

轻印刷的中间产品可开发成全 文数据库并可作为电子出版物发行

陈 光 祚

当前,我国出版界采用激光照排和轻印刷的技术日益普遍。这种生产加工的程序是:把作者的书稿录入计算机,经校对改错后,采用电子排版软件(如科印、华光、方正等软件)排定书稿的字体、字型和版式,然后采用激光输出、制版、印刷、装订成册,向社会发行。这种轻印刷流程的最终目标,目前仍是生产印刷版的图书。至于电子文本录入、电子排版,仅仅被看作是加工的手段。

书稿录入计算机,不管是否采用电子排版软件处理,都是计算机可以阅读的电子文本。这种电子文本以二进制代码形式出现,只能存贮在计算机的硬盘或软盘上,或是存贮在CD-ROM激光光盘上。它们只能借助计算机才能显示和阅读。

目前出版社对这种电子文本,即轻印刷生产过程中的中间产品的认识不足。在出版印刷版图书之后,把它们仅仅作为纸型保存,以备在书籍再版重印时使用。有的出版社甚至认为保存这种中间产品需要占用不少软磁盘,经济上不合算,因此不愿这样做。

其实,这种电子文本是一种重要的潜在资源,只要对它进行进一步的加工,即可开发成全文数据库,从而作为电子版的图书——完整意义上的电子出版物而向社会发行。

一、机读版图书是一种新型的出版物类型

出版物就其载体和使用方法分,可分为印刷型、缩微型、声相型及计算机可读型。前三种形式的出版物,我国的出版界与读者较为熟悉。而对于计算机阅读型的出版物(或称机读版图书)的了解相对较少。

在国外,60年代初,美国《化学文摘》社开始采用电子计算机技术来改进印刷版《化学文摘》的生产速度和质量。其办法是,抛弃了过去铅排的作法,而将书稿录入计算机,成为机读文本。并采用一定的软件进行索引的自动编制。然后,将机读文本排定版式,进行激光输出、制版、印刷、装订成册。通过这种技术改造,《化学文摘》的出版速度大大提高。出版质量方面,由于计算机能对电子文本进行自动检错、自动编制索引等,也有很大提高。这种技术改造的目的,仍然是生产印刷版书刊。在这一点上,我国出版界当前采用激光照排与轻印刷技术的状况,正经历美国60年代初的这种过程。

但是,在美国,人们很快地发现,作为中间产品的电子文本可以在计算机上直接使用——对出版物中的各种信息进行自动检索、自动选取其中的部分文本进行显示打印,或对其进行重新排序处理。这样,磁带版的《化学文摘》就出现了。这是一种新型的电子出版物。从此,《化学文摘》通过磁带,装入情报系统的计算机成为《化学文摘》数据库,实现了计算机化的情报检索服务。

以世界上规模最大计算机化的情报服务系统——美国DIALCG系统为例,以电子文本为基础的数据库达300多个,而包含1907年以来世界化学化工文献线索一千多万条记录的《化学文摘》数据库仅仅是其中一个数据库。尤其值得重视的是,在DIALCG系统中,象百科全书、报纸、杂志等全文数据库近年来增长十分迅速,已达200多个。如,其中有《美国学术百科全书》、《芝加哥论坛报》、《银行家杂志》等等。读者可从中查找和获得这种全文数据库中的任何篇章、片断的文本和信息。

国外的出版家现在对同一著作,既发行印刷版,又发行以磁带或磁盘形式出现的电子版。目前的发展趋势是,印刷版书刊的出版日渐减少,而电子版的书刊日趋增多。越来越多的出版物只有电子版而无相应的印

刷版。这是与电子版的优点分不开的。电子版的优点在于:

首先,电子版书刊的存储密度大。一张5英寸直径的普通软磁盘可记录18万个汉字,一张高密度磁盘可记录50多万汉字。至于一张直径为12厘米的CD-ROM光盘可存贮十部托尔斯泰《战争与和平》的全文。

其次,电子版可以用计算机高速查找情报。对于全文中的任何一个人名、地名、关键词都可以在数秒内查到查全,并可按用户的要求,把有关段落章节进行显示或打印,也可按用户的要求对它进行重新排序和套录。

电子版可以通过通讯线路进行远距离的传输。

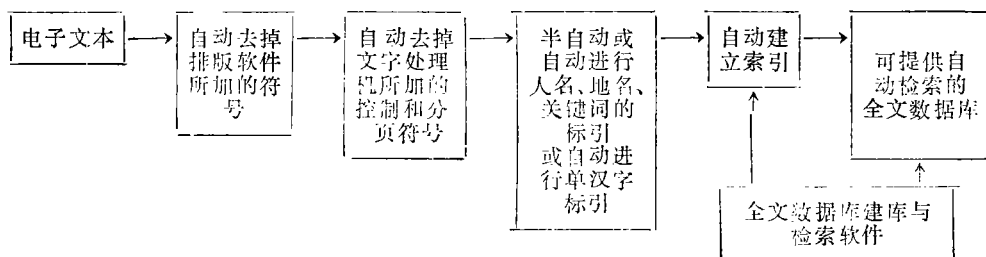
最后,电子版书刊的内容文字可以允许经常修改更新,这要比印刷版几年出一次新版快得多。也就是说,用户借助电子版可以获得最新的情报内容。

因此,电子版书刊有强大的生命力,是信息化时代的产物和标志之一。

二、全文数据库的加工

要使电子文本成为全文数据库,即使它具有自动检索的功能,只需对它进行必要的标引加工和配置相应的建库与检索软件。

要将目前仅作为纸型保存的电子文本加工成全文数据库,其必要的处理过程是:



从这个加工流程图中可以看到,这种加工过程都是自动或半自动的。一部100万字的电子文本书稿,只需要一个人在微型计算机上加工10天左右即可。其中,去掉排版软件所加的符号,约需半天时间。去掉文字处理机所加的控制和分页符号,只需几分钟即可。对书稿中人名、地名、关键词的半自动或自动标引所需时间最多,约4—7天。至于建立索引,形成全文数据库,只需一至二个小时。如果采用单汉字标引的方式,加工时间只需两天。因此,从加工时间来看,所投入的代价很低。

从加工所耗物质的代价来说,主要的消耗是需购买软磁盘。100万字的书稿连同建库形成的索引,只需三片高密度软磁盘。目前每片空白软磁盘价为13元,共需不到40元。

消耗较大的是微型计算机的机时,但不少出版社已购置了微型计算机作为书稿录入之用,可资使用。

因此,由于轻印刷的中间产品——电子文本已经是现成的,对它们进行加工,使之形成全文数据库的代价应该说是极低的。对这种中间产品的开发,是现有资源的再利用。

用于全文数据库建库与检索的软件,我们已研制成功,1990年10月18日通过了国家教委的鉴定。我们已利用了这个软件,试验建成了《湖北省地方志大事记》、《中国人民解放军六十年大事记》、《山东省基本情况》等的全文数据库,并与武汉大学出版社合作出版《国共两党关系通史》的全文检索电子版。事实证明,对上述全文数据库中的任何人名、地名、关键词、年代都可进行检索,甚至对电子全文中的任何一个字也可进行检索。我们愿意将这个软件贡献给出版界,为我国出版业的现代化作贡献。

至于目前发行电子出版物的市场销售量问题,也应该有所估计。

首先应看到它的社会效益。电子出版物的发行,将是我国出版界走向现代化的一个标志,也能使出版物更好地为四化建设服务,适应信息时代读者新的要求。

其次,从经济效益来看,那些资料性丰富、具有较高科学与实用价值著作的全文数据库,其发行的市场是存在的。

目前我国高等学校、科研部门、图书情报部门、报刊编辑部门、行政部门所拥有的微型计算机已有一定

(下转第123页)

劳模、革命烈士的名单和生平事迹介绍,读后令人感叹不已,精神振奋。它为人们树立了学习的楷模,促使人们振作起来,奋发向上,为建设武汉,建设具有中国特色的社会主义而共同努力。在这个意义上,《武汉年鉴》又是一块进行爱国主义、社会主义思想教育的阵地。至于那些获得各级各类奖励的产品、厂家介绍,无疑起到了促进企业狠抓产品质量、提高效益和指导消费的作用。

然而,《武汉年鉴》并不满足于现状。每年新版,都要根据武汉市的具体情况和社会要求等因素,对某些栏目、分目或子目进行适当的增删、调整,以使其能更科学更全面地反映武汉市上一年度的最新情况,给人们提供最新信息,更适合广大读者的需求,并进一步增强其使用、保存价值和服务于社会的功能。

1991年版的《武汉年鉴》,首先登载了赵宝江市长的专文:《团结奋进,为实现十年规划和“八五”计划而奋斗》。该文既简要回顾了“七五”的成就,又着重对“八五”计划和90年代武汉发展前景提出宏观展望,是1991年版年鉴的纲。在“武汉综述”栏,还专题介绍了武汉从1949年以来产业结构变化、形成的过程和1990年的现状。还增加了几个大方面的内容:“七五”期间发展概况、计划生育工作、第四次全国人口普查资料等。值得特别指出的是,1991年版的《武汉年鉴》,新辟了《社会生活》栏,增加了“婚姻·家庭”、“社会保险”等内容,有利于人们对社会、经济发展作更加深入全面的观察。

由于《武汉年鉴》能根据每年具体情况进行调整和改进,因而,它始终紧紧与时代同步,具有极强的时效性和时代感,它使人们从中不仅能领略到浩如烟海的知识资料和信息,而且能从中获得珍贵的精神享受,看到武汉的时代风貌,感受到时代跳动的脉搏。

三、体例科学 编排合理

《武汉年鉴》内容涵括面极广,如果稍有不慎,在全书编排体例上很可能出现逻辑混乱,脉络不清,主次不分,材料堆砌的情况。然而,这些可能出现的问题却并没有出现。年鉴编纂者善于组织材料,通过分析,采用那些能反映普遍性、规律性的材料,剔除那些可有可无或非本质的材料。记述体系以栏目、分目、条目顺序编排,以条目为基本记叙单位。这样,使得各栏目内容归属明确,避免了内容的疏漏、交叉和重复。这种有纲有目、纲举目张式的编排体例,用于这种百科性大型工具书,是十分得体的。

值得一提的是,《武汉年鉴》装帧、印刷质量尤为精美,色彩艳丽,图文并茂,纸张光可鉴人。阅后令人神清气爽。这大概是它倍受读者青睐的又一个原因。

《武汉年鉴》至1991年版止,共出了七部。随着时间的推移,卷数的增多,它的独特的参考使用和保存价值将愈来愈重要。它将永远是全面、系统、真实、快捷而又清晰地反映武汉风貌的一面不可缺少的镜子。

(上接第121页)

的数量,而且正在不断增多。这些微型计算机可拿出一部分机时用于全文数据库的检索。他们使用全文数据库,要比使用相应印刷版书刊在检索速度和完备性方面优异得多。对于所查到的篇章片断,可由计算机打印,这也要比人工抄录方便准确得多。

三、全文数据库的检索功能

全文数据库的检索功能大致如下:

1. 按人名、地名、年代、关键词进行检索。例如,在《国共两党关系通史》全文库中,可查找到所有出现“列宁”、“武昌”、或“北伐”的所有段落或句子。
2. 可按人名、地名、年代、关键词之间的组配进行高专指度的检索。例如可查找“英国的在武汉的租界”,“周恩来与黄埔军校的关系”,“黄兴在1911年的活动”等。
3. 可在上述检索到的段落中,进行任何一个字的查找,并可显示或打印包含该字的所有段落与句子的文字。这可用于查找某人物的某句名言或诗词。
4. 可进行法定数检索。即用户事先确定检索的资料量,检索系统可引导用户逐步逼近其检查目标。
5. 具有人名、地名、关键词等的聚类功能。例如,与“西山会议派”有关的人物名单,与“收复租界”有关的地名,与“李之龙”有关的事件等。

总之,由于形成了全文数据库,著作全文中的各个知识项目和文字都可进行单项或多项组配检索,从而使文献全文成为一个“活体”,大大有利于资料信息的获取。