基于形式概念分析的本体研究

郭佳伟

（大连海事大学 辽宁大连 计算机技术系）

摘要:由于传统的本体构建方法主要是由人工完成，因此会造成耗时耗力、主观思想的干扰比较大、对隐含的概念和关系选取模糊等问题，从而提出了形式概念分析构建本体的方法，该方法有很多类，例如：基于结构化资源、基于非结构化资源、异构资源的合并本体构建方法等。其中，每个方法否有优点和缺点，根据客观世界、形式背景进行构建概念格对问题的信息进行提炼和抽取是解决这一问题的一小步，结合具体的实际情况和应用领域在构建时对对象和属性的取舍、针对不同的特点进行构建形式背景进行进一步研究。

关键字:本体；形式概念；形式背景

**Ontology Research Based on Formal Concept Analysis**

GUO JiaWei

（Dalian Maritime University Dalian, Liaoning Department of Computer Technology）

Abstract: As traditional ontology construction method is mainly done by human, it will cause time-consuming and labor-intensive, the interference of subjective thoughts is relatively large, the implicit concepts and relations are ambiguous, and so on. Method, which has many classes, such as: based on structured resources, based on unstructured resources, heterogeneous resources, such as merging ontology building methods. It is a small step to solve this problem, which is combined with the actual situation and application field. At the time of construction, the method of constructing concept lattice according to the objective world and formal background is the best way to solve the problem. The choice of objects and attributes, for different characteristics of the background to build further study.

Keyword:Ontology;Formal concept;Formal Background

# 1 概述

随着语义网和计算机信息的飞速发展，对信息的提取和抽象对于系统的研究越来越关键。通过对客观世界的实际问题中获取形式概念，将其置于形式背景中进而构建概念格，通过对形式概念表的对象约简、关联规则抽取、概念推理从而进行知识优化、知识发现、知识推理，进而达到获取知识的目标。

概念是反映对象的特有属性的思维形式，是从对象的属性中抽取特有属性概括而成，表达的语言形式多样，都有内涵（特定结构具体内容）和外延（适用对象涵盖范围），通过对概念的抽取后，还应当用一些概念表示法：表达式法、二维表法、图示法。因此，对于概念的提取和抽象的研究至关重要。

# 2 形式概念分析中的相关定义

形式概念分析是由R.Wille于1982年提出，是应用数学的一个分支，是信息处理的一种理论，是知识处理的一种理论，是对现实世界的对象和属性的提取，关于形式概念和概念的主要定义如下：

## 定义1

设 形式对象集：U（X属于U），形式属性集：A（B属于A）

二元关系R属于（U X A）

若 X={x|x属于U，对任意的a属于B，xRa}

B={a|a属于A，对任意的x属于X，xRa}

则 二元组（X，B）被称为形式概念

X中x 每个x都有全部属性

B中a 每个x都有的属性

## 定义2

一个形式背景K={G，M，I}是由2个集合G和M以及G和M之间的关系I组成。G的元素称为对象，M的元素称为属性。（g，m）属于I或者glm表示对象g具有属性m。

## 定义3

若（A1，B1），（A2，B2）是某个形式背景的2个概念，而且A1属于A2，则称（A1，B1）是（A2，B2）的子概念，（A2，B2）是（A1，B1）的父概念，并记作（A1，B1）<=（A2,B2）。形式背景中的所有概念用这种序组成几个称为概念格，记作：L（G，M，I）K=（G，M，I）。

# 3 本体的定义

给出构成相关领域词汇的基本术语和关系，以及利用这些术语和关系的构成的规定这些词汇外延的规则的定义，本体是概念模型的明确的规范说明，是共享概念模型的形式化规范说明，其中基本构建元语：类，关系，函数，公理，实例。概念可以指任何事物；关系表示概念间的相互作用；函数是一种特殊的关系；公理表示永真断言；实例表示元素。

本体的目标是获取、描述和表示相关领域的知识，提供对该领域知识的共同理解，确定该领域内共同认可的词汇，并且在不同层次的形式化模式上给出这些词语和词语之间相互关系的明确定义，本体论是反省客观存在的概念模型。用一个公式表示：本体=概念+关系+函数+公理+实例。

# 4 构建本体的准则和方法

为了使本体构建更加的贴近实际且实用性强，构建本体有许多的规则：清晰性；完全性；一致性；可扩展性；最小承诺（只定义必要的术语，只定义约束最弱的关系）；最小编码偏好（不指定术语形式化用何编码）。

构建本体的层次可以划分为数据源/技术层，处理层，输出层，并且这5种方法中所使用的数据源和技术，在数据源/技术层-处理层以及处理层-输出层之间建立联系。

## 4.1 Uschold和King方法

建立在企业本体基础之上，是相关企业间术语和定义的集合，该方法只提供开发本体的指导方针，目前企业本体在爱丁堡大学人工智能研究所及他的合作伙伴，具体的步骤：

1. 确定本体应用的目的和范围：根据所研究的领域和任务，建立相应的领域本体或过程本体，领域越大，所建立本体越大，因此需限制研究的范围。
2. 本体分析：定义本体所有术语的意义及其之间的关系，该步骤需领域专家的参与，对该领域越了解，所建的本体就越完善。
3. 本体表示：一般用语义模型表示本体。
4. 本体评价：建立本体的评价标准是清晰性、一致性、完善性、可扩展性。清晰性就是本体中的术语应该无二义性；一致性指的是本体中的关系逻辑上应当一致；完整性指的是本体中的概念以及关系应当是完整的，应该包括该领域的所有概念，但是很难达到，需要不断完善；可扩展性本体应用能够扩展，在该领域不断发展时能加入新的概念。
5. 本体的建立：对所有本体按照以上标准进行检验，符合要求的文件的形式存放，否则转到（2）。

## 4.2 Gruninger和Fox方法

1. 定义直接可能的应用和所有解决方案，并且提供潜在非形式化的对象和关系的语义表示。
2. 能力问题作为约束的条件，把偶偶能够解决什么样的问题和这个问题应当怎样去解决，这里的问题用属于表示，答案用公理和形式化定义回答，由于在没有形式化Ontology之前进行的，所以叫非形式化的能力问题。
3. 术语的规范化：从非形式化能力问题中提取非形式化的术语，然后用Ontology形式化语言进行定义。
4. 形式化的能力问题：一旦能力问题脱离了非形式化，Ontology术语已经定义，则能力问题自然形式化了。
5. 形式化公理：术语定义所遵循的公理用一阶谓词逻辑表示，其中包括定义和语义或者解释。
6. 说明问题的解决方案必须是完全的。

## 4.3 Berneras方法

这种方法开发本体由应用开发控制。所以每一个应用都有相应的表示该应用所需的Ontology。这些本体既能重用其他的Ontology，也能被后继应用集成，应用于电子网络的开发。开发过程如下：

1. 应用的说明：提供应用的上下文和应用模型所需要的组件。
2. 相关本体论范畴的初步设计：首先搜索已经存在的Ontologies,然后进行抽象、提炼、扩充。
3. Ontology的构造：最小关联原则用来确保模型既能相互依赖，有尽可能一致，一直得到最大同构。

## 4.4 Methontology方法

这种方法由马德里大学工艺分校开发人工智能图书馆使用。它分为三个阶段：

1. 管理阶段：这一阶段的系统规划规则包括任务的进展情况、需要的资源、如何保证质量等问题。
2. 开发阶段：规范说明y概念化y形式化y执行y维护。
3. 维护阶段：包括知识的获取、系统的集成、评价、文档的说明与解释、配置的管理与改善。

## 4.5 基于Sensus方法

这个Ontology用于自然语言程序，有ISI自然语言组企图为机器翻译提供广泛的概念结构，共有5万多个概念，为了能在Sensus基础上构造特定领域的Ontology，必须把不相关的术语从Sensus中剪出掉，具体过程如下：

1. 定义/叶子术语；
2. 把叶子属于手工的和Senses术语相连；
3. 找出叶子节点到Sensus根的路；
4. 增加和域相关并且没有出现的概念；
5. 用启发式思维找出全部的特定的域的术语：对于某些有两条以上路经过的结点必须是一颗子树的父节点，那么这颗子树以上的所有结点都和该域相关，是要增加的术语。对于高层结点通常有多条路经过，则很难判断。

# 5 结束语

由于本体提供通信双方的公共理解，类似于网络协议在通信双方的地位，只不过本体是人工智能角度出发构造的软件。正由于此，国外研究本体异常活跃，国内则处于起步阶段。本体的范围从形成所有领域的知识表达的基础的非常通用的术语到限定特定知识领域的专业术语，本体可以应用于许多领域，如：电子工程、化学、远程教育、电子商务等。信息检索系统、数字图书馆、易购信息的集成、以及Internet搜索引擎都需要领域本体来组织信息和指导搜索过程。

现有的教学系统缺乏知识的工程化，针对教学系统进行设计的常用词汇表和框架，在恰当的抽象层次上给智能教育的任务加以形式化。使学习者通过教学系统发现自己的不足，实现异构的自治系统间的互操作，设计本体，充分利用现有的教学资源使形式概念分析更加实用，将其作用尽可能最大化。

# 参考文献

[1]Mike Uschold,Michael Gruninget The Knowledge Engineering Review[J]1 1996,11(2):93

[2]刑平平，施鹏飞，熊范纶1数据挖掘技术在农业数据中的有效应用[J]1 计算机工程 与应用，2011（2）：

[3] 张文秀，朱庆华.领域本体的构建方法研究[J].图书与情报，2011，（1）：16-19,40.

[4]Knowledge Interchange Format[EB/OL]1 NCIT T2/004

[5]吴泉源1 人工智能与专家系统[J]1 长沙：国防科技大学出版社

[6]Uschold M.Building Ontologies:Towards A Unified Methodology[J]1 Expert System 96,Cambrindge,1996

[7]滕广青，毕强.国外本体协调研究前沿进展及热点分析[J].中国图书馆学报，2012， （1）：113-120.

[8]张云中.一种基于本体构建方法[J].现代图书馆情报技术，2011，27（12）：15-23

[9]滕广青.基于概念格的数学图书馆知识组织研究[D].长春.吉林大学，2012

[10]刘萍，高慧琴，胡月红.基于形式概念分析的情报学领域本体构建[J]