

电子出版物及其制作技术

陈光祚 主编



武汉大学出版社

电子出版物及其制作技术

主 编	陈光祚
编写人员	陈光祚
	王在平
	陈 睿
	任世勇

武汉大学出版社

(鄂) 新登字 09 号

图书在版编目 (CIP) 数据

电子出版物及其制作技术/陈光祚主编; 陈睿等著

· 一武汉: 武汉大学出版社, 1994. 8

ISBN 7-307-01836-6

I. 电...

I. ①陈... ②陈...

Ⅲ. ①电子出版物-理论-研究

②电子出版物-技术-研究

N. G230.7 TP399: G23

武汉大学出版社出版

(430072 武昌 珞珈山)

湖北省崇阳县印刷厂印刷

新华书店湖北发行所发行

1994 年 8 月第 1 版 1994 年 8 月第 1 次印刷

开本: 787×1092 毫米 1/32 印张: 5.375

字数: 136 千 印数: 1—2000

ISBN 7-307-01836-6/G·250 定价: 4.70 元

前 言

电子出版物是继印刷型出版物之后出现的一种新型出版物。它体现了现代信息技术在出版业中的应用。与印刷型出版物比较，它的信息载体不是纸张，而是磁性或光学信息存储介质；它的信息表达形式不是常规的文字符号，而是计算机可读形式的数据；它能提供计算机化的情报检索、文献中信息量的计量分析和知识项的聚类等功能；它能借助于通讯线路实现高速传输和网络条件下的出版发行。

电子出版物的出现与发展，是出版业的一次革命。它是信息时代的产物和标志。它不仅为出版业增添了一个新的物理形态的出版物，而且它能缩短出版周期、不受印数的束缚、降低出版成本。电子出版物的出现与发展，是人类走向“无纸信息系统”的重要一步。

电子出版物对社会将产生深远影响。它将提高人们读书治学、获取有用信息的效率；它能使文献信息从一出版就置于计算机的控制之下，使出版与信息处理实现了有机结合；它将使图书馆面临一个新的收藏对象，将改变文献处理与服务工作的流程。

我们从1987年开始，就一直从事全文数据库的电子出版物的制作研究。先后得到了三项国家科研基金的支持：国家教委作为文科博士点基金项目1986年下达了“全文数据库研究”计划；中华社会科学科研基金于1991年下达“全文数据库与电子出版物”项目的任务；国家新闻出版署于1992年作

为重点科研项目立项，下达“电子出版物制作技术研究”的项目任务。

从1987年以来，我们的科研项目取得了一系列成绩：1990年10月制作完成《湖北省地方志·大事记》、《中国人民解放军大事记》（解放战争时期）等二个全文数据库，并通过了国家教委的鉴定，发表了有关的研制报告和技术报告；1991年7月与武汉大学出版社合作，制作成功《国共两党关系通史》（150多万字）的电子出版物，并向社会公开发行。这是利用出版过程中的中间产品激光照排的电子文本加工而成、并由出版社正式发行的电子图书。《新闻出版报》及其他中央与地方报纸、中央电视台及地方电视台对此作了报导，并称这是我国第一部电子出版物；经技术上的改进，又制作发行了《中国名胜诗词大辞典》（1993年）、《市场经济大辞典》（1994年）等电子出版物。

制作电子出版物，包括下述技术处理：电子版图书的总体设计；文本格式化；自动与半自动标引；全文数据库建库与检索软件的编制；电子出版物评价准则的确定等等的研究。

我们将研究中的成果在有关刊物中作了发表。例如“论全文检索系统”（《武汉大学学报》（哲学社会科学版）1989年第6期）；《轻印刷的中间产品可以开发成全文数据库并可作为电子出版物发行》（《武汉大学学报》（哲学社会科学版）1991年第6期）；《论我国电子版图书及其制作技术》（《中国出版》1992年第5、6、7期）；“利用全文数据库技术制作电子出版物”（《情报学报》1993年第1期）等。

本书就是在八年来有关电子出版物研究成果的基础上写成的。书中的三个制作电子出版物软件的论述是由研究生撰写的。即第四章，LOCALHIS系统，由陈睿撰写；第四章

FULLSEA 系统由王在平撰写；第四章 TEXT 系统由任世勇撰写。此外，第三章第八节、第四章第四节、第五章第三节由王在平撰写。其它各章节由陈光祚撰写。

我们特别感谢湖北省新闻出版局田胜立副局长、武汉大学出版社牛太臣社长的支持与指导。他们作为新闻出版署下达项目课题组的领导，对本书的编写出版给予了积极的鼓励和支持，并审阅了书稿；武汉大学出版社的毕卫东、李钢等同志，他们作为新闻出版署下达项目的课题组成员，也给此书的问世给予了帮助；武汉大学图书情报学院马费成院长对此书的编写也一直给予了关心。另外，研究生杨熔、朴英哲、何静同学参加了部分校对工作。在此，我们对上述同志表示衷心的感谢。

由于著者水平有限，书中的缺点乃至错误，敬请读者指正。

陈光祚（武汉大学出版社特约编审、
武汉大学图书情报学院教授）
1994 年 5 月于武汉大学

目 录

前言	1
第一章 电子出版物是一种新型出版物	1
第一节 电子出版物的基本特征	1
第二节 电子出版物的范围与种类	6
第三节 电子出版物的独特功能	10
第四节 电子出版物给出版业本身带来的深刻影响	13
第五节 电子出版物发展概况	17
第二章 电子出版物对图书情报工作的影响	23
第一节 电子出版物的出现使文献从一开始就具有 计算机化的情报处理能力	23
第二节 电子出版物将是图书馆藏书的一个新品种	24
第三节 电子出版物是参考咨询的有力工具	27
第四节 电子出版物使文献计量学获得 新的应用领域	31
第五节 电子出版物有助于实现图书内容 索引编制的计算机化	36
第三章 利用全文检索技术制作电子出版物	40
第一节 电子出版物的主体是全文数据库系统	40
第二节 中文电子版图书的基本设计要求	41

第三节	电子文本的格式化处理	46
第四节	自动标引与半自动标引相结合的标引处理 ...	47
第五节	电子版图书应有的基本检索功能	58
第六节	设置后控词表的机制	62
第七节	电子版图书编制的另一些可能的方案	65
第八节	电子出版物功能核对表	70
第九节	我国电子出版物制作中尚待解决的技术问题	76
第四章	全文数据库系统软件实例剖析	84
第一节	《湖北省地方志》、《国共两党关系通史》全文检索系统 LOCALHIS	84
第二节	《中国名胜诗词大辞典》、《市场经济大辞典》全文检索软件 FULLSEA	94
第三节	非 B 树索引结构的电子出版物系统—TEXT	104
第四节	全文检索系统的结构模型设计	120
第五章	三部电子出版物简介	129
第一节	《国共两党关系通史》全文检索电子版简介	129
第二节	《中国名胜诗词大辞典》电子版简介	138
第三节	《市场经济大辞典》电子版简介	145
附一	湖北省地方志全文计算机检索系统鉴定资料	153
附二	参考文献目录	160

第一章 电子出版物是一种新型出版物

电子出版物(Electronic Publication)是高科技的产物,是体现现代信息技术的新型出版物。更具体来说,它是计算机化的信息存贮与处理技术同文献出版与流通的结合。电子出版物的产生与发展是信息化社会的标志之一。它的存在与普及将加速信息的传递、推动社会科学文化的进步。

第一节 电子出版物的基本特征

所谓电子出版物,是继印刷型出版物之后出现的一种新的出版物类型。它将著作的文本转换成计算机可读形式的信息,记录在磁性载体(磁盘、磁带等)或光学载体(CD—ROM光盘)上,并在计算机软件的支持下,自动形成全文数据库,并提供从著作中的人名、地名、年代、关键词等各种知识项出发对文本进行单项检索或多项组配检索的功能,甚至著作中的任何一个字都可以由计算机检索出来。所检索到的文本段落,可根据用户的要求按句、段、节、章等进行输出,即在计算机屏幕上加以显示,或由打印机打印在纸上,或由计算机对其进行“套录”(downloading),将检索结果以计算机可读信息的形式记录到磁盘或光盘上。电子出版物的内容,除了与印刷版相当的著作文本之外,还包含由计算机自动编成的各种索引,以及计算机情报检索等软件。借助于这种索引和检索等软件,电子出版物允许按照用户的要求,对著作中

的各个知识项进行抽取、排序、重新组织，因而使原先散见于各个篇章段落中的知识项能够从逻辑上得到集中，提供从各种角度进行知识项聚类的能力。因此，电子出版物的文本成为置于读者控制之下的、具有全文检索功能的“活体”。它远比印刷版图书拥有更多的使用价值。

电子出版物的基本特征是：

1. 文献的文本是以计算机可读数据的形式加以表示的。这种机读形式的数据，是计算机可以识别、理解、处理的数据。归根结底，它们是用二进制码表示的。在遥远的古代，人们就发明了文字、符号（包括象形文字和拼音文字），它们可以刻在甲骨、石头上，书写在泥版、贝叶上，书写或印刷在纸张上，这些文字符号是供人们识别和理解的。因此这些文字和符号可以说是“人能够阅读的数据”。“白纸上印黑字”，长期以来构成了出版物的基本特征。但是，“人读数据”的形式并不是唯一的。录音带、电影片、录像带可以记录声频视频的模拟信息。在这种情况下，所用的载体不再是纸张，而是胶片和磁带，以及称为 CD 的光盘。对于这种形式的数据，人们不能直接“阅读”，而必须借助于录音机、播放机与机具才能使人们看到、听到并理解这些信息。在计算机信息处理的时代，一般使用二进制数据的表示形式，即“计算机可以阅读”形式。这种“机读数据”的出现，是二十世纪科学技术伟大成就之一。这是因为：第一，它的记录密度高；第二，它可以用计算机进行高速准确的处理；第三，它可以借助数据传输网络进行远距离的传输。所有的电子出版物必须以机读数据的形式出现。这是它首要的特征。

2. 电子出版物的信息载体是磁性或光学信息存贮介质。不同载体的物理存贮信息的密度是不同的。随着科学技术的

进步,载体的信息存贮密度越来越大:

泥版	$10-10^4$	mm^3 (每 1 字符存贮密度)
印刷图书	1	mm^3
磁性载体	10^{-4}	mm^3
光学载体	10^{-6}	mm^3

有人估计,若按每个字符以 10^{-8}mm^3 的密度来存贮信息,那末全球的全部信息可以存贮在一立方米的体积中,可以通过一条光导纤维用二至三周的时间完成传输。

信息载体的进步,直接促进了出版业的进步。例如在古代,纸张以其质轻价廉,并与印刷术相结合,创造了古代文明,奠定了出版业的基础。在今天,电子数据处理、光学存贮介质以及通讯系统的应用,使电子出版物得以产生。信息技术不仅改变了信息存贮的载体,而且也直接导致新型出版物的出现。目前,电子出版物以 CD-ROM (高密度只读光盘存贮器) 和多媒体光盘为主,同时也以磁盘或磁带为载体。电子出版物不再需要纸张,从而使文献的产生、阅读、传播进入了“无纸时代”。

3. 电子出版物中的信息是有结构的,是经过一定的格式化处理而组成的信息的集合。也就是说,电子出版物必须具备数据库的结构,否则,零乱的信息无法被计算机程序有效地加以处理。电子出版物的基本单元是记录。一篇文章、一条词条、一个自然段、一个内容相对完整的章或节均可作为一个记录。一部电子出版物就是由许多这样的记录组成的。而一条记录之中,又可进一步区分若干字段。一个字段中,又

可进一步区分若干个子字段或重复字段。这都需要根据文献文本的本身结构来规定。可以说，电子出版物的主体就是数据库，特别是全文数据库。

4. 电子出版物应有相应的检索软件和其它编辑软件的支持。计算机软件是“激活”文献文本数据的动力。在电子出版物中，无论是随机检索，还是文中顺序扫描，或者对检索到的文本进行排序、按一定格式进行显示、打印或套录，都丝毫离不开软件的支持。如果仅仅有机读数据的存在而无相应的软件，这些数据就是“死”数据，是不可能实现电子出版物的功能的。电子出版物的软件（包括实现一定数据库结构的建库软件、提供各种检索功能的检索软件，以及实现检索结果各种处理的编辑、打印、套录软件）往往根据电子出版物的文本结构特点和用户的需要而定。因此，各种电子出版物的软件不尽相同，有时甚至相差很大。许多电子出版物都有自己特定的软件（专用软件）。电子出版物软件的标准化和通用性是一个需要解决的问题。

5. 电子出版物的使用——即检索或浏览，必须借助于电子计算机进行。这里所说的电子计算机，包括大、中、小型计算机，微型计算机。在微型计算机中，还包括掌上型电脑等。电子出版物之所以离不开计算机，一是因为它的文本是机读数据，只有计算机才能识别；二是因为浏览、检索过程是必须借助软件进行的，而软件的运行是不能离开计算机的。这种情况，正如缩微胶片不能没有显微阅读器、录像带不能没有录像播放机一样。电子出版物对于计算机的要求是：一是应有较大的硬盘存贮容量。这是因为电子出版物的文献，少则数十万字，多则数百万乃至数千万字，加上相应的索引，本身容量较大；同时，检索软件本身也有数十K至一、二百K

字节之多；再者，在检索和对检索结果的处理中还可能产生一些中间工作文件，也需占有一定的磁盘容量。二是要求有较高分辨率的显示器和打印机。

6. 电子出版物必须以出版为目标。所谓“出版”，就是公诸于世，以文献的形式参与社会的知识交流。电子出版物毕竟是出版物，与印刷型出版物一样，有自己的著作权。在出版物的内容方面，作者与出版社要承担责任，接受社会的检验与评论。那些不是以出版为目标的电子文本，例如许多单位用微机或四通打字机形成的内部使用的电子文书文件，是不属于电子出版物之列的。即使是配上建库与检索软件，也只能称为电子文书档案全文系统。

7. 电子出版物的发行方式，既可以是作为封装型的磁盘或光盘出售给读者用户，也可以是作为联机数据库的形式，以联机情报检索服务的方式，以有偿服务为基础，向公众提供检索利用。与印刷型图书的发行相比较，作为联机提供的方式是一种新增加的发行方式。联机发行方式有许多优异之处：首先，是借助于通讯线路，能够使设有联机系统终端机地方的读者及时地利用电子出版物。当新的电子出版物或修改更新后的电子出版物装载到联机检索系统之日，就是各地读者可以存取之时。这样，地理上的障碍和发行上的延迟都不再存在。也就是说，读者可以获得最新的信息，加速了知识的传播。第二，由于联机检索系统可以同时装载许多电子出版物，例如美国的 DIALOG 联机情报检索系统装载有数以百计的电子出版物，从《圣经》到电子版的报纸、手册、年鉴、百科全书等，这样读者在使用电子出版物时有广泛的选择余地。第三，通过联机情报检索系统检索利用电子出版物，读者只为实际的检索时间与检索量付费，而不象购买电子出版物那

样为整个电子出版物付费。显然，这种方式的发行对某些读者来说是经济合算的。当然，联机检索方式的费用，包含下列几方面的费用：通讯线路费；作为联机数据库的电子出版物利用费；计算机处理的机时费；检索命中结果的打印费；终端租用费等等。从这个角度来说，费用还相当昂贵。对于频繁使用电子出版物的读者来说，购买封装型的电子出版物却是经济合算的。但是，电子出版物增加了一种发行方式，对读者来说，就有更大的灵活性与选择性。电子出版物开拓了一个新的文献发行渠道。

第二节 电子出版物的范围与种类

根据前面所说的电子出版物的基本特征，大致可以确定电子出版物的范围。（即以出版为目标、通过市场供应方式发行的、以磁盘、磁带、光盘为载体、以机读数据形式表示的、体现数据库结构并辅有计算机检索软件等信息处理手段的、通过计算机加以利用的出版物。符合这个范围的就是完全意义上的电子出版物。）

但是也有一些例外：

1. 书目数据库。例如美国《化学文摘》、《科学文摘》等大量的文摘、索引、书目等数据库，是最早出现的机读数据的集合。它其中包含了二次文献的全部信息，即题名、作者姓名、作者单位、发表出处、页数、文种、主题标引词、分类号、文摘等等项目。每篇文献的描述信息为一记录。各记录按顺序号（基本上是文献加工的时间）排列。这些文献磁带从 60 年代以来，一直在发行，是数据库工业的主要产品之一。这种书目数据库只包含数据，而无计算机检索软件。图

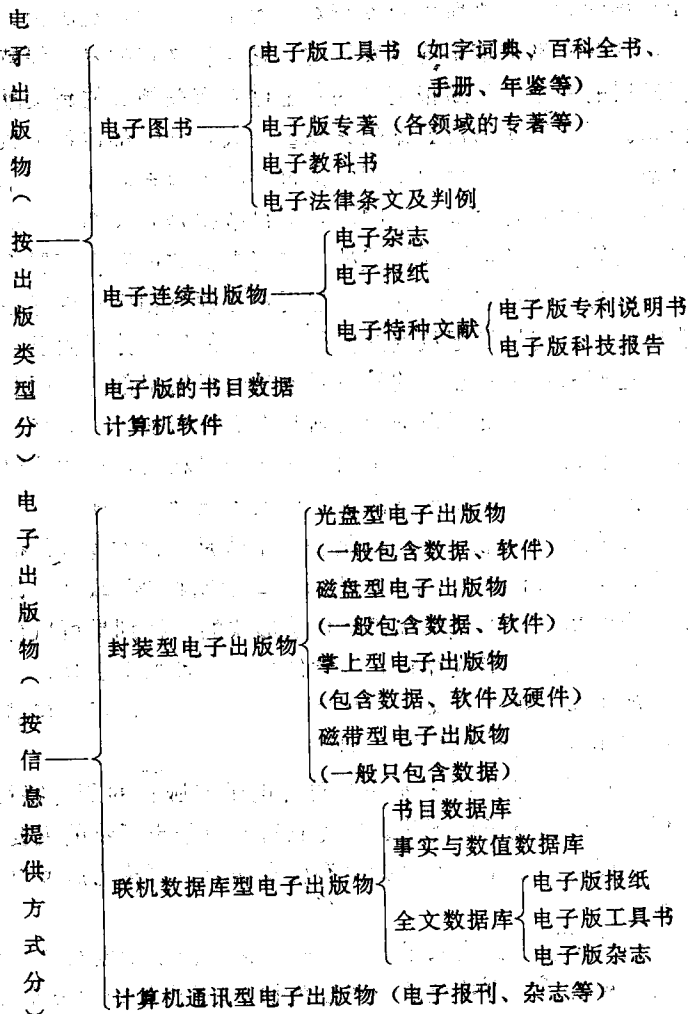
书情报部门购买或租用了这些文献书目磁带之后，需自配建库与检索软件。之所以只发行数据而不带软件，是因为：第一，这些文献磁带是连续发行的；它与新文献的发表保持平行发展的关系，不断报道与揭示全世界有关领域的新文献的信息。因而这些书目数据的累积量很大，一般都达到了数百万条记录；第二，用户采用的情报存贮与检索软件互相不尽相同，也没有必要统一。同时，书目数据是有长期使用价值的，而软件是经常更新换代的。这些书目数据库也应属于电子出版物。尽管它不带软件，但用户必须有相应软件存在才能使用。它们只是把数据的发行与软件的发行分开而已；第三，这些书目文献磁带的发行对象是图书情报部门，而非最终情报用户。图书情报部门必须对它们进行加工，即将它们装入计算机，建立数据库索引，才能加以利用。正如从粮店购买的大米面粉，需经过锅灶的加工才能食用一样。

2. 软件出版物。软件出版物的主体是软件本身（包括源程序和可执行程序），同时也附有一些说明材料。软件是计算机指令的集合，本身就应该是机读形式的。特别是可执行程序，只能以机读形式存在。但软件只是一种加工工具，它本身一般不包含加工对象——数据。在这一点上，正好与书目数据库相反。它们都各取一端。

尽管书目数据库未能完全符合电子出版物的所有特征，但仍然可归属电子出版物。因为它符合电子出版物的主要特征。机读书目数据库不直接具有检索功能，可称为原型电子出版物。

软件出版物是为满足人们的软件需要而出版发行的，可列为一种有其自身特色的电子出版物。电子出版物的分类，既可按出版物的类型来划分，也可按信息服务方式来划分。现

列表如下:



所谓封装型电子出版物，是相对联机数据库型电子出版物而言的。它将数据及软件作为一个出版单位，予以“包装”而发行到用户手里，这种出版物大多数是电子图书，它是相对独立的整体。数据库的内容相对稳定。

联机数据库型的电子出版物装载在联机情报检索服务系统中，用户只能通过联机检索方式进行利用。这种电子出版物出现最早的是书目数据库，例如各种科技文献的文摘、索引、目录。其后，一些工具书的机读版也出现在联机系统中。例如百科全书、手册等。联机全文数据库出现较晚，七十年代末才崭露头角，但现在发展很快。以电子报纸为例，美国的全国与地方一级的数十种报纸都作为 DIALOG 联机系统的全文数据库而提供检索服务。作为联机数据库的电子出版物，一般来说属“开放型”出版物，需要对数据库的内容经常予以更新，数据库内容的动态性较强。报刊是连续出版的，必须连续更新其数据库。手册、大全之类的工具书，也是连续出新版的。至于象经济商情方面的联机数据库，其更新周期更短，最短的每九十秒钟更新一次。联机数据库型的电子出版物便于读者获得最新的信息与资料。

计算机通讯型电子出版物，是一种在计算机网络条件下，通过各个终端用户发送和接受的电子出版物，类似于分布式数据库。这种出版方式主要有电子邮政、电子会议系统、电子杂志等。这一类电子出版物出版的典型方式是作者在自己的终端上创作、编辑论文，其它读者通过网络存取并利用它；读者可以对某一篇论文发表意见，或者与作者对话。所有这些论文包括有关它的意见、对话均可被网络中的其它用户阅读利用。读者和作者可以通过网络实现直接的联系。每一个终端用户可以是作者、读者，也可以是自己作品的出版者。在