

# 基于本体和云服务的交通异构数据集成方法研究<sup>①</sup>

阳其凯<sup>②</sup> 王 坚<sup>③</sup> 凌卫青

(同济大学 CIMS 研究中心 上海 201804)

**摘要** 考虑到智能交通系统(ITS)建设的需要,研究了交通异构数据集成的问题,重点研究了系统异构与语义异构的交通异构数据集成。在分析基于数据维、服务维、流程维的交通异构数据集成模式的基础上,引入本体与云服务技术,提出了以服务封装、语义检索、流程驱动、服务组合、面向用户为导向的交通异构数据集成方法。该方法运用云服务打破了异构数据的系统壁垒,引入本体解决了异构数据的语义检索问题。为了验证该集成方法的有效性,构建了大城市交通异构数据云服务体系,该系统的示范为目前交通领域异构数据集成问题提供了解决方案,为我国智能交通系统建设提供了借鉴与导向。

**关键词** 交通异构数据, 数据集成, 语义网, 本体, 云服务

## 0 引言

进入 21 世纪以来, 我国交通信息化建设得到了快速发展, 特别是智能交通系统(intelligent transport systems, ITS)的建设与推广<sup>[1]</sup>, 但是也存在不少问题。由于交通领域内各单位在系统设计与实施时缺乏统一的规划和协调, 交通信息系统具有深度异构特征, 系统间互操作性差, 另外, 由交通领域的特点所决定, 大城市构建各种交通管理系统涉及到不同的交通数据, 数据模型和数据的存储格式也不相同<sup>[2]</sup>。因而, 如何保证单个的交通系统内数据的一致性、准确性、无歧义性和交通子系统之间信息的交互与共享, 实现综合、全面、深入的异构数据集成, 是 ITS 建设的关键。

交通异构数据集成一直是智能交通系统建设需要考虑的重要问题。国内外学者对交通异构数据的集成开展了部分研究, 大体将交通异构数据的集成分为基于中间件的异构数据集成<sup>[3]</sup>、基于 Web 服务的异构数据集成<sup>[4]</sup>、基于本体的异构数据集成<sup>[5-8]</sup>、

基于数据仓库的异构数据集成等。如 Li 等<sup>[5]</sup>提出了基于本体的 ITS 数据集成框架, 该框架将不同数据源的异构数据映射为具有统一语义描述的本体数据, 最大程度地减小 ITS 内部对共享数据的二义性理解, 并在文献[6]中将基于本体的思想应用于城市快速路网的数据集成中。曹妍<sup>[7]</sup>提出了基于本体的 ITS 集成框架 OITSIF, 该框架通过智能交通领域本体的构建, 实现了关系数据库与本体的映射, 解决了 ITS 异构数据的语义集成问题, 同时实现 ITS 语义信息检索和数据集成的应用。王坚等<sup>[8]</sup>提出了基于本体的交通状态统一语义理解机制, 对交通数据语义集成做了积极探索。但是上述的研究缺乏从数据管理、数据应用方面对异构数据集成做出全面的、系统的阐述。本研究在基于数据维、服务维、流程维的交通异构数据集成模式<sup>[9]</sup>的基础上, 提出了以服务封装、语义检索、流程驱动、服务组合、面向用户为导向的交通异构数据集成方法, 该方法运用云服务打破了异构数据的系统壁垒, 通过引入本体解决了异构数据的语义异构问题。

① 国家自然科学基金(71273188)资助项目。

② 男, 1991 年生, 硕士生; 研究方向: 云计算, 语义网; E-mail: yangqikai5235@163.com

③ 通讯作者, E-mail: jwang@tongji.edu.cn

(收稿日期:2015-02-02)

## 1 交通数据异构性分析

交通数据的异构性包含多方面,主要分为四类:结构异构、系统异构、语法异构与语义异构,本文重点研究系统异构与语义异构的交通异构数据集成问题。

**系统异构:**是指由多个交通子系统带来的硬件、系统软件(如操作系统)和通信系统之间的差异,同时也包含数据的存储管理机制不同,如采用不同的数据库进行存储(DB2、Oracle、SQL Server、MySQL等)。

**语义异构:**产生语义异构的根源在于不同数据源的设计者和实施者对于交通领域知识与数据的看待角度不完全相同。主要包括以下几个方面:(1)同一个术语在不同的信息源表示不同的含义(同名异义);(2)在不同信息源中,许多不同的术语表示相同的概念(异名同义);(3)不同的信息源中,使用不同的结构来表示相同(或相似)的信息;(4)各信息源中的概念之间存在诸多的关联,但由于各信息源的分布自治性,这种隐含的联系不能体现出来;

(5)不同信息源对同一事物或对象的抽象程度不同;(6)不同信息源对概念的形式化描述方法不同,这就造成了信息表达上的异构。

## 2 交通异构数据集成方法

文献[9]提出了基于数据维、服务维、流程维的智能交通异构数据集成化管理体系架构,本文对该集成模式进行深入的阐述。如图1所示,该集成模式的数据维描述面向交通领域的多源多维异构数据(如视频数据、GPS数据、线圈数据等),服务维描述交通系统的功能性以及数据服务(交通系统的功能性包括路况信息、交通资讯、交通管制等,这些粗粒度的服务是由更小粒度的服务或数据构成),流程维描述面向智能交通的业务过程。根据交通领域的细化程度,流程分为通用流程、专有流程和特定流程三个层次。通用流程即一般流程;专有流程是对某一特定的事件(如虹桥枢纽交通疏散)而制定的,也就说专有流程是面向事件的;特定流程是针对某一行业与领域(如交通、消防)而定制的,它是面向行业的。

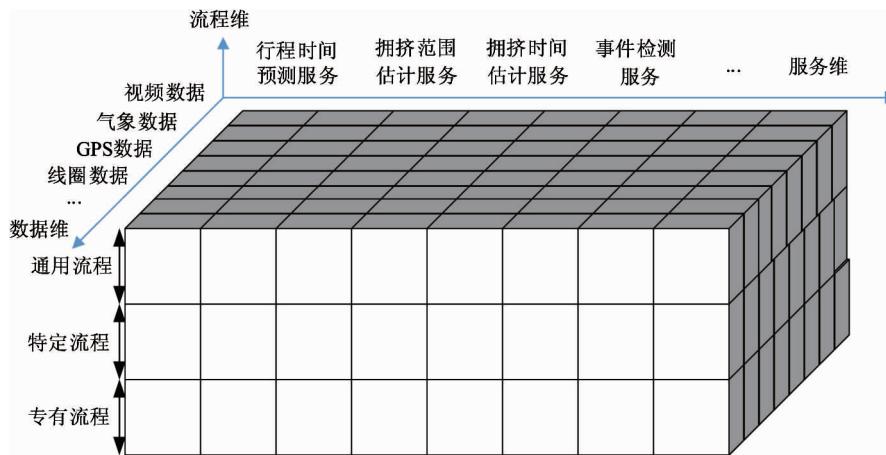


图1 交通异构数据集成模式

该集成模式的三个维度之间的耦合关系可以揭示数据、服务、流程三者之间实现异构数据集成的内在机理。面向数据维-服务维,研究服务与数据之间的关联以及各类服务对不同数据的调用需求和影响;面向流程维-数据维,研究不同层次业务流程对数据结构的细化程度;面向服务维-流程维,研究根

据通用流程、专有流程和特定流程建立交通领域的业务过程和集成交通领域服务,以适应各项交通服务及数据服务的快速组合与协同工作。

通过对基于数据维、服务维、流程维的交通异构数据集成模式的分析,本文给出了从数据维到服务维、服务维到流程维的交通异构数据集成的方

法——以服务封装、语义检索、流程驱动、服务组合、面向用户为导向的交通异构数据集成方法(以下简称本文方法)。

智能交通面向不同的群体、不同的领域,每一特定的领域具体活动都需特定的流程,每一种流程都需串接不同的服务,每一种服务可能需要不同的数据源作为支撑。本文方法以异构数据为基础,以数据的服务化为依托,以流程驱动为导向,进行数据服务的组合,从而实现异构数据的集成,进而满足用户多样化、个性化的智能交通信息业务需求。

本文方法涉及服务封装、语义检索、流程驱动、服务组合及面向用户的问题。服务封装是指利用云服务技术将异构数据封装为服务发布至云端,打破异构数据的系统壁垒。语义检索是指对异构数据服务进行本体建模,利用本体丰富的语义表达能力,解决异构数据服务的语义异构问题。流程驱动与服务组合是指利用工作流技术对异构数据服务进行组合,生成更大粒度的服务或者应用。面向用户是指异构数据集成的最终目的是面向用户,这里的用户可以包含三类,一类是利用语义检索进行异构数据服务查询的普通用户,另一类用户利用检索模块得到相关服务,在选择好流程模板之后,将服务推入流程,生成更大粒度的服务或者应用;第三类用户是应用的使用者,而这些应用是建立在异构数据集成的基础之上的。简单地说,可以将以上关键点归结为 4 步:(1)交通异构数据云服务化,负责异构数据服务的封装发布;(2)交通异构数据服务本体建模,使得领域概念以及概念之间的关系在语义的层次得到统一;(3)实现服务语义检索,向异构数据服务请求者与使用者提供语义信息检索接口;(4)实现面向服务的流程驱动机制。

## 2.1 交通异构数据云服务化

### 2.1.1 交通异构数据云服务体系结构

图 2 所示为交通异构数据云服务的体系结构,它借鉴了现有的 Web 服务的体系架构和相关技术与标准,涉及了三种角色(云服务请求者、云服务提供者、云服务中心)与三种操作(云服务发布、云服务发现、云服务交互)。

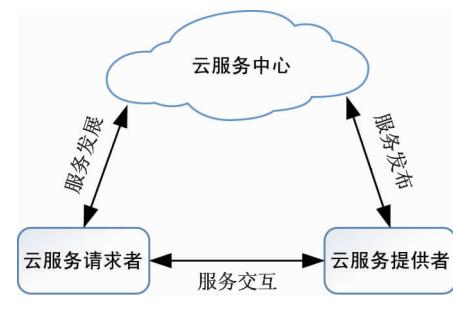


图 2 云服务体系结构

**云服务提供者:** 云服务提供者是异构数据云服务的开发者和所有者,它需要对异构数据进行包装、处理,最后以服务的形式发布到云端;同时也需要提供服务的描述(包括语法层面,也可以含语义层面);服务发布到云端后,生成 Web 服务描述语言(WSDL)文档。

**云服务中心:** 在传统的 Web 服务体系架构,云服务中心角色为通用描述、发现与集成(universal description, discovery and integration, UDDI)服务。云服务中心与 Web 服务 UDDI 注册库并非等效,该部分可以通过 UDDI 技术实现,也可通过其他技术实现。云服务中心提供一个供云服务提供者发布异构数据云服务的 WSDL 描述的平台,云服务请求者可在此平台查找所需要的异构数据云服务。云服务中心同时可以包含记录服务描述信息的服务描述数据库。

**云服务请求者:** 云服务请求者是服务的消费者、使用者,也是调用服务的用户。它向云服务中心发送请求,检索相关的服务,云服务中心执行请求者的请求信息,返回结果。云服务请求者通过得到的 WSDL 文档查看服务的调用方法,根据其调用方法绑定服务,调用服务的功能,完成服务提供者和服务请求者之间的交互过程。

### 2.1.2 交通异构数据云服务的封装与发布

#### (1) 交通异构数据的封装

交通异构数据的封装的关键在于交通异构数据源的封装模块的构建。封装模块的主要作用在于从数据源获取数据和返回可扩展标记语言(XML)格式的数据。不同的数据源进行封装时,需要不同的封装模块。封装模块的结构如图 3 所示。

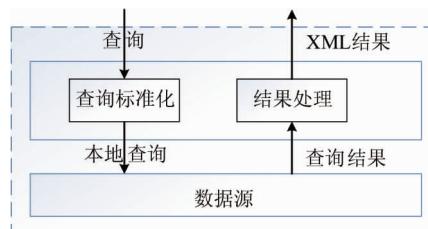


图3 封装模块的结构

封装模块主要由两部分组成,一是查询标准化模块,二是结果处理模块。查询标准化模块主要将来自用户的查询转化为对应异构数据源的本地查询句,利用标准化的本地查询语句在对应的本地数据源上执行。因此,查询标准化模块在理论上应该支持多种本地查询化语言(如结构化查询语言(SQL)、MYSQL等)。结果处理模块是对本地的查询结果进行处理,转化为统一的、标准的XML形式的结果集。

### (2) 交通异构数据服务的发布

在进行交通异构数据的封装之后,需要对各个封装模块进行服务化,使用Web服务描述语言(WSDL)文件描述这个服务,并在云服务中心进行注册。本文利用Axis2<sup>[10]</sup>进行异构数据的云服务发布。封装模块的类文件提供三个接口方法:一个主接口方法public String Core (Core方法),两个辅助接口方法即public ResultSet Standardization (Standardization) 和 public String Result (Result方法)。Core方法的输入为用户查询,输出为标准的可扩展标记语言(XML)结果集合,它顺序执行了Standardization方法和Result方法。Standardization方法的功能是解析用户查询语句,并执行标准化的本地查询语句,返回查询结果集。而Result方法是对本地查询结果集合进行处理,返回XML格式的数据。

## 2.2 交通异构数据服务本体建模

本体建模是实现语义检索的基础。本文采用单本体方法(也称全局本体方法),即所有的数据源都和全局本体相关,通过建立交通异构数据服务领域本体,使得领域概念以及概念之间的关系在语义的层次得到统一,进而解决其语义异构集成问题。

### 2.2.1 交通异构数据服务的本体模型

交通异构数据服务本体模型(Traffic Heteroge-

neous Data Service Ontology Model, THDSOM) = (C, P, R, I, A)是一个五元组,其中:C 表示交通异构数据服务领域相关的概念集,如快速路数据服务、曹安公路线圈数据服务都属于交通异构数据服务领域的概念集的元素;P 表示属性集,是对概念集中元素的特征以及性质进行的描述,如快速路数据服务所具有的特点、性质等;R 表示交通异构数据服务领域概念与概念的关系,如曹安公路线圈数据服务与曹安公路嘉松路路口线圈数据服务是一种包含关系;I 表示概念所覆盖所有实例的集合;A 表示公理集合,用于约束交通异构数据服务领域中概念、关系的一阶逻辑谓词集合。

### 2.2.2 交通异构数据服务的本体建模

交通异构数据服务的本体建模需要解决三个主要的问题:定义交通异构数据服务类的层次结构;定义交通异构数据服务类的属性;定义交通异构数据服务类或实例的语义关系。

(1) 交通异构数据服务分为静态数据服务和动态数据服务两大类。静态数据服务包括道路交通网络基础信息(如道路等级、长度、收费、立交连接方式等)、道路基础地理信息(如路网分布、功能小区的划分、交叉口的布局等)、车辆保有量信息(包括分区域、时间、不同车种车辆保有量信息等)及交通管理信息(如单向行驶、禁止左转、限制进入等)。动态数据服务主要包括交通流数据服务(占有率、车流密度、流量、运行速度、车头间距、自由流车速等)、环境状况信息(降雨强度、结冰面积率、降雪量、气温、风速、可视距离、能见度等)、交通事件(交通事故、施工、道路阻塞、交通管制等)<sup>[11]</sup>。图4所示为本体的等级结构。

(2) 交通异构数据服务类的属性可以参照一般服务的属性,如服务的名称、服务ID、服务的定义、服务的发布者、服务的隶属(如归属于什么系统)、服务的输入描述、服务的输出描述等。

(3) 交通异构数据服务类或实例之间的关系是指本体中不同类或实例之间的联系。下面列举几种比较常见的解决语义关系的术语与机制:不同的交通异构数据服务使用不同术语表达同一概念,可以通过网络本体语言(Web ontology language, OWL)



图 4 本体的等级结构

的语义表达机制 owl:equivalentClass 解决；类的外延异构，类的外延由属于该类的实例或个体集合定义，集合间的关系有相交（owl:intersectionOf）、不相交（owl:disjointWith）、子类（rdfs:subClassof）等；与类之间的关系一样，属性之间同样存在属性等价（owl:equivalentProperty）、继承关系（rdfs:subPropertyOf）等。

下面以 owl:equivalentClass 解决异名同义语义异构为例来说明。如 FiveMinutesStepFlow 为 A 系统下五分钟交通量表，FiveMinutesStepTrafficFlow 为 B 系统下五分钟交通流量表，将其分别封装为 FiveMinutesStepFlow 数据服务与 FiveMinutesStepTrafficFlow 数据服务，如图 5 所示为 FiveMinutesStepTrafficFlow 与 FiveMinutesStepFlow 的本体模型。

### 2.3 基于 Jena 的语义检索

本文提出的交通异构数据集成方法语义检索模块是基于 Jena 搜索引擎实现的。Jena 是用来构建语义网应用的 Java 框架，它为资源描述框架（RDF）、RDFS、OWL 提供一个程序开发环境<sup>[12]</sup>。

Jena 包括：对 RDF 文件和模型进行处理的 RDF API，

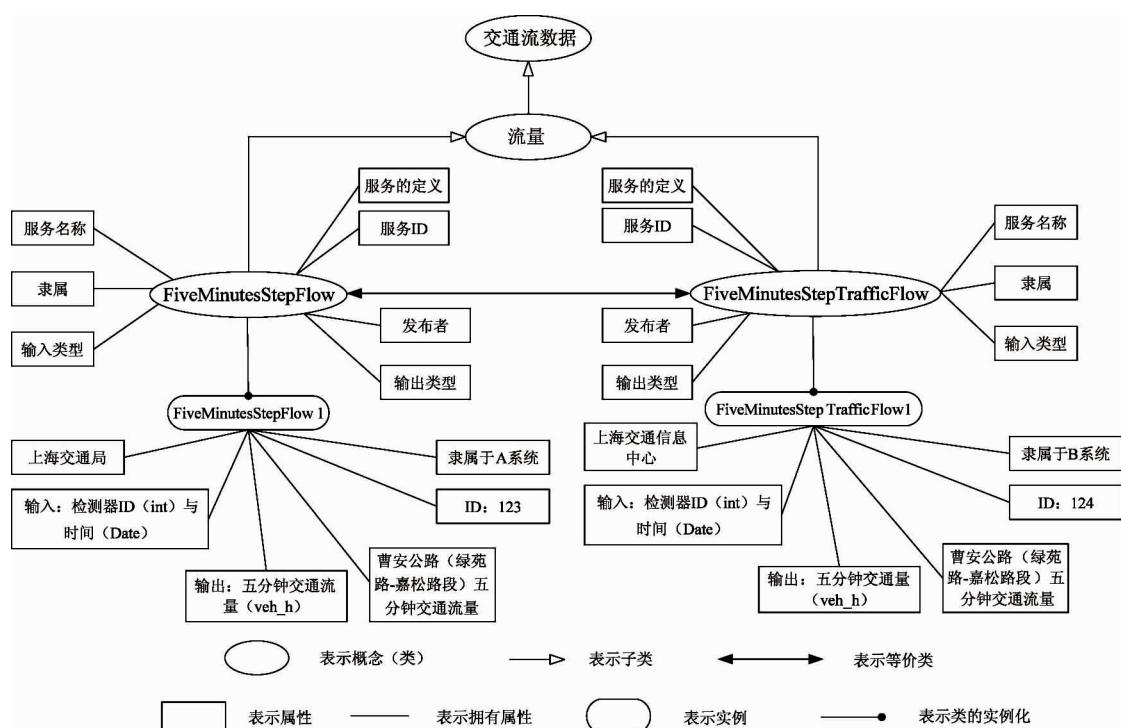


图 5 FiveMinutesStepFlow 与 FiveMinutesStepTrafficFlow

对 RDF、RDFS、OWL 文件进行解析的解析器,用于检索过程推理的推理机子系统,本体子系统,以及用于对本体模型中的语义数据进行检索的简单协议和 RDF 查询语言(Simple Protocol and RDF Query Language, SPARQL)。

图 6 所示为语义检索的一般流程是:获取服务请求者的输入语句;对查询语句进行预处理;在预处理的基础之上将其转化为标准的查询语句 SPARQL,或者在此之前对其进行语义扩展(如通过扩展概念,增加查询关键词与本体中相关知识的映射粒度);标准化的检索语句经 Jena 引擎之后,返回语义查询结果;在将结果呈现给用户之前可做适当的处理。

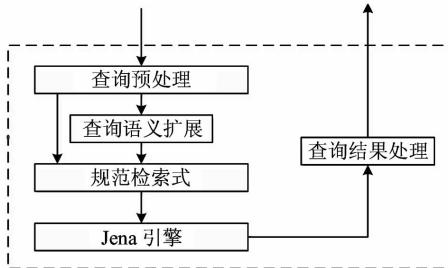


图 6 语义检索一般流程

其中,基于 Jena 的语义检索实施流程如下:(1)建立本体模型,读取本体文件:首先建立一个本体模型(OntModel),本体模型提供处理本体数据的功能。Jena 通过 com.hp.hpl.jena.rdf.model 包中的 ModelFactory 创建本体模型。(2)生成查询语句,执行查询:通过字符串拼接的方式生成 SPARQL 查询语句,最后通过 com.hp.hpl.jena.query 包中的类来创建和执行 SPARQL 查询。(3)查询结果处理:查询结果为 ResultSet 类型,ResultSet 支持在查询返回的每个 QuerySolution 上进行迭代。

## 2.4 面向服务的流程驱动机制

本文采用 JBPM(Java Business Process Management, 精备流程管理)<sup>[13]</sup>作为流程语言框架,基于 JBPM 的服务整合机制包含以下要点:明确交通异构数据(服务)与交通应用服务(包含支撑应用的业务流程)的组织形式,以及流程的通用实施过程。

(1) 交通异构数据与交通应用服务(业务流程)的组织形式

图 7 所示为交通异构数据与交通应用服务的组织形式,它揭示了以面向服务的流程驱动机制进行

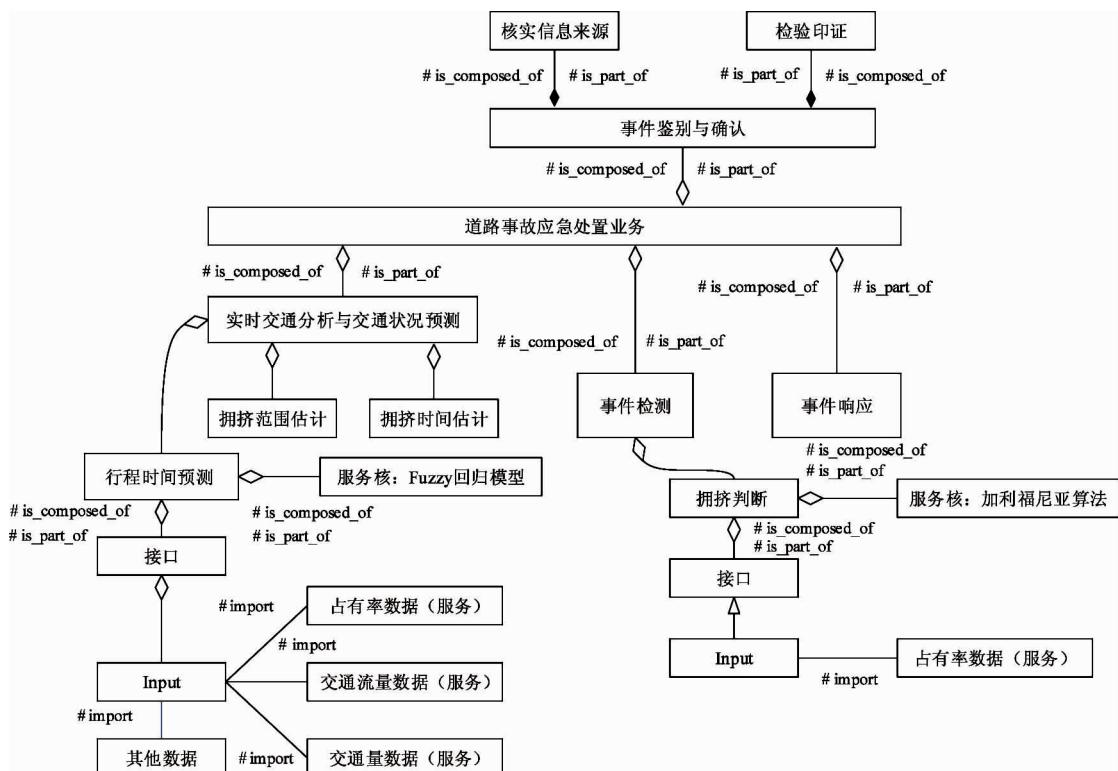


图 7 交通异构数据与交通应用服务的组织形式

异构数据集成的原理。图中为道路事故应急处置业务,按照处置的一般业务流程集成了实时交通分析与交通状况预测服务、事件检测服务、事件鉴别与确认服务、事件响应服务。在事件检测服务中包含拥挤判断服务,拥挤判断服务可以按照从数据(如占有率数据)到模型(如加利福尼亚算法)的一般流程进行服务的组织。

#### (2) 流程的通用实施过程

流程的通用实施过程包括流程的设计与发布、服务的搜索与绑定。流程的设计主要由流程设计人员根据特定的业务流程进行流程的编制,既可以包含一般的流程,也可以包含流程模板;服务的搜索基于语义实现,服务的绑定是指将特定的服务与流程节点进行绑定,这里既包含静态绑定也包含基于流程模板的动态绑定。

##### (i) 流程定义

在进行流程的设计时,主要考虑的是流程模板的设计。流程模板确定了工作流程的基本框架,但是其中的具体活动无法事先确定。在运行时,各个活动节点均与具体的服务绑定。MyEclipse 中可以集成 JBPM 的图形化流程设计器(GPD)插件,开发人员和业务人员利用 GPD 可以对业务流程进行可视化建模,建模的结果以格式化.xml 文件形式输出形成流程定义文件 process\_name.jpdl.xml。如图 8 所示为交通拥挤监测通用流程模板。



图 8 交通拥挤监测通用流程模板

##### (ii) 流程部署

流程模板定义好后,相关人员可以对流程定义文件进行部署,即将其存储到数据库中。流程部署方式有多种,一种常见的方式是利用 JBPM 的插件来完成,但是使用范围有限。由于本文所阐述的流程为系统内置,所以使用系统自动部署(或管理员部署),将打包的 Zip 文件(包含 jpdl.xml 和 png)进行上传。

##### (iii) 流程启动

流程部署完毕之后即可启动流程。启动流程

时,必须创建一个流程实例,流程实例(ProcessInstance)是基于流程模板产生的,一个流程模板可以产生多个流程实例,但是一个流程实例只能有一个流程模板。当一个流程实例创建出来时,一个 Token 也就被创建,此时的 Token 指向开始节点。

#### (iv) 流程执行与服务动态绑定

流程的执行就是流程中每个节点都正确的执行。当流程引擎进入任务节点时,首先创建任务实例,任务完成后离开当前的任务节点。任务在执行时根据流程实例 id 号获取该流程实例对应流程的服务。节点任务执行完毕,相应的 Web 服务执行返回结果,任务结束,保存相应的结果到数据库中。

## 3 应用

基于本文提出的交通异构数据集成方法,设计并开发了大城市交通异构数据云服务系统(以下简称 THDCSP V1.0),THDCSP V1.0 中包含的两个子系统分别是云计算体系结构中 PaaS(Platform as a Service)层与 SaaS(Software as a Service)层的实现,其中 PaaS 模式的交通信息服务系统主要面向交通平台开发人员,在 PaaS 平台上进行数据与服务的发布,为交通的应用提供新的服务功能,同时 PaaS 负责对平台的服务进行统一的管理;SaaS 模式的交通信息服务系统主要面向 SaaS 服务或者应用的开发者以及使用者,开发主要在平台上通过语义检索相关的服务,根据流程组合成更大粒度的服务或者应用,而使用者根据自己的需求进行服务(应用)的定制与使用。如图 9 为 THDCSP V1.0 的系统框架。

THDCSP V1.0 系统框架为五层架构体系,它包含数据层(Data Layer)、服务层(Service Layer)、本体层(Ontology Layer)、流程层(Flow Layer)与用户层(User Layer)。数据层体现了交通异构数据集成模式的数据维,其主要功能是提供交通异构数据源,异构数据源分散于各个单位部门、各个子系统内部,也可以为同一系统内部的数据源;服务层体现了交通异构数据集成模式的服务维,体现了本文方法的“服务封装”的思想,该层将数据封装成 Web 服务或云服务;本体层构建在服务层之上,体现了本文方法的“语义理解”的思想,它包含了本体管理模块(包

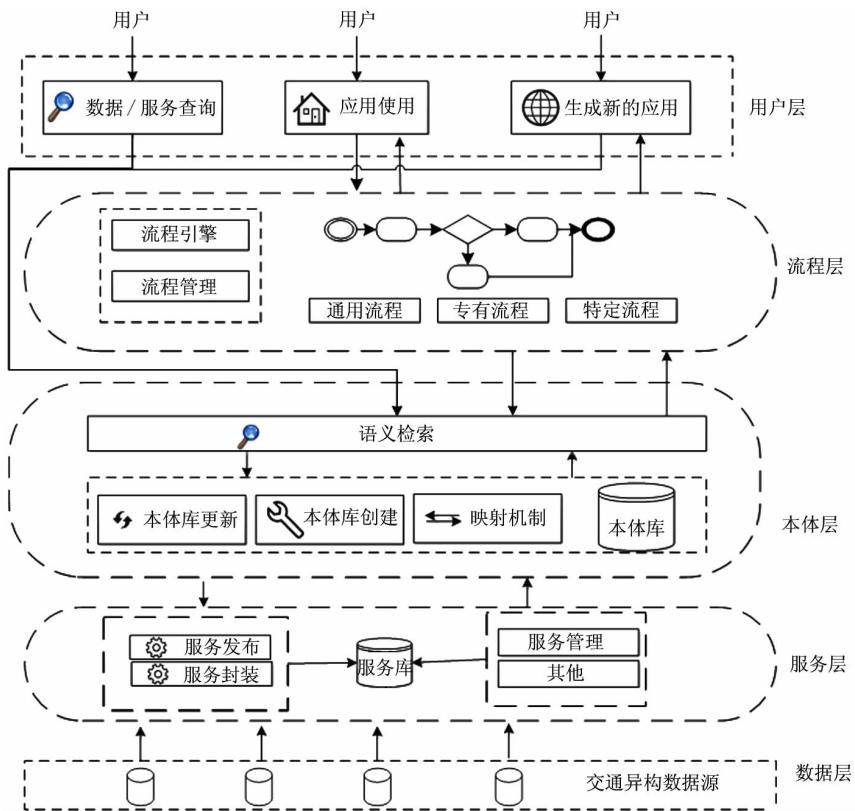


图 9 THDCSP V1.0 系统框架

含本体库创建、本体库更新等)与语义检索模块,而语义检索模块基于 Jena 实现,其目的是提供给用户或者应用程序一个统一的语义查询接口;流程层体现了交通异构数据集成模式的流程维,体现了本文方法的“流程驱动、服务组合”的思想,它包含了流程驱动引擎、流程模板库以及流程管理模块;用户层体现了本文方法的“面向用户”的思想,这里的用户是多种类型,有服务语义检索的用户,有普通应用使

用者,也有新应用开发者。

图 10 为服务语义检索界面,图中下部分为使用流程,例如在输入框中输入“时间占有率为”,点击“search”,其结果列表的操作栏中点击“导入”,即可将相应的服务导入流程。当然,在生成新的应用之前,需要对所选的服务进行流程校验,如果满足,将申请接入 SaaS 平台;如果不满足,系统将会报错,并提示开发者重新选择服务。



图 10 语义检索界面

## 4 结 论

本文在分析交通数据异构性的基础上,重点解决系统异构与语义异构问题,通过结合本体与云服务技术,在深入分析数据维、服务维、流程维的交通异构数据集成模式的基础上,提出了以服务封装、语义检索、流程驱动、服务组合、面向用户为核心的交通异构数据集成方法。为了验证框架的有效性,构建了大城市交通异构数据云服务系统。本文成果为进一步实现交通系统的集成化与异构交通数据集成发展打下基础,具有很强的理论与实践意义。同时,本文方法的服务本体自动更新以及面向 SaaS 模式的基于流程驱动的服务开发模式等有待进一步深入研究。

### 参考文献

- [ 1 ] 《中国公路学报》编辑部. 中国公路交通学术研究综述. 中国公路学报, 2012, 25(2):2-50
- [ 2 ] 王静远, 李超, 熊璋等. 以数据为中心的智慧城市研究综述. 计算机研究与发展, 2014, 51(2):237-259
- [ 3 ] 胡德敏, 李抒娅. 面向对象的分布式异构交通信息平台的设计研究. 上海理工大学学报, 2008, 30(5):497-500
- [ 4 ] Lv Bin, Huimin Niu. Transportation heterogeneous data integration based on Web Services. In: 2007 International Conference on Transportation Engineering, Chengdu, China, 2007. 1347-1352
- [ 5 ] Li W X, Yan M D, Wang J W. Ontology-based data integration for intelligent transport systems. *Journal of Central South University (Science and Technology)*, 2013, 44(7): 3038-3045
- [ 6 ] Li W X, Wang J W. Ontology-based transportation data integration and application for urban fast road networks. *Journal of Chang'an University (Natural Science Edition)*, 2013, 33(5):93-100
- [ 7 ] 曹妍. 本体理论在城市智能交通系统语义集成中的应用研究[博士学位论文]. 大连:大连海事大学, 2010. 15-17
- [ 8 ] 王坚, 阳其凯, 凌卫青. 基于本体的交通信息统一语义理解机制. 高技术通讯, 2014, 24(12):1211-1220
- [ 9 ] 徐恩虎, 凌卫青, 王坚等. 基于云计算的面向智能交通海量信息的高性能计算支撑公共服务框架. 机电产品开发与创新, 2013, 26(1):87-89
- [ 10 ] Apache. Apache Axis2/Java Version 1.6.2 Documentation Index. <http://axis.apache.org/axis2/java/core/docs/contents.html>; Apache, 2012
- [ 11 ] 杨兆升. 基础交通信息融合技术及其应用. 北京:中国铁道出版社, 2005. 1-3
- [ 12 ] Apache. Getting started with Apache Jena. <http://jena.apache.org>; Apache, 2013
- [ 13 ] Liu P, Zhou B S. Research on workflow patterns based on JBPM and JPD. In: 2008 Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan, China, 2008. 838-843

## Research on traffic heterogeneous data integration based on semantic web and cloud services

Yang Qikai, Wang Jian, Ling Weiqing

(CIMS Research Center, Tongji University, Shanghai 201804)

### Abstract

The integration of traffic heterogeneous data, especially under the circumstances of system heterogeneity and semantic heterogeneity, was studied in consideration of the needs of the construction of an intelligent transportation system (ITS). Based on the analysis of the traffic heterogeneous data integration mode considering data dimension, service dimension, process dimension, a new method for integration of traffic heterogeneous data with the directions of service package, semantic search, process drive, service composition and serving users was proposed by adopting the notology and cloud service technology. The method breaks the system barrier of heterogeneous data by using cloud service, and solves the semantic heterogeneity of heterogeneous data by introducing the ontology. A large city traffic heterogeneous data cloud service system was built to verify the effectiveness of the integration method. The result gives systematic solutions to the heterogeneous data integration problems in the traffic field at present, and provides reference and guidance for the construction of ITS in China.

**Key words:** traffic heterogeneous data, data integration, semantic web, ontology, cloud services