·信息技术·

基于本体的多 Agent 企业数据集成模型研究*

Study on Enterprise Data Integration Model Based on Ontology and Multi－agent

唐晓波 刘婷婷

（武汉大学信息管理学院 武汉 430072）

摘 要 针对分布异构环境下企业数据集成的难点问题，将XML、本体和多Agent技术相结合，构建了一个基于本体的多Agent企业数据集成模型，分析了模型的层次结构，并详细探讨了系统实现的关键技术问题，最后以一个实例描述了系统的流程。采用本体技术来实现信息资源的组织，描述信息资源的特性，获取数据的模式，可解决企业异质信息问题。采用多Agent技术实现数据的分发，向用户提供信息处理和检索，可解决企业信息资源的动态性和分布性问题。

关键词 数据集成 本体 Agent 语义

中图分类号 TP311 **文献标识码** A **文章编号** 1002－1965(2009)10－0119－05

企业信息资源是企业在经营活动中累积起来的以信息为核心的信息技术、设备和信息生产者等信息要素的集合，是管理者做出决策的基础，具有广泛性、动态性、分布性和异构性等特点。在众多企业的发展过程中，其内部的信息化工作是分时期、分部门完成的，不同的部门形成了自己的信息资源库，并且信息资源形式多样，包括文本、表格、源代码、服务、视频、音频等，但这些信息缺乏统一的描述，信息资源库也没有相互集成和整合。此外，很多企业在一定程度上实现了主要业务的信息化管理，但各个系统之间，如ERP、PDM、OA、SCM、CRM、DSS、KM等，其信息不能很好的集成和共享。这些情形使得企业形成一个个“信息孤岛”，从而很难为企业领导层的管理和决策提供必要的数据和信息。因此，企业信息资源集成显得异常重要，它是将企业原本离散的、多元的、异构的、分布的信息资源通过物理的或逻辑的方式组织成一个整体，解决“信息孤岛”问题，实现信息资源的优化配置和共享，最终提升组织核心竞争力的动态过程。

随着商业和技术的发展，现存的解决方案往往不能掌握现实环境中的复杂性，很多集成还停留在应用层面，不能实现重用，且对于数据源的改变不能同步，尤其是对于语义异构问题还没有很好的解决，这就影响了集成后数据的质量[1]。IDC的《2008年中国企业数据集成和数据质量调查》显示，超过70%接受调查的中国企业已经建设或正在建设数据集成项目，而数据质量问题导致大部分企业数据集成项目难以达到预期。没有成功的数据集成，就无法实现管理的集成，风险控制、产品经营、决策支持也无从谈起。为了适应企业信息资源异质、企业业务变迁、企业组织形态变化等企业信息资源集成的特点，解决企业信息资源广泛性、分布性和异构性带来的企业数据集成和整合难题，本文提出了基于本体和多Agent来实现企业数据集成。采用本体技术来实现信息资源的组织，描述信息资源的特性，获取数据的模式，可解决企业异质信息问题，采用多Agent技术实现数据的分发，向用户提供信息处理和检索，可解决企业信息资源的动态性和分布性问题。

1 相关技术简介

1.1 基于XML技术的数据集成 XML是W3C提出的一种定义其它语言的元语言，它克服了HTML结构性和扩展性差的缺点，具有平台无关、易于扩展、自描述、语义性强等特性，这些特点使得它可以在数据集成中为结构化数据、半结构化数据、对象数据库等多种数据源的数据内容加入标记，可作为统一的数据描述工具。目前国内外基于XML的异构数据集成的研究有很多，如AT&T实验室的SilkRoute，INRIA的Agora等。这些应用了XML技术的异构数据集成系统都

收稿日期：2009－05－27 修回日期：2009－06－25

基金项目：教育部人文社会科学研究一般项目“企业信息资源集成研究”(编号：08JA870013)成果之一。

作者简介：唐晓波(1962－)，男，教授，硕士生导师，研究方向为管理信息系统；刘婷婷(1986－)，女，硕士研究生，研究方向为企业信息资源集成。

能够较好地解决数据源之间语法上的异构性，但是由于XML文档中缺少语义信息，因此对于语义异构的数据源间的集成与交换显得无能为力，如果使用纯XML方法定义映射方式，系统必须编写额外的映射信息，开发的难度将大大增加。

1.2 **初步引入本体的数据集成** Ontology是一种能在语义和知识层次上描述信息系统的概念模型建模工具，它能用来描述概念及概念之间的关系，并能通过概念之间的关系来描述概念的语义。较之于XML，Ontology对概念的定义更加严格、精确。在数据集成过程中，它一方面可以担任知识共同理解的载体的角色，使得人与人、人与机器之间能达成事物概念的一致性理解。另一方面，以本体作为数据集成中的虚拟视图，可以方便对集成数据进行统一、快捷的信息查询和数据挖掘服务。在初步引入本体的数据集成系统中，本体被用作数据源语义的直接描述，一般情况下，存在以下三种方法来对数据源进行集成。

a. 单本体方法。采用一个全局本体作为所有信息源的通用语义模型，对应各分布异构的信息源。采用这种方法存在的问题在于全局本体的构建较为困难，缺乏足够的灵活性，不适合领域视角不同的信息源的集成。由Yigal Arens等人开发的SIMS是基于本体的信息集成领域早期研究的重要成果[2]，该系统采用单本体方法，构建了一个全局本体。

b. 多本体方法。在多本体方法中，每一个信息源对应于一个局部本体，不同局部本体间建立映射关系。多本体支持动态性较强的信息源集成，但是因为没有共享的词汇表，不利于本体间的互操作。

c. 混合本体方法。采用了共享词汇表来描述领域内基本术语，局部本体中的每个术语都是源自共享词库的，因此解决了复杂本体间的映射问题，从一定程度上克服了单本体和多本体方法的缺点。

国外随着研究的不断深入，对于本体的应用也越来越充分，本体应用在Web信息集成中最有代表性的项目有(Onto)Agent、Ontobroker和SKC。国内研究的重点在于集成的框架、本体的构建、本体间的映射，并取得了一定的成果，但是在应用方面显得不足，与国外有一定的差距。

1.3 **多Agent技术** Agent是驻留在某一环境下能够自主、灵活地执行动作以满足设计目标的行为实体，因其具有自治性、主动性、代理性、智能性和移动性等特征，所以被广泛应用于分布式信息检索系统中。单个Agent的能力往往受到它的知识库和资源的限制，因而无法适应开放的、动态的分布环境，而多Agent系统是多个Agent的集合，每个Agent拥有不同领域、不同程度的问题求解能力，各个Agent相互通信、协商、协作，可以共同完成单个Agent所不能实现的任务，更适合复杂问题的求解[3]。在多Agent系统中，移动Agent是解决分布式检索的关键，它能代表用户完成特定的任务，携带自身的程序、数据和状态一起移动，代表源节点在目的主机上运行直至得到结果并返回，可大大降低分布式应用中由于中间结果带来的负载和通信开销。

由于企业的信息源是分布存在的，必须依靠网络来传输，并且网络中站点的内容经常变化，这就存在检索结果的实效性问题。传统的分布式计算是基于消息传递和远程调用的，对网络带宽的依赖性强。为解决这一问题，在集成系统中引入多Agent技术，其优势主要体现在减轻网络负载、降低网络延迟和并行性上。目前，关于Agent技术在信息检索中的应用正在积极的研究和开发中，具有代表性的项目是由微电子和计算机协会(MCC)采用多Agent开发的InfoSleuth[4]信息集成系统，这也是作者在本文提出的数据集成模型的基础。

可以看出，上述各种技术在不同的侧重点和应用上解决数据集成，利用XML描述企业信息资源，利用本体实现信息资源语义层集成，利用多Agent优化异构分布式信息资源检索和发现机制，这三种方法的综合应用将是未来企业信息资源集成机制的发展方向。

2 MODIS模型设计

本文采用三级体系结构，在Mediator/Wrapper方式的基础上进行改进，构建了基于Ontology的多Agent企业数据集成系统(MODIS, Multi - agent and Ontology based Data Integration System)的框架模型，如图1所示，用户只需要通过简便的操作就能实现对屏蔽掉分布异构特性的各种数据的统一操作。

2.1 **数据接口层** 该层处于系统的最低层，由各种分布的、异构的数据源(如结构化数据，非结构化数据，文本文件)组成。每个数据源都有自己的包装器，它可以对外呈现一个统一的形式，向上接受来自中介层的子查询，转化为针对特定数据源的查询语句执行访问，向下把数据源中的数据转化为统一的XML格式，并基于局部本体把XML格式的数据映射成基于局部本体的实例文档，把转化后的文档传给中介层。

2.2 **数据集成中介层** 中介层是系统的核心部分，由资源描述模块和查询模块组成，主要实现了对资源层异构数据源的统一查询访问，对用户屏蔽了各数据源的分布性和异构性。

2.2.1 资源描述模块。该模块主要由本体库实现，它存储了利用本体描述语言描述的全局本体和局部本体，以及它们之间的映射关系。利用本体的概念

定义和语义表达能力，提供给用户语义驱动的查询方式，解决不同数据源间的语义冲突。

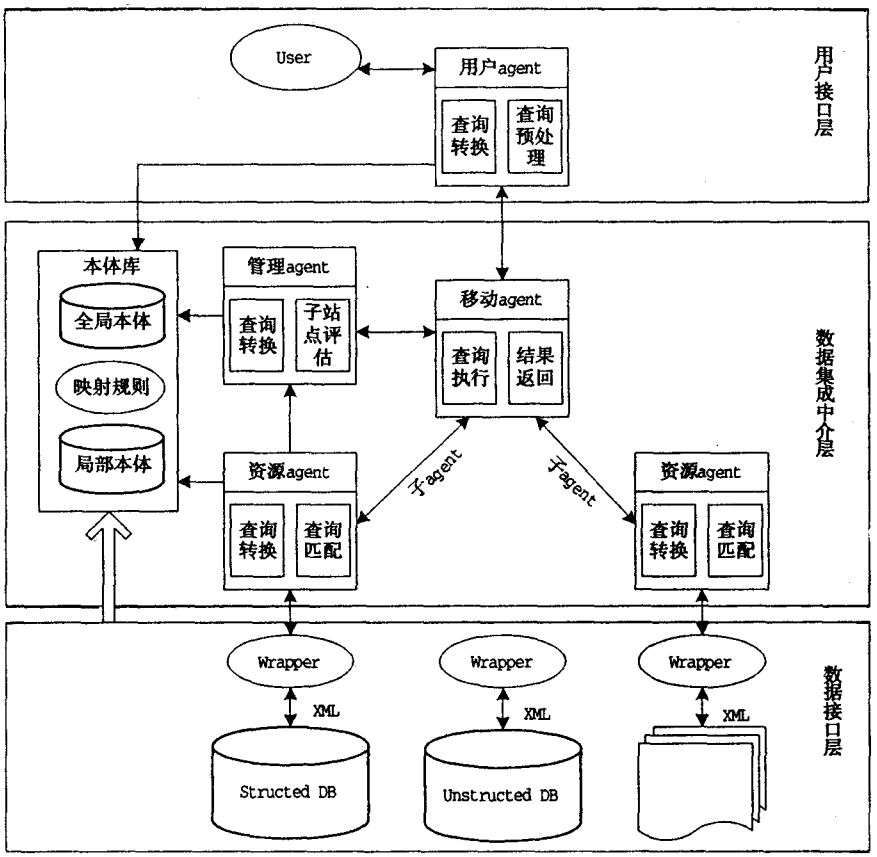

图 1　基于本体的多 Agent 企业数据集成模型

a. 全局本体。全局本体通过精确表达领域内使用的公有概念以及概念的属性和他们之间的关系，使得用户的查询都是建立在全局本体描述的视图上，为查询异构数据源提供了一个语义统一的接口。同时全局本体强大的推理支持能力也为数据集成提供了便利，特别是在确定局部本体概念之间关系上，不但解决了局部本体之间的语义异构性，还满足了他们之间相互查询的请求[5]。

b. 局部本体。局部本体描述具体数据源中的概念和关系。每一个局部本体概念都将对应到全局本体的相关概念上，这个对应的结果就是本体库中存储的映射规则。系统能够根据全局本体与局部本体之间的映射规则将全局的查询重构为对应每一个具体数据源的子查询，当增加新的数据源时，只需要增加新的映射信息，而不需要对原有映射和全局本体进行过多的更改，增加了系统的灵活性。

2.2.2　查询处理模块。该模块主要向用户提供信息处理和检索，此模型中采用了两种 Agent：一种是静态类型，用于各个站点的管理，为移动 Agent 提供资源和各项服务；一种是移动类型，作为移动的智能查询实体往返于各站点之间，与其它静态 Agent 交互，最终完成用户提交的任务。

a. 资源 Agent。资源 Agent 属于静态 Agent，提供从本体概念到本地概念及术语、从全局查询语言到本地查询语言的映射，将查询从通用的查询语言翻译成本地可以理解的语言，并将查询的结果翻译成通用的格式传送给移动 Agent。一个包装器对应于一个资源 Agent，资源 Agent 通过包装器来进行对数据源的控制与访问。

b. 移动 Agent。接受由用户 Agent 传来的检索任务，将子移动 Agent 派往合适的领域，并结合本体对搜索结果进行分类过滤后反馈给用户 Agent。推理控制中心是核心控制模块，它通过交互模块与其他移动 Agent 进行消息通信，并根据信息库和路由表的支持，决定下一步的移动目标。

c. 管理 Agent。管理 Agent 属于静态 Agent，负责模块内的安全，并保证移动 Agent 可以安全地访问模块内的信息以及移动 Agent 之间、移动 Agent 与资源 Agent 间的通信。包括为静态 Agent 和移动 Agent 提供注册服务，为移动 Agent 提供迁移服务，限制模块内移动 Agent 的数量避免模块拥塞，定期通过派遣出去的移动 Agent 通知其他管理 Agent 此领域的存在[6]。

2.3　**应用接口层**　应用接口层由用户 Agent 来完成本层功能，用户 Agent 能产生友好的用户界面，负责和终端用户交互，接收用户的检索请求，借助本体知识，对检索信息进行语义扩展、分类、规范化描述等工作，再提交给数据集成中间件层，并将从中介层返回的 XML 格式的数据转化为相应显示格式展示给终端，此外，用户 Agent 具有自身的知识库和学习机制，可通过学习不断调整自己以适应用户的偏好。

3　系统关键技术

3.1　**提取数据模式**　数据源模式是对各异构数据源的描述，鉴于目前 XML 技术比较成熟，本文所提到的模型将利用它来进行底层数据描述，即利用 XML 定义异构系统间传递数据的结构。这就将集成结构化数据、半结构化数据的问题转化为集成异构 XML 数据的问题，大大降低了问题的复杂性。

XML 和数据库的转换，主要是 XML 模式和关系

模式之间的转换:关系模式到 XML 模式的转换是将关系模式中相关信息(字段信息、数据等)用 XML 文档的层次关系表达出来;XML 模式到关系模式的转换是将规范的 XML 文档中的数据属性、内容部分转换成关系模式中字段信息,并将元素之间的位置关系转换成关系模式中的外键等应用关系。

3.2 **本体及其映射的构建** 本系统从本体的复用性出发,采用混合本体的方法,对各分布异构的数据源建立对应的局部本体,再对系统所要应用的领域进行分析,进而构建全局本体。

3.2.1 本体的构建。局部本体是利用局部数据源的 XML Schema 并依据相应的转化关系构建。首先对各数据源分别进行全面的分析,从数据模式中抽取出概念,经过针对该数据源使用环境的领域知识进行提升,形成最终的局部本体。

全局本体的建立主要是通过对数据源和局部本体的分析,将多个要集成的局部本体中语义上等价的类可以抽象为全局本体中的一个类,类的概念相同的属性被抽象为全局对应类的一个属性,一个类与另一个类概念相等的关系被抽象为全局本体的一个关系[7]。如果只在一个数据源中出现的类,则直接将它的类和属性放在全局本体对应的位置。所提取出的全局概念可使用具有丰富表达能力的 OWL 表示出来,若采用支持 OWL 的本体定义的可视化图形工具 Protégé,可简化本体定义的复杂度。

3.2.2 映射关系的建立。对于不同类型的异构,解决方式也不一样,有些可以在构建本体的过程中消除,有些则需要通过本体与本体、本体与数据源的映射解决。映射不但实现了不同数据源的语义一致,而且为以后本体的共享和重用提供了重要的接口。

局部本体到数据源映射的办法是从本体有向图中的路径映射到数据源的 XPath 路径,本体到数据源的映射也用 OWL 表示。为了完成用户查询,映射层应具备记录数据源的具体信息的功能,包括数据源的逻辑名称到其物理地址的映射等。

在全局本体的构建中,对局部本体分析形成全局本体的过程就是映射逐步明确的过程[8]。每一个局部本体的概念都将它对应到全局本体的一个语义相同的概念上,对于不能直接一一对应的概念,则使用对应功能函数对概念进行合并后再映射。

3.3 **基于本体论技术的多 Agent 协调** 在开放、动态的多 Agent 环境下,协商是实现协同、冲突消解和矛盾处理的重要环节,是系统成功运转的关键。基于此,在 MODIS 中,Agent 通信与协作机制是 Agent 结构设计中必须着重考虑的问题。每个 Agent 可以代表所在站点的资源管理策略以及用户的查询请求参与到协商过程中,在站点资源管理和用户查询请求中寻找到最大的共同利益,Ontology 则充当中介者的作用,共同的本体成为不同资源库的 Agent 之间交互的共同基础,使得查询请求的相互理解成为可能[9]。为了让 Agent 拥有更丰富的知识推理能力,实现异构 Agent 之间进行自动协商,本系统在协商机制中引入本体论的技术。

a. 基于 OWL 的 Agent 通信语言。要让多个 Agent 之间能够进行沟通协商,首先必须要让 Agent 之间具有标准的沟通语言,目前 Agent 普遍采用的通信语言规范是 FIPA ACL。为了使 Agent 对语义有更好的理解性,本文结合 OWL 来改进 Agent 之间的通信语言,应用本体能够很好的公式化表达及其推理而不必考虑应用系统的结构。

b. 构建协商协议本体。为了使 Agent 之间具有相同的语义认知能力,Agent 必须具备相关协商领域的知识。在动态的电子商务环境中,协商协议是协商各方需遵守的约束彼此行为的规则,因此本文将协商协议以本体建模,用来表现 Agent 在协商过程中所具备的相关知识,Agent 只要以协商协议本体作为不同互动规则的共同认知,即可参与到协商活动中,理解到互动规则的描述,并通过规则引擎来得知活动中每一个阶段可以采取的行动,从而进一步实现了系统的自动化和智能化[10]。

4 实例分析

为了更加形象地描述系统流程,本文将举出一个实例。假设某银行有 10 个分行,各分行所用的数据库类型不一定都相同,有 Oracle、Sybase、SQL Server,银行的业务数据分散地存储在这些数据库中。总行通过构建 MODIS 平台,可实时了解各分行的业务情况。现监管部门的负责人提出"2008 年度所有不良贷款信息"的查询请求,系统处理流程如下:

a. 用户 Agent 接受到这个请求,借助本体库中的领域知识,对查询进行预处理,找出检索信息所在的领域,将"不良贷款"扩展为"呆账贷款(贷款利息拖欠逾期 3 年)"、"呆滞贷款(贷款人走死逃亡或经国务院批准)"和"逾期贷款(贷款本息拖欠超过 180 天)",从而将查询转换为基于本体的查询语句。

b. 管理 Agent 在用户模块内初始化一个移动 Agent,此移动 Agent 携带用户 Agent 传来的任务,并将任务分解为各个子任务,对于每一个子任务向管理 Agent 请求能够提供该服务的资源 Agent 的句柄(如 IP 地址、消息格式、接口名称等内容)。管理 Agent 返回符合条件的 Agent 的句柄,并结合本体库的信息,返回子站的优先级列表。

c. 移动 Agent 获得句柄后,在迁移过程产生子移

动 Agent，与之并行工作，其数量可根据网络实际情况进行调整。这些携带有用户任务的移动 Agent 按照一定的路由策略，与各模块中的资源 Agent 进行交互。

d. 资源 Agent 通过包装器将查询转换为对相应分行的数据库查询语句，将执行结果返回给各个移动 Agent，从而实现了对各个分行的异构数据源的访问和控制。

e. 各移动 Agent 通过协商，将结果进行过滤，并结合本体库的语义推理的功能对结果进行分析，按照信贷风险的等级，将结果按次级、可疑和损失分组，返回给用户 Agent，再由用户 Agent 呈现给用户。这样就完成一次 Agent 的协作过程。

在数据集成之前，很难对分支机构的不良贷款进行监控，而只有实现了数据的高度集成，才能够真正对所有企业、所有个人贷款的审批实行有效监控，尤其是对分行反常的数据信息的监控和跟踪，从而降低不良贷款率，实现防范和化解金融风险的目的。

5　结束语

现代企业面对多变的市场环境、先进管理理念和信息技术在企业的应用不断扩展，企业信息资源集成显得尤为重要，未来企业是全面集成的企业，具有协作性、虚拟化、敏捷性、学习型和精良性等特征才能适应未来市场的变化，提高企业竞争力和生存力。针对传统集成方法的各种局限性，本文提出了一个基于本体的多 Agent 企业数据集成模型，该模型的特点是：一方面，引入本体对各分布异构数据源进行描述，并借助本体及相关知识对用户的检索请求进行规范化描述，提高检索系统的语义处理能力，实现了异构数据源语义级的无缝集成。另一方面，利用多 Agent 技术实现数据的分发，优化了异构分布式信息资源检索和发现机制。通过对集成方法和关键技术的研究，为企业信息资源的集成提供了理论和技术支持。本文作者认为下一步的研究重点在于：

a. 增加该模型的实际可操作性，研究基于本体的多 Agent 企业信息资源集成在 Web2.0 环境下的实用性。

b. 在模型中引入处理信息的个性化 Agent，研究以用户为主导、面向决策、服务导向的信息资源集成的实现方案。

c. 研究数据挖掘技术在数据集成中的应用，从大量数据中提取有用信息，为决策者提供有效的数据分析和决策支持。

参 考 文 献

[1] Cruz I F, Xiao H. The Role of Ontologies in Data Integration[J]. Journal of Engineering Intelligent Systems, 2005, 13(4): 12 - 16

[2] 李凌志，张玉婷. 基于本体的信息集成研究[J]. 情报杂志，2008(1):68 - 70

[3] 张云勇. 移动 Agent 及其应用[M]. 北京：清华大学出版社，2002:25 - 31

[4] 史海燕，毕　强. 国外主要信息集成项目介绍与评析[J]. 情报科学，2004，22(7):839 - 844

[5] 刘文斌. 基于本体的信息集成[D]. 南京：南京航空航天大学，2006

[6] 曲卫红. 基于移动 Agent 的分布式信息检索的研究[J]. 现代情报，2006(1):159 - 161

[7] 鱼　滨，郑娅峰. 基于本体的异构数据集成方法及其实现[J]. 计算机应用与软件，2007，24(9):30 - 33

[8] Marie - Christine Rousset, Chantal Reynaud. Knowledge Representation for Information Integration[J]. Information Systems, 2004(29):3 - 12

[9] 叶海军，熊筱芳，姚力文等. 基于本体的多 Agent 交互模型[J]. 计算机与现代化，2008(4):32 - 35

[10] 蒲秋梅. 基于 Ontology 和 Agent 的电子商务协商研究[D]. 武汉：武汉理工大学，2007

（责编：刘武英）

（上接第 114 页）

Knowledge Cereation[J]. California Management Review, 1998, 40(3):40 - 54

[7] Scharmer CO. Self - transcending Knowledge: Sensing and Organizing Around Emerging Opportunities[J]. Journal of Knowledge Management, 2001(5):137 - 150

[8] Nomura T. Design of Ba for Successful Knowledge Management - How Enterprises Should Design the Places of Interaction to Gain Competitive Advantage[J]. Journal of Network and Computer Application, 2002, 25:263 - 278

[9] Malin Brannback. R&D Collaboration: Role of Ba in Knowledge E - creating Networks[J]. Knowledge Management Research & Practice, 2003(1):28 - 38

[10] Knowledge Research Institute, Inc. The Personal Knowledge Evolution Cycle[DB]. http://www.knowledgeresearch.com\articles.htm

[11] 张建华. KM 中的双线知识集成策略[J]. 科学学与科学技术管理，2006(9):103

[12] Borst WN. Construction of Engineering Ontologies[M]. Ph thesis, University of Twenty, Enschede, 1997:48 - 66

[13] Wright R W. The Effects of Tacitness and Tangibility on the Diffusion of Knowledge - based Resources[J]. Academy of Management Proceedings, 1994:52 - 56

[14] Nonaka I, Takeuchi H. The Knowledge - creating Company[M]. Oxford: Oxford University Press, 1995

（责编：王平军）