

基于本体的通用知识网格架构研究

李玉华 卢正鼎 廖振松

(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

摘要: 提出了一个基于本体的通用知识网格(UKG)架构模型,用于在网格环境下创建大规模的分布式知识发现和知识集成系统,并对网格基本的安全机制进行扩展,引入安全策略本体服务和防火墙本体服务,提供较高的安全性.讨论了本体构造的原则和方法,介绍了外汇监管集成本体的设计和实现.给出了基于UKG的外汇监管的应用举例,展示用户在解决问题时如何从UKG服务获益.

关键词: 通用知识网格; 本体; 架构; 安全

中图分类号: TP393.08 **文献标识码:** A **文章编号:** 1671-4512(2006)03-0021-04

Research on the architecture of ontology-based universal knowledge grid

Li Yuhua Lu Zhengding Liao Zhensong

Abstract: An ontology-based grid architecture model was proposed in terms of the universal knowledge grid (UKG) to build large-scale distributed knowledge discovery and integration system on the grid. The high security was obtained by extending basic grid security mechanism and using security policy ontology services and firewall ontology services in this model. The rules and methods of ontology construction were discussed and the design and implementation of the foreign exchange supervision integration ontology were introduced. A practical example of foreign exchange management (FEM) was presented to show the benefits gaining from UKG in solving problems.

Key words: universal knowledge grid; ontology; architecture; security

Li Yuhua Assoc. Prof.; College of Computer Sci. & Tech., Huazhong Univ. of Sci. & Tech., Wuhan 430074, China.

本体^[1]最早是一个哲学上的概念,是客观存在的一个系统的解释或说明. Studer 等认为本体是共享概念模型的明确的形式化规范说明.本体的目标是捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并不同层次的形式化模式上给出这些词汇(术语)和词汇间相互关系的明确定义.

本文提出了一种基于本体的网格架构模型,称为通用知识网格(Universal Knowledge Grid, UKG),可在网格环境下提供高性能的知识发现和知识集成服务,并对网格基本的安全机制进行扩展,提供较高的安全性^[2].

1 通用知识网格的架构模型

UKG(见图1)是定义在基本的计算网格工具集和网格服务基础上的.其主要功能如下:

1.1 智能知识浏览器

当前的 Web 浏览器是为人们浏览 Web 文档而设计的,它只能解释 HTML 标签并以纯文本的文档形式展示.本研究提出一种智能知识浏览器作为用户接口.智能知识浏览器提供一组图形化的工具,用户使用可视化的工具或自然语言创造自己的知识需求.智能知识浏览器可显示机器

收稿日期: 2004-12-16.

作者简介: 李玉华(1968-),女,副教授;武汉,华中科技大学计算机科学与技术学院(430074).

E-mail: yuhua.yy@163.com

基金项目: 国家自然科学基金资助项目(60403027);国家十五重大科技攻关项目(2001BA102A06-11).

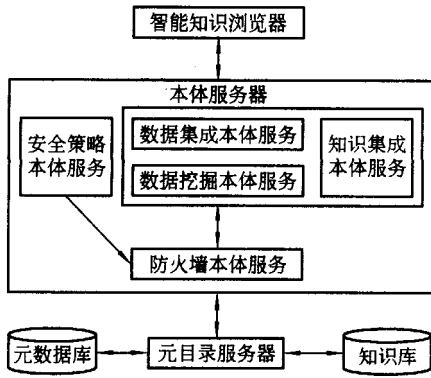


图 1 通用知识网络的架构

可理解的语义并解释显示的内容. 用户可按语义操作资源.

a. 自动推荐服务. 智能知识浏览器使用同义词和分类学来自动引导用户到推荐的搜索项. 例如: 一个人想搜索与“网格”相关的文档, “数据网格”和“知识网格”相关的文档也将被搜索. 关于“网格”的分类学会帮助推荐相关的搜索项.

b. 高级导航. 应用本体提供高级信息和知识导航. 用户通过浏览模型化的概念和它们之间的关系可很容易地访问相关的信息.

c. 推理能力. 内嵌的推理引擎帮助用户搜索和呈现隐含的知识.

1.2 本体服务器

本体服务器是整个架构的核心模块. 它负责本体的管理和查询. 本体服务器提供下面的服务: 数据集成本体服务, 数据挖掘本体服务, 知识集成本体服务, 安全策略本体服务, 防火墙本体服务.

本体服务器使用户不必直接与物理数据源进行交互, 因此用户不必关心实际的物理数据源的形式和位置. 本体服务器也能隔绝任何数据源的变化, 使用户不受数据源变化的影响. 本体由本体服务器进行创建和维护. 本体服务器在创建和维护本体时, 可应用“导入”功能, 就像数据库管理系统的“import”功能一样, 实现本体的可重用. 可从数据库、数据仓库的元数据、已存在的本体中导入应用域概念, 形成目标本体. 在将新的本体集成到本体服务器时可通过限制或约束等方法以消除冗余, 并进行一致性检查.

1.2.1 数据集成本体服务

数据集成本体服务描述 Web 文档的语义, 在半结构化数据和结构化数据库之间架起一座桥梁, 帮助进行数据清洗和数据准备, 进行异构数据源集成.

1.2.2 数据挖掘本体服务

数据挖掘本体服务可给领域专家提供一个解

决特定问题的不同种数据挖掘功能、算法和软件参考模型. 图 2 示出了数据挖掘本体的功能类. 图 3 示出了数据挖掘本体的部分算法类层次.

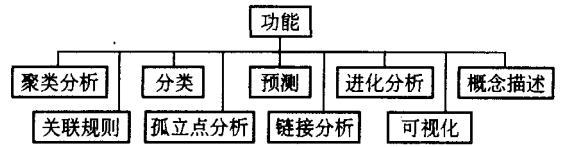


图 2 数据挖掘本体的功能类

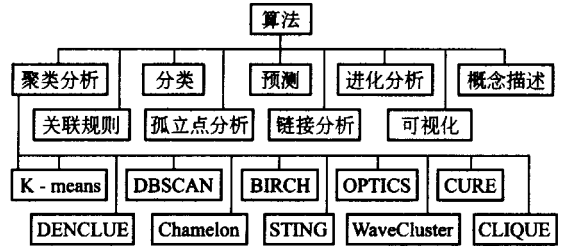


图 3 数据挖掘本体的部分算法类层次

1.2.3 知识集成本体服务

知识集成本体服务包括以下功能: 维护公共有效本体集; 负责查询条目之间的关系; 方便不同知识库之间的通信和知识交换; 不同本体之间的翻译和映射; 对数据挖掘本体的挖掘的模式进行评分, 并把有效的知识补充到知识库; 提供按需服务以支持问题解决和决策支持; 集成不同水平的知识源(例如概念、公理、规则和方法), 以支持问题的分析和解决.

1.2.4 安全策略本体服务

在网格架构下, 一个实际的问题是协调的资源共享以及在动态、异构的虚拟组织上的问题解决. 这种共享需要高度的可控性, 在资源的提供者和使用之间明确地定义共享什么资源、谁允许使用以及什么情况下才共享资源. 使用这些安全的、协同共享策略的个人或者团体组成的动态集合称为虚拟组织.

对认证、组成员和类似相关的许多问题, 在虚拟组织中成员之间的关系, 代表了存在于那些成员及其父组织之间关系的覆盖. 这种覆盖不仅存在于涉及信任、安全机制中, 也存在于那些父组织某些地方的策略中. 在网络安全技术中, 虚拟组织用作参与到某一特定组织或团体的实体的桥梁.

安全策略本体服务包括域安全策略和虚拟组织的安全策略. 域安全策略涉及到资源访问、服务质量、可接受的信任、内部的角色以及内部的用户权利等. 虚拟组织的安全策略有成员规则、用户角色、用户权利、争论解决等策略.

1.2.5 防火墙本体服务

防火墙本体服务接受描述任务需求的任务配

置文件,访问应用实例所属的网格共同体和防火墙要保护的资源的那些域的安全策略,在语义水平对呈现给它的相关的信息和策略进行推理.只有那些与安全策略一致的应用请求语义防火墙才允许其通过,而不管它使用什么样的协议和接口.

1.3 元数据目录服务器

元数据目录服务器维护 U KG 中描述所有数据、工具和知识的元数据.元数据信息用 XML 文档表示,存储在元数据库和知识库中.

元数据库存储着各种数据源、数据集成和数据挖掘工具、算法的元数据.

知识库存储应用领域收集的各种知识及数据挖掘的结果,如学习的模型和发现的模式等.

元数据目录服务是 U KG 的元目录.它识别特定的发现和注册服务,也提供一种集成机制,集成多个知识库到一个高级知识库,作为一个集成单元服务于某一特定领域的功能目的.在满足这些需求时,元数据目录服务提供如下丰富的管理知识库集的功能:协同选择,一个用户可同时选择一个或多个目录服务器维护的知识库的知识;知识传播,一个用户或虚拟角色可通过智能代理把知识传播到多个目录服务器;知识管理,知识流网络帮助在动态的虚拟组织中完成有效的知识共享^[3];集成的知识库,一个或多个目录服务器的知识库子集可联合组成高级的集成知识库.

2 本体的构造

出于对各自问题域和具体工程的考虑,构造本体的过程各不相同.目前没有一个标准的本体的构造方法.最有影响的是 Gruber 在 1995 年提出的 5 条规则^[4].为了 Web 上应用程序使用方便,需要有一个通用的标准语言来表示本体,就像 XML 作为标准的数据交换语言一样.目前正在开发中的语言有:SHOE, OML, XOL, Riboweb, RDFS, OIL 和 OWL.在本体的编码中采用 OWL (Web Ontology Language) 语言.

创建本体的过程如下:确定目标域的目的和范围;获取关键的概念和概念间的关系;本体编码;与已存在的本体集成,即消除冗余并检查一致性.

对于外汇监管知识域而言,要进行综合分析需集成多个内部业务系统和外部数据源的数据,而且各个业务系统的数据分布是很复杂的,数据分布变化很大:a. 单数据源,数据集中存放在一个物理数据源.b. 水平分布的联邦数据源.对于

一个给定问题域的部分数据和语义的部分数据分布在不同的物理地点.这种数据分布要求提供高性能的服务,查询所有的数据源并将结果组成一个完整的数据结果集.c. 垂直分布的联邦数据源,这种数据源有确定每个事务的公共属性,但其他一些属性分布在多个物理地点.d. 异构的数据源,这种数据源在结构和设计上都是不同的,即对于相同语义的属性的名字、描述、应用或数据类型都是不同的.

在与领域决策支持专家一起进行深入分析的基础上构建一个数据集成本体,来实现各种信息资源的集成.外汇交易的时间、地区、交易分类、业务类别、交易金额是外汇交易决策分析的几个很重要的特征.

时间分为三个层次:年、季度、月;地区分为省(或直辖市)、市(或直辖市所属区)、县(或区)三级;度量维包括收入、支出金额,及对应的同比、环比和比重等;业务系统包括国际收支、结售汇、外汇账户、外债等业务系统和贸易进出口等外部数据源.交易代码维是最复杂的,它需要将各个业务系统的交易代码整合在一起,集成一个统一的各方认可的交易代码标准.每个维映射成数据集成本体的一部分.下面以地区为例介绍本体的构造和实现.

给出四种不同的外汇交易区域:省(或直辖市),例如北京市,湖北省;市(或直辖市所属区县),例如北京市海淀区,北京市昌平区,湖北省武汉市;县(或区),例如湖北省武汉市武昌区,湖北省黄冈市红安县.而且必须为不同外汇交易地域之间的关系建模:市是省的一部分,如武汉市在湖北省;县是市的一部分,如红安县在湖北省黄冈市.

3 一个应用实例

下面讨论一个简单的基于 U KG 服务在外汇监管域中用户要对某一地区的外汇交易情况进行分析检测的实例,有益于展示用户在解决问题时如何从 U KG 服务获益.

外汇监管域可由几个虚拟组织(金融机构虚拟组织、外汇管理虚拟组织、海关虚拟组织、经贸虚拟组织和执法机构虚拟组织)组成,每个虚拟组织有它自己的安全策略,外汇监管域有域的安全策略.

本体开发工具采用 Stanford 的 protégé³.实验采用本体开发工具进行本体输入和管理, Jena

2.3 本体解析器进行本体的解析和开发推理程序.本体的存取和查询,是项目中一个非常基础的工作之一,目前采用 Mysql 作为数据库.以 Jena 作为存取引擎.主要利用 Jena 和 protégé 的 java API, Jbuilder 开发工具开发. 网络开发平台采用 Globus Toolkits 3.

在这个应用实例中,开发了一个外汇监管域的数据集成本体 OntFC,用来集成各个异构的数据源;一个基于角色的访问控制域安全策略本体 OntRBAC 和一个数据挖掘本体 OntDM.基于本体的异构数据源的集成采用混合本体的方法^[5],每个数据源定义一个局部本体,然后各个局部本体再通过本体映射和上下文集成到全局域本体 OntFC,这种方法可屏蔽数据源,向用户提供统一的数据视图,又能方便地增加数据源.

在实例的情节中, U KG 执行的步骤如图 4 所示,图中顺序号说明如下:

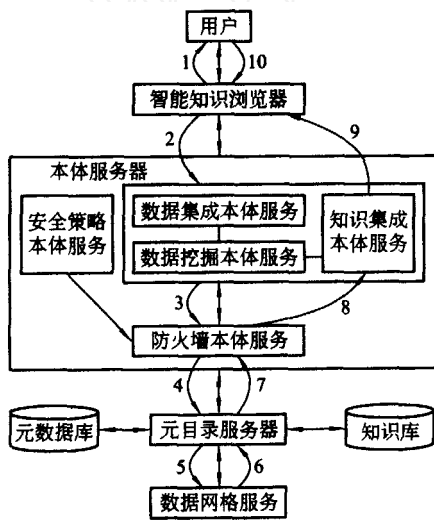


图 4 U KG 的一个应用实例

1 利用智能知识浏览器,基于本体 OntFC 提供的统一的数据视图,用户选择“湖北省”的交易信息,选择进行孤立点检测,创建监管任务,了解其当前的交易是否正常;

2 任务被传送到本体服务器;

3 本体服务器的数据集成本体 OntFC、数据挖掘本体 OntDM,解释任务的语义和资源需求,生成任务需求配置文件,包括所涉及的数据源,数

据挖掘算法等信息,传送给防火墙本体服务;

4 防火墙本体服务根据任务需求配置文件和安全策略本体 OntRBAC 的安全策略,进行推理判断,看用户是否具有访问数据源和执行数据挖掘任务的相应权限,将许可的需求配置文件传给元目录服务器;

5 元目录服务器定位所需的网格节点的交易数据库和孤立点检测算法等,传送元数据到数据网格服务;

6 数据网格服务执行数据挖掘任务,结果返回元目录服务器;

7 元目录服务器将挖掘结果返回防火墙本体服务;

8 防火墙本体服务根据返回结果及安全策略本体服务的安全策略,进行推理判断,将许可的结果传给知识集成本体服务;知识集成本体服务进行模式评价,有效的模式通过元目录服务集成到知识库;

9 有效的模式返回浏览器;

10 用户通过浏览器获得数据挖掘的结果,了解有哪些交易异常情况发生,采取措施进行监管.

参 考 文 献

[1] Gruber T. A translation approach to portable ontology specifications [J]. Knowledge Acquisitions, 1993(5): 199-220.

[2] Surridge M, Upstill C. Grid security: lessons for peer-to-peer systems [C] Peer-to-Peer Computing. Linköping: IEEE Computer Society, 2003: 2-6.

[3] Zhuge Hai. A knowledge flow model for peer to peer team knowledge sharing and management [J]. Expert Systems with Applications, 2002, 23(1): 23-30.

[4] 邓志鸿,唐世渭,张 铭,等. Ontology 研究综述 [J]. 北京大学学报:自然科学版, 2002, 38(5): 729-738.

[5] Wache H, Voögele T, Visser U, et al. Ontology-based integration of information-a survey of existing approaches [C] Bernhard Hebel, ed. Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence. Washington: Morgan Kaufmann, 2001: 108-117.