

一种果品领域本体库的构建方法

刘 琼, 李宝敏

(西安工业大学 计算机科学与工程学院, 陕西 西安 710032)

摘 要:果品领域知识涉及到的学科知识多、知识量大, 如何使这些知识得到共享和重用, 目前仍是一个重要的挑战, 通过构建果品领域本体库来解决这一问题。本体是一种知识重用、知识共享和建模的重要工具。果品领域本体库是果品领域本体的集合; 果品领域本体是果品领域知识中概念、属性、过程及相互关系的形式化描述, 这种形式化描述使知识可以重用和共享, 让计算机能够理解; 为人与人之间、组织与组织及计算机之间通信提供了统一术语与概念, 使知识共享成为可能, 同时也有利于知识的获取, 可提高获取的速度和可靠性。

关键词:本体; 本体库; 本体的构建方法

中图分类号: TP302.1

文献标识码: A

文章编号: 1673-629X(2009)01-0197-03

A Construction Method of Ontology Library in Fruit Areas

LIU Qiong, LI Bao-min

(Institute of Computer Science and Engineering, Xi'an
Technological University, Xi'an 710032, China)

Abstract: The fruit areas knowledge involves much academic knowledge and a big quantity of knowledge. It is still an important challenge that how to make these knowledge share and reuse at present. To solve this problem, constructs the ontology library in fruit areas. Ontology is an important tool for knowledge reuse, knowledge sharing and modeling. Ontology library contains much ontology in fruit areas; ontology in fruit areas is the formal descriptions of the concept, the attribute, the process and the relationship. This kind of formal description may make knowledge reuse and share, so that machines can understand; provides a uniform terminology and concepts to make it possible to share knowledge between the human and the human, the organization and the organization and the computers, simultaneously is also advantageous to the acquisition of knowledge, and enhance the speed of access and the reliability.

Key words: ontology; ontology library; construction method of ontology

0 引 言

本体是一种知识重用、知识共享和建模的重要工具。目前, 本体在信息系统、自然语言处理和基于知识系统等方面得到广泛的应用。果品知识与其它领域的知识相比有以下特点: 果品知识涉及到的学科知识多, 这就加大了对果品知识的处理难度, 如何使这些知识得到共享和重用, 特别是在计算机间和计算机与人类之间通信, 使信息得到有效的共享和重用, 目前仍是一个重要的挑战; 另外, 果品领域知识量大, 但以前这些知识都是以一种自然语言的形式存在于文本载体中, 很难处理这些自由结构的知识, 使知识无法得到有效

的应用, 要开发出有效的、具有智能的、高效的系统, 离不开知识理论作为基础, 而本体正好能够帮助解决这些问题。主要介绍果品领域本体库的构建过程。

1 本 体

1.1 本体的定义

本体(Ontology)的概念源于哲学。在计算机科学中, 本体是共享概念模型的明确的形式化规范说明^[1]。这一概念包含四层含义:

(1) 概念模型: 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, 概念模型表现的含义独立于具体的环境状态;

(2) 明确: 概念和概念的约束都有明确的、无歧异的定义;

(3) 形式化: 本体能通过本体语言编码, 使计算机可读, 且可以被计算机处理;

(4) 共享: 本体体现的是共同认可的知识, 反映的

收稿日期: 2008-05-23

基金项目: 国家“星火计划”项目(2004EA850069)

作者简介: 刘 琼(1984-), 女, 陕西西安人, 硕士研究生, 研究方向为语义网、智能检索与语义网络; 李宝敏, 教授, 研究方向为计算机系统结构、计算机网络与语义网。

是相关领域内公认的概念集。

本体的目标是捕获相关领域的公有知识,提供对该领域知识的共同理解,确定该领域内共同认可的术语,并从不同层次的形式化模型上给出这些术语和术语间相互关系的明确定义,实现对领域知识的推理。

1.2 本体的建模元语

Perez 等人用分类法组织了本体,归纳出 5 个基本的建模元语:

(1)类或概念:指任何事务,如工作描述、功能、行为、策略和推理过程等;

(2)关系:在领域中概念之间的交互作用,形式上定义为 n 维笛卡儿积的子集: $R: C_1 \times C_2 \times \dots \times C_n$;

(3)函数:一类特殊的关系。在这种关系中前 $n-1$ 个元素可以唯一决定第 n 个元素,可以形式化的定义为 $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$;

(4)公理:代表永真断言;

(5)实例:代表元素。

在实际的应用中,不一定要严格按照上述五类建模元语来构造本体,同时概念之间的关系也不仅仅是上述四类基本关系,可以根据特定领域的具体情况定义相应的关系,以满足特定领域的应用要求。

1.3 本体的构建原则

最有影响力的本体构建准则是 Gruber 在 1995 年提出的五条准则^[2]:

(1)清晰性:本体应该能有效地传达其中所定义的术语的含义。此外,给出的定义应该尽可能完整。

(2)一致性:即由本体得出的推论与原有的定义是相容的,不能产生矛盾和冲突。本体中定义的公理和概念也同样要满足一致性的要求。

(3)可扩展性:本体应该能保证添加新的通用或专用术语,而不需要修改原有的定义,即能支持在已有的概念基础上定义新术语。

(4)编码偏好程度最小:概念应该在知识层次上说明,而不应依赖于特定的符号层次的编码。编码偏好应该是最小化的,因为不同的知识系统可能采用不同的表示系统或表示风格。

(5)最小本体承诺:对待建模对象给出尽可能少的约束。

1.4 本体的构建方法

Stanford 大学的 Noy 和 Hafner 给出了一种构建本体的具体方法^[3]。Noy 等人认为:本体的开发过程通常是迭代的,即先给出初步的本体框架,然后在对本体进行不断修改和精炼的过程中补充细节。本体的构建过程分为以下七步进行:(1)确定本体的领域和范围;(2)考虑重用现有本体;(3)列出本体中的重要术语;

(4)定义类和类的继承;(5)定义属性和关系;(6)定义属性的限制;(7)构建实例。

2 果品领域本体库的构建方法

2.1 果品领域本体库框架

在果品领域本体库中建立三个本体,它们分别是果树本体、网站本体和地区本体。本体库框架如图 1 所示。

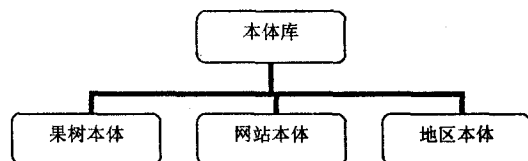


图 1 果品领域本体库框架结构

2.2 果品领域本体库的实现

2.2.1 实现工具——Protégé3.3.1

选用 Protégé3.3.1^[4]作为本体建模工具,Protégé 是斯坦福大学开发的一个开放源码的本体编辑器,是用 java 编写的,支持基于 RDF、OWL 和关系数据库格式的本体开发,其界面风格与普通 Windows 应用程序风格一致,易学易用。Protégé 提供了十分方便的可视化界面来开发本体,省却了手工书写 RDF 或 OWL 文档的麻烦,极大提高了开发本体的效率,而且提供了强大的插件,可生成相关的本体结构图、知识图以及进行知识推理等。在 Protégé3.3.1 编辑器中,本体结构以树形的层次目录显示,用户可以通过点击相应项来编写或添加类、子类、属性、实例等本体元素,另外,用户可以不用考虑具体的本体描述语言,而在概念层次上设计领域本体模型。

2.2.2 定义本体元数据

本体元数据包括本体头信息、命名空间(Namespace)、本体导入(imports)等。命名空间和本体导入为本体的分布应用提供了基础机制。

有四个命名空间对每个本体都是默认的,在 Protégé 中,会自动在每个工程中添加下面四个命名空间前缀。它们是:

`xmlns: rdf = "http://www. w3. org/1999/02/22 - rdf - syntax - ns #"`

`xmlns: xsd = "http://www. w3. org/2001/ XMLSchema #"`

`xmlns: rdfs = "http://www. w3. org/2000/01/ rdf - schema #"`

`xmlns: owl = "http://www. w3. org/2002/07/owl #"`

为网站本体建立一个命名空间:

website="http://www.website.cn#"

2.2.3 建立本体的类

在定义好本体元数据后,就可以在工程中建立本体的类了。采用 Protégé3.3.1 本体编辑器建立三个本体:果树本体、网站本体和地区本体。将果树本体分为两类:常绿果树类和落叶果树类;同时将落叶果树类又分为:仁果类、核果类、坚果类、浆果类和柿枣类;将常绿果树类可分为:柑果类、荔枝类、荚果类、草本类和藤本蔓生类。将地区本体分为七类:华北、西北、中南、西南、东北、华东和港澳台。

2.2.4 建立本体的属性

在描述属性时,将属性分为两类:

(1) 数据属性:

网站名称 (website: websiteName): 基数为 1, 定义域: 网站, 值域: string(字符串);

网站的网址 (website: webAddr): 基数为 1, 定义域: 网站, 值域: string(字符串);

网站完整描述 (website: fullDesc): 基数为 1, 定义域: 网站, 值域: string(字符串)。

(2) 对象属性:

网站相关水果 (website: relatedFruit): 基数为 ">=1", 定义域: 网站, 值域: 果树;

网站所属地区 (website: regionOf): 基数为 ">=1", 定义域: 网站, 值域: 地区。

在描述好类和属性的数据结构后,就可以编写相应的本体代码文档了,由于采用了 Protégé 作为设计工具,省却了手工逐行书写代码的麻烦。

2.2.5 定义语义关系

定义同义关系、上位关系、下位关系和属关系以提高系统的查全率和查准率^[5]。

(1) 同义关系:表示相似数据源之间对称的等价关系,即不同本体论中的两个术语有同样的语义。如桔子和橘子,山楂和山里红,君迁子和黑枣等。

(2) 上位关系:表示一个本体中术语的语义比另一本体中另一术语的语义更普通或更抽象。

(3) 下位关系:表示一个本体中术语的语义比另一本体中另一术语的语义更专业或更特殊。如水果对橘子,华东对安徽、福建、山东等。

(4) 属关系:表示一类事物属于另一类事物。

2.2.6 添加本体实例

在设计好所有类和属性后,就可以创建各种类的实例了。设计本体的类和属性有点类似于设计数据库中的表结构,要考虑各张表的字段、取值类型、关键字等,而创建类的实例则类似于往数据库中的表输入数据。一个完整的本体是由类、属性和类的实例组成的。

果树本体中,仁果类的实例为苹果、梨、枇杷、木瓜和山楂;核果类的实例为桃、杏、李、樱桃和榔;坚果类的实例为核桃、栗、银杏、阿月浑子和榛子;浆果类的实例为葡萄、草莓、醋栗、猕猴桃和树莓;柿枣类的实例为柿、君迁子(黑枣)、枣和酸枣;柑果类的实例为橘、柑、柚子、橙、柠檬、枳、黄皮和葡萄柚;荔枝类的实例为荔枝、龙眼和韶子;荚果类的实例为酸豆、角豆树、四棱豆和苜蓿等;草本类的实例为香蕉和菠萝;藤本蔓生类的实例为西番莲和南胡颓子^[4]。

地区本体中,华北的实例为北京市、天津市、河北省、山西省和内蒙古自治区;东北的实例为辽宁省、吉林省和黑龙江省;华东的实例为上海市、江苏省、浙江省、安徽省、福建省、江西省和山东省;中南的实例为河南省、湖北省、湖南省、广东省、广西省和海南省;西南的实例为重庆市、四川省、贵州省、云南省和西藏自治区;西北的实例为陕西省、甘肃省、青海省、宁夏回族自治区和新疆维吾尔自治区;港澳台特区的实例为香港特别行政区、澳门特别行政区和台湾省。

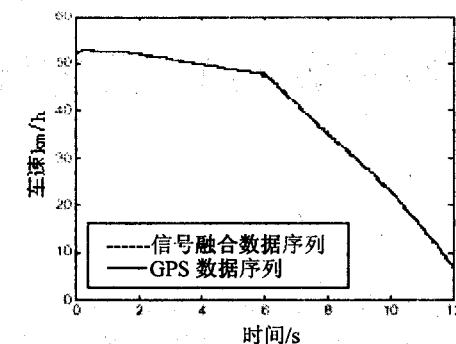
2.3 本体库的存储方式

目前对于本体的存储方法主要有三种:纯文本存储方式、数据库方式和专用管理工具方式。以文件方式存储本体轻便快捷,适合于小型的本体库。缺点在于不适合较大的本体库,因为它每次都需要读入内存操作,因此需要较多的内存,同时如果修改模型需要一次保存全部模型,尽管修改的是局部,导致效率降低。专用管理工具方式通过特殊的数据存储模型和映射机制支持本体的存储管理,但其模型难以具有普遍性和一般性,扩展性较差,不是 W3C 组织关注的发展方向。鉴于以上两种方式的缺点,采用数据库存储方式。

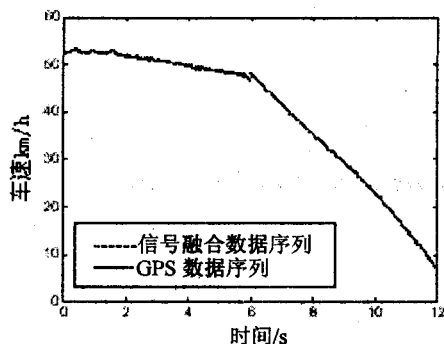
由于基于本体库的检索对本体库存在着较大流量的访问,为了提高访问性能,将本体存储于关系数据库中。这样可以利用关系数据库成熟的技术,但存在模型不匹配的问题,关系数据库采用的是相对简单的关系模型,以二维表的形式通过元组和属性来表示数据,而 OWL 本体的描述虽然也具有类和属性,但同时还有很多基于语义的描述,其模型是 N 维的。如何把 N 维模型变为二维模型,这是接下来要研究的内容。

基于数据库的本体库存储方式需要配置一个数据库连接,然后从这个连接创建新的本体库或者打开已有的本体库。配置数据库连接的参数包括数据库 URL,用户名,密码,数据库类型。数据库类型是一个字符串,用来区分是哪种类型的数据库,如“MySQL”代表 MySQL 数据库。将本体库存储到数据库中后,就可以在应用程序中使用本体库的数据了。

(下转第 203 页)



(a) 滤波加速度计信号



(b) 未滤波加速度计信号

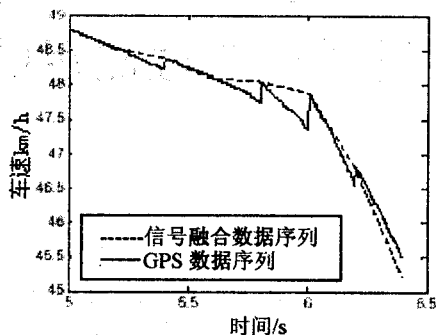
图3 滤波前后车辆对地速度曲线比较

- 11. <http://www.novatel.com/Documents/Manuals/om-20000086.pdf>.

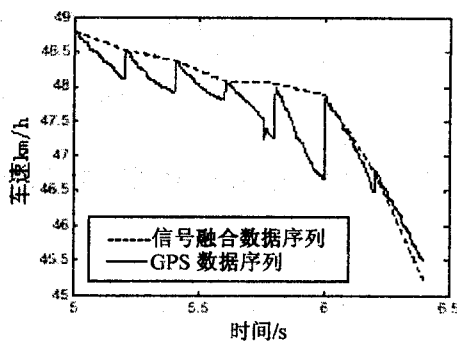
- [3] Sensor Products MMA7260Q XYZ Three-Axis Low g Acceleration Sensor [EB/OL]. 2005 - 06. <http://www.freescale.com/files/sensors/doc/fact-sheet/MMA7260QFS.pdf>.

- [4] 董绪荣, 张守信, 华仲春. GPS/INS 组合导航定位及其应用 [M]. 长沙: 国防科技大学出版社, 1998.

- [5] 陆晶晶, 张为公, 周耀群. 基于 GPS 和加速度计的车速测



(b) 滤波加速度计信号



(c) 未滤波加速度计信号

图4 滤波前后车辆对地速度局部放大曲线比较

量系统的开发与研制[J]. 测控技术, 2007, 26(8): 15 - 18.

- [6] Cygnal Integrated Products, Inc. C8051F 单片机应用解析 [M]. 北京: 北京航空航天大学出版社, 2002.

- [7] 付梦印, 邓志红, 张继伟. Kalman 滤波理论及其在导航系统中的应用 [M]. 北京: 科学出版社, 2003.

- [8] Zhou H R, Kumar K S P. A current statistical model and adaptive algorithm for estimating maneuvering targets [J]. Journal of Guidance, Control and Dynamics, 1984, 7(5): 596 - 602.

- [9] 赵力. 语音信号处理 [M]. 北京: 机械工业出版社, 2003.

(上接第 199 页)

3 结束语

从本体论的角度出发探讨了果品领域本体的建立过程。随着对果品领域本体的进一步研究, 许多问题和目标也将呈现出来, 总的来说主要有以下几点:

(1) 目前农业本体库主要是由手工开发。半自动获取虽然取得了一些进步, 但仍然有很大的局限。自动获取, 包括基于本体的自动获取还处于起步阶段。

(2) 需要建立一套完整的本体评估标准, 主要是保证知识的清晰性、一致性、完整性、可扩展性, 使知识得到有效的共享和重用。

(3) 果品领域本体通过对类和类之间关系的严格定义来确定类的精确含义, 表示果品领域中共同认可的、可共享的知识。

在语义网中, 是解决语义层次上 Web 信息共享和交换的基础, 具有非常重要地位。因此, 如何在果品本

体基础上开发 Web 本体是果品语义网研究的任务。

参考文献:

- [1] Studer, Benjamins V R, Fensel D. Knowledge engineering: Principles and methods [M] // Data and Knowledge Engineering. [s.l.]: [s.n.], 1998: 161 - 197.
- [2] Gruber T R. Towards principles for the design of ontologies used for knowledge sharing [J]. International Journal of Human and Computer Studies, 1995, 43: 907 - 928.
- [3] Noy N F, McGuinness D L. Ontology development: a guide to increasing your first ontology [R]. USA: Stanford University, 2001.
- [4] Musen M, Noy N, O'Connor M, et al. Stanford Center for Biomedical Informatics Research [EB/OL]. 2008. <http://protege.stanford.edu/>.
- [5] 冯志勇, 李文杰, 李晓红, 等. 本体论工程及其应用 [M]. 北京: 清华大学出版社, 2007: 88 - 89.