

文章编号:1001-9081(2009)01-0315-04

基于复用的工作流过程定义工具——PDTBR

龚晓庆, 刘 锋, 葛 玮, 郝克刚

(西北大学信息科学与技术学院, 西安 710069)

(gqx@nwu.edu.cn)

摘要:为了提高工作流过程定义的效率,提出了一种基于复用的工作流过程定义方法。文中讨论了构建和描述可复用资产——领域业务本体和工作流模板的方法,给出了基于领域业务本体检索工作流模板的方法。在此基础上,开发了基于复用的过程定义工具 PDTBR,并在实际使用中验证了该方法的可行性与有效性。

关键词:工作流; 工作流过程定义; 软件复用; 工作流模板; 领域业务本体

中图分类号: TP311.52 **文献标志码:**A

Workflow process definition tool based on reuse – PDTBR

GONG Xiao-qing, LIU Feng, GE Wei, HAO Ke-gang

(College of Information Science and Technology, Northwest University, Xi'an Shaanxi 710069, China)

Abstract: An approach to define workflow process on the basis of reuse was proposed in order to improve its efficiency. The process for producing and describing the reusable assets, domain business ontology and workflow template were presented here. Moreover, a workflow template retrieval method based on domain business ontology was stated. A reuse-based process definition tool was developed as a result. It has demonstrated that the approach is feasible and more productive in application.

Key words: workflow; workflow process definition; software reuse; workflow template; domain business ontology

0 引言

工作流管理技术是实现企业业务过程重组、过程管理与过程自动化的核心技术。工作流过程定义就是将业务过程描述成计算机可处理的形式化表示,它一直是工作流技术研究的重要课题之一^[1]。工作流过程定义既要明确描述业务过程中的活动及活动间的关系,又要定义活动间所传递的信息,执行活动的应用程序、活动所需要的资源等内容。只有从不同的角度详尽地了解了业务过程,才能完整地描述业务过程,因此过程定义被普遍认为是一项耗时而困难的工作。

软件复用被公认为是提高软件开发效率和开发质量的有效途径,而面向领域的复用更容易取得成功。本文研究了工作流过程定义中的复用问题,提出了一种基于复用的过程定义方法,该方法不再采用一切从零开始的方式进行过程定义,而是复用过去积累的知识与经验,能够降低过程定义的复杂度,提高过程定义的质量。本文考虑了过程定义在两个不同层次上的复用。在概念层次上,利用领域业务本体帮助开发人员理解领域业务、捕获具体系统的业务需求,促进特定领域通用业务知识的共享;在实现层次上,利用工作流模板作为通用业务过程的标准化设计。二者相结合,利用领域业务本体为工作流模板提供语义描述,提出了基于领域业务本体检索工作流模板的方法,建立了构建领域业务本体、生产工作流模板以及基于它们进行工作流过程定义的过程。在此基础之上开发了基于复用的过程定义工具 PDTBR,并在实际使用中初步验证了本文研究成果的可行性和有效性。

1 工作流技术和软件复用

工作流过程定义是对业务过程进行形式化的描述。如何快速、准确地进行过程定义关系到是否能满足企业的业务需求,直接影响到企业业务过程的运行效率。

西北大学软件工程研究所多年来对工作流过程定义进行了深入的研究,提出了信牌驱动式工作流计算模型^[2],并不断进行完善,使其具有了较强的表达能力^[3]。基于该过程模型,研究所与西安协同数码公司合作开发了工作流管理系统 SynchroFlow,它已经成功地被电子商务、办公自动化等领域中多个项目所采用。在实际应用中,我们发现过程定义仍是耗时而且困难的,完整地描述业务过程涉及到资源、信息等不同方面的视图,需要付出大量的劳动。同时还发现过程定义是一项依赖于经验的工作,一些有经验的开发者进行过程定义的效率远高于新手,原因是他们曾经进行过类似系统的过程定义,一方面,他们对业务领域知识较为了解,另一方面,他们常常能凭记忆在以前的项目中寻找可复用资产,例如已有的过程定义和 jsp 页面等。由此可以看出,如果能将软件复用的思想引入过程定义,为工作流过程定义提供系统化的复用机制,将大大提高过程定义的效率。

软件复用是指重复使用“为了复用目的而设计的软件”的过程。软件复用中的一些成熟理念(例如面向特定领域考虑复用问题)为研究工作流过程定义中的复用问题提供了重要的指导作用。“领域”是指一组具有相似和相近软件需求的应用系统所覆盖的功能区域^[4]。软件复用的经验显示:领域具有内聚性和稳定性、可复用信息具有领域特定性^[5],因此同一领域中容易获得复用度较高的资产,可复用资产的存

收稿日期:2008-09-09。基金项目:国家 863 计划项目(2007AA010305)。

作者简介:龚晓庆(1974-),女,陕西泾阳人,讲师,博士,CCF 会员,主要研究方向:软件工程、工作流; 刘峰(1981-),男,陕西西安人,硕士研究生,主要研究方向:软件工程; 葛玮(1960-),男,陕西西安人,副教授,主要研究方向:软件工程、工作流; 郝克刚(1936-),男,陕西西安人,教授,博士生导师,主要研究方向:软件工程、软件理论。

储、检索等也比较容易,面向领域的软件复用容易取得成功。

目前,市场上大多数工作流管理系统都是各个领域通用的产品,然而同一领域中业务功能需求、组织机构等方面都较为类似,可以在过去成功的项目中分析整理、抽象出可以复用的资产,对相同领域新系统进行过程定义时定制、复用它们,这样在通用的工作流管理系统上提供针对特定领域的可复用资产,避免了从零开始设计工作流的低效方式。

针对工作流过程定义中的复用问题的研究成果比较少,其中比较有影响的是文献[6]提出的工作流模板(workflow template)的概念。但现有文献对工作流模板的讨论都比较框架化,对于工作流模板的表示、定义、描述、存储、检索、生产及使用过程都没有进行进一步的研究,而且对工作流过程定义中较高层次的复用(如业务知识等)都没有考虑。总体而言,目前对过程定义中复用问题尚没有一个系统化的解决方案。

2 基于复用的工作流过程定义方法

基于复用的工作流过程定义方法的框架如图1所示。

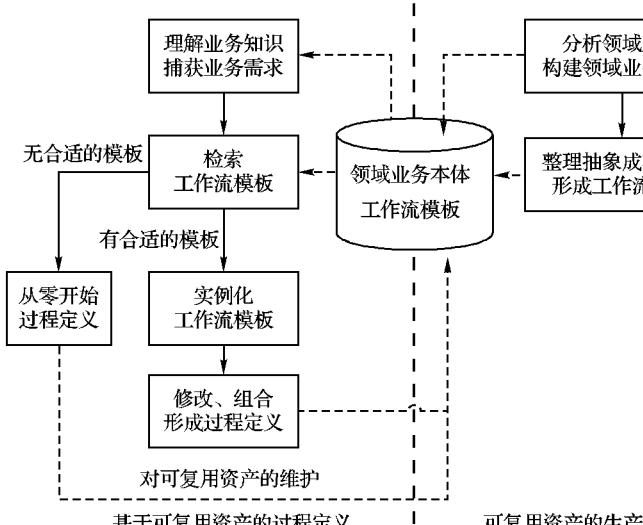


图1 基于复用的工作流过程定义方法的框架

图1中垂直虚线右边体现了可复用资产的生产,需构建领域业务本体,生产工作流模板并进行不断积累和维护。虚线左边的是基于可复用资产进行过程定义的过程。在该过程中,具体系统的开发者以领域业务本体为背景和用户进行交流,理解业务知识,捕获业务需求,检索合适的工作流模板,对其实例化,然后通过适应性修改后以及组合后形成具体系统的过程定义。

基于可复用资产进行工作流过程定义包括四个阶段,它是一个迭代的、逐渐精化的过程,简要描述如下:

- 1) 基于领域业务本体对具体系统进行业务分析,形成业务功能层次模型,确定具体系统的角色、组织单元及其相互关系。
- 2) 根据具体系统的业务需求检索工作流模板并进行适应性修改。
- 3) 开发工作流模板库中不存在的过程定义。
- 4) 对可复用资产进行维护,不断促进可复用资产的完善和充实。

在这个框架下还需要解决以下问题:

- 1) 领域业务本体的定义、描述以及构建领域业务本体的方法。
- 2) 在继承工作流模板思想的基础上,进一步研究工作流

模板的表示、工作流模板间的关系、工作流模板生产过程、工作流模板的管理、工作流模板的检索等相关问题。

3 领域业务本体

3.1 领域业务本体的定义

本体是“共享概念模型的明确的形式化规范说明^[7]”。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,并从不同层次的形式化模式上给出这些词汇(术语)和词汇之间相互关系的明确定义。本体可以作为领域知识的形式化描述,这种形式化描述可以作为软件系统中可复用的组件。

本文提出领域业务本体目的在于帮助开发者从宏观上理解特定领域的业务知识,并用其描述工作流模板的语义,引导开发者去获取体现了业务过程具体实现的工作流模板,因此,本文限定主要从以下几个方面刻画特定领域的业务:

1) 业务功能。描述领域中需要完成的工作或者任务。

2) 资源。指业务功能执行中需要被使用或者产生、消耗的实体,包括信息、人员、工具等方面。

3) 组织单元。指人们根据组织职能和组织实际而设计的结构化元素,如某机构、部门等。

4) 角色。是对人员按所承担职责的一种划分。

根据以上分析,可抽象出业务功能、信息、组织单元、角色这几个顶层的概念类型。其中对业务功能概念类型可进一步划分为原子业务功能和复合业务功能。识别出属于这几种概念类型的概念集合后还需要建立概念间的关系,表1列出了一些概念之间的语义关系,关系由箭头表示,形成规则箭头两边分别是源概念以及导出概念的概念类型。表1定义了概念类型及概念间的关系,它并不涉及到具体领域的业务知识,而是规范化和系统化了特定领域中业务知识的描述,领域业务本体可看作是它在特定领域的实例化。

3.2 领域业务本体的描述

OWL语言是目前较为通用的本体描述语言,本文采用OWL描述领域业务本体。限于篇幅,不在此举例说明。

3.3 构建领域业务本体

领域业务本体的构建必须对特定领域的业务知识有着详尽地理解,从而进行概念的获取。领域中业务知识的来源一般有领域专家、领域相关的业务文档和领域中已经成功应用的工作流系统。

构建领域业务本体是领域业务知识概念化和形式化的过程,其任务包括:

- 1) 建立概念和概念间的关系。

自顶向下的方法:a)识别领域中的顶层业务功能,将其加入概念集合中。b)对可分解的业务功能进行分解,将分解得到的业务功能加入概念集合中。对其中的原子业务功能,分析其输入、输出信息以及角色和组织单元,也将这些概念也加入概念集中,并建立与业务功能以及其他概念的关系。c)对可分解的业务功能不断重复b的操作。

自底向上的方法:从各种知识来源中提取有关的概念,加入概念集合中。然后不断地对概念归类,确定概念之间的关系,自底而上,循环反复抽取、归并概念及其关系的方法获取领域业务本体。

这两种方法可以综合使用,细致、反复地进行,以免遗漏相关概念而造成知识的不完整。

2) 定义概念的属性。针对概念集合中每个概念定义属性,确定属性取值的类型以及属性约束。这一步与上一步是密不可分、相互交织的,两者必须同时进行,成为领域业务本

体构建中最为重要的两步。

3) 使用 OWL 语言表示领域业务本体。

需要说明的是,本体的构建和完善是一个反复迭代的过程,不可能一蹴而就,必须不断地对领域业务本体进行维护和完善。

表 1 一些概念间关系的定义

关联	形成规则	说明
part-of	业务功能→业务功能 组织单元→组织单元	整体和部分的关系,表明业务功能、组织单元可以分解
Is-a	业务功能→业务功能 角色→角色 信息→信息	表示角色、信息等概念间的泛化关系
performer	原子业务功能→角色 原子业务功能→组织单元	表示参与业务功能执行的角色或者组织单元
Input	信息→原子业务功能	表示为执行业务功能输入的信息
Output	原子业务功能→信息	表示执行业务功能后输出的信息

4 工作流模板

4.1 工作流模板的定义

一个工作流模板是一个通用业务过程的标准化定义,设计者可根据实际需求对其进行剪裁和修改从而产生新的过程定义^[6]。我们将工作流模板理解为实现了某业务功能的工作流过程定义的封装,外在的等同于一个活动节点。具体地说,工作流模板实际上是一个参数化的过程定义,它的实现中可保留一些与具体系统相关的待定参数,复用时再根据具体情况给予指定,这就是实例化工作流模板的过程。

领域业务本体中的概念由其相互关联刻画了内涵,含有公认的语义。工作流模板实现了某个业务功能,它和领域业务本体中某个业务功能类型的概念相关联;工作流模板的输入、输出数据,以及参与的角色、组织单元都引用领域业务本体中的概念进行描述。这样就通过领域业务本体中的概念刻画了工作流模板的语义。

具有语义的工作流模板由描述、接口和实现三部分组成:

1) 描述。

a) 基本信息的描述:如工作流模板的名称、id 号、作者、提交日期、版本、描述等。

b) 语义的描述:包括一个四元组: $\langle function, input, output, participant \rangle$,其中: $function$ 对应于领域业务本体中某个业务功能类型的概念; $input, output$ 是领域业务本体中信息概念类型的概念集合,体现了输入、输出数据的接口的语义。 $participant$ 是领域业务本体中组织单元、角色概念类型的概念集合,是工作流模板涉及到的参与者。

2) 接口。

a) 工作流模板输入、输出数据的接口。

b) 参数化接口:工作流模板实例化时根据具体系统的情况,通过这个接口对工作流模板实现中待定的参数加以制定,如具体指定活动的参与者等信息。

3) 实现。

它是某业务过程形式化描述的文本,其中保留了一些待定参数。实例化时,通过对参数的替代,该部分成为一个具体

的过程定义。本文使用信牌驱动的工作流模型,它定义的业务过程能转化为 XML 格式的文本。由于工作流模板的实现外在地等同于一个活动节点,因此在信牌驱动工作流模型中加入一类新的活动节点(模板活动)就可以实现工作流模板的嵌套,即一个工作流模板的实现中可以使用其他的工作流模板。

本文采用 XML 描述工作流模板,图 2 给出工作流模板描述的 XML Schema,由于页面有限,没有展开所有的图中的所有 Element。工作流模板的整个 XML Schema 文档篇幅较长,这里不再给出。

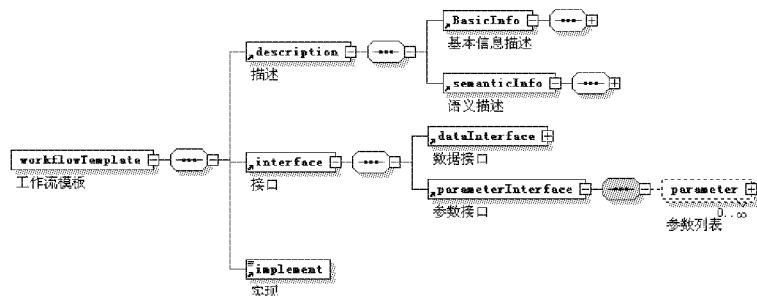


图 2 工具 XML SPY 中显示的工作流模板的 XML Schema 结构

4.2 工作流模板的生产

工作流模板的生产一般由领域工程师来完成。针对领域中曾经成功应用的工作流项目的过程定义,对其进行抽象,将和具体系统相关的部分定义为参数(如活动的参与者),从而形成工作流模板。

4.3 基于领域业务本体的工作流模板检索

特定领域中工作流设计者能根据以下两种方式从库中检索相应的工作流模板:

1) 关键字检索。根据基本信息的关键字,如作者、入库日期等检索满足条件的工作流模板。

2) 语义匹配。工作流模板入库时利用领域业务本体中的概念描述了其语义信息,可以计算用户查询的关键字在领域业务本体上和工作流模板语义描述信息的相似度来检索工作流模板。

语义相似度用以度量概念间语义的相似程度,本文采用了如下的计算方法^[8]:

设 t_1 和 t_2 是本体中的两个概念, $\text{Sim}(t_1, t_2)$ 表示这两个概念之间的相似程度, 则有公式:

$$\text{Sim}(t_1, t_2) = \sum_{i=1}^n \delta_i(t_1, t_2) \theta_i \quad (1)$$

其中: n 是术语 t_1 和 t_2 在本体中 kind-of 关系的层次中所具有的最大深度; θ_i 是权重(可简单地取 $\theta_i = 1/n$); 当 t_1, t_2 前 i 个父类相同时, $\delta_i(t_1, t_2) = 1$; 当 t_1, t_2 前 i 个父类不相同时, $\delta_i(t_1, t_2) = 0$ 。 Sim 返回 $[0, 1]$ 的实数表示两个概念间的语义相似度。

两个概念集之间的相似度计算方法: 设 $M = \{m_1, m_2, \dots, m_x\}$ 和 $N = \{n_1, n_2, \dots, n_y\}$ 是两个概念的集合, $\text{SimOfSet}(M, N)$ 表示这两个概念集之间的相似度, 则有:

$$\text{SimOfSet}(M, N) = \sum_{i=1}^x \sum_{j=1}^y \text{Sim}(m_i, n_j) \quad (2)$$

基于领域业务本体检索工作流模板的方法为: 用户查询时输入一个概念集合, 对于库中的每个工作流模板的语义描述的概念集, 利用式(2)计算得出两个概念集的相似度, 进行排序后显示出来供用户选择。

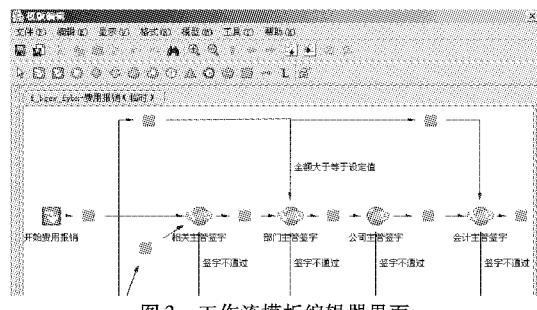
在具体实现上, 我们使用文献[9]提供的 API 对 OWL 描述的本体模型进行操作, 利用 Jena 中的方法可以获得相应概念集合来求解概念语义相似度:

- 1) 读入 OWL 描述的领域业务本体。
- 2) 使用 `ontModel` 的 `getOntClass()` 得到两个需要计算相似度的概念的 `OntClass`。
- 3) 使用 `OntClass` 的 `getSuperClass` 方法可用来得到这两个概念的超类列表。
- 4) 计算两个超类列表中超类相同的个数 n 。
- 5) 利用上面的公式计算出两个概念的语义相似度。

5 基于复用的工作流过程定义工具(PDTBR)

SynchroFlow 已有的过程定义工具主要包括组织机构建模、过程定义和界面设计几个功能模块。PDTBR 在原有 SynchroFlow 的过程定义工具扩充了管理和使用特定领域的可复用资产的机制, 主要包括以下功能:

- 1) 查看工作流模板。
- 2) 选择工作流模板。可以通过浏览和检索两种方式从工作流模板库中根据需要选择合适的模板。
- 3) 细化、组合、修改工作流模板。可以在工作流模板的编辑环境中实现这三个功能。



- 4) 删除工作流模板。
- 5) 添加工作流模板。添加工作流模板时要通过入库登记对工作流模板的以下几方面对其描述:a) 工作流模板的基本信息;b) 语义描述(使用领域业务本体中的概念描述工作流模板)。
- 6) 实例化工作流模板。进行具体系统的工作流过程定

义时, 系统的开发人员根据具体的需求选择工作流模板后进行实例化, 即根据具体的情况指定工作流模板中的参数, 例如模板中活动的参与者等。实例化后的工作流模板成为一个能被工作流管理系统执行的过程定义, 被存入项目的过程定义库中供使用。

- 7) 使用情况反馈。
- 8) 领域业务本体的编辑。基于工具包 Jena 实现了一个简单的本体编辑工具, 用户编辑本体后可以将本体存储成为 OWL 语言, 这样可在概念的层次上设计领域业务本体, 而不需要熟悉具体的 OWL 描述语言。

9) 用户管理。



图 4 简单的本体编辑器

6 结语

本文研究了特定领域工作流过程定义中的复用问题, 提出了基于复用的工作流过程定义方法, 该方法以已有的工作为基础, 充分利用过去工作流系统开发中积累的知识和资产, 降低了过程定义的复杂度。工作流模板等可复用的资产是领域中经受过实践检验的解决方案, 对它们进行复用能有效地提高过程定义的质量。基于该方法实现的过程定义工具 RBPDT 在实际开发中的使用效果也证明了其可行性与有效性。当然, 基于复用的过程定义方法依赖于可复用资产的积累, 随着特定领域中可复用资产不断地积累与完善, 在同一领域工作流过程定义的效率也将不断地提高。对于工作流模板在入库前的质量评价和后期度量也是有待研究的问题。

参考文献:

- [1] 罗海滨, 范玉顺, 吴澄. 工作流技术综述[J]. 软件学报, 2000, 11(7): 899 - 907.
- [2] 岳晓丽, 杨斌, 郝克刚. 信牌驱动式工作流计算模型[J]. 计算机研究与发展, 2000, 37(12): 1513 - 1519.
- [3] 龚晓庆, 葛玮, 郝克刚. 信牌驱动模型中的工作流模式[J]. 西北大学学报, 2004, 34(1): 13 - 18.
- [4] LI K Q, CHEN Z L, MEI H, et al. An introduction to domain engineering [J]. Computer, 1999, 26(5): 21 - 25.
- [5] 杨美清, 王千祥, 梅宏, 等. 基于复用的软件生产技术[J]. 中国科学: E 辑, 2001, 31(4): 363 - 371.
- [6] van der AALST W M P, BASETN T. Inheritance of workflow: an approach to tackling problems related to change [J]. Theoretical Computer Science, 2002, 270: 125 - 203.
- [7] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering, principles and methods[J]. Data and Knowledge Engineering, 1998, 25(1/2): 161 - 197.
- [8] EUZENAT J. State of the art on ontology alignment[EB/OL].[2005-09-19]. <http://www.jarrar.info/Publications/D2.2.3-Alignment.pdf>.
- [9] Mc BRIDE B. Jena: Implementing the RDF model and syntax specification[EB/OL].[2004-10-27]. <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-40/mcbride.pdf>.