

基于本体论的农业知识的 OWL 描述

杨保明, 刘晓东, 姚 兰, 赵飞蓉

(西安交通大学 计算机系, 陕西 西安 710049)

摘 要: 基于本体的信息交换特性, 采用知识本体来表示农学领域知识。首先开发出农业知识本体; 然后建立概念、概念之间的继承、实例, 以及多种约束关系来表达高度共享和重用的农业领域知识; 最后针对虚拟农作物生长系统开发平台中知识引擎对作物生长模型脚本的要求, 用 OWL 描述了农业知识本体, 通过实际应用, 证明了该方法是可行且有效的。

关键词: 本体论; OWL; 知识表示; XML; 脚本

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2007)05-0058-03

The OWL Description of Ontology - Based Agriculture Domain Knowledge

YANG Bao-ming, LIU Xiao-dong, YAO Lan, ZHAO Fei-rong

(Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Based on the information exchange through ontology, this paper expresses agriculture domain knowledge with the ontology. Firstly, agriculture knowledge ontology has been proposed. Then concepts in agriculture domain and relations among these concepts, such as inheritance relationship, instance relationship and restrictions, are built by the Agriculture Ontology to represent highly sharable and reusable domain knowledge. Finally, in demand for the crop growth model script to the knowledge-engine in the platform, this article has also described the AO with OWL, which is proved to be feasible and effective by application.

Key words: ontology; OWL; knowledge representation; XML; script

1 引言

本体模型 (Ontology) 是“一个对有共识的、已经概念化的事物规范的、明确的定义”即把现实世界中的某个领域抽象成一组概念 (如对象、属性、进程等) 及概念之间的关系。OWL 的开发目标不是针对人来描述 Web 信息内容、而是为了解决面向应用程序和环境之间的语义互操作的描述语言, 它通过本体来显式地表达词汇、术语的意义以及它们相互之间的关系。在表达概念的语义灵活性方面, OWL 比过去的 XML、RDF、RDF-S 等语言都要强; OWL 采用面向对象的方式来描述领域知识, 即通过类和属性来描述对象, 并通过公理 (Axioms) 来描述这些类和属性的特征和关系^[1,2]。

文中研究了农业领域本体在虚拟农作物生长系统开发平台中的具体应用。知识本体能够以一种显式、形式化的方式来表示语义, 提高异构系统之间的互操作性, 促进知识共享。农业知识本体 (A-

griculture Ontology) 是农业领域中信息语义的基本单位, 它在本质上就是对农业领域信息资源进行分类与描述的概念化体系。AO 是虚拟农作物生长系统开发平台输入脚本的重要来源。

2 农业知识本体的开发

虚拟作物生长系统开发平台是基于生长模型构建虚拟作物的, 其所需的知识来源于农学领域的实践, 而农业知识本体是对农业领域信息资源的科学分类与描述的概念化表示, 在本质上就是对所有入库的农业领域信息资源进行分类与描述的概念体系。因此, 平台可以认为是知识本体驱动的一个应用。

从技术实现的角度来分析, 知识本体构造经历了概念抽取、OWL 描述、XML 序列化这样 3 个阶段。知识本体驱动的应用程序使用知识本体的过程与此相反, 需要经历 XML 解析, 然后向应用程序提供应用编程接口。Ontology 的定义目前还没有统一的说法。但大多数研究者认为“Ontology 是关于可共

收稿日期: 2006-06-12

基金项目: 国家“863”计划项目 (2003AA209021)

享概念化体系的明确的形式化规格说明”。Ontology 主要由概念和关系组成。概念表示领域中的一组或一类实体。关系描述概念之间或概念的属性之间的互相作用,如 IS_A、IS_PART_OF 等。关系可以带属性,如关系是否可选、是否有传递性等。Ontology 还可以包含函数、实例、公理等组成部分。下面对 Ontology 作形式化定义^[3]:

定义 1 不继承任何概念的概念为初始概念 RootConcept, 在特定的 Ontology 中初始概念是惟一的。

定义 2 Ontology 中概念和概念之间的关系组成 Ontology 继承关系图 $G(V_{concept}, Einberit-relatic)$, 其中 $V_{concept}$ 是 Ontology 中所有概念的节点集合, Einberit-relatic 对应于 Ontology 中继承关系的有向边集合, 有向边 $\langle V_{concept_i}, V_{concept_j} \rangle$ 表示概念 $concept_i$ 为概念 $concept_j$ 父节点。

定义 3 继承关系图 G 中, 节点 $V_{concept_i}$ 到 $V_{concept_j}$ 经过的有向边的序列为 $V_{concept_i}$ 的一种继承关系路径, 路径长度 $L_{rootconcept, concept_i}^k$ 为第 k 条路径上的边数, m 表示有 m 个继承关系路径, 则初始概念 RootConcept 到概念 $Concept_i$ 的最大路径长度为:

$$L_{rootconcept, concept_i} = \max_{k=1, \dots, m} L_{rootconcept, concept_i}^k$$

2.1 AO 子集-玉米栽培知识本体

从玉米的栽培技术中提炼出了一些核心概念、实例及其属性相关关系, 建立了一个有关玉米栽培的验证性知识本体。

步骤 1: 本体捕获。抽取出玉米栽培技术中的基本概念和术语的识别, 然后给出概念和属性之间的关系, 这些概念和关系为本体建立提供基本的结构和框架。

步骤 2: 确定概念的属性及基本的约束关系。在确定了关系之后, 才能把先前分散的概念联系在一起, 形成一个有机的整体。

(1) 玉米栽培知识本体中主要的属性:

- 玉米(property: 品种、根、茎、叶、生长时间、需肥量、需水量...);
- 根(property: 形状、半径、长度、生长量...);
- 肥料(property: 钾含量、磷含量、施肥密度、吸收情况...);
- ...

(2) 玉米栽培知识本体中约束关系:

- 播种(domain: 种子, range: 玉米): 先决条件

约束关系;

- 生长(domain: 播种, range: 玉米): 时间约束关系;
- 危害(domain: 害虫, range: 农作物): 互为条件约束关系;
- ...

步骤 3: 对 AO 进行形式化。根据核心概念及相关关系, 用下面的六元素方法, 将概念和关系一一对应, 可得出 AO 形式化的表达。

Yo- Myo Naing 等人提出采用六元素方法来描述知识本体系统, 六元素方法比一般的有向图方法更能清晰地表达出概念之间的关系, 并且很容易对应到 owl-rule 的表达能力上, 是一种强有力的形式化表达本体的方法。

2.2 玉米栽培知识本体

为了继续用 RDF 或者 OWL 等本体语言描述本体, 方便计算机对语义的抽取, 下面给出一个玉米生长的农业知识中的一个简单的 Ontology 片断, 如表 1 所示。

第一部分是概念的层次定义, 是用基本表达式 $C1::C2$ 声明概念之间的子类关系, 农作物 corn 是 crop 的子类, water 是 Object 的子类, 等等。

第二部分是概念的属性定义。农作物 corn 定义了种类 species、种子 seed、根 root、茎 stem、生长时间 growstage 5 个属性。并规定第一个属性的值是原子类型 String, 种子、根、茎的值必须是 org 类的实例, 属性 growstage 的值是 stage 类的实例, 即通过属性建立概念之间的关联。而 org 类和 stage 类等也都有它们自己的属性定义。

第三部分包括规则或公理定义, 依据这些规则可以推理出一些分散的、隐含的知识, 对原有信息加以补充和完善。

表 1 中的 Ontology 简单定义了两条规则, 用推导符 “->” 连接, 其含义是, 如果 corn1 是玉米, 则它应该具备玉米所定义的各个部分。

3 农业知识本体的 OWL 描述

W3 组织在总结 RDF、RDFS 等其他本体语言的基础上提出了 OWL, 不过 OWL 也不是完全摆脱 RDFS, 而是建立在 RDFS 之上, 利用 RDFS 提供的许多便利机制, 并加入更强大的语义表述功能机制对 XML 数据的语义信息加以描述^[4,5]。

OWL 继承了 RDFS 中的包括 Rdfs: Class, Rdfs: subclass, Rdfs: Property, Rdfs: subProperty,

表 1 ontology 片断

概念层次定义	属性定义
Object[crop::Object corn:: crop. org::Object stage::Object. water::Object. fertilizer::Object. ...	corn[species ==>>String; seed=>>org; root=>>org; stem=>>org; growstage=>>stage]. org[name=>>String; shape=>>String]. Stem[long ==> Millimeter; width ==> Millimeter; growL=>> Millimeter; growW=>> Millimeter]. Stage[name=>>String; time=>>Day]. Water[amount=>>float]. Fertilizer[Namount=>>Integer; Pamount=>>Integer; Kamount=>>Integer]. ...
推理规则定义	
1.FOR ALL corn1 corn1:corn[species ->>白糯二号] --> corn1:corn[[seed- >>[shape- >>扁圆]] AND[root- >> [shape- >>须根]]AND [stem- >>[shape- >>杆形, long- >> 200, width- >> 5]]. 2. FOR ALL corn1 corn1:corn [species ->>白糯二号 AND growstage ->>发芽期] --> corn1:corn [stem ->>[growL ->>water*0.0005] ANDstem- >>[growW- >>water*0.0001]]. 3. ...	

Rdfs：domain, Rdfs：Range 等在内的一系列建模原语。并在此基础上加入了一系列的建模原语使得能够对类、特性，以及它们之间的关系做更加细致的定义。

OWL 语言本身设计用于网络环境，按照 XML 语言格式实现。它遵循面向对象的思想，按类和属性的形式描述领域知识所包含的结构，即一个本体由类和属性构成。相对于 XML、RDF 和 RDFS, OWL 拥有更多的机制来表达语义，从而超越了 XML 、RDF 和 RDFS 仅能够表达网上机器可读的文档内容的能力。下述例子就是用 OWL 表示的公理“玉米是产量很高的农作物”。

```
<owl : Class rdf : ID = "Corn"> "  
<rdfs : label > Corn< / rdfs : label >
```

```
<rdfs : subClassOf rdfs : resource = "# Crop " />  
<rdfs : subClassOf  
<owl : Restriction >  
<owl : onProperty rdf : resource = "# yield " />  
<owl : hasClass rdf : resource = "# high " />  
</owl : Restriction >  
</rdfs : subClassOf >  
</owl : Class >
```

根据应用的需求, OWL 实际上包括 3 种复杂程度递减的本体语言:

- OWL FULL：包括所有的 OWL 词汇和 RDFS 提供的原语，能够提供最大程度的知识描述能力,但是由于过于复杂,且还不成熟,因此还在不断的更新中。
- OWL DL：提供大部分 OWL 词汇支持和 RDFS 支持，并在语义上等同于描述逻辑 DL (Description Logics)。
- OWL Lite：是 OWL DL 中相对容易实现部分的子集合,只提供层次分类和简单的约束功能。

基于 OWL DL 的特点, 以及它与描述逻辑之间的等同性, 将主要使用 OWL DL 来进行研究。由于采用 OWL 语言描述的本体比较庞大，所以就针对上一节中玉米知识本体的片段分别从概念实体、概念属性及概念间关系 3 个方面来详细说明, 三者分别与 OWL 语言中的类 (class)、属性 (datatype-property、objectProperty) 和公理(axioms) 对应。

上述描述中, 使用类公理(subclassof)描述了两个类(概念) 之间的继承关系。在描述类属性时, 使用了数据属性(datatypeproperty) 和关系属性(objectproperty)，前者描述了类的某个属性, 而后者在具备前者功能的同时, 也表示了两个类之间的某种关系。另外，使用属性公理 domain 和 range 表示属性的应用领域和属性的取值范围。

由上可知，基于本体的信息建模所包含的概念、关系和公理数量都较庞大且结构复杂。因此本体的完备性有赖于其构造前领域专家对领域知识的分析和抽取, 对该领域越了解, 所建的本体就越完善。

虚拟作物平台在得到以上输入后, 通过内置的知识引擎就可实现作物形态的矢量化表征, 建立作物形态可视化的数学模型和作物形态显示模型, 再结合非规则几何形体的真实感显示技术就可真实再现作物的生长发育过程。

(下转第 65 页)

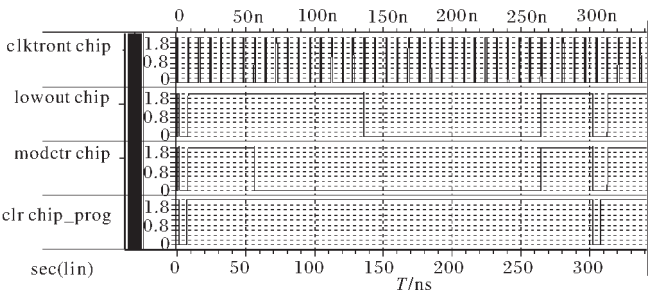


图5 可编程分频器瞬态分析波形($N_s=6$)

表 2 芯片功耗测试结果

电源电压(1.8V)	仿真结果	测试结果
电流	6.8mA	7.5mA
功耗	12.25mW	13.4mW

划、基于时序的布局和参数提取等前后端设计,设计并实现了 PLL 频率综合器中的可编程分频器。芯片测试结果表明,通过改变芯片控制逻辑,分频器能精确地完成与控制端口相匹配的分频功能,从而成功实现了应用于 802.11a 频率综合方案,达到了设计要求。

参考文献:

[1] Yu X Y, Do M A, Jia L, et al. Design of a low power wide -band high resolution programmable frequency divider [J]. IEEE Transactions on VLSI Systems, 2005, 13 (9): 1098~1103

(上接第 60 页)

4 结束语

AO 是农学领域知识语义形式化的表示,具有表述的精确性和很强的推理能力,以此为基础的生长模型脚本,符合自然界的客观规律,是虚拟作物生长平台开发的重要基础。文中以玉米栽培知识本体为例,提出了一种一般农业本体的建立方法,并在最后将本体以 OWL 描述,方便开发平台使用以此为基础的生长模型脚本。在实际应用中也得到了良好的效果。理论上讲,对于其他领域的知识本体,只要平台的组件库支持,都可以开发出相应的模型。

参考文献

[1] Web Ontology Language (OWL) Guide Version 1. 0,W3C Working Draft [EB/ OL]. <http://www.w3.org/TR/2002/WD-owl-guide,2002-11-04>

[2] 黄浏祥, 刘晓东, 朱林, 等. 基于组件技术的虚拟作物生

[2] Ram Singh Rana. Dual - modulus 127/128 FOM enhanced prescaler design in 0.35 μ m CMOS technology [J]. IEEE Journal of Solid State Circuits, 2005, 40(8): 1662~1670

[3] Hung C M, K K O. A fully integrated 1.5- V 5.5- GHz CMOS phase- locked loop [J]. IEEE J. Solid- State Circuits, 2002, 37(4): 521~525

[4] Tournier E, M Si é Graffenil J. High- speed dual - modulus prescaler architecture for programmable digital frequency dividers [J]. IEEE Electronics Letters, 22nd November 2001, 37(24): 1433~1434

[5] Cicero S Vaucher, Igor Ferencic, et al. A Family of low- power truly modular programmable Dividers in Standard 0.35 μ m CMOS Technology [J]. IEEE Journal of Solid- State Circuits, 2000, 35(7): 1039~1045

[6] David A Hodgesl. Analysis and design of digital integrated circuits in deep submicron technology. 北京: 电子工业出版社, 2005

[7] 时昕, 王东辉, 侯朝煊. 深亚微米 SoC 中的电源/地网络设计[J]. 微电子学与计算机, 2004, 21(12): 198~202

[8] Jan M Rabaey, Anantha Chandrakasan. Digital integrated circuits: a design perspective (Second Edition)[M]. 北京: 清华大学出版社, 2004

作者简介:

何小虎 男, (1981-), 硕士研究生。研究方向为 VLSI 设计。
胡庆生 女, (1964-), 副教授。研究方向为数字系统设计。

长系统开发平台的设计 [J]. 微电子学与计算. 2005, 22 (1): 13~16

[3] 张维明. 语义信息模型及应用 [M]. 北京: 电子工业出版社, 2002: 77~79

[4] Language Ontology Web Overview, W3C Working Draft [EB/ OL]. <http://www.w3.org/TR/2003/WD-owl-features,2003-3-31>

[5] 王颖, 张丽霞, 刘晓东, 等. 关联规则挖掘技术在 Web 预取中的应用[J]. 微电子学与计算, 2005, 22(4): 166~169

作者简介:

杨保明 (1977-), 研究生。研究方向为人工智能和虚拟现实技术应用。
刘晓东 (1954-), 副教授。研究方向为人工智能与虚拟现实技术的研究与应用。
姚 兰 (1982-), 研究生。研究方向为人工智能和虚拟现实技术应用。
赵飞蓉 (1982-), 研究生。研究方向为人工智能和虚拟现实技术应用。