

## Construction of extensible markup language-based MR brain imaging database of Chinese

LI Wei<sup>1</sup>, YANG Jin-zhu<sup>1</sup>, CHEN Nan<sup>2</sup>, WANG Xing<sup>2</sup>, ZHAO Da-zhe<sup>1\*</sup>, LI Kun-cheng<sup>2\*</sup>

(1. Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang 110179, China; 2. Department of Radiology, Xuanwu Hospital, Capital Medical University, Beijing 100053, China)

**[Abstract]** Medical image database is the important and effective means of medical image data retrieval and analysis in clinical diagnosis, education and research. In this study, a methodology of building extensible markup language (XML)-based medical image database was introduced, which supported medical image data modeling, storage, retrieval and mining. Finally, the establishment of China's normal brain MR image database had good application value. The MR brain image database system is extensible, reliable, and independent, and has high performance which can meet the requirement of image storage and retrieval.

**[Key words]** Database; Information storage and retrieval; Extensible markup language; Data model

## 基于可扩展标记语言技术构建中国人大脑 MR 影像库

栗伟<sup>1</sup>, 杨金柱<sup>1</sup>, 陈楠<sup>2</sup>, 王星<sup>2</sup>, 赵大哲<sup>1\*</sup>, 李坤成<sup>2\*</sup>

(1. 东北大学医学影像计算教育部重点实验室, 辽宁 沈阳 100179;  
2. 首都医科大学宣武医院放射科, 北京 100053)

**[摘要]** 医学影像数据库及信息系统是医学影像数据有效检索与分析的重要手段, 是医学临床、教育与研究的基础。本文提出一种基于可扩展标记语言(XML)的医学影像数据库及信息系统的构建方法, 实现中国正常人脑 MR 影像库及信息系统, 该系统支持医学影像数据的有效组织、存储、检索和挖掘。临床实际应用证明该系统具有可扩展性强、结构稳定、功能独立、存取效率高等特点, 满足人脑 MR 影像数据存储、检索和处理分析的需求, 具有良好的应用价值。

**[关键词]** 数据库; 信息存储与检索; 可扩展标记语言; 数据模型

**[中图分类号]** R445.2; TP312 **[文献标识码]** A **[文章编号]** 1003-3289(2010)09-1640-04

随着医学成像技术的快速发展与应用, 医学影像在临床诊断和医学研究中占据重要地位。医疗机构每天都产生大量医学影像数据, 如何对这些数据进行有效地管理和维护, 充分利用和挖掘其中的重要辅助信息, 以支持深入的医学研究, 是亟待解决的重要课题<sup>[1]</sup>。医学影像库是建立在图像数据库、图像处理、计算机网络技术及医学领域知识基础上, 支持医学影像数据有效组织、存储、检索和挖掘的数据库技术, 能使现

有的影像资源得到充分利用<sup>[2]</sup>, 其发展将促进临床诊断、医学教育和科研工作的进步。

近年来, 面向医学影像特定领域的数据库技术的研发倍受业界重视, 一些研发成果已逐渐向应用于临床。国外比较具有代表性的系统有北美放射学会的 MIRC<sup>[3]</sup>、德国 Aachen 大学开发的 IRMA<sup>[4]</sup> 和 DePaul 大学开发的 BRISC<sup>[5]</sup>。MIRC 是建立在 Internet 基础上的医学信息库, 支持多个影像库的协作。IRMA 存放多种设备的影像, 定义了通用的基于医学影像内容检索系统的流程。BRISC 是肺部 CT 结节影像检索库, 为临床提供计算机辅助诊断。此外还有病理学者 Monarchic 首先开创的肿瘤影像数据库、美国卡耐基梅隆大学的影像数据库 (Carnegie Melon Image Database) 等。国内张凯等<sup>[6]</sup> 研究了一种基于 DICOM 的医学影像数据库系统; 岳文军等<sup>[7]</sup> 研究了基于 Internet 的医学影像成像教学互动平台。然而目前医学影像库大多在现有的成熟的关系数据库或扩展关系数据库基础之上管理医学影像数据, 尚无统一的抽象

**[基金项目]** 国家 863 计划项目 (2006AA02z391)、国家 973 计划项目 (2005CB522800、2004CB318101)、国家自然科学基金 (30621004)。

**[作者简介]** 栗伟 (1980—), 男, 河南驻马店人, 在读博士。研究方向: 医学影像检索与挖掘。E-mail: l-w@neusoft.com

**[通讯作者]** 赵大哲, 东北大学医学影像计算教育部重点实验室, 110179。E-mail: zhaodz@neusoft.com

李坤成, 首都医科大学宣武医院放射科, 100053。

E-mail: likuncheng1955@yahoo.com.cn

**[收稿日期]** 2010-06-23 **[修回日期]** 2010-07-14

数据存储模型和交换模型,不具备系统的数据通用性和功能扩展性。针对这些问题,可扩展标记语言(extensible markup language, XML)技术越来越得到研究人员的重视。XML 是全球信息网联盟(World Wide Web Consortium, W3C)1998 年制定的标准,是标准通用标记语言(standard generalized markup language, SGML)的一个子集,是一种元语言,不仅适合数据的表示和描述,也适合作为数据交换的中介<sup>[8]</sup>。本文提出一种基于 XML 的医学影像库构建方法,支持医学影像数据的有效组织、存储、检索和挖掘,建立中国正常人脑 MR 影像库,具有良好的应用价值。

1 医学影像库数据模型

数据模型通常由数据结构、数据操作和完整性约束三部分组成。数据结构是所研究对象类型的集合;数据操作是对系统动态特性的描述;数据完整性约束是一组完整性规则的集合,规定数据库状态及状态变化所应满足的条件,以保证数据的正确性、有效性和相容性<sup>[9]</sup>。本部分从上述三个方面描述医学影像数据库模型。

1.1 数据结构 针对医学影像本身、临床和研究中应用的特点,本文提出基于案例的医学影像数据模型。该数据模型采用 XML 进行结构描述和存储,每个案例包含若干个信息单元,称为案例模块。在临床和科研中,与医学影像相关的数据还有患者人口信息、临床检查、实验室检验、辅助检查报告以及影像研究中的测量参数等数据,这些数据结构复杂,变更频繁。每个独立的信息作为案例中的一个模块,每个模块具有独立的结构。

案例中存储的数据类型包括结构化和非结构化数据,增强了用户组织、存取数据的灵活性。医学影像库中案例编号是惟一的,每个案例都包含一个头(head)结构,描述该案例的编号、创建者、创建日期、权限等信息;还包括若干个信息模块(module),每个信息模块根据信息内容组织相应的结构。

1.2 数据操作 医学影像库中数据操作包括案例增加、案例删除、案例更新和案例查询,案例模块操作包括模块增加、模块删除、模块更新和模块版本管理。①案例增加:系统自动创建案例的编号、创建时间和创建用户,用户可设置案例访问权限。用户可从模块定义库中选择案例所包含的模块,也可根据

需要后续不断增加新模块。②案例删除:系统根据案例编号删除该案例包括的所有信息,包括案例头信息、案例模块信息(包括各个版本的模块信息)、模块查询信息和索引信息。③案例更新:系统根据案例编号和新的内容来更新案例的头信息。④案例查询:包括元数据查询、影像查询和文本查询;元数据查询包括患者的人口统计信息、影像元数据的查询;影像查询指基于内容的影像查询,是基于影像特征的相似性查询,提供范例影像作为输入;文本查询是基于诊断报告等非结构化文本的查询,提供关键字作

为输入。⑤模块增加:系统根据用户添加的模块类型,自动添加到案例的存储空间中,并根据索引内容,更新索引模块中的数据以备查询。⑥模块删除:系统根据用户删除的模块名称,将模块所有版本从案例存储空间中移除,并更新索引中相关数据。⑦模块更新:根据模块修改的内容,更新当前存储库中最新版本模块的信息,并更新索引中数据。⑧模块版本管理:模块结构的变化会引起模块版本的变更,模块可以有多个版本,每个版本都有对应的版本定义,系统更新和查询都是以最新版本模块的数据内容为基准。

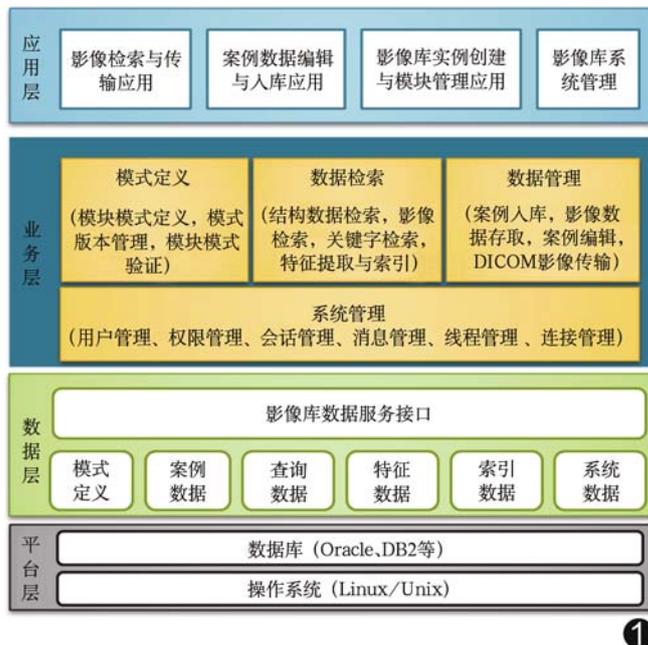


图 1 医学影像库体系结构

1.3 数据完整性约束 案例数据的完整性是影像库数据管理的重要组成部分。本研究采用 XML Schema 来约束案例数据的合法性和正确性,主要包括模块模式定义,模块模式验证和模式版本管理。①模块模式定义:案例中的每个模块的每个版本都对应惟一的模式定义规范,同时根据用户实际的

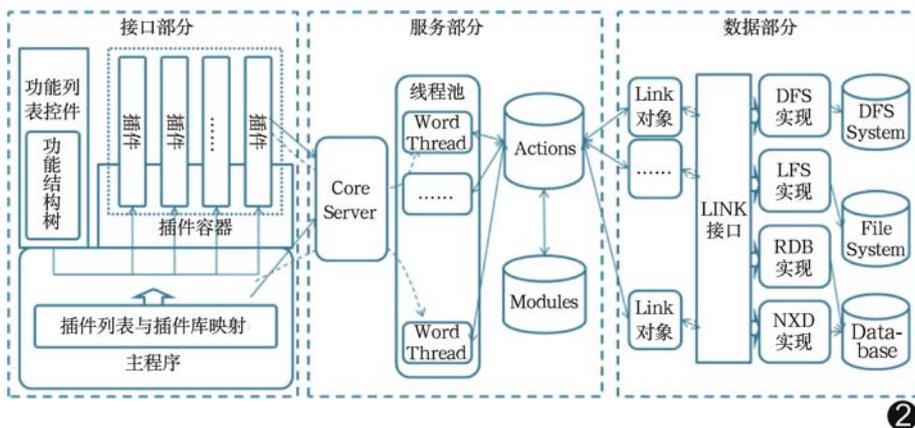


图 2 医学影像库逻辑架构

需求,系统提供新案例模块的模式定义,用户可根据新定义的模块模式为相应的案例添加新的案例模块。②模块模式验证:案例模块的增加、更新之前,系统都会对该模块数据进行模式验证,对不合法的数据系统会提示用户。③模式版本管理:案例的每一个模块都对应该模块若干个版本的一个模式版本,模块增加和更新都必须指明遵循对应的模式定义版本。

### 2 医学影像库系统架构与设计

医学影像库采用分层的系统结构,构建组件化的医学影像库业务处理模块,提供了规范、灵活的开发和用户接口,满足接口灵活、信息共享、标准规范、安全高效的要求。

#### 2.1 医学影像库系统体系架构

自下而上分为四层,即系统平台层、数据服务层、业务服务层和系统应用层(图 1)。

针对系统四层内容分别描述如下:①系统平台层:由操作系统、数据库管理库和网络管理系统等组成,支持分布式、客户机/服务器计算能力。②数据服务层:提供统一的数据支撑服务接口和服务,包括数据模式定义、案例数据、查询数据、特征数据、索引数据和系统数据等数据的存储和管理。③业务服务层:提供支持影像库操作的核心服务,包括模式定义、数据检索、数据管理和系统管理等服务。④系统应用层:根据用户需求向其提供相应服务,如影像数据检索应用、影像传输应用、影像数据管理应用、影像库管理应用等。

#### 2.2 医学影像库系统逻辑架构

分为三部分:用户接口、服务处理和数据服务(图 2)。

影像库逻辑架构的三部分功能和采用的技术包括:①接口部分采用了基于插件的服务接口结构,根据用户的权限系统自动加载对应的功能插件,保证系统数据操作的安全性。每个功能插件通过服务接口和服务端进行交互。②服务部分采用线程池的服务处理方式,每个用户请求服务端都对应一个处理活动(Action),并对业务逻辑处理映射到相应的业务模型。③数据部分采用了数据连接池技术和接口技术,给上层提供统一访问接口,对各种物理存储介质设计相应的操作接口,包括文件系统、关系数据库、原生 XML 库、分布文件系统等。

### 3 正常人脑 MR 影像库系统实现

基于上述影像库架构和设计,本文建立了正常人脑 MR 影像库系统,以验证该医学影像库构建方法的合理性和有效性。正常人脑 MR 影像库的影像数据来自全国 15 家医院约 3000 名志愿者的头部 MR 扫描数据。正常人脑 MR 影像库为用户提供影像案例数据入库、数据检索与下载和数据保存三个方面的应用支持。

3.1 数据入库应用 所有 MR 扫描原始图像按 DICOM 标准组织后得到,包含 DICOM 文件头和至少一个图像数据的集合。每个志愿者案例包括人口统计模块、临床检查模块、影

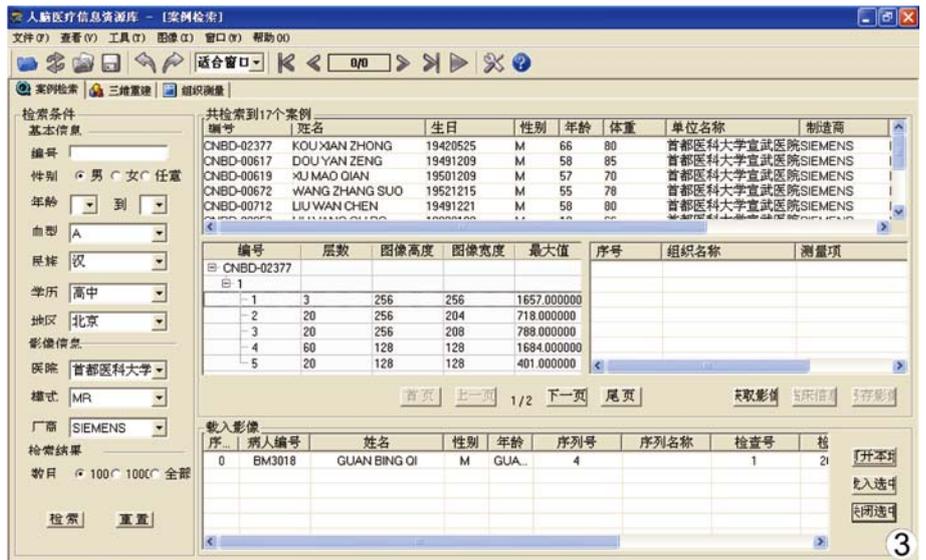


图 3 影像检索与下载操作界面

图 4 脑组织测量标定结果存储界面

像检查模块和测量标记模块四个模块,其中人口统计模块记录了志愿者的人口统计数据,包括年龄、性别、民族等;临床检查模块记录了志愿者的临床检查数据,包括耳鼻喉科、眼科、外科和实验室检查等;影像检查模块记录了 MR 图像属性数据,包括宽高、层间距、层厚等;测量标记模块记录了脑部组织分割测量参数,包括脑部组织的体积、影像组织标定等信息。系统通过案例编辑和入库程序完成 3000 名志愿者数据编辑和入库,其中影像数据通过 DICOM 协议传输标准存入服务器端。

3.2 影像检索与下载应用 正常人脑 MR 影像库为用户提供了结构数据检索,检索条件包括:案例编号、性别、年龄范围、地区、医院、模式等。用户操作界面见图 3,用户对检索结果可以分批显示,系统自动根据用户选中的结果进行该案例影像的序列信息显示,显示的信息包括层厚、层间距等影像信息,同时能够显示该案例中组织测量的结果,包括组织名称、测量项和值。用户选中其中一个影像序列,影像获取操作可以使系统自动将服务器端的影像下载到本地,供研究人员进行脑部组织的测量和标定。

3.3 脑组织测量标定存储应用 研究人员根据多种脑部组织分割算法将 MR 影像中组织区域标定出来,并根据分割结果统计该组织的表面积、体积、长度等测量参数。用户利用如图 4 的操作界面接口将测量标定结果更新到正常人脑 MR 影像库中对应案例的测量标定模块。同时,用户也可以直接统计出影像库中存储的各种测量指标,挖掘影像库中新信息,如通过测量得到杏仁核体积的正常参考值是相关神经科学研究和临床应用的基础<sup>[10]</sup>。对组织标定结果进行训练和分析,进而建立正常人的脑部组织模板供临床使用。

#### 4 讨论与分析

本研究提出的医学影像库架构在 Linux 系统平台之上,保证服务器高效稳定地不间断运行,服务端的用户访问接口给客户端的开发部署提供了极大的灵活性。本文提出的医学

影像库的构建方法为中国正常人脑 MR 影像库系统提供了较高的系统扩展性,有效减少了数据占有空间和数据冗余,提高了数据处理的速度和效率,经临床实际应用证明,该医学影像库具有可扩展性、结构稳定性、功能独立性、存取效率高等特点,具有良好的科研应用价值。

#### [参考文献]

- [1] 朱歆华,赵大哲,于亚新,等.基于本体的医学资源库系统的设计与实现.东北大学学报:自然科学版,2007,28(1):26-30.
- [2] 杨焯.基于 DICOM 标准的医学影像数据库的建立.长春理工大学计算机科学与理论硕士论文,2008:2-4.
- [3] Prevedello LM, Andriole KP, Khorasani RR. Informatics in radiology: integration of the medical imaging resource center into a teaching hospital network to allow single sign-on access. Radiographics, 2009,29(4):973-979.
- [4] Lehmann TM, Güld MO, Thies C, et al. Content-based image retrieval in medical applications. Methods Inf Med, 2004, 43(4): 354-361.
- [5] Lam MO, Disney T, Raicu DS, et al. BRISC—an open source pulmonary nodule image retrieval framework. J Digit Imaging, 2007, 20(Suppl 1):63-71.
- [6] 张凯,吕扬生.基于 DICOM 标准的医学图像数据库.中国生物医学工程学报,2002,21(6):548-551.
- [7] 岳文军,蹇朴,李春平,等.医学影像成像教学互动平台 CAI 的研究与开发.医学影像学杂志,2003,13(3):202-204.
- [8] 王静.XML 路径查询处理关键技术研究.中国科学院研究生院计算技术研究所博士论文,2003:1-5.
- [9] 王珊,萨师焯.数据库系统概论.北京:高等教育出版社,2006:12-14.
- [10] 潘克棣,陈楠,王星,等.两种测量软件在杏仁核体积测量中的比较.中国医学影像技术,2009,25(5):890-893.

## 三线表的规范格式

▲表序和表题:表序即表格的序号,一篇论文中如只有 1 个表格,则表序编为表 1,表题即表格的名称,应准确得体并能确切反映表格的特定内容且简短精练。

▲项目栏:指表格顶线与栏目线之间的部分,栏目是该栏的名称,反映了表身中该栏信息的特征或属性。

▲表身:三线表内底线以上,栏目线以下的部分叫做表身,是表格的主体表身内的数字一般不带单位,百分数也不带百分号,均归并在栏目中表身中不应有空项,如确系无数字的栏,应区别情况对待,在表注中简要说明,不能轻易写“0”或画“—”线等填空,因可代表阴性反应“0”代表实测结果为零。

▲表注:必要时,应将表中的符号标记代码,以及需要说明的事项,以最简练的文字,横排于表题下作为表注也可附注于表下。