

一种本体编辑及可视化工具^①

张欢欢^{1,2}, 宋良图¹, 魏圆圆^{1,2}, 朱毅^{1,2}

¹(中国科学院 合肥智能机械研究所, 合肥 230031)

²(中国科学技术大学 自动化系, 合肥 230026)

摘要: 随着本体的广泛应用, 构建实用本体的重要性愈加突出。在中文领域, 支持中文的本体构建工具不够完善, 迫切需要高效、符合中文习惯的本体开发工具。利用 protégé-owl API 对本体进行底层操作, 利用 Prefuse 库实现可视化, 采用 MVC 框架开发了一款中文本体编辑及可视化工具。实验表明, 该软件能够有效地对中文本体进行编辑和可视化, 为中文本体的构建者提供了更加适合中文的本体构建平台。

关键词: 本体; protégé-owl API; 可视化; Prefuse

Tool for Editing and Visualization of OWL Ontologies

ZHANG Huan-Huan^{1,2}, SONG Liang-Tu¹, WEI Yuan-Yuan^{1,2}, ZHU Yi^{1,2}

¹(Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China)

²(Department of Automation, University of Science and Technology of China, Hefei 230026, China)

Abstract: With the extensive application of ontology in many fields, it is more and more important to build practical ontologies. However, available editing tool for Chinese ontologies is rare, so efficient ontology development tool which is designed for Chinese is needed. In this paper, we present a tool for editing and Visualization of Chinese ontologies on MVC mode, which reads and writes Chinese ontologies with the protégé-owl API, and implements ontology visualization with prefuse library. The experiments show that the design can edit and realize the visualization of chinese ontologies efficiently, which provides a more convenient platform for chinese ontology researchers.

Key words: ontology; protégé-owl API; visualization; prefuse

随着本体在知识工程、语义网 (Semantic Web)、系统建模等领域的广泛应用, 构建实用本体 (Ontology) 的重要性愈加突出。而本体的构建离不开本体开发工具的支持, 它使用户不需要了解本体语言的细节就可创建、解析、存储和重用本体。国外对本体开发工具的研究较早, 也产生了大量的本体开发工具。Michael Denny 在^[1]中指出, 截至 2004 年为止本体开发工具有 94 个, 从最早的 Ontolingua、OntoSaurus、WebOnto 到 protégé^[2]、oilEd、OntoEdit 等, 现在的本体开发工具数目还在不断的增加。然而, 这些本体开发工具只是对英文提供了很好的支持, 而对中文本体进行可视化、推理、解析等方面或多或少存在一些问题。例如, 当前较流行的本体开

发工具 protégé 在处理中文时存在各种编码问题, 同时无法提供对中文本体的可视化功能。而国内至今还没有一款完备的本体开发工具, 为了更好的支持中文本体的开发, 因此迫切需要高效、符合中文习惯的本体开发工具。

对于本体开发工具而言, 其所支持的本体描述语言的选择非常重要, 而 W3C 提出的 OWL^[3]语言具有非常强的本体语义描述功能。现有对 OWL 语言操作的接口有三个: Jena、OWL API、Protégé-OWL API。其中 Protégé-OWL API^[4]是本体开发工具 protégé 提供的专门针对 OWL 和 RDF(S)本体语言的开源 Java 类库。Protégé-OWL API 提供了加载和保存 OWL 文件、查询和操作 OWL 数据模型的接口, 以及执行基于描

① 基金项目:国家科技支撑计划(2008BAK49B05);中国科学院知识创新工程青年人才领域前沿项目资助

收稿时间:2011-07-21;收到修改稿时间:2011-08-29

述逻辑推理的接口。

本文设计了中文本体的编辑及可视化工具，它结合中文环境下的本体特征，采用 GBK 编码、MVC 设计模式，利用 protégé-OWL API 对本体进行读写操作，采用数据可视化工具 Prefuse 实现本体的可视化。该工具能更好的支持中文，完成对中文本体的编辑和可视化。

1 中文本体编辑及可视化工具的相关概念

1.1 OWL 本体模型

Studer 在文献[5]中提出本体 ontology 是共享概念模型的明确的形式化规范说明，包含概念模型、明确性、形式性和共享等含义。Perez 在文献[6]中提出按分类法来构建本体，本体包含 5 个基本的建模要素，分别是类（概念）、关系、函数、公理和实例。OWL (Web Ontology Language) 是 W3C 提出的一种本体描述语言，可以用来描述 Web 文档和应用中的内在类和关系^[7]。OWL 采用了面向对象的方式，通过类和属性来描述资源，并且通过公理来描述类和属性的特性及关系。OWL 本体包括了类、属性、实例的描述^[8]。

Protégé-OWL API 中的 OWL 本体模型是个数据结构类图。最上层的接口是 OWLModel，对 owl 文件的操作均是通过先创建该 owl 文件的 OWLModel 模型，而后通过该 OWLModel 进行创建类、属性和实例以及它们之间的关系。OWL 本体模型的组成元素众多，下图只是给出了一个比较简单的结构，如图 1 所示。主要组成有以下几部分：

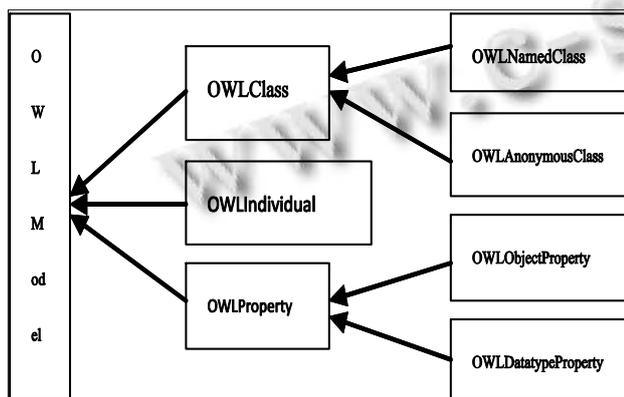


图 1 OWL 本体模型

1) OWLClass，即本体中的类或者概念，其中类主要是命名类（OWLNamedClass）和匿名类

（OWLAnonymousClass）。

2) OWLProperty，即本体中的属性，包含对象属性（OWLObjectProperty）、数据类型属性（OWLDatatypeProperty）。

3) OWLIndividual，即本体中的实例。

除此之外，OWL 本体模型中还包含了本体的命名空间（NameSpace），本体的名称等。

1.2 中文编码

OWL 文件使用 xml 编写，编码使用 xml 的 encoding 特性指定，默认编码是 UTF-8。UTF-8 是 Unicode 的变长字符编码，用 1 到 6 个字节编码 Unicode 字符。大部分的本体开发工具为节约存储空间采用基于 UTF-8 编码，例如 protégé 软件，这些本体开发工具能很好的处理采用 UTF-8 编码的 OWL 文件。但是当 OWL 文件中的元素是以中文命名元素名称时，如 <省级 rdf:about="#天津市"> </省级>，采用 UTF-8 编码的本体开发工具无法识别该中文 OWL 文件，会以乱码出现，使这些本体开发工具对中文的支持能力不尽如人意。

为正常显示以上出现编码问题的中文 OWL 文件，有两种解决方案：一是将 OWL 文件的 encoding 属性指定为 GBK 编码，protégé 等本体开发工具能正常显示中文本体，解决乱码问题；二是将 protégé 等本体开发工具的编码更改为 GBK，如在 protégé 安装目录下的 protege .lax 文件中加上一行:protege.file.encoding =GBK，这样也能正确显示中文本体。第一种解决方案，对每个中文本体均要修改编码，而第二中解决方案是从本体开发工具出发，根本解决编码问题。为了更完美地处理、显示中文本体，本文采用 GBK 编码。

1.3 可视化

本体可视化的工具主要分为两种，一种是基于 protégé 的可视化插件；另一种是独立于本体领域的通用可视化工具^[9]。

Protégé 官网^[10]上的可视化插件有 21 种，常用的有 OWLViz 和 OntoGraf。在 protégé4.1 版本中集成了 OWLViz、OntoGraf 两种可视化插件，以树的图示显示本体元素间的关系。OWLViz 显示了类层次结构间的父子关系；OntoGraf 展示了本体中包括类、属性、实例等元素的图示关系。这两者在处理中文本体时遇到了不同现象的编码问题：OWLViz 无法对中文本体图示进行布局，生成的图形堆叠在左上角；OntoGraf 无

法生成中文图示。究其本质原因,插件是基于 UTF-8 编码,java 语言对 I/O 流(字符流和字节流)处理存在的编码问题。

独立的可视化工具拥有更加强大的可视化功能和扩展接口,如美国加利福尼亚州伯克利大学计算机科学开发的 java 开源可视化工具包 Prefuse^[11]。Prefuse 是一个用户界面用来把有结构与无结构数据以具有交互性的可视化图形展示出来。它是一个比较强大的数据建模显示工程,支持多种数据的建模、显示,以及与用户的交互。它提供了表(Table)、图(Graph)和树(Tree)作为数据的基本结构,且图和树是表的另一种形式。在数据显示方面,它提供了大量的布局以及视觉编码技术,并且支持动画、动态查询等。

本文的可视化系统是基于 Prefuse 可视化工具包进行二次开发,实现了中文本体的可视化。运用树(Tree)数据结构,由节点(Node)和弧(Edge)组成,其中节点映射本体中类、实例元素,弧映射本体中的属性元素。

2 本体编辑及可视化工具的设计与实现

2.1 总体架构设计

系统框架采用 MVC (Model-View-Controller) 设计模式,分为三个部分:Model 层、View 层和 Controller 层,如图 2 所示。Model 层是 Protégé-OWL API 中描述的 OWL 本体模型,包括类(OWLClass)、属性(OWLProperty)和实例(OWLIndividual)等元素;View 层显示和操作底层模型的图形用户界面,包含类视图、对象属性视图、数据属性视图、实例视图和可视化等五个视图;Controller 层实现 Model 层和 View

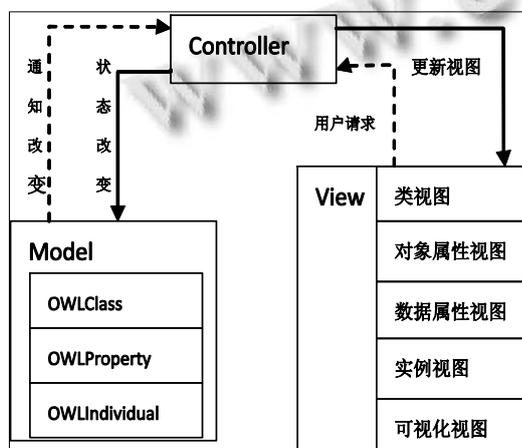


图 2 系统结构图

层之间沟通的桥梁,将 View 层的修改传递到 Model 层数据模型中,或把 Model 层的数据变化精确更新到 View 层显示中。该系统结合中文环境下的本体特征,采用 GBK 编码、MVC 设计模式,能更好的支持中文,实现中文本体的编辑和可视化。

2.3 本体编辑

本体的编辑是通过操作本体四个元素的图形化界面来实现,四个视图分别为:类视图、对象属性视图、数据属性视图和实例视图。类视图面板分两部分,左边部分以树的形式显示本体中包含的所有类,右边是类属性(包括 Annotation 属性、父类、互斥类)编辑区域,实现了类及其属性的“添加”和“删除”功能。对象属性视图、数据属性视图、实例视图的布局与类视图相似,实现了以树状结构显示本体元素及其 Annotation 属性、定义域类、值域类的“添加”和“删除”相关操作。具体实现以类的显示为例,读取文件到 OWLModel 中,先获取该 OWLModel 模型的根节点 Thing 类(Thing 类是所有类的父类);而后通过递归调用 showClassTree()函数在左侧树中显示出所有用户自己定义的类。以下代码为递归函数 showClassTree()的部分代码。

算法 showClassTree(treeRoot, extendedIterator)

输入:父节点 treeRoot, 包含该父节点的所有子节点的迭代器 extendedIterator

输出:树状结构

```
for (迭代器 extendedIterator) {
```

```
    获取该类节点 node, 添加到树中
```

```
    If node 有子节点
```

```
        do showClassTree(node, 子节点迭代器)
```

```
    return 树}
```

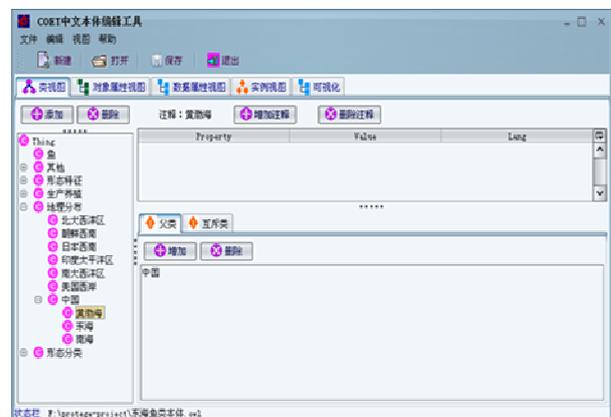


图 3 类面板界面

本体编辑的图形化界面如图 3 所示,以“东海鱼类本体”为例。用户只需要在四个面板之间进行切换即可顺利地对本体的编辑,步骤清晰地构建所需要的领域本体。

2.5 本体可视化

可视化视图,通过调用数据可视化工具 Prefuse 库中 Tree 型数据结构显示出本体数据对象。Prefuse 的数据处理对象是 Protégé-OWL API 中所创建的 OWLModel 本体对象。根据图 4 所示的映射关系,将本体数据对象映射到 Prefuse 的数据对象,本体类和实例分别映射为对应的节点(Node),对象特性映射为对应的边(Edge)。

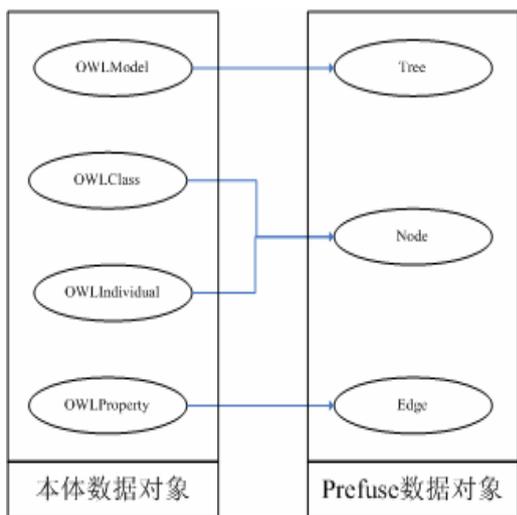


图 4 数据对象映射图

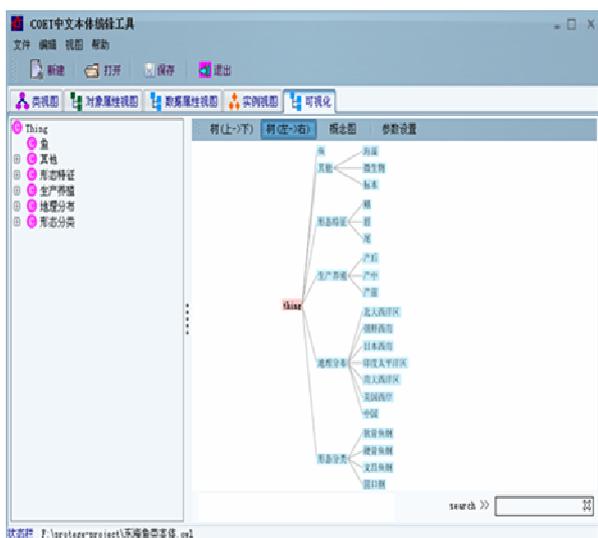


图 5 可视化界面

本体可视化实现了从上到下、从左到右的视图显示,并支持用户在画布上对本体的对象进行直接拖拽编辑,从左到右的结果如图 5 所示,以“东海鱼类本体”为例。用户可以从图示中清晰地看到概念的层次结构、属性关系、实例等内容,并可以直观、简便地进行正确性检查。

3 中文本体开发工具比较

中文本体开发工具的研究在国内的最有代表性的是苏晓路在文献[12]中提出的 LODE 本体开发系统客户端和王晓盈在文献[13]中提出的中文本体构建及可视化研究。在文献[12]中的 LODE 系统中,本体的浏览方式是将类、属性和实例以树形结构显示,并按类、当前类的相关属性及当前类的实例的顺序排列同一棵树中。这样的显示方式可使用户浏览本体时,完整地看到当前类的所有相关元素,提高了本体浏览的整体性,但此方法的在类和属性的节点较多时,受屏幕显示范围的局限,利用鼠标编辑会有所不便。在文献[13]中提出的中文本体构建及可视化工具是,汉化 protégé3.4 的核心功能,修改 OntoViz 插件的接口解决编码问题,实现了中文本体的构建和可视化。这样的方法可以缩短开发周期,但是 protégé 3.4 版本的基于 OWL 本体建模方式仅支持 OWL1.0,不支持 OWL2.0;对象属性和数据属性在一个视图中,不利于查看和编辑。本文中提出的本体编辑及可视化工具采用 GBK 编码、MVC 设计模式,利用 protégé-OWL API 对本体进行读写操作,利用数据可视化工具 Prefuse 实现本体的可视化。该工具实现了图形化编辑本体文件,用户只需要在四个面板之间进行切换即可顺利地对本体的编辑,步骤清晰地构建所需要的本体,比文献[12]中的开发工具更优越,当不同类存在相同的属性或实例时,不需要重复地构建属性和实例;并且,可以从可视化图示中清晰地看到概念的层次结构、属性关系、实例等内容,并可以直观、简便地进行正确性检查,比文献[13]中仅显示类之间关系的工具更完善。因此该工具能更好的支持中文,实现中文本体的编辑和可视化。

4 结语

文中对当前较流行的本体开发工具进行了研究分

析,针对当前中文本体开发工具不够完善的现状,通过 Protégé-OWL API、数据可视化 Prefuse 技术,采用 MVC 设计模式设计并实现了中文本体编辑及可视化工具。它提供友好的图形化界面,可以为用户提供方便快捷的中文本体开发平台。

由于一个完善本体的开发是一项长期而艰巨的工程,其构建工具的性能直接影响到本体的开发效率和质量。本体的推理技术还需要进一步的完善,利用 protégé-owl API 实现语法检查、一致性检查,以及查询等功能。随着系统功能的不断开发改进,该中文本体编辑及可视化工具将会成为一个完善的中文本体开发工具。

参考文献

- 1 Denny M. Ontology tools survey, revisited. Retrieved February, <http://www.xml.com/pub/a/2004/07/14/onto.html> 2004, 20:2006.
- 2 Noy N, Ferguson R, Musen M. The knowledge model of Protege-2000: Combining interoperability and flexibility. Knowledge Engineering and Knowledge Management Methods, Models, and Tools, 2000.69-82.
- 3 钱平,郑业鲁.农业本体论研究与应用.北京:中国农业科学技术出版社,2006.90-91.
- 4 Knublauch H. protégé-owl api programmer's guide. URL: <http://protege.stanford.edu/plugins/owl/api/guide.html>, Last Accessed Date,2005,15(06):2005.
- 5 Studer R, Benjamins VR, Fensel D. Knowledge engineering: principles and methods. Data & knowledge engineering, 1998,25(1-2):161-197.
- 6 Pérez AG, Benjamins VR. Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods, 1999. Citeseer, 1-15.
- 7 McGuinness DL, Harmelen VF. OWL web ontology language overview. W3C Recommendation, 2004,10:2004-2003.
- 8 Welty C, McGuinness DL. OWL web ontology language guide. W3C recommendation, W3C (February 2004) <http://www.w3.org/TR/2004/REC-owl-guide-20040210>.
- 9 董慧,王超.本体应用可视化研究.信息系统,2009,12(32):116-120.
- 10 Topic-Visualization. <http://protegewiki.stanford.edu/wiki/Visualization>.
- 11 Prefuse. URL: <http://prefuse.org/doc/manual>.
- 12 苏晓路,李景,孟学宪,等. LODE 本体开发系统客户端的主要功能与特点. 农业网络信息, 2009, 8: 63-65.
- 13 王晓盈,王晓璇,刘鹏.中文本体构建及可视化研究.计算机技术与发展,2010,2,20(2):121-124.
- 14 朱丽娟,王康元,张洁.基于 SVG 和 Java 的电力系统节点电压可视化.继电器,2006,34(5):60-64.

(上接第 155 页)

良好的推广前景。以西门子 S7-200 PLC CPU226 为中央处理单元,不仅增强了系统的稳定性及抗干扰性而且建立了良好的人机界面,可以及时地了解系统的运行情况,从而减少查找、排除故障的时间,提高了工作效率。

参考文献

- 1 黄静波,牟龙华.矿用组合开关主控单元的设计.工矿自动化,2007,2(1):74-77.
- 2 李萍.矿用隔爆兼本质安全型八组合开关的研制.现代电子技术,2008,9(17):185-187.
- 3 徐达,侯友夫,荣延平,刘良勇.新型矿用隔爆四回路智能组合开关的研究.中国矿业,2007,16(1):87-89.
- 4 陶学仪,尚药世.现代化矿井安全高效综采工作面供电技术.煤炭学报,2010,35(11):1930-1934.
- 5 郁青林,王东,汤东旭,李翠兰.基于 PLC 的矿用组合开关的设计.工矿自动化,2011,3(3):19-21.
- 6 赵洪瑞,孙凡,李仲强,王淼.基于 PLC 的矿用隔爆兼本质安全型智能真空组合开关控制系统.电气防爆,2009,4(4):40-43.