# 林产品本体的构建方法研究

杨　抒[1,2],武　刚[1],王　欢[1],彭恩强[1]
(1.北京林业大学 信息学院,北京 100083;2.新疆农业大学 计算机与信息工程学院,新疆 乌鲁木齐 830001)

**摘要**:针对林产品商务 Web 信息整合的过程出现的林产品语义冲突、共享困难等问题,从现有的本体相关理论出发,使用本体的形式化定义形式,对林产品的概念进行语义形式化描述,确定林产品本体的六元组。进而建立了林产品的核心本体。并给出该本体的关联可视图。

**关键词**:Web 信息整合;林产品;领域本体;概念语义
**中图分类号**:S126;TP3　　**文献标识码**:A　　**文章编号**:1002-2767(2010)07-0147-05

将海量的 Web 数据库中的信息进行整合的核心问题是要解决不同数据源数据的异构问题。由于分布式的 Web 系统采用不同的知识表示方法来解决相同领域内的问题,因此造成了从一个基本概念衍生出了多种不同的语义。其主要表现为缺乏可共享的理解。近年来,本体(Ontology)的概念引入了 Web 信息整合领域,研究者通过本体用来解决知识表示、知识组织等问题[1-3]。

如何构建林产品领域本体是解决林产品领域内语义异构的研究课题。也是面向林产品商务信息 Web 信息整合需要解决的核心问题。林产品本体的构建也有助于解决相关知识领域内的推理等问题。

## 1　林产品本体的应用背景分析

国内林产品商务信息数据的质量是业内用户非常关心的问题[4]。作为信息的载体,用本体描述相关数据是解决该问题的方法之一。用本体对林产品以及相关信息进行形式化说明,首先使纷繁复杂的林产品的名称得以规范、避免“二义性”的发生,提高数据的质量,从而提高用户获得林业相关产品、服务、技术等信息的效率;其次可通过本体本身所具有的语义知识关系,挖掘出和林产品与其相关的生产企业中最具有价值的林产品商业信息;最后,对可对提供林产品相关服务、技术的企事业单位之间的相关知识进行关联。通过本体对数据进行描述以及对知识进行关联,提高用户获得林业相关产品、服务、技术等信息的效率,图 1 是基于本体的 Web 信息服务系统的框架,可以看出本体在其中的作用,其主要表现在 4 个方面,即(1)规范、并且形式化领域知识;(2)明确领域知识逻辑结构和业务流程;(3)在计算机之间共享对于形式化信息的理解;(4)进行领域信息的知识推理。

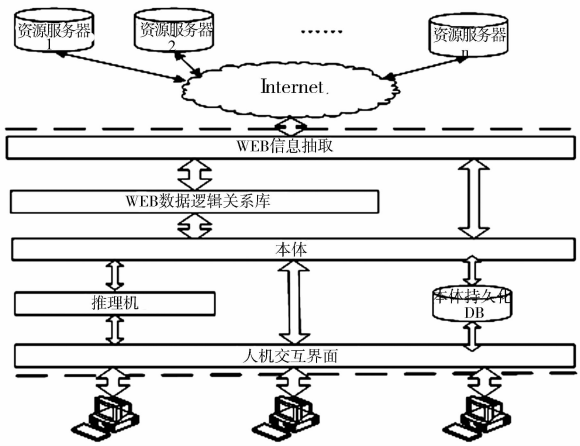


图 1　基于本体的 Web 信息系统结构示意图

该系统按照本体的领域知识结构,进行知识挖掘、关联。从而展现知识结构化。围绕这个目的所建立基于本体的林产品商务 Web 信息整合系统具有语义存取结构和知识推理功能,在实现这个系统之前需要对林产品以及相关商务信息做清晰的语义、概念化的形式描述,构建相关领域知识整合系统需要以领域本体为基础。

## 2　林产品领域本体核心概念集的建立

### 2.1　林产品领域本体、概念本体和属性本体的定义

在引进本体概念构建林产品领域本体进行

**收稿日期**:2010-04-08
**第一作者简介**:杨抒(1979-),男,新疆维吾尔族自治区乌鲁木齐市人,在读博士,讲师,从事林业信息化研究。E-mail:sleepingys@sohu.com。
**通讯作者**:武刚(1946-),男,北京市人,博士,教授,博士生导师,从事机器学习与森林经营管理。

Web 信息整合的工作中将林产品领域本体分为两个层次。

描述概念，对林产品领域的相关概念进行精确化和形式化；建立属性关系，刻画概念本体之间关系属的属性本体。

对林产品本体进行形式定义

定义 1：$O=\{C,P,A,H^c,Prop,att\}$

林产品本体是一个六元组[4]。其中，C 表示林产品本体中所涉及概念的集合；P 表示所有关系的集合；A 表示所有属性集合；$H^c$ 表示所有概念之间的层次联系，其中，$H_c(C_1,C_2)$ 表示 $C_2$ 和 $C_1$ 是上下位的关系，$Prop(p)=(C_1,C_2)$ 表示 $C_1$ 和 $C_2$ 概念之间存在 P 联系；函数 $att:A\rightarrow C$ 将概念与相关解释对应起来。在林产品领域引进了本体的概念，能够通过概念集的明确、关系集的确立、属性刻画和约束，更好地描述林产品领域的数据这一概念。

## 2.2 林产品领域本体构建步骤

在构建林产品领域本体时，采用了自顶向下迭代的构建方法。

迭代渐进的原则：首先，根据应用需求的需要设计一个最小的核心本体；然后由用户需求不断地提出，逐步地修正定义林产品概念并完善他的属性刻画。在本体评价的过程中发现核心本体存在的问题并加以改进，从而形成以用户为中心且达到应用需求较完善的本体。

自上而下的方法：采用骨架法设计本体，首先需要领域专家给出林产品领域内的顶层概念，其次，在根据定义好的骨架本体的范围，从林产品领域内相关专业词典中抽取关键词，同时为概念的分面(facet)进行分面分析。最后，在细化概念的各种语义关系的过程中分层进行构建各层子类，得到本体的概念集合。在对本体输出形式化描述可采用本体构建软件工具完成。

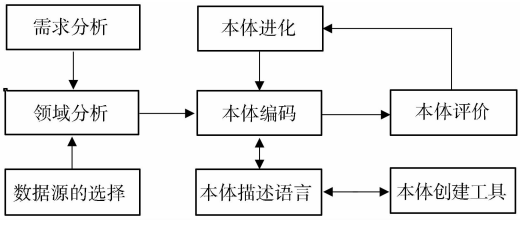


图 2 本体设计流程

2.2.1 林产品概念的获取 相关领域本体的构建首先要获得林产品本体的重要概念、关键概念；其次确定上位概念、较为显著的概念和常用概念，以建立核心概念集[5-6]最后在此基础上进行扩展。领域概念的获取要以应用为中心同时坚持用户保障原则、文献保障原则。领域专家给出的骨架本体也为概念的进一步获取圈定了范围[7]，概念获取的途径有 2 个：专业词典：从《中国林业产业与林产品年鉴》《国民经济行业》和《全国主要产品分类与代码》等中抽取林产品相关部分概念；关键词记录：抽取提供林产品商务服务相关 Web 页面的关键词，保留频次大于 2 的词[8]。

该研究主要从 Web 信息整合系统需求为基础对林产品及相关商务信息进行语义分析并确立了本体间的属性关系。对专业词典 Web 数据库提供信息经过语义分析，得到了林产品顶层类下 18 个相关概念作为该本体的一级核心概念[9]。木材生产、人造板、制浆造纸、木制品制造、竹、藤生产与加工、木本粮食与油料生产、水果生产、林木种苗生产、林产化学加工、森林蔬菜、饮料、饲料、花卉、驯化野生动物及其产品加工、药材生产、森林旅游、森林狩猎、林业机械制造、林业产业国际合作、林业投资。

2.2.2 概念层次的确定 对林产品领域本体的概念层次确定是从林产品领域的基本概念出发，逐步细化进而逐层确定各层子类的概念。

在林产品本体中，根据当前我国林产品生产及企业或个人对林产品的供应、需求等情况，进行类的确定及划分。以我国林产品分类为例，既要考虑到所建的类要全面覆盖整个林业生产领域内的概念，也要考虑对普通互联网用户对林产品的认知程度。以此为前提参考了《林产品年鉴》《国民经济行业》等诸多国家级权威分类法和“苗木网”“阿里巴巴”等业内用户经常使用的电子商务网站。在对我国林业行业产品进行分类、再分类的过程中要注意同层的类与类之间要相互独立，互不交叉。

在该研究中，根据业务分析和领域专家确定，将顶层类确定为“社会实体”类、“林产品”类、“市场信息”类和“地理位置”类。然后分别在顶层类下面建立相应的子类。如一个林业相关的产品可能是企业提供的技术或生产的实体产品，或者政府、事业单位提供的与林产品相关的服务和技术等。所以又可将顶层概念林产品的子概念可分为“实体产品”“技术产品”“服务”子类。

如上所述，社会实体的顶层类构建思路与林产品本体构建相似。对建立好的社会实体类再建

立各自的子类,然后逐步细化。

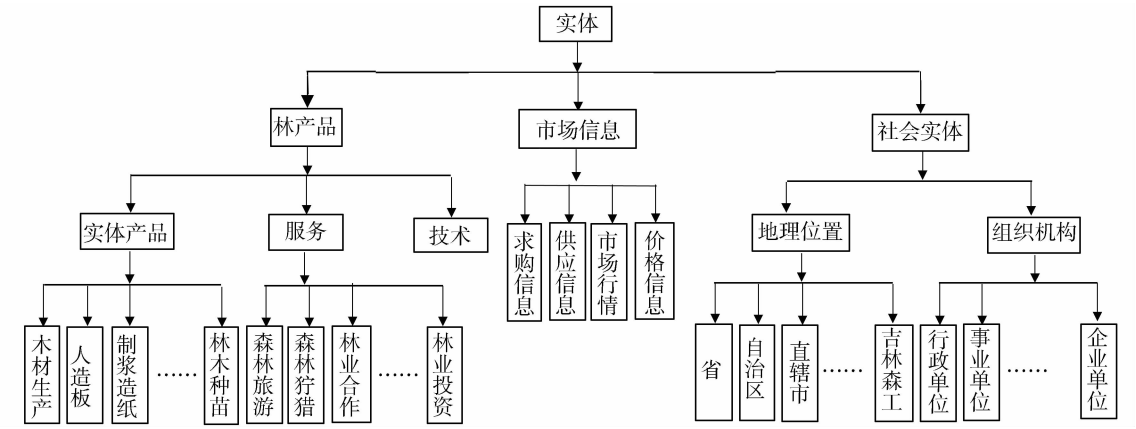


图 3　林产品的类和林业行业机构类的层次结构图

2.2.3　概念语义关系的确立　概念间的语义关系是领域本体的重要组成部分,也是进行内容挖掘和语义推理的基础。其中包括同义关系、上下位关系、实例关系、包含关系这样的等级关系,以及表示概念,需要专家确定的非等级关系。由定义 1 可知林产品领域本体形式化定义中的 P 是领域本体概念和概念间的关系集合。在林产品领域本体中顶层类的语义关系可向下层分支类进行映射。

(1)同义关系:由于领域专家和业内用户或用户间描述产品的角度各有差异,对同一概念的林产品可能有不同的表述。如人们提到板材的“材质”和板材的“材料”都是描述相同的产品属性。因此确认概念的语义关系可以扩大林产品领域本体的外延。

(2)上下位关系:一般指的是领域本体中上层本体与子本体之间的关系,属于本体概念的层次关系。由定义 1 可知,其中 $H^c \subseteq C \times C$ 表达了概念之间的层次联系。就具体实例而言,人造板类是实体产品类的子本体,它们之间是继承关系。这两个本体之间是直接的父子关系。上下位关系也可以定义为间接的语义关系。

(3)包含关系:指的是一个本体是另外一个本体的属性,由定义 1 可知,其中 $H_c(C_1, C_2)$ 表达了 $C_1$ 是 $C_2$ 的子概念这样的包含关系。

2.2.4　添加概念属性　类的概念层次结构主要用于描述领域知识的框架,在确定了类的概念和类的概念层次结构后,还必须描述概念间的内在结构来丰富领域知识,即类的属性。类的属性也可从顶层类向下层分支类进行映射,因此属性应该被定义在拥有该属性的最大的类上。类的属性包括对象属性和数据属性。

以面向林产品市场关注人群同时基于 Web 数据库的林产品商务信息知识需求类型进行属性设置为例。对目前互联网用户中对林产品商务信息关注较多的信息进行分类后,将市场信息分为“供应信息”“求购信息”“市场行情”“价格信息” 4 种类型知识,因此在林产品本体顶层类“市场信息”中设置了这 4 个相关属性。同时为使得市场信息本体的属性覆盖面广、针对性强,需要从用户需求角度考虑与上述 4 类知识的延伸属性。使之能够满足不同用户对市场信息不同类型的需求[10]。

2.2.5　属性约束分析　属性的约束条件和属性可以继承于父类,也可以重新为子类添加新的属性约束。一般情况下属性约束分析只能通过领域专家进行确定。以原木本体为例。

```
原木本体: 继承 木材生产本体
{
属性: 平均直径
      类型 float
属性: 小头直径
      类型 float}
  条件:平均直径≤小头直径
```

2.2.6　添加实例　在由类的层次结构、语义关系、属性等条件构成的林产品领域知识体系框架中描述领域概念中的个体就可以逐步建立起林产品领域的本体模型了。其中包括类添加实例,设置相应的属性并加以约束等。以原木本体添加实例为例,在“林产品”—“原木”—“杨木”的层次结构中确定“林产品”为起始,而“杨木”为最低粒度水平。至于类的起始和实例的最低粒度水平是由应用范围确定的。

2.2.7 本体的确认评价与进化 本体开发的确认与评价采用迭代增量的构建方法。构建好了核心本体原型后，参照现有的本体评价准则，其中包括明确性、清晰性、一致性、可扩展性、约束最小等方面。再由领域专家对本体进行评价和确认以达到应用的目的。

领域知识很多且存在交叉，同时领域知识层出不穷。因此构建完成的核心本体原型是构建的结束也是该本体进化的开始。在修正与完善本体的方式有多种，其中包括集成新本体、机器学习、内容挖掘等多种方法。当然本体的进化主要以相关领域内用户的需求为驱动的。

## 3 林产品本体 OWL 形式化表示与可视关联图

Web Ontology Language(OWL)是 W3C 为语义网应用定义得本体语言。它是结合了 DAML+OIL 应用经验而改进的修订版，建立在 RDF 基础上，以 XML 为书写工具。除了能表达概念间的语义关系外，还用于计算机与计算机之间的交流，相比较于 XML、RDF 和 RDFS 拥有更多的机制来表达语义[11]。OWL 可以通过本体建模可视化工具 Protégé[12]及其相关的插件来构建林产品本体。林产品领域本体的部分源代码见图 4。

构建的林产品产品核心本体原型，包含了林业企业、事业单位以及其相关实体产品和技术的实例。利用 Protégé 中的 TGVizTab 插件，可以显示构建的林产品本体层次效果(见图 5)。

## 4 结论

针对基于 Web 的林产品商务信息整合过程中，由于分布、异构而造成的知识表达、知识共享等困难。采用本体的形式化定义方法，提出了构建林产品本体的六元组逻辑结构。并采用自顶向下迭代的方法对该领域的概念进行描述。根据应用需求，确定了该领域概念间的语义关系、属性及约束等工作。进而建立了林产品的本体原型。该原型有利于消除语义冲突，为知识推理和内容挖掘等应用提供语义框架基础。

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
    xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
    xmlns:swrl="http://www.w3.org/2003/11/swrl#"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns="http://www.owl-ontologies.com/Ontology1269850235.owl#"
    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#"
  xml:base="http://www.owl-ontologies.com/Ontology1269850235.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="林业投资">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >林业投资</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="林产品"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="林业机械制造">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#林产品"/>
    </rdfs:subClassOf>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >林业机械制造</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="森林狩猎">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >森林狩猎</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#林产品"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="药材生产">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#林产品"/>
```

图 4 林产品本体 OWL 代码(部分)

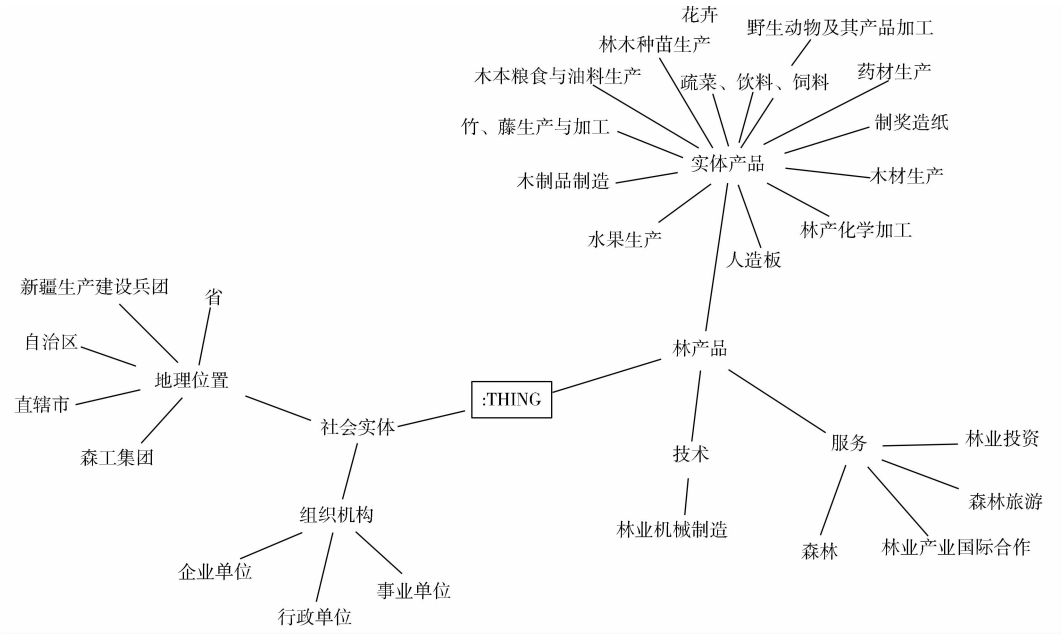


图 5　林产品类和林业机构类的关联知识显示图(部分)

## 参考文献:


[1] Ushold M. Knowledge levelmodeling:concep ts and terminology[J]. Knowledge Engineering Review, 1998, 13(1): 25-29.

[2] 陈刚,陆汝钤,金芝. 基于领域知识重用的虚拟领域本体构造[J]. 软件科学,2003,14(3):10-13.

[3] PerezA G, Benjamins V R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem Solving-Methods[C]//BenjaminsV R, Chandrasekaran B, Gomez Perez A, et al. Proceedings of the IJCA I-99Workshop on Ontologies and Problem-Solving. Methods (KRR5). Stockholm:Sweden, 1999:1-15.

[4] 蒋维,郝文宁,橱晓恕. 军事训练领域核心本体的构建[J]. 计算机工程,2008(5):191-192.

[5] 张钢,倪旭东. 从知识分类到知识地图:一个面向组织现实的分析[J]. 自然辩证法通讯,2005(1):59-60.

[6] 苏新宁,任皓,吴春玉,等. 组织的知识管理[M]. 北京:国防工业出版社,2004,42(3):103-126.

[7] 李景. 本体理论在文献检索系统中的应用研究[M]. 北京:北京图书馆出版社,2005.

[8] 张云涛,龚玲,王永成. 面向自然语言提问的检索技术[J]. 广西师范大学学报:自然科学版,2003,21(1):629.

[9] 陈强,廖开际,奚建清. 专家知识地图的关键技术与设计[J]. 计算机工程与科学,2008,30(2):96-114.

[10] 郑业鲁,李泽,何绮云,等. 农业生产技术和市场信息本体构建及其初步应用[J]. 农业网络信息,2009(8):47-49.

[11] 高文,张小栓,傅泽田. 基于 OWL 的鱼病诊断本体模型[J]. 计算机工程与设计,2007,28(19):4470-4471.

[12] Protégé[EB/OL]. http://Protege. stanford. edu,2007-05-17.


# An Approach for the Design of Forest Products Ontology


**YANG Shu[1,2], WU Gang[1], WANG Huan[1], PENG En-qiang[1]**

(1. Information Science and Technology College of Beijing Forest University, Beijing 100083; 2. Computer And Information Science Department of Xinjiang Agriculture University, Urumqi, Xinjiang 830001)



**Abstract**: The key problem that affects the multi-source information integration and the difficulty of share caused by semantic inconsistency. From the present theory of semantic Web and Ontology, Formal semantic representation about the concept of forest products is made by using formal definition method of ontology. Then the logic structure of the forest products ontology was proposed according to the properties of forest products. On this basis, the ontology model of forest products was built. Finally, the relevant visible figure of this Ontology was presented.



**Key words**: Web information integration; forest products; domain ontology; concept semantic