



跨世纪计算机实用技术丛书

文图编辑排版新技术

——原理、标准与应用

徐福培 主 编
宋增民



东南大学出版社

文图编辑排版新技术

——原理、标准与应用

主 编 徐福培 宋增民

编著者 (以姓氏笔划为序)

吕纪力 刘山鹏 戴建东

主 审 蔡杰



C0405747

东南大学出版社

内 容 提 要

计算机进入出版业,引发了排版印刷技术的革命,而文图编辑排版系统正是这场革命的助推器。衡量一个编辑排版系统优劣的最主要的标志就是文字输出质量的好坏与速度的快慢。为此本书首先介绍了编辑排版与输出的基础技术:字符集与编码,字形描述与字库组织;接着介绍了国际标准 CJK 大字符(ISO 10646)的编码规则、原理和使用方法;为进一步与国际接轨,还详细介绍了与编辑排版有关的一系列标准及其使用方法;最后以北大方正为例,具体介绍了北大方正系统的组成、排版语言及各种操作。

本书取材新颖、内容丰富、实用性强,适用于从事电子出版、办公自动化等领域的工作人员使用;也可以作为大专院校有关专业的教学参考书。

责任编辑 张 煦

责任校对 王小然

文图编辑排版新技术 ——原理、标准与应用

徐福培 宋增民 主编

*

东南大学出版社出版发行

(南京四牌楼2号 邮编210018)

江苏省新华书店经销 常熟市文化照相制版彩印厂印刷

*

开本 787×1092 毫米 1/16 印张 14.875 字数 381 千

1998 年 7 月第 1 版 1998 年 7 月第 1 次印刷

印数:1—3000 册

ISBN 7—81050—340—5/TP·51

定价:20.00 元

如遇印装质量问题,可直接向承印厂调换

地址:常熟市梅李镇通江路21号 邮编:215511

《跨世纪计算机实用技术》丛书总序

我们正处在一个世纪之交的伟大历史时期。当今社会的计算机应用出现了很多新的特征:办公自动化系统和管理信息系统的实现使社会、生产的管理完全改变了传统的方式;集计算机辅助设计、制造和决策管理为一体的计算机集成制造系统使生产制造过程得到新的飞跃;计算机网络 INTERNET 将世界连成一体,并以一种神奇的方式推动着社会的变革;计算机进入出版业,出现了排版印刷技术的革命;多媒体技术融入社会,更以一种非同寻常的力量推动着信息社会的发展,人们就在这瞬息万变的发展中即将跨入一个新的世纪。

随着 21 世纪的临近,人们面临着各种严峻的挑战,我们必须把人才的培养作为迎接挑战最紧迫的对策。为了适应世纪之交计算机人才培养的需要,我们向读者奉献一套《跨世纪计算机实用技术丛书》。作者编写这套丛书的宗旨为:

1. 介绍 20 世纪最后 10 年计算机技术在其重要应用领域的现状、发展方向以及前景展望。希望读者从这套丛书中,得到感知,获得鼓舞和力量。

2. 力求面向实际,作为实用技术的工具和指南。希望读者得益于它,在世纪之交的计算机应用工作中,能够得到具体的帮助,取得实效。

3. 从本世纪之末到下世纪之始,计算机科学与技术领域新事物肯定层出不穷,本丛书将不断把新颖实用的内容奉献给读者,作者们将与读者一起跨越世纪,在各领域计算机应用中不断前进。

这一套丛书的内容涉及数据库系统、网络通信、多媒体、图像处理、电子出版、系统仿真、计算机辅助设计、计算机硬件及软件新平台、操作系统以及语言等方面。我们尽了很大的努力期望本丛书能很好地为读者服务,但由于水平所限,难免有错误和疏漏之处,希望读者和专家们不吝赐教。

《跨世纪计算机实用技术丛书》编委会

1997 年 5 月

《跨世纪计算机实用技术》丛书

编委会名单

主 编：徐福培

副主编：夏德深 孙志挥

编 委：(以姓氏笔划为序)

孙志挥	吴乐南	陈天授
陈廷标	陈金水	周先华
夏德深	范建新	徐福培
黄凤英	傅德胜	

前 言

文图编辑排版系统已在办公自动化、电子出版等领域得到了广泛的应用,方正、华光等桌面出版系统是我国计算机应用领域中的成功范例之一。近年来如雨后春笋般发展起来的光盘或网络出版物,其基础技术之一仍是文、图、声、影像等元素的编辑排版技术。如何进一步提高文图等元素的编辑排版质量,加快彩色图文编辑排版和多媒体电子出版系统的推广应用,又如何进一步走国际标准化道路,开拓国际市场,扩大应用范围,仍然是人们关心的热点之一。为此,我们在多年来研究开发各类文图编辑排版系统和相关国际标准的基础上,编写此书,以进一步推动文图编辑排版系统在我国推广使用。

无论在何种电子出版系统中,编辑排版操作的对象大量的仍然是文字信息,因此,文字输出质量的好坏与速度的快慢是衡量一个编辑排版系统优劣的最主要的标志。为此本书首先介绍了编辑排版与输出的基础技术:字符集与编码、字形描述与字库组织。本书还介绍了国际标准 CJK 大字符集(ISO 10646)的编码规则、原理和使用方法,以帮助读者对大字符集的了解和使用。Adobe Type 1 和 True Type 是用于字形描述与字库组织事实上的工业标准,字形质量好、精度高。本书详细叙述了这两类字库技术的原理、字形的描述方法、字库的组织结构以及各自的特点,以帮助读者对这两种字库技术的学习与理解。

开放性、走国际标准化道路,是当今计算机发展的根本道路,编辑排版系统也是一样。为此,本书详细介绍了与编辑排版有关的一系列标准和它们的使用方法,这包括如何书写文档的通用标准置标语言 SGML 和如何组织版面输出的标准页描述 SPDL(包括 PostScript 语言)。本书还给出了适合于超媒体、超文本编辑排版的事实上的标准,即超文本置标语言 HTML 和超媒体/基于时间的结构语言 HyTime。最后还简单介绍了与文图彩色输出有关的彩色空间技术和彩色输出标准。

本书取材新颖,内容丰富实用,还给出了一些应用实例,对从事电子出版、办公自动化等领域的应用人员有很好的参考价值,它可以作为大专院校有关专业的教学参考书。

本书的 1~7 章由徐福培教授主持编写,参加编写的还有吕纪南、戴建东、刘鹏等同志,8、9 两章由宋增民教授编写,蔡士杰教授审阅了全书,纪晓云同志在书稿录入方面做了大量艰苦细致的工作,在此一并表示感谢。由于时间仓促及限于水平,书中错误和不妥之处在所难免,敬请读者批评指教,以便再版时改正。

编 者

1998.5

目 录

1 概论	1
1.1 历史与发展	1
1.1.1 告别“铅与火”,迎来“光与电”	1
1.1.2 多媒体光盘与网络出版物	2
1.2 系统组成与工作流程	3
1.3 标准化	4
2 字符集与编码	7
2.1 ASCII 码与 ISO 646 标准	7
2.2 汉字编码字符集	8
2.2.1 汉字编码基本集 GB2312-80(简称 GO)	9
2.2.2 汉字编码辅助集	10
2.3 ISO 10646/GB 13000 通用多八位编码字符集	10
2.3.1 背景	10
2.3.2 UCS 诞生及其主要特点	12
2.3.3 UCS 总体结构	12
2.3.4 UCS 子集	14
2.3.5 实现等级	15
2.3.6 统一的中、日、韩汉字	15
2.3.7 推广应用	18
2.4 汉字扩展内码规范——GBK	19
2.4.1 扩展汉字内码规范的原则	19
2.4.2 规范的内容	20
2.4.3 应用	22
3 字形与字库技术	23
3.1 概述	23
3.1.1 历史回顾	23
3.1.2 点阵字	24
3.1.3 向量字	24
3.1.4 曲线字	24
3.2 Type 1 字库	25
3.2.1 字库结构	25
3.2.2 字形设计	30

3.2.3	Type 1 的编码	32
3.2.4	Type 1 的加密	34
3.2.5	Type 1 的 hinting 技术	35
3.2.6	Type 1 操作符	36
3.2.7	字库实例	38
3.3	True Type 字库	43
3.3.1	曲线描述的数学基础	43
3.3.2	True Type 字库文件格式	47
3.3.3	字形数据的读取	55
3.3.4	True Type 汉字库的应用	57
4	彩色空间技术与标准	60
4.1	引言	60
4.2	彩色模型	61
4.2.1	HSV 模型和 HLS 模型	61
4.2.2	RGB 模型	62
4.2.3	CMY(或 CMYK)模型	62
4.2.4	YIQ 模型	63
4.2.5	CIE 模型	63
4.3	彩色空间转换	63
4.3.1	设备彩色空间之间的转换	63
4.3.2	CIE 彩色空间到设备彩色空间的转换	64
4.4	彩色生成	65
4.5	分色	66
4.6	半色调加网	66
4.6.1	有序抖动技术	66
4.6.2	牺牲分辨率的半色调技术	67
4.6.3	调频网技术	68
4.7	彩色标准	68
4.7.1	显色表色标准	68
4.7.2	混色表色标准	69
5	标准通用置标语言 SGML	71
5.1	引言	71
5.2	SGML 产生背景	71
5.3	SGML 基本概念	73
5.3.1	几个重要概念	73
5.3.2	SGML 的四类置标	74
5.4	SGML 文档组成	75

5.4.1	SGML 说明	75
5.4.2	文档类型定义 DTD	77
5.4.3	文档实例	83
5.5	SGML 系统	86
5.5.1	SGML 语言分析器模型	86
5.5.2	初始化	86
5.5.3	动态过程映射	87
5.5.4	错误处理	87
5.6	SGML 应用与发展方向	88
5.6.1	SGML 应用	88
5.6.2	应用发展方向	90
6	页面描述语言 PostScript 和 SPDL	91
6.1	引言	91
6.1.1	背景	91
6.1.2	特性与功能	91
6.2	通用程序设计语言	95
6.2.1	解释器	95
6.2.2	语法	95
6.2.3	数据类型和对象	96
6.2.4	栈	98
6.2.5	执行	98
6.2.6	存储管理	98
6.2.7	资源	99
6.2.8	编码	99
6.2.9	过滤器	100
6.3	图形与图像	100
6.3.1	成像模型	101
6.3.2	图形状态	102
6.3.3	坐标系和变换	102
6.3.4	着色	103
6.3.5	彩色空间	104
6.3.6	图像	107
6.3.7	彩色加网(半色调技术)	113
6.4	正文输出	115
6.4.1	字库的组织与使用	115
6.4.2	字库词典	116
6.4.3	字库类型	118
6.5	中文页面描述语言及其解释器	119

6.5.1 中文页面描述语言解释器·····	119
6.5.2 CPDL 的程序设计·····	123
7 超文本置标语言 HTML ·····	130
7.1 引言 ·····	130
7.2 HTML 的由来及特点·····	130
7.3 HTML 的结构和语句概要·····	131
7.4 文件头元素 ·····	133
7.5 文件体元素 ·····	136
7.5.1 块格式化元素 ·····	137
7.5.2 列表元素·····	139
7.5.3 超文本链接元素 ·····	141
7.5.4 图像元素·····	143
7.5.5 数据采集元素 ·····	143
7.5.6 字符格式化元素 ·····	147
7.5.7 信息类型元素 ·····	148
7.5.8 表格元素·····	149
7.5.9 JAVA APPLET 嵌入元素 ·····	151
8 北大方正电子出版系统 ·····	153
8.1 系统概况 ·····	153
8.2 方正 BDDOS ·····	156
8.2.1 BDDOS 的构成和功能 ·····	156
8.2.2 动态键盘的功能与操作 ·····	156
8.2.3 系统中符号的输入 ·····	157
8.3 方正 FE 编辑系统·····	157
8.3.1 进入编辑·····	157
8.3.2 各种编辑功能 ·····	158
8.3.3 FE 的其它功能·····	163
8.4 方正书刊排版系统及其操作 ·····	164
8.4.1 排版功能和操作 ·····	164
8.4.2 版式定义功能和操作 ·····	165
8.4.3 显示功能和操作 ·····	168
8.4.4 打印功能和操作 ·····	169
8.4.5 图表处理功能和操作 ·····	169
8.4.6 工具功能和操作 ·····	170
8.5 方正发排系统 ·····	171
8.5.1 93 发排系统的启动 ·····	171
8.5.2 使用方法·····	172

8.6 方正 PSP 曲线系统	177
8.6.1 输出(File)菜单	178
8.6.2 显示(Show)菜单	185
8.6.3 选项(Options)菜单	185
9 北大方正书刊排版语言	187
9.1 概述	187
9.1.1 BD 排版语言基础	187
9.1.2 文件名的约定	189
9.1.3 基本概念	190
9.2 基本排版命令	193
9.2.1 强制结束类命令	194
9.2.2 字体字号类命令	195
9.2.3 字距、行距类命令	197
9.2.4 行调整类命令	199
9.2.5 简单表格	200
9.3 6.0 版的其它排版命令	203
9.3.1 字符控制类命令	203
9.3.2 期刊杂志的排版命令	205
9.3.3 数学、化学类命令	208
9.3.4 版面控制类命令	212
9.3.5 框线类命令	214
9.3.6 书眉、脚注及其它命令	215
9.4 7.0 版新增功能	218
9.4.1 文字符号的处理	219
9.4.2 版面控制	220
9.4.3 7.0 版更改注解对照表	222
参考文献	226

● 电子出版 ● 编辑排版 ● 标准化

1.1 历史与发展

1.1.1 告别“铅与火”,迎来“光与电”

中国古代四大发明中的两项——造纸术和活字印刷术(均与出版印刷业密切相关)奠定了出版印刷业的基础,是人类文明发展史上的重要里程碑。从汉代发明造纸术后出现的石刻,到隋唐出现的纸版和木版雕刻印刷术,成为世界上印刷术的发展源头。宋代毕昇发明活字印刷,为推动印刷业的发展作出了巨大贡献。约四百年后,德国的 Gutonberg(1395 - 1468)发明了用铅合金制作活字,开创了现代铅字排版技术。本世纪初,英国人发明了机械式打字机,后来日本人又发明了汉字打字机。直到 70 年代初,由于计算机技术(特别是微型机)的发展和应用于出版印刷业,开创了现代出版业(电子出版)的先河,从此传统的出版印刷业完全进入了一个崭新的时代——告别“铅与火”,进入了“光与电”的时代。

我国对电子出版技术的研究起步较早,1974 年 8 月,电子工业部、新华社、一机部(机械工业部)、国家出版局、中科院等单位向国务院和国家计委联合申请“汉字信息处理”重大工程项目,定名为“748 工程”。该工程分为精密汉字照排系统、中文情报检索系统和汉字通信系统三部分。经过 20 多年众多单位的联合攻关,在精密汉字照排系统方面取得了突出的成果。

从 1975 年开始,北京大学以王选教授为代表的一批科技工作者,开始了对精密汉字照排系统的研究,以后并与山东潍坊计算机厂等单位协同攻关,于 1979 年推出了华光 1 原理样机,并排出了 8 开报纸底片,这是国产电子出版系统的重要突破。这一凝结了中华科技人员智慧和心血的华光 1 在性能上超过了当时国外的先进产品,确定了我国在中文电子出版领域的领先地位。从此以后的华光 2 至华光 4,逐步实现了产业化。在此期间,经济日报利用华光 4 系统正式排报,成为我国第一家整页编排输出的中文报纸。在 1987 年 12 月隆重召开的华光系统生产定型会和《经济日报》照排系统国家验收会上,当时的中国科协主席周培源先生热情洋溢地说:“现在计算机能处理汉字了,能排版,可以宏扬中华文化,它的意义不亚于原子弹爆炸”。

《经济日报》的成功,树立了一个好样板。到了 90 年代初,随着华光 5 和方正 91 系统的分别推出,精密照排系统逐步在我国普及。

从 80 年代中期开始,国际上电子出版系统开始向彩色系统发展。特别是 90 年代初随

着页面描述语言 PostScript Level 2 的推出,高质量的彩色电子出版系统出现了蓬勃发展的局面,外国大公司凭借 PostScript 和新一代硬件正在窥视着中国市场。在这新形势下,随着 748 工程发展起来的我国电子出版业又面临着严峻的挑战。

华光和方正集团依靠 20 年来集聚的经济和技术实力开始向 PostScript 和彩色出版冲击。1993 年两大集团分别推出了以 PostScript Level 2 为核心的方正 93 和华光 6 电子出版系统,实现了与国际先进技术的接轨,印刷业的“彩色革命”也逐渐到来。

通过以上简单回顾可以看到,从 748 工程的立项,在国家有关部门的大力扶植下,经过众多科技工作者的不懈努力,我国的电子出版技术不断发展和进步。由于实现了从“铅与火”到“光与电”的革命,目前我国报纸已从 1500 余种发展到 4800 余种,各种期刊从不到 2000 种发展到 6500 种,准刊率由 93% 提高到 96%;与 60 年代相比,我国年排字能力从 90 亿字提高到 130 亿字,突破了排版瓶颈;书刊出版周期从 300 多天缩短到 50~60 天,出版印刷质量提高,我国的文化事业出现了空前的繁荣局面。

1.1.2 多媒体光盘与网络出版物

进入 90 年代,随着多媒体技术、大容量光盘技术和网络技术的飞速发展,推动和促进了多媒体电子出版物的诞生、应用与发展。什么是多媒体电子出版物呢?国家新闻出版署对电子出版物的定义是“电子出版物,系指以数字代码方式将图、文、声、像等信息系统存储在磁、光、电介质上,通过计算机或类似设备阅读使用,并可复制发行的大众传播媒体。”这个定义明确了多媒体电子出版物的几个重要特点:信息多媒体化,存储形式和载体电子化、数字化,阅读计算机化。多媒体电子出版物与传统出版物究竟有哪些区别呢?我们可以归纳为以下几个方面:

① 出版载体不同:传统出版物以纸张作为唯一载体,而电子出版物以 CD-ROM 光盘、IC 卡等作为载体,其容量大。一张光盘可存放 20 卷本的百科全书或一二年内每天数十个版面的报纸,体积小,容易收藏与管理。

② 信息表现方式不同:传统出版物以表现图文为主,信息类型是平面的、静态的;而多媒体电子出版物除图文以外,还表现声音、动画、影像等立体的、多元的、动态的信息,因此更具吸引力。

③ 制作方式不同:传统出版物在印刷以前的工作(称为印前作业,包括录入、排版、编辑、制版等),目前基本上已可由计算机完成,而印刷、装订和发行主要由人工完成;而电子出版物从录入、排版到出版主要均由计算机完成,制作效率高,出版迅速。

④ 使用方式不同:电子出版物翻阅方式全部电子化,且可以借助超文本技术和计算机交互能力,容许读者在任一节点作纵向或横向的深入探索或查询。在网络方式下,还可以满足多人同时阅读一篇作品或一本书的需要。

⑤ 发行方式不同:特别是网络出版物,一经出版立即上网,发行速度快。

正因为电子出版物有上述特点,从 80 年代后期到今天不到 10 年时间内,电子出版物得到了迅速的发展。以美国为例,已有 2100 多家 CD-ROM 出版商,许多著名的出版社纷纷改编并推出光盘版图书。如美国著名的 Grolier 百科全书、Compton 百科全书、Columbia 百科全书等。欧洲市场也很活跃,其电子出版物销售总额已从 1992 年的 2.55 亿美元上升到 1996 年的 11.9 亿美元。

我国光盘出版物也得到了迅速发展,目前已有 500 多种光盘出版物,其中比较有影响的有《中国药典》、《人民日报全文数据库》、《中国法律大全》和《邓小平文选》等。网络出版中,以电子报纸最为迅速。目前世界上已有 2000 多种报纸上网,如美国的“纽约时报”、“华尔街日报”,我国的“人民日报”、“广州日报”和“中国计算机报”等。

1.2 系统组成与工作流程

电子出版系统一般由三部分组成:文图输入、编辑排版和输出系统。

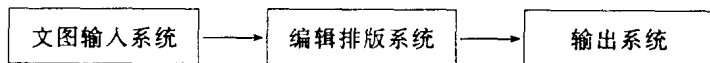


图 1.1 电子出版系统组成

1) 文图输入系统

对于传统的出版系统(以纸介质为出版载体),该部分包括文字的录入,图形的绘制,图像或照片的扫描输入等;而对于多媒体电子出版物,除了完成上述输入功能外,还应包括声音的采集,动画的制作,影视资料制作等部分。这些工作均可以在计算机平台上通过相应的软件来完成。文字的录入可以通过 WPS、Word 等字处理软件来完成;图形与图像的制作可以通过 CorelDraw、Freehand、Photoshop、Photostyle 等图形、图像软件来实现;动画制作可以借助 Director、3Dstudio、Animator pro 等软件完成。对于多媒体光盘出版物,则有专门的写作工具,目前较著名的有美国 Macromedia 公司的 Authorware, Asymetrix 公司的 Toolbook 等;国内也推出了一些有一定影响的产品,如方正“奥思多媒体制作系统”,北京汉声公司的“洪图多媒体编著系统”等。对于网络出版物,则有 WWW 文档的写作工具,如微软的 FrontPage、Netscape 的 Navigator Gold 等,它们可以直接书写 HTML 格式文档,也可以通过文档格式转换程序把由 Word 等字处理软件书写的文档转换成 HTML 格式来完成。

2) 编辑排版系统

编辑排版系统的主要功能是把由输入系统送来的文字、图形、图像、声音、动画等元素组织排列成出版者需要的格式。例如,整个文档组织格式的指定,标题的选择,字体、字号的确定,图形与图像的大小、位置和颜色的指定等等。这一般通过在这些元素的前后加上适当的标记(mark)来完成,生成相应的置标文件,最后送给输出系统处理。

对于传统的出版物而言,编辑排版系统根据其功能和应用范围的不同分成以下几类:办公文书处理系统、桌面出版系统和精密照排系统等。

办公文书系统,可以完成一般办公室内文件、表格、简单图形和图像的编辑、排版等功能,利用 WPS、Word 等软件就可以完成。

桌面出版系统(Desktop Publishing System—DTP)和精密照排系统从编辑排版的功能来看,没有太大的差别。它们均要求编辑排版功能比较丰富完善,排出的版面规范、美观,其主要差别是系统的精度要求和规模不同。这两类系统的编辑排版软件有 PageMaker、QuarkXpress 以及国内的华光、方正书报排版软件等,它们均能较好地满足用户的多种需要。

出版物中还有各种图形、图像和照片等,需要与文字混合编排,这可以通过一些专用的图形、图像编辑软件来完成,如前面提到的 CorelDraw、Photoshop、Photostyle 等,这些软件还可

以对图形、图像进行缩放、修饰、校色等多种处理。

3) 输出系统

输出系统的主要功能是对由编辑排版后生成的文件进行还原输出。对于不同类型的出版系统,由于其功能、应用范围不同,输出系统也有多种不同的类型。

办公文书处理系统输出比较简单,一般编辑排版好的文件经过解释还原后可以在打印机(点阵打印机、激光印字机、喷墨打印机等)上直接输出。

桌面出版系统输出以激光印字机为主,输出分辨率在 300~600dpi 之间。其输出介质以纸张为主,也可以输出到透明聚脂薄膜或描图纸上,直接用于轻印刷制版,所以这类系统也叫“轻印刷系统”。

精密照排系统(包括彩色电子排版系统)主要用于各大出版社、报社、印刷厂等单位,专业性强,精度要求高、速度快,所以其输出设备一般为大幅面、高精度的激光照排机或电分机,其分辨率至少在 1000dpi 以上,甚至 3000dpi 以上。这类输出设备的输出载体一般为胶片(彩色输出时为 4 张分色片),胶片经过冲洗后,再用来制作印版(PS 版),最后上胶印机印刷。随着印刷制版的工艺革命,目前已开始推广计算机直接出印版(无胶片)的系统。如 Agfa 公司的 Lithostar 电脑直出印版系统,由计算机编辑排版好的版面文件经解释还原后直接曝光在铝版上,经冲洗后便可进行印刷。目前还出现了计算机直接上印刷机印刷的系统,省去了制作胶片和印版的过程。如 Agfa 公司最近推出的 Chromapress 彩色数码印刷系统,它每分钟可以输出 35 张 A4 双面彩色印刷品,其效果可以与胶印 2400dpi、150 线的印刷品媲美。

对于精密照排系统而言,无论它采用上述哪一种方式输出,经过编辑排版生成的版面文件,目前一般均转换成由页面描述语言 PostScript Level 2 描述的格式,再经过页面描述语言解释器(或称删格图像处理器 RIP)的解释,把 .PS(或 .EPS)文件解释还原成全点阵信息,然后由照排机生成分色片或由直印系统直接输出。因此,这类输出系统中解释器或 RIP 的处理质量和速度至关重要。

对于网络和光盘出版物,读者阅读这些出版物一定得借助于计算机才能进行,因此提供一个好的阅读器(或称浏览器)是关键。阅读器必须要有良好的用户界面、导航能力和检索能力。在 WWW 浏览器中,Netscape 的 Navigator 和微软的 Internet explorer 均是较成功的产品,颇受用户的青睐。

1.3 标准化

传统的出版业在告别“铅与火”,进入“光与电”的时代后,意味着出版业将包括以下三个电子化过程:出版过程电子化、出版物电子化和出版信息服务电子化。在这三个电子化过程中,为了便于电子文档的传输、交换、阅读和检索,走开放式、国际标准化道路具有十分重要的意义。例如,在早期的华光或方正电子出版系统中,都是通过专有的系统排版语言对需排版的内容进行过程性置标,然后利用专用解释器对这些过程性置标进行处理,从而产生页面输出。这一过程存在以下两点不足之处:

- 已完成置标的文档很难在不同的系统之间进行交换,为了使得该文档内容能在不同系统上应用,必须重新置标,这必然造成极大的浪费;

- 不便于对文档的内容进行信息检索,也不便于与数据库进行交互。

为了克服上述弊端,走国际化道路是唯一出路。本书介绍了文图编辑排版方面的有关标准,包括字符编码标准和字形描述的事实上工业标准,以及电子文档结构格式标准和版面描述的有关标准等,而有关图、声、影像信息本身的表现、压缩、存储等标准不在本书内涉及。

这里,首先简单叙述本书中所涉及的各类编辑排版标准的功能与特点,详细内容可见有关章节。

(1) 通用多八位编码字符集 UCS(Universal Multiple - Octet Code Charactor Set)

ISO10646 (GB 13000),这是国际标准化组织历经八载,数易其稿,最后于 1993 年 5 月正式发布的大字符集编码标准。该标准定义了两种字符编码格式:

- UCS—4:一个字符用组、面、行、字位共 4 个字节来表示;
- UCS—2:选用 00 组、00 平面(基本多文种平面)中的行、字位两个字节来表示一个字符,目前 ISO 也只定义了该平面,这也是工业界和用户目前最关心的一种格式。

(2) Adobe Type 1 和 True Type 字形与字库技术

它们分别是由 Adobe 公司和 Apple 等公司先后推出的用于字形描述和字库组织的事实上的工业标准。在这两种格式的字库中,其字形均用曲线来描述,精度高、字形质量好;且小字均采用了 hinting 处理技术,以保证小字不断笔、不粘笔;字库的组织 and 编辑方式灵活,以适应不同用户和多种场合的需要。

(3) 标准通用置标 SGML

ISO8879 - 1986,这是用于描述文档逻辑结构和内容的语言。SGML 用描述性置标在文档信息中区分不同的元素,用文档类型定义(简称 DTD)描述一类文档的结构,说明该文档有哪些元素及元素之间的关系,用 SGML 说明来指出文档中所用字符集、置标用字符和语法等。由于 SGML 描述文档的结构,它将文档结构和内容与对该文档的处理区分开来,从而使文档与具体应用无关,并可多次利用。用 SGML 置标的文档,出版商可以毫不困难地将同一内容组成不同的出版物。

(4) 文档类型语义和规格说明语言 DSSSL

ISO/IEC 10179 (1994 年),SGML 描述了文档的逻辑结构,但并没有指出如何对 SGML 中的元素进行格式化,而 DSSSL 正是描述文档格式化的处理语言,从而可在不同系统之间实现格式化文档的交换。DSSSL 提供了两种不同类型的变换:SGML 树转换处理 STTP 和 SGML 树格式化处理 STFP。

- STTP:通过对 SGML 树型结构文档进行重组、复制或裁剪等操作,组合成新文档;
- STFP:对 SGML 文档加注格式化要求,最终形成版式文档(如形成 SPDL 文档)。

(5) 页面描述语言 PostScript 和 SPDL

页面描述语言 PostScript 由 Adobe 公司于 1982 年正式推出(Level 1),1990 年又推出了面向彩色印刷的 Level 2。该语言的主要功能是在输出页面上描述文字、图形和图像。由于这种描述高效,且与设备无关,因此,目前已广泛应用于电子出版领域,成为事实上的工业标准。1991 年,国际标准化组织 ISO 在吸取 PostScript 等页面描述语言优点的基础上,推出了标准页面描述语言 SPDL。

图 1.2 示出一个建立在上述标准基础上的电子出版系统模型。

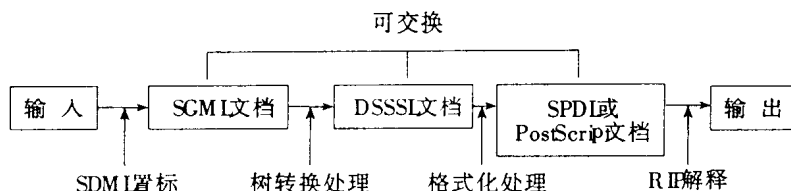


图 1.2 电子出版系统模型

输入信息经过 SGML 置标形成 SGML 文档,该文档可用于交换或存入数据库,也可以经过 DSSSL 语言的树转换处理(重组、复制或裁剪)形成新的 SGML 文档。这类文档再经过格式化处理转换成 SPDL 或 PostScript 页面描述语言描述的版面文件,最后经过解释器(RIP)还原输出,生成分色片或直接送彩色数码印刷系统。

这一系统中的每一步生成的文档均符合国际标准,便于不同系统之间的交换。

(6) 超文本置标语言 HTML

HTML 是 SGML 的一个子集或者说一个典型应用,1996 年 4 月 WWW 联盟已推出了 HTML3.2 版。HTML 除了具有 SGML 通用、跨平台等特点外,还具有小巧、灵活、易掌握等优点,从而被 WWW 采用为创建超文本的语言。HTML 通过置标和属性对一段文本的语义进行描述,此外它还提供了由一个文档到另一个文档,或在一个文档内部不同部分之间的超文本链接,从而实现环球漫游。

上述电子出版模型(图 1.2)中的 SGML 置标过程可以由 HTML 语言来完成,生成的 HTML 文档在 Internet 网上发行传送,而文档内容的显示风格、字符大小、行间距等由浏览器决定(即完成格式化过程)。

HTML 与 SGML 一样,着重于文档结构的描述,对文档中各类信息的关系、表现、交换等缺少支持。为此,国际标准化组织正在研究制订用于多媒体信息编码和交换的国际标准——MHEG,以解决在不同平台、不同应用环境下多媒体文档信息的通用编码表示、表现和交换等问题。

(7) 超媒体/时基结构化语言 HyTime

ISO/IEC 12083,1991 年 10 月推出。HyTime 是 SGML 在吸取 SMDL(Standard Music Description Language)标准有关内容后扩展而成,用以表示超文本、多媒体、超媒体和时基文档的逻辑结构。由于 HyTime 实现较为复杂,目前厂商支持尚少。