

# E-Book 发展研究

E-Book Development Research

吴於茜

(北京大学国家文字信息处理实验室 北京 100871)

**Abstract** E-Book is called electronic book, it is a new thing of publish in these two years, it's also the result of the Internet technology. This article introduces the history and status of E-Book, the Open e-Book Organization and Open eBook Publication Structure V1.0, an E-Book document format based on XML defined by OEB.

**Keywords** E-Book, eBook, Electronic book, XML, OEB, Electronic publish

## 一、E-Book 介绍

E-Book 可译为电子书,是英文 electronic-book 的缩写,也可写为 eBook,顾名思义是指以电子文档方式发行、传播、阅读的电子书籍,它的传播媒介为 Internet 网络,是信息时代技术发展的必然产物。它的出现,极大地推动了网络出版的发展,对传统的纸介质出版及印刷业带来了一定的冲击。

从本世纪八十年代初,人们就开始关注 E-Book 的发展。那么, E-Book 与电子文档有什么分别呢?目前电子文档已经非常流行,人们也编写和制定了一些有关书写电子文档的软件和文档格式,但电子文档仍然没有取代纸介质的印刷品,“无纸化办公”成了一句空话。事实证明,随着计算机应用的普及,纸张的消耗不但没有降低,反而急剧增多,这与我们目前所使用的计算机显示器性能与电子文档松散的排版方式有着直接的联系。每个使用计算机的人都有过这样的体验,连续地长时间阅读电子文档会使我们的眼睛感到非常疲劳,这是因为显示器文字与背景之间的对比度与纸张印在纸上的文字的对比度相差很多,文字在显示器上的显示效果比较模糊,而印在纸面上的字就好很多,看起来柔和而且清晰。另一方面,现在的电子文档的浏览器在排版方面较传统印刷有很大区别,没有很严格的排版规则,没有边空设置,没有页面间隔,随意的断行和不规范的行距字距等,都给阅读者带来很大的阅读障碍。在这种条件下读者无法进入一种沉醉状态,专心致志于文档本身的内容。由于上述以及其它种种原因,抛弃纸介质而直接在显示屏上读书还只能是个理想。而 E-Book 拉近了现实与理想的距离,它提供了一种

更为接近人们阅读习惯的阅读方式,使得人们在阅读 E-Book 书籍时,能够不受外在条件的干扰而能专注于吸收书的内容。它不仅包含了电子文档的记录形式,而且还包括了文档的显示方式、传播形式等。从某种意义上讲,电子文档也属于 E-Book 的一部分,但并不等同于 E-Book, E-Book 是一个更为广义的概念。

E-Book 首先包含了文章或书籍的数据文档,以一定的数据格式存储文章或书的内容,另外,与数据文档密不可分的是它为读者提供了特殊的阅读手段,目前,已经有多种供读者阅读 E-Book 的硬件设备——阅读机已经问世,正在市场上销售,另有一些硬件厂商也在开发新产品。E-Book 阅读机是一个便携式机器,形状像一个笔记本,采用高清晰度的液晶显示屏,可以存放 5000 页或更多的内容,通过与 PC 机相连,可以下载网络上各种文档和书籍。E-Book 的普及将大大地降低由于印刷、复制文件、书刊而带来的纸张消耗,还方便书籍的携带和传播。

## 二、OEB 组织与 E-Book 发展现状

E-Book 近两年来发展很快,受到了包括作家、出版商、书刊销售商、软件开发公司、硬件制造厂商的普遍关注,并成立了一个由四十多个组织组成的 Open e-Book Authoring Group(以下简称为 OEB 组织),参加的组织中有 E-Book 先锋 Nuvomedia Inc.、SoftBook Inc.、GlassBook Inc.、Libius Inc.、EveryBook Inc.,以及 Microsoft、Adobe、Motorola、Nokia 等一大批软件及硬件厂商,还包括了一些著名的印刷出版商。经过 OEB 组织的努力,于 1999 年 9 月 21 日完成了 OEB1.0 标准的制定,从此 E-Book 有了统一的文档数

吴於茜 讲师,硕士,长期从事桌面电子出版系统的软件开发及研究工作,1999 年开始进行 XML 及 E-Book 的研究。

据格式标准。

制定 OEB1.0 标准的目的是为了以防在 VCR 市场上曾经出现的 Beta 对 VHS 的格式大战的重演,避免由于厂商的对立给消费者带来不利影响,进而阻碍 E-Book 市场的建立及发展。OEB1.0 标准的出台,从根本上统一了 E-Book 的文档数据存储方式,保证了消费者购买任何一个品牌的阅读机都可正确阅读 E-Book 文档,使得消费者不用承担由于购买阅读机而带来的格式兼容方面的风险,奠定了建立 E-Book 市场的坚实基础。

目前市场上已经推出了几种阅读机,一种是 SoftBook 公司推出的 SoftBook,另一种是 NuvoMedia 推出的 RocketBook。SoftBook 阅读机可以将 50,000 页的信息容纳在它宽 8.5 英寸、高 11 英寸、厚 1 英寸、重 2.9 磅的机器中。合作开发者可使用 SoftBook 的印刷工具软件制作 E-Book 文档,然后将其下载和显示到阅读机中。SoftBook 还提供了 SoftBook Express,它作为一个图标显示在 Internet 站点,供读者进行文档下载,SoftBook Network Services 将对下载文档进行鉴定和加密。



SoftBook



RocketBook

NuvoMedia 的 Rocket eBook 宽 5 英寸、高 7.5 英寸、厚 1.5 英寸、重 22 盎司,外形上较 SoftBook 更小巧,它可以容纳 4000 页的信息,通过连线与计算机串口相连,通过 PC 机将信息下载到 RocketBook 中。RocketBook 为读者提供了为文档加批注、着重线以及标记书签等功能,还为读者提供字典,并提供了 Rocket Writer 软件让读者将自己编写的 HTML 文档转换为 E-Book 文档并加入到 E-Book 阅读机中去。

阅读机的背景均为白色,文字显示为黑色或灰色,具有较高的可阅读性。

SoftBook 和 NuvoMedia 都有自己的网上售书连锁店。

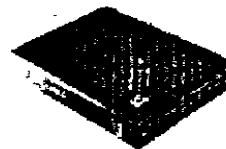
除了上述最早销售的两种阅读机外,EveryBook、GlassBook、Librius、Microsoft 也已经推出自己的阅读机和阅读软件。

EveryBook 推出号称“真正的 E-Book”的售价为 1500 美元的阅读机,与上述两种单液晶显示屏的阅读机不同,它首先采用双显示屏设计,外形更向传统的书

的造型,采用这种设计,扩大了阅读器的显示窗口,更利于信息查找与浏览。



打开的 EveryBook



关闭时

除此之外,Microsoft 公司正在开发一个名为 Reader 的软件,用于阅读 E-Book 文档,该软件首次包含了 Microsoft 公司最新研究的 ClearType 字符显示新技术。2000 年 4 月 19 日 Microsoft 宣布集成卡西欧、康柏和惠普的手持式 PC 电脑,Reader 的 Pocket-PC 版首次亮相,该软件的 PC 版本将在年内的晚些时候发布。

为促进 E-Book 的发展,Microsoft 还设立了十万美元的奖金,用来奖励为 E-Book 发展作出卓越贡献的组织和个人,并得到了其它公司的支持和加盟。

### 三、OEB1.0 简介

OEB1.0 是“Open eBook Structure Specification 1.0”的缩写,是 OEB 组织成员为更好地开拓 E-Book 市场而联合制定的 E-Book 文档标准。它规定了 E-Book 的文件格式和结构,保证了 E-Book 文档可以正确显示在任何阅读系统中。该标准的使用群体是作家、编辑、出版商、E-Book 文档创建者、电子书出版系统及阅读机厂商。

OEB 文档基于 HTML 和 XML 语言,是标准的 XML 文档。OEB 出版物就是一系列 OEB 文档的组合,另外还配有一个包装描述文件,主要记录 E-Book 的创建信息,包括创建者、创建日期、出版者、版权声明等。

OEB1.0 结合了 XML1.0 标准、XML 命名标准、HTML4.0 标准、CSS1 样式语言及部分 CSS2 样式语言、Dublin Core 和 USMARC 描述语言、Unicode 字符集以及 MIME 媒体类型等一些标准,制定出 OEB 的 DTD 文件。OEB 文档要以(<?xml version='1.0'?>)开始,表明这是一个 XML 文档。OEB 标准中规定了每个 E-Book 出版物要有且只有一个包装描述文件(Package file),并定义了 Package file 中所包含的描述项目。

OEB1.0 的制定虽然统一了 OEB 的文档结构和格式,但并没有包含有关版权保护的内容,而版权保护问题也是决定 E-Book 能否健康发展的一个关键问

题,目前,有很多公司正在致力于电子出版物的版权保护工作,如 GlassBook 公司,并成立了相关组织 EBX。

#### 四、关于 Microsoft Reader 与 Clear Type

Microsoft 参与制定了 OEB1.0 标准,并在 OEB 组织内起到很关键的作用。在推广 E-Book 的实际工作方面,Microsoft 已经宣布在明年初发布一个用于阅读 E-Book 的应用软件 Microsoft Reader,并且宣布在软件 Reader 中首次使用最新的字符显示技术 ClearType。

据介绍,ClearType 是一种能使文字在显示器上的显示效果象印在纸上一样平滑清晰的字符技术。ClearType 的使用将加速书刊出版从纸介质转移到电子书的进程,由此出版商可以为 E-Book 的读者提供高质量的阅读体验,字体设计者也有了更多的发挥余地,而便携式设备制造商将从能够同时提供高质量的色彩和分辨率的彩色液晶显示器上获利。

传统的计算机字符显示面对的是每个类似小方格一样的象素值,通过设定每个方格的开关(填充或不填充)来显示字符,这种方式使得一些字符的边缘带有明显的锯齿。ClearType 技术改进了这种选择方格开关似显示方法,获取字符的显示区域超出了传统意义上的象素界线,使得字符在显示器上的显示效果要平滑得多,没有锯齿。ClearType 技术对 CRT 显示器的字符显示有一定的改善,但它给数字式液晶显示屏(LCD)的字符显示带来的是一种飞跃,这将极大地改善掌上电脑和高质量平面桌面系统的显示。一些业界人士在看了 ClearType 技术后都给予了很高的评价,一致认为该技术将字符技术的研究加快了至少十年,是字符技术发展的里程碑,真正使得在电脑上阅读书籍成为可能。

Microsoft Reader 是一个可供人们阅读 E-Book 的应用软件,它的目标是给所有的使用者一种类似于读纸介质书籍的高质量的、舒适的体验。该软件结合了传统的排版方式和最新的字符显示技术,主要功能如下:ClearType 字符技术;清晰、规范的显示;充足的页边空;合适的字距和缩排;强大的加着重标记、注解、书签功能;内置词典;类似书架的 E-Book 管理系统;具有复制保护系统。

Microsoft Reader 的推出将改善 E-Book 领域目前没有一个大型的、普及的、具有影响力的阅读软件的局面,它将鼓励出版商在 E-Book 领域进行更大胆的投入。

随着 Microsoft Reader 发布日期的确定,一些出版商纷纷作出响应。1999 年 10 月,Penguin 发布了 Penguin eBook CD 以响应 Reader,Mondadori 选择

Reader 配合意大利语 E-Book 的推出,Havas 选其做为法语 E-Book 的支持,00h00.com 也将扩大其在线书籍的发行以支持 Reader 的推出。

目前,Reader 的 PocketPC 版本已经发布,并集成在卡西欧、康柏和惠普的手持式 PC 电脑中,其 PC 版本将在稍后推出。

#### 五、关于 E-Ink 与 E-NewsPaper

E-Book 向我们走来的时候,另一种电子媒体也离我们越来越近了,它就是 E-NewsPaper,电子报纸,电子报纸与电子书的初衷都是一致的,就是要取代纸介质的出版物,代之以数据存储的电子文档形式。它也有自己的便携式阅读器,与 E-Book 阅读机不同的是,E-NewsPaper 阅读器在外形设计上更像是一张报纸,而且它也不采用液晶屏作为显示屏,而是使用一些新的材料如一些类似玻璃纤维的物理材料以及 E-Ink(电子墨)等来显示报纸内容,这样制作的 E-NewsPaper 厚度非常薄,甚至可折叠。



从目前已经看到的资料了解到,IBM 开发出世界上第一份电子报纸(如上图)。它是一本很薄、很轻的“电子书”,有一个金属框,上面有一层透明的保护层,内有 16 页用玻璃纤维加工的纸,下面则有几个功能键,用来更新内容和传递资讯。

它使用 IBM 研制的“电子墨”,以黑白两种颜色印刷,颜色可以保持一段时间。当它模糊不清时,读者必须给它充电,之后会恢复原色。换句话说,报纸可以反复使用,这将节约大量的纸张。

电子报纸比现在的报纸稍大,可以平放在提包内,也可以像现在的报纸一样折起来或剪贴。IBM 的科学家在设计时,充分考虑到人们阅读报纸的习惯,如把报纸叠起来带走等。读者通过 Internet 网为它更新内容,将新闻与图片下载到电子报纸上,之后使用“电子墨”将内容展示在 16 页玻璃纤维纸上。

电子墨由成千上万个充满着亮色和暗色染料的圆球状小胶囊组成,通过电压的控制快速地变化以显示不同的颜色。电子墨将由于它低廉的造价和轻薄的外形而对液晶显示屏造成冲击。

(上接第 113 页)

在处理查询时,假设我们有一个原子查询要求查询有关 A 和 B 的内容,我们考虑所有从属于 A, B 的对象,假设有一片段的描述具有属性 C, D, 且有  $C \xrightarrow{(a)}$  A,  $D \xrightarrow{(b)}$  B, 则我们定义这个片段符合该查询条件的相似度为 ab。

知识可以由人工传授,即手工输入,也可以通过对访问历史记录进行知识发现来获得。

由于非语义信息的匹配结果通常也是一个相似度的数值,此时我们就把基于非语义信息的查询和基于语义信息的查询统一起来了,如果原子查询间用“与”连接,则相似度取较小的一个,如果原子查询间用“或”连接则取较大的一个。最后将相似度大于给定阈值的片段按照相似度排序后作为查询结果返回。

### 3.4 与其他查询系统的比较

现有的多媒体资料查询系统可以分为两类,一类是基于如颜色、纹理、形状等特征的查询,通过特定算法与库存资料的相应特征进行比较,以相似度为衡量标准判断是否将其作为查询结果。另一种做法沿袭了针对文本或类文本(例如超文本)资料的查询检索方法,使用关键字,与预先对多媒体资料标注时所用的关键字进行匹配,如有匹配,则认为该段资料符合查询要求。

本系统中对象属性包含非语义属性,也就是说能够将第一种做法纳入这个查询体系中去。与第二种做法相比,它将元数据先横向分为内容元素和谓词,再纵向地分为多个层次,使语义的表达更为精确,使元数据能够比关键字更有效地得到利用,同时让用户也以这种方式精确地描述自己感兴趣的多媒体资料,使查询的结果更准确。这种查询方式还进一步考虑了领域知识的引入,通过一定的推理返回相关的查询结果,与现有的查询系统相比更加智能。通过引入推理的肯定系

数,让语义信息的匹配也有了一个相似度的概念,从而较好地非语义信息和语义关系纳入了一个统一的查询框架之中。

**结束语** 本文提出了一个分布预取式 VOD 系统的体系结构,并基于这个结构提出了合适的调度算法,其优势在于硬件的基本要求不高,可扩充性强,响应请求快,吞吐量较大。另外,针对多媒体信息检索这个问题,本文提出了一个灵活的基于内容的检索办法,综合考虑非语义信息和语义信息,将其纳入统一的框架之中,对元数据进行了分类、分层,使查询的表达及其结果更准确,它还进一步考虑了领域知识的引入,通过一定的推理返回相关的查询结果,与现有的查询系统相比更加智能。

### 参考文献

- 1 Gemmell D J, et al. Multimedia Storage Servers: A Tutorial. IEEE Computer, 1995, 28(5): 40~49
- 2 Requirements Group. MPEG-7: Context and Objectives (version-10 Atlantic City), MPEG98 October 1998/Atlantic City, USA
- 3 James A, et al. Multiple Level Classification of Visual Descriptors in the Generic AV DS. MPEG99 Melbourne, Australia, Oct 1999
- 4 Brubeck D W, Rowe L A. Hierarchical Storage Management in a Distributed Video-On-Demand System. IEEE Multimedia, 1996, 3(3): 37~47
- 5 Chang S -F, et al. Visual Information Retrieval from Large Distributed On-line Repositories. CACM, 1997, 40(12): 63~71
- 6 Benitez A B, et al. Using Relevance Feedback in Content-based Image Metasearch. IEEE Internet Computing, 1998, 2(4): 59~69
- 7 向哲,钟玉琢,龙白滔. 视频服务器的体系结构. 计算机世界, 1999, 13
- 8 Available at: <http://www.walkingdog.com>
- 9 Available at: <http://www.davic.org>
- 10 Available at: <http://www.openh323.org/standards.html>

## 计算机科学

(1974年1月创刊)

第28卷第1期(月刊)

2001年1月25日出版

中国标准刊号: ISSN 1002-137X  
CN50-1075/TP

定价: 12.00元 国外定价: 5美元

邮发代号: 78-68

发行范围: 国内外公开

主管单位: 国家科学技术部

主办单位: 国家科技部西南信息中心

编辑出版: 《计算机科学》杂志社

重庆市渝中区胜利路132号 邮政编码: 400013

电话: (023) 63500828 E-mail: [jsjcx@swic.ac.cn](mailto:jsjcx@swic.ac.cn)

主 编: 朱宗元

印刷者: 国家科技部西南信息中心印刷厂

总发行处: 重 庆 市 邮 政 局

订购处: 全 国 各 地 邮 政 局

国外总发行: 中国国际图书贸易总公司(北京399信箱)

国外代号: 6210-MO