

王心裁 吕元智

超媒体数据库技术与档案信息组织^{*}

摘要 超媒体是超文本的多媒体扩充,它既可以处理文本、图像等静态信息,又可以处理声音、视频、动画等动态信息,是一种以全新的方式组织信息并对相关信息提供关联的技术集合。超媒体技术是按照人类联想思维的习惯来进行网状的非线性的信息管理,目前,已在科学、文化、教育、地理信息等方面被广泛运用并取得了成就。图2。参考文献8。

关键词 超媒体数据库技术 档案信息组织 信息管理

分类号 G250.73

ABSTRACT Hypermedia is a multimedia extension of hypertext. It is a set of technologies to organize information in a brand new way and to correlate related information, and can process not only still information, such as text and image, but also dynamic information, such as sound, video and animation. It realizes web-like nonlinear information management according to human association thinking, and has been widely applied in science, culture, education, geographical information, etc. 2 figs. 8 refs.

KEY WORDS Hypermedia database technology. Organization of archival information. Information management.

CLASS NUMBER G250.73

1 超媒体数据库技术

1.1 超媒体

超媒体是超文本的多媒体扩充,它既可以处理文本、图像、图形等静态信息,又可以处理声音、视频、动画等动态信息,是一种以全新的方式组织信息并对相关信息提供关联的技术集合^[1]。其最基本的组成单元是节点和链,注重的是信息间的相互关系。在超媒体中,信息被编码成一个个节点,相关的节点通过机器所支持的连接方式链接在一起,用户通过关联信息的线索来寻找关系以获取信息。超媒体技术是超文本和多媒体技术的结合。在功能上,它既保留了超文本灵活、非线性的信息管理优势,又利用了多媒体生动、自然的表现特点。超媒体技术是按照人类联想思维的习惯来进行网状的非线性的信息管理^[2]。目前,超媒体技术已在科学、文化、教育、新业务推广、地理信息系统等方面被广泛运用并取得了令人瞩目的成就。

1.2 超媒体数据库

超媒体数据库是一种能够有效实现超媒体数据的存储、读取、检索等功能的数据库系统,是多种式

样信息的载体。与传统数据库相比,超媒体数据库具有数量大、集成性、实时性、非解释性、非结构性等特性^[3],其功能也有了很大改进。超媒体数据库是多媒体数据库的延伸,它继承了多媒体数据库的优点,如数据的独立性、利用数据库查询语言进行高层次查询、并发控制、容错技术等,又拓展了新功能,如对时空关系的数据进行同步管理等。它还具有友好的交互式用户接口和可视化的编程环境,使用起来形象直观,用户可以与计算机通过超媒体数据进行交互对话。

超媒体数据库是开发现代多媒体档案管理库管理系统的基础。它在档案信息存储、管理、开发、检索、利用、编研、宣传等方面的作用已初现端倪。

2 现代档案的特点及其信息组织要求

2.1 现代档案的特点

计算机技术和通讯技术的出现与发展,使得人类记录和传递信息的方式发生了革命性变化。在人类社会活动中直接形成的记录——档案,也随之发生变化。与传统档案相比,现代档案除了档案的基本特点——原始性、凭证性外,还有一些新特点。

^{*}本文为教育部人文社科规划项目“网络信息资源组织与开发模式”论文,项目编号:2000ZDXM870001。

(1) 由于国家、地区和行业发展不平衡,现代化技术应用程度不同,在各项活动中形成的档案载体也不尽相同。目前既有传统纸质档案、胶片(卷)档案,也有现代数字化载体档案。档案载体呈现多样化局面。

(2) 为了全面记录某项活动,人们往往采用多媒体技术同步记录某项活动的文字、声音、视频等多样信息,形成多媒体文件。由这种多媒体文件转化来的档案,是多种媒体综合记录的结果。随着多媒体技术的广泛应用,档案的媒体出现复合化趋势。

(3) 现代社会,人们往往借用现代科学技术对事件的全过程进行完整、真实的记录。如用数码摄像机可以对一次会议的全过程进行全方位立体式记录。由这种记录方式形成的档案,能形象、逼真地再现当时的活动。

(4) 档案是人类在社会实践活动中形成的原始记录,是一种知识化的信息记录。在知识经济社会,档案的知识性将会更加突出。

2.2 档案信息组织的要求

(1) 精确性要求。由于现代档案范围广泛、数量大、类型多、内容复杂,大量档案信息湮没在其中。为了把档案信息准确揭示出来,在组织档案信息时,要严格遵守精确性这一要求。精确性要求一方面是指标引的档案信息具有专指性,能够准确反映问题;另一方面是指标引人员应客观地分析出文献所实际具有的主题,避免主观,切忌仅依文献题名进行主题分析。同时,应根据档案的具体内容和用途,选定适当的标引深度。

(2) 全面性要求。这包括两个方面:其一指主题分析与标引要全面反映文献所含的信息;其二是指档案著录要全面。既要著录反映档案内容特征的项目,也要著录反映档案形式特征的项目。全面的著录能提供多途径检索入口,是提高档案利用率的基础和前提。

多媒体数据所含的信息量丰富。媒体数据内及其各种组合包含有难以用符号化方式描述的信息和信息线索。信息线索是基于内容查询的重要条件,存在于媒体数据之中和媒体数据之间,可以用特征词及其组合来描述,或通过计算机自动进行特征抽取,这与用户的经验和知识以及特定的应用领域及上下文有关^[4]。

(3) 一致性与标准化要求。档案著录的一致性是指在著录项目的选择及标识符号的使用等问题上应尽量保持一致。标引的一致性是指对赋予文献的标引词及标引深度应始终保持一致与稳定。标准化是档案信息组织的基准,是档案信息能够共享的基础。

(4) 知识化和智能化要求。为了准确提供档案信息,让人们在利用档案时能最大限度获取知识,档案信息组织要尽量地把档案信息转化为可识别、可

利用的信息,便于人们掌握,并在实践中转化为指导行动的知识。在档案信息组织时,还应提供智能化导航,减少信息“迷路”,为人们查找和利用档案信息提供智能化支持。

(5) 符合思维习惯要求。传统档案信息组织方式如档案目录组织法一样是线性结构方式,只有掌握了档案信息检索的基本知识,严格遵守检索格式,才能对其进行有效利用。为了提高档案信息的利用价值,现代档案信息组织要克服传统档案信息组织方式的缺点,按照人们的思维习惯,利用非线性的网状结构进行组织,以方便人们及时有效地查询利用。

3 超媒体数据库技术在档案信息组织中的应用实现

超媒体数据库技术在档案信息组织中的应用,主要就是构建档案信息超媒体数据库。

3.1 思想设计

思想设计是构建档案信息超媒体数据库的基础和首要环节,是超媒体数据库的灵魂,它的质量好坏直接影响到后来工作的成败。构建一个好的档案信息超媒体数据库,功能及其实现是设计者必须考虑的两个重要方面。

功能方面的要求有:第一,能从多维角度创建档案信息节点,并且能够进行自由链接。同时还可以进行修改、删除、追加节点、链等操作,并能方便地把创建的档案信息节点及链接节点的链,加入到数据库中。第二,支持对数据库的多角度浏览与精确化检索和模糊检索。第三,具有初步智能化,能够做简单的推理,能按人们的思维习惯提供服务,如信息页面自由转换等。

对功能的实现,要求第一,在档案信息组织过程中采用超文本结构,按照人们的思维习惯,从多维角度精确地组织档案信息。第二,在超媒体数据库表现层(用户接口层)提供触发热标或按钮等便于用户自由选择查看内容的手段,自主地控制查阅过程,查询相关档案信息或进行重点演示,并提供在线帮助。

3.2 档案信息超媒体数据库模型构建

采用三层结构模型来构建档案信息超媒体数据库(图1)。在该模型中,节点是档案信息存储的基本单位,可以存放文字、图形、声音、视频、多媒体信息等。在档案信息组织过程中,通过链把档案信息节点链接成页,即相关档案信息的集合,并通过各种交叉关系形成一个网状结构。

(1) 存储层:由分立的媒体信息文件构成。这些媒体信息文件可以是单媒体文件,也可以是多媒体

文件。它们构成档案信息超媒体数据库的基本素材,是基础层。档案信息,无论是单媒体形式还是多媒体形式,都以节点的形式存储在该层。

(2)逻辑层:是存储层与表现层(即用户接口层)的中介。它对节点和链的基本特点起决定作用,它记录了节点之间链的关系及节点属性、链类型,是实现信息媒体输入输出格式标准化的最理想层次。这一层,在存储层的基础上,通过链接节点形成档案信息页,

档案信息页可以根据需要建立多层,另外还要建立节点与文本之间的引用关系,建立页与节点之间的交叉索引关系使档案信息形成网状结构^[5]。

(3)表现层——人机交互的窗口:是档案信息超媒体数据库的外层。这一层直接体现了档案信息组织的成果——档案信息数据库能否被有效利用,也直接体现了用户对系统的需求^[6]。这一层需要建立导航图和热标。用户借助导航图和热标,访问数据库。

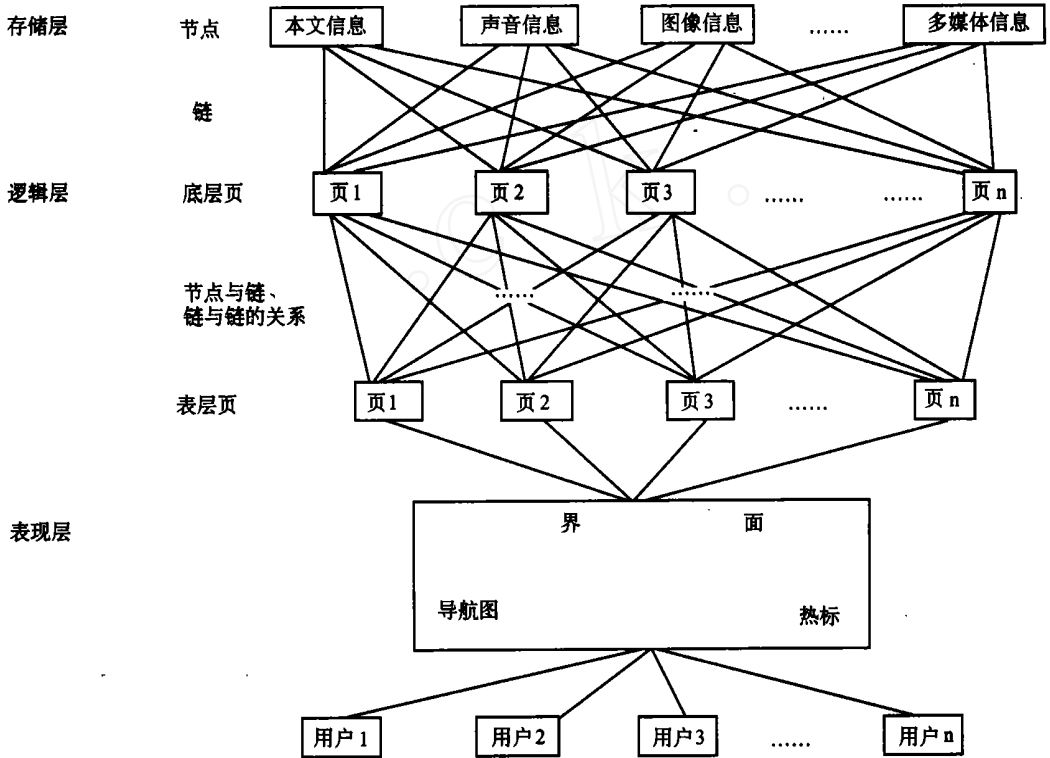


图 1 档案信息超媒体数据库模型

3.3 具体实现

(1)对档案信息进行分析。对档案信息进行组织,首先要分析档案信息的内容特征和外形特征。档案信息的内容特征,是指文件或案卷所包含的主题内容,在著录中主要体现在题名及提要项上;在主题标引中体现为相关的主题词。档案信息的外形特征,是指文件或案卷的责任者、产生时间、文本、载体、密级等特征。其次,对析出的内容特征和外形特征作出相应的标识,为下一步档案信息整理作准备。

(2)对档案信息进行整理,形成信息节点。在分析的基础上,通过采样、量化、编码等手段,对各种类型的档案信息进行处理,形成标准化的数字化信息。然后,对数字化的档案信息进行分类,利用多媒体编辑

工具,如 Animator、3Dstudio、Premiere 等,把文字、图形、图像、声音、动画等类型的档案信息转化成相应的信息节点。最后,对档案信息节点进行层次排列。

(3)对档案信息节点进行链接,是档案信息组织实现的最基本环节,它关系到档案信息组织的质量好坏。对档案信息节点进行链接,并不是无原则地链接,它要求根据档案信息的内在关联如主题等进行链接。链接过程分两步。第一步是对档案信息节点进行链接,形成一个个档案信息页,即相关档案信息的集合。第二步是档案信息页之间的链接。它可以根据需要,按照一定原则形成多层的页面关系。通过节点与页、页与页之间的链接,构成一个多层次的网状结构(如图 2)。

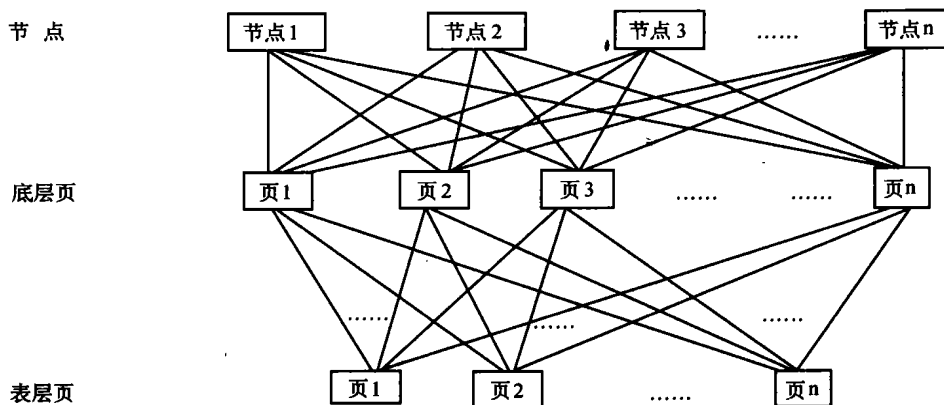


图2 档案信息节点的链接

(4) 把档案信息节点、链、页以及它们之间的关系存储到超媒体数据库中。基于文件(File-Based)的方法和二进制域(BLOB: Binary Large Object, 二进制大对象)方法是档案信息存入超媒体数据库的两种具有代表性的方法^[7]。

基于文件的方法主要是用于超媒体数据信息, 以外部文件的形式存储在硬盘、光盘或网络上。这一方法是将包含信息的档案文件的名称与路径存入数据库中。在访问数据时, 根据数据库记录中存储的名称和路径, 检索出文件定义并送到多媒体控制例程, 再由多媒体控制例程完成多媒体信息的显示或(和)播放工作。这种方法直接简便, 易实现, 并且其信息可被数据库管理系统之外的外部程序访问和使用。

二进制大对象方法是把超媒体数据直接存放于数据库中的BLOB字段中, 而不是存放在外部文件中的方法。这种技术将超媒体数据看作同文本(汉字、字母或数字)式的数据一样。使用BLOB字段, 所有数据都通过数据库表格存储在一起, 不必担心文件被删除或移动。BLOB字段有两种类型: Memo(备注)字段和OLE(对象链接和嵌入)字段。前者用于大文本对象, 后者用于大二进制对象, 如声音和图像等。

(5) 提供人机界面。提供人机界面, 是档案信息组织的“外围”环节。人机界面是人们利用档案信息的交互平台。它的好坏直接影响到人们利用档案信息的难易程度。这一步, 主要是制作导航图和热标来指导和实现相应档案信息页间的跳转^[8]。

导航图是整个系统的总体结构图, 目的在于给用户一个总体性的指导。本文在此拟用树状层次结构的导航图, 因为它既可以避免线性结构思维的局限, 又不会产生网状结构常导致的“迷路”弊端, 给用户一个总体的概念和清晰的思路。这一步, 可以按

页的基本顺序来实现和完成链接, 在总体上保持导航图的层次结构, 使其明了、清晰。

另外, 还要建立热标, 包括热字、热区、按钮等, 通过热标维持档案信息组织的非层次关系, 实现档案信息的交叉自由检索。

(6) 测试与维护。这是档案信息组织的辅助环节, 也是保障档案信息组织准确和完善的基本环节。档案信息经过组织后, 存放在超媒体数据库中, 难免有些错误或不完善的地方。对超媒体档案信息数据库进行测试是很有必要的。通过测试, 可以优化数据库的结构和功能。纠正档案信息组织中的错误, 保障档案信息数据库正常运行。

最后, 还要对超媒体档案信息数据库进行维护, 根据新的需要, 对档案信息进行再组织, 保障档案信息发挥其最大效用。

参考文献

- 1 李淑华, 朱颜军. 超媒体系统. 中国工程师, 1997(1)
- 2,3 林仪, 刘云生. 超媒体数据库的研究与设计. 计算机工程, 1997(2)
- 4 李国辉等. 分布式超媒体数据库系统的系统结构和组成. 中国图像图形学报, 1997(6)
- 5,8 刘瑞祥. 超媒体技术在文档管理中应用的研究. 中国矿业大学学报, 1998(3)
- 6 钱钢等. 超媒体数据库三层数据模型. 上海交通大学学报, 2000(7)
- 7 贾拥民. 析档案多媒体数据库及其建立方法. 山西档案, 1998(1)

王心裁 武汉大学信息资源研究中心副教授, 博士.
通讯地址: 武昌珞珈山. 邮编 430072.

吕元智 武汉大学信息管理学院研究生. 通讯地址同上.
(来稿时间: 2002-06-25)