

面向语义网的中文本体应用研究

凌绍东,霍 林,王 超

(广西大学 计算机与电子信息学院,广西 南宁 530004)

摘 要:自从 Protégé 出现以来,它成为唯一一款支持中文本体构建的工具,而且能够进行可视化操作。目前在部分领域已经构建了中文本体,但是对所构建本体在语义网中的应用缺乏研究。因为本体的应用主要依赖于推理工具,而国内尚无一款专业的基于中文本体的推理机或者推理工具,这阻碍了基于中文本体信息检索的发展。针对中文本体的应用,深入研究了本体在信息检索领域应用的现有技术、方法、工具。文中通过使用 Jena 开发包对中文本体的应用进行实验,实验表明,利用现有的推理工具等可以实现中文本体在语义网中的应用。

关键词:中文本体;可视化;本体应用;推理机

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2014)02-0194-04

doi:10.3969/j.issn.1673-629X.2014.02.048

Research on Chinese Ontology Application Oriented Semantic Network

LING Shao-dong, HUO Lin, WANG Chao

(College of Computer and Electronic Information, Guangxi University, Nanning 530004, China)

Abstract: Protégé has become the only tool which can build Chinese ontology and support visual operation since it appeared. At present, there already have built some Chinese ontologies in some areas, but researches on ontology application in semantic network are very rare. Development based on Chinese ontology's information retrieval is suffocate, because ontology application depends on inference machine, but there has no one technical inference machine or tool based on Chinese ontology. In allusion to the applications of Chinese ontology, research existing methods, technologies and tools that apply to information retrieval by ontology. Try experiments with Chinese ontology's applications by using Jena software developer's kit, the results explain that using existing inference machine can achieve Chinese ontology's applications in semantic network.

Key words: Chinese ontology; visualization; ontology application; inference machine

0 引言

随着语义网和 Web 本体语言的发展,基于本体的信息检索技术逐步发展起来。2004 年 2 月 10 日, OWL (Web Ontology Language) 正式被 W3C 定为推荐的标准 Web 本体语言。

在信息检索领域,当本体构建完成之后,就可以在实际应用中使用它们进行语义检索,本体的出现为真正实现语义检索提供了可能。本体应用是面向语义网的信息检索中很重要的一部分,通过查阅和分析相关文献资料,得出本体在基于语义网的信息检索领域的应用主要包括本体的持久化存储、本体的扩展、本体的查询和推理等几个方面。所谓本体的持久化存储,就

是为了便于对本体进行共享和互斥操作而存入数据库,而后在数据库中完成对本体的一系列操作;本体构建之后,在实际使用中需要对本体进行扩充和完善,添加应用所需的实例、类、属性和关系等,即本体扩展,这是一个周期性、反复性的过程;在检索时,使用本体对检索词进行查询和推理,这是一个语义化的过程,把检索词转化为语义词,最后使用语义词完成检索请求。

经研究发现,本体的持久化存储、扩展、查询和推理都依赖于推理工具。由于 Jena^[1] 是一个基于 Java 语言的开发包,可以挂接各种推理机或者推理引擎,可以在 Java 开发平台中调用 Jena 开发包,更重要的是对用 Protégé^[2] 构建的 OWL^[3] 本体具有很好的查询和推理性能。因此,文中使用 Jena 开发包对中文本体的应

收稿日期:2013-05-01

修回日期:2013-08-04

网络出版时间:2013-11-29

基金项目:国家自然科学基金资助项目(61262072)

作者简介:凌绍东(1986-),男,硕士研究生,研究方向为计算机网络与并行分布式计算;霍 林,教授,研究方向为并行分布式计算、信息安全。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20131129.0857.012.html>

用进行研究,并解决过程中出现的问题。

1 推理机

推理机^[4]在本体构建中的功能作用主要分为三个方面:初始本体构建完成之后对本体进行分类和一致性检查;随着领域知识的扩展和应用需求的增加,需要使用推理机进行本体的扩展(本体的扩充和完善);在检索时,对用户的检索词除了进行本体概念的匹配之外,还对其进行推理,找出隐含概念,即查询扩展。

一般来说,在 Protégé 里构建了本体,就会想在应用程序里使用它,这就需要一些开发接口。用程序操作本体是很必要的,因为要自动生成本体,靠人手通过 Protégé 创建本体的所有概念和实例是不现实的,一般都是通过本体构建工具构建本体的一个基本架构,完成当前应用所需的概念和实例的本体化,以后根据实际应用需求来扩展本体。用本体构建工具完成本体构建之后,生成本体语言文件如 RDF/RDFS、OWL 或 DAML 等^[5-6],在实际应用中使用针对本体的推理机(Jess、Racer、Jena、FaCT、Pellet 等)来对本体文件进行查询和推理等操作。

Jena 是面向语义 Web 的应用开发包,它是 HP 公司语义网研究小组开发的一套 Java 工具包,用来支持人们进行语义网的相关研究和应用开发。它是基于 Java 语言的,包含的内容比较全面,推理机只是其中一部分,还支持对本体的解析。Jena 提供的推理机也和 Racer、FaCT、Pellet^[7]等一样,是针对本体的推理机,但 Jena 本身并不是“推理机设计专家”,它自身包含的推理机基本上就是一种 CLISP 配合本体领域产生式规则的前向推理系统。因此,它的运行效率不是很高。好在现在有 DIG 接口,DIG 有点像数据库中的 ODBC,允许前端挂接到后台不同的推理引擎上。这样,在 Jena 中,也可以使用 Racer、FaCT、Pellet 这样更“专业”些的推理机。Jena 中的推理引擎和 Racer 都是针对具体本体语言的推理机,针对性强,效率高。当用户使用 OWL 表示本体时,Protégé 会用 Jena 或 Racer 等推理。

同时,利用 Jena 开发包也可以完成本体的存取等操作。文中使用 Jana2. 10. 0 开发包和其自带的推理引擎对本体的持久化存储、扩展、查询和推理进行研究和实验。

2 本体持久化存储

为了对本体的应用进行研究,首先使用 Protégé 4. 1 构建一个领域本体 Animal. owl,层次结构如图 1。其中,圆点代表类,菱形代表实例,猫科、犬科和马科是动物科属的子类,貉、狼、犬和狐是犬科

的子类。

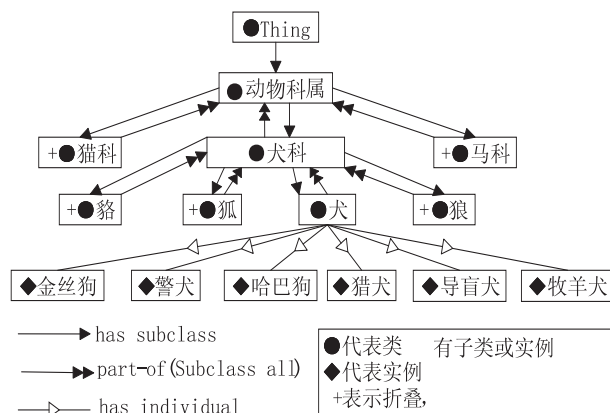


图1 本体层次结构

在实际应用中,往往需要把本体持久化入库^[8-9],常用的数据库有 MySQL、DB2 和 ORACLE 等。选用哪种数据库根据实际应用而定,文中使用数据库 MySQL5. 1 持久化存储本体文件 Animal. owl,在这里要特别注意,需要将字节流文件转换为 UTF-8 编码格式,不然就很可能使中文显示为乱码。

首先创建数据库 animal,然后把 owl 读入数据库 animal,可以通过数据库对本体进行读写等操作,部分代码如下:

```

IDBConnection connection = new DBConnection ( strURL,
strUser, strPassword, strDB );

Class.forName ( " com. mysql. jdbc. Driver " );

//从此处开始读入一个 OWL 文件并且存储到数据库中,使用
数据库连接参数创建一个模型制造器

ModelMaker maker = ModelFactory. createModelRDBMaker
( connection );

//创建一个默认模型,命名为 animal,这里首先要创建一个
数据库命名为 animal

Model defModel = maker. createModel ( " animal " );

InputStreamReader in = null;
File file = new File ( " E: \Animal. owl " );

//读取本体文件,为防止无法识别中文而出现乱码,需要把
文件流格式转换为 UTF-8

in = new InputStreamReader ( new FileInputStream ( file ), "
UTF-8 " );

//数据入库

defModel. read ( in, " http:// www. owl-ontologies. com /An-
imal " );

```

执行程序,这时就可以看到在数据库 animal 中就出现了 7 个表,如图 2 所示。

而 owl 文件的主要数据就保存在 jena_g1t1_stmt 这张表中,本体存储了之后就可以对本体进行操作了,例如可以读取本体中的类,部分代码如下:

```

IDBConnection con = new DBConnection ( strURL, strUser,
strPassword, strDB );

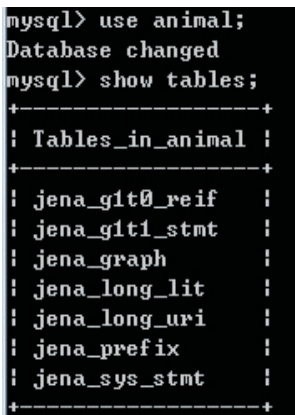
OntModel model;

```

```

ModelMaker maker = ModelFactory. createModelRDBMaker
(con);
//从数据库 animal 中读取本体
Model base = maker. openModel(" animal" );
OntModelSpec spec = new OntModelSpec ( OntModelSpec.
OWL_MEM );
spec. setImportModelMaker( maker );
model = ModelFactory. createOntologyModel( spec, base );
//读取本体中的类
for ( Iterator i = model. listClasses(); i. hasNext(); ) {
OntClass c = ( OntClass ) i. next();
System. out. println( " Class:" +c. getLocalName());

```



```

mysql> use animal;
Database changed
mysql> show tables;
+-----+
| Tables_in_animal |
+-----+
| jena_glt0_reif    |
| jena_glt1_stmt    |
| jena_graph        |
| jena_long_lit     |
| jena_long_uri     |
| jena_prefix       |
| jena_sys_stmt     |
+-----+

```

图 2 数据库存储表

得到部分结果如下:

```

Class : 斑马 Class : 犬 Class : 狐 Class : 狮子
Class: 狼 Class : 猫 Class : 老虎...

```

3 本体扩展

本体扩展^[10-11]就是在本体的扩充和完善,是一个反复叠加的、周期性的过程。随着领域知识的日渐丰富,当前本体不能再满足实际应用需要,就需要对本体进行扩展。对于本体的构建,不可能一次性完成,因为要自动生成本体,靠人手通过 Protégé 创建本体的所有概念和实例是不现实的,一般都是通过本体构建工具构建本体的一个基本架构,完成当前应用所需的概念和实例的本体化,随着领域知识应用需求的增加,使用推理机不断地扩充和完善本体。

Jena 推理机的 Jena API 定义了本体中与概念(也就是类)和实例的相关操作,对概念之间的各种关系都有相应的定义方法,典型的有添加子类、添加约束、创建互斥概念、迭代返回某种类型的概念以及相关的逻辑判断等。

这一节使用推理机 Jena2. 10. 0 版本来对本体进行一系列的操作,它是基于 Java 语言开发的,因此,可以使用 Eclipse3. 2 和 Jena 的 API 接口来完成对本体扩展的目的。首先需要把 Jena 的所有 Jar 包都引入工程

目录下,然后就可以通过 Jena API 往已建好的本体中添加类、创建实例、创建属性、设置属性等,完成扩展操作之后保存为 OWL 文件。

接下来,在本体中添加一个类“猴科”,它是“动物科属”类的子类,然后添加“猴科”的子类“猕猴”和“疣猴”,接着创建“猕猴”和“疣猴”的实例对象,最后设置类关系和属性关系等,部分代码如下:

```

//把本体文件读入一个 OntoModel
OntoModel ontModel = ( OntoModel ) ModelFactory. create-
FileModelMaker( "E:\Animal. owl" );
String ns = new String( " http://www. owl-ontologies. com/
animal#" );
//读取根类
OntClass animal = ontModel. getOntClass ( ns + " 动物科
属" );
//添加子类 and 实例等
OntClass cercopithecidae = ontModel. createClass( ns + " 猴
科" );
OntClass macaque = ontModel. createClass( ns + " 猕猴" );
OntClass colobus = ontModel. createClass( ns + " 疣猴" );
cercopithecidae. addSubClass( macaque );
cercopithecidae. addSubClass( colobus );
Individual chlorocebus = macaque. createIndividual( ns + " 绿
猴" );
.....
//存入磁盘
String savePath = "E:\Animal. owl" ;
DataOutputStream out= null;
try {
out = new DataOutputStream ( new BufferedOutputStream
(new FileOutputStream( savePath)));
} catch ( FileNotFoundException e ) {
e. printStackTrace();
};
ontModel. write( out );

```

然后使用本体构建工具 Protégé3. 3 打开该 OWL 文件,就可看到扩展的部分类、实例、属性及关系等。

4 本体查询与推理

在本体查询中,主要有三大查询语言^[12-13]: RDQL、SPARQL 和 OWL-QL,前两种是基于 RDF 的,第三种是基于 OWL 的,OWL 作为 RDFS 词汇上的扩充,完全可以由 RDF 查询语言对 OWL 本体进行查询,不过必须要有支持 OWL 词汇集的推理机的支持。SPARQL 是通过 RDF 图模式匹配实现查询的,它的语法形式和 SQL 比较相似。关于查询语言,可以查阅相关文档资料。

本体推理必须依据一定的推理规则来进行,每种推理机都定义了一套推理规则,用来完成该推理机的

基本推理功能,这些规则可以说是通用的,主要是针对本体的特点制定的,主要用来检测本体中类、实例、属性、关系等的可满足性。

以 Jena 为例,它只支持基于规则的简单推理,其推理机制支持将推理器(inference reasoner)导入 Jena,创建模型时将推理器与模型关联以实现推理。当用推理机进行推理时,必须依据一定的推理规则^[14],推理机都有一个规则集定义其行为,主要都是针对本体特点定义的一些规则,用于检测本体中概念的可满足性,以及不同类、实例之间的关系等。在推理机 Jena(…\jena-2.10.0-source\jena-core\etc 目录下)的源文件中就定义了很多规则,如:

```

rdfs: (? a rdfs:subClassOf ? b), (? b rdfs:subClassOf ?
c) ->( ? a rdfs:subClassOf ? c);

rdfs: (? x rdfs:subClassOf ? y), (? a rdf:type ? x), (?
b rdf:type ? y) ->( ? a rdfs:subClassOf ? y);

thing: (? x rdf:type ? C), (? C rdf:type daml:Class) ->
(? x rdf:type daml:Thing);

conflict: (? X owl:sameIndividualAs ? Y), (? X owl:dif-
ferentFrom ? Y) -> contradiction( 'same/different', ? X, ? Y);

```

这些规则是基于 OWL 的查询语言 OWL-QL 的。它们是推理机所具有的通用规则,但是,对于具体的应用领域来说,这些规则并不完整,只能完成一些基本关系的推理;对于本体中的一些隐含关系以及两个或者两个以上本体的推理,则显得无能为力。所以,根据实际应用需求,还要定义自己的推理规则,以满足本体推理的需求,如:

```

rdfs: (? a rdfs:subClassOf ? c), (? b rdfs:subClassOf ?
c) ->( ? a rdfs:siblingOf ? b);

rdfs: (? a rdfs:instanceOf ? c), (? b rdfs:instanceOf ? c) -
>( ? a rdfs:siblingOf ? b);

```

文中,使用 SPARQL 来查询本体,获得本体中警犬的兄弟名称,先是创建查询语句,再制定推理规则,最后调用推理机进行推理和查询,得到查询结果为:

```

http://www.owl-ontologies.com/Animal#哈巴狗
http://www.owl-ontologies.com/Animal#导盲犬
http://www.owl-ontologies.com/Animal#牧羊犬
http://www.owl-ontologies.com/Animal#猎狗
http://www.owl-ontologies.com/Animal#金丝狗

部分代码如下:

//SPARQL 查询语句

String queryString =

" PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syn-
tax-ns#>\n" +

" PREFIX : <http:// www.owl-ontologies.com /Animal#>\n" +

" SELECT ? name WHERE { " +

"? 犬 :siblingOf ? 警犬." +

```

```

"}";

//推理规则

String rule = " rdfs: (? a rdfs:subClassOf ? c), (? b rdfs:
subClassOf ? c) ->( ? a rdfs:siblingOf ? b)"

+" rdfs: (? a rdfs:instanceOf ? c), (? b rdfs:instanceOf ?
c) ->( ? a rdfs:siblingOf ? b)";

//在 Jena API 中找到 JenaReasonerFactory

Reasoner reasoner = RDFSRuleReasonerFactory. theInstance
(). create( null );

//获取标准的 OWL_DL 语言规则并加载 Jena 推理机

OntModelSpec ontModelSpec = OntModelSpec. OWL_DL_
MEM;

ontModelSpec. setReasoner( reasoner );

OntoModel ontModel = ( OntoModel ) ModelFactory. create-
FileModelMaker( " Animal. owl" );

//创建查询集

InfModel inf = ModelFactory. createInfModel( reasoner , ont-
Model );

//添加查询语句

Query q = QueryFactory. create( queryString );

//执行查询并获取结果集

QueryExecution qexec = QueryExecutionFactory. create ( q,
dataset );

ResultSet resultSet = qexec. execSelect();

while( resultSet. hasNext() ) {

QuerySolution row = ( QuerySolution ) resultSet. next();

//输出查询结果

System. out. println( row );

```

5 结束语

文中首先对本体推理工具做了比较分析,重点对 Jena 开发包进行了研究,然后使用 Protégé3.3 构建一个中文本体,接着使用 MySQL 数据库完成对本体文件的存取操作,随之对构建本体进行扩展,往本体中添加类、实例和属性等,最后通过制定一些推理规则对本体进行查询和推理操作。在实验过程中,解决了所出现的一些问题,并对一些关键问题进行了重点介绍和分析。

实验证明,使用现有的推理工具和数据库等可实现中文本体在信息检索领域的主要应用,为实现中文本体在基于语义网的信息检索领域的应用奠定了基础。如果有一款功能强大的中文本体推理工具,那么基于中文本体的信息检索技术将会快速地发展起来,开发一款中文本体推理工具显得意义重大,值得期待。

参考文献:

- [1] 耿科明,袁 方. Jena 推理机在基于本体的信息检索中的应用[J]. 微型机与应用,2005,24(10):62-64.

(下转第 201 页)

主要识别水平左移、水平右移、垂直上移、垂直下移、右上移、右下移、左上移、左下移几种动作,每个动作测试 10 次。则正确识别次数如表 1 所示。

表 1 肢体动作识别结果表

实验结果/次	水平左移	水平右移	垂直上移	垂直下移	右上移	右下移	左上移	左下移
5 ~ 10/(cm/s)	6	6	7	8	7	8	7	7
10 ~ 15/(cm/s)	8	9	9	8	9	8	8	8
大于 15/(cm/s)	9	10	10	10	10	9	10	10

根据 MMA8452Q 线加速度计公式: $A(\text{加速度}) = F(\text{惯性力})/M(\text{质量})$, 结合实验结果可以看出, MMA8452Q 移动速度越大,根据惯性定律,通过电磁力去平衡就可以测得 F 值就越明显,从而可以得到 F 与电流的关系值也就越大,识别效果也就越理想。在速度大于 10 cm/s 时识别率达到 91%,可基本满足家电肢体动作控制要求。

4 结束语

该系统使用加速度传感器 MMA8452Q 作为数据采集模块,STC89C52RC 单片机最小系统作为数据处理模块,MAX232 作为串行数据传输模块来设计实现了肢体动作控制系统。系统具有性能稳定、实时性强、集成度高、扩展性灵活等特点,体现了简单、安全、有效、可靠的肢体动作识别系统的设计理念,适用于家电控制等各种场所。实验表明,系统运行稳定、实时性

好、安全性高。

参考文献:

[1] Mantyjarvi J, Juha K, Panu K, et al. Enabling fast and effort-less customisation in accelerometer-based gesture interaction [C]//Proc of the 3rd international conference on mobile and ubiquitous media. Maryland, USA: [s. n.], 2004:25-31.

[2] 龚陶波. 基于计算机视觉的静态肢体动作识别系统[D]. 武汉:华中师范大学,2008.

[3] 熊书来. 基于 ARM7 TDMI 模拟微控制器的 I2C 接口应用[J]. 软件导刊,2010(9):126-127.

[4] 张会敏,谢泽奇,张云龙. 红外图像报警系统的研究与实现[J]. 计算机测量与控制,2011,19(2):432-435.

[5] 杜树春. 单片机 C 语言和汇编语言混合编程[M]. 北京:北京航空航天大学出版社,2008.

[6] 张会敏,卢兰涛,崔 鹏,等. 双阈值背景差分算法在嵌入式报警系统中的应用[J]. 传感器与微系统,2010,29(10):138-140.

[7] 张秋余,姚开博,吴佩莉. 基于矩形特征和改进 Adaboost 的手势检测[J]. 计算机工程,2008,34(14):176-178.

[8] 刘 蓉,刘 明. 基于三轴加速度传感器的手势识别[J]. 计算机工程,2011,37(24):141-143.

[9] 陈 意,杨 平,陈旭光. 一种基于加速度特征提取的手势识别方法[J]. 传感技术学报,2012,25(8):1073-1078.

[10] MMA8452Q 使用手册[M]. 出版地不祥: Freescale 公司, 2010.

[11] 赵全利,张之枫. 单片机原理及应用[M]. C51 版. 北京:机械工业出版社,2012.

[12] 王东锋. 单片机 C 语言应用 100 例[M]. 北京:电子工业出版社,2009.

(上接第 197 页)

[2] Gennari J H, Musen M A, Ferguson R W, et al. The evolution of Protégé: An environment for knowledge-based systems development[J]. International journal of human-computer studies, 2003, 58(1):89-123.

[3] McGuinness D L, van Harmelen F. OWL Web ontology language overview[M]. [s. l.]: [s. n.], 2004.

[4] 郁 鹏,周 济,温淑红. 知识处理中的推理机设计[J]. 小型微型计算机系统,1999,20(1):20-24.

[5] 高 琦,陈华钧. 互联网 Ontology 语言和推理的比较和分析[J]. 计算机应用与软件,2004,21(10):73-76.

[6] 杨 力. 从 RDF, DAML+OIL 到 OWL-Ontology 语言比较[J]. 农业图书情报学刊,2005,17(11):108-110.

[7] 徐德智,汪智勇,王 斌. 当前主要本体推理工具的比较分析与研究[J]. 现代图书情报技术,2006(12):12-15.

[8] 陈光仪. 基于关系数据库的本体存储研究[D]. 长沙:中南

大学,2009.

[9] 李 勇,李跃龙. 基于关系数据库存储 OWL 本体的方法研究[J]. 计算机工程与科学,2008,30(7):105-107.

[10] 何 伟,杨小平. 基于词间语义关联性的本体扩展[J]. 计算机应用与软件,2011,28(11):73-76.

[11] 李 渝. 本体查询及自动完善系统研究[D]. 重庆:重庆大学,2010.

[12] 张宗仁,杨天奇. 基于自然语言理解的 SPARQL 本体查询[J]. 计算机应用,2010,30(12):3397-3400.

[13] Welty C, McGuinness D L. OWL Web ontology language guide [EB/OL]. 2004-02. <http://www.w3.org/TR/2004/REC-owl-guide-20040210>.

[14] 刘 宁,李冠宇,邵 彬. Jena2 推理机制的研究[J]. 微机计算机信息,2010,26(11-3):173-175.

面向语义网的中文本体应用研究

作者：[凌绍东](#)，[霍林](#)，[王超](#)，[LING Shao-dong](#)，[HUO Lin](#)，[WANG Chao](#)

作者单位：[广西大学 计算机与电子信息学院, 广西 南宁, 530004](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

ISTIC

年，卷(期)：2014(2)

本文链接：http://d.g.wanfangdata.com.cn/Periodical_wjfz201402049.aspx