

Web 服务在多源异构农业数据融合上的应用研究

倪芳^{1,2,3}, 曾辉^{1,4}, 卓辉^{1,4}, 廖桂平^{1,2}

(1. 湖南农业大学 信息科学技术学院, 湖南 长沙 410128;

2. 湖南农业大学 农业信息研究所, 湖南 长沙 410128;

3. 湖南省科学技术厅信息研究所, 湖南 长沙 410001;

4. 湖南省国家农村农业信息化工程技术研究中心, 湖南 长沙 410128)

摘要:从农业物联网中海量数据、繁杂信息等诸多平台之间存在农业数据库交互与集成的现状出发,结合其分布式 Web 服务可对不同地域、不同库区以及不同系统平台之间进行知识交互,采用 Web 服务技术对其数据内容进行定义、修改、查询等特殊方式对其当前农业信息系统中的用户需求和标准可生成相关数据仓库和模式预测模型;同时可对其采用 XML 在任意应用系统中读写数据,实现在各种系统中数据交换与信息结合,同时使得数据能安全独立地在程序中运行与分析且不影响系统平台的性能,为应用终端用户提供多源数据,并使得数据之间异构融合、信息相互畅通传输,打破前期不同数据库之间存在的“数据鸿沟以及数据壁垒”。这有助于提高农业信息服务,以及更好更快地促进农业物联网中数据的传递与交互。

关键词:数据;可扩展标记语言;用户;农业物联网

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2016)08-0129-05

doi:10.3969/j.issn.1673-629X.2016.08.027

Research on Application of Web Services in Multi-source Heterogeneous Data Integration on Agriculture

NI Fang^{1,2,3}, ZENG Hui^{1,4}, ZHUO Hui^{1,4}, LIAO Gui-ping^{1,2}

(1. College of Information Science and Technology, Hunan Agricultural University, Changsha 410128, China;

2. Agricultural Information Institute, Hunan Agricultural University, Changsha 410128, China;

3. Scientific and Technical Information Institute of Hunan, Changsha 410001, China;

4. National Rural Agricultural Information Engineering Technology Research Center of Hunan, Changsha 410128, China)

Abstract: From the situation that agricultural database exists interaction and integration in many platform like huge amounts of data, miscellaneous information and others in Agricultural Internet of Things, combined with its distributed Web services interacting between different regions, different reservoir area and different platforms, the special way that the Web services technology is used to define, modify and query their data contents can generate relevant data warehouse and predictive model for user requirement and standard in its current agricultural information system, while using XML to read and write data to any application system, the data exchange and information merging is implemented in a variety of systems. At the same time the data can safely and independently operate and analyze in the program and does not affect system performance, providing multi-source data for the end-user application, making data heterogeneous fusion and information smooth transmission with each other, breaking the “digital divide as well as the data barrier” between the pre-issue different databases. It will help improve agricultural information services and promote the data transmission and interaction in Agricultural Internet of Things better and faster.

Key words: data; XML; user; Agricultural Internet of Things

收稿日期:2015-10-10

修回日期:2016-03-03

网络出版时间:2016-07-29

基金项目:国家自然科学基金资助项目(10674045, F050501); 国家科技部“十二五”重点课题(2011BAD21B03, 2012BAD35B05); 国家科技部“国家星火计划”项目(2011GA770001); 国家科技部科技支撑计划课题(2012BAD35B00); 高等学校博士点基金资助项目(200805370002); 湖南省科技计划项目(2013GK3106); 湖南省教育厅学位基金项目(YB2010B024); 湖南农业大学人才引进基金项目(08YJ02); 新型农业信息传感技术创新团队项目(92020200004)

作者简介:倪芳(1985-), 女, 硕士, 研究方向为数字图像技术在农业物联网中的应用。

网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20160729.1833.006.html

0 引 言

随着计算机科学的不断发展,数据、知识、信息^[1-3]在农业生产、指挥、管理、流通等领域的作用愈来愈突出。首先,在农业生产方面,由于受地理环境、水分、土壤资源等约束^[4],使得农业数据在采集、存储、加工、处理方面变得更加复杂;与此同时,数据也由简单定量收集向满足当前用户所需的信息服务方面发展,其数据呈单一、散片等形式存在^[5]。因此,如何使得多源数据异构融合^[6-7]以实现农业数据集成与共享迫在眉睫。为此,文中采用 Web-XML 技术^[8-9]对其不同用户终端的不同需求做出相应的数据处理,实现农业多源数据异构整合,使其达到数据与信息共享。

1 系统架构设计

1.1 系统运行环境及设计思路

系统以 Oracle 10g 和 SQL Server 分别作为 (Client 和 Server) 的数据库设计结构。采用 Web Logic Server 12c 作为动态 JSP 引擎^[10];在 Linux 下运用 Visual C++ 6.0 编写设计模型与推理库,生成人机界面与数据库综合管理系统。首先,农业生产自身诸多因素(如作物生长环境、地域、湿度、气温)使得农业多源异构特征表现异常突出,其中关于此异构不仅仅指数据库内部表结构异构,而且还包含操作系统的异构。其次,在应用终端数据使用的装置(RMV、PDA、语音识别、文字表达)不同,其数据存储、处理、加工形式也不同,这将使得数据仓库中的数据格式变得多元化;同时,在实际

生产过程中异构数据应用相对同源数据更为显著。因此,实现农业数据多源异构是业务逻辑层与应用层之间中间件的核心问题。

为此,有效解决多源异构问题对提高农业数据集成与共享日趋重要。实现方案(见图 1)主要有以下几点:

(1)构建统一的业务逻辑范畴;提高数据可靠性。其具体方案为利用 XML、SOAP\WSDL 或使用农业信息系统的成熟架构,解决各个数据库中语义冲突以及通讯协议标准不一致的问题,实现所有农业信息数据在此平台下两两数据库之间的数据同步,从而实现共享与集成,达到无缝链接作用。

(2)运用中心数据库,保证系统访问透明性。其具体实现方法为构建一个中心数据库对其所有农业信息数据库进行存储和 Web Service 认证,使得农业应用终端需求多元化以及数据挖掘得以实现;在此过程中,用户不必考虑其数据模型、管理系统、应用系统之间的异构问题,只需运用某种工具,掌握具体使用规则对其进行检索,获取满足自己需要的信息服务。

(3)数据独立性充分保证各个农业数据库系统中数据在集成与共享时平台的独立。其具体路线为当数据仓库中某个或某些数据发生改变,其共享平台保持原有架构不变,将其 XML 元数据映射技术作为中转仓库对数据进行处理,保证程序与数据两者充分隔离。这样既保证了数据的真实性,也保障了系统平台的可靠性。

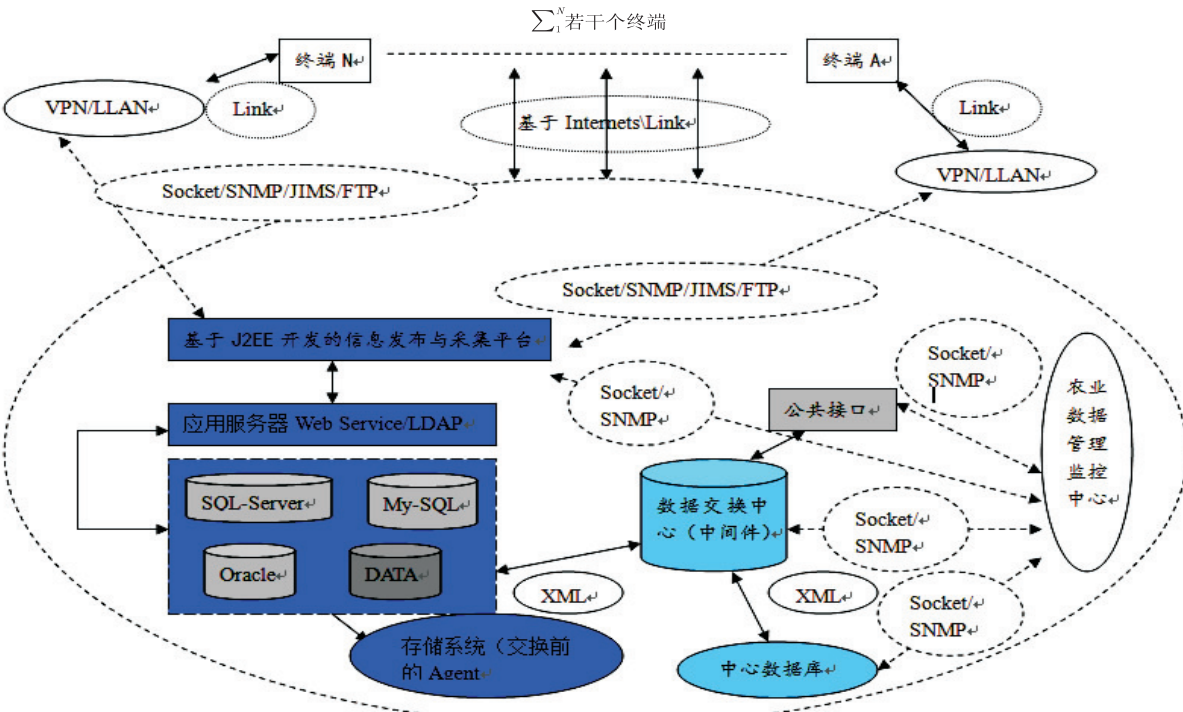


图 1 面向 Web 服务农业数据集成与共享拓扑图

1.2 底层模块架构设计

农业信息综合平台建设是集多因素、多层次、多目标、关系复杂且在时间与空间受其限制的复杂庞大系统。而这一系统的数据复杂性、实时更新性、海量数据模糊性,使得该系统相对其他专家系统要求更高。

(1)在农业基础数据动态变化时,要考虑数据管理库系统、知识推理库、模型解释器(见图 2)中的数据更新以及技术(GPS,GIS,RS)^[11]更新。为系统中基础数据库以及知识库提供数据支撑和保障,加速农作物生产管理即时化、空间化、立体化^[12]发展。同时,对其被提供的数据无论在量还是在质上都要求精准,扩充其三库(知识库、数据库、模型库)的强大生命力,这将进一步丰富 ES 系统的功能与开发价值。

(2)关于系统功能集成化^[13]问题:在农作物生产管理过程中受气候、湿度、病虫害以及时空差异、地域差异等因素的影响,考虑将不同领域的多个农业专家系统相结合,综合其作物生长态势模型与专家经验模型、推理机对其作物在不同地域、气候以及培育环境变化时动态模拟作物产前、产中、产后全过程,这将大大降低农资成本,为科学生产管理提供依据。

(3)数据模型多元化^[14]:当前农业专家系统针对知识简单、单一的线性回归模型;而随着 BP 神经网络^[15-16]、模糊数学^[17]、随机模拟^[18]、非线性科学^[19]等学科的产生,对进一步模拟实际农作物生长全过程起到了巨大帮助。

其系统开发平台运用的 C/S 服务体系结构模型,以及后台 SQL Server 数据库管理系统间模块设计方案如下:

1.2.1 封装构建

针对接口文件处理的语义和软构件结构管理器的信息与知识,将交互界面(脚本语言编写界面、模板套用)生成的界面与数据文件、知识封装为一个实用的系统。文中采用组件技术(统一的接口标准)对其构件进行编写,方便跨平台调用与加载,同时使得开发平台模块化性能突出。如数据库构件管理器/知识库构件管理器运用于自身平台的同时兼容开发的应用系统,提高了可重用性与规范性。

1.2.2 农业知识库构建

所谓农业知识库是关于农业相关领域理论、现有数据、核心算法、数值模型等若干个知识散片集群存储在控制器中,其知识经过获取、表达、管理、组织、传递使知识群变得结构化、交互化、易运用。首先,在农业知识获取上,需借助知识工程师、专家经验、智能编辑程序(如 MYCIN 中 TEIRESIAS 知识获取方式)。其次,在农业知识表达方面,为了使得客户终端的农业用户便于对知识以及知识间关联的理解与操作,在 ES

系统中采用数据结构或语义处理机制对农业知识进行表达。其表达过程:在语义 Web 中以 RDF 资源框架、XML 语法、URI 机制^[20]对各种抽象农业知识进行集成与分析,处理当前自然语义、人工逻辑推理以及机器学习处理不了的信息。再次,在农业知识管理上,由于当前农业数据存在多源性,其表达形式存在多样性,如文字、表格、图像等,为此,采用本体技术实现对领域知识与任务本体间的集成与共享^[21],推理出与该领域无关结果(顶层本体),进而加速应用本体(农业知识解释与具体化)的发展;同时,采用描述元数据语义与规则,对隐性农业知识进行充分挖掘,消除同一领域间的不同学科瓶颈和不同领域间的语义障碍,为异构数据融合与知识交互^[22]提供便利。其目的旨在实现农业知识查询、农业知识匹配、农业知识重组、农业知识结构化与非结构化以及消除农业知识冗余。最后,在农业知识传递方面,通常在考虑传递成本以及显性知识编码时常采用单向传递模型,转移知识载体过程中不需要依赖知识接收方;一旦传递过程中存在隐性知识或隐性和显性混合知识时需用到往复传递模型,使得显性知识与隐性知识相互转化,实现知识接受质量与成本均衡;随着隐性知识复杂度增加,进而不需要考虑知识源与知识接受者的主次地位,实现隐性知识与隐性知识间的交互。此时采用交融传递模型,保证知识传递质量与可靠性。

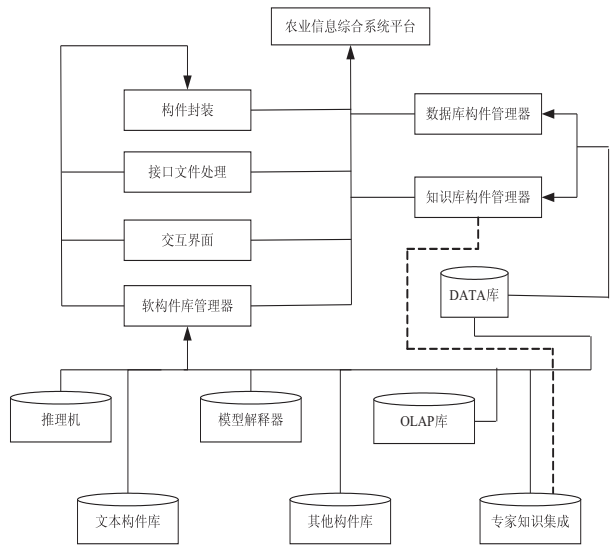


图 2 农业信息综合系统流程模块图

2 农业数据异构关键节点

2.1 农业数据接口层和访问层

为了使农业用户终端能直观对系统服务功能进行操作、访问,常常需要考虑服务接口(在 UUDI 查询中心具有注册认证、实名登陆、数据上行与下行)。由于 Java Web Services Developer Package 为 Java 实现 Web

服务提供开发软件包(在数据交换平台中经常用到 JAXP(Java API for XML Processing),为用户提供描述信息、数据解释和 SOAP 消息机制,SOAP 消息的接收与发送采用 JAXM(Java API XML Messaging),JAX RPC(Java API for XML-based Remote Procedure Calls)调制 SOAP 消息网络对象,XML 数据模式交互运用 Schema 文档把可标记扩展数据变换为本地关系数据,从而降低整个农业综合系统中服务运营成本、减少数据应用的难度,更好地对不同农业用户提供不同的信息反馈。

多源异构农业数据融合图如图 3 所示。

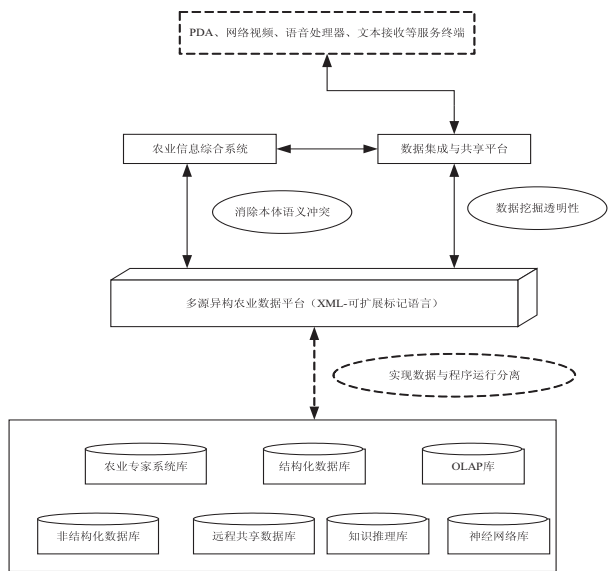


图 3 多源异构农业数据融合图

2.2 农业数据交互关系模式

在数据关系异构中使得 C/S 双方都具备独立的数据库系统是数据融合的关键节点。在 Web 服务体系架构下要求服务器后台提供网络访问接口,用户端提供相应的交互前端;为此采用 XML 格式对农业综合系统中本地数据(源数据、公共模式数据、局部模式数据、用户模式中数据)进行交换,要求使用相关程序对其进行转化。转化过程如下:

(1)在 Client 方采用 Web 服务封装 UUDI 注册和查询中心发布:其中这些数据以(关系模型、半结构化、数值模型、文本格式)源数据形态保存,之后进入局部模式层采用 SQL Server 2008 JDBC Driver 和 Oracle 的 JDBC Driver 访问 SQL Server 2008 和 Oracle 10g,对其数据进行描述、融合、映射等一系列加工;再通过映射配置将局部模式转化为用户模块,进入面向终端用户的公共模式层,根据不同需求、权限生成若干个子模块以满足作物农户、农产品终端市场、农业科研院所以及农业政府部门的需求。

(2)针对运行过程中 Server 的接收请求:在 UUDI 数据注册中心保存着登陆数据源节点,通过对源数据

抽取完成源数据向局部数据模式的转变,进而在融合中心完成公共模式数据到局部模式数据的映射和排序。

不同用户根据自身需求和权限向融合中心数据库发送业务请求;服务器系统接收到指令之后将融合中心系统库中的数据进行转换并反馈给信息请求方。例如,Linux 下 C/C++系统开发平台中基于 Web Service 的异构数据融合部分代码如下:

```
#include "bayesFtL.hpp"
#include <boost/limits.hpp>
#include <vector>
/* Filter namespace */
namespace Bayesian_filter
{
    const Bayes_base::Float Numerical_rcond::limit_PD_init =
std::numeric_limits<Bayes_base::Float>::epsilon() * Bayes_
base::Float(1e5);
    Bayes_base::~Bayes_base()
    {}
    void Bayes_base::error (const Numeric_exception& e)
    {
        throw e;
    }
    void Bayes_base::error (const Logic_exception& e)
    {
        throw e;
    }
    .....
    .....
    .....
    while( ssi < sortR.end() )
    {
        if( ColProxy::less( *ssp, *ssi ) )
            ++u;
        ssp = ssi;
        ++ssi;
    }
    return u;
}
Sample_filter::Sample_filter (size_t x_size, size_t s_size) :
Sample_state_filter(x_size,s_size)
{
}
void Sample_filter::predict(Functional_predict_model& f)
{
    const size_t nSample =S.size2();
    for (size_t i=0;i!= nSamples;++i) {
        FM::ColMatrix::Column Si(S,i);
        FM::noalias(Si)= f.fx(Si);
    }
```

```
}  
} //namespace
```

3 结束语

文中提出采用分布式 Web 服务中的 XML 技术对多源异构农业数据实现融合,能将抽象的农业数据变得模型化,从而有效解决当前农业不同领域的数据整合与交互问题;同时,通过智能农业知识库挖掘其中的隐性知识,有利于提高农业数据价值,对农业物联网中多源异构数据融合与集成产生深远影响,加速现代农业发展步伐。

参考文献:

- [1] Bellinger G, Castro D, Mills A. Data, information, knowledge, and wisdom[EB/OL]. 2004. <http://www.systems-thinking.org/dikw/dikw.htm>.
- [2] Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back to metabolism in KEGG[J]. *Nucleic Acids Research*, 2014, 42: 199–205.
- [3] 倪 凯, 祝晓东, 张 超. 基于关联规则的空间数据知识发现及实现[J]. *计算机应用与软件*, 2005, 22(12): 34–35.
- [4] Pimentel D, Houser J, Preiss E, et al. Water resources: agriculture, the environment, and society[J]. *Bioscience*, 1997, 47(2): 97–106.
- [5] Weiser M. The computer for the 21st century[J]. *Scientific American*, 1991, 265(3): 94–104.
- [6] Wang H, Liu X, Lai J, et al. Network security situation awareness based on heterogeneous multi-sensor data fusion and neural network[C]//Proc of second international multi-symposiums on computer and computational sciences. [s. l.]: IEEE, 2007: 352–359.
- [7] Ren Y, Peng D, Wu J, et al. The research and application of multi-resource heterogeneous data fusion on dynamic traffic routing system[M]//Foundations and practical applications of cognitive systems and information processing. Berlin: Springer, 2014: 375–387.
- [8] Appelbaum M S, Garvett J H. System, method and computer program product for data event processing and composite applications: U. S. , 7774791[P]. 2010–08–10.
- [9] Sheth A, Thirunarayan K. Semantics empowered Web 3.0: managing enterprise, social, sensor, and cloud-based data and services for advanced applications[M]. US: Morgan & Claypool Publishers, 2012.
- [10] Greenwald R, Stackowiak R, Stern J. Oracle essentials: Oracle database 12c[M]. [s. l.]: O'Reilly Media, Inc, 2013.
- [11] Deren L. On definition, theory and key technics of the integration of GPS, RS and GIS[J]. *Journal of Remote Sensing*, 1997, 1(1): 64–68.
- [12] Fernandez-Galvez J, Simmonds L P. Monitoring and modelling the three-dimensional flow of water under drip irrigation[J]. *Agricultural Water Management*, 2006, 83(3): 197–208.
- [13] Bonaudo T, Bendahan A B, Sabatier R, et al. Agroecological principles for the redesign of integrated crop – livestock systems[J]. *European Journal of Agronomy*, 2014, 57: 43–51.
- [14] Pellegrini L, Tasciotti L. Crop diversification, dietary diversity and agricultural income: empirical evidence from eight developing countries[J]. *Canadian Journal of Development Studies*, 2014, 35(2): 211–227.
- [15] Zhao Lixi, Shui Pengbo, Fang Jiang, et al. Using monitoring data of surface soil to predict whole crop-root zone soil water content with PSO-LSSVM, GRNN and WNN[J]. *Earth Science Informatics*, 2014, 7(1): 59–68.
- [16] Ma L, Liu F, Chen L, et al. Simulation of Crop Evapotranspiration Based on BP neural network model and grey relational analysis[J]. *Diabetes Care*, 2012, 2(1): 15–21.
- [17] Chatterji S, Tiwary P, Sen T K, et al. Land evaluation for major crops in the Indo-Gangetic Plains and black soil regions using fuzzy model[J]. *Current Science*, 2014, 107(9): 1502–1511.
- [18] Mavromatis T, Hansen J W. Interannual variability characteristics and simulated crop response of four stochastic weather generators[J]. *Agricultural and Forest Meteorology*, 2001, 109(4): 283–296.
- [19] 曾 辉, 卓 辉. 超高速光纤通信系统中孤子间稳定传输滤波器补偿效应研究[J]. *激光杂志*, 2015, 36(1): 87–89.
- [20] Bizer C, Heath T, Berners-Lee T. Linked data—the story so far[J]. *International Journal on Semantic Web and Information Systems*, 2009, 5(3): 1–22.
- [21] Camporeale C, de Nicola A, Villani M L. Semantics-based services for a low carbon society: an application on emissions trading system data and scenarios management[J]. *Environmental Modelling & Software*, 2015, 64: 124–142.
- [22] Hecker M, Lambeck S, Toepfer S, et al. Gene regulatory network inference: data integration in dynamic models—a review[J]. *Biosystems*, 2009, 96(1): 86–103.