

基于领域本体面向问题的需求分析与领域建模

刘燕玲, 华庆一, 郭晓娟

(西北大学 软件工程研究所, 陕西 西安 710127)

摘要:为了解决传统的以应用领域为目标、以问题解决方案为中心的需求分析方法造成的开发人员与需求人员以及客户三者之间的误解,需求分析不可再次使用进而导致系统开发效率低下,客户满意度低等问题,文中提出了基于本体面向问题进行需求分析的方法。阐述了基于本体面向问题进行需求分析的必要性、本体的相关概念和设计原则、本体的描述语言以及基于 LAU 建立领域本体的实例与需求分析过程,并论证了利用该方法可进一步提高需求分析的质量。

关键词:本体;领域本体;面向问题

中图分类号:TP311.52

文献标识码:A

文章编号:1673-629X(2007)08-0099-04

An Ontology - Based, Problem - Oriented Requirements Analysis and Domain Modeling Method

LIU Yan-ling, HUA Qing-yi, GUO Xiao-juan

(Institute of Software Engineering of Northwest Univ., Xi'an 710127, China)

Abstract: The traditional requirements analysis takes the application domain as the goal and takes the solution of the questions as the center, which induces the misunderstanding among developers, requirement analyzers and users and the result of requirements analysis is non-reusable and ulteriorly causes the system development's efficiency and the satisfaction degree of customers low. To solve these problems, proposed a based on ontology and problem-oriented requirements analysis. Expatiates the necessity to carry on the based on ontology and problem-oriented requirements analysis, the related concepts, the principles of design and description language, offers the domain ontology instance using LAU and the process of requirements analysis and demonstrates that the based on ontology and problem-oriented requirements analysis can improve the quality of requirements analysis.

Key words: ontology; domain ontology; problem-oriented

0 引言

随着软件工程的不断发展,开发软件的规模和复杂度不断增加,软件开发的中心任务逐渐由过去的代码实现为主变为现在以需求分析为主。需求分析成为软件开发中的难点,往往花费大量的资源仍无法保证需求分析的效率和质量。需求分析中最大的困难之一是开发者对目标领域背景知识的缺乏,以及目前的研究是以问题的解决方案为中心,忽略了对问题域的研究,造成开发者与用户之间缺乏共同的交流语言。一方面开发者无法从用户那里得到系统的和前后一致的需求;另一方面,在开发者和用户之间经常产生误

解。因此需求提取具有很大的难度,而且难以重用。

为了摆脱上述困境,开展基于本体面向问题域的需求分析方法的研究是十分必要的。Jawed Siddigi 等人在文献[1]中指出,软件是某种目标机器的描述,它的开发就是这种机器的构建,其需求描述了目的,一个机器的目的是在机器之外发现的,即在问题域的上下文中发现的,所以需要转向一种面向问题的开发方法。而实践证明对问题领域的描述分析,本体论的方法较之分类理论方法是更有效的。

Ontology 最早是一个哲学上的概念,从哲学的范畴来说,Ontology 是客观存在的一个系统的解释或说明,关心的是客观现实的抽象本质。在计算机领域关于本体的定义没有标准,以下三种是比较有代表性的:“Ontology 是概念模型的明确的规范说明”(Gruber^[2]);“Ontology 是共享概念模型的形式化规范说明”(Borst^[3]);“Ontology 是共享概念模型的明确的形式化规范说明”(Studer^[4])。本体的定义概括起来主

收稿日期:2006-10-19

基金项目:国家自然科学基金资助项目(60073050);陕西省自然科学基金资助项目(2000X15)

作者简介:刘燕玲(1982-),女,河南尉氏人,硕士研究生,研究方向为人机交互;华庆一,教授,博士生导师,研究方向为人机交互、软件工程。

要包含 4 层含义:概念模型 (conceptualization), 明确 (explicit)、形式化 (formal) 和共享 (share)。领域本体作为一种语义模型能够对领域知识进行详细的描述, 表达了领域概念之间的语义关系, 它的提出实现了需求的重用。和传统的需求分析方法相比较, 基于本体的应用需求的获取是基于应用本体的, 而应用本体可以根据领域本体获得, 大大节省了需求分析的时间。

1 本体的相关概念和设计原则

1.1 相关概念

在知识工程界, 最早给出本体论概念的是 Nehces 等人, 他们将本体定义为: 相关专题的基本术语和关系, 以及利用这些术语和关系构成该专题规则的集合。即某个领域的本体就是关于该领域的公认的概念集, 该概念集包含确定的语义和概念之间的关联, 概念是以本体表征出来的。概括地讲, 本体是面向领域的概念化描述, 它是多个代理之间的概念及其关系的描述, 被用于不同代理之间的知识共享、集成及重用。基于以上对本体的相关描述, 给出本体的相关定义:

定义 1 本体是某一领域 D 的概念化描述, 它包含一系列概念及其关系。即 $O = \{CS, RS\}$, 其中 CS 为概念集合, RS 为关系集合。

定义 2 概念 (Concept) 总与一类对象集合相对应, 可视为领域 D 中一类对象的抽象描述, 即 $C_i = \{O_{i1}, O_{i2}, \dots, O_{in}\}$, 其中 O_{ij} 为领域 D 中的对象。

定义 3 关系 (Relation) 用于描述概念所包含的对象之间的个别联系, 设 C_1, C_2, \dots, C_n 为 n 个不同的概念, 称任意属于笛卡儿乘积 $C_1 \times C_2 \times \dots \times C_n$ 的子集 $R (R \subseteq C_1 \times C_2 \times \dots \times C_n)$ 为 C_1, C_2, \dots, C_n 之间的一个关系。

1.2 本体的设计原则

当本体被设计的时候, 要选择和决定应该在本体中表示什么, 为了指导本体的设计和评估, 不少研究人员从实践出发, 提出了各自的本体设计标准, 其中最有影响的是 Gruber^[2] 提出的 5 条规则:

1) 明确性 (Clarity) 和客观性: 即本体应该用自然语言对所定义词汇 (terms) 给出明确的、客观的语义定义。

2) 一致性 (Coherence): 即由本体中词汇得出的推论不能和词汇定义相矛盾。

3) 可扩展性 (Extendibility): 一个本体被设计时应该要预料到共享词汇表的用途, 便于扩展。

4) 最小代码依赖 (Minimal encoding bias): 概念模型应该定义在知识级上而不依赖于一个特定的符号级编码。因为共享的概念模型可能用不同的表示系统和

表示形式来实现。

5) 最小本体约束 (Minimal ontological commitment): 本体应该在充分达到目的基础上满足最小本体约束准则。即对建模对象应给出尽可能少的约束, 以便让本体用户可按需要自由地特殊化和实例化本体。

2 本体的描述语言 LAU

在软件开发中, 图形化语言直观、形象的特点在用户、领域专家以及开发人员交流受到了青睐, 像 LINGO 语言^[4]。LINGO 语言的优点: 简单易用, 而且具有强语义关系的优势。但这种语言不为多数人所熟悉。文中研究的领域本体主要用于软件需求分析, 所以应该找一种为该领域大多数人所熟悉的图形化语言, 我们可能会想到 UML 语言。Cranefield 在论文中提出了一种用 UML 来表示本体的方法^[5], 即把本体当作一个静态模型, 模型中包含了类图、关系符号和对象图。在类图中, 每个类由类名、类属性 (包括名字、类型和可见性) 和类的操作 (函数) 组成。但是 UML 过于复杂和庞大, 违反了最小本体约束原则。而且 Cranefield 用面向对象的方法来表示本体, 这是不可取的, 本体是一种形式化的概念; 另外, 本体应该是领域的形式化模型, 而 UML 语言缺乏一定的语义功能。基于这两种语言存在的问题以及可用性, 根据本体的定义和设计原则提出了 LAU 语言, 该语言抽取必用的 UML 图形符号, 引入 LINGO 的强语义符号, 并在此符号上定义约束, 和 Cranefield 的方法不同, 此处并不引入面向对象的类、属性等概念。

LAU 的基本图形符号包括以下四种: 概念 (concept) 符号、普通关系 (relation) 符号、强语义关系符号和约束注释符号, 如图 1 所示。

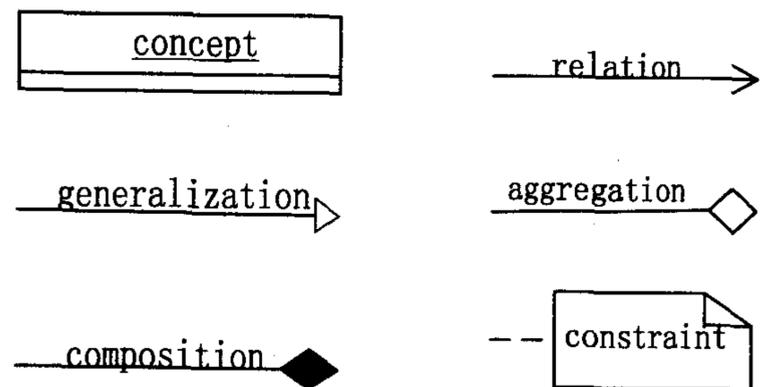


图 1 LAU 的基本图形符号

概念符号使用 UML 中的对象符号, 在此其仅仅代表一个概念, 没有面向对象中对象的含义, 在符号上可以注明概念名, 普通关系符号其上可以注明关系名。强语义关系符号只定义了三种: generalization, aggregation 和 composition。可以在此符号上定义约束条件, 以增强图形符号的语义, 如果在实际应用中需要, 还可以定义其他的强语义关系符号。最下方是约束注释符

号:用于将约束加在概念和关系上。

3 领域本体的创建

采用面向问题域的方法,可以将应用系统的需求看作是问题域中现象之间的关系,而需求规范就是一种限定性的需求。对问题域中的现象进行表达,需要对问题域本身进行建模。问题域建模要涉及人们认识事物、解决问题的方法。

利用本体论的方法对问题域进行分析是比较有效的。用户的业务环境可以看作是一个社会系统,组织理论和角色理论是进行社会系统行为及特征分析的重要方法。组织的存在是因为人们为了完成单个人无法完成的任务而形成的一种相互依赖的关系。角色理论告诉人们相互依赖意味着一个人的活动可能受其他人活动的影响,这种相互依赖、相互影响的关系受组织内部多种规则的约束,这种约束是保证组织有效运转的重要前提。组织规则表达了通用的、全局的需求,组织结构定义了组织的类型及控制方式。所以组织中的本体是角色、结构及规则。

目前领域本体的创建还没有一个统一的标准,本体对领域知识的表示,综合很多学者提出的各种本体创建的方法,如 Uschold^[6], Gruninger 和 Fox^[7], Falbo^[8], 本体对领域知识的表示是对领域知识的概念化和形式化。结合面向问题域本体的概念,提出一种实用的面向问题的领域本体的创建方法,如图 2 所示。

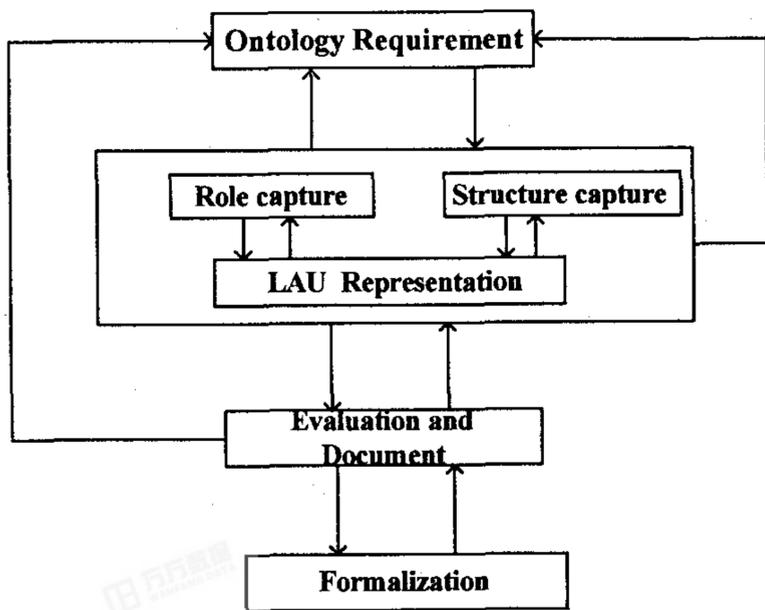


图 2 面向问题的领域本体的创建方法

图中,最上方是本体的需求分析(Ontology Requirement)。中间的方框代表了本体获取(Ontology Capture),方框内包含本体获取的三个子活动:角色获取(Role Capture),应用 LAU 语言的图形化表示(LAU Representation)和结构获取(Structure Capture)。下方是本体评估和文档化(Evaluation and Document)和本体形式化(Formalization)。

以某电信运维流程支撑系统的 ROU 为例,图 3 所示的本体模型图是关于 ROU(Role of user,用户角色)的一个实例。

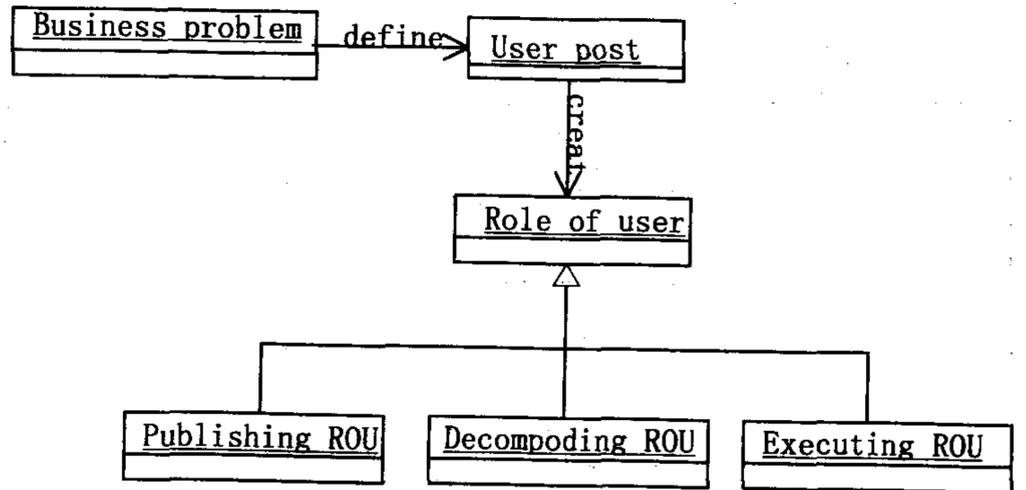


图 3 ROU 本体模型图

ROU 提供了关于用户岗位在业务处理中的相关信息,是业务顺利进行的保证,也是各电信系统进行业务数据组织和交换的主要形式。ROU 本体模型图定义了概念 ROU 表示以及它与用户岗位之间的关系,其它类型的 ROU 的概念是 ROU 的子概念,在图中以泛化(generalization)关系描述。各个类型的 ROU 之间还有约束关系,例如在电信运维流程支撑系统中只有作业计划发布了,才能进行作业计划的分解,分解后才能按照系统根据作业计划自动生成的作业任务进行执行,这就需要结构来制约,规则来规范。

4 基于领域本体面向问题域需求分析过程

基于本体的领域需求分析方法是以领域本体作为需求分析的原模型,给领域本体内的所有人员提供共享以及可相互沟通的领域知识,并以领域本体来指导和规范整个领域需求分析过程。而由上所述可知,面向问题域的需求分析是解决当今软件开发所面临问题的一种方法,文中采用 LAU 语言来表示需求分析的结果,以求能更好地将需求分析作用于系统开发,更好地实现软件的复用。具体的基于本体面向问题域的需求分析过程模型如图 4 所示。

为了全面、深入地分析领域需求,文中将领域分析分为不同的阶段来进行,如图 4 所示,按顺序共分为:业务分析、逻辑分析、本体表示分析 3 个阶段,最后输出领域需求分析的结果:领域需求模型。以下为各个阶段的具体内容。

4.1 业务分析

业务分析阶段的目标是从概念上描述系统的业务框架和过程。这一阶段需求分析人员应该以领域本体提供的领域知识为支撑,与企业业务人员、企业领导等进行全面反复的讨论研究,以获得对业务流程的全面、准确的分析结果。

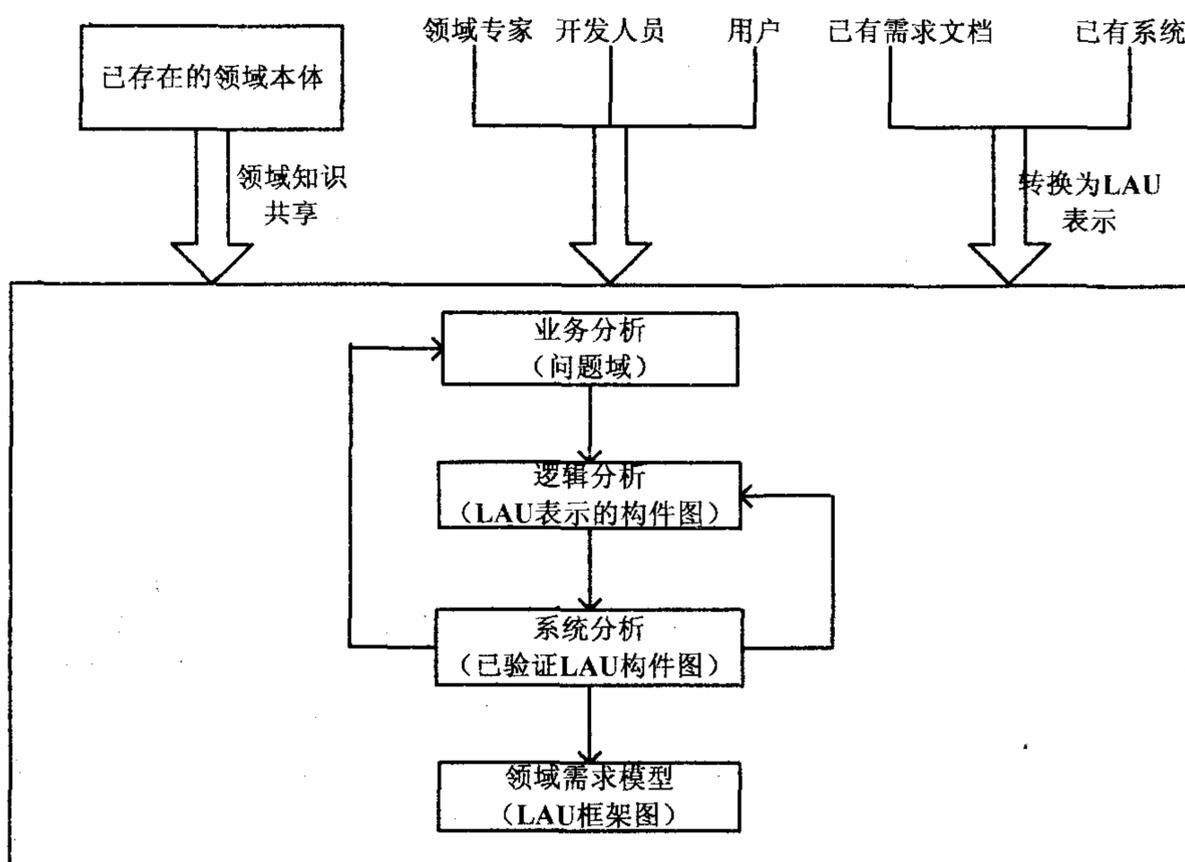


图 4 基于本体面向问题域的需求分析过程

输入:业务需求

输出:领域的问题域

4.2 逻辑分析

逻辑分析在系统业务分析的基础上,对业务中的每个问题进行详细的分析,以确定问题需要参与的角色、角色的作用和角色之间的关系,并且以 LAU 构件图来描述。

输入:问题域

输出:LAU 构件图

4.3 系统可行性分析

系统分析的目的是根据特定的需求识别问题域中的各种构件以及各构件之间的交互关系和约束条件,对问题域的规则进行获取分析、验证。

输入:LAU 构件图

输出:经过验证的 LAU 构件图

4.4 领域需求模型的构造

经过系统分析验证过的构件图准确地表达了问题域对应的需求,各个问题域的构件图的有机结合,可有效地表达系统需求,输出该应用领域的领域需求模型,可用 LAU 框架图来表示。

输入:构件图集合

输出:LAU 框架图

5 结语与未来的工作

提出了一种基于本体面向问题域的需求分析方法,在笔者所参与的某电信运维流程支撑系统需求分析中得到了应用。对于开发者来说,利用面向问题域的需求分析进行领域建模可以提供更一般化的框架。

它由相应的构件类型及结构所组成,通过实例化可以为一个具体问题提供解决方案。利用角色理论对问题域和软件进行建模,可以使领域模型及软件模型基于共同的本体的概念,容易理解,而且问题域的描述及解决方案的描述是同构的不必进行语言的转换,所以更加直接地导向解决问题的方向。

由于目前还没有统一的构建本体的方法,构建者往往采用不同方法、不同工具、不同语言生成本体,因此在同一领域内由于目的不同,使用者不能使用已有的本体,甚至由于语言的不同不能得到确切的映射本体,共同领域内可共享和重用的

本体,由于缺少一个交流的平台,使得其他使用者不了解或者得不到已有本体,因此未来需要更进一步地研究如何将不同的本体间相互转化。

参考文献:

- [1] Siddigi J, Shekaran C. Requirements Engineering: The Emerging wisdom[J]. IEEE Software, 1996, 13(2): 15 - 19.
- [2] Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human Computer Studies, 1995, 43: 907 - 928.
- [3] Borst W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse[D]. Enschede: University of Twente, 1997.
- [4] Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998, 25(1 - 2): 161 - 197.
- [5] Cranefield S, Purvis M. UML as an Ontology Modelling Language[C]// in Proceedings of the IJCAI - 99, Workshop on Intelligent Information, 16th International Joint Conference on AI. Stockholm, Sweden: [s. n.], 1999.
- [6] Uschold M, King M. Towards a methodology for building ontologies[C]// Proc. of the Workshop on Basic Ontological Issue in Knowledge Sharining, International Joint Conference on Artificial Intelligence. Montreal, Canada: [s. n.], 1995: 78 - 92.
- [7] Gruninger M, Fox M S. Methodology for the Design and Evaluation of Ontologies[R]. Toronto: University of Toronto, 1995.
- [8] Falbo R A, Menezes C S, Rocha A R C. A Systematic Approach for Building Ontologies[C]// in Proceedings of the IBERAMIA'98. Lisbon, Portugal: [s. n.], 1998.