

# 基于 Jena 及其本体推理的研究

柴留祥, 何 丰

(北方民族大学 计算机科学与工程学院, 宁夏 银川 750021)

**摘 要:** Tim Berner-Lee 于 2000 年提出了语义 Web, 随后语义网技术作为普遍研究的热点。语义 Web 作为下一代万维网的发展方向, 它的实现能有效提高互联网使用效率, 为实现本体层 Web 信息的共享和交换, 人们引入了“本体”。本体作为语义网的基石, 在各个领域发挥着重要的作用。文中主要介绍本体开发以及推理方面的相关知识, 通过本体编辑器 protégé3.4 扩建动物本体, 并应用 RACER 推理机推理实现以及生成本体关系图。利用 Eclipse 开发软件结合 Jena 和 Java 插件实现对 OWL 文件的个体实例提取。通过实施为概念分类的检测提供了依据。

**关键词:** 本体; Racer; protégé; Jena

**中图分类号:** TP31

**文献标识码:** A

**文章编号:** 1673-629X(2011)11-0117-03

## Research on Jena-Based and Ontology Reasoning

CHAI Liu-xiang, HE Feng

(College of Computer Science and Engineering, Northern University for Nationalities, Yinchuan 750021, China)

**Abstract:** Tim Berner - Lee puts forward the semantic Web in 2000, semantic Web technology is a hot spot. The semantic Web as the next generation of the world wide Web development direction, and its realization can effectively improve efficiency. For the realization of ontology layer of Web information sharing and exchange, people introduced "ontology". As the cornerstone of the semantic Web, ontology plays an important role in various fields. It mainly introduces the ontology development and reasoning related knowledge, through the ontology editor protege3.4 expands animal ontology, and using RACER reasoning machine implements and creates ontology relationship chart. Using Eclipse development software combined Jena and Java plugin realize the individual instance extraction OWL files. Through the implementation of the concept for classification of detection provide the basis for classification of detection.

**Key words:** ontology; Racer; protégé; Jena

## 0 引言

随着语义网技术的发展, 本体作为语义网的基石, 也越来越多的受到人们的关注, 然而领域本体是用于描述特定领域知识的一种专门本体, 现在它的应用已经深入到各个领域。为此, 文中首先引入本体和 Jena 的相关知识, 通过本体编辑器构建本体, 最后给以实现。

## 1 相关知识

### (1) 本体。

本体的概念最早起源于希腊哲学。从哲学的角度来说, 本体是关于自然界存在之间如何相互组织的一门哲学。20 世纪 90 年代以来, 本体引起了计算机领域的广泛关注, 相关领域有: 知识工程、知识表示、语言

工程、数据库设计、信息集成、信息检索于抽取等<sup>[1]</sup>。随着 Ontology 的逐渐发展, “本体是明确的共享概念模型的形式化规范说明”, 作为本体的定义: 本体定义包含 4 层含义。

1) 概念模型 (conceptualization) 是指客观世界中一些现象的相关概念抽象出来所得的模型。

2) 明确 (explicit) 是指明确定义了所使用的概念及其概念的约束。

3) 形式化 (formal) 是指能理解和处理计算机信息。

4) 共享 (share) 是指本体是共同认可知识的体现和公认的概念集在相关领域的反映。

本体最终是对相关领域内知识的捕获, 并共同理解的领域知识, 确定共同认可的概念在领域内, 对不同层次模型中概念间的相互关系给出明确定义。

概念的集合作为本体的本质, 把现实世界中某个领域的一组概念及概念关系抽象成集合中的概念, 其定理就是对概念中所具有的属性进行限制<sup>[2]</sup>。本体把知识表述语言作为它的主要贡献并提供了一种更为

收稿日期: 2011-04-18; 修回日期: 2011-07-25

基金名称: 宁夏自然科学基金(71061001, G011201)

作者简介: 柴留祥(1981-), 男, 硕士研究生, 研究方向为语义网技术; 何 丰, 博士, 从事语义 Web 和数据挖掘的研究。

详尽的规范,语义才能够被更全面地表达<sup>[3]</sup>。

(2) 领域本体。

领域本体<sup>[4]</sup>就是描述一种学科概念,包括概念间的关系、概念的属性、属性和关系的约束。由于知识具有显著的领域特性,所以领域本体能够更合理而有效的进行知识的表示。

领域本体<sup>[5]</sup>就是表示某一个特定领域范围内的特定知识。这里的“领域”是根据本体构建者的需求来确立的,它可以是一个科学领域,可以是某几个领域的一种结合,也可以是一个领域中的一个小范围。

领域本体的模型及建模原语。

用  $C$  表示类集,  $I$  表示类集中的实例集,  $R$  表示关系集合,  $is\_a$  表示类间的继承关系,公理集合用  $A_x$  表示,来约束类间的属性和关系,所以本体是一个  $O = (C, I, R, A_x)$  四元组。

若把上述关系集合  $R$  划分成关系和函数,那么就把四元组变成五元组<sup>[6]</sup>  $O = (C, R, F, A, I)$ , 作为基本的建模原语,即加入函数  $F$ , 公理  $A$ 。

(3) 本体编辑器(protégé3.4)。

Protégé3.4<sup>[7]</sup>是一个开放源码的本体编辑器,采用 JAVA 语言编程实现的,由斯坦福大学的 Startford Medical Informatics 研发的,便于操作和使用。并以树形层次结构显示,而且编辑类、子类、属性等可以通过点击项目来添加,领域模型还可以在概念层次上设计。

(4) Jena 及 Jena 推理规则的构造语法。

Jena 起源可以追溯到 SIRPAC API 的工作,由 HP 公司的 Brian Me Bride 开发的 Java 工具包。应用程序可以通过 Jena 进行 RDF 模型的解析、创建和查询工作。模型中的 RDF statement 可以通过 Jena 定义的接口来访问和处理,接口有:RDF Node、常量、属性等,由于它的可靠性和持续稳定性,被首选作为应用程序开发的工具。近几年采用 Jena API 来处理 RDF 作为成功开发的语义网工程在 WWW 会议上被提出。

Jena 推理规则的构造语法<sup>[7]</sup>:

用户一方面可以利用上面的结构实现本体中的推理,另一方面,如果用户对 Java 很熟悉,则可以自己开发推理规则。规则的语法格式如下:

```

Rule      := bare-rule
           or [ bare-rule ]
           or [ ruleName; bare-rule ]
bare-rule := term, ...term->hterm, ...hterm //向前推理
           or term, ...term<-term, ...term //向后推理
           hterm := term
           or [ bare-rule ]
term := ( node, node, node ) //三元组模式
       or ( node, node, functor ) //扩展三元组模式
       or buihin( node, ...node ) //调用处理元语

```

functor := functorName( node, ... node ) //结构化的文字表述

```

node := uri-ref //例如 http://foo.com/eg
       or prefix:localname //例如 rdf:type
       or varname //变量名
       or 'a literal' //字符串
       or number //例如 42 或 25.5

```

用户根据上述语法格式写出针对特定问题的规则,就可以更加方便、快捷地查询各种信息了。

(5) RACER。

RACER<sup>[8]</sup>是一个 OWL 推理机(Reasoner),RACER 推理能完全将本体中存在的隐性知识呈现出来。对本体中隐藏的知识,可以通过搜索 TBox 和 ABox 将本体论中的隐性知识显示出来。通过 RACER 推理可以产生新的知识,但 RACER 只能推论采用描述逻辑架构的本体,也就是 OWL DL 语法,若使用 OWL Full 语法则超出了 RACER 的推理范围。

2 本体的构建方法

目前比较流行的领域本体的构建方法<sup>[9]</sup>主要有以下几种:骨架法、企业模型法、七步法等<sup>[10]</sup>。文中构建方法主要采用斯坦福大学开发的七步法构建本体。

建立概念分类层次的可行性方法有:①自顶向下法:开始于最顶层概念,同时逐步添加子类概念并细化。②自底向上法:开始于最小的底层概念,将在概念下对细化的类进行综合的组织。③综合法:首先定义概念,然后对所定义的概念进行归纳和细化。可从顶层概念开始,并与一些中级概念进行关联。因为领域概念中“中级概念”更具有代表性,所以对于开发者而言综合法最便捷。自底向上方法更加适合收集更多更广泛的实例。任何方法,都要从类定义开始。

根据上述特点,本文扩展了原有动物本体,具体概念之间的关联进行 UML 建模。进而更加直观说明类集之间的关系如图 1 所示。

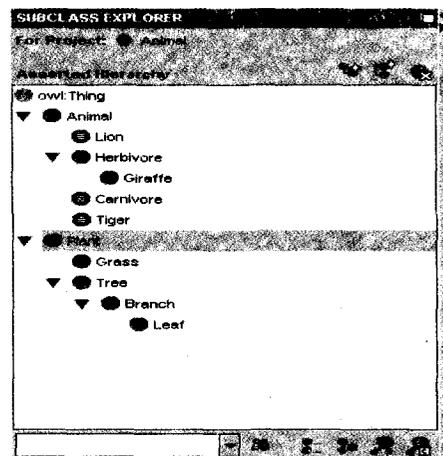


图 1 扩展本体类图

选择本体编辑器 (protégé) 自身的插件,此时会出现 OWL 图标,将自动生成关系图如图 2 所示。

### 3 本体的进化

本体构建成功后,先通过 RACER 将本体作整理和分类,RACER 是一个 OWL 推理机,通过 RACER 推理可以将本体编辑器 protégé 构建的本体以树状的形式呈现出来,使概念之间的包含关系更加明确如图 3 所示。

经过 RACER 推理后的关系图如图 4 所示。RACER Service 可以使用在 Linux 或 Windows,并可和多个本体系统配合,图 5 所示是 RACER 监听<sup>[11]</sup> port 时的界面,RACER 运行时使用 8080 端口,本研究使用 protégé 和 RACER 结合使用,由此推出本体中隐藏的知识。

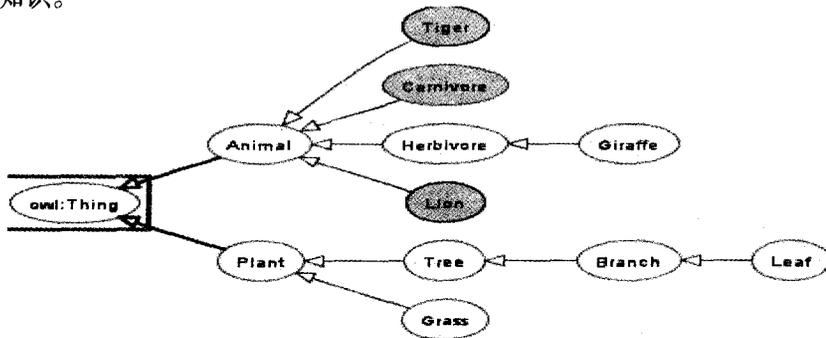


图 2 关系图

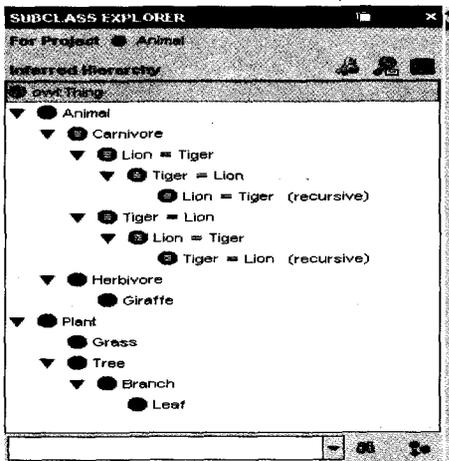


图 3 推理后的本体类图

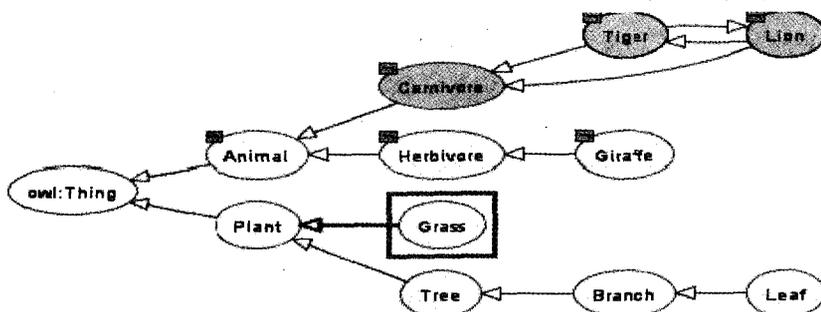


图 4 推理后的关系图

图 4 显示 Lion 和 Tiger 同属于肉食动物,由于肉食动物可以吃素食动物和肉食动物,经过 RACER 推理后 Lion 和 Tiger 之间可以相互吃与被吃。

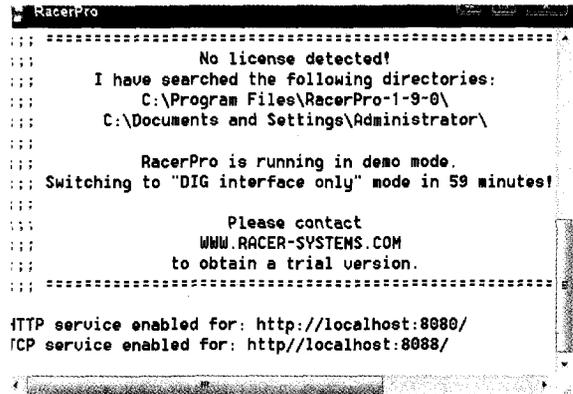


图 5 RACER 监听界面

根据 Jena<sup>[12]</sup>语法的定义规则,在下面程序中,自定义了 Jena 规则来测试概念之间的传递包含关系。用 a 代表树 (Tree),b 代表树枝 (Branch),c 代表树叶 (Leaf)。p 代表之间的包含关系。可以表示如下的三元组 (其中 demo: 代表名称空间“http://jena.hpl.hp.com/demo#”):

```
demo:a demo:p demo:b
demo:b demo:p demo:c
```

定义一个规则来支持关系 demo:p 的传递属性:

```
[rule1:(? a demo:p ? b)(? B demo:p ? c)
->(? a demo:p ? c)]
```

然后连接 Jena 自带的推理机后,可以推理概念 c 即属于概念 b 也属于概念 a。

```
for (StmtIterator i=inf. listStatements ( a, p, c ); i. hasNext ();
i. hasNext ()) {
Statement s=i. nextStatement ();
System. out. println ( s );
For (Iterator id=inf. getDerivation ( s ); id. hasNext ()) {
Derivation deriv = ( Derivation ) id. next ();
deriv. printTrace ( out, true );
}
}
```

其结果可以根据规则事实集并依此类推可以推出

三元组 (demo:a demo:p demo:c)。

本体推理主要包括:计算本体类层次关系,检查概念一致性,检查实例个体是否是合法的本体实例等。

### 4 结束语

文中所采用的本体是基于原

(下转第 123 页)

(b)、(e)分别为经过 Canny 算法处理后的结果图,Canny 算法的两个阈值分别为 50 和 150。

(c)、(f)分别是应用文中轮廓提取算法后的结果图,取固定参数  $\beta_1 = 20, \beta_2 = 52, T_1 = 150, T_2 = 50$ 。

由图可以看出 Canny 算法虽然能提取出大部分的边缘,但是无法准确地提取出焦炭光学组织的轮廓,同时不能得到较好的轮廓曲线。而文中方法能够提取到精确地焦炭组织的轮廓。

## 5 结束语

文中提出的基于边界加权的轮廓提取算法能够很好地对图像进行聚类平滑,很好地保留图像的边缘信息,提取出焦炭多类显微组织图像的轮廓,且当原始图像中存在噪声干扰时,能够消除噪声,轮廓提取效果优于 Canny 算子等方法,且抗噪能力更强,不会提取出过多的杂点。利用文中方法提取出焦炭显微图像的轮廓后,为焦炭光学组织的进一步特征提取和分类奠定了一定的基础。

### 参考文献:

- [1] Lang B Q, Hu L T, Zheng Z C, et al. Carbonization behavior of coal-tar pitch modified with divinylbenzene and optical texture of resultant semi-cokes[J]. Journal of analytical and applied pyrolysis, 2004, 71(2): 77-93.
- [2] Zhang Z L, Zhao M M, Wang P Z, et al. Study on Factors of Influencing on Coke Thermal Property [J]. Iron & Steel, 2009, 44(10): 10-13.

- [3] 刘小除. 图像处理技术在焦炭光学组织自动分析软件中的应用[D]. 上海:复旦大学, 2005.
- [4] 夏京城. 图像处理与模式识别技术在焦炭颗粒识别中的应用研究[D]. 上海:复旦大学, 2005.
- [5] 张代林. 利用图像分析法测定焦炭气孔结构的研究[J]. 燃料与化工, 2003, 34(4): 175-178.
- [6] John C. A Computational Approach to Edge Detection [J]. IEEE-PAMI, 1986, 8(6): 679-698.
- [7] Cheng Y Z. MeanShift, Mode Seeking, and Clustering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790-799.
- [8] Meer P. Edge Detection with Embedded Confidence [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(12): 1351-1365.
- [9] Comaniciu D. MeanShift: A Robust Approach Toward Feature Space Analysis [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.
- [10] 李乡儒, 吴福明. 均值偏移算法的收敛性 [J]. 软件学报, 2005, 16(3): 365-374.
- [11] Wu Kuo-Lung, Yang Miin-Shen. MeanShift-based Clustering [J]. Pattern Recognition, 2007, 40: 3035-3052.
- [12] Chen Jie, Wang Zhen hua, Dou Li hua. Scale Adaptive Canny Edge Detection Method [J]. Opto - Electronic Engineering, 2008, 35(2): 79-84.
- [13] 张志刚, 周明全. 一种轮廓曲线的多边形近似算法 [J]. 计算机应用, 2006, 26(3): 576-579.
- [14] 谢凤英. Visual C++ 数字图像处理 [M]. 北京: 电子工业出版社, 2008.

(上接第 119 页)

有动物本体的扩展本体,以实现在 RACER 基础上的推理和 Jena 自定义规则的传递包含关系。特别是在领域本体的构建上,可以给加入适当的限制条件、包含关系等可使本体的构建更加完善和具体,使其在后期工作中推理和查找中更加全面。

下一步的工作对推理规则加入限制条件实现其推理、可将建立好的本体存入知识库,可以使用 nRQL 进行查询等。

### 参考文献:

- [1] Guarino N. Formal Ontology in Information System [C] // Proceedings of FOIS98. Trento, Italy: IOS Press, 1998: 3-15.
- [2] 武成岗, 焦文品. 基于本体论和多主体的信息检索服务器 [J]. 计算机研究与发展, 2001, 38(6): 641-647.
- [3] 曹 锐, 陈 刚. 基于本体的网络化资源检索 [J]. 计算机工程, 2004, 30(3): 143-146.
- [4] 张志刚. 领域本体构建方法的研究与应用 [D]. 大连: 大连

海事大学, 2008.

- [5] protégé [EB/OL]. [2009-03-01]. <http://protégé.stanford.edu>.
- [6] 邓志鸿, 唐世渭, 张 铭, 等. Ontology 研究综述 [J]. 北京大学学报(自然科学版), 2002, 38(5): 730-738.
- [7] Jena 2-A Semantic Web Framework for Java [EB/OL]. 2006-05-04. <http://jena.sourceforge.net/index.html>.
- [8] Haarslev V, Moller R. RACER system description [C] // Proc of the Int Joint Conf on Automated Reasoning (IJCAR 2001). [s. l.]: Springer, 2001: 701-705.
- [9] 袁 媛. 领域本体建设的方法论和工具研究 [D]. 北京: 中国人民大学, 2004.
- [10] 关庆珍, 周竹荣. 基于 ontology 的用户模型研究 [J]. 计算机应用, 2007(10): 56-60.
- [11] RacerPro [EB/OL]. [2009-03-01]. <http://www.racer-system.com/>.
- [12] 肖 敏. 领域本体的构建方法研究 [J]. 情报杂志, 2006(2): 70-72.