

Web 知识本体集成

李 剑

(中科院软件所软件工程技术中心 北京 100080)

摘 要 WWW(World Wide Web)是一个开放、分布式的环境,其中包含了大量的信息,使用知识本体可以帮助用户进行网络信息获取。在 WWW 上存在许多版本的知识本体,知识本体之间可以通过引用关系来实现知识本体之间的互操作和重用。由于 WWW 的快速增长特性以及 Web 知识本体之间的相对独立性,如何集成知识本体成为更好地支持 Web 知识本体重用的关键问题。本文提出了一种将知识本体表示转换为一阶谓词逻辑形式表示的方法,并在此基础上提出了检查知识本体与其引用的知识本体一致性的方法以及实现知识本体集成方法。

关键词 Web 知识本体,知识本体集成,知识本体重用

Web Ontologies Integration

LI Jian

(Technology Center of Software Engineering, Institute of Software, Chinese Academy of Sciences, Beijing 100080)

Abstract The World Wide Web is an open repository with lots of information, in which the ontologies can help users to get their required information. There are varies of ontologies in the WWW. By reference, ontologies can be reused by others to implement interoperability of them. As the rapid evolution and autonomy of Web ontologies, how to integrate ontologies becomes the key to reuse Web ontologies. This paper presents a method of mapping ontologies to first order logic. Then, it also presents methods to check the ontologies' consistency and integrate the ontologies based on the former method.

Keywords Web ontology, Ontologies integration, Ontology reusing

1 介绍

WWW(World Wide Web)上包含了大量的知识信息,HTML 网页是表示这些信息的主要载体,它可以提供不同形式的知识表示(例如:文本,图片,图像,声音等)。HTML 文档之间通过超链接(HyperLink)相联系,用户可以通过点击超链接来浏览分布的网页。但是,WWW 现在还不能很好地为用户提供良好的信息检索服务,其主要提供的信息查找方式为 Web 搜索引擎或者 Web 目录系统,用户主要通过输入关键词在 Web 上搜索相关的网页信息,这样很容易返回给用户大量的无关网页。

使用知识本体(Ontology)可以帮助进行 Web 信息查找,提高信息查找的准确性。本体本来是一个哲学的概念,其为“现实世界中各种类型的对象、属性、事件、过程和关系的哲学表示”^[1]。在计算机科学中,知识本体可以看作作为一种知识共享和重用的方式,多个智能代理(Agent)共同遵从某个知识本体,共享和重用这一知识本体的知识。在 Web 上,用户可以用知识本体来标记 HTML 网页,这种标记表达一种原知识,也可以看作是网页内容的某种语义信息。这样支持知识本体的网络智能代理可以根据这些语义信息来更准确地获取所需的知识信息。

WWW 上存在着不同形式的知识本体,它们可以采用不同的表达方式来标注与之相关的知识,例如 SHOE(Simple HTML Ontology Extensions)^[7], Ontobroker^[9]等。为了更好

地支持知识本体的重用,在这些语言的内部一般都支持本知识本体对其他知识本体的引用。但由于 WWW 是一个典型的开放、异质性的环境,这些知识本体之间存在着很大的自治性。如何检测这些知识本体之间的一致性,以及如何集成已有的知识本体以更好地支持知识本体的重用,这些都是亟待解决的问题。本文提出了一种将知识本体转换为一阶谓词逻辑表示的方法,并在此方法的基础上提出了引用知识本体一致性检测以及 Web 知识本体集成的方法。

2 Web 知识本体

WWW 主要由大量的 HTML 网页组成,Web 知识本体通常表示某一领域的实体、关系、属性等知识。使用这些知识本体来标注 HTML 网页就能为其信息附加语义上的标记,特定的网络智能代理可以利用这些语义信息来更好地搜索用户所需的信息。下面以 SHOE^[7]语言为例来说明 Web 知识本体的一般特性和使用方式。

SHOE 语言是一种基于知识本体的知识表示语言,它通过一些表示原知识的标签来扩展 HTML 语言,并将这些标签和一个或多个 SHOE 知识本体联系起来。这样,因特网智能代理就能够根据这些标记所提供的知识来进行搜索查询。另外,知识本体还能够使用相关规则来提供隐藏的知识表示,智能代理可以根据已有的知识以及相关规则推导出隐藏的知识。图 1 中就表示用 SHOE 描述的一个知识本体,其中:〈ONTOLOGY〉中表示所定义的知识本体。〈USE-ONTOLOGY〉

*)本文研究得到国家重点基础研究发展规划 973 资助项目(G1999035806)和国家自然科学基金重点基金(69833030)的资助。李 剑 博士研究生,主要研究方向为基于内容的网络信息处理。

表示这一知识本体引用的其他知识本体,一般这些知识本体也是 SHOE 形式的。在引用到其他知识本体时,有时需要使用其他知识本体的术语,SHOE 使用<DEF-RENAME>将被引用的知识本体中的术语定义更名,以便在本知识本体中使用,它们在语义上保持一致。SHOE 知识本体使用<DEF-CATEGORY>定义目录,目录表示实体所属的范围,可以将其看作为一个类定义的概念,图 1 中所定义的“FamilyMember”目录可以看作是家庭成员集合。在 SHOE 中,<DEF-RELATION>为关系定义,关系表示对象实体之间的联系,图 1 中所定义的“MarriedTo”关系就表示家庭成员之间的婚姻关系。另外,在 SHOE 中还可以用<DEF-INFERNCE>来定义推论规则,例如:图 1 的例子中所定义的推论规则表示的意义为:“如果 X,Y 都是家庭成员,并且 X 和 Y 结婚了,那么 Y 也和 X 结婚了”。这些推理规则可以看作表示知识本体中实体间潜在关系的公理,根据这些公理可以推导出一些有用的相关信息。

```
<HTML><BODY>
<ONTOLOGY ID="FamilyRelation-ontology" VERSION="1.1"
  BACKWARD-COMPATIBLE-WITH="1.0">
<USE-ONTOLOGY ID="People-ontology" Version="1.0" PREFIX
  ="u" URL="...">
<DEF-RENAME FROM="u. Person" TO="Person">
<DEF-CATEGORY NAME="FamilyMember" ISA="Person">
<DEF-RELATION NAME="MarriedTo">
  <DEF-ARG POS=1 TYPE="FamilyMember">
  <DEF-ARG POS=2 TYPE="FamilyMember">
</DEF-RELATION>
<DEF-INFERNCE>
  <INF-IF><CATEGORY NAME="Person" VAR FOR="X">
  <CATEGORY NAME="Person" VAR FOR="Y">
  <RELATION NAME="MarriedTo">
    <ARG POS=1 VAR VALUE="X">
    <ARG POS=2 VAR VALUE="Y">
  </RELATION></INF-IF>
  <INF-THEN><RELATION NAME="MarriedTo">
    <ARG POS=1 VAR VALUE="Y">
    <ARG POS=2 VAR VALUE="X">
  </RELATION></INF-THEN>
</DEF-INFERNCE>
</ONTOLOGY>
</BODY></HTML>
```

图1 SHOE 知识本体定义

使用知识本体来标注 HTML 网页可以表示网页内容的语义信息。图2表示 George 个人主页的一部分片段,其使用图 1 中的知识本体来标记本网页。从这一标记可以知道 George 是一个家庭成员,并且有关其配偶的信息可以在网页 <http://localhost:8080/helena> 上找到。通过这种知识本体标记,在查询诸如 George 配偶的有关信息时,识别 SHOE 语言网络智能代理会自动地在对应的网页上查询相关信息。相对于利用关键词的检索而言,这种利用语义信息的查询方式明显会增加信息查找的准确性。

```
<META HTTP-EQUIV="FamilyRelation-ontology" CONTENT="
  ...">
<USE-ONTOLOGY "our-ontology" VERSION="1.0" PREFIX="
  our" URL="...">
<CATEGORY "our. FamilyMember">
<RELATION "our. MarriedTo"="http://localhost:8080/helena">
```

图2 使用 SHOE 标记的 HTML 片段

3 知识本体一致性

知识本体之间通过引用重用其他知识本体的知识,由于 Web 知识本体的自治性,被引用到的知识本体和本知识本体之间可能存在冲突,我们采用将知识本体转换为一阶谓词逻辑的方式来检测知识本体的一致性。

3.1 知识本体表示

首先我们给出知识本体一致性的形式化定义:

定义1 一个知识本体 $o(o \in O, O$ 为知识本体的集合)可以表示为一个三元组 $\langle V, A, I \rangle$, 其中 V 是 A 中使用到的术语定义的集合, A 是一系列公理的集合。 $I \subset O$ 表示本知识本体所引用到的知识本体集合。

定义2 假设两个知识本体 $o(o = \langle V, A, I \rangle), o'$, 如果 $o' \in o. I$ ($o. I$ 表示知识本体 o 的引用集合), 则称知识本体 o 引用知识本体 o' (或知识本体 o' 被知识本体 o 引用到), 记为 $o \rightarrow o'$ 。

在 SHOE 知识本体中, 我们可以将所定义的目录和关系申明部分都看作为知识本体 o 的术语定义 (V) 部分。而将所定义的推理规则看作是知识本体 o 的公理集 (A) 部分。

定义3 对于一个知识本体 $o = \langle V, A, I \rangle$:

1. 如果 $I = \Phi$ (Φ 表示空集), A 是一致 (公理集内部不冲突), 则称知识本体 o 是一致的。

2. $I \neq \Phi$, 假设 $o_1, o_2, \dots, o_n \in I$, 如果对于每个 $o_i \in I$, o_i 是一致的, 并且 $A \cup o_1. A \cup \dots \cup o_n. A$ 是一致的 ($o_i. A$ 表示知识本体 o_i 的公理集), 则知识本体 o 是一致的。

定义3表示如果一个知识本体没有引用其他知识本体, 并且知识本体本身的公理集是一致的, 则此知识本体是一致的。如果一个知识本体引用到其他知识本体, 则要求这些被引用到的知识本体是一致的, 并且它们的公理集和本知识本体的公理集没有冲突, 那么此知识本体是一致的。

需要说明的是: 在知识本体引用中, 我们通常认为是不允许循环引用的, 不允许一个知识本体直接或间接引用的知识本体引用到自己, 也就是不存在这样一个知识本体引用序列, 使得 $o \rightarrow o_1 \rightarrow o_2 \rightarrow \dots \rightarrow o_i \rightarrow o$ 。定义3就是在此假设的基础上定义知识本体的一致性。

3.2 映射到一阶谓词逻辑

为了检测 SHOE 知识本体之间的一致性, 我们可以采用一种映射方法将 SHOE 表示的知识本体按照表1所对应的形式转换到一阶谓词形式。

术语定义集 V : MarriedTo, Person, FamilyMember;
引用集 I : {People-ontology};
公理集 A :
 $Instance(People-ontology. Person, x) \leftrightarrow Instance(Person, x), Sub(FamilyMember, Person)$
 $MarriedTo(x_1, x_2) \rightarrow Instance(FamilyMember, x_1)$
 $MarriedTo(x_1, x_2) \rightarrow Instance(FamilyMember, x_2)$
 $Instance(FamilyMember, x) \wedge Instance(FamilyMember, y) \wedge MarriedTo(x, y) \rightarrow MarriedTo(y, x)$

图3 知识本体的一阶谓词表示

这一映射转换方法将 SHOE 中的目录和关系定义放入术语申明集合中, 还将它们转换为一些逻辑公式定义, 其中: Sub 表示目录之间的父子关系; Instance 表示实体属于某目录。另外, 这一方法将 SHOE 中的推论规则转换为一阶谓词逻辑中的蕴涵关系来表示。对于在引用其他知识本体时所用到的换名定义, 本方法采用等价关系来建立原名和现在所实用名之间的等价关系。同时, 此转换方法中还加入了一系列附加公理, 这些公理表示目录之间以及目录与实体之间普遍存在的关系, 例如: 子目录关系的传递性等, 它们构成附加公理集合 A_v , 在检查知识本体之间的一致性的时候会使用到这一公理集合。

图3中所表示的就是将图2表示的 SHOE 知识本体转换为用一阶谓词方式所表示的知识本体。

表1 SHOE 语言与一阶谓词逻辑的映射转换

SHOE 语言部分	术语定义 V	一阶谓词逻辑公式 A	说明
知识本体内部定义			
$\langle \text{DEF-CATEGORY NAME} = X \text{ ISA} = Y \rangle$	X	$\text{Sub}(X, Y)$	表示一个目录是另外一个目录的子目录。
$\langle \text{CATEGORY NAME} = X \text{ VAR FOR} = x \rangle$		$\text{Instance}(X, x)$	表示一个实体属于一个目录。
$\langle \text{DEF-INFERENCE} \rangle$ $\langle \text{INF-IF} \rangle \dots \langle \text{INF-IF} \rangle$ $\langle \text{INF-THEN} \rangle \dots$ $\langle \text{INF-THEN} \rangle$ $\langle \text{DEF-INFERENCE} \rangle$		$(\dots) \rightarrow (\dots)$	其推理规则对应于蕴涵关系。其中,在 $\langle \text{INF-IF} \rangle$ 或者 $\langle \text{INF-THEN} \rangle$ 中定义的各子元素所对应的子句之间是逻辑与的关系。
$\langle \text{DEF-RELATION NAME} = A \rangle$ $\langle \text{DEF-ARG POS} = 1 \text{ TYPE} = X_1 \rangle$ $\dots$ $\langle \text{DEF-ARG POS} = n \text{ TYPE} = X_n \rangle$ $\langle \text{DEF-RELATION} \rangle$	A	$A(x_1, \dots, x_n) \rightarrow \text{Instance}(X_1, x_1)$ $\dots$ $A(x_1, \dots, x_n) \rightarrow \text{Instance}(X_n, x_n)$	根据 SHOE 中定义的关系定义推导出关系的参数实体属于对应的目录。
$\langle \text{RELATION NAME} = A \rangle$ $\langle \text{ARG POS} = 1 \text{ VAR VALUE} = x_1 \rangle$ $\dots$ $\langle \text{ARG POS} = n \text{ VAR VALUE} = x_n \rangle$ $\langle \text{RELATION} \rangle$		$A(x_1, \dots, x_n)$	给出关系在一阶谓词逻辑中的定义。
引用定义			
$\langle \text{DEF-RENAME FROM} = "u. A" \text{ TO} = B \rangle$		(1) $\text{Instance}(A, x) \leftrightarrow \text{Instance}(B, x)$ 或者 (2) $A(x_1, \dots, x_n) \leftrightarrow B(x_1, \dots, x_n)$	u 表示本知识本体所引用的知识本体。根据 A 在 u 知识本体中的类型,如果 A 为一个目录定义,则翻译为(1),如果 A 为一个关系定义,则其翻译为(2)。
附加公理 A_u			
自反性	$\forall X, \text{Sub}(X, X)$		
传递性1	$\forall X, Y, Z \text{Sub}(X, Y) \wedge \text{Sub}(Y, Z) \rightarrow \text{Sub}(X, Z)$		
传递性2	$\forall X, Y, Z \text{Sub}(X, Y) \wedge \text{Instance}(X, z) \rightarrow \text{Instance}(Y, z)$		
单一性	$\forall \text{Sub}(X, Y) \wedge \text{Sub}(Y, X) \rightarrow (X = Y)$		

3.3 知识本体的一致性检测

通过将知识本体转换为一阶谓词表示形式,我们可以利用一阶谓词逻辑的一致性检测方法检测知识本体的一致性,具体方法如下:

1. 对于一个知识本体 o ,如果它不引用其他知识本体,则使用表1中知识本体内部定义转换部分,将知识本体转换为一阶谓词形式,构成其公理集合 A ,检测 $AY A_u$ 的一致性。如果其是一致的,则根据定义3可知这一知识本体 o 是一致的。

2. 对于一个引用其他知识本体的知识本体 $o = \langle V, A, I \rangle$,我们首先检测其所引用到的知识本体的一致性,对于每个被引用到的知识本体 $o_i \in I$,重复使用本转换检测方法,检查 o_i 的一致性。然后,在每个知识本体都一致的基础上,检测 $AY o_1, AY \dots Y o_n, AY A_u$ 所对应一阶谓词公式集合的一致性,如果这些一阶谓词公式集合都是一致的,根据定义3可知本知识本体 o 是一致的。

由于知识本体之间的引用链可能很长($o \rightarrow o_1 \rightarrow o_2 \rightarrow \dots$),则在使用以上方法进行检测时,其迭代检测次数会很多,并且检测的知识本体数目也会相当多。因此,可以考虑在引用链某一长度限制内(n)检测知识本体的一致性,也就是检测从知识本体 o 出发,经过 n 个引用链接步内所到达的知识本体。因为 WWW 是一个分布性的环境,我们很难保持其在全局上的一致性,采用上面这种方法,我们可以检测和保持其在某局部的一致性。使用这种方法可以依靠工具自动地进行知识本体表示的转换和一致性的检测,因此可以在较大的领域范围内检测知识本体的一致性。

4 Web 知识本体集成

Web 知识本体通过前面所提到的知识本体引用方法可以重用已定义的知识本体,从而实现互操作性,这是 Web 知识本体的一个重要的特性。我们可以在已有的知识本体的基础上构建新的知识本体,引用这一知识本体,从而实现重用。

然而,由于 WWW 环境的分布性和开放性,知识本体是在并发的环境下开发的,这就必然会导致在已有可重用的知识本体的情况下,人们还继续重新开发新的知识本体,从而造成知识本体的冗余重复,影响知识本体之间的重用性和互操作性。另外,也可能导致这些知识本体所表示的知识之间存在互相冲突的地方。因此有时人们需要进行知识本体之间的集成,实现知识本体之间的重用,并检测解决知识本体之间的冲突。

要实现知识本体之间的集成首先需要建立知识本体之间的联系,然后在此基础上,生成一个或多个新的知识本体来包容需要集成的知识本体。

4.1 知识本体之间的关系

Web 上存在的知识本体具有分布性,知识本体之间还可能在语义上有直接或者间接的联系。这些联系可以分为以下几个方面:

1. 知识本体之间术语定义的重复性:在一个知识本体上的术语定义实际在语义上等同于另一知识本体中的术语定义。对于 SHOE 语言来说,存在两种知识本体术语定义等价性:目录等价性和关系等价性。例如:在 o_A 知识本体中有

$\langle \text{DEF-CATEGORY NAME} = \text{"pupil"} \text{ ISA} = \text{"u. people"} \rangle$ 目录定义,而在 o_B 知识本体中有 $\langle \text{DEF-CATEGORY NAME} = \text{"student"} \text{ ISA} = \text{"u. people"} \rangle$ 目录定义,这两个知识本体定义中的 pupil 目录和 student 目录在语义上保持一致,因此这两目录在语义上等价。又如:在 o_A 知识本体中有 $\langle \text{DEF-RELATION NAME} = \text{"IsChild"} \rangle$ 关系定义,而在 o_B 知识本体中有 $\langle \text{DEF-RELATION NAME} = \text{"ChildOf"} \rangle$ 关系定义,两者所定义的关系在语义上是等价的,其称之为关系等价性。

2. 知识本体之间术语定义的包容性:在一个知识本体上的术语定义所表示的概念可以包容另一知识本体的术语定义的语义范围。对于 SHOE 来说,存在两种知识本体术语定义的包容性:目录包容性和关系包容性。例如:在 o_A 知识本体中有 $\langle \text{DEF-CATEGORY NAME} = \text{"Vehicle"} \rangle$ 目录定义,而在 o_B 知识本体中 $\langle \text{DEF-CATEGORY NAME} = \text{"Car"} \rangle$ 目录定义,在概念上 o_A 知识本体中的 Vehicle 目录包容 o_B 知识本体

中 Car 目录,也可以说 Car 是 Vehicle 的一个子类,其可以称为目录包容性。又如:在 o_A 知识本体中有 $\langle \text{DEF-RELATION NAME} = \text{"IsOffspring"} \rangle$ 关系定义,而在 o_B 知识本体中有 $\langle \text{DEF-RELATION NAME} = \text{"IsChild"} \rangle$ 关系定义,在 o_A 知识本体中定义的 IsOffspring 关系定义实际上包容 o_B 知识本体定义中的 IsChild 关系。假设这两个关系定义都有且仅有两个参数定义,如果有两个实体满足 o_B 知识本体中定义的 IsChild 关系,那么它们也一定满足 o_A 知识本体中的 IsOffspring 关系,这可以看作关系包容性。

我们可以将知识本体之间这些对应的联系关系转换为一阶谓词的表示形式:对于知识本体之间的等价和包容关系,我们将其转换为一阶谓词的等价和蕴涵关系表示。在进行知识本体之间的集成时,可以利用到这些逻辑公式。在表2中,我们列出了 SHOE 知识本体之间所可能存在的对应关系,同时也列出了它们所对应的一阶谓词公式表示形式。

表2 知识本体集成转换

对应关系	等价性		包容性	
	目录等价性	关系等价性	目录包容性	关系包容性
o_A 中定义	$\langle \text{DEF-CATEGORY NAME} = A \rangle$	$\langle \text{DEF-RELATION NAME} = A \rangle$	$\langle \text{DEF-CATEGORY NAME} = A \rangle$	$\langle \text{DEF-RELATION NAME} = A \rangle$
o_B 中定义	$\langle \text{DEF-CATEGORY NAME} = B \rangle$	$\langle \text{DEF-RELATION NAME} = B \rangle$	$\langle \text{DEF-CATEGORY NAME} = B \rangle$	$\langle \text{DEF-RELATION NAME} = B \rangle$
对应关系	在 o_A 中的 A 目录定义与 o_B 中的 B 目录定义等价	在 o_A 中的 A 关系定义与 o_B 中的 B 关系定义等价	在 o_A 中的 A 目录定义包容 o_B 中的 B 目录定义	在 o_A 中的 A 关系定义包容 o_B 中的 B 关系定义
集成公理—一阶谓词逻辑	$\text{Instance}(o_A, A, x) \leftrightarrow \text{Instance}(o_B, B, x)$	$o_A. A(x_1, \dots, x_n) \leftrightarrow o_B. B(x_1, \dots, x_n)$	$\text{Instance}(o_B, B, x) \rightarrow \text{Instance}(o_A, A, x)$	$o_B. B(x_1, \dots, x_n) \rightarrow o_A. A(x_1, \dots, x_n)$

4.2 知识本体集成方式

知识本体集成是指集成利用两个或多个本无联系的知识本体,重用其表示的知识。在知识本体集成过程中,如果简单地生成一个新的知识本体,将两个或多个知识本体中的知识以一种统一的形式表示在新的知识本体中,那么其直接引用的知识本体或者被其间接引用的知识本体也需要进行调整,以表现在新的知识本体中。由于其知识本体引用链接的复杂性,很难完成这项工作。因此一般采用引用的方式实现知识本体的集成和重用,其集成方法主要分为三种,下面以 SHOE 知识本体的集成为例说明这三种集成方式:

(1) 生成新的知识本体 对于待集成的两个知识本体 o_A, o_B , 我们可以通过生成新的知识本体 o_{new} , 这一知识本体引用 o_A, o_B 两个知识本体 ($o_{new} \rightarrow o_A, o_{new} \rightarrow o_B$), 从而实现知识本体的集成和重用。具体集成过程如下:

首先通过第3节中所介绍的方法检查知识本体 o_A 和 o_B 的一致性,然后建立知识本体之间的关系,找出知识本体 o_A, o_B 之间术语定义的等价或者包容关系,找出 $o_A. V, o_B. V$ 术语定义集合中相对应的术语子集 V_A, V_B (V_A 和 V_B 中的术语有对应的等价和包容关系)。按照表2中所定义的转换关系,将这些等价或者包容关系转换为一阶谓词的表示形式,这些对应关系构成集成公理集合 A_I 。接着根据表1中所定义的映射转换关系生成知识本体 o_A, o_B 用一阶谓词表示的公理集 $o_A. A$ 和 $o_B. B$ 。接着对这些一阶谓词公式进行换名处理,在表示关系的 $R(x_1, \dots, x_n)$ 以及表示目录的实体 X 前都需要加入所属知识本体的名称作为前缀,这样可以避免命名冲突。接着检查 $o_A. AY o_B. BY A_I Y A_U$ (A_U 为集成的附加公理,见表1)的一致性,如果其是一致的,则生成一个新的知识本体 $o_{new} =$

$\langle V_A Y V_B, A_I, \{o_A, o_B\} \rangle$, 它引用知识本体 o_A, o_B, A_I 就作为新生成知识本体的公理集。在将 o_{new} 转换为 SHOE 形式时,除了对应的术语定义转换外,可以将集成公理集合 A_I 中的一阶谓词公式表示都转换为 SHOE 中的推理规则 $\langle \text{DEF-INFERENCE} \rangle$ 表示。一条表示等价关系的一阶谓词公式可以转换为两条推理规则表示(相当于正向和反向的两条蕴涵关系)。根据定义3可以知道这样建立的新知识本体是一致的。

(2) 扩展知识本体 对于待集成的两个知识本体 o_A, o_B , 我们可以通过扩展其中的一个知识本体 o_A , 形成新的知识本体 o'_A , 这一知识本体引用 $o_B (o'_A \rightarrow o_B)$, 从而实现知识本体的重用和集成。具体集成过程如下:

首先还是和集成方法1一样检测知识本体 o_A, o_B 的一致性,找出它们之间的联系,并找出 $o_A. V, o_B. V$ 中相对应的术语子集 V_A, V_B , 构成集成公理集合 A_I , 对知识本体 o_A, o_B 中的公理进行转换处理后,生成公理集 $o_A. A$ 和 $o_B. B$, 检测 $o_A. AY o_B. BY A_I Y A_U$ 的一致性。然后在知识本体 o_A 中加入对知识本体 o_B 的引用,同时在知识本体 o_A 加入 A_I 集成公理集合,形成新的知识本体 $o'_A = \langle o_A. VY V_B, o_A. AY A_I, o_A. IY \{o_B\} \rangle$, 这一知识本体通过引用 o_B 重用知识本体 o_B 的知识。最后可以将其转换为 SHOE 表示形式,这样生成的新知识本体 o'_A 也是一致的。

在集成知识本体时,通常会发现待集成的两个知识本体 o_A, o_B 会有一些重复性:两个知识本体 o_A, o_B 之间在术语定义等价的前提下,其在公理定义上具有重复性,也就是说知识本体 o_A 中的公理集 $o_A. A$ 和知识本体 o_B 中的公理集 $o_B. B$ 都可以推导出某一公理集 A_N , 这样公理集 A_N 可以看作两个知识本体 o_A, o_B 公理集的重复部分。

术语定义的重复部分和公理定义上的重复部分都可以看作是二个知识本体之间的共有部分,也是更具有普遍性的部分。这也意味着其他知识本体相对于重用知识本体 o_A, o_B 而言,更有可能重用这一更具有普遍性的部分。因此,如果将这些共有的部分集成为一个新的更有普遍性的知识本体,这样就更好地支持了在 Web 上的知识本体重用。基于这种思想,我们提出第三种知识本体集成方法。

(3)生成更普遍性的知识本体 对于待集成的二个知识本体 o_A, o_B ,我们可以生成新的 o_N ,这一知识本体为 o_A, o_B 的公共部分,同时改造知识本体 o_A, o_B ,生成新的知识本体 o'_A, o'_B ,它们都引用新生成的知识本体 $o_N(o'_A \rightarrow o_N, o'_B \rightarrow o_N)$ 。具体集成过程如下:

首先和方法1、2一样检查二个知识本体的一致性,然后找出知识本体之间术语定义上的联系,并将其中的等价关系作为术语定义之间的公共部分 V_N 。根据知识本体的映射转换,生成公理集 $o_A \cdot A$ 和 $o_B \cdot B$,检测它们的一致性。然后根据术语定义上的联系找出公理集 $o_A \cdot A$ 和 $o_B \cdot B$ 之间的公共部分 A_N 。

于是可以生成一个新的知识本体 $o_N = \langle V_N, A_N, \Phi \rangle$,在知识本体 o_A, o_B 中去除以上部分,并加入对新生成的知识本体 o_N 的引用,形成新的知识本体 o'_A, o'_B ,其中 $o'_A = \langle o_A \cdot V, o_A \cdot A - A_N, o_A \cdot IY \{o_N\} \rangle, o'_B = \langle o_B \cdot V, o_B \cdot A - A_N, o_B \cdot IY \{o_N\} \rangle$ 。根据定义2可以得出 o'_A, o'_B 和 o_N 都是一致的。

表3中表示的是对于 o_A, o_B 二个知识本体,采用以上三种不同的集成方法集成所获得的结果。由于篇幅限制, o_A, o_B 二个知识本体及其生成的新知识本体都是采用一阶谓词的形式表示的。在此基础上,我们可以将其一阶谓词逻辑表示转换为对应的 SHOE 知识本体表示。

在以上三种方法中,除了需要用户帮助定义知识本体之间的联系外,其他集成步骤都可以通过从本知识本体表示转换为一阶谓词表示的方法,依靠工具实现自动或者半自动的实现知识本体之间的一致性检测和集成。同时,这三种方法都是实现知识本体之间的两两集成的,对于多个知识本体,我们可以以知识本体两两集成为基本方式,逐步完成多个知识本体之间的集成。

表3 三种知识本体集成方法

待集成的知识本体	o_A		o_B	知识本体间联系
	$V: \text{MarriedTo}, \text{IsFather}, \text{IsMother}, \text{IsChild};$ $A: \text{MarriedTo}(x, y) \rightarrow \text{MarriedTo}(y, x)$ $\text{IsFather}(x, y) \rightarrow \text{IsChild}(y, x)$ $\text{IsMother}(x, y) \rightarrow \text{IsChild}(y, x)$		$V: \text{MarryWith}, \text{IsHusband}, \text{IsWife};$ $A: \text{MarryWith}(x, y) \rightarrow \text{MarryWith}(y, x)$ $\text{IsHusband}(x, y) \leftrightarrow \text{IsWife}(y, x)$	$o_A. \text{MarriedTo}(x, y) \leftrightarrow$ $o_B. \text{MarryWith}(x, y)$
集成方法1集成	o_A	o_B	o_{new}	
	不变	不变	$V: o_A. \text{MarriedTo}, o_B. \text{MarryWith}; I: \{o_A, o_B\};$ $A: o_A. \text{MarriedTo}(x, y) \leftrightarrow o_B. \text{MarryWith}(x, y)$	
集成方法2集成	o_B	o'_A		
	不变	$V: \text{MarriedTo}, \text{IsFather}, \text{IsMother}, \text{IsChild}, o_B. \text{MarryWith}; I: \{o_B\}$ $A: \text{MarriedTo}(x, y) \rightarrow \text{MarriedTo}(y, x), \text{IsFather}(x, y) \rightarrow \text{IsChild}(y, x),$ $\text{IsMother}(x, y) \rightarrow \text{IsChild}(y, x), \text{MarriedTo}(x, y) \leftrightarrow o_B. \text{MarryWith}(x, y)$		
集成方法3集成	o'_A		o'_B	o_n
	$V: \text{IsFather}, \text{IsMother}, \text{IsChild};$ $I: \{o_n\}$ $A: \text{IsFather}(x, y) \rightarrow \text{IsChild}(y, x)$ $\text{IsMother}(x, y) \rightarrow \text{IsChild}(y, x)$		$V: \text{IsHusband}, \text{IsWife};$ $I: \{o_n\}$ $A: \text{IsHusband}(x, y) \leftrightarrow \text{IsWife}(y, x)$	$V: \text{MarriedTo}; I: \Phi;$ $A:$ $\text{MarriedTo}(x, y) \rightarrow \text{MarriedTo}(y, x)$

5 相关工作

Wiederhold^[2]从一般意义上讨论了本体之间的复合问题,一个本体表示了某一领域内的知识,他认为不同领域之间的差别主要表现在以下几个方面:实体(相当于对象实例)以不同形式表示的差别;不同知识本体中范围(相对于类,或者是 SHOE 中的目录定义)表示方式的差别;值的单位不同而导致的编码差别;在不同实体中同一术语的含义不同而导致的上下文差别。他采用了和本文类似的建立知识本体间等价关系映射的形式来解决上面所述的几种差异,继而实现知识本体之间的交集、并集、差集等操作。在本文的知识本体集成方法中,生成的包含知识本体引用的本体可以看作是原知识本体的并集。另外,采用第3种形式所生成的更具有普遍意义的新知识本体可以看作是二个知识本体之间交集的一个子集。

Heflin 和 Hendler^[3]分析了 SHOE 知识实体在 WWW 上的动态变化的特点,同时也涉及到了知识本体的集成问题,然而其在集成时没有注意考察知识本体之间的一致性问

题。仅仅初步探讨了知识本体之间的几种集成形式。A. Farquher^[4,5]中描述了在 Ontolingua 中集成知识本体的方式。Ontolingua 是一个知识本体开发环境,它提供了一系列知识本体的编辑工具以及一些模块化和可重用的知识本体,用户可以在这一基础上构建新的知识本体。和 SHOE 不同的是,Ontolingua 支持知识本体之间的循环引用。通过复合方式构建新的知识本体时,Ontolingua 首先解决的也是命名差异性的问题,另外文章讨论了在复合时实体之间的包容性,实体限制以及实体语义的多态性问题。

Decker^[6]描述了一种领域知识表示集成方法,其将用 Frame-Logic 以及用 Chronolog 时态逻辑表示的公式都转换为一阶谓词逻辑公式,然后采用集成这些一阶谓词逻辑公式的形式来集成领域之间的知识。对于采用不同形式表示的知识本体,我们可以使用将其都转换为一阶谓词表示的方法实现知识本体之间的集成。

结论 本文提出了 Web 知识本体之间的集成方法,并采用了 SHOE 知识本体的集成作为实例来说明了这一集成方

(下转第19页)

- bra. Data & Knowledge Engineering, 2003, 44: 165~192
- 7 Clarke B L. Individuals and points. Notre Dame Journal of Formal Logic, 1985, 26(1): 61~75
- 8 Biacino L, Gerla G. Connection structures. Notre Dame J. Formal Logic, 1991, 32: 242~247
- 9 Randell D, Cui Z, Cohn A. A spatial logic based on regions and connection. In: Proc. of the 3rd Intl. Conf. on Knowledge Representation and Reasoning, Morgan Kaufmann, 1992. 165~176
- 10 Duntsch I, Wang H, McCloskey S. Relation algebras in spatial reasoning. In: E. Or lowska, A. Szatas, eds. Extended Abstracts of the 4th Seminar on Relational Methods in Algebra, Logic, and Computer Science, 1998. 63~68
- 11 Stell J G. Boolean connection algebras: A new approach to the Region-Connection Calculus. Artificial Intelligence, 2000, 122: 111~136
- 12 Duntsch I, Wang H, McCloskey S. A relation algebraic approach to the region connection calculus. Theoretical Computer Science, 2001, 255: 63~83
- 13 Gotts N M. An axiomatic approach to topology for spatial information systems; [Research Report 96. 25]. University of Leeds, School of Computer Studies, 1996
- 14 Pratt I, Lemon O. Ontologies for plane, polygonal mereotopology. Notre Dame J. Formal Logic, 1998, 38: 225~245
- 15 Li Sanjiang, Ying Mingsheng. Region Connection Calculus: Its models and composition table. Artificial Intelligence, 2003, 145: 121~146
- 16 Roy A J, Stell J G. Spatial relations between indeterminate regions. I. J. of Approximate Reasoning, 2001, 27: 205~234
- 17 王生, 刘大有, 胡鹤. 近似空间关系代数 ASRA 及应用. 中国图像图形学报, 2003(8): 946~950
- 18 McQueen M, Mann S. A Relational Algebra Approach to Problems of Spatio-temporal Representation. In: Proc. of the Spatial Information Research Centre's 10th Colloquium. 1998. 219~228
- 19 Medak D. Lifestyles - an algebraic approach to change in identity. In: Bohlen M H, Jensen C, Scholl M, eds. Spatio-temporal database management. Springer, Berlin Heidelberg New York, 1999. 19~38
- 20 Viqueira J R R. Relational algebra for Spatio-temporal data management. EDBT PhD Workshop 2000
- 21 Claramunt C. Extending Ladkin's Algebra on Non-convex Intervals towards an Algebra on Union-of Regions. In: 8th ACM Symposium on GIS, 2000. 9~14
- 22 Claramunt C, Jiang Bin. An integrated representation of spatial and temporal relationships between evolving regions. J Geograph Syst, 2001, 3: 411~428
- 23 Lee S Y, Hsu F J. 2D C-string: a new spatial knowledge. representation for image database system. Pattern Recognition, 1990, 23: 1077~1087
- 24 Lee A J T, Chiu H-P, Yu P. 3D C-string: a new spatio-temporal knowledge representation for video database systems. Pattern Recognition, 2002, 35: 2521~2537
- 25 Coenen F P, Visser P. A General Ontology for Spatial Reasoning. In: Miles, R., Moulton, M., Bramer, M., eds. Research and Development in Expert Systems XV, proc. of ES'98, Springer, London, 1998. 44~57
- 26 Stock O. Spatial and Temporal Reasoning. Kluwer Academic Publishers, 1997
- 27 Escrig M T, Toledo F. Qualitative Spatial Reasoning: Theory and Practice [M]. Ohmsha published, 1999
- 28 Cohn A G, Hazarika S M. Qualitative Spatial Representation and Reasoning: An Overview. Fundamental Informatics, 2001, 46(1-2): 1~29
- 29 Gerevini A, Renz J. Combining topological and size information for spatial reasoning. Artificial Intelligence, 2002, 137: 1~42

(上接第5页)

法。在已有的知识本体集成方法中,还没有方法注意到在集成过程中知识本体一致性的问题。本文提出的集成方法将知识本体转换为一阶谓词表示形式,利用一阶谓词逻辑来检测知识本体之间的一致性,这样保证了集成知识本体的一致性。同时这一方法也可以使用工具自动或者半自动地完成这一集成过程。

由于 WWW 具有分布性、异质性、开放性等特点,在其中会存在各种不同表示形式的知识本体。通过将多种表示形式的知识本体映射转换到一阶谓词逻辑,我们可以实现多种知识本体之间的集成。今后的工作将重点集中于如何扩展本方法,使之能支持更多表示形式的知识本体集成,以扩展本集成方法的应用范围。

参 考 文 献

- 1 Smith B, Welty C. Ontology: Towards a New Synthesis, Formal Ontology and Information Systems. New York: ACM Press, 2001
- 2 Wiederhold G. An Algebra for Ontology Composition. In: Proc. of 1994 Monterey Workshop on Formal Methods, U.S. Naval Postgraduate School, 1994
- 3 Hefin J, Hendler J. Dynamic Ontologies on the Web. In: Proc. of 17th National Conf. on Artificial Intelligence (AAAI-2000), 2000
- 4 Farquhar A, Fikes R, Rice J. Tools For Assembling Modular Ontologies in Ontolingua. In: Proc. of 14th National Conf. on Artificial Intelligence (AAAI-97), Menlo Park, CA: AAAI Press, 1997
- 5 Gruber T. Ontolingua: A mechanism to support portable ontologies, version 3.0. [Technical report]. Knowledge Systems Laboratory, Stanford University, California, 1992
- 6 Decker S. On Domain-Specific Declarative Knowledge Representation and Database Languages. In: Proc. of the 5th Knowledge Representation Meets Databases Workshop (KRDB98), 1998
- 7 Luke S, Spector L, Rager D, Hendler J. Ontology-based Web Agents. In: Proc. of the First Intl. Conf. on Autonomous Agents, New York, NY: Association of Computing Machinery, 1997
- 8 Staab S, Erdmann M, Maedche A, Decker S. An Extensible Approach for Modeling Ontologies in RDF(S). In: Workshop on Semantic Web Associated to ECDL'2000, 2000
- 9 Fensel D, Decker S, Erdmann M, Studer R. Ontobroker: Or How to Enable Intelligent Access to the WWW. In: Proc. of 11th Banff Knowledge Acquisition for Knowledge-Based System Workshop (KAW98), Banff, Canada, 1998
- 10 Blazquez M, Fernandez M, Garcia-Pinar J M, Gomez-Perez A. Building ontologies at the knowledge level using the ontology design environment. In: Proc. of the 11th Intl. Workshop on Knowledge Acquisition, Modeling and Management (KAW'98), Banff, Canada, 1998
- 11 Gruber R. A translation approach to portable ontology specifications. Knowledge Acquisition, 1993, 5(1)