

# 基于 OWL 的软件需求获取

李 伟, 陈昊鹏

(上海交通大学软件学院 DSL 实验室, 上海 200240)

**摘 要:** 基于网络本体语言 OWL 支持软件需求获取过程, 通过 OWL 语言建立领域知识的本体, 获取该领域的共享知识, 并基于该领域知识引导需求涉众获取软件需求。该需求获取方式更加系统, 并具有高可复用性和精确性等特点。

**关键词:** 网络本体语言; 资源描述框架; 本体

## Acquirement Software Requirements Based on OWL

LI Wei, CHEN Haopeng

(DSL Lab, School of Software, Shanghai Jiaotong University, Shanghai 200240)

**【Abstract】** This paper tries to support the process of requirement elicitation based on OWL (Web Ontology Language). By creating ontology of the domain knowledge using OWL, the shared domain knowledge can be obtained and thus it can guide stakeholders to acquire software requirement on the basis of such knowledge. This method is more systematical and has the character of high reusability and accuracy etc..

**【Key words】** Web ontology language(OWL); Resource description framework(RDF); Ontology

### 1 概述

在软件系统开发过程中, 有效的需求工程对项目的成败至关重要。软件需求的获取 (elicitation)、分析 (analysis)、协商 (negotiation)、文档化 (specification) 是需求工程中的 4 个主要步骤。在实际开发中, 特别是在基于敏捷软件开发方法的软件项目中, 这个过程往往是一个相互交织、迭代、增量的过程。需求获取过程是其中一个非常关键、重要却又较困难的活动, 其往往与需求分析过程交融在一起。

软件需求难以被捕获, 原因大抵归结为如下 3 点<sup>[1]</sup>:

(1) 基于各种现实目的, 涉众 (stakeholders) 和开发者使用不同的“语言”交流。

(2) 经常难以约束客户, 关于他们想要的东西, 他们不知道, 他们认为他们知道, 但是他们不能清晰地描述系统, 他们时常改变他们的想法, 他们自己很矛盾。

(3) 需求文档模棱两可, 含混不清, 有冗余, 而且内部彼此矛盾。

在实际项目中, 基于情景实例 (Scenario-based) 的软件需求获取方法用得比较广泛, 其采用客户较熟悉的情景实例来引导客户逐步提供系统信息<sup>[2]</sup>。故事版 (storyboard), 情景实例被作为客户与开发者相互交流的“语言”。通过这种方式形象而又不失客观地抽取客户头脑中关于新系统的特征 (功能性和非功能性的需求)。标准化的用例规范 (use case specification) 也较易理解。不过, 这种需求获取方法, 其缺点主要是不能保证所有的系统构建目标都能被捕获; 而且开发人员也要深刻地了解领域系统的知识。

基于知识, 特别是基于本体 (Ontology) 的软件需求获取是目前研究较多的方法<sup>[2]</sup>。其目的就是通过开发者和客户之间建立共享的领域知识平台, 进而引导客户完成需求获取过程, 并且形成准确无二义的需求规范; 与此同时开发团队可充分复用基于共享领域知识的软件需求模型。这种基于

领域共享知识的需求获取方法更加系统, 并具有高可复用性和精确性等特点。

### 2 基于 RDF 数据模型的信息描述

在大型软件开发过程中, 用数学符号形式化的需求文档规范可以使得需求文档无冗余、精确、一致, 减少了文档的二义性。但对开发者来说, 需要接受培训、掌握复杂难懂的形式化的知识 (Z, Petri Net, SOFL 等); 对用户来说, 形式化的文档非常不友好, 难于理解、沟通。如何使得需求文档既能用户友好, 又精确, 一直是软件需求工程界研究的一个热点问题。

资源描述框架 (Resource Description Framework, RDF) 是一种描述 Web 上资源信息的语言 (已被 W3C 组织定义为标准)。RDF 定义了一个机器能处理的数据语义描述的数据模型。这个基本模型的形式定义如下:

- (1) 存在一个名为资源 (R) 的集合;
- (2) 存在一个名为字面值 (L) 的集合;
- (3) 存在一个名为属性 (P) 的集合,  $P \subseteq R$ 。

有一个名为命题 (S) 的集合, 它是三元组 {sub, pred, obj} 的集合, 其中主体 (sub)  $\in R$ , 谓词 (pred)  $\in P$ , 对象 (obj)  $\in L$ 。在 RDF 中, 可以用命题: AuthorOf (基于 OWL 的软件需求获取, 李伟) 来表示本篇文章的作者是李伟这样一个命题。RDF 也可以表示成无序的边带有标识的有向图, 其实质上就是人工智能领域的语义网络<sup>[3,4]</sup>, 如图 1 所示。

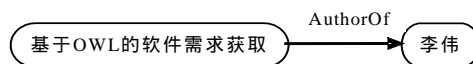


图 1 RDF 数据模型

**作者简介:** 李 伟 (1979 - ), 男, 硕士, 主研方向: 形式化工程方法, 本体工程; 陈昊鹏, 博士后、讲师

**收稿日期:** 2006-02-03 **E-mail:** dragonlw@situ.edu.cn

其所对应的 RDF/XML 的语法为：

```
<?xml version="1.0" encoding="UTF-16"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:mydomain="http://www.mydomain.org/my-rdf-ns">
  <rdf:Description rdf:ID="基于 OWL 的软件需求获取">
    <mydomain:AuthorOf>
      李伟
    </mydomain:AuthorOf>
  </rdf:Description>
</rdf:RDF>
```

RDF 数据模型是领域无关的，它能够描述任何信息，比如一个组织、商业活动等。可以用 RDF 来描述企业业务领域的知识，也可以用来描述软件系统的特征信息。RDF 的语法可被表示成为如上所示的 XML 序列化语法格式(RDF/XML 语法)，这样 RDF 就提供了一种更易被应用程序处理的陈述的方法，并且它具有强大的表述能力，而且 RDF 本身也具有一定的推理能力，例如，图能推出子图等。虽然 RDF 文件难以阅读和编辑，但我们不必直接去写 RDF 文件，可以通过各种用户友好的工具来编辑、修改 RDF 文件，这样既可以精确地描述信息，又可方便地与用户进行沟通。

### 3 OWL 语言与需求获取过程

#### 3.1 本体与本体描述语言 (RDFS/OWL)

RDF 允许使用自己的词汇来描述任何资源，但 RDF 没有假定任何特定的应用领域，也没有定义任何应用领域的语义。这些应该是本体所要完成的事情。

本体原是一个哲学的概念，指的是对客观存在的系统的解释和说明。随着人工智能的发展，本体被人工智能界给予了新的定义。采用 Struder (1998 年) 的定义：本体是共享概念模型的明确的形式化规范说明<sup>[3]</sup>。

概念模型 (conceptualization) 通过抽象出客观世界中一些现象 (Phenomenon) 的相关概念而得到的模型，其表示的含义独立于具体的环境状态。

明确 (explicit) 所使用的概念及使用这些概念的约束都有明确的定义。

形式化 (formal) 本体是计算机可读的。

共享 (share) Ontology 中体现的是共同认可的知识，反映的是相关领域中公认的概念集，它所针对的是团体而不是个体。

Ontology 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。关于 Ontology 研究的主题、Ontology 技术的状态、自动化及半自动化的 Ontology 构建、Ontology 的映射和 Ontology 的进化等请参阅文献[5,6]。

RDFS (RDF Schema) 是一个较简单本体语言。RDFS 提供了核心类(Core Class)、核心属性(Core Attribute)和核心限制(Core Constraint)等机制来定义资源的类(例如书是一个类，具体的书是一个实例)，资源可能的属性、资源和资源之间的继承等关系。在某个特定的知识系统中，例如图书分类系统，可用 RDFS 定义 RDF 中出现的资源、属性等，然后用 RDF 描述命题，在上述的过程中由于没有引入任何额外的算子和操作，因此和 RDF 具有同样的可判定性<sup>[3,4]</sup>。

2004 年 2 月 10 日，W3C(The World Wide Web Consortium) 宣布 OWL 和 RDF 为该组织的推荐标准。OWL (Web Ontology

Language) 是在 DAML(The DARPA Agent Markup Language) 的基础上发展起来的。相比 RDFS，OWL 具有更强大的描述能力。OWL 在 RDFS 的基础上添加了更多用于描述属性和类型的词汇，例如类型之间的不相交性(disjointness)、基数(cardinality)、等价性(equipollence)、属性特征(例如对称性 symmetry)、枚举类型(enumerated classes)以及类之间的布尔运算 (union、intersection and complement) 等。OWL 语言有 3 个子语言：OWL-LITE，OWL-DL 和 OWL-FULL。本文采用的是基于描述逻辑 (Description Logic) 的 OWL-DL 子语言，其有丰富的表达能力，并且具有可判定的推理能力<sup>[3]</sup>。

基于 OWL-DL 语言，给出了 ATM 取款机的简单 Ontology 示例 (从语法角度来说，一个 OWL 文档完全是一个 RDF 文档)：

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">
  <rdfs:comment>ATM 取款(OWL 片断) Ontology 示例作者: 李伟
  要点:主要描述了 ATM 取款的 RMB 和 Dollar 两种取款方式及简单的取款规则.
</rdfs:comment>
  <owl:Ontology rdf:about="">
    <owl:Class rdf:ID="查询">
      <rdfs:subClassOf>
        <owl:Class rdf:ID="ATM 取款机的功能">
          </rdfs:subClassOf>
        </owl:Class>
      <owl:Class rdf:ID="取款 (RMB)">
        <owl:disjointWith>
          <owl:Class rdf:ID="取款 (Dollar)">
            </owl:disjointWith>
          <rdfs:subClassOf>
            <owl:Class rdf:ID="取款">
              </rdfs:subClassOf>
            </owl:Class>
          <owl:Class rdf:ID="转帐">
            <rdfs:subClassOf rdf:resource="#ATM 取款机的功能">
              </owl:Class>
            <owl:Class rdf:about="#取款">
              <rdfs:subClassOf rdf:resource="#ATM 取款机的功能">
                </owl:Class>
            <owl:Class rdf:about="#取款 (Dollar)">
              <rdfs:subClassOf rdf:resource="#取款">
                <owl:disjointWith rdf:resource="#取款 (RMB)">
                  </owl:Class>
                <owl:Class rdf:ID="存款">
                  <rdfs:subClassOf rdf:resource="#ATM 取款机的功能">
                    </owl:Class>
                  <owl:DatatypeProperty rdf:ID="每天取款限额">
                    <rdfs:range
                      rdf:resource="http://www.w3.org/2001/XMLSchema#int">
                      <rdfs:domain rdf:resource="#取款">
                    </owl:DatatypeProperty>
                  <owl:DatatypeProperty rdf:ID="每次取款限额">
                    <rdfs:range
                      rdf:resource="http://www.w3.org/2001/XMLSchema#int">
```

```

<rdf:domain rdf:resource="#取款"/>
</owl:DatatypeProperty>
</rdf:RDF>

```

### 3.2 基于 OWL 的软件需求获取过程

图 2 显示了基于 OWL 的软件需求获取过程。

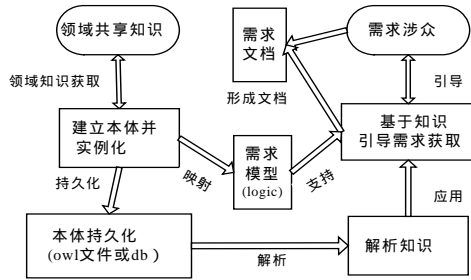


图 2 基于 OWL 的软件需求获取过程

(1)首先要建立相关领域知识的本体，并实例化本体得到相关领域的共享知识。领域共享知识最终以 RDF 文档的形式或数据库中表的记录的形式存储。在此过程中，我们可以基于本体建立具有高复用性的领域需求模型，详见文献[2]。

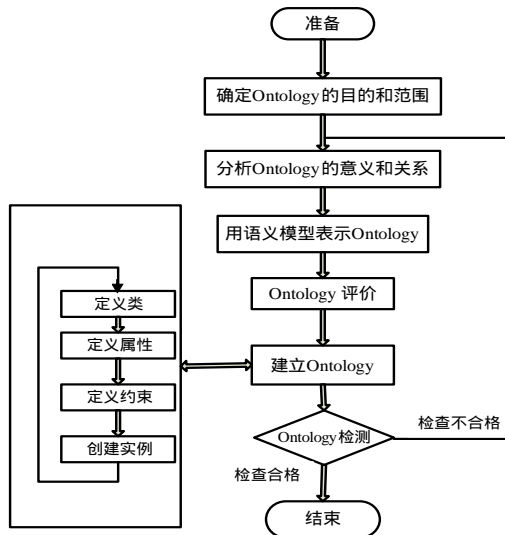


图 3 Ontology 的建立过程

基于图 3 所示的过程建立Ontology，这里Ontology的检测就是对照本体定义的 4 个方面的要求，并且检查所建立的Ontology能否满足要求为标准来确定是否合格。我们也可以通过Racer、Fact等DL（Description Logic）推理机来检测基于OWL-DL语言所建立的本体的一致性和正确性<sup>[3]</sup>。本体建立并实例化的过程就是获取共享领域知识的过程，这个过程需要领域专家的参与。基于OWL语言描述的本体最终可以以RDF文档的形式或以数据库表的记录的形式持久化储存。

(2)解析基于 RDF/XML 语法的 OWL 文档或从数据库中检索领域知识的信息。

HP 实验室的 Jena2 是一个专门针对基于 RDF，OWL 的语义网的第 2 代 Java API，它提供了以 RDF API 为核心的语义 Web 工具集。在这个工具集里，Jena 提供了对 RDFS 和 OWL 推理的多方面的支持，用这个工具可以较方便地实现对共享领域知识的解析和应用。

(3)基于该领域知识引导需求涉众获取软件需求，并最终形成软件需求文档。

可以采用基于问题的方式、基于领域知识体系模板的形式来引导需求涉众理清其领域业务知识的脉络，并给出正确的软件需求。对于开发者可充分复用基于本体建立起来的领域需求模型，充分理解需求涉众的软件需求，进而开发出真正满足用户需求的软件制品。

### 4 结束语

基于 OWL 的软件需求获取在一定程度上解决了需求捕获中的困难。这种基于领域共享知识的需求获取方式使得开发人员和需求涉众之间有了彼此交流的共同“语言”。通过需求引导的方式使得需求涉众对系统的功能性需求有了清晰的认识，从而更快、更系统、更准确地达成对需求的共识。相比之下，用户的需求文档也会减少二义性和不一致性。而领域需求模型具有高可复用性的特点，使得软件开发事半功倍。

领域Ontology的建立、映射、进化和维护等都是需要深入研究的课题<sup>[5,6]</sup>，而如何更好地引导需求用户获取软件需求也有许多工作要做。但这种基于OWL的软件需求获取方式还是有着良好的发展前景。

### 参考文献

- 1 Kendall S. The Unified Process Explained[M]. Addison Wesley, 2002.
- 2 金 芝. 基于本体的自动需求获取[J]. 计算机学报, 2000, 23(5): 486-492.
- 3 W3C. World Wide Web Consortium Issues RDF and OWL Recommendations[EB/OL]. <http://www.w3.org/2004/01/sws-pressrelease.html.en>, 2004-02-10.
- 4 何杭军, 龚丽霞, 陆汝占. 语义网知识表示的评价标准[J]. 计算机工程, 2002, 28(6): 108-110.
- 5 Ding Y, Foo S. Ontology Research and Development: Part 1——A Review of Ontology Generation[J]. Journal of Information Science, 2002, 28(2): 123-136.
- 6 Ding Y, Foo S. Ontology Research and Development: Part 2——A Review of Ontology Mapping and Evolving[J]. Journal of Information Science, 2002, 28(5): 383-396.

(上接第 51 页)

算法比传统 k-means 方法更适合于对空间对象的聚类分析。

需要指出的是，对于空间对象的聚类，随机选择初始聚类中心并不是一个合适的方法，进一步的研究应结合研究区域的空间格局，如考虑局部的“热点”或“冷点”区域对象，使初始聚类中心的确定与空间分布建立联系。

### 参考文献

- 1 Han J, Kamber M. Data Mining: Concepts and Techniques[M]. San Francisco: Academic Press, 2001.
- 2 Cressie A C. Statistics for Spatial Data[M]. New York: Wiley, 1991.

- 3 Fischer M, Getis A. Recent Developments in Spatial Analysis: Spatial Statistics, Behavioral Modeling and Computational Intelligence[M]. Berlin: Springer, 1997.
- 4 骆剑承, 周成虎, 梁 怡等. 多尺度空间单元区域划分方法[J]. 地理学报, 2002, 57(2): 167-173.
- 5 刘旭华, 王劲峰. 空间权重矩阵的生成方法与实验[J]. 地球信息科学, 2002, 4(2): 39-42.
- 6 李新运, 郑新奇, 闫弘文. 坐标与属性一体化的空间聚类方法研究[J]. 地理与地理信息科学, 2004, 20(2): 38-40.

