

# 基于形式概念分析领域本体构建方法研究

王璐

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

**摘要:** 本文的目的是提出基于形式概念分析的领域本体构建的新方法, 旨在解决领域本体在信息科学等领域内广泛应用但同时又缺少成熟有效的领域本体构建方法对其支撑的矛盾。本文的创新点是构建了基于形式概念分析的领域本体构建方法的新模型, 提出了基于形式概念分析的领域本体构建的新方法, 在领域本体方法理论上进行了创新。本文的研究成果是明确了基于形式概念分析的领域本体构建方法的基本原理, 阶段划分, 构建了基于形式概念分析的领域本体构建方法模型, 并将模型具体化为基于形式概念分析的领域本体构建方法的具体流程。本文的研究成果为领域本体在相关学科各个领域的应用起到了良好的促进作用, 为领域本体的构建方法的发展与创新提供了启发。

**关键词:** 形式概念分析 概念格 领域本体 领域本体构建方法

**中图法分类号** TP311

**文献标识码** A

## Research on Domain Ontology Construction Method Based on

## Formal Concept Analysis

Wanglu

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)

**Abstract :** The purpose of this paper is to propose a new method of domain ontology construction based on formal concept analysis. It aims to solve the contradiction between domain ontology which is widely used in information science and other fields, but lacks mature and effective domain ontology construction method. The innovation of this paper is to construct a new model of domain ontology building method based on formal concept analysis, propose a new method of domain ontology construction based on formal concept analysis, and innovate on domain ontology theory. The results of this paper are as follows: the basic principle and phase division of domain ontology construction method based on formal concept analysis are defined, and the domain ontology construction method model based on formal concept analysis is constructed, and the model is embodied as domain ontology construction based on formal concept analysis Method of the specific process. The research results of this paper can promote the application of domain ontology in various fields of related disciplines, and provide enlightenment for the development and innovation of domain ontology construction methods.

**Key words:** formal concept analysis; concept lattice; domain ontology; domain ontology construction method

## 1 研究背景

本体的作用的捕捉, 描述和表达相关领域知识, 提供对该领域知识的共同理解, 确

收稿日期: 2016-10-9

作者简介: 王璐(1994-)女, 硕士生在读。

定该领域内共同认可的词汇，并从不同层次的形式化模式上给出这些词汇和词汇间相互关系的明确定义。尽管本体的功能已经被认识，并逐渐得到应用，但目前本体构建的研究还处于探索阶段，没有形成成熟，统一的方法作为指导，不能满足本体应用的需求。目前仅有的一些方法也是从个案的开发过程中通过逆向工程总结出来的，并存在着很多不足和固有的局限性，这就产生了一对矛盾：本体技术在信息科学中的广泛应用在客观上要求必须有先进的领域本体构建技术来对其支撑，而当前的构建方法由于各自种种的缺陷并不能很好的满足这种客观需求，因此，要解决这对矛盾，就必须在本体构建方法方面做出积极的探索，领域本体的构建方法，必须向标准化，工程化，规范化，细节化的方向来发展。

就国内外当前的情况来看，将形式概念分析理论应用于领域本体的构建理论都尚在起步阶段，形式概念分析与领域本体结合的方式不尽相同，结合的深度各异，尽管有些理论存在着一定的可取之处，但现有的结合方式仍存在着很多不足，将形式概念分析引入到领域本体构建方法当中，还有大量的工作要做，还需深入地进行研究。

## 2 研究意义

本选题的理论意义体现在：丰富了领域本体构建方法的理论，提出了新的基于形式概念分析的领域本体构建方法，弥补了当前领域本体构建方法的不足之处，同时以新的思路。新的视角为领域本体构建方法的理论的发展与创新提供了启发，完善了领域本体构建的理论基础。

本选题的现实意义体现在：提出了领域本体构建的新方法。提高了所构建领域本体的效率和质量，为领域本体在各个领域的应用起到了良好的促进作用。

## 3 形式概念分析用于领域本体构建的现有方法分析

国外再将基于形式概念分析用于本体构建方面已经取得一些成果，具有代表性

的方法主要有：Cimiano 的方法 [1-2]，GuTao[3]的方法，Haav[4-5]的方法，Marck Obitko[6]方法，将形式概念分析用于领域本体构建现在仍处在一个探索的阶段。

### 3.1 Cimiano 的方法

Cimiano 等提出了一种采用形式概念分析分析词语在文本中的使用方式来获得相应的背景知识进而生成本体的领域本体构建方法。该方法的基本思想是：首先，使用一个自然语言的解析器，通过该解析器从领域本体中的每一个句子都可以得到一颗语法树；其次，由语法树直接得到动词对象间的依赖关系；再次，通过进一步的词典查询，对提取的动词和对象用词进行规范化表示。最后，将形式概念分析中的概念和本体中的概念直接等同，得到概念格，由概念格得到领域本体。

### 3.2 GuTao 的方法

GuTao 提出的形式概念分析用于本体构建的方法如 3.2.1 图所示：

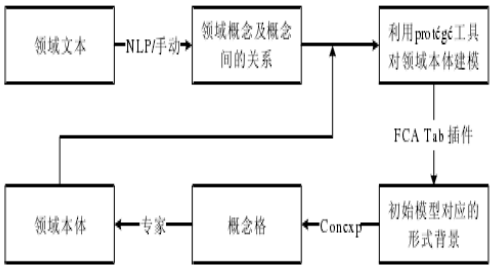


图 3.2.1 方法

### 3.3 Haav 的方法

Haav 提出了将形式概念分析与基于规则的语言相结合，进行半自动的领域本体构建的方法，该方法适应的两个前提是：①领域文本内容比较短的情况；②假设领域本体描述了某一实体，里面包含了描述领域的术语。该方法将概念格节点的属性作为本体中的概念，对本体的属性表达也就成为空缺，不利于表达领域本体属性间的关系。该方法在形式背景的建立上存在不合理性。

### 3.4 Marek Obitko 的方法

Marek Obitko 等人在 GACR 项目中提出了一种将形式概念分析用于领域本体构建的方法，该方法约定：概念由属性来描述；属性决定概念的层次结构；当两个概念的属性相同时，这两个概念也相同；可以直接有

修改过的概念格作为本体表示。基于以上规则，该方法的具体步骤如下：

- （1）方法的使用从空的对象和属性开始；
- （2）由领域本体开发人员根据需要把对象和属性添加到形式背景当中；
- （3）显示形式背景所对应的概念格；

在可视化基础上，按需要做如下操作：一是直接编辑，包含增加或删除对象，增加或删除属性，给对象增加属性或从对象中删除属性；二是根据提示编辑概念格，例如两个对象在概念格上同一个位置，要么被对象添加属性，以区别对象。另外，FCA 可以产生新的对象，它们不在形式背景中，可以增加这些概念。

最后重复这一过程，直到领域本体开发人员满意为止。

Marek Obitko 的方法只适合小领域本体的构建，总的来说，该方法的自动化程度较低。

## 4 领域本体构建方法的详细流程

基于形式概念分析的领域本体构建方法模型是基于形式概念分析的领域本体构建方法[7]的高度抽象，有利于把握该方法的本质。但仅仅从模型的层次来认识基于形式概念分析的领域本体构建方法是不够的，要进一步深入研究该方法，就必须将模型细化成流程，下面将一一详述四个详细流程：

### 4.1 制定开发计划

制定领域本体开发计划是基于形式概念分析的领域本体构建方法的第一步，该步骤的主要任务包括：明确领域本体开发项目的进度安排情况；调度领域本体开发项目所需要的资源；对要建立本体的相关领域进行初步调查，同时充分考虑建立该领域的本体所受到的各种约束，研究开发建设该领域本体的必要性和可行性，提出领域本体构建的可行性研究报告；对确保领域本体项目管理工作的顺利实施的相关问题进行安排。制定开发计划是整个基于形式概念分析的领域本体构建方法的起点，也是运用该方法构建领域本体的基础，是不可或缺的一个步骤。

### 4.2 领域本体分析

在该步骤中，最重要的是对领域本体需求的分析，对需求分析的具体内容包括：领域本体需求的细化问题，领域本体需求追溯能力，领域本体需求满足问题，领域本体需求文档的生成，领域本体需求升级五个方面。通过对上述几个方面的需求分析，对上述信息给出完整，准确，清晰，具体的定义，并编制领域本体需求分析说明书。

### 4.3 领域本体设计

基于形式概念分析的领域本体设计的目的是生成领域本体的概念层次模型，其具体包括四个步骤：

- （1）对领域文档的预处理

本体时明确的形式化说明，实际上，大多数形式化并不是明确的，而是相当含糊的存在于领域文档或大脑的活动中。所以，要构建领域本体，必须先对领域文档进行相关处理，将领域文档转换成领域的关键概念及概念间的关系。本文将领域文档分为三类：结构化数据，半结构化数据和非结构化数据，这对三类文档的预处理要分别对待。

- （2）依据“对象—属性”关系生成领域形式背景

形式背景是应用形式概念分析法的基础，是形式概念分析的基本柱石。领域形式背景的生成，其目标是建立对象，属性及对象与属性关系的三元关系表。

- （3）依据单值背景构造概念格

该步骤的主要任务是依据生成的单值形式背景，选用合适的造格算法，将单值形式背景转换成概念格。

- （4）将概念格进行转换处理得出领域本体概念层次模型

本步的任务是将概念格进行转换处理，得到领域本体概念层次模型。其做法是将概念格中的概念(即 Hasse 图中的节点)和本体中的概念直接等同，通过底端节点处理、概念关系处理、顶端节点处理三个处理步骤，将概念格转换成描述本体概念层次的领域本体概念层次模型。

### 4.4 领域本体编码实现

领域本体的编码实现环节也包括三个方面的步骤：

- （1）根据领域本体概念层次模型映射

出领域本体原型

本步骤的作用是根据构建出的领域本体概念层次模型,映射出领域本体原型。映射主要包括两部分的内容,一是对本体概念层次模型的节点的标示,二是对概念层次关系的转换。

### (2)对领域本体原型进行扩充

对领域本体原型的扩充是在领域专家的参与下完成的。领域本体原型的扩充主要包涵四个方面的内容:对领域本体属性的扩充、对领域本体实例的扩充、对领域本体公理的扩充、对领域本体非分类关系的扩充等。

### (3)对扩充后的领域本体进行形式化描述

本步骤的主要作用是选择合适的本体描述语言,对扩充后的领域本体进行形式化描述,即完成本体的编码过程,最终得到领域本体。

到目前为止,本体描述语言呈现出一种多样化的趋势,在这种多样化趋势下,本体描述语言的选择就成为一个需要关注的问题。本文的观点是,本体描述语言的选择并非唯一的,而是需要与具体的项目结合起来,综合考虑各种本体描述语言与所构建领域本体的关系,然后做出选择。一般情况下,建议选择 OWL 描述语言对本体进行描述。

开发本体的主要目的是知识共享和重用。如果开发出的本体没有较强的支持复用的能力,从某种意义上说,所开发的本体是失败的。

## 5 领域本体的测试

领域本体测试[9]是保证领域本体可靠性的主要手段。领域本体测试阶段的主要任务是测试应用基于形式概念分析的领域本体构建方法所构建好的领域本体,排除已构建好的领域本体的错误,确保其达到规定的要求,并形成领域本体测试报告。

领域本体测试的一般准则是:①所有的领域本体测试都应能追溯到领域本体需求中去;②远在领域本体测试工作开始之前就应当制定领域本体测试计划;③测试中应当重视帕累托法则(即二八法则)的应用;④先开始“小规模”的测试,并逐步开展“大

规模”测试;⑤穷举测试是不可能的;⑥推荐独立的第三方来进行领域本体构建的测试工作。

## 6 领域本体的维护

领域本体的维护是领域本体生命周期的最后一个阶段,也是持续时间最长花费代价最大的一个阶段。所谓领域本体的维护就是将所开发的领域本体交付使用之后,为了改正错误或满足新的需求而修改领域本体的过程。

领域本体的维护通常包括四类活动:①为了纠正在领域本体使用过程中暴露出来的错误而进行的改正性维护;②为了适应领域本体的应用环境的变化而进行的适应性维护;③为了改进原有领域本体而进行的完善性维护;④为了改进将来的可维护性和可靠性而进行的预防性维护。

从某种意义上说,基于形式概念分析的领域本体构建方法的每一个步骤都和领域本体的维护有着密切的关系。良好的设计,完整准确易读易理解的文档资料,以及一系列严格的测试等等,因此,在方法进行的每一个步骤,都应该充分的考虑领域本体的维护问题,为领域本体维护做准备。

## 7 领域本体构建方法的优越性

较之传统的领域本体构建方法[8],本文所构建的基于形式概念分析的领域本体构建方法有很大的优越性,同时本文所构建的方法又对国外所提出的一些形式概念分析用于领域本体构建方法中所存在的问题进行了修正及完善,因此,本文提出的方法在这个意义上,也具有优越性。总的来说,本文提出的基于形式概念分析的领域本体构建方法有着以下诸多方面的优点:

(1)充分考虑了领域本体的各类结构数据及隐性知识,提出了依据领域结构化数据、半结构化数据及非结构化数据向领域关键概念及概念关系的转换机制;

(2)提出了形式背景的处理和完善机制,充分考虑了对多值背景的转换及形式背景的编辑修改完善机制;

(3)在方法具体操作的过程中,引入了循

环反馈机制,使得领域本体的构建过程能够更加合理、更加完善;

(4) 充分考虑了领域本体的非分类关系、本体公理及推理规则,为本体的智能推理奠定了基础

## 参考文献

- [1] Cimiano P, Stumme G, Hotho A, Tane J. Conceptual knowledge processing with formal concept analysis and ontologies. In: Proc. of the The Second Intl. Conf. on Formal Concept Analysis (ICFCA 04), Springer, 2004, 189-207
- [2] Cimiano P, Staab S, Tane J. Automatic Acquisition of Taxonomies from Text: FCA meets NLP. In: Proc. of the Intl. Workshop on Adaptive Text Extraction and Mining, 2003, 10 — 17
- [3] Gu Tao. Using formal concept analysis for ontology structuring and building. ICIS, Nanyang Technological University, 2003
- [4] Haav H M. A semi-automatic method to ontology design by using FCA. In: Snasel V, Belohlavek R, eds. Concept Lattices and their Applications. Proceedings of the 2nd International CLA Workshop, TU of Ostrava, 2004, 13-2\_5
- [5] Haav H M. An Application of Inductive Concept Analysis to Construction of Domain-specific Ontologies. In: B. Thalheim, Gunar Fiedler (Eds), Emerging Database Research in East Europe, Proceedings of the Pre-conference Workshop of VLDB 2003, Computer Science Reports, Brandenburg University of Technology at Cottbus, 2003, 14(3), 63-67
- [6] Obitko M, Snasel V, Smid J: Ontology Design with Formal Concept Analysis, Edited by Vaclav Snasel, Radim Belohlavek. In: Proc. of the CLA 2004 Intl. Workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sep. 2004. 111 — 119
- [7] 陈禹. IDEF 建模分析与设计方法 [M]. 北京:清华大学出版社, 1999
- [8] 张志刚. 领域本体构建方法的研究与应用 [D]. 大连海事大学硕士学位论文

文, 2008

[9] 张海藩. 软件工程导论 [M]. 北京:清华大学出版社, 2003 年 12 月第 1 版