

本体的查询与推理研究

纪兆辉

(淮海工学院 计算机工程学院, 江苏 连云港 222005)

摘 要: 基于 OWL 所构建的旅游本体为数据源, 借助 SPARQL 查询语言, 对其进行各类相关查询, 以此为基础, 对本体的推理机制也进行了相关研究, 为基于本体的相关语义应用提供很好的借鉴。

关键词: 本体; 查询; 推理; OWL; SPARQL

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2011)10-0052-04

Research on Ontology Querying and Inference

JI Zhao-hui

(School of Computer Engineering, Huaihai Institute of Technology, Lianyungang 222005, China)

Abstract: It uses the tourism ontology as data source and makes relevant query by the aid of SPARQL language. Besides, the preliminary study on ontology inference is also illustrated. The views and results put forward by us can provide good examples for relevant semantic applications based on ontology.

Key words: ontology; query; inference; OWL; SPARQL

1 引言

本体是共享概念模型的明确的形式化规范说明^[1]. 本体的本质表现在: (1)概念化: 指通过抽象出客观世界中一些现象的相关概念而得到的模型; (2)明确: 指概念及它们之间的联系都有明确的定义; (3)形式化: 指精确的数学描述, 计算机可读; (4)共享: 指本体中体现的知识是被不同的人和组织所共同认可的, 可以跨应用系统进行互操作。

形式化定义的本体不仅能够描述特定领域内的相关概念, 而且能够描述概念之间的语义联系. 使用本体作为信息检索的载体, 用户能够方便地在概念上描述信息需求, 构造复杂的语义查询. 目前, 本体在知识管理、语义检索、自然语言理解等诸多领域得到广泛的研究和应用. 围绕本体所蕴含的知识进行有效地查询与智能化推理, 是本体在各个领域能够成功应用的基础和关键。

目前大部分开发和研究人员都基于 OWL 进行本体描述. 文中, 我们以基于 OWL 所开发的旅游本体为数据源, 借助 SPARQL 查询语言, 对其进行各

类相关查询, 以此为基础, 对本体的推理机制也进行了相关研究。

2 OWL 本体概述

2.1 本体的构成

可将本体 O 定义如下:

$O = \{ \text{Concept}, \text{Property}, \text{Relation}, \text{Constraints}, \text{Instances} \}$. 其中:

(1) $\text{Concept} = \{ c_1, c_2, \dots, c_n \}$ 是概念的集合. 如在我们所构建的旅游本体中, “地区”、“旅游景点”、“公交”、“宾馆”、“游客”、“土特产”等都属于概念。

(2) $\text{Property} = \{ p_1, p_2, \dots, p_m \}$ 是概念的属性集合, 包括数据类型 (DataType) 属性和对象 (Object) 属性. 如: 对“旅游景点”这个概念来说, 它拥有“景点名称”、“景点级别”、“所在区域”等数据类型属性, 同时具有“公交线路”、“附近宾馆”、“风土人情”、“适合人群”等对象属性. 通过对象属性可以在不同个体之间建立语义关联。

(3) $\text{Relation} = \{ r_1, r_2, \dots, r_l \}$ 是概念之间各种联系的集合, 如父子关系、等价关系、互斥关系等. 如

“自然景观”、“人文景观”两个概念和“旅游景点”这个概念之间就存在父子联系.而“山”、“水”和“自然景观”之间也存在父子联系.

(4) $Constraints = \{cs_1, cs_2, \dots, cs_k\}$ 表示属性的取值范围和概念实例间的基数约束.

(5) $Instances = \{is_1, is_2, \dots, is_q\}$ 是概念的实例,即个体的集合.如“花果山”、“孔望山”是概念“山”的实例.

本文的本体基于 OWL DL 来描述, DL 知识库主要由 TBox 和 ABox 组成, 前面是描述概念关系的术语公理集, 后者是描述个体实例的断言公理集. OWL 本体库由公理和事实组成, 公理对应于 TBox、事实对应于 ABox. 为了使读者对下文的查询和推理有更直观的理解, 以下结合旅游本体, 对 OWL 的类特征和属性特征进行详细阐述^[2].

2.2 OWL 的类特征

(1) 父子类(subClassOf)

若有类 A、B, 其中 B 是 A 的子类, 则称 A、B 是父子关系. 若实例 I 是 B 类的对象, 则其一定是 A 的对象. 如:“生态旅游”类为“旅游景点”类的子类.

(2) 等价类(equivalentClass)

两个类如果具有不同的名字但拥有相同的个体集合, 则被称为等价类. 如:“旅游”类可以被说成是“travel”类的等价类. 因此, 任何“旅游”类的实例同时都是“travel”类的实例, 反之也一样. 在此, 可以称“旅游”和“travel”两个概念为同义词, 在信息检索系统中, 可以通过同义词对查询关键字进行扩展, 从而提高检索的查全率.

(3) 互斥类(disjointWith)

互斥类是指不相交的两个类. 如:“山”和“水”两个类为互斥类, 这样即可推断出: 如果“花果山”为“山”的实例, 那么“花果山”一定不是“水”的实例.

(4) 互补类(complementOf)

由 A、B 两个事件构成一个整体事件, 当有事件发生时, A、B 两者之间有且仅有一个发生. 如: 设连云港仅有“山”和“水”两种形态的旅游方式, 那么如果某人去连云港旅游过, 但是他没有看到连云港的“水”, 则他一定看到了连云港的“山”.

(5) 相交类(intersectionOf)

相交类是指两个类之间存在交集. 如:“连岛”既属于“水”这一类旅游景点, 又属于“生态旅游”类景点, 因此在本知识库中“水”和“生态旅游”就是两个相交类.

2.3 OWL 的属性特征

(1) 函数属性(FunctionalProperty)

对于任意函数属性 P, 若存在 $P(x, y)$ 和 $P(x, z)$, 则有 y 和 z 必是同一个个体或文字. 如: X 宾馆有一营业执照 Y, 且 X 宾馆有一营业执照 Z, 由于同一家宾馆的营业执照具有唯一性, 则可以推导出 Y, Z 是同一个营业执照.

(2) 对称属性(symmetricProperty)

对于任意对称属性 P, 若存在 $P(x, y)$, 则有 $P(y, x)$. 如: adjacent_area 是对称属性, 如果“东磊”是“渔湾”的 adjacent_area, 则可以推断“渔湾”也是“东磊”的 adjacent_area.

(3) 逆反属性(inverseOf)

一个属性 A 被称为另一个属性 B 的逆反属性, 如果对任意个体 x、y 间有 A 关系, 当且仅当 y、x 间有 B 关系. 如:“地区”概念的 contains 属性与“旅游景点”概念的 located_in 属性是逆反属性. 即由某地区“contains”某个旅游景点, 可推断出这个旅游景点“located_in”该地区.

(4) 传递属性(TransitiveProperty)

对于任意传递属性 P, 如果存在 $P(x, y)$ 和 $P(y, z)$, 则有 $P(x, z)$ 成立. 如: 设 isNear(a, b) 表示 a 是 b 的邻近景区, 且有 isNear(花果山, 渔湾), isNear(渔湾, 东磊), 则可推断 isNear(花果山, 东磊) 成立.

3 OWL 本体的查询

目前, SPARQL^[3] 的查询规范已经成为 W3C 的推荐标准, 其相对于本体的作用类似于 SQL 对于关系数据库的重要性, 是对 RDF、OWL 本体进行查询的首选语言. SPARQL 查询包含三元组模式的集合, 称为基本图模式. 三元组模式类似于 RDF 三元组, 不同之处在于主语、谓语和宾语可能是变量. 一个基本图模式匹配 RDF 数据的一个子图, 其条件是子图的 RDF 术语可以被变量置换, 而结果为一个和原来子图等价的 RDF 图.

对于本体的查询 $Q = q_1 \wedge q_2 \wedge \dots \wedge q_n$, 其中的每个原子查询 q_i , 其抽象语法形式^[4]如表 1 的第 1 列所示, 其 RDF 三元组形式如表 1 的第 2 列所示.

3.1 TBOX 查询

TBOX 查询是指对概念本身、概念的属性及约束条件和概念间相互关系的查询^[5-6]. 在实际的语义检索系统中, 经常需要对查询的关键词做上下位、同义和关联性等方面的扩展, 这时就要借助于 TBOX 术语体系的查询. 目前与本体查询和推理相关的应

用程序接口主要有 Jena OWL API、Protégé OWL API 等,在程序中都需要借助于 SPARQL 的三元组语法来查询 OWL 本体知识库。以下结合我们所建立的基于 OWL DL 的旅游本体知识库,给出具有代表性的查询实例,希望对读者有所启发。

表 1 原子查询与对应的三元组形式

原子查询的抽象语法形式	RDF 三元组形式
Type(a, C)	<a, rdf: type, C>
PropertyValue(a, p, v)	<a, p, v>
sameAs(a, b)	<a, owl: sameAs, b>
differentFrom(a, b)	<a, owl: differentFrom, b>
subClassOf(C ₁ , C ₂)	<C ₁ , rdfs: subClassOf, C ₂ >
equivalentClass(C ₁ , C ₂)	<C ₁ , owl: equivalentClass, C ₂ >
Disjointwith(C ₁ , C ₂)	<C ₁ , owl: Disjointwith, C ₂ >
complementOf(C ₁ , C ₂)	<C ₁ , owl: complementOf, C ₂ >
subPropertyOf(p, q)	<p, rdfs: subPropertyOf, q>
equivalentProperty(p, q)	<p, owl: equivalentProperty, q>
ObjectProperty(p)	<p, rdf: type, owl: ObjectProperty>
DataTypeProperty(p)	<p, rdf: type, owl: DataTypeProperty>
FunctionalProperty(p)	<p, rdf: type, owl: FunctionalProperty>
InverseFunctionaly(p)	<p, rdf: type, owl: InverseFunctional>
TransitiveProperty(p)	<p, rdf: type, owl: TransitiveProperty>
SymmetricProperty(p)	<p, rdf: type, owl: SymmetricProperty>
AnnotationProperty(s, p _a , o)	<s, p _a , o>
Domain(p, C)	<p, rdfs: domain, C>
Range(p, C)	<p, rdfs: range, C>

例 1: 查询某个概念的子类。

如查询“旅游景点”类的子类,其对应的 SPARQL 语句如下:

```
SELECT ? 子类
```

```
WHERE { ? 子类 rdfs: subClassOf travel: 旅游景点 }
```

以上语句中的前缀 travel 代表旅游本体知识库中元素的名字空间,而 rdf、rdfs、owl 则分别是 W3C 有关 rdf、rdfs 和 OWL 规范的定义。上述语句从知识库中提取出“旅游景点”的所有子类:包括:“传统文化”、“生态旅游”、“度假旅游”、“人文景观”、“自然景观”等。

例 2: 查询某个概念的等价类。

如查询概念“旅游景点”的等价类,其对应的 SPARQL 语句如下。该语句以及后续的查询语句中用到的前缀同上例,所以全部省略。

```
SELECT ? 旅游景点等价类
```

```
WHERE { travel: 旅游景点 owl: equivalentClass ? 旅游景点等价类 }
```

例 3: 查询某个概念的对象属性 (Object Property)。

如查询概念“旅游景点”的所有对象属性,其对应的 SPARQL 语句如下:

```
SELECT ? 旅游景点对象属性
```

```
WHERE { ? 旅游景点对象属性 rdfs: domain ? travel: 旅游景点 }
```

例 4: 查询与某个概念具有关联的概念。

如查询与概念“自然景观”有关联的其他概念,其对应的 SPARQL 语句如下:

```
SELECT ? 自然景观相关概念
```

```
WHERE { { ? x rdfs: domain ? travel: 自然景观 }
{ ? x rdfs: range ? 自然景观相关概念 } }
```

例 5: 查询知识库中某个概念的标注信息。

在语义信息检索系统中,某个概念的标注属性信息,同样可以用来对查询进行扩展。OWL 本体中概念的标注属性包括: rdfs: seeAlso、rdfs: isDefinedBy、rdfs: Comment、rdfs: label。

对“游客”概念注释属性的查询语句如下:

```
SELECT ? 游客的注释信息
```

```
WHERE { travel: 游客 rdfs: comment ? 游客的注释信息 }
```

3.2 ABOX 查询

ABOX 查询是对知识库中已有事实进行查询,涉及到实例本身、实例的属性值及实例间的关系^[5]。

例 6: 查询某个实例的属性值。

如查询实例“花果山”邻近宾馆的相关信息,其相应的 SPARQL 语句如下:

```
SELECT ? 宾馆名称 ? 房间类型及价格 ? 联系电话 ?
地址 ? 星级
```

```
WHERE { { travel: 花果山 travel: hasNearHotels ? x }
{ ? x travel: 名称 ? 宾馆名称 }
{ ? x travel: 房间类型及价格 ? 房间类型及价格 }
{ ? x travel: 联系电话 ? 联系电话 }
{ ? x travel: 地址 ? 宾馆地址 }
{ ? x travel: 星级 ? 星级 } }
```

3.3 混合查询

该类查询是将术语体系和事实进行联合查询,篇幅所限,此处不给出示例。

4 OWL 本体的推理

本体中蕴含了丰富的语义知识,对本体知识库进行推理,是以本体作为信息组织形式比传统数据库更智能化的方面^[7]。对于本体的使用者来说,推理的作用主要是为了获得本体中特定形式的知识集合以及运用本体中的知识来辅助解决涉及语义的应用。

本体推理机由本体解析器、查询分析器、推理引擎、结果展现等模块组成,如图 1 所示。

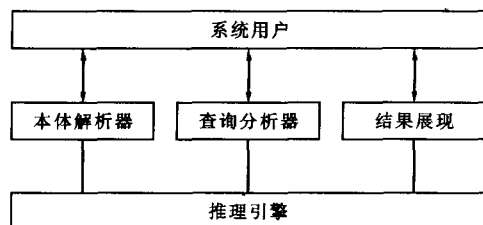


图1 本体推理机的构成

其中本体解析器提供了对文件形式或者存储在数据库中的本体进行解析,查询分析器为用户提供了分析用户查询命令的接口,推理引擎是整个推理机的核心,其内部结构对本体推理能力的影响至关重要。推理的结果可以以 RDF 或其他形式展现出来。就使用推理机的方式而言,既有操纵本体文件的接口 OWL API,又有以 DIG 标准实现的服务器接口,可以让外部程序通过 http 协议与服务器通信从而达到使用推理机的目的。

推理机常用的规则主要分为以下两种:

(1)RDFS, OWL 内置的规则,包括类的公理、属性公理等语言自身定义的关系限制,是进行本体分类等操作的主要方法。如:旅游知识库中已经把“山”和“水”定义为互斥类,且已知“花果山”为“山”的实例,则推理机可以推出“花果山”一定不是“水”的实例。

(2)自定义规则:按照某种规定语法所明确定义的规则,能够灵活地构造各种关系,可以将此规则写在应用程序里或者单独保存在规则文件中。

目前一些研究机构已开发了一些基于本体的实际使用的推理机,比较流行的有 Racer、Pellet 和 Jena。我们对旅游本体知识库进行推理使用的是 Jena^[8]自带的推理机,为了对旅游知识库进行深层次的应用,让系统更加智能化,我们围绕知识库中的术语体系和公理定义了一系列的规则,并将这些规则保存在项目的若干个 rules 文件中。其中第 i 个规则文件的格式如下:

$[rule_i: (三元组_1), \dots (三元组_n) \rightarrow (三元组_{n+1})]$

其中,前 n 个三元组为规则的前提,第 $n+1$ 个三元组为规则的结论,此处采用 Jena 的前向推理 (Forward Chaining) 引擎。每个三元组具有形式: subject predicate object。其中 subject 和 object 分别为主语和宾语,一般为概念,而 predicate 为谓词,一般为对象属性。

定义好规则以后,运行推理引擎,则基于规则和

OWL 知识库中已有的公理和事实进行推理,然后用推理后得到的新事实更新 OWL 知识库。

5 结束语

本文结合旅游本体知识库,在对 OWL 本体的类特征和属性特征进行分析的基础上,通过示例,对典型的基于 SPARQL 的本体查询进行阐述,同时对本体的推理进行了初步探讨。通过我们对本体查询和推理的初步实践,认为:本体的查询与推理是相关语义应用的基础与关键所在,可以在一定范围内实现语义扩展、关联搜索等智能推理的应用。

今后的研究将针对所创建的旅游本体,利用 Jena 和 Pellet 推理机实现一个基于 Web 且能发挥本体智能推理优势的实用系统。

参考文献:

- [1] Knowledge System Laboratory, Stanford University. What is an ontology[EB/OL]. (2009-01-28)[2011-05-10]. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- [2] World Wide Web Consortium. OWL web ontology language overview [EB/OL]. (2009-11-12)[2011-05-10]. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
- [3] World Wide Web Consortium. SPARQL query language [EB/OL]. (2008-01-15)[2011-05-12]. <http://www.w3.org/TR/rdf-sparql-query/>.
- [4] Evren Sirin, Bijan Parsia. SPARQL-DL: SPARQL query for OWL-DL[C]// 3rd OWL Experiences and Directions Workshop (OWLED-2007). Innsbruck, Austria, 2007.
- [5] 高志强,潘越,马力. 语义 Web 原理及应用[M]. 北京:机械工业出版社,2009.
- [6] 王维,解争龙,许青林. 觉察上下文计算平台的研究[J]. 微电子学与计算机,2010,27(6):170-173.
- [7] 宋晓峰,唐发根. 基于本体的推理技术的相关研究[EB/OL]. (2007-09-27)[2010-8-22]. <http://www.paper.edu.cn>.
- [8] Ian Dickinson. The Jena Ontology API[EB/OL]. (2009-02-24)[2010-08-20]. <http://jena.sourceforge.net/ontology/index.html>.

作者简介:

纪兆辉 男,(1971-),硕士,副教授。研究方向为语义 Web 与本体、数据挖掘。