

一个基于元数据导航的服务 workflow 装配模型

王月龙¹⁾ 王文俊¹⁾ 罗英伟^{1),2)} 汪小林^{1),2)} 许卓群¹⁾

¹⁾(北京大学信息科学技术学院计算机科学技术系 北京 100871)

²⁾(石河子大学信息科学技术学院 石河子市 832003)

摘 要 以城市应急处置业务作为应用背景,提出了一个 workflow 的分层概念模型和一个与之关联的元数据分层描述规范,在此基础上构建了一个基于元数据导航的、从高层的业务应用到底层的分布、动态资源的逐层绑定的运行机制,基于该机制实现了一个城市应急联动系统 IERS(Integrated Emergency Response System)的实验原型.该机制解决了从应急处置业务 workflow 到底层分布服务和资源的装配问题,增强了应急系统对动态、分布服务环境的支持力度,提高了应急处置业务在执行过程中的自动化水平和自适应性,同时也分解和简化了 workflow 问题的复杂度.

关键词 应急联动系统; Web 服务; 网格服务; 元数据; workflow

中图法分类号 TP311

An Assembly Model for Service Workflows Navigated by Metadata

WANG Yue-Long¹⁾ WANG Wen-Jun¹⁾ LUO Ying-Wei^{1),2)} WANG Xiao-Lin^{1),2)} XU Zhuo-Qun¹⁾

¹⁾(Department of Computer Science and Technology, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871)

²⁾(School of Information Science and Technology, Shihezi University, Shihezi 832003)

Abstract Aimed at the emergency response application, this paper puts forward a layered concept model for workflow and a hierarchical descriptive model of metadata, and upon which establishes a mechanism for workflow assembly and execution navigated by the metadata. The mechanism is illustrated by using an application scenario of IERS(Integrated Emergency Response System). The mechanism described here solves the problem of mapping and binding from emergency tasks to distributed services and physical resources, enhances adaptability of an emergency system to dynamic and distributed environment, improves self-adjustability and stability of the emergency system in the course of deployment and execution, and at the same time decomposes complexity and reduces abstruseness of the problems related with workflow.

Keywords IERS; Web services; Grid services; metadata; workflow

1 引 言

本文主要研究面向分布式环境的应急处置业务 workflow 的部署和实现. workflow 是一种重要的协同工

作机制^[1,2],目的是实现整个或部分业务过程的自动化描述和运行. workflow 也是 Web 服务环境和网格(Grid)服务环境中一种重要的协同工作机制. 服务可以组合成服务 workflow. 服务 workflow 是指为完成某一业务功能而对一个或多个 Web/Grid 服务的调用

收稿日期:2006-02-17;修改稿收到日期:2006-02-20. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2002CB312003)、高校博士点基金(20020001015)、国家自然科学基金(60203002)以及新疆生产建设兵团博士基金资助. 王月龙,男,1967年生,博士研究生,主要研究方向为 GIS 及应用集成. 王文俊,男,1970年生,博士,主要研究方向为应急联动、GIS 及应用集成. 罗英伟,男,1971年生,博士,副教授,CCF 高级会员,主要研究方向为分布式计算及 GIS. E-mail: lyw@pku.edu.cn. 汪小林,男,1972年生,博士,副教授,主要研究方向为分布式计算及 GIS. 许卓群,男,1936年生,教授,博士生导师,主要研究领域为科学计算可视化、并行计算、人工智能及 GIS.

序列. 本文研究的重点是如何将一个应急处置业务工作流映射成一个分布式服务流并绑定具体的物理资源部署和执行的问题.

Web 服务与 Grid 服务构成了服务环境, 即工作流的运行环境. 服务环境包含了服务以及服务所使用的资源. 服务环境中的所有实体都是资源, 如数据文件、处理能力、存储系统、数据库等. 服务也可看作一类特殊的资源.

在 Web 环境中, 针对服务和资源的动态、分布和异构等特性, 如何实现业务工作流的快速部署, 将一个业务过程绑定到具体的服务和资源实例, 保证工作流自动、稳定而且高效地执行是一个迫切需要解决的问题.

以应急处置业务领域为背景, 本文提出了一个基于元数据导航的从高层业务工作流到底层服务工作流的分层部署和执行的解决方案, 并以此为基础实现了一个城市应急联动系统 IERS(Integrated Emergency Response System)的实验原型. 本文主要内容如下:

首先, 提出了一个工作流的分层描述模型. 按照任务粒度和抽象层次将工作流从高到低逐层进行建模, 自上而下分别定义了业务工作流、抽象工作流和执行工作流等.

其次, 提出了一个面向工作流部署和执行的元数据分层描述规范. 为了能对工作流的映射和匹配提供导航服务, 将元数据自上而下划分为业务层元数据、任务层元数据、规范层元数据和物理层元数据. 每一层元数据描述了相邻两层工作流原子节点之间的映射关系. 元数据的分层建模为分层工作流的映射和装配提供了一个有益的基础设施.

基于上述工作, 设计了一个基于元数据导航的、从高层的应急处置业务过程到底层的分布动态服务和资源的映射和执行的机制, 实现了一个 IERS 原型系统, 并进行了实验模拟. 该机制较好地解决了应急处置业务工作流在动态服务环境中的自动映射和半自动部署和执行问题, 提高了应急处置业务的自动化水平和运行效率.

2 应急处置领域的本体模型

如何让计算机能够理解和处理工作流是工作流能否自动部署和执行的关键, 解决思路之一就是给工作流文本添加语义. 本体技术^[3~5]是当前知识建模的重要工具, 是描述语义模型的通用手段. 本文尝试基于本体技术对工作流建模进行研究. 使用本体

有以下几个方面的优势:

- (1) 它使得工作流的描述文本具有更多的语义信息, 便于机器理解和处理;
- (2) 基于本体概念的工作流支持语义映射, 便于不同抽象层次工作流的映射和匹配;
- (3) 它支持跨领域的语义共享与互操作;
- (4) 工作流的描述更为严谨和理论化, 便于工作流的普适和复用.

2.1 基本概念

本体主要描述概念与概念之间的关系以及基于这些概念的公理和业务规则等. 对相关领域的概念及其内涵进行梳理和界定是本体建模的一项重要工作. 面向突发性公共事件的应急处置环境是一个跨业务领域、跨部门的虚拟组织(VO), 下面给出虚拟组织(VO)及与 VO 有关的几个概念的定义.

定义 1. 虚拟组织(Virtual Organization, VO)^[6]是指在某种共享规则下的一组个体或组织, 即多个互相独立的参与者或自治的系统通过共享资源来共同完成某个任务. 虚拟组织实现了有条件的、高度可控的共享. 共享关系随着时间的推移可动态变化. VO 成员及资源是动态、异构的和分布的. VO 的组成要素主要有资源、服务、组织机构、共享规则、安全规则、VO 本体(库)等. 为支持分布式 Web 环境各类资源的共享与互操作, 虚拟组织(VO)的概念被引入到 Web/Grid 服务领域, 并得到了广泛的应用.

应急处置业务领域的虚拟组织 VO(下文中如不特别说明, 所有的 VO 均指应急处置 VO)是指应急处置过程所涉及的组织、人员、资源和系统等所构成的一个动态的“应急事件处置环境”. 该“环境”为完成某个突发事件的处置往往需要涉及到多个分布、自治的业务领域、多个组织、多类资源、多类服务或多个系统等. 建立应急处置 VO 的目标是为应急处置业务提供协同工作的环境.

定义 2. 资源是对世界进行概念化后所得到的对象和关系. 在虚拟组织 VO 中, 所有基于某种共享规则用于共享目的的实体都可以称之为资源. 在基于网络/网格环境的 VO 中, 常用的 VO 资源包括处理能力、存储系统、目录服务、Web 服务、Grid 服务、数据服务、缓存服务等.

在一个 VO 中, 具有相同功能的资源实例可能会有多个, 为支持对同一类资源的统一描述, 本文引入了“抽象资源”的概念, 抽象资源是指 VO 环境中具有相同功能的同一类资源的抽象描述, 与之相对应的是“具体资源”(或“物理资源”、“资源实例”等), 描述的是抽象资源的一个具体实现. 以 WebGIS 系

统中的图层概念为例,比如“城市道路图层”是抽象资源,它描述的是一类 GIS 数据,而在该市的测绘局、消防局、交通局和公安局等部门都会有相应的道路网图层数据.这些数据就是具体资源.

使用抽象资源的概念可以对应应急处置预案和应急处置业务方案的描述提供语义和概念的支持,因为在高层的应急处置业务工作流程文本中不可能绑定具体的资源实例.

定义 3. 服务一般是指可以在网络或网格环境中调用的具有某种功能的(软件)实体^[7],是对 VO 中可以远程调用并实现某类功能的软件模块的抽象. VO 中的服务主要有 Web 服务、Grid 服务以及某些可以远程调用的(软件)应用系统等.

具有相同接口和相似功能的服务构成一个“服务类”;某一服务类的一个具体实现,即一个具体的服务叫做“服务实例”.广义地说,服务也可以看作是一类特殊的 VO 资源,因此服务类可以称为“抽象服务”.服务实例也可以称为“具体服务”或“物理服务”.

VO 环境中 Web 服务与 Grid 服务的主要区别如下:

(1) Web 服务以一种标准的方式在 Internet 上发布和被调用. Web 服务一般没有状态,用户只关心服务的可用性.不过,同样的接口可能会有不同的实现;

(2) Grid 服务增强了 Web 服务的功能,使得服务的状态通过服务实例的生命周期体现出来. Grid 服务通过标准的 Web 服务发布方式提供相应的操作,允许用户通过服务工厂创建、运行和终止服务实例.而且,通过服务实例提供的自描述信息服务数据,不同 Grid 服务之间能够进行中间状态的信息交换.

定义 4. 数据是对某一实体或现象的描述,该描述具有一定的逻辑结构和存储格式.描述同一类事物的多个数据对象构成一个“数据类”.数据类是对数据对象的抽象描述,可以看作是数据对象的模板;数据对象也可以称为“数据实例”,它是某一数据类的一个具体实现.同一个数据类的不同数据对象可能会有不同的格式.如“城市道路图层”是数据类,它的数据实例可以是 MapInfo 格式、Arc/Info 格式或 GML 格式.数据也是一类 VO 资源.数据类又可以称为“抽象数据”.数据实例也称为“具体数据”或“物理数据”.

定义 5. 角色是某一类对象性质、行为、权利和责任等方面属性的综合描述,是对事件参与者作用和职能等特征的抽象.角色作为责任和权利的统

一,一般分为职务角色和功能角色.角色是一个主动的、相对独立的抽象单元,具有一定的任务目标,能按程序完成一系列操作,实现某一业务功能.角色一般由 Agent(组织、人员和系统等)实现.一个角色可能会有多个 Agent 承担,一个 Agent 也可能会担当多个角色.

在应急处置领域引入角色的概念可以对应应急预案和应急处置业务方案提供概念和语义的支持,因为在高层的业务工作流程文本中使用各类角色有关的概念更为合理.这些角色在生成可执行的工作流后再与具体的参与者(如部门、人员等)相关联.

2.2 本体、领域本体和 VO 本体

本体的抽象有多个层次,按概念的粒度和适用的范围划分,本体可以分为上位本体(Upper Ontology)和领域本体(Domain Ontology)^[8,9],也有专家在上位本体和领域本体中间又定义了一层中位本体(Mid-Level Ontology),以更好地支持跨领域的语义共享与互操作.

上位本体是一类抽象层次较高的通用本体,它描述了现实世界最基本的概念及概念之间的关系,是一种领域无关的、基础性的概念模型和知识结构,这使它具有广泛的通用性和较强的表达能力.上位本体的作用更多地是为多领域本体之间的映射和相互理解提供支持. SUMO(Suggested Upper Merged Ontology)^{①②③}是由 IEEE Standard Upper Ontology Working Group 创建的一类通用的上位本体. SUMO 是一类领域无关的、基础性的概念模型和知识结构. SUMO 的作用更多地是为多领域本体之间的映射和相互理解提供支持.目前, SUMO 已和其它一些通用本体建立了映射,如 WordNet, Lexical Database 等,目前已经成为一类被普遍接受的、主流的上位本体.

领域本体(Domain Ontology)是面向特定领域、用于描述领域知识的概念模型,目标是捕获相关领域的知识,确定该领域内共同认可的词汇,并对这些词汇(术语)和词汇间的相互关系形式化地描述,以支持对该领域知识的共同理解.领域本体要求模型中的概念至少在某个特定的领域或者某个范围内是公认的.领域本体有助于消除在概念和术语上的分歧,对领域内的概念理解达成共识.

适应应急处置业务领域的实际需要,我们实现了一个分层的本体模型.这个分层本体按抽象层次依次

① <http://www.ontologyportal.org/>

② <http://ontology.teknowledge.com>

③ <http://virtual.cvut.cz/kifb/en/>

是:基于 SUMO 的上位本体、面向突发性公共事件处置的本体、面向火灾事故处置的本体、面向化工火灾处置的本体。最上面一层是面向各领域间语义共享的通用本体,下面三层可以看作是领域本体,其中第二层本体(面向突发性公共事件及应急处置的本体)也可以看成是支持应急处置各子业务领域(110, 119, 120, 122 等)间语义共享与互通的中位本体。

应急处置 VO 中的各成员一般都有自己的词汇和术语集。这些术语之间有时可能会出现语义方面的歧义甚至混乱。如一词多义的情况,匪警(110)业务领域的“第一出动”≠火警(119)业务领域的“第一出动”;一义多词的情况,按照国家有关的消防法规,火灾划分为“一般火灾”、“重大火灾”和“特大火灾”三级。同时,按照国务院 2006 年 1 月 8 日最新颁布的《国家突发公共事件总体应急预案》,火灾事故又是“突发公共事件”中的一类“事故灾难”,可以分为“一般”、“较大”、“重大”和“特大”四个级别,这样,按消防法规规定性为“重大火灾”的一场火灾,如果按突发公共事件的标准划分,则可能是“重大事件”,也可能是“较大事件”。为更好地支持跨领域、跨组织的语义共享和业务集成,我们实现了应急处置业务领域 VO 的本体模型。

VO 本体(VO Ontology)是指虚拟组织中各管理域一致认可的词汇和词汇之间相互关系的明确定义。VO 本体主要描述 VO 内各业务领域之间的语义关系,包括各领域/组织中的专业术语、概念之间的映射关系以及它们之间的区别与联系等。我们建立的 VO 本体同时还扩展了描述事件、方案、预案和抽象服务工作流等方面的词汇。VO 本体是 VO 的重要组成部分,明晰的 VO 本体描述便于消除跨组织、跨领域、跨系统间的语义冲突,是实现多管理域协同工作的语义基础。

3 工作流的分层模型

一个工作流从生成到部署执行会表现为不同的形态。为了简化问题的复杂度,更方便地对工作流问题进行分析与设计,本文对工作流进行分层建模研究,提出了一种工作流的层次模型,按照抽象层次以及组成工作流的业务过程的粒度大小,将工作流划分为业务工作流、抽象工作流和执行工作流三个层次。

3.1 用户分析

每一类工作流都对应着一类用户,在对工作流分层建模之前,有必要首先对用户群体进行分类研究。以应急处置领域为例,一个应急处置业务工作流

从生成到部署和执行,需要 4 类不同的用户参与,他们分别是业务应用者、VO 管理者、资源管理者和资源提供者。

3.1.1 业务应用者

应急处置领域的业务应用者一般是指应急指挥员或领域专家。领域专家拥有丰富的业务知识和应急处置经验,负责制定领域知识库。该知识库包括业务规则、应急预案和应急范例等。当某一事件发生后,应急指挥员根据事件动态,参照知识库生成一个用业务过程序列描述的处置方案。该业务过程序列是一个工作流,这样一个工作流完全用业务术语描述,与底层的实现无关。因此,这一类用户在创建或触发一个业务工作流时,不需要考虑资源或服务状态问题。

3.1.2 VO 管理者

应急处置领域是一个涉及多个业务领域、多部门、多子系统的 VO 环境,其中各个部门和子系统可能都会有各自的服务和数据资源,这些资源的状态是分布和动态变化的,这时如果直接将用业务术语描述的工作流匹配到某一具体的服务或资源会是一件困难的事情。VO 管理者可以看作是面向抽象资源和抽象服务的管理者,他只关心当前 VO 环境中服务和资源的种类、数量和功能等,而不必关心其具体的技术细节。

VO 管理者负责将用业务术语描述的工作流转换为抽象资源描述的工作流。同时还要基于本体知识库解决跨管理域的语义共享与互操作问题,如在必要时,可能需要明确某一个“第一出动”过程是火警(119)的“第一出动”,还是匪警(110)的“第一出动”。

3.1.3 资源管理者

资源管理者是指某一业务部门或业务子系统的资源管理者。他负责该部门或业务子系统内部所有资源的管理与发布,并向上一级用户(VO 管理者)注册当前资源的描述信息。资源管理者负责把使用抽象资源和抽象服务描述的工作流绑定到具体的资源实例,最终完成工作流在分布式服务环境中的部署和执行。

3.1.4 资源提供者

资源提供者提供资源实例和服务实例,这些资源和服务实例构成了底层的服务环境,即工作流的运行环境。服务环境中的所有实体都可称之为资源,如 Web/Grid 服务、数据文件、处理能力、存储系统、数据库等。服务也可看作一类特殊的资源。服务环境中的资源一般是分布、异构、自治而且是动态变化的。

3.2 工作流的层次

3.2.1 业务 workflow

业务 workflow (Business WorkFlow, BWF) (或应用 workflow) 是完成特定业务功能并采用业务术语描述的工作流。业务 workflow 通常由业务应用者 (应急指挥员或领域专家) 基于领域知识制定, 这一层 workflow 主要是面向业务领域的通用预案和业务方案等。

业务 workflow 主要由任务节点组成。为支持跨业务领域、跨系统的语义共享与互操作, 业务 workflow 中的任务节点均使用规范的业务术语描述, 这些术语来源于领域本体和 VO 本体。

在这一层, 需要考虑业务 workflow 运行过程中, 业务用户如何与 workflow 进行交互, 及时控制 workflow 的恰当执行, 保证 workflow 按照要求完成预定的处置计划。

3.2.2 抽象 workflow

抽象 workflow (Abstract WorkFlow, AWF) (或抽象服务 workflow) 是为完成某一特定的业务功能, 采用一个组织内定义的标准资源类 (抽象服务或抽象资源) 描述的工作流。workflow 中的任务节点采用抽象服

务 (服务类、接口等) 描述。数据节点采用抽象数据 (数据类) 描述。这一层的工作流, 一般是由业务 workflow 部署到服务环境中执行时所生成的中间层 workflow, 为下面一层可执行的工作流绑定具体资源和服务提供支持。

3.2.3 执行 workflow

执行 workflow (Executable WorkFlow, EWF) (或可执行 workflow) 是抽象 workflow 实现服务和数据的选取与绑定后生成的 workflow。该层 workflow 可以直接部署到服务环境中执行。

因为服务环境是动态变化的, 所以执行 workflow 在运行过程中的自适应性及健壮性是一个需要重点考虑的问题。问题包括资源调用失败、流程停顿、流程死锁、远程系统崩溃后的自动调整与恢复、资源的多重选择策略等, 从而向下保证资源的有效调度, 向上保证整个业务过程的可靠执行。

执行 workflow 中的任务节点和数据节点分别使用服务实例和数据实例描述。表 1 给出了一个三层 workflow 词汇描述的示例。

表 1 三层 workflow 的词汇描述示例

workflow 层次	workflow 中的任务	workflow 中的数据
业务 workflow	用业务术语描述, 如“计算编队顺序”	用业务术语描述, 如“道路”
抽象 workflow	用抽象服务描述, 如“最短路径计算”	用虚拟资源描述, 如“市区 1:2000 道路网图层”
执行 workflow	用服务实例描述, 如 http://168.168.168.216:8080/ogsa/services/cn/edu/pku/gis/pai/bestpath/BestPathFactoryService	用资源实例描述, 如 \\168.168.168.1\abc\abc.shp

3.2.4 各层 workflow 与各类用户的关系

三层 workflow 与各类用户的关系如下:

首先, 业务应用者 (应急指挥员或领域专家) 根据某一应急任务需求使用业务术语书写一个业务 workflow 文本。该业务 workflow 描述了一组业务过程的运行顺序和运行条件, 目标是完成某突发事件的处置任务;

其次, 业务 workflow 被提交给 VO 管理者。VO 管理者将业务 workflow 中的每一个业务过程表示为对服务类与数据类的调用序列, 形成抽象 workflow, 并提交给资源管理者;

最后, 资源管理者根据当前服务环境中每个资源和服务的实时状态确定对具体资源和服务的调用, 将抽象 workflow 中的数据类和服务类映射为具体的数据实例和服务实例, 抽象 workflow 变成可在网络环境中运行的执行 workflow;

资源提供者负责网络、服务、数据等资源及运行环境的开发与维护工作。

3.3 工作流分层建模的意义

对 workflow 进行分层建模之后, 分解了 workflow 从

生成到部署执行的复杂度, 便于不同的用户群分别从不同的层次对 workflow 问题进行分析、设计和研究。

首先, 将 workflow 分为业务 workflow 与服务 workflow 两层, 使业务应用者专注于业务过程的设计和优化, 不必考虑采用哪一类服务或者资源。例如应急指挥员或领域专家只要在业务 workflow 中使用“计算编队顺序”这一任务节点, 而不用关心调用什么算法和服务来完成这一任务。

其次, 将服务 workflow 分为抽象 workflow 和执行 workflow 的主要益处有: (1) 降低一次性映射的难度, 如果将 BWF 直接映射到 EWF, 不仅要考虑业务术语的转换, 还要考虑服务环境的技术细节, 而服务环境又是动态变化的; (2) 从 BWF 到 AWF 的映射可以预先进行, 而且这种映射结果是可重用的, AWF 到 EWF 的映射可以在执行时进行, 这时只需考虑资源和服务的选取策略, 这样既提高了运行效率, 又隔离了因服务环境变化对高层 workflow 的影响。

workflow 分层建模后, 各类用户的角色会更加明确和清晰。应急指挥员和领域专家只负责业务 workflow 的设计和优化; VO 管理者主要负责抽象 workflow

的设计和抽象资源的管理,他不必有太多的业务知识,也不必熟悉太多底层服务环境的物理细节,只需在宏观层次上了解当前的分布环境中哪几类资源和服务,它们分别能实现哪些业务功能即可;资源或网络管理者主要负责从抽象服务到具体服务和具体资源的映射和匹配,不必具备某一业务领域的深厚知识,但他必须熟知该业务子系统内有哪些具体的资源和服务。

3.4 一个分层工作流的实例

下面以城市应急联动系统 IERS(Integrated Emergency Response System)为例来说明工作流的分层模型。IERS 是综合各种城市应急服务资源,采用统一的号码,用于公众报警和紧急求助,统一接警,统一指挥,联合行动,为市民提供紧急救援服务,为城市公共安全提供保障的应急处置指挥系统。IERS 是一个跨业务领域的综合管理系统,它集成了 110,119,120,122,LBS 等业务子系统,能够统一调度各业务子系统的服务资源、数据资源及其它资源等以完成一个应急救援任务。

以某市 IERS 为例,不失一般性,假设“某大型商场发生特大火灾事故”,这样一个火警信息上报到应急联动指挥中心以后,将触发一个处警方案,即应急处置业务工作流,该工作流的逐层装配、部署和执行如图 1 所示,下面是每一层的工作流描述。

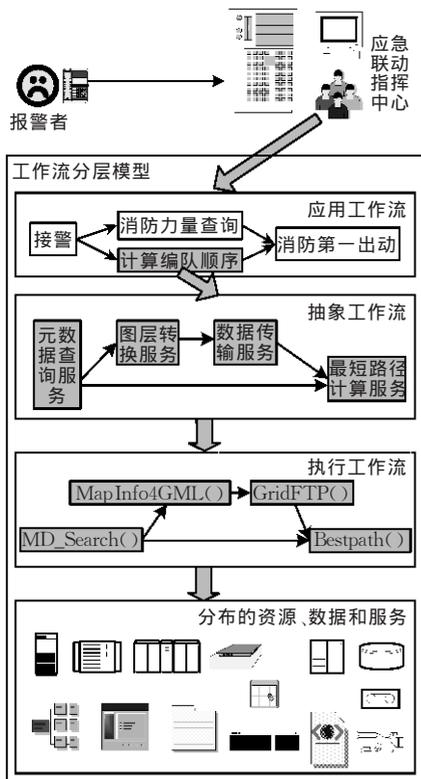


图 1 IERS 的一个分层工作流实例

(1) 应急联动指挥中心接到报警后,指挥员根据当前的事件动态和警情特征,参照处警业务规则及相关的应急业务知识库,制定一个处警方案,即业务工作流。首先查询当前可用的消防实力(人员、车辆、器材等),然后根据各消防中队距离火灾现场的距离,计算出拟出动的消防力量,即编队顺序,根据这个出警顺序,下达“第一出动”的命令。这一层的工作流完全由消防业务术语描述,不涉及抽象的或具体的资源和服务概念。

(2) 业务工作流中的每一个业务节点都需要映射为下一层的抽象工作流,以“计算编队顺序”为例,它对应的抽象工作流如下:首先“查询元数据信息”,如果本地有 GML 格式的“城市道路图层”,则直接调用“最短路径计算服务”;如果本地没有该图层,则首先调用“格式转换服务”,将其它格式的“城市道路图层”转换为 GML 格式,然后调用“数据传输服务”将该数据传送到本地,再调用“最短路径计算服务”。该层工作流所涉及到的服务或者资源等都是一抽象的描述,不涉及具体的资源或服务实例。

(3) 抽象工作流中的抽象服务和抽象资源需要映射成具体的服务和资源实例,形成执行工作流,并部署到服务环境中,才能最终完成工作流的执行。上一层的抽象服务“元数据查询服务”、“图层转换服务”、“数据传输服务”和“最短路径计算服务”在这里分别映射为“MD_Search()”、“MapInfo4GML()”、“GridFTP()”和“BestPath()”等服务实例。注意,这些服务都是指在某一特定主机环境中且绑定具体资源的服务。由于服务环境是一种分布、动态的环境,而且其中大部分服务和资源可能都是冗余的,因此,服务和资源实例的选取需要根据资源选择策略确定,这时需要考虑负载情况、传输带宽、服务质量等因素。

4 元数据的分层模型

元数据是关于数据的描述性信息^[10],是说明数据内容、质量、状况及其它有关特征的描述信息,从机器的角度看,信息是机器可理解的资源描述。Web 环境的元数据主要包括各类 Web 资源的描述,如主机、数据、服务、硬件、带宽、用户等,也包括语法、语义等方面的描述性信息。元数据技术主要对 Web 环境中的服务、数据、资源等进行描述和管理,提供相应的导航和选取服务来帮助用户选择合适的资源。针对应急处置业务的应用背景,本文提出了一种分层的元数据组织模型,基于这样一种规范组织的元

数据基础设施,能够帮助业务工作流程经过逐层映射,最终在服务环境中便捷地部署和执行。

为了能对工作流的装配和部署提供导航服务,本文提出了一个分层的元数据描述规范,将元数据内容分为业务层、任务层、规范层和物理层等几个层次,如图 2 所示。各个层次的元数据内容都是相互关联、相互映射的。元数据的信息不仅包括分布式服务和数据的描述,也包括领域知识的描述、领域本体、VO 本体、用户/主机的描述,还有对工作流功能节点和控制节点的描述等。

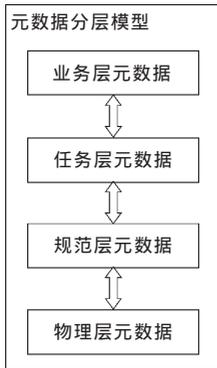


图 2 一个元数据的分层模型

4.1 业务层元数据

业务层元数据的主要目标是支持业务方案的半自动化触发,在应急处置的临机关头帮助业务用户(决策者或领域专家)快速生成业务 workflow。以应急处置业务领域为例,该领域的业务层元数据主要包括两方面的内容:业务知识元数据和语义信息元数据。

业务知识元数据主要包括各类应急知识如预案、范例以及业务规则等的描述和索引信息,该类元数据帮助用户基于一个事件的描述信息自动查找和匹配到相关的业务知识,如适用的预案、范例和业务规则等。这些适用的预案、范例和规则经过用户适当调整和改造,可以变成一个初步的处置方案,即业务 workflow。

语义信息元数据是某类应急知识(如应急处置范例)的语义描述信息。该类元数据能够为用户提供必要的语义联想和语义导航支持,语义信息元数据基于本体创建,具有一定的语义推理功能,能够指导用户对业务 workflow 片断进行调整和优化,并能帮助用户自动生成一些业务 workflow 片断。

4.2 任务层元数据

业务 workflow 是一个任务节点序列,其中的任何一个任务节点可以映射成抽象 workflow 中的抽象服务和抽象资源序列。任务层元数据主要描述任务节点与

抽象资源和服务的映射关系,基于这样一种映射关系,一个业务 workflow 文档就可以自动映射成使用抽象资源和抽象服务描述的抽象 workflow。

图 3 描述了一个任务节点“计算编队顺序”和一个抽象服务序列的映射关系,相应的描述文档,即任务节点“计算编队顺序”的元数据内容如下:

```

<TaskNode Name="UnitSequence">
  <DomainTerm>
    <Term>计算编队顺序</Term>
  </DomainTerm>
  <AbstractWorkFlow>
    <variables>
      <variable type="string" name="url4gml"/>
      <variable type="string" name="gml4road"/>
      <variable type="string" name="gml4result"/>
    </variables>
    <activity>
      <sequence>
        <invoke portType="元数据查询服务"
          operation="GetGISMeta">
          <parameters/>
          <outputDirection name="url4gml"
            index="1"/>
        </invoke>
        <switch>
          <case condition="url4gml="">
            <sequence>
              <invoke portType="图层转换服务"
                operation="Mapinfo2gml">
                <parameters/>
                <outputDirection name="gml4road"
                  index="1"/>
              </invoke>
              <invoke portType="数据传输服务"
                operation="SendGML">
                <parameters name="gml4road"
                  index="1"/>
                <outputDirection name="url4gml"
                  index="1"/>
              </invoke>
            </sequence>
          </case>
        </switch>
        <invoke portType="最短路径计算服务"
          operation="Multi2One">
          <parameters>
            <parameter name="url4gml"
              index="1"/>
          </parameters>
        </invoke>
      </sequence>
    </activity>
  </AbstractWorkFlow>
</TaskNode>
  
```

```

</parameters>
<outputDirection name="gml4result"
  index="1"/>
</invoke>
<return>
  <output>gml4result</output>
</return>
</sequence>
</activity>
</AbstractWorkflow>
</TaskNode>
...
</Operation>
</Service>
<Service ServiceType="GridService">
  <ServiceProvider Name="LXP"
    Type="">
    <LocatorType="Factory"
      Handle="http://gis.pku.edu.cn/ogsa/gis/
        pai/bestpath/BestPathFactoryService"/>
    </ServiceProvider>
    <Operation Name="calculateBestPathLength">
      ...
    </Operation>
  </Service>
</InstantiationServiceList>
</AbstractService>

```

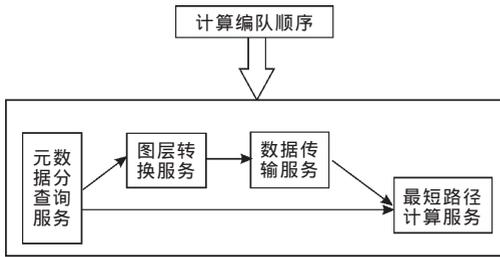


图 3 一个任务节点和抽象服务序列的映射

4.3 规范层元数据

规范层元数据是对分布式服务环境中同一类服务、数据等资源的抽象描述。该层的抽象服务元数据,是各业务领域实现服务的依据,也是 workflow 调用服务的标准,这一部分的元数据标准一旦确定,就要尽量减少改动,否则所有已经注册的服务实例可能都要做相应的改动,这样对系统的维护很不利。

该层元数据主要是一些集成信息或抽象信息,是对一些标准的资源类的统一描述,每一个资源类可以有不同的具体实现(或实例)。比如抽象服务类可能有“图层格式转换服务”、“道路图层分段服务”、“最短路径计算服务”等;抽象数据类可能有“道路图层”、“政区图层”、“消防中队图层”、“重点单位图层”等。

“最短路径计算服务”是一个抽象服务类,下面的文档片断是该抽象服务的描述元数据,该元数据文档描述了该抽象服务和两个服务实例的映射关系:

```

<AbstractService Name="最短路径计算服务"
  Description=""/>
<InstantiationServiceList>
  <Service ServiceType="WebService">
    <ServiceProvider Name="WYL" Type="">
      <LocatorType="Service"
        Handle="http://168.168.168.225:8080/axis/
          services/bestpath/BestPathService"/>
    </ServiceProvider>
    <Operation Name="getBestPathLengthOrder">

```

“最短路径计算服务”这一抽象服务被映射到两个可用的服务实例:一个是 WYL 提供的 Web Service,它采用 Dijkstra 算法实现,是一个标准的 Web Service 接口;一个是 LXP 提供的 Grid Service,它采用 A* 算法计算最短路径,是一个 Grid 服务的工厂方法。

4.4 物理层元数据

物理层元数据主要是对分布式服务环境中服务实例、数据实例等具体资源的描述。各业务领域将自己的共享资源和服务等通过注册接口注册到物理层元数据库,形成物理层元数据记录,这部分记录的内容主要包括:资源 URL、IP、服务容器类型等。物理层元数据不仅描述了各个共享的服务或数据资源实例的信息,同时还描述了资源实例宿主机器的硬件信息以及系统负载情况等,供资源选择模块调用。

以 GIS 数据为例,其元数据内容主要包括图层元数据和地图元数据。图层元数据以美国联邦数据委员会(FGDC)的元数据规范^[1]为基础,描述了图层的名称、空间范围、专题以及图层所在的空间数据库(包括该空间数据库的类型、访问方法)等;地图元数据内容主要包括地图的名称、空间范围、专题以及组成图、存放位置、访问方式等。

服务实例元数据的设计参考了 Web 服务技术的 UDDI 规范和 Globus Toolkit 3. x 的 Index Service 规范,描述了服务的名称、功能、部署地点、服务端口、服务接口、服务参数信息以及访问权限等。具体地说,一个服务实例的元数据主要包括以下几个方面的内容:

(1) Resource, 描述了服务执行所需要的资源;

(2) Service profile, 描述了服务的基本轮廓信息,如服务提供者的信息、服务的功能信息、服务自动选取时所需要的几个功能属性信息等;

(3) Service model,描述了服务的工作机制,说明了服务调用之后产生的数据流和控制流等;

(4) Service grounding,描述了调用一个服务实例的细节信息,规定了调用该服务实例的通信协议、URI、端口号、入口参数和出口参数等细节.

5 一个应急处置业务流程的部署和执行

下面以某市 IERS 系统处置一次重大火灾事故为例,介绍基于元数据导航的工作流的逐层映射、部署和执行过程,各层工作流的映射和匹配如图 4 所示,详细介绍如下.

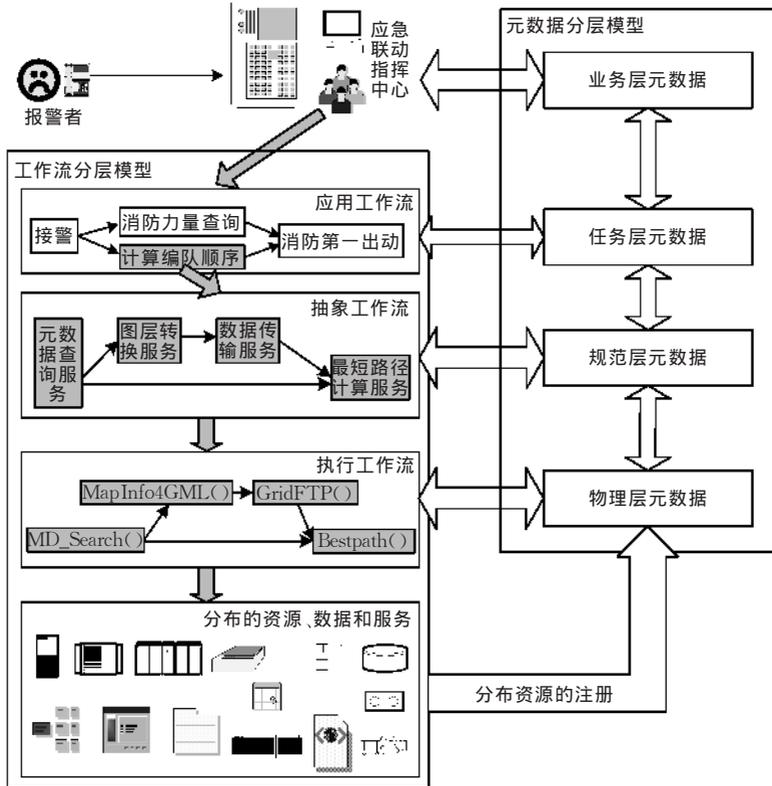


图 4 基于元数据导航的工作流逐层映射

5.1 业务工作流的触发

当“应急联动指挥中心”接到火灾警报后,需要首先了解以下一些灾情要素:“时间”、“地点”、“事故等级”、“伤亡人数”等,根据这些事件信息,就能基于业务知识元数据信息在应急知识库中检索到相关的预案或者范例.应急指挥员根据事件的动态对相关的预案或范例中的任务节点进行调整,生成一个初步的处置方案.

指挥员通过语义浏览工具,对相关的语义信息元数据进行快速浏览,获得必要的语义提示和启发,以此对处置方案进行进一步补充和完善,最终生成一正式的处置方案,即一个应急处置业务 workflow,该 workflow 主要包括以下行动序列:“查询当前消防实力”、“计算编队顺序”、“消防第一出动等”.

5.2 抽象工作流的生成

以业务 workflow 中的一个任务节点“计算编队顺

序”为例,介绍如何基于元数据信息将该层 workflow 中的任务节点映射为一个抽象 workflow 片段.根据任务层元数据的相应描述,任务节点“计算编队顺序”能够自动映射为如下的抽象 workflow 片段:

首先“查询(图层)元数据信息”,如果本地存在 GML 格式的“城市道路图层”,则直接调用“最短路径计算服务”;

如果本地没有相应的图层,则首先调用“格式转换服务”,将其它格式的“城市道路图层”转换为 GML 格式;

然后调用“数据传输服务”将该数据传送到本地;最后调用“最短路径计算服务”.

5.3 执行工作流的生成和执行

抽象 workflow 到执行 workflow 的映射主要基于规范层元数据的描述信息.以“计算多点到一点的最短路径”为例,这是一个抽象服务,查询规范层元数

据库中对该类服务的描述,可知有多个服务实例可以使用,综合考虑多种因素,我们选择了如下的服务实例: <http://168.168.168.132:8080/axis/services/bp?wsdl>, 再为该服务实例绑定数据资源“tjroad.xml”,就实现了“最短路径服务”。同样的方法将其他抽象资源和抽象服务分别映射成相应的资源实例和服务实例,最后生成一个可以部署和执行的执行工作流。

执行工作流根据物理元数据对资源和服务实例的调用细节的描述,如通信协议、权限、入口参数、出口参数等,调用相应的服务实例并绑定相应的数据,就完成了执行工作流的部署和执行,这样就成功实现了从业务工作流到服务环境的逐层映射、装配和运行。

6 相关研究工作

InfoPAE 是 Tecgraf 公司与 Brazilian Oil 公司合作开发的一个应急方案部署系统。XPAE^[12] 是 InfoPAE 系统的一个方案定义语言。一个应急方案 Plan 描述了应对某一个设定危机情节所采取的行动集合,表现为一个工作流。在语义层次,一个 Plan 可划分为若干个 subPlan,在部署执行时,一个 Plan 或 subPlan 可以作为一个独立的 Process 执行,或由多个 Process 完成。该系统没有考虑事件处置参与者的动态变化以及跨业务领域互动时产生的语义冲突。

ISI 研究的 Pegasus^[13,14] (Planning for Execution in Grids) 是 GriPhyN[®] 和 SCEC/IT 项目的一部分,它是一个可配置的系统,能将复杂的业务工作流映射到网格环境中执行。Pegasus 中的工作流有两个层次:用描述逻辑应用功能和逻辑数据文件名组成的抽象工作流,绑定具体数据文件和服务的具体工作流。Pegasus 主要目标是获取数据产品,只是根据工作流执行时服务环境的实时状态,从冗余的服务和资源中选取更合适的服务和资源。两个层次的工作流都用网格术语描述,没有涉及不同领域术语的映射。

加利福尼亚大学的 SciDAC-SDM 项目^[15,16] 将用业务术语描述的抽象工作流与底层的由服务组成的执行工作流严格区分,并借用与 GAV (Global-As-View) 查询重写的数据库协调技术相类似的方法将抽象工作流转换成执行工作流。该项目中,从业务领域到服务环境的映射只是静态地转换,不能根据服务的实时状态在执行时选取更合适的服务。

Pegasus 和 SciDAC-SDM 都是研究如何将大

规模科学计算的工作流部署到分布式计算环境中。而应急处置领域的业务流程具有和科学计算工作流不同的几个特点:

(1) 科学计算工作流的主要目标是获取终级数据,而应急处置工作流更注重的是业务过程的有效协同和集成,以达成某一业务目标;

(2) 科学计算工作流一般只限于某一专业领域,而应急处置工作流多涉及跨管理域、跨部门的业务应用。更多的工作应侧重于跨领域的语义的共享与互操作问题的解决。

7 结论与展望

以应急处置业务领域为应用背景,本文提出了一个分层工作流模型,同时提出了一个支持工作流逐层映射的元数据分层模型,以此为基础构建了一个基于元数据导航的服务工作流的动态映射和装配执行的实现机制。该机制能够实现从业务工作流到抽象工作流、执行工作流,最终到分布、动态、异构资源和服务实例的逐层自动映射,并指导用户临机进行半自动化选择和绑定,并最终完成部署和执行。

该机制的先进性如下:建立了一个分层的应急处置领域本体和 VO 本体模型。该本体模型的实现为基于语义网技术研究工作流问题提供了一个良好的基础设施;基于本体概念对应急处置工作流和元数据进行分层建模,较好地支持了工作流的分层映射和跨业务领域的语义共享与互操作问题;文章中所提出的基于元数据导航的服务工作流的装配机制,主要针对应急处置领域的业务特点,侧重于解决应急业务过程的有效协同和集成问题,以达成某一业务目标,原型系统的实验结果,也较好地验证了这一模型的先进性和有效性。

下一步我们将重点研究以下几个方面的内容:基于语义推理的业务工作流的自动生成;工作流的逐层装配和编辑(人机交互)工具;该机制对临机决策问题(如突发事件的类型、规模等超出预期,常规的处置手段无法有效地进行化解,要求决策者临机调整处置方案进行应对)的支持等。

参 考 文 献

- 1 Workflow Management Coalition. Workflow Standard—Interoperability Wf-XML Binding. Document Number WFMC-TC-1023, Version 1.0. <http://www.wfmc.org>, May 2000

- 2 Anderson M. . Workflow Interoperability-Enabling E-Commerce. <http://www.wfmc.org/standards/docs/IneropChallengePublic.pdf>, 2003
- 3 Uschold M. , Gruninger M. . Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 1996, 11 (2): 93~155
- 4 Gruber T. R. . A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993, 5(2): 199~220
- 5 Gruber T. R. . Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 1995, 43(5/6): 907~928
- 6 Guarino N. . *Formal Ontology and Information Systems*. Trento, Italy: FIOS Press, 1998, 3~15
- 7 Foster I. , Kesselman C. , Tuecke S. . The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 2001, 15 (3): 200 ~ 222
- 8 Nilsson N. J. . *Artificial Intelligence: A New Synthesis*. Beijing: China Machine Press, San Francisco, CA: Morgan Kaufmann Publishers, Inc, 1999, 215~316
- 9 Russell S. , Norvig P. . *Artificial Intelligence: A Modern Approach*. Pearson Education North Asia Limited. Beijing: People's Posts & Telecommunications Press, 2002, 229~231
- 10 Toward a unified theory of meta-data. By Bill Inmon for The Data Administration Newsletter, 12/19/2002. <http://www.tdan.com/i022fe03.htm>
- 11 FGDC: Content Standard for Digital Geospatial Metadata. Version 2, 1998
- 12 Casanova M. A. , Coelho T. , Carvalho M. T. M. , Corseuil E. T. L. , Nóbrega H. , Dias F. M. , Levy C. H. . The design of XPAE—An emergency plan definition language. In: *Proceedings of the IV Workshop Brasileiro de Geoinformática (GEOINFO)*, Caxambu, Minas Gerais, Brasil, 2002, 25~32
- 13 Deelman E. , Blythe J. , Gil Y. , Kesselman C. , Mehta G. , Patil S. , Su M. , Vahi K. . Pegasus: Mapping scientific workflows onto the grid. In: *Proceedings of the 2nd European Across Grids Conference*, Nicosia, Cyprus, 2004
- 14 Deelman E. , Blythe J. , Gil Y. , Kesselman C. . Pegasus: Planning for Execution in Grids. *GriPhyN Technical Report 2002-20*. <http://www.isi.edu/~deelman/Pegasus/pegasus-overview.pdf>
- 15 Ludäscher B. , Altintas I. , Gupta A. . Compiling abstract scientific workflows into Web service workflows. In: *Proceedings of the 15th International Conference on Scientific and Statistical Database Management (SSDBM)*, Boston, Massachusetts, 2003, 251~254
- 16 Altintas I. , Memon A. , Ludäscher B. *et al.* Design and execution of scientific workflows using Web Services. In: *Proceedings of the Presentation*, San Diego Supercomputer Center and San Diego Software Industry Council to Host Web Services Conference, 2004
- 17 Altintas I. , Bhagwanani S. *et al.* A modeling and execution environment for distributed scientific workflows. In: *Proceedings of the 15th International Conference on Scientific and Statistical Database Management (SSDBM)*, Boston, Massachusetts, 2003
- 18 Altintas I. , Ludäscher B. . A modeling and execution environment for scientific workflows. In: *Proceedings of the Presentation*, Supercomputing 2003, Phoenix, AZ, 2003



WANG Yue-Long, born in 1967, Ph. D. candidate. His research interests include WebGIS and joint emergency response system.

WANG Wen-Jun, born in 1970, Ph. D. . His research interests include WebGIS and joint emergency response sys-

tem.

LUO Ying-Wei, born in 1971, Ph. D. , associate professor. His research interests include distributed computing, software agent and WebGIS.

WANG Xiao-Lin, born in 1972, Ph. D. , associate professor. His research interests include component technology, software agent and WebGIS.

XU Zhuo-Qun, born in 1936, professor, Ph. D. supervisor. His research interests include computer visualization, parallel computing, AI and WebGIS.

Background

This work is supported by the National Basic Research Program (973 Program) of China under grant No. 2002CB312003; the National Research Foundation for the Doctoral Program of Higher Education of China under grant No. 20020001015; the National Natural Science Foundation of China under grant No. 60203002. All those projects are studied around spatial information processing, sharing and integrating, so as to realize an integrating platform for distributed and heterogeneous spatial information that possesses author's own copyright.

The group has focused on research and development of WebGIS for over ten years. They studied some key technologies in spatial information science, and developed several GIS

platform and tools. Also, they applied their platform and tools in many application fields, such as digital city, joint emergency response system, forest, land use and so on.

According to the requirements of city emergency response, this paper aims to design a layered emergency workflow model based on domain ontology and VO ontology, and at the same time a hierarchical metadata infrastructure is designed to provide navigation for assembly and deployment of emergency workflow based on distributed services and resources. A prototype system of IERS based on the idea is implemented and a demonstration of fire emergency response scenario is given.