

电子病历存储模式研究^{*}

宇文姝丽

杨小平

(河北大学管理学院 保定 071002) (中国人民大学信息学院 北京 100082)

〔摘要〕 在分析传统电子病历存储方式的基础上, 根据 HL7 CDA 提出细颗粒度结构化的 XML 模式存储方案, 不仅可以有效降低病历数据的存储空间和查询时间, 同时也可以方便临床医生和科研工作者将病历数据应用于医学信息统计、临床辅助诊断等其他领域, 从而拓展电子病历应用空间。

〔关键词〕 电子病历; HL7 CDA; 存储模式

Research on the Electronic Medical Records Storage Pattern YUWEN Shu-li, School of Management, Hebei University, Baoding 071002, China; YANG Xiao-ping, School of Information, Renmin University, Beijing 100082, China

〔Abstract〕 Based on the analysis of the traditional storage pattern of electronic medical records (EMR) data, the paper proposes another storage plan with fine-grained structure of XML mode according to the HL7 CDA. This storage plan can not only reduce the medical records data storage space and the query time effectively, but also facilitate the clinicians and researchers to use medical records data for medical information statistics, clinical auxiliary diagnosis and other areas, so as to expand the application spaces of the EMR.

〔Keywords〕 Electronic medical records (EMR); Health Level 7 Clinical Document Architecture; Storage pattern

1 引言

当前数字化医疗与医疗信息化成为医疗卫生健康领域的重要发展方向, 医学领域得以快速发展的同时, 医疗卫生服务体系也面临很多严峻的问题, 因此医院不同科室之间、不同医院之间、医院和其他的医疗机构之间的信息共享使电子病历开始成为研究重点。在电子病历及其相关技术的研究和应用

领域中, 病历数据的存储问题成为影响电子病历快速发展的瓶颈: 电子病历系统要求病人的信息长期保存、随时可以获取。在传统的医院信息系统中, 病人信息是以关系数据表的形式进行存储, 这些信息的存储管理是以支持日常业务管理为目的, 因此基于数据库效率和容量管理方面的考虑, 不可能长期将历史信息联机保存。如何使病人信息在脱离数据库后仍然能够支持以病人为中心的结构随时获取, 是病历存储要解决的重要问题。

国际医疗组织 Health Level 7 (HL7) 在其应用研讨交流会上提出了基于 XML 语言开发电子病历的观点^[1-2]。本文采用 HL7 标准的 CDA 对病历数据进行分类结构化, 利用 XML 模式有效地解决对病历信息的描述、存储、处理等问题, 最终将病人的信息用关系表和 XML 文档相结合的形式保存下来用于信息处理。通过实验验证了这种混合存储方式比当前单纯的关系数据表存储方式可以更好地服

〔收稿日期〕 2010-11-09

〔作者简介〕 宇文姝丽, 讲师, 博士研究生, 主要研究方向为电子医疗、信息系统工程, 发表论文 10 余篇, 参编论著 3 部; 杨小平, 博士生导师。

〔基金项目〕 国家社科基金 (项目编号: 10BTQ039); 河北省社会科学发展研究课题 (项目编号: 201004020)。

务于患者,能为医生提供及时、准确的信息,同时也有助于临床的科学研究。

2 相关研究

2.1 国内外研究现状

2.1.1 国外 医疗数据的存储最早始于 20 世纪 60 年代末到 70 年代初的美国基于麻省总医院多用途程序设计系统 (Massachusetts General Hospital Utility Multi-programming System, MUMPS) 构架的医疗数据库,随后出现的关系数据库以其强大的结构化数据的处理能力成为医院早期以管理和收费为主要目的的医院信息系统。随着医院信息系统向临床的转变,越来越复杂的数据关系和数据类型需要存储和处理,电子病历的设计思想开始向面向对象转变,如 HL7 设计的 RIM 对象和 PREN13606 的信息对象模型。典型的有英国 Whittinongton 医院心血管科和伦敦大学共同开发的用于医疗的面向对象数据库 Objectstore^[3],它在给数据库应用程序开发带来方便的同时也限制了数据操作的灵活性。因此在面向对象数据库存储病历数据的同时,鉴于关系数据库的成熟技术和医院信息系统集成的需要,在关系数据库厂商提供扩充的大对象存储或数据拆分等功能的情况下,依然采用关系数据库进行病历数据存储。在美国病历数据处理领域常用的还有 Post-relational 数据库 Caché,它是基于 MUMPS 的面向对象的医疗数据库,该数据库在电子病历数据的处理速度上具有明显的优势,是一种比较有潜力的数据库技术。随着网络技术和信息交互的广泛应用,XML 的出现给电子病历数据存储带来了新的生机,并引发了原生 XML 数据库 (Native XML Database, NXD) 的迅速发展^[4],DB2, Oracle 等相继发布商业版本宣布对 XML 的支持,可以把标准的 XML 文档作为一个基本的单元来进行存储,而对底层的物理存储模型不再加以限制。同时国际医疗组织也相应提出了基于 XML 开发电子病历的观点,并给出了相关的技术支持框架 HL7 临床文档架构 (Clinical Document Architecture, CDA)^[2]。

2.1.2 国内 我国医院信息系统迅速发展始于 20

世纪 90 年代,统计显示几乎所有的 HIS 都是基于关系数据库的。电子病历技术兴起于 2000 年左右,临床和科研界仅仅是对电子病历进行局部性的实验性电子病历工作,即只是将传统的纸病历利用相关技术把部分医疗数据库完成电子化。最早的是由总后卫生部主持开发的军卫一号,然后相继出现了四川银星、北京安博维、北京中科奥、南京海泰、北京东华等从事医院信息系统和电子病历开发的公司,除了安博维和东华采用了 Caché 数据库外,其他的都还是使用关系数据库进行存储。

2.2 病历数据存储方式

XML 是一种可以定义描述对象结构的结构化标识语言,它能够有效地解决对电子病历的描述、存储、处理等问题。同时由于 XML 文件格式非常灵活,很难直接存入关系数据库的字段中,因此如何存储 XML 类型的病历数据成为当前病历研究的一个热点。

2.2.1 将 XML 数据作为长文本或者大对象存储 这种存储方式只是存储 XML 数据,失去了 XML 原有的语义和结构化的嵌套层次信息,当需要访问 XML 数据时需要对长文本或大对象的内容通过 DOM 或 SAX 方式进行解析,会影响系统效率和查询的灵活性。另一方面也丢失了 XML 自身的一些信息,后续对 XML 的访问完全依赖于开发人员编写代码,无论是使用 DOM 还是 SAX 解析对开发人员来说都是一种负担,最重要的是数据库的性能无法保证。

2.2.2 将 XML 数据拆分成关系数据表 XML 拆分存储方式简单地说就是把 XML 文档分解成多个具有关联关系的二维表来描述^[4-6]。实际上任何具有一对多关联的实体,在 XML 文档中仅仅用一个嵌套层次就可以解决,映射到关系数据库中却是两个具有主外键关系的表。这样对于一个很简单的 XML 文档可能需要大量的具有关联关系的二维表来表示,当对 XML 文档进行查询时需要将这些具有关联关系的表连接起来再检索,影响了检索效率。XML 文档原本只需要通过对 XML 节点的调整就能很好地适应需求的变化,而这种拆分存储方式面对

需求变化实现起来非常被动，常常需要改变关系表的结构，代价太高，因为任何开发人员都不愿意在项目实施中或实施后再去调整底层的数据表结构。

2.2.3 原生 XML 数据库 原生 XML 数据库是以标准 XML 格式存储信息的数据库，不需要进行数据的转换，可以发挥 XML 的优势。在功能方面与关系数据库一样，原生 XML 数据库产品也支持安全、并发访问、查询、检索、增删改、事务管理的编程接口，区别于关系数据库的关系模型的是原生 XML 数据库内部数据模型是 XML，这种特殊的数据库大大提升了 XML 的可管理性和可操作性。但是企业对这种非主流的、尚未经过业界验证的数据库的性能的质疑和不认可态度令病历开发厂商望而止步，并且这种原生 XML 数据库基本上不支持关系型数据，而目前关系数据库已渗透到企业的各个方面，这使得原生 XML 数据库应用受限。

2.3 HL7 CDA 简介

2.3.1 结构框架 CDA^[2,7-8] 是 HL7 标准的一部分，是以临床文档交换为目的的描述临床文档的结构和语义的文档标记。CDA 本身不对文档内容建模，只是对需要交换的临床文档的结构和语义制定了标准。一个 CDA 文档由一个文档头和一个文档体组成，文档头确定了文档的分类，提供了鉴定信息、受访数据、患者、提供者等基本信息；而文档体则用来描述临床报告，由可嵌套的内容集组成。完整的 CDA 文档就是一套分层级的文档规范，而这个层级就形成“文档结构”，具体结构框架，见图 1。

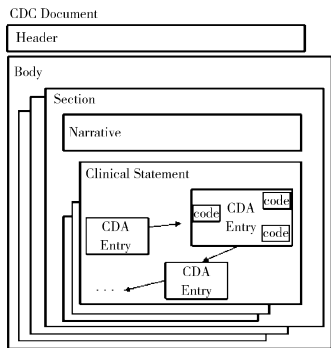


图 1 CDA 结构

2.3.2 结构化层次 CDA 的文档结构定义了一个可扩展分层次的文档规范集，它是一个为了表示临床文档而对其语义和结构进行约束的定义的集合。在 CDA 结构中，文档规范之间的关系是通过特殊化和扩展来定义的。使用该结构可以消除医疗局部文档中遇到的强迫接受的文档约束的情况，以便实现互操作。文档的层次同样是一整套大量的规范，可以用来进一步实现约束。CDA 文档根据其内容可以用于计算机处理的程度不同和对内容进行的约束不同分为 3 个层次，见表 1。

表 1 CDA 文档结构化层次

层次	结构化粒度
第 1 层	无约束的 CDA 框架
第 2 层	Section 级约束的 CDA 结构
第 3 层	Entry 级约束的 CDA 结构

CDA 文档的第 1 个层次面向高结构化颗粒度，文档体中包含了大量的结构信息，虽然可以使用 HL7 提供的受控词汇表，但是没有语义信息的表达；第 2 个层次是对第 1 个层次的特殊化，在第 2 个层次 CDA 开始对文档的内容和结构通过模板的形式结构化，并包含了一套严格的 CDA 结构和语义信息；第 3 个层次则是第 2 个层次的进一步扩展，在第 2 个层次的基础上增加了另外的标记，可以使医疗临床信息在 HL7 定义的范围内被完整地表示，而从数据共享和集成的角度对更细粒度的文档内容进行结构化。需要说明的是在不同的结构层次下，临床文档内容是不变的，而每个不同的层次使用计算机处理时共享的语义也有所不同。目前国内外对于数据病历结构化的粒度集中在第 2 个层次，对于第 3 个层次的结构化粒度不同的国家、不同的系统给出了自己不同的理解和实现方式，多以结构化数据和非结构化数据叙述性描述并存的形式来出现。

3 存储方案^[9-11]

3.1 病历数据结构化

3.1.1 建立多维信息结构 从前面 CDA 的介绍中可以看出，即便采用 CDA 第 3 个层次来实现病历数据结构化，其粒度还是比较粗略的，因此要想在实

际医疗系统中加以应用，需要再加以具体的细化。根据纸质病历的 SOAP（Subjective data, Objective data, Assessment, Plan）描述方式，对当前 HIS 系统的改进引入基本数据、组合数据、组件和模板来对病历数据进行结构化的分割和组合，建立起一个多级的灵活多维信息结构。具体定义如下：基本数据是病历信息中不可分割的单元，如白细胞、红细胞等；组合数据是临床中经常使用的具有特定业务含义的基本数据集合，如某个诊断结果片段，某项检查片段等；组件是组合多个基本数据和组合数据形成的具有业务含义的临床数据，用于存储病人常规信息、体检信息、入院诊断等信息；模板是由医疗单位的业务人员事先建立好的不同科室病人的病历样式。对于组件来说，各个科室的医生可以选择相应的模版以填写病历信息，动态地实现各个组件组合。这样一套基于 XML 的病历数据存储模式体系是完全面向业务人员设计的，医生能够动态地定义语义明确的病历模版并可以自行调整。

3.1.1.2 模式设计分解细化规则

鉴于上述病历数据的结构化思想,依据 CDA XML 模式的层次结构以及病历数据的更新频率,将一个病人的完整病历信息分解为基本信息、现诊疗过程、既往史(包括过去史、个人史和家族史)、过往病历 4 个部分,除了基本信息相对独立外,其他 3 项又可以进一步细化,具体的模式设计分解细化规则如下:将病历文档作为根节点,包含所有的病历信息,即由它决定病历的模板;将病历文档进行细化,分成各个组件,每个组件代表一个独立的项目,如病人基本信息、现诊疗过程、既往史、过往病历等;将组件进一步细化成更小的部分成为组合数据,它们之间可以相互组合形成更大的组件,如主诉、现病史、各项检查项目等;将组合数据元素进一步细化成最小的单元,成为基本数据,它是不可再分割的数据单元,也就是病历信息树的叶节点,如病人姓名、现病史的起病时间等。

3.2 病历数据存储方案

3.2.1 混合存储模式 电子病历数据的存储方式将直接关系到电子病历处理的效率和复杂度,而电

子病历数据的复杂性和可扩展性使得上述 3 种存储方式都有所欠缺。为了提高系统的效率,降低数据冗余,本文采用混合关系数据表和 XML 数据存储模型来实现病历数据的存取:一方面对于数据结构化程度比较高,结构不太复杂而且相对固定的部分采用关系数据表的形式存储;另一方面对于非结构化程度比较高,数据比较复杂而且结构经常变动的部分,采用 XML 的半结构化数据存储模式,便于实现病历系统跨平台查询功能和不同医疗系统之间的数据交换。两种描述同时存在并保持相对独立,依用途不同相互配合使用。

3.2.2 XML 基本存储模式 使用 CDA XML 描述病历数据内容,首先要定义模式结构 XML Schema^[12-13],对病历数据进行相关的约束,主要包括:可以出现的子元素、元素的属性、子元素出现的顺序、数量、数据类型等,这些约束在图形化的模式设计中都有标识。例如数据类型显示在图表的左上角,用多个矩形框表示该元素可以在 XML 文档中出现多次,用虚线框表示其在 XML 文档中可选择有无等等。根据病历数据的结构化方法,对病历数据按照结构化粒度从大到小的方法逐步组合而成基本的存储模式,见图 2。

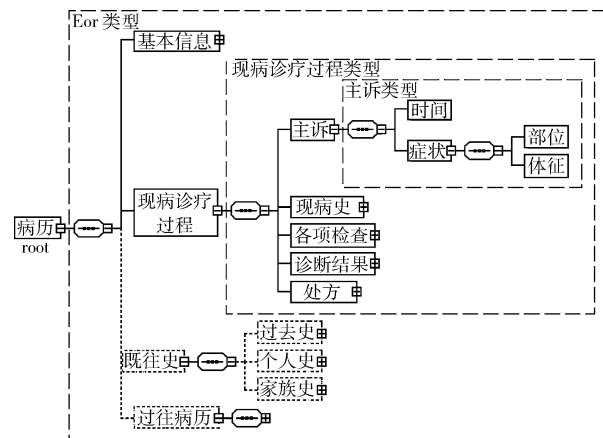


图2 病历数据 XML 模式截图

每个基本数据作为一个属性或者子元素，如姓名、性别等。为了数据处理的方便，尽量作为子元素处理。组合数据和组件一般定义为由多个属性或子元素组成的复杂数据类型，也可以引用其他子元素或者复杂数据类型。模板就是一个完整的 XML

Schema, 它用根节点以说明此文档的基本信息, 一个 Schema 表示一个患者的一次诊疗过程。

3.2.3 对不同内容按需采取不同存储方式 依据临床数据实践, 病人的基本信息相对来说变化较小, 数据的结构化程度比较高, 而且涉及到病历系统的各个部分, 对于这些需要经常与其他系统交互的信息采用关系数据库表的形式存储, 以提高系统的效率。对于像过去史、个人史和家族史这部分变化不大结构化程度比较高的数据, 由于个人情况的不同, 需要存储的数据也会有所不同, 采用关系数据库表会加大病历数据之间交互的复杂性, 同时也会增加系统的冗余和存储空间, 故而把它们集中起来使用既往史这个组件采用 XML 模式存取。相对来说病人的每一次诊治过程是变化最大的, 同时也是最重要的, 因此系统主要是针对病人每次的诊治过程作独立的 XML 模式存储。例如门诊病人就医的过程一般为挂号、分诊、就诊、一般检查、实验室检查、检查结果反馈、诊断、开具处方等, 根据病历规范手册这些过程对于单个病人来讲有些是可有可无的, 这样就需要医生动态地调整该 XML 模式以适应需求。

4 实验设计

4.1 实验方法

以上分析结果表明关系数据库表和原生 XML 数据库相结合的方式是病历存储的一种更好的方式。针对支持原生 XML 存储的数据库 DB2 V9, 设计了两种存储方法, 一种是把病历数据拆分成关系数据库表的形式进行存储, 另一种是采用关系数据库表和 XML 模式相结合的方式存储。为了简化实验均使用 100 份门诊病历, 选取以下 3 组典型的数据库操作分别在两种存储方式上运行以比较其性能: 插入病人病历; 使用简单查询规则对单个病人进行查询, 例如姓名; 使用多条件查询规则对病人信息进行查询, 例如姓名 + 时间, 姓名 + 科室, 姓名 + 医生姓名等。

4.2 插入操作

对于病历的插入操作, 分别测试了一次性插入 1 个、5 个、10 个、20 个病人病历的情况, 发现随着插入的量的增加混合数据存储方式明显地优于单纯的关系数据库表的操作, 见图 3。这是由于单纯的数据库表的存储方式需要建立各个表之间的连接, 从而增加了数据的冗余, 而混合存储方式只需要根据 XML 模式插入内容即可, 减少了不必要的操作。

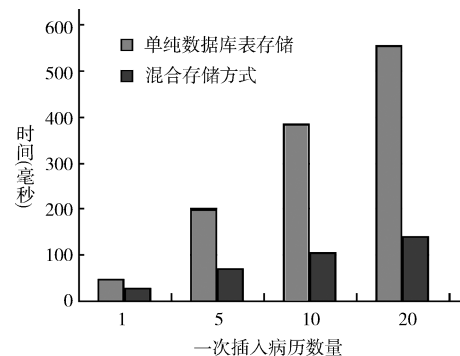


图 3 插入操作比较结果

4.3 查询操作

电子病历在很大的程度上主要是用于后期的查询统计工作, 所以设计了 5 个查询条件进行查询: Query 1 是最简单的病人基本信息的查询, Query 2 是对病人所有病历主要信息的查询, Query 3 则是根据时间 (或某一时间段) 对病人病历的查询, Query 4 是根据病人和科室信息的组合查询, Query 5 是根据病人和医生信息的组合查询。查询结果的比较, 见图 4。对于都是用关系数据库表存储的病人基本信息的简单查询来说, 两种存储方案的查询速度基本一致, 但是当查询涉及到病人病历数据信息时混合存储方案就会明显地优于纯关系数据库表存储方法, 这是因为将层次化的病历数据分解到关系表, 关系查询必然会有多个连接操作来实现, 过多的数据库表的连接势必会影响系统的效率。Query 4 和 Query 5 同为组合查询但结果相差太多, 原因在于医院科室信息测试数据直接列入病人病历中, 而医生信息则是独立成表。

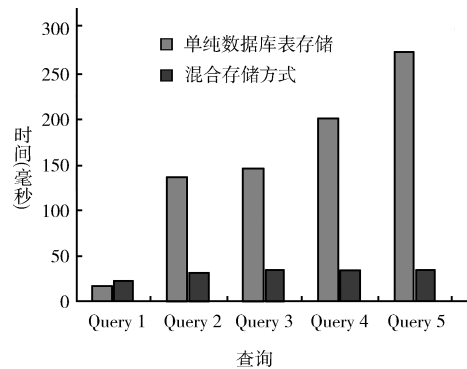


图 4 查询操作比较结果

5 结语

本文根据国际医疗标准 HL7 CDA 设计了一个集关系数据库和原生 XML 数据库于一体的混合病历数据存储方案,设计层次上由于 XML Schema 的使用减少了数据表的数量而使得设计比较简洁,应用上从医院提取了 100 个门诊病历进行了测试,测试结果表明该存储方案明显地优于传统的纯关系数据库表形式的存储方法。尤其是电子病历数据多以后期的统计分析为主,混合存储的方式有利于提高系统的查询速度使得该存储方案的优势更加明显^[14-15]。当然目前系统仅是测试了门诊病历部分,而且数据量比较小,使得测试结果还有一定的误差,后期将使用住院病历部分对该存储方案加以完善,并开展多用户的使用测试。

参考文献

- 1 HL7 Organization 著,梁铭会,俞汝龙主译. 医疗健康信息传输与交换标准 v2.4 [M]. 北京:北京大学医学出版社,2007.
- 2 HL7 Organization. HL7 Vision 3 [EB/OL]. [2010-10-31]. <http://www.hl7.org/v3ballot/html/welcome/environment/index.htm>.
- 3 Jeffrey L S, Jeffrey A L, Matvey B P. "Smart Forms" in an Electronic Medical Record: documentation - based clinical

decision support to improve disease management [J]. Journal of the American Medical Informatics Association, 2008, 15 (4): 513 - 523.

- 4 Sudhanshu S, Kunal V, John A M, etc. Designing a High - performance Database Engine for the 'Db4XML' Native XML Database System [J]. The Journal of Systems and Software, 2004, (69): 87 - 104.
- 5 Duclos C., Venot A. Structured Presentation of Drug Indications: lexical and semantic analysis and object - oriented modeling [J]. Methods of Information in Medicine, 2000, 39 (1): 83 - 87.
- 6 IBM DB29. 攻克电子病历技术瓶颈 [EB/OL]. [2010-09-30]. <http://www.ibm.com/db2/index.html>.
- 7 Robert H D, Liora A, Sandy B, etc. HL7 Clinical Document Architecture [J]. Journal of the American Medical Informatics Association, 2006, 13 (1): 30 - 39.
- 8 Dolin RH, Aschuler L, Boyer S, et al. HL7 Clinical Document Architecture [EB/OL]. [2010-10-31]. <http://www.hl7.org/v3ballot/html/welcome/environment/index.htm>.
- 9 彭兆丽, 孟浦, 成于珈. 医疗机构病历书写规范 [M]. 北京: 科学出版社, 2003.
- 10 安徽省卫生厅. 病历书写规范 [M]. 合肥: 安徽科学技术出版社, 2004.
- 11 吴殿源, 戴志鑫. 病历书写基本规范使用手册 [M]. 北京: 军事医学科学出版社, 2004.
- 12 Artem C, Mustafa A, Shiyong L, et al. XML Subtree Reconstruction from Relational Storage of XML Documents [J]. Data&Knowledge Engineering, 2007, (62): 199 - 218.
- 13 Joseph F, San K C, Herbert S. The XML Tree Model - toward an XML conceptual schema reversed from XML schema definition [J]. Data&Knowledge Engineering, 2008, (64): 624 - 661.
- 14 William W S, Kenneth D M, Isaac S K. The PING Personally Controlled Electronic Medical Record System: technical architecture [J]. Journal of the American Medical Informatics Association, 2005, 12 (1): 47 - 54.
- 15 Zheng L, Peter B, Michael S. Storage Model for CDA Documents [C] //Proceedings of the 36th Hawaii International Conference on System Sciences, 2003.