

数字化医院多源异构数据集成平台研究与设计*

张红伟 翟鹏程 冯思度 杨健叶 贾贻然

(徐州医科大学医学信息学院 徐州 221004)

[摘要] 介绍数据集成相关研究情况, 阐述多源异构数据集成技术以及基于 ADO.NET 和 Web service 的多源异构数据集成方案、功能模块及其实现, 指出其有助于实现数字化医院中多源异构数据交换、共享和集成分析。

[关键词] 医院; 多源异构数据; 集成; ADO.NET; Web 服务

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2019.01.005

Study and Design of Multi-sources Heterogeneous Data Integration Platform in Digital Hospitals ZHANG Hongwei, ZHAI Pengcheng, FENG Sidu, YANG Jianye, JIA Yiran, School of Medicine Information, Xuzhou Medical University, Xuzhou 221004, China

[Abstract] The paper introduces related study on data integration, elaborates on the multi-sources heterogeneous data integration technology, the integration protocols, functional modules and their realization of multi-sources heterogeneous data based on ADO.NET and Web service, as well as points out that it is conducive to the realization of the exchange, sharing and integration analysis of multi-sources heterogeneous data in digital hospitals.

[Keywords] hospital; multi-sources heterogeneous data; integration; ADO.NET; Web service

1 引言

在医院的数字化建设过程中需规划和建设大量信息系统, 如医院信息系统 (Hospital Information

System, HIS)、B 超、医学影像存储与传输系统 (Pictures Archiving and Communication System, PACS)、电子病历 (Electronic Medical Records, EMR)、病理分析、体检、检验信息系统 (Laboratory Information System, LIS)、财务管理、智能楼宇等。这些在数字医院建设中功能各异而又彼此协同的诸多系统产生大量数据, 来源不一且种类多样, 如何将其集成来进行信息互通和数据共享, 从而充分利用现有信息资源并提高数据价值和利用率是亟待解决的问题。

2 数据集成相关研究

2.1 研究起源

数据集成是指将不同来源、格式、特点的数据通过各种技术手段进行无缝连接, 从而实现多种数

[收稿日期] 2018-09-04

[作者简介] 张红伟, 讲师, 博士, 发表论文 5 篇。

[基金项目] 2017 年国家级大学生创新创业训练计划项目“数字化医院中多源异构数据集成平台的设计与开发”(项目编号: 201710313021); 江苏省 2017 年高等学校大学生创新创业训练计划项目“数字化医院中多源异构数据集成平台的设计与开发”(项目编号: 201710313021Z); 徐州医科大学 2015 年校级课题“基于 Web Service 的数字化医院中多源异构数据集成技术研究”(项目编号: 2015KJ15)。

据的统一访问和管理。在研究多源、异构数据集成技术方面，国内外不少学者提出不同方案。数据集成研究起源可追溯到 Mcleod 等于 20 世纪 80 年代中期提出的联邦式数据库系统^[1]。随着计算机网络的迅速普及和发展，传统的数据库数据外又涌现出大量的半结构化数据。传统数据库集成方法已经无法满足人们获取更多、更新数据的需求，此时迫切需要一种新的数据集成系统。这种系统不仅能集成传统数据库中的结构化数据，而且还可以集成网络上应用日益广泛的半结构化和无结构化数据。

2.2 最新进展

随着许多新技术（如 CORBA^[2]，XML 等）的涌现，通用异构数据集成系统的研究也正在如火如荼地进行。近年来随着因特网和万维网等网络技术的发展，特别是可扩展标记语言（Extensible Markup Language，XML）的出现进一步推动数据集成发展，对差异更大的数据源进行集成已经成为研究人员的新课题，将重点转移到基于 XML 的数据集成方面^[3-5]。此外有学者将本体技术应用于多源异构数据集成^[6-7]，以及使用中间件技术进行异构数据集成^[8-10]，在一定程度上解决相关领域多源异构数据集成问题，但是没有针对数字化医院的多源异构数据集成方案。

3 数字化医院中数据集成技术与方案设计

3.1 集成技术

ADO.NET 起源于 ADO（ActiveX Data Objects），是为 .NET Framework 程序员提供数据访问服务的组件，提供对 SQL Server 和 XML 此类数据源以及通过对象链接和嵌入数据库（Object Linking and Embedding Database，OLEDB）和开放式数据库链接（Open Database Connectivity，ODBC）公开的数据源一致访问，此外共享数据程序可以通过 ADO.NET 连接到这些数据源并可以检索、处理和更新其中包含的数据^[13]，对多种不同数据源、不同结构数据进行访问。近年来 Web Service 技术已得到快速发展和应用，其采用 XML 定义一组 Web Service 协议栈，通过 SOAP、WSDL、UDDI、WSFL、BPEL4WS 等开放协议和标准提供面向因特网应用的统一服务注册、发现、绑定和集成机制，成为广域环境下实现

互操作的一种主要机制，得到学术界和产业界的广泛认可。将 Web Service 技术引入数据集成研究领域有助于解决所面临的应用集成、资源共享、系统互操作和标准化等问题^[11-12]。针对不同结构的数据源开发相应的数据集成 Web Service 并将其发布以供对应的应用程序或数据源使用，完成该数据源向指定目标数据库集成。

3.2 集成方案

首先分析数字化医院多源异构数据特点，然后根据其特点针对性地开发相应 Web Service 方法并发布，接着各系统向目标平台提供源数据，平台对不同来源的数据进行特点分析并判断其种类，最后选用对应的方法或基于 ADO.NET 功能模块进行数据处理并将处理后的数据集中到一个结构化的数据仓库。在数据完成集成后可以信息进行信息的统一查询、共享、分析等工作。方案中对多源异构数据集成的处理流程，见图 1。

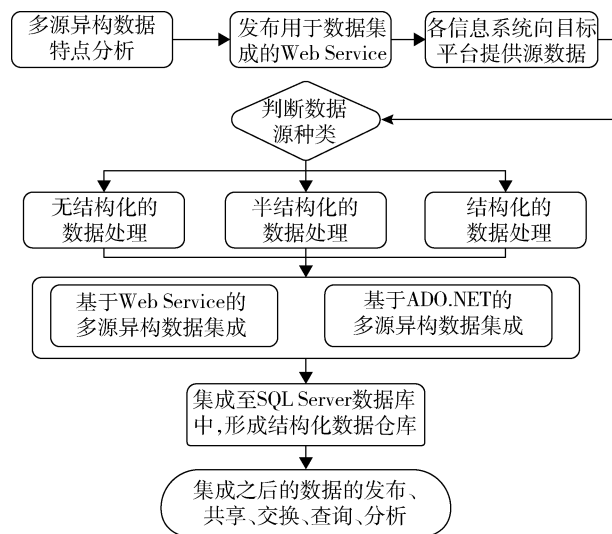


图 1 多源异构数据集成方案技术路线

4 数据集成平台系统实现

4.1 功能模块

在分析数字化医院中异构数据特点、集成需求基础上共设计系统数据集成的 5 个功能模块和数据利用的两个功能模块，见图 2。其中数据集成模块用于将不同来源、结构的数据进行读取，然后分析并根据结构决定是否集成到目标数据库中；数据利用模块主要

对集成后的数据进行统一查询、共享、统计分析等。

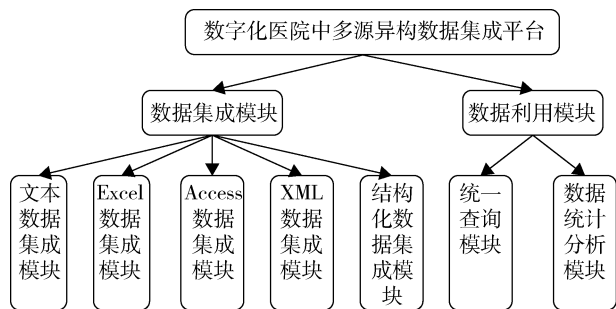


图 2 系统功能模块

4.2 集成功能实现

多源异构数据集成功能是整个平台的核心。为直观显示以及后续数据的集成，该系统分别将患者基本信息数据（Excel 格式）、患者献血信息数据（Txt 格式）、患者诊断信息数据（XML 格式）、患者体征信息数据（Access 数据库格式）以及病房温湿度信息数据（字符串格式）显示在 DataGridView 中，将集成后的数据显示在另外一个 DataGridView 中。对于 Excel 数据源、文本数据源、Access 数据库源以及字符串源等数据的集成，采用 ADO.NET 技术。其实现过程是首先将这些数据做一定的预处理，从而形成结构化的数据集（DataSet），在集成功能函数中，系统会自动判断要集成的具体数据；逐行判断 DataSet 中的数据是否存在目标数据库中，若存在则根据关键字进行数据更新，若不存在则进行数据插入。对于 XML 数据的集成，采用 Web Service 方式。首先设计并发布针对 XML 数据集成方法的 Web Service，可根据需要发布到内网或外网上，以供其他程序调用，能够满足跨编程语言以及跨操作系统的需求。对于 XML 数据集成，首先根据其节点进行数据元素提取，从而将这些元素放进一个数组进行统一预处理，最后通过调用 Web Service 完成数据集成。

5 结语

本文采用 ADO.NET 以及 Web Service 技术对数

字化医院中的数据进行集成，实现不同信息系统之间的数据交换、共享、统一查询和统计分析。此外 Web Service 技术的运用有利于跨平台和跨操作系统的系统集成，这是传统数据集成的一大进步，也为后续的大数据分析提供基础。多源异构数据集成涉及领域极广，还有待于进一步深入研究，如如何利用集成的数据进行大规模数据分析和挖掘、充分发掘数据价值等。

参考文献

- 1 W Bright M, R Hurson A, Pakzad S H. A Taxonomy and Current Issues in Multidatabase Systems [J]. Computer, 1992, 25 (3): 50 - 60.
- 2 王宁, 陈滢, 俞本权, 等. 一个基于 CORBA 的异构数据源集成系统的设计 [J]. 软件学报, 1998, 9 (5): 378 - 359.
- 3 Bertino E, Catania B. Integrating XML and Databases [J]. IEEE Internet Computing, 2001, 5 (4): 84 - 88.
- 4 王伟伟, 孙庆鸿. 基于 XML 的分布异构数据集成平台 [J]. 东南大学学报 (自然科学版), 2006, 36 (5): 715 - 719.
- 5 黄辉, 于爱军, 李永生, 等. 基于 XML 的虚拟法异构数据集成方法研究 [J]. 计算机应用研究, 2009, 26 (1): 172 - 174.
- 6 马新建, 夏士雄. 基于本体的多源异构数据集成方法研究 [J]. 微计算机信息, 2008, 24 (25): 291 - 293.
- 7 潘佳云. 基于本体的异构数据集成技术研究 [D]. 上海: 东华大学, 2013.
- 8 舒后, 何薇, 陈晓鹏. 基于 XML 数据集成中间的查询处理器设计 [J]. 北京印刷学院学报, 2008, 16 (6): 64 - 67.
- 9 裴玉玲, 庞佑兵. 基于中间件的异构数据集成系统设计 [J]. 微电子学, 2010, 40 (2): 260 - 264.
- 10 缪谨励, 陶留锋, 尹兆峰. 多源异构国土规划空间数据集成技术 [J]. 地理空间信息, 2014, 12 (6): 5 - 7, 1.
- 11 胡春明, 怀进鹏, 孙海龙. 基于 Web 服务的网格体系结构及其支撑环境研究 [J]. 软件学报, 2004, 15 (7): 1064 - 1073.
- 12 尚龙华, 安毅生, 张绍阳, 等. 基于 Web 服务的交通数据交换过程 [J]. 计算机系统应用, 2015, 24 (3): 260 - 265.
- 13 Microsoft. ADO.NET [EB/OL]. [2018 - 08 - 20]. <https://docs.microsoft.com/en-us/dotnet/framework/data/adonet/index>.