

综述·评论

●谢新洲

情报科学技术的新发展

ABSTRACT Information science and technology has in the recent years make a very fast progress. The paper introduces new progresses in: parallel computing technology, new pattern access technology, user interface technology, multimedia technology, hypertextual technology, object-orientation programming, fulltext retrieval and electronic publication. 10 refs.

SUBJECT TERMS Information technology — Progresses Computer technology — Progresses

CLASS NUMBER G356

情报技术的发展在情报科学发展史中起着极其重要的作用,以至国外有的专家认为,情报科学的发展史在某种程度上就是情报技术的发展史,计算机用于处理文献就出现了情报科学^[1]。当然,此观点有失之偏颇之处,但它从一个侧面说明了情报技术的重要性。近几年来,随着计算机技术和信息处理技术的不断发展,情报技术已发生了日新月异的变化,其中给人耳目一新的超文本和多媒体系统的出现、影响出版业发展未来的全文本和电子出版技术的结合、新一代的面向对象的程序设计、海量存贮设备和并行技术等,已经或将影响情报科学的发展方向,扩大情报科学的内涵和应用范围,使情报科学更加普及和具有更大的活力,为人类获取所需信息提供更加方便的条件,并将为信息的开发利

用在深度上大大推进一步。

一、并行计算技术

传统的信息检索大都采用顺序处理方式,利用计算机进行顺序存贮和顺序检索,这是由于传统的冯·诺伊曼计算机是在完全确定顺序的数据值上完成一系列预定指令所决定的。在这种方式中,只要合理地组织文档数据,设计优化的检索软件,采用大容量的宿主计算机,系统可以在一定程度上保证用户的检索响应时间。然而,在系统的数据库中,随着数据量的不断增大和全文等新型数据库的出现,这一方式就难以满足用户需求。于是人们就开展了信息检索的并行处理研究。这对信息检索系统特别有利,因为数据库查询适合于并行处理,数据库分区和数据库的不同

部分可在不同的处理器上并行查找,结果将大大提高检索速度,还可采用协处理器,进行信息检索的文本压缩和字符串匹配。文献[2]指出,利用并行计算机实现全文检索,有利于改进响应时间,有利于查找大型数据库;允许采用超级算法,还可降低查找成本。

并行计算机用于信息检索,带来了新的研究课题,如并行算法研究、并行体系结构设计和数据库多处理机分布。并行检索算法的设计有 3 种可能的途径,即在现有的顺序信息检索算法中标识和探讨并行性,采用或修改已有的并行算法,设计全新的并行检索算法。并行算法研究主要集中在模式匹配算法的实现,基于正文标记的检索以及倒排文档检索等方面。对处理信息检索任务的并行体系结构设计的研究主要表现在数据库专用机和检索机上。数据库专用机就是从单指令单数据的体系结构进入面向单指令多数据的机器。对数据库的多处理机分布的研究主要集中在将数据库的数据分散到分布式信息检索系统中的各台机器中去,通过开发数据处理软件改进系统性能,以达到使系统具有响应快、可用性好、资源共享、适应负载变化、文档分割安全及易扩充等目的。

二、新型存贮技术

在存贮技术中,最显著的发展是以 CD-ROM 为代表的海量存贮技术。海量存贮技术的发展为多媒体技术的出现提供了物质条件。据国外一些人士预测,除了光盘存贮技术外,可能大大影响信息存贮介质的是磁光盘和数字纸^[3]。

(一)CD-ROM 光盘。光盘技术是一种将计算机技术与精密激光技术结合起来的大规模数据存贮技术。CD-ROM 是其中的一种主要产品,此外还有可擦写一次的光盘 WORM 和可擦写多次的光盘等。

CD-ROM 是一种只读存贮介质,很适合于存贮大容量的信息。较之磁盘,CD-ROM

存贮容量要大得多,是一种海量存贮器,其存贮容量达到 650MB,相当于 5.25 英寸低密盘的 1800 倍,可将文字、静态图象、声音和动态图象等多媒体信息结合在一起。CD-ROM 的可靠性高、耐用性强,而磁盘受热、湿、空气和污染物的影响较大。数据衰变,杂散磁场,机械振动都会直接影响磁盘记录的可靠性。CD-ROM 主要缺点是制造母盘复杂、成本高、数据访问速度低、速度传输率低和查找的响应时间长。

CD-ROM 的出现极大地促进了电子出版物的发展。自从 1985 年首张 CD-ROM 问世以来,只读型光盘获得了迅速增长。截至 1989 年,全世界已有各种 CD-ROM 数据库 496 种,1990 年 861 种,1991 年 1522 种,1992 年 2214 种,已有 2602 家公司在从事这种产品生产。

(二)数字纸(digital paper)。数字纸是继 CD-ROM 后出现的一种新型只读光存贮介质。数字纸记录和读取信息的原理与 CD-ROM 的原理基本相似。它首先将信息数字化,变为二进制码,再用红外激光照射数字纸,激光在记录层上烧蚀出凹坑。烧蚀出的凹坑与未经烧蚀的部分对红外激光产生不同的反射光,根据反射光的强弱,转换成机器码,再转换成能辨认的信息。目前,数字纸产品有光带盘和软光盘。数字纸较之 CD-ROM,性能有很大改进。首先,数字纸有更高的存贮密度和容量。数字纸利用激光的高聚性和能使激光束准确定位的装置,使得每英寸能容纳 16000 道,一个直径为 12 英寸的光带盘由 731 米长、35 毫米宽、75 微米厚的数字纸卷成,可存贮 10000000 兆字节,一个软光盘的容量也达 1000 兆字节。其次是查找和数据传输速度较快。光带盘的平均查找时间为 28 毫秒,最大查找时间为 60 毫秒。软光盘的平均查找时间为 40 毫秒,而 CD-ROM 的平均查找时间为 1 秒,相当于硬磁盘的速度,比 CD-ROM 要快得多。再次是价格低廉,数字纸存

贮密度达到每平方英寸 10 兆字节,存贮 1 兆字节的成本为 0.5 美分,而用纸张存贮 1 兆字节的费用要 3~6 美元。与盒式磁带每 1 兆字节 18 美分相比,也要便宜 36 倍。另外,数字纸的信息保存时间长,记录在数字纸上的信息至少可保存 15 年,CD-ROM 只有 10 年左右,而磁带只能保存 2~10 年。数字纸的不足之处是对环境有一定要求,光带的存贮温度为 5℃~65℃,操作期间温度要控制在 16℃~32℃之间。另外,数字纸作为一种产品,设备的价格太高。

(三)磁光盘。磁光盘是一种可擦写光盘,近几年来发展强劲,大有与硬盘并驾齐驱之势,并成为多媒体数据库的主要存贮介质。目前,磁光盘有 14 英寸、5.25 英寸和 3.5 英寸等类型。其中以 5.25 英寸磁光盘发展最快,应用最广。5.25 英寸的磁光盘大都采用 SCSI 接口与 ISO 标准盘盒。目前世界上有佳能、理光、索尼、夏普、日立等一些著名厂家生产的 5.25 英寸的磁光盘。索尼公司的 SMO-D501 型驱动器的双面格式化容量为 600MB (512 字节/扇区, 34 扇区/道)或 650MB (1024 字节/扇区, 17 扇区/道),1991 年索尼试制成了双面容量高达 4GB 的 5.25 英寸磁光盘。这种磁光盘比目前一般的磁光盘记录密度高 6 倍,其数据传输率也较目前的 0.925MB/S 提高 3 倍,存取时间为 30ms。1988 年,完成了 5.25 英寸磁光盘驱动器的标准化工作。目前,磁光盘的存贮成本为 0.65~4 美元/MB,驱动器成本约为 5.5 美元/MB。3.5 英寸磁光盘特别适合 PC 机与膝上型机。磁光盘利用磁光存贮技术存取信息,而磁光存贮技术又利用聚焦激光束在垂直磁记录介质上写入和读出信息。磁光盘的主要优点表现为^[4]:

1. 存贮容量大,有中等水平的平均存取时间;

2. 可以随意在驱动器上装卸盘片,便于携带、保存与交换,有良好的可换性;

3. 采用非接触读写技术,因此可保证有较高的数据可靠性与完整性;

4. 有较强的抗震与耐冲击性,不会破坏数据和驱动器,安全性高;

5. 可在盘上反复读写与更新数据,反复使用次数可高达 $10^6 \sim 10^7$ 次;

6. 具有较高的数据传输率与信噪比。

由于具有以上优点,磁光盘的销售额在快速增长。预计到 90 年代中后期,大容量可擦写磁光盘将在性能价格比上达到当时硬盘驱动器的水平,在外存领域中取得与磁盘同等重要地位。

未来磁光盘的发展,主要有以下几个方面^[5]:进一步提高存贮容量与记录密度;通过提高盘片转速与实现直接重写来提高数据记录与改写速度;改进光头设计,减轻光头重量,缩短存取时间;进一步改进记录介质;采用多光束并行读写的磁光盘驱动器,进一步提高数据记录与传输速度;开发多功能光存贮媒体与多功能光盘驱动器,生产自动换盘多光盘库系统等。

三、用户接口技术

用户接口技术是近年来研究人员和信息用户关心的一个重要领域。用户接口设计的好坏直接关系到检索系统的成败。人们要求接口做到“用户友好”。这一方面体现出信息检索技术越来越成熟,另一方面也体现了信息检索技术越来越普及。命令驱动和菜单选择是两种传统的用户接口。命令驱动面向有经验的用户,可以比较自由灵活地表达检索策略;菜单选择适合无经验用户,简单明了,但有时显得繁琐,不适用于复杂的提问。人机接口技术经历了字符命令、菜单选择、窗口显示、多窗口系统、图形用户界面和智能接口等几个阶段。人机接口的设计必须考虑系统与用户的交互,方便用户对系统的理解,达到交互自然、美观、易用的目的,并具有容错等功能。目前,信息检索系统的人机界面技术主要

集中在多窗口界面方面,如弹出式窗口、下拉式窗口,以达到用户对屏幕的美感和方便。至于对用户和系统间的交流研究,还做得很不够。人机接口技术的今后发展重点将是智能接口和图形用户界面(GUI)技术。

智能接口技术主要解决两个问题,即如何克服用户和系统之间的语言障碍,以及如何在概念上帮助用户分析其检索需求,制定合适的检索策略和评价检索结果。为此,在智能接口设计中要解决以下几个技术问题:1. 自然语言的理解和对话管理;2. 检索中对专业主题(包括可用的数据库及其内容和术语)的表达;3. 检索技术的表达,如询问语言、检索策略制定;4. 中间用户的技巧和经验的引导及表达;5. 知识处理与推理管理的求解方法设计。

图形用户界面技术主要实现以下目标^[6]:1. 允许用户通过操纵屏幕显示的图形、图象元素而直接操纵应用系统;2. 用户通过鼠标或其它指针移动光标来完成对应用系统的直接操纵;3. 允许用户对界面进行一定的剪裁和定制;4. 提供对用户操作错误的保护机制和强有力的帮助机制;5. 提供强有力的开发工具,使开发者能够方便地开发出在其环境下运行的应用程序;6. 提供一个界面设计标准,使各种应用界面显示风格与用户交互的方式具有较好的一致性,便于用户学习和使用。

四、多媒体技术

多媒体技术是近年来迅速发展的一個领域。由于其灵活、宽松和亲切的服务效果而引起人们越来越大的兴趣,并使许多研究人员投身到该研究领域中去。多媒体技术的出现与发展,使得计算机除了能处理文字信息外,还具有综合处理声音、文字、图形、图象和动画的能力。多媒体技术已应用于许多领域,深刻地影响着信息处理的发展,并在信息服务市场上占有重要的地位。用于信息检索的多

媒体系统,在设计时除了数据模型及其实现问题外,还包括以下几个具体问题^[7]:1. 多类型存贮设备的管理;2. 多类型I/O设备的管理;3. 系统的分布性和可移植性;4. 开放的系统结构,允许用户将应用程序装入系统,或具有与其它软件相连的接口;5. 具有多种提高检全率和检准率的措施,如索引语言、检索策略和算法;6. 联机更新;7. 按多种版式和用户定义格式进行灵活输出;8. 具有方便系统人员、最终用户与系统进行交互的友好环境。

五、超文本技术

超文本技术是近几年迅速发展起来并与传统技术不同的一种非线性技术,它可以按人脑的联想,把相关的信息联系起来,供用户浏览和检索。超文本技术不仅注重信息本身,更注重信息之间的联系,把相关的事物联系在一起,形成网络结构。

超文本技术模拟人脑的联想功能,将计算机的存贮信息能力和人脑的筛选信息能力结合在一起,一方面提高了检索效率,另一方面让用户控制系统的建立,使用户感到更友好。超文本的关键技术是超文本数据库,数据库存贮结构中有用于存贮信息的结点和表示信息之间关系的链。近年来,超文本技术被广泛用于多媒体的管理,出现了所谓的超媒体技术。超文本系统与普通信息系统比较,具有以下特征^[8]:

(一)采用了动态的、开放式的设计方法。允许用户借助链路,从一篇文献的主体思想转移到任意大小或任何内容的相关信息上,添加新的信息,只需键入连接到其它信息结点即可,无须重新设计记录格式。

(二)超文本的内容排列,与普通系统不同,它是按知识单元及其关系建立的知识结构网络。

(三)超文本系统除了处理一般的文字材料外,还包括图片、地图和其它直观材料,把文字资料和图象资料有机地结合在一起。

(四)超文本是一种人机交互的友好系统。用户利用计算机可以增删超文本的内容,加注评语,修改或重建立体知识网络。

(五)超文本系统是一种通讯工具,它可以组织许多研究人员通过各自的终端沟通情况、交流思想。

六、面向对象的程序设计

面向对象的程序设计方法是 90 年代以来流行的方法,许多程序设计语言,如 Smaltalk、C++ 等都采用面向对象的方法,从而产生了面向对象的程序设计语言。有人预言,面向对象方法在 90 年代将占居软件技术的主流地位。面向对象的方法已经在诸如 CAD、办公信息系统、图形化用户接口、多媒体应用、分布式系统等非传统应用中发挥了卓有成效的作用。

在面向对象的系统中,所有的概念客体都被模型化为对象,对象由给定的数据结构和对这些数据结构进行的某些操作所构成。面向对象系统内部的所有操作都通过对象间的消息传递而实现。面向对象方法和对象模型中的另一概念为类,类由一组对象构成,它提供了对具有共性的对象进行聚类的机制。面向对象方法的优越性主要表现在^[9]:

(一)支持不同层次的数据抽象概念,对特定数据类型或“类”的所有操作都可以集中起来,以增加数据库的模块化程度。根据这一特性,可以方便地获取一类对象的共同性质。

(二)支持封装性。封装性是一个源自抽象数据类型的概念,对象封装性具有两方面的含义,一是对象将状态和操作集中起来,二是对象的操作只能由自身的操作存取。

(三)支持继承性,一个类可以成为另一个类的子系,从而允许程序员根据需要而在继承原有的基础上定义更具体的类。

(四)支持多继承性。允许类的任意相交,即允许一个对象属于不同的类,这个特性导

致格类型的概念结构。

(五)支持多态性,同一类的所有对象在收到同一条消息时将采取同样的动作,不同类的对象在接收到同一条消息时可能采取不同的动作。

七、全文检索和电子出版

全文检索技术的发展应归结于信息检索技术、海量存贮设备和用户需求几方面。海量存贮设备为全文数据库提供了物质基础,检索全文本章、节、段、句甚至字的技术为全文检索系统提供了技术条件,而用户希望能获得原始信息的需求则是全文检索技术发展的直接社会条件。一个全文检索系统包括全文数据库和全文检索软件两部分。目前,全文检索在厂商名录、产品介绍、合同规范、工程标准、证券股票、科学技术甚至文学领域都得到了迅速发展,显示出了其强大的生命力。全文数据库在数据库行业中所占的比重越来越大。全文检索提高了检索系统的检全率,但由于多数系统采用自由文本检索(有些全文系统采用受控词检索,对全文进行标引)而降低了检索系统的检准率,并为联机用户,尤其是无经验的用户确定检索用词带来了严重困难。目前,国外一些全文数据库已以 CD-ROM 等电子出版物形式出版发行。

电子出版技术是近年来迅速发展的一门技术,受到了人们的广泛重视,被认为是继活字印刷术之后出版业的一次革命。电子出版技术借助计算机和远程通信技术,吸收全文检索、多媒体和超文本等技术而发展起来。电子出版有 3 方面的含义^[10],其一是指图书、期刊或报纸生产过程中的计算机排版技术,实际上是指出版物生产过程的计算机化,最终产品仍以纸张为载体;其二是不仅采用电子技术手段,而且最终产品是以磁盘和光盘等作为载体,并要借助于计算机才能阅读的

(下转第 66 页)

心,其中的亚洲地区中心设在日本国会图书馆。

4、国际资源共享计划(UAP)。中心设在英国不列颠图书馆文献提供中心,负责人为D·布雷德伯里。

5、第三世界图书馆事业发展规划(ALP)。1984年拟订。中心设在瑞典的乌普萨拉大学图书馆,负责人为B·贝格达女士。

纵观国际图联近70年的发展历史可以看出,国际图联已在两个方面发生重大变化:一是成员方面,已从以欧洲为中心转向国际图书馆界大家庭;二是活动方式,已从务虚转向务实,即已从以会友转向全面开展科研活动。显然,国际图联在提高图书馆社会地位、塑造图书馆社会形象、推动世界图书馆事

业发展方面做了大量工作,已被各国公认是图书馆界的国际组织。

参考文献

- 1 丘东江. 国际图书馆协会联合会参考资料. 南京: 江苏省图书馆学会, 1993: 1~58
 - 2 董小英. 国际图联的历史、现状与前瞻. 津图学刊, 1993, (3): 28~37
 - 3 柯平. 国际图联的组织与活动概要. 图书馆建设, 1993, (5): 65~68
 - 4 李月琳. IFLA的产生和发展概况. 津图学刊, 1994, (1): 81~91
 - 5 柯平. IFLA“国际书目控制与国际机读目录”计划的回顾与展望. 大学图书馆学报, 1993, (5): 18~23, 29
 - 6 国际图联执委会制定; 董小英译. 国际图联的长期政策. 图书馆, 1991, (6): 38
- (来稿时间: 1994—06—06. 编发者: 丘峰。)

(上接第53页)

电子出版物; 其三是泛指一切以电子形式传递信息的技术, 即所谓“无纸”出版, 包括可视图文和电子邮件等。目前, 世界上的许多国家已在出版电子图书、电子期刊和电子报纸等。据有些人士认为: 电子杂志由于提供信息快, 用户使用方便, 有可能成为下世纪科学杂志的主流; 同时, 芯片式电子图书也将获得迅速发展。与此同时, 版权保护将是一个值得重视的社会问题。

参考文献

- 1 D. B. luly, R. W. Trice. A History of Information Science: 1945-1985. San Diego, Academic Press, Lnc, 1989
- 2 赵棒未等. 信息检索的新发展——并行信息检索. 见: 第十届全国计算机情报管理学术讨论会会议文集. 宜昌, 1993: 53~54
- 3 Miles M. Jackson. The Future of Books in an Electronic Era. The International Information & Library Review. vol24(4), 1992, 307~313

- 4 郑登理. 可擦重写磁光盘的现状及其发展动向. 情报学报, 1993, 12(2): 145~154
- 5 郑登理. 从与硬磁盘机的比较看磁光盘技术的进展. 见: 第十届全国计算机情报管理学术讨论会会议文集. 宜昌, 1993: 29~210
- 6 曾民族, 陈豫. 迎接新信息检索技术的挑战. 见: 第十届全国计算机情报管理学术讨论会会议文集. 宜昌, 1993: 4~5
- 7 谢新洲, 陈睿. 多介质情报检索系统的基础问题. 情报学报, 1991, 10(6): 429~436
- 8 陈光祚主编. 计算机情报检索系统导论. 北京: 书目文献出版社, 1993: 284~288
- 9 陈睿, 谢新洲. 新一代信息系统——面向对象信息的分析与设计. 北京: 航空工业出版社, 1993: 50~62
- 10 谢新洲. 电子出版物的发展及其影响. 北京大学学报(哲社版), 1993(6): 109~116

谢新洲 武汉大学科技情报专业毕业。现为北京大学信息管理系讲师。已发文30余篇, 出版专著1种, 译著1种. 通讯地址: 北京大学, 邮编100871。
(来稿时间: 1993-03-07. 编发者: 丘峰。)