

基于医院系统的多源数据集成技术研究^{*}

葛敬军

张宝昌

(1 济宁医学院医学信息工程学院 日照 276825
2 济宁医学院健康大数据研究中心 日照 276825)

(1 济宁医学院健康大数据研究中心 日照 276825
2 济宁医学院信息技术中心 日照 276825)

王其华 胡珊珊

李庆玲

(济宁医学院医学信息工程学院 日照 276825)

(1 济宁医学院医学信息工程学院 日照 276825
2 济宁医学院健康大数据研究中心 日照 276825)

〔摘要〕 介绍多源医疗数据集成相关研究情况,在已有技术基础上,结合嵌套对象模型、虚拟视图技术和 Mashup Web 服务等关键技术,研究多源数据如何从各自系统中分离、表示、提取以及集成,实现系统和数据的互操作,对应用场景进行阐述。

〔关键词〕 多源系统;嵌套对象模型;虚拟视图;数据集成

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.08.008

Study on Multi-source Data Integration Technology Based on Hospital System GE Jing-jun, 1Department of Information Engineering, Jining Medical University, Rizhao 276825, China, 2Research Center of Big Data and Healthcare, Jining Medical University, Rizhao 276825, China; ZHANG Bao-chang, 1Research Center of Big Data and Healthcare, Jining Medical University, Rizhao 276825, China, 2Information Technology Center, Jining Medical University, Rizhao 276825, China; WANG Qi-hua, HU Shan-shan, Department of Information Engineering, Jining Medical University, Rizhao 276825, China; LI Qing-ling, 1Department of Information Engineering, Jining Medical University, Rizhao 276825, China, 2Research Center of Big Data and Healthcare, Jining Medical University, Rizhao 276825, China

〔Abstract〕 The paper introduces related study on multi-source data integration. Based on the existing technical foundation, through key technologies such as nested object model, virtual view technology and Mashup Web service, it studies on the segregation and extraction of multi-source data from, expression of multi-source data in, and integration of multi-source data with their respective system, realizes the mutual operation of the system and data, and elaborates on the application scenarios.

〔Keywords〕 Multi-source system; Nested object model; Virtual view; Data integration

〔收稿日期〕 2018-04-04

〔作者简介〕 葛敬军,博士,副教授,发表论文 16 篇;通讯作者:张宝昌,讲师。

〔基金项目〕 国家自然科学基金项目(项目编号:81571334);济宁医学院博士科研基金项目(项目编号:JY2015BS20);济宁医学院青年教师科研扶持基金(项目编号:JY2016KJ031Y);济宁医学院青年教师科研扶持基金(项目编号:JY2017KJ054);济宁医学院教育科学研究课题(项目编号:16005)。

1 引言

医院经过多年的信息化发展在不同时期会遗留大量的系统,同时针对不同临床业务积累了大量的数据,这些系统和数据能够为各类医学应用和研究提供支持。将不同系统的数据加以综合,消除多源系统之间的冗余和矛盾,开展医疗数据综合应用越来越受到重视。从 20 世纪 90 年代开始,各级医院开始关注医疗数据的价值,对医疗数据系统进行整合。从早期的数据库^[1-2]到面向服务架构(Services-oriented Architecture, SOA)集成、数据网格^[3]、面向云计算的医院数据集成^[4],先后经历了部门集成、系统集成、数据集成阶段,各类集成方法不断推陈出新,在不同医院业务中发挥着重要作用。现今的各医院中存在不同时期的系统,这些系统由于缺少互操作标准和行业规范,数据很难获取及与周围的系统进行融合。另外医疗数据正面临异构、分散等诸多挑战,目前医疗数据集成还不能建立通用的数学模型、算法和框架,需要根据具体情况选择合适的集成方法。医疗数据集成是一个持久的课题,在实际应用中有待进一步的研究实践。

2 相关研究

2.1 医疗数据集成

医疗数据随处可见,如医院信息系统(Hospital Information System, HIS)、电脑断层扫描(Computer Tomography, CT)、核磁共振(Magnetic Resonance Imaging, MRI)、数字减影血管造影(Digital Subtraction Angiography, DSA)、ECT、PET、DR、脑电图、多导心电图、人体扫描、监护仪数据等,为临床诊断提供依据。多源数据集成是一项多层面的数据处理工程,最早应用于军事领域,后来在多个领域中得到发展。其原理类似于综合处理信息,将来自人体各个感官(眼、耳、鼻、四肢)的信息(外物、声音、气味、触觉)组合起来,理解周围环境和事件。对多源数据进行关联组合处理,利用这些不同来源和形式(如数值型文本型、图形图

像、音视频等)的数据来描述同一主题,为共同的任务集成在一起。近年来医疗数据集成发展迅速,在影像诊断^[5]、三维重建^[6]、肿瘤治疗等应用中发挥重要作用。

2.2 类型

医疗数据集成分为数据、模型和系统集成。在数据集成中,将多个医疗系统中的影像数据加以综合,形成清晰、完整、准确的信息描述。其中三维重建根据 CT、MRI 等图像构建人体立体信息模型。在模型集成中,通过数据模型(联邦数据库、数据仓库、数据空间^[7]等)组合实现更多有效的数据存储与交换。在系统集成中,基于 Web 互操作服务规范,利用 Web Services^[8]将各业务系统提供医院信息门户访问。

2.3 发展方向

目前医疗数据集成的方法、技术和框架研究已积累不少的成果^[9-11],随着新的数据集成方法不断推出,尤其是大数据和人工智能技术的兴起,不少学者将虚拟技术^[12]、神经网络与人工智能技术引入医疗数据集成,对患者信息进行分割、集成以及智能识别,为临床诊断、远程监护、可视化手术等提供及时、准确的人体病征信息。本文是在已有的数据集成技术^[13-15]基础上研究多源数据如何从各自的业务系统中剥离、表示,将数据提取、集成,使数据以可靠、精准的形式融入到医疗系统,实现数据互操作和数据到系统的平滑移动。这对现有的集成方法起到较好的补充作用,在实际应用中还有待于不断完善。

3 关键技术

3.1 嵌套对象模型

3.1.1 嵌套对象模型对资源的描述 医疗数据集成缺少统一描述,给查询、共享带来很多困难,为此设计嵌套对象模型。关于数据集成的研究有很多成果:H. Mizuguchi 等人^[16]提出基于嵌套关系的 WebNR/SD 公共数据模型;T. Lahiri 等人^[17]扩展对

象数据库管理组（Object Data Managment Group, ODMG），提出对象交换模型（Object Exchange Model, OEM）。这两种模型对数据和操作方面具有很强的描述能力，本文借鉴上述模型的优势提出嵌套对象模型（Nested Object Model, NOM）。NOM 的每个对象由 3 元组（Key, A - VS, Object）表示，其中 Key =（MD - ID, MD - T）表示对象名，MD - ID 描述文件的物理大小、存储位置、类型等模式信息；MD - T 是资源类型，包括结构化关系数据库类型、非结构化文件类型、Web 类型等；A - VS =（Attribute, Values）是数据源的属性信息；Object 是嵌套对象主题，通常采用目录列表、数据库名和文件名来描述数据源对象的内容特性。模型以资源类型标识、内容属性、嵌套主题 3 个层次来描述和存储资源，提供对这些资源的操作。嵌套对象模型对资源的描述，见图 1。

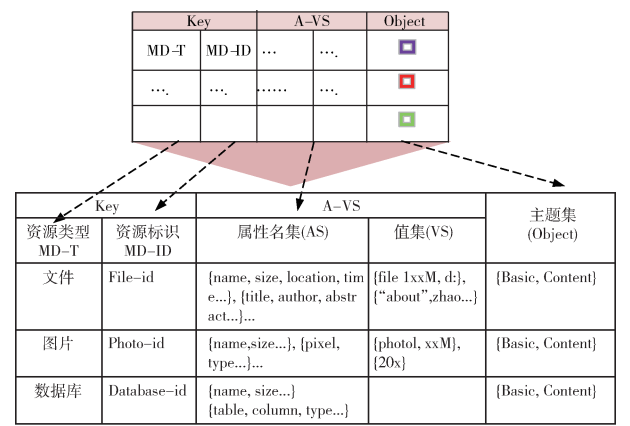


图 1 嵌套对象模型对资源的描述

3.1.2 关系模型和 NOM 对象模型的转换规则

(1) 关系表中的记录映射为 1 组 NOM 对象（具有父子关系），NOM 父对象为集合类型，值为记录的属性，各属性分别建立相应的 NOM 子对象。（2）使用外键子对象表示关联关系，值为关联关系名。如 NOM 对象模型中的记录集！属性！关系对应于关系模型中的关系！属性！引用。同理 NOM 对象模型中的记录集！属性！关系对应于半结构化/非结构化数据可扩展标记语言（Extensible Markup Language, XML）数据中的元素（Element）！属性（Attribute）！元素嵌套关系。NOM 对象模型是以表形式存储，将 NOM 对象映射成关系数据库中的表，

属性映射成关系数据库中的字段，而对象之间的关系映射比较复杂，具体包括对 ValueConstraint、Datasource、Recordset 的映射。从 NOM 定义可以看出 NOM 可以刻画每个对象实体，每个对象实体是嵌套的数据项，这些数据项是查询、存储和处理的基本单位。可以是指定的关系数据表，视图、列或一个数据单元区域，也可以是 XML 文档中标记的指定区域等。NOM 对象模型打破表模型的束缚，可以纵向或横向伸缩，具有灵活的可扩展性。从共享数据角度看每个 NOM 对象只是对象化的共享数据表，提供统一的访问接口，对象名和被封装数据表的名称具有一一对应关系。

3.2 虚拟视图技术

虚拟视图定义为 $R(S) = (A1, A2, \dots, Am)$ 。其中 $R \in A$ 是对象名， S 是属性列表，如果令 A 为全局的属性名称和关系名称集合，则 Ai 是 $Ri(Si)$ 的子属性或者其子关系的虚拟视图。 Ri 被称为 R 的子对象。虚拟视图描述的数据很容易被呈现为嵌套表格，其由表头、行和列组成。表头描述虚拟视图的模式信息，表头中的每一行描述一个虚拟视图；虚拟视图中的属性名显示为列名，顶层的列名为虚拟视图名称，子关系属性对应的列被称为复合列，子属性对应的列被称为原子列；虚拟视图的实例被显示为行。虚拟视图是融合的一部分，具备简单、直观、较强表达、易于人机交互使用的特点。虚拟视图可以对不同的数据源进行映射，能够使多个嵌套对象表格中的数据合并，构建基于全局/局部模式的虚拟视图。虚拟视图的生成过程，见图 2。首先对各类实体数据源进行模式提取，将各数据源的模式信息集中到 1 个中心服务器上，利用这些抽取到的信息建立对嵌套对象模型的转换。图 2 中嵌套对象模型的阴影部分为虚拟视图的表头，反映虚拟视图的数据模式，非阴影部分为虚拟视图的实例；其次完成实体数据源向虚拟数据库的映射。底层的多源特性将对用户屏蔽，实现集成中关键的步骤。通过这一系列的转换、抽取和分析，所有数据源将以多个界面展示给用户。虚拟视图实质是对各种数据源的属性或文本信息的抽取过程。虚拟

视图代替数据库,扮演虚拟数据库的角色,虚拟视图上的数据仍以原有的格式和状态保存在原位置。业务系统通过虚拟视图系统选择要融合的数据,提取虚拟视图与数据源的映射,进行虚拟视图查询,再到数据源中提取数据,发送并做进一步的处理。

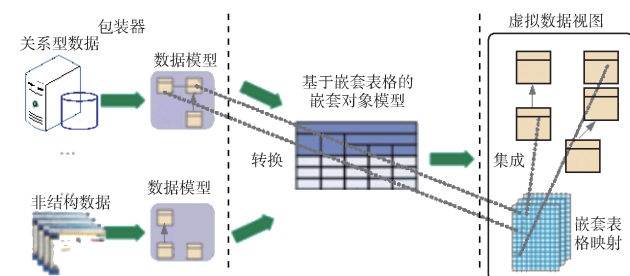


图2 虚拟视图生成过程

3.3 Mashup Web 服务

对于医疗系统融合而言主要解决系统和数据连通问题,在分布式系统环境下保证系统和数据的平滑移动。因此本文将各个医疗系统视为服务,为每个系统设计专有的 Web Services 适配器,将系统和数据进行注册和发布,定义符合 Web Services 标准的服务。在不改变系统的情况下将系统和数据以及运行环境等包装为服务。在 Mashup^[18] Web 平台上进行服务组合。Web Services 适配器的功能包括:

(1) 将底层消息和简单对象咨询协议 (Simple Object Access Protocol, SOAP) 消息互相转换。(2) 将 XML 格式的请求转换为应用系统、应用程序能解析的格式,同时将处理后的结果转换为 XML 格式,然后传递到 Mashup 发布平台上,完成服务集成和协调。(3) 将应用系统包装成业务服务。Mashup 发布平台,见图 3。利用开放应用程序接口 (Application Programming Interface, API) 和工具能够非常快速地将数据与某个应用整合在一起,以满足情景式应用的需求。主要由 4 部分组成:(1) 应用管理器负责提供 Mashup 的安装、卸载、生命周期管理、应用情景管理以及事件交换管理等。(2) 布局管理器负责按照需要灵活排列 Mashup 的定义和应用陈述。(3) 数据管理器提供透明的数据服务访问。(4) Mashup 运行管理器建立客户端与后台的 REST/HTML 连接,支持用户调用 Mashup 组件。

从技术层面看 Mashup 是基于 HTML、CSS、ECMAScript、XML 等技术规范编写的应用。通过服务组件技术、REST/JOSN/HTML 协议、公共接口 API 与各个应用进行请求响应的通信。从应用角度上可以将 Mashup 分为数据集成、系统融合。其中表示系统融合即指 Web 集成。一方面利用 Mashup 技术将系统分解封装,形成服务组件支持调用;另一方面支持医疗应用需求,随时从 Mashup 服务组件中选择组件构建应用系统。Mashup 集成平台能更好地体现以用户为中心的特点,因而受到广泛关注,成为企业 2.0 建设的核心技术。

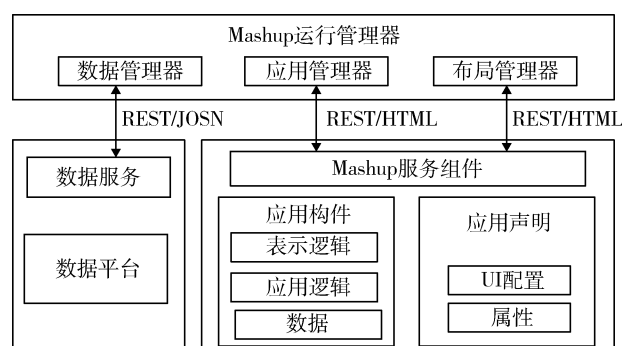


图3 Mashup 发布平台

4 应用场景

目前医疗监护、病症诊断、社区医疗服务等应用环境一般都配备床边监护仪、中心监护仪、心电图机等多种监测系统。这些系统采集不同的生理指标,如血氧、脉率、体温、心电等,由于各系统是对同一对象不同层面的监测,所以这些系统之间存在一定的相关性,需要充分考虑不同系统指标的连系,在信息全面的基础上对患者生理数据集成及综合评估,最大化利用多个系统协同工作。多源数据集成成为建立智能医疗监护、社区医疗服务系统提供从数据到系统级的融合。远程监测、诊断和医疗应用数据集成框架,见图 4。医疗监护、社区医疗服务系统依据生命体征信息得到相应数据流,这些数据流分布、异构而且相关联,不同数据流被包裹在不同模型和系统中,采用嵌套对象模型从不同系统中提取和分离数据流,利用虚拟视图和 Mashup 服务组件技术对数据流进行集成和转储。从数据流中

提取特征,通过决策判别树或神经网络分类器,根据正常及异常时的特征数据建立监护模型。利用监护模型进行分类及判别患者的异常情况,在 SOA 服务机制上对医疗监护、社区医疗服务进行业务分解和组合等,建立相应的预置服务,通过 Web Services 连接到监护中心,根据患者病史、病历等基本信息,以及患者状况变化的各种人体图像和生物电信号对其病情进行远程医疗监护。

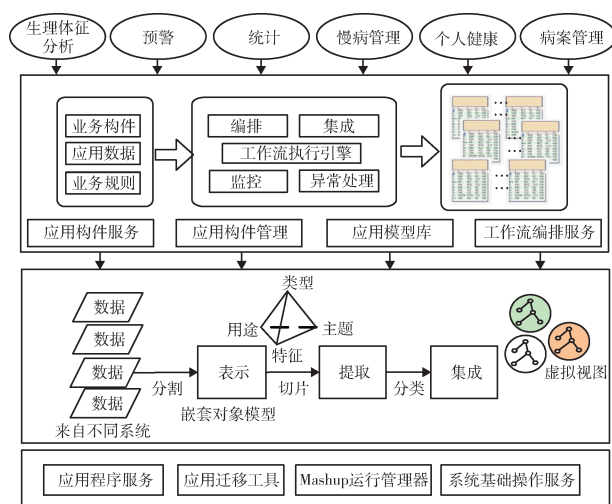


图 4 远程监测、诊断和医疗应用数据集成框架

5 结语

目前许多医院系统的架构和数据平台存在缺陷,将应用、需求与数据打通,将医疗病历、案例或者诊断标准化,解决数据平台兼容、底层数据复杂、存储形式差异等问题,实现数据和应用的平滑移动。同时医疗数据集成需要考虑多层次的知识积累,将知识与数据结合起来,通过人工智能方面的研究和发现进一步去挖掘、预测,从而提高疾病诊断与预防水平。

参考文献

- 1 刘高军,鲍晓琦. 基于联邦数据库的数据集成平台研究与改进 [J]. 计算机光盘软件与应用, 2012, (2): 45-46.
- 2 马少锋,温明锋,陈健超. 基于数据仓库的医院数据分析平台建设与应用 [J]. 医学信息学杂志, 2017, 38 (7): 18-21.
- 3 蒋历军. 基于网格的医院数据系统 [J]. 国外医学. 生

- 物医学工程分册, 2005, 28 (6): 361-365.
- 4 侯佳音,史淳樵. 云计算技术在医院的信息化建设中的应用研究 [J]. 电子设计工程, 2016, 24 (5): 35-39.
- 5 孙雪梅. 区域医疗影像诊断中心的建设方案探讨 [J]. 中国医疗设备, 2016, 31 (4): 97-99.
- 6 高鹏,刘鹏,苏红森,等. 基于 HTML5 与可视化工具包的医学影像三维重建及交互方法研究 [J]. 生物医学工程学杂志, 2015, 32 (2): 336-342.
- 7 刘莉,赵霞. 医院信息系统中访问控制机制的研究 [J]. 医疗卫生装备, 2009, 30 (11): 48-49, 56.
- 8 高振辉,敖继威,范鹏飞. 基于 Web Services 的医院信息资源共享环境构建 [J]. 信息与电脑 (理论版), 2016, (16): 87-89.
- 9 孙小康,吴思竹,张泽,等. 基于移动医疗的可穿戴设备数据集成平台典型案例探析 [J]. 医学信息学杂志, 2016, 37 (9): 13-20.
- 10 HealthKit. Apple Inc [EB/OL]. [2016-07-20]. <https://www.healthkit.com/>.
- 11 Mezghani E, Exposito E, Drira K, et al. A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare [J]. Journal of Medical Systems, 2015, 39 (12): 185.
- 12 邹得强,潘志强,吴庆斌. 基于 VMware 虚拟化技术的医院信息系统建设 [J]. 医学信息学杂志, 2014, 35 (10): 22-25.
- 13 葛敬军,孔繁之,张明,等. 数据密集型应用中的异构数据集成服务研究 [J]. 计算机工程, 2015, 41 (7): 60-65.
- 14 张鹏,王桂玲,季光,等. 基于数据服务的数据组合视图的优化更新 [J]. 计算机学报, 2011, 34 (12): 2344-2354.
- 15 温彦,刘晨,韩燕波. 利用数据服务即时生成跨域数据视图 [J]. 计算机科学与探索, 2012, 6 (3): 221-236.
- 16 Mizuguchi H, Kitagawa H, Ishikawa Y, et al. A Rule-oriented Architecture to Incorporate Dissemination-based Information Delivery into Information Integration Environments [M]. Heidelberg: Springer, 2000: 185-199.
- 17 Lahiri T, Abiteboul S, Widom J. Ozone: integrating structured and semistructured data [M]. Heidelberg: Springer, 2000: 297-323.
- 18 刘建勋,石敏,周栋,等. 基于主题模型的 Mashup 标签推荐方法 [J]. 计算机学报, 2017, 40 (2): 520-534.