**果树学领域本体构建研究**

**结题报告**

(陈国秀, 山西农业大学图书馆, 2010027 )

1. **研究背景、目的及意义**

**1.1研究背景**

随着计算机及信息技术的发展、因特网上的学术信息越来越多，大量的以文字、图像、数据库、音、视频等形式存在的信息资源非常丰富，形成了广阔的信息海洋。人类如何充分利用计算机来解读和表达这些海量信息，使其得到高效、准确的采集、存储与利用，使分布、异构的信息资源之间实现数据、信息和知识的交换、重用和共享？这些问题不仅是图书馆数字化进程中必须面对的，也是全人类知识的数字化进程中必须解决的难题。语义网的构想正是针对这一难题的。

语义网的构想是在WEB中引入语义知识表示，为网页增加具有语义的信息，从而保证WEB页面能被计算机理解和自动处理。因此，如何表示语义信息对于语义网就显得非常关键。本体正是描述语义网中语义知识的建模手段，它形式化地定义了领域内共同认可的知识，是语义网体系中的核心。

领域本体是用于描述指定领域知识的一种专门本体。它给出了领域实体概念及相互关系、领域活动以及该领域所具有的特性和规律的一种形式化描述。目前,本体模型的研究已经进入实际应用阶段,许多研究领域都建立了自己标准的本体。

**1.2研究目的**

本研究旨在通过本体构建实践加深对本体、语义网等概念的正确理解，熟悉本体构建的具体步骤和方法，为探索语义网环境下的图书馆发展方向积累知识和经验。

1. **研究内容、思路及创新点**

**2.1 研究内容**

本研究拟以作者熟悉的果树学为目标领域，充分了解现有各种本体编辑器的功能特色，选择适合的本体编辑工具，尝试构建果树学本体的具体步骤和方法。

**2.2 研究思路**

通过大量阅读现有本体构建相关论文，选择1-2个成功案例，借鉴其方法和要点；充分掌握本体构建基本要领的基础上，着手果树学本体构建实践，并邀请果树学的专家参与其中，力求对目标领域的知识描述做到准确、全面。

**2.3 创新点**

本研究的研究目标、研究内容、研究方法等都具有一定的创新性，因为语义网作为第三代互联网的发展方向，是最近几年才被提出的新思想，语义网的概念及相关知识均处于研究阶段，本项目的研究涉及语义网的核心概念——本体构建，无论是构建方法、构建过程、构建内容均属于创新性探索。

另外，虽然相似的本体构建研究已在多个领域被发现，如中医针灸学、图书情报学等领域，但均处于探索研究阶段，通过大量的查询，没有发现果树学领域的本体构建研究的相关报道。

1. **研究过程、方法及步骤**

## 3.1 研究过程：

 对本馆馆藏以及国内主要数据库CNKI、维普中文期刊数据库及万方数据等进行了“领域本体”、“构建”等主题检索，获得参考文献（书和论文）若干，进行了认真学习研究，初步掌握了构建领域本体的一般方法和基本步骤。

对常见本体编辑工具进行了认真研究，确定选择美国Stanford大学开发的开源软件Protégé作为果树学本体编辑工具。该软件版本非常多，而且处于动态变化中，发展很快。通过反复试用，观察其功能和特性，重点是观察其对中文建模的图像显示的支持程度，最后，确定选择了Protege3.4.4版。

果树学本体构建基本流程：

1.确定果树学科的核心概念；

2.围绕核心概念，搜集、罗列所有与该学科相关的概念，构成学科概念集；

3.确定各学科概念的层级关系，在Protégé 中建立类、子类；

4.分析类之间的属性联系，建立属性关系，并给每一属性定义域、值域及特性（定义公理）；

5.为类概念建立实例；

6.进行推理；

7.形式化：图像显示，表达。

1. **结论与建议**

**4.1 果树学本体模型核心概念集**

 果树学主要包括分类学、栽培学、遗传育种学、贮藏加工学、病虫害学5大学科。以此5大学科为核心，建立核心概念集。

借助《中图法》第四版、《中国分类主题词表》第二版、及《中国果树分类学》（渝德俊）、《果树栽培学总论》（郗荣庭）、《果树遗传育种》、《果树病理学》、《果蔬贮藏加工学》（隆锐柏）《果树病理学》（曹若彬）、《果树昆虫学》（北京农业大学等）等收集关键概念，确定上位概念、较为显著的概念和常用概念，通过分析，确定概念特性（名词性概念、动词性概念、物种概念、病害名称等），建立核心概念集。

**4.2 概念之间的属性关系**

本研究分析出的概念之间的关系主要有以下几种：

|  |  |
| --- | --- |
| **名称** | **含义** |
| 包含关系 | A包含B, B是A的部分； |
| 互逆关系 | A有（品种）B, B是A的（品种）； |
| 亲子关系 | A是B的亲本， B是A的子代， |
| 同位关系 | A有兄弟B, B有兄弟A, A是B的姐妹，B是A的姐妹， |
|
| 等同关系(别称)  | A是B, B是A， |
|

**4.3 用Protégé构建果树学本体模型（见附图）**

研究实践证明，学科领域本体构建是一项非常复杂的工程。建模所必需的知识基础包括：离散数学和概率论、逻辑学、领域知识、AI领域的推理知识、知识表示、计算语言学、编程和数据库技术、对知识进行组织的知识（分类法和叙词表）和编码学等9个领域的知识。由于课题组人员对逻辑学、概率论、计算语言及编程方面的知识积累不够，使本研究难以向深度进展。

由于时间有限，本课题只以果树分类学为例，对学科领域本体的构建过程进行了实践性探索。旨在了解领域本体构建的基本方法和步骤，进一步理解构建领域本体的重要意义和其应用价值。

本课题的本体构建过程选用了Protege3.4.4版，完全在中文环境中进行，类名称和属性名称均使用中文。研究证明该软件版本对中文具有很好的支持功能。

1. **支撑材料列举（发表的文章、相关照片、系统、软件、获奖情况等等）**

图1 Protégé3.4.4构建的果树学本体截图

图2 类、子类、实例全图

图3 属性关系图

1. **参考文献**

[1] 董慧.本体与数字图书馆. 武汉: 武汉大学出版社，2008.8

[2] 陆建江等.语义网原理与技术. 北京: 科学出版社，2007.3

[3] 李景.本体理论在文献检索系统中的应用研究. 北京: 北京图书馆出版社，2005.3

[4] 中国图书馆分类法编辑委员会.中国图书馆分类法(第四版). 北京: 北京图书馆出版社,2004,6

[5]国家图书馆中国图书馆分类法编辑委员会.中国分类主题词表(第二版).北京: 北京图书馆出版社,2005,9

[6] 杜小勇, 马文峰, 武文娟,等. 学科领域本体的构建与进化——以经济学领域为例，现代图书情报技术，2007,3, 7-12

[7] 孙瑾.本体编辑工具的分析与研究.图书情报工作，2006,12, 26-29

[8] HORRIDGE M. A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools. Edt.1.2 Manchester: The University of Manchester, March 13, 2009

**专家验收表**

|  |  |
| --- | --- |
| 项目名称 |  |
| 主持人 |  | 职务/职称 |  |
| 所在单位 | （加盖公章） |
| 专家意见 |  |
| 专家签名 |  |  |  |  |  |
| 职务/职称 |  |  |  |  |  |

\*如需要可增加页数