔穿带、螺旋形丝圈贯穿等。

2. 图书的版本、装帧和载体形态

同一部书由于编辑、印刷、出版者以及出版时间、装订形式的不同而形成的各种本子，称为版本。现代图书的版本一般有三种：

- （1）同一种书内容相同的版本，有初版本、重版本，单行本、合订本，影印本等；
- （2）同一种书内容有所不同的版本，有增订本、修订本、缩写本、节本、洁本、重写本、通俗本等；
- （3）同一种书装帧形式不同的版本，有平装本、精装本、线装本、袖珍本等。

版本常识介绍：

【初版本】图书第一次出版称为初版。也叫第一版。

【重印本】是在初版的基础上重印的书。从第一版第一次印刷计算起，每重印一次就在印次上累计标明，版次不变。

【重版本】是指图书重新出版，内容有重要变更，版次更改，印次累计。

【修订本】指书籍的文字和内容经重大修改订正后重排再印的版本。

【增订本】指书籍内容做了较大的增补改订（如增添章节、补充材料、修正文字）后，另行排印出版的版本。修订本和增订本的版次均应改为第二版，但印次仍与第一版相接累计标明。

【单行本】是指将全集、选集或丛书的某一篇单独出版的书。将报刊的长篇文章编印出版的小册子也是单行本。

【合订本】是将多册单行本（包括活页文选）合订为一册出版的图书。报纸、杂志合订成集，也称合订本。

【汇编本】是将有关文章或有关文件汇编在一起的图书。

【缩写本】是在保留原著的主要情节和特色的前提下，对中外名著的内容压缩编写而成的版本。

【选本】从大部头著作或从多种著作中选出部分篇章编印出版的图书。

【节本】是对原著进行删节或节取某些章节编印出版的图书。

【洁本】指将某些文学作品中夹杂的淫秽、迷信等不健康内容删除后出版的版本。

【译本】将外文译成中文、将中文译成外文，或将不同民族的文字互译。

【珍本】指有历史、艺术或科学价值而又罕见的珍贵图书版本。

【善本】指版本较好的古本书。即经过精细校勘，错误较少，在学术或艺术价值上比一般本子优异的刻本或抄本。

【孤本】指一种图书的某一版本仅有一部流传于世。也指未刊印问世的手稿，或原物已散失的碑贴拓本仅留存部分。