

●杨道良 沈红芳

从博克电子图书到电子图书馆^{*}

ABSTRACT The "BOOK Electronic Book" Compilation Environment (WDBOOK)" developed by the wuhan University has put forward the EBML of Chinese electronic book which conforms to the special characteristics of Chinese information and provides the platform of making automatic electronic book and the electronic book reading platform, the latter brings about the bottom support environment of electronic library. 4 illus 4 refs

KEY WORDS Electronic books production Electronic libraries Network work

CLASS NUMBER G258.94

1996年2月,武汉大学出版社研制的“博克电子图书编著环境(WDBOOK)”通过了国家新闻出版署组织的专家鉴定。“博克电子图书编著环境”是一套通用的中文电子图书制作与阅读工具系统。它能将轻印刷的副产品——排版文件直接加工成电子图书,较好地解决了电子图书置标规范、机器辅助标引、汉字库扩充、多机制快速检索、出版发行控制等问题。该成果提出了一套语法严谨、描述功能完备、符合中文信息特性的中文电子图书置标语言规范(EBML),提供了自动化的电子图书制作平台、高效率的电子图书阅读与研究平台,彻底改观了我国手工作坊式的电子图书生产方式,是电子图书自动化生产的新突破。与会专家一致认为“该项成果属国内领先水平”。

“博克电子图书编著环境”的问世,使得几代人梦想的电子图书馆接近变成现实。首先,它从源头解决了文献的电子化问题。一旦

出版社采用了“博克电子图书编著环境”,该出版社在出版印刷版图书的同时,其中间副产品——排版电子文本,可以很容易地加工成电子图书,这样图书的印刷版和电子版可同时发行。甚至有些发行量较小的图书可以只发行电子图书,以降低出版成本,增加作者出书的机会。同样,杂志社、报社也可以用“博克电子图书编著环境”来发行电子杂志、电子报纸,企事业单位和家庭亦可将办公文件、有保存价值的资料、个人档案等制作成图文并茂的博克电子文档。这些产品(电子图书、电子杂志、电子报纸、电子文档)都将成为电子图书馆的收藏对象。其次,博克电子图书阅读平台支持网络阅读和选书功能,已经实现了电子图书馆的底层支撑环境。

1 博克电子图书的制作

1.1 博克电子图书的作用流程

* 国家自然科学基金资助项目研究论文之一

制作人员首先对源文本(华光、方正排版文本、WPS 文本或 OCR 扫描识别的一般文本)按标引语言的语法规则作初始标引(如标引出卷、章节、关键词等),然后利用博克电子图书编著环境的机辅标引工具对其进行完全标引。如果书中有插图、缺字(国标字库中没有的汉字),还要用扫描仪或绘图软件制作好所需的图片、用补字工具造好所缺的汉字。启动博克电子图书编著环境的制作生成工具进行制作合成,便生成了格式化的电子图书文件。这时可以对电子图书文件进行试阅读,如果发现错误或版式不美观等问题,可根据需要修改相应的标引文本,并重新制作,直到满意为止。然后就可交给出版发行管理工序进行出版发行,打包、复制并装帧成最终产品——电子图书(见图 1)。

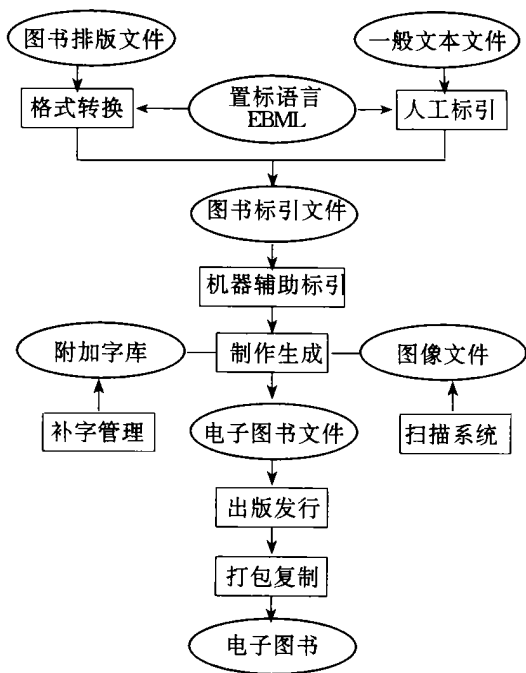


图 1 博克电子图书制作流程

从图 1 可以看出,博克电子图书的生产效率是相当高的,除了 3 个环节——文本标引、造字、图像处理需要人工参与外,其余的工作都是由软件自动完成的,就是工作量最

大的文本标引,在博克机器辅助标引工具的协助下亦变得非常轻松。制作人员只需将责任编辑在责编文稿时挑选出的关键词列表放在标引文件的开头并标引一遍,然后运行机器辅助标引工具,整本书的标引工作就完成了。如果标引时不够细心,挑选出的关键词较少,问题也不大。因为博克制作系统自动抽取了每个汉字的位置信息,形成了单汉字索引,即图书中的每个字词都是可以检索的,标引关键词的目的只是为给读者阅读检索时提供主题内容导向和便于进行统计分析。

用“博克电子图书编著环境”制作电子图书,一个熟练的制作人员一天可以制作一部 50 万字的电子图书。

1.2 博克电子图书编著环境的特点

(1) 提出了一套规范化的中文电子图书置标语言(EBML),不仅反映了印刷版图书所包含的全部要素,而且还反映了出版管理要求、阅读研究要求等。

(2) 制作电子图书的自动化程度高。博克电子图书的生产是以软件自动化处理为主的,与手工作坊式的电子图书生产相比,出版周期缩短了 95%,生产成本降低了 80%。博克电子图书编著环境把电子图书的生产速度和效率提到了惊人的程度。

(3) 制作出的电子图书质量高。博克电子图书具有超级的检索功能,还提供各种信息计量功能(如用字、用词频率统计、聚类分析等)及插书签、做笔记、摘录等读书研究所需的辅助功能,大大地提高了读者的读书深度和研究效率。

2 博克电子图书的阅读方式

2.1 博克电子图书阅读平台的工作机理

读者阅读博克电子图书是通过博克电子图书阅读平台来进行的。这个阅读平台是一个用户友好、操作简单的电子图书阅读工具,整个界面和操作都具有 Windows 风格。

读者阅读电子图书时,首先要进行注册、选书,然后阅读。阅读平台中有两个文件专门支持选书功能:书目文件和CFG文件。书目文件中记录了书库中所有电子图书的简单书目信息(如书名、作者、出版社、图书文件所存放的位置——文件路径或网络节点、图书文件的名称等)。CFG文件是书目文件的子集,相当于用户书目,即书目文件中用户感兴趣的电子图书书目。书目文件是读者购回电子图书后进行安装时,安装程序自动抽取相应参数形成的,而且是不断追加的。

读者注册完毕后,阅读平台便列出CFG文件中的书目列表,读者可以从中选取一本书进行阅读,也可以进入阅读平台的简易书目查询功能,从书目文件中检索出自己所需图书,组成新的CFG文件,然后再任意选择阅读。读者选中某一本书后,阅读平台便根据CFG文件中相应的位置指示去访问该书的数据文件,供读者阅读研究。其工作机理见图2。其中,博克电子图书阅读平台就好比书案,书目文件和CFG文件就好比图书目录卡片盒,而博克电子图书库就好比藏书的书柜,它们组合在一起就构成一个电子书房。

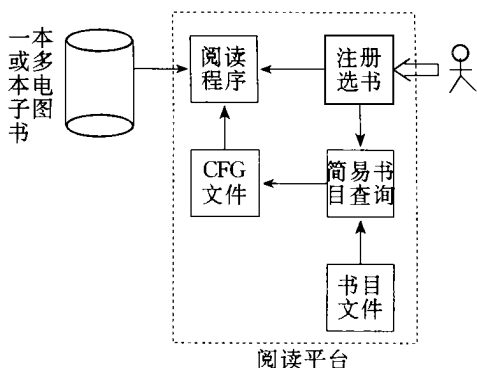


图2 博克电子图书阅读平台的工作原理

2.2 博克电子图书阅读平台的功能

博克电子图书阅读平台所提供的功能非常丰富,概括起来可分为:浏览、检索、研读与信息计量。

(1) 浏览。博克电子图书的读书概念与传统的读书概念是一致的,符合传统的读书习惯。读者可以逐页阅读,也可以任意翻到某一章节阅读,或直接翻到某一页阅读。

(2) 检索。检索功能是电子图书区别于传统印刷图书的最大特点。博克电子图书的检索功能是由一个独立的检索进程来完成的。阅读平台在获得读者的提问式后,通过调用检索进程来得到检索结果。检索进程支持关键词和非关键词(自由词)的检索,支持后控词表(同义词表)检索,可以设置不同的检索匹配单元(段、句、小句),还可以进行各种复杂的组配检索、位置检索、截词检索、模糊检索等,检索速度平均为秒级。其他程序(如书目查询)亦可调用检索进程来完成某些深入查询功能。

(3) 研读。博克电子图书的研读功能给读者提供了一个电子书案,上面有“剪刀浆糊”(即摘录功能)、“笔记本”(即笔记功能)、“书签”(即书签功能)供读者任意选用。在这样一个良好的读书研究环境中,读者的写作研究效率将是空前的。

(4) 信息计量。这是博克电子图书为文献计量学所作的一大贡献。在博克电子图书中,不仅可以进行用字、用词频率的统计,还可以按条件进行聚类分析(如分析某一个人物与其他人物的亲疏关系等)。

3 从博克电子图书到电子图书馆

以博克电子图书来构筑电子图书馆可根据规模大小分为3个档次。

3.1 单机环境下的电子图书室

单机环境下的电子图书室其实就是一架博克电子图书阅读平台。换言之,博克电子图书阅读平台本身已经具备了电子图书室的功能(见图2)。因此,单机环境下的电子图书室的硬件配置要求低,只需一台486计算机(带硬盘、CD-ROM驱动器)即可,维护简单,投

资少, 很适合小型企事业单位和家庭。

博克电子图书的发行多是以 CD-ROM 光盘形式发行的, 一张 CD-ROM 光盘可以存贮《红楼梦》大小的电子图书 200 本, 售价不过百元, 如果单位或家庭收藏上千张博克电子图书光盘, 一个颇具规模的电子图书室或电子书房便建立起来了。当然也有一些用磁盘形式单本发行的电子图书, 这样的书可以安装到计算机的硬盘中, 与 CD-ROM 互为补充。

在单机环境下, 由于图书数量有限, 博克电子图书阅读平台本身所具备的书目功能已经基本够用了。这时, 书目文件可以存放所有光盘及硬盘电子图书的书目, 这样便于读者进行统一的书目查询。书目文件中的“图书文件位置”项还将记录光盘编号, 当读者检索出所需图书进行阅读时, 如果该书不在当前光盘或硬盘上, 阅读平台会自动提示更换相应编号的光盘。

3.2 局域网络环境下的中小型电子图书馆

这种中小型电子图书馆可以建设在原有的图书馆自动化网络基础上。其中, 网络服务器的超大容量硬盘或光盘组驱动器将成为电子图书库, 读者将通过工作站上的阅读平台来阅读服务器上的电子图书。在这种情况下, 博克电子图书阅读平台本身的简易书目功能已经不够用了, 必须建立配套的书目数据库管理系统来管理整个电子图书馆的书目数据库。可以利用现有的书目检索系统, 只需增加与博克电子图书阅读平台的接口即可。书目检索系统还可以利用博克电子图书的自动编目功能(即自动抽取电子图书文件中的图书在版编目数据及电子图书文件所在的位置信息功能), 自动形成书目数据库。书目检索系统将支持多种检索途径和复杂的组配检索及集合运算。读者可以通过它检索出满意的书目集合, 然后进行阅读; 也可以将书目集合保存下来(如打印或套录), 供以后阅读或使用。

读者通过工作站进行阅读时, 可以先进

入书目检索系统, 从服务器上的书目数据库中检索出与自己的研究主题相关的若干书目集合, 并指定一个最满意的集合来形成工作站阅读平台的简易书目文件及 CFG 文件。这时, 读者既可以直接从 CFG 文件中选取一本图书阅读, 也可以进入阅读平台的简易书目查询功能, 进一步缩小书目范围, 然后再从较少的书目中选取一本图书阅读。其运行模式如图 3。

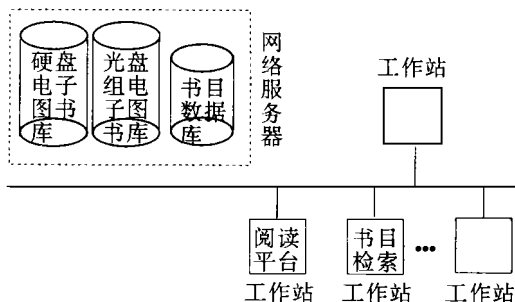


图 3 局域网络环境下的中小型电子图书馆

图 3 中的光盘组驱动器是一新型的光盘存储设备, 可以容纳数张 CD-ROM 光盘同时进行存取, 盘片还可以随时更换。每个工作站上可以只安装阅读平台或只安装书目检索系统, 也可以两者都安装, 视工作站的用途而定。

工作站还可以延伸到读者家里, 使读者足不出户即可阅览图书馆所收藏的所有电子文献。

3.3 分布式网络环境下的大型电子图书馆

这种大型电子图书馆将是以信息高速公路为依托, 以全国范围或地区范围的电子信息资源共享为前提的极其复杂的分布式系统。在这种大型电子图书馆中, 电子图书库和书目数据库的物理存储将分散在网络的各个不同节点, 但对读者而言, 并不需要知道它们存放的位置。因为在高速通讯网络的支持下, 读者阅读千里之外的电子图书与阅读本地的电子图书没有什么两样。

这种大型电子图书馆的书目检索系统功能将更加强大, 能够支持分布式的书目数据

库查询, 而且还要求能根据图书的相关程度对命中书目进行排序。但当读者通过书目检索系统查询到一组相关书目后, 这些图书的重要性读者还是无法区分。因此, 要求书目检索系统能帮助读者对命中书目逐本进入图书的内部进行查询, 区分出图书的相关程度。阅读平台的检索进程可以完成这个任务。当读者检索出一组书目集合后, 根据读者的要求,

书目检索系统可以对读者指定的一个书目集合进行相关程度排序。这时, 书目检索系统将启动阅读平台的检索进程深入该集合中的每本图书内部, 用读者给出的主题词进行检索, 并根据命中结果数的多寡来重新排列书目集合中的书目顺序, 排在最前面的图书将是最相关的。这种大型电子图书馆运行模式如图 4 所示。

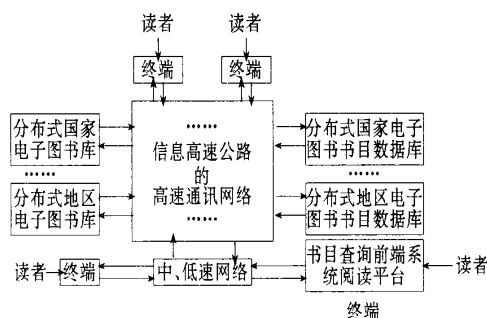


图 4 分布式网络环境下的大型电子图书馆

从图 4 可以看出, 这种大型电子图书馆是由若干中小型电子图书馆在信息高速公路的连接下组成的。因此, 读者阅读的过程与在中小型电子图书馆中是一样的。读者可以将自己的书目查询限定在本地节点或本地区节点, 也可以查询整个分布式网络的所有节点。查询的最终结果将传送到读者的终端, 并可形成阅读平台的简易书目文件和 CFG 配置文件, 以便读者阅读。

用户家庭的电脑亦可通过电话线路或网络通讯线路连入电子图书馆网络, 在安装了书目查询前端系统和阅读平台之后, 各种查询或阅览在家中随时可以进行。

电子图书馆具有无比的优越性, 但要建设它也是一项艰巨的工程, 需要社会各界的共同努力, 特别是新闻出版界和图书信息界的通力合作方能完成; 作为信息源头的新闻出版界需要采用统一标准的电子图书制作工具制作出更多的统一标准的电子文献, 为电子图书馆和信息高速公路提供充足的“货源”; 作为“信息用户界面”的图书信息界则要

在传统图书馆向现代电子图书馆的改造和用户电子信息服务方面发挥主导作用。其巨大的建设资金, 除了政府投入外, 还应该大力提倡“社会办图书馆”, 鼓励全社会对图书馆的投资。

参考文献

- 1 博克研制小组“博克电子图书编著环境(WD-book v1.0)”鉴定会材料, 1996
- 2 万齐鸣 信息高速公路与未来图书馆 中国图书馆学报, 1996(4)
- 3 陈光祚 电子出版物及其制作技术 武汉: 武汉大学出版社, 1994
- 4 Paut Nieuwenhusen Criteria for the evaluation of text storage and retrieval software The Electronic library, 1988(6)

杨道良 沈红芳 武汉大学图书情报学院 94 级硕士研究生。通讯地址: 武汉大学图书情报学院。邮编 430072。

(来稿时间: 1996 10 3。编发者: 翟凤岐)