

# 基于语义网的旅行代理研究

程泽凯, 许建国, 秦 锋, 刘 凯

(安徽工业大学, 安徽 马鞍山 243032)

**摘 要:**随着旅游信息的快速增长, 旅游质量已不能满足旅游者的要求。为提高旅游者的旅游质量, 文中给出一种旅行代理的方法。根据软件代理的思想提出了旅行代理, 简述旅行代理的概念和作用, 描述语义网架构和本体概念, 通过本体构建工具 Protégé 来构建旅游领域本体库, 研究语义网环境下旅行代理的意义。但是由于时间等诸多原因, 研究内容尚处于基础性的阶段, 进一步遵照标准应用本体技术, 创建实用、可靠的基于语义网的旅行代理系统, 还有待进一步工作。

**关键词:**语义网; 旅行代理; 本体

**中图分类号:** TP393

**文献标识码:** A

**文章编号:** 1673-629X(2010)11-0201-04

## Research of Travel Agent Based on Semantic Web

CHENG Ze-kai, XU Jian-guo, QIN Feng, LIU Kai

(Anhui University of Technology, Maanshan 243032, China)

**Abstract:** With the rapid growth of tourism information, the quality of tourism can not satisfy the requirements of tourists. The method of travel agent is provided to improve the quality of tourism of tourists. Put forward the travel agent based on the idea of software agent, depict the concept and the role of travel agent, describe the structure of the semantic Web and the concept of ontology, construct ontology library of the field of tourism with Protégé, study the significance of the travel agent in the environment of semantic Web. Construct the system on the standard from W3C, and try to provide new thinking about the application of ontology in travel agent. But the study is just a beginning step of the research in this field. There are a lot of problems in the model and need a further study.

**Key words:** semantic Web; travel agent; ontology

## 0 引 言

随着计算机技术和 Internet 的迅猛发展, Web 信息资源数量飞速增长, 其信息的组织呈现多种模式并存的局面。当前, 亟需在大量信息中迅速寻找到满足用户要求的信息, 而旅游是一个“集吃住行游购娱”六大要素的一个综合性产业, 一次旅游活动涉及到众多的服务设施和旅游资源, 旅游信息又有着丰富的空间和时间的内涵, 再加上旅游过程是一个受人为、自然等多种因素制约的复杂过程。随着旅游信息的几何级数的速度增长, 旅游质量已不能满足当前旅游者的要求。

针对用户访问旅游网站购物问题, 希腊赫拉斯研究与技术基金会计算机科学研究所 Grigoris Antoniou 教授与荷兰 Vrije 大学人工智能系 Frank van Harmelen 教授提出软件代理概念<sup>[1]</sup>。文中针对软件代理的思想提出一种基于语义网的旅行代理构想。

## 1 旅行代理

旅行代理是关于旅游领域能够自主地和主动地运作的软件。软件代理所做的工作是从旅游网站上选择最好的物品, 而旅行代理则可以根据用户偏好寻找合适的旅游地点、旅游线路。

语义网上的旅行代理(如图 1 所示)将根据用户给出的任务和偏好信息, 在语义网上查找相关内容, 进而分析做出选择, 最后将查找的答案反馈给用户。

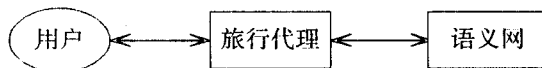


图 1 旅行代理

旅行代理不会取代语义网上的人类用户, 也不是非得由它们来做决定。在很多情况下, 旅行代理的作用是收集和整理信息, 为用户提供备选方案。

## 2 基于语义网的旅行代理

### 2.1 语义网架构

Berners-Lee 在 2000 年提出了语义网<sup>[2]</sup>架构(如图 2 所示)。

收稿日期: 2010-01-22; 修回日期: 2010-05-21

基金项目: 安徽省自然科学基金资助项目(KJ2007A051)

作者简介: 程泽凯(1975-), 男, 安徽马鞍山人, 副教授, 研究方向为人工智能、数据挖掘。

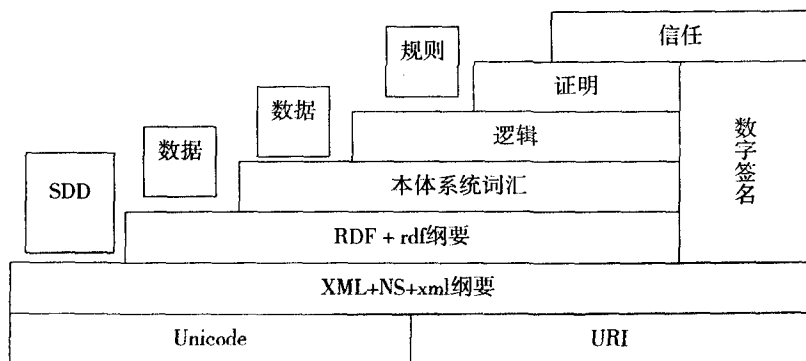


图 2 语义网架构

由图 2 可见,语义网由下至上共分七个层次以及位于图 2 右方的数字签名和位于左方的四个部分:自我文献描述 SDD(Self Description Documents)、资源描述框架 RDF (Resource Description Framework)和 rdf 纲要数据、本体系统词汇(Ontology Vocabulary)数据以及逻辑规则。

由下至上,语义网的第一层包括两部分,其中,泛编码使计算机自动表示任何语种文字成为现实,省略了过去在编码表示上不可或缺的变换选择。第一层的另一部分为统一资源标识 URI,即以统一的方式(如 <http://www.w3c.org>)标识各种各样的网络信息资源。

语义网的第二层为可扩展标识语言 XML 和与之对应的名称空间及 XML 纲要,为语义网的实现奠定最根本的语义基础。网络资源经过以上两层处理,初步生成文献描述,为随后的深层语义表述做好准备。

语义网架构的第三和第四层分别为资源描述框架和本体系统词汇。在此,rdf 纲要的功用与 XML 纲要相似,用于促成依据不同 RDF 而生成的网络信息资源描述之间的兼容。与 XML 和 RDF 相比,本体系统词汇则从更深的层次,对网络信息资源的语义进行表述和规范。

第五层的逻辑层用于进一步增强本体语言的能力,以便表达应用相关的陈述性知识。

语义网架构的第六层为证明层,证明层涉及实际的演绎过程、证明的(低层)网络语言表示以及证明的验证。

第七层为信任层,信任层将随着数字签名和其他种类知识的使用而出现,并以其他可信任的代理的推荐或其他机构以及顾客群体的排名和认证为基础。处于分层结构这座金字塔的顶端,信任是关于高层而且至关重要的概念:只有当用户信任它的操作(安全)和它所提供信息的质量时,万维网才能发挥它的全部潜力<sup>[3]</sup>。

## 2.2 本体论

### 2.2.1 本体的概念

本体的概念最早出现在哲学领域,即“客观存在的系统化的解释和说明”,用于描述事物的本质<sup>[4]</sup>。在信息科学与人工智能界,Neches 等人最早对本体给出定义,即“构成相关领域词汇的基本术语和关系,以及由这些基本术语和关系构成的解释这些词汇外延的规则<sup>[5]</sup>”。

以后越来越多的人研究本体,并给出了许多不同的定义。其中斯坦福大学知识系统实验室的 Cruber,把本体定义为“概念模型的明确的规范说明”<sup>[6]</sup>,是最著名的并被广泛引用的。Studer 等人对本体进行深入研究后,认为“本体是共享概念模型的明确的形式化规范说明<sup>[7]</sup>”。

本体概念中包括着四个主要方面:概念模型、明确、形式化、共享。

(1)概念模型:把客观世界的一些概念进行抽象而得到的模型;

(2)明确:明确定义概念以及概念之间的关系;

(3)形式化:通过数学的方式来精确描述,使得本体具有可读性;

(4)共享:本体中的知识具有客观性、共同性,是相关领域中普遍认可的概念集,它体现的是全体的共识。

### 2.2.2 本体的分类

Guarino 根据概念的概括与抽象程度或称依赖程度,把本体划分为四类<sup>[8]</sup>:

(1)顶级本体:对一些普遍性的概念以及这些概念之间的关系进行说明,独立于一个特定的问题或领域,与具体应用无关,能根据大的用户群将顶层本体一体化,至少在理论上是合理的。

(2)领域本体:对一些特定性的概念以及这些概念之间的关系进行说明,如:医药、汽车领域中的术语,要用顶级本体的术语进行约束和限定。

(3)任务本体:对一些指定性的概念以及这些概念之间的关系进行说明,如:诊断、销售,也需用顶级本体的术语进行约束和限定。

(4)应用本体(Application Ontologies):描述的概念和概念之间的关系是依赖于特定的领域及任务的,并且常常是与领域实体所起的作用相联系的。

这四类本体的层次关系如图 3 所示。

### 2.2.3 本体的构建准则

Perez 等人<sup>[9]</sup>提出利用分类法(Taxonomy)来对本体进行组织,并归纳了 5 个基本的建模元语(Modeling

Primitives):

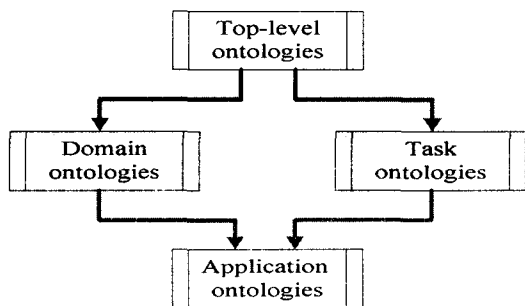


图 3 本体间的层次关系

(1)概念 C:指的是广义上的概念,既可以是一般性的概念,也可以是具体性的概念。从语义层面上来讲,可以将概念描述为相关对象的集合。通常使用框架的结构来定义概念。

(2)关系 R:指的是领域中的概念之间的联系,通常用  $n$  维笛卡儿积的子集  $R: C_1 \times C_2 \times \dots \times C_n$  来形式化定义。从语义上讲,基本的关系有 4 种,如表 1。

表 1 概念之间的基本关系及其描述

关系名	关系的描述
part-of	表达概念之间部分与整体的关系
kind-of	表达概念之间的继承关系,类似于面向对象中的父类与子类之间的关系
instance-of	表达概念的实例与概念之间的关系,类似于面向对象中的对象和类之间的关系
attribute-of	表达某个概念是另一个概念的属性

(3)函数 F:指的是一类特殊的关系。该关系的第  $n$  个元素可以通过前  $n-1$  个元素来唯一决定,将函数 F 形式化定义为  $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$ 。

(4)公理 A:代表永真断言,例如,概念之间的从属关系、概念的上下位关系等等。

(5)实例 I:从语义上讲,实例表示的是对象,对概念集合来讲,可以表示其中某个特定的概念。

### 2.3 OWL 本体描述语言

本体语言用于对领域模型进行显示的形式化描述,需要具有良定义语法、高效率的推理支持、形式语义以及充分的表达能力和表达的方便性<sup>[10]</sup>。

基于 RDF 和 RDFS 构建的 OWL,通过添加基于描述逻辑的语义原语来描述和构建各种本体模型,从而使用更多的建模原语来描述本体的性质、类以及它们之间的关系,使得 OWL 的表示能力比 XML、RDF、RDFS 更强。OWL 是以 DAML + OIL 为起点发展的,其目的是为了支持更加丰富的语义表达。

### 2.4 Ontology 在旅行代理中的应用

Ontology 是语义网中的关键技术,用来描述概念之间的语义关系。Ontology 被广泛应用的主要原因是

“它可以使人和机器之间进行有效的沟通”<sup>[11]</sup>。在语义网环境中,本体提供了对给定领域的一种共识。

旅行代理的本体思想:

(1)首先要建立旅游领域内的本体,然后把这些相对孤立的本体链接起来。

(2)存储与表示本体。用网络本体语言 OWL 对本体进行描述和储存。

(3)搜索和浏览有关旅游本体的信息。在语义网上通过旅行代理对所需的旅游本体信息进行查询,当然这种查询不是现在基于关键词的查询,而是基于语义的查询。

采用美国斯坦福大学的本体编辑工具 Protégé 来构建旅游领域本体库。图 4 为 Protégé 中建立的类,分别为旅游地点的 entertainment、room、traffictool 等。同时定义了类之间的对象属性以及类自身的数据类型属性,并定义了类之间的各种关系以及对属性的限制<sup>[12]</sup>。

利用语义网环境下的旅行代理,可以更全面更准确地查找到更多相关的旅游信息,提高旅游质量。

当游客需要查找马鞍山市内酒店时,基于关键词的查询有时是不能满足的,可能通过酒店的实例,查询到海外海皇冠假日酒店和格林豪泰酒店。由于安工大宾馆并不存在于酒店的具体实例中,此时就查询不到安工大宾馆。然而安工大内设有宾馆,同时安工大宾馆也是酒店的一个子类,基于语义的旅行代理则认为安工大宾馆也作为酒店实例,并把安工大宾馆作为查询结果返回给游客。

当统计马鞍山市内的酒店时,由于安工大宾馆和酒店并无直接实例关系,借助传统的数据库系统无法得出安工大宾馆也是酒店的一个子类。而通过语义网,可以认为安工大内宾馆是一酒店子类,自然也认为安工大宾馆属于酒店,同时安工大又位于花山区,通过旅行代理可以方便地查找到花山区内有一酒店为安工大宾馆。

同时,当游客需要查找从马鞍山火车站到旅游景点的公共交通时,在基于关键词的查询中,凡是包括“公共交通”字样的信息,都满足查询结果,这样有可能提交给游客的信息是不全面且不具体的。而通过语义网环境下的旅行代理来进行查询,只要信息语义是关于“公共交通”的,无论在信息中出现的字样是“中北巴士”、“公共汽车”,还是“出租车”或其它公共交通工具名称,都作为查询结果返回给游客。

## 3 结束语

文中从理论上提出旅行代理的概念,探讨语义网

对旅行代理的意义,下一步的工作就是开发一个实验系统,并为基于语义网的旅行代理系统提供技术支持。

文中所研究的内容依然只是基础性研究,希望在这个领域有更多专业人士参与研究、开发。

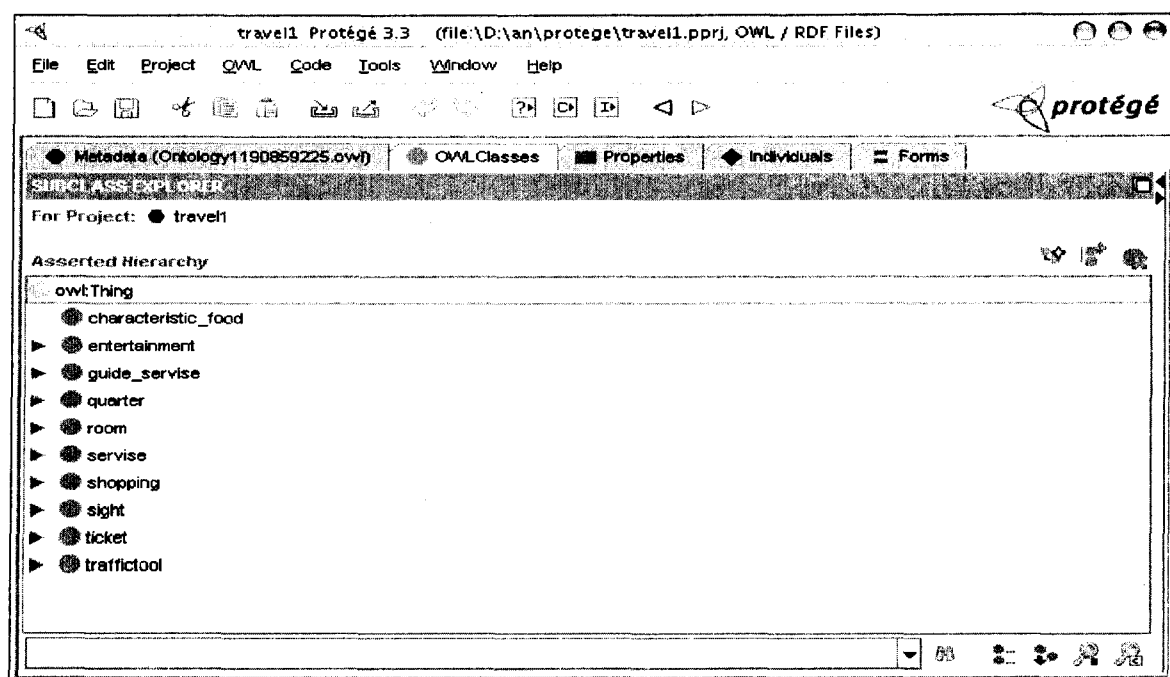


图 4 旅游领域本体库分类图

#### 参考文献:

- [1] Antoniou G, van Harmelen F. A Semantic Web Primer[M]. 北京:机械工业出版社,2008.
  - [2] Berners-Lee T, Hendler J, Lassila O. The semantic Web[J]. Scientific American, 2001, 284(5): 34-43.
  - [3] 宋 伟. 语义网简明教程[M]. 北京:高等教育出版社, 2004: 117-118.
  - [4] 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.
  - [5] Neches R, Fikes R E, Gruber T R, et al. Enabling Technology for Knowledge Sharing[J]. AI Magazine, 1991, 12(3): 36-56.
  - [6] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2): 199-200.
  - [7] Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998, 25(1-2): 161-197.
  - [8] Guarino N. Understanding, Building and Using Ontologies: A Commentary to Using Explicit Ontologies in KBS Development[J]. International Journal of Human and Computer Studies, 1997, 46: 293-310.
  - [9] Preze A G, Benjamins V R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods[C]//In: Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRRS). Stockholm: [s. n.], 1999: 1-15.
  - [10] Antoniou G, van Harmelen F. Web Ontology Language: OWL [M]//Handbook on Ontologies. [s. l.]: Springer Verlag, 2004: 67-92.
  - [11] Gruber T. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human-Computer Studies, 1995, 43(5/6): 907-928.
  - [12] 王文晶, 李 茹, 宋小香. 基于问题分析的旅游咨询系统[J]. 计算机工程, 2009, 35(12): 226-228.
- 
- (上接第 200 页)
- 雷达信号模拟器实现[J]. 北京航空航天大学学报, 2003, 29(1): 5-8.
  - [9] 张志安, 陈荷娟. 基于 PCI 总线的高速数据信号模拟器的实现[J]. 测试技术学报, 2007, 21(5): 450-454.
  - [10] 孙乐义, 王 英, 付菊英. 基于 ARM 的声学海流剖面仪信号模拟器的设计与实现[J]. 浙江理工大学学报, 2009, 26(2): 241-244.
  - [11] 张 辉, 刘 峥. 基于 CPCI 总线的通用雷达回波信号模拟器[J]. 信息与电子工程, 2007, 5(6): 418-423.
  - [12] 徐 涛, 姜本清. 基于高速 D/A 卡的复合视频信号发生器[J]. 海军航空工程学院学报, 2009, 24(2): 207-210.