

# 基于形式概念分析的知识集成建模研究

亓泽东

作业	分数[20]
得分	

2020 年 11 月 13 日

# 基于形式概念分析的知识集成建模研究

元泽东

(大连海事大学 信息科学技术学院 大连 中国 116026)

**摘要:**随着互联网的快速发展,信息共享和数据交换不断扩大。近年来,本体学习逐渐为研究人员熟知,原因是获取信息较为简单且能提供可共享的高级结构。本体作为共享化概念的形式说明,其应用不断被采纳,采用形式概念分析的方法,对知识集成建模过程进行研究和分析,在此基础上利用本体来描述知识模型,提出基于形式概念分析的知识集成过程模型,结果证明在建模过程中引入 FCA 可以产生创新知识概念,提高网络资源的查全率,还可以对知识概念异构整合,消除同形异义词,提高网络资源的查准率。

**关键词:**形式概念分析 知识集成 本体

中图法分类号 TP18

## Research on Knowledge Integration Modeling Based on Formal Concept Analysis

Qize Dong

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

**Abstract:** With the rapid development of the Internet, information sharing and data exchange continue to expand. In recent years, ontology learning has gradually become familiar to researchers because it is relatively simple to obtain information and can provide a high-level structure that can be shared. As a formal description of shared concepts, the application of ontology has been continuously adopted. The method of formal concept analysis is adopted to study and analyze the knowledge integration modeling process. On this basis, the ontology is used to describe the knowledge model, and a method based on formal concept analysis is proposed. Knowledge integration process model, the results prove that the introduction of FCA in the modeling process can generate innovative knowledge concepts, improve the recall rate of network resources, and can also integrate heterogeneous knowledge concepts, eliminate homographs, and improve the precision rate of network resources.

**Keywords:** Formal concept analysis knowledge integration ontology

## 1 引言

随着万维网的蓬勃发展,知识的传播更加迅速,知识的有效集成和共享对企业的长远发展起到关键性作用并能够提升企业竞争力。知识集成是组织获取知识所带来的收益并创造竞争优势的根本过程,这表明了知

识集成具有重要的战略意义和实践价值。知识集成模型主要是用来描述知识与知识之间、知识与人之间、知识与过程之间的关系,是对如何进行知识集成的一种模式或指导,主要的模型是基于认识论、本体论和知识范围的知识集成三维模型,即企业中的内部知识和外部知识在个体、团队和组织不

同层面的知识集成,该模型侧重于知识模型的扩展

## 2. 基于形式概念分析的知识集成建模

### 2.1 形式概念分析的基本概念

形式概念分析 (Formal Concept Analysis FCA) 是 Wille 提出的一种从形式背景进行数据分析和规则提取的强有力的工具。形式概念分析建立在数学的基础之上,对组成本体的概念、属性以及关系等用形式化的语境表述出来。一个概念就是最大限度地收集对集合中共同特点有帮助的元素,并且运用形式概念分析的方法,可以发现构造和展示由属性 (Attributes) 和对象 (Objects) 构成的概念 (Concept) 及其之间的关系。它的涵义是一个形式背景 (Formal Concept), 由一个三元组  $D = (A, B, R)$  构成, 其中  $A$  是对象的集合,  $B$  是属性的集合,  $R$  是  $A$  和  $B$  之间的一个二元关系,  $a R b$  表示  $a \in A$  与  $b \in B$  之间存在关系  $R$ 。

### 2.2 本体的概念

本体 (Ontology) 的概念最早源于亚里士多德 (公元前 384-322 年) 对事物存在本质的研究,在哲学上的定义为“对世界上客观存在物的系统地描述,即存在论”,是客观存在的一个系统的解释或说明,关心客观现实的抽象本质<sup>[1]</sup>。近十几年来,这个词被用在了计算机界,并在人工智能和、计算机语言以及数据库理论中起到越来越重要的作用。本体被哲学定义过的范畴,在现今被人工智能赋予了新的定义,这使得它被带入了信息科学中。信息科学对本体的定义是慢慢走向成熟的。1991 年 Neches 等人最早给出 Ontology 在信息科学中的定义:“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构成的规定这些词汇

外延规则的定义。”后来在信息系统、知识系统等领域,随着越来越多的人研究 Ontology, 产生了不同的定义。1993 年 Gruber 定义 Ontology 为“概念模型的明确的规范说明”。1997 年 Borst 进一步完善为“共享概念模型的形式化规范说明”。Studer 等人对上述两个定义进行了深入研究,认为 Ontology 是共享概念模型的明确的形式化规范说明,这也是目前对 Ontology 概念的统一看法。下面对本体进行举例来更好的理解本体。

### 2.3 形式概念分析与知识集成模型的结合点

形式概念分析与知识集成模型之间的结合点主要有两项: (1) 在形式概念分析中,概念格由概念层次结构组成,内涵和外延构成了概念;而知识集成模型是用来体现知识与知识之间关系的,可以利用本体对知识与知识之间的关系进行描述,进而形成公认的概念集合。(2) 本体将形式概念分析与知识集成模型紧密联系起来。本体体现的是共同认可的知识,反映的是相关领域中公认的概念集,本体的目标是捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的概念,并从不同层次的形式化模式上给出这些概念 (知识术语) 和概念之间相互关系的明确定义。知识集成模型的目的是描述知识与知识之间、知识与人之间、知识与过程之间的关系,因此知识集成模型的最佳描述工具就是本体。

### 2.4 基于形式概念分析的知识集成模型及框架

原始数据中的知识概念经过 5 个步骤的处理,减少了知识模型中概念节点之间的冗余性,同时也保证了它们之间的唯一性。这样在基于形式概念分析的知识集成建模过程中,不但进行了有效的知识创造和知识扩展,而且实现了对知识模型进行异构整合,以便更加方便地检索网络资源。基于形式概念分析的知识集成模型及其框架如图 1 所示:

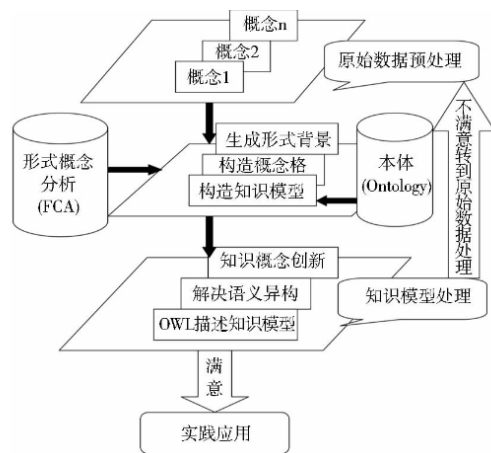


图 1 基于形式概念分析的知识集成模型

### 3 基于形式概念分析的知识集成模型实例研究

以网络资源中“计算工具：珠算的历史”和“算盘电，子算盘，定位算盘，速算盘类技术资料”两文为例，叙述基于形式概念分析的知识集成建模过程。由于知识集成建模过程较长，只列出其 5 个步骤的主要框架部分，具体情况如下

#### 3.1 原始数据预处理

算盘、乐音算盘、电子算盘、算盘（如意），其中算盘、乐音算盘、电子算盘、算盘（如意）之间是并列关系，同时它们和计算工具又是从属关系。其中算盘的属性是能计算、木制和有算珠；乐音算盘的属性是能计算、能奏乐、木制、金属制成和有算珠；电子算盘的属性是能计算、木制、金属制成和有算珠；算盘（如意）的属性是能推理。

#### 3.2 形式概念分析

形式概念分析一般分为三个步骤：（1）根据原数据处理的结果生成形式背景：通过对计算工具的历史进行分析，共有 4 个对象：算盘、乐音算盘、电子算盘和算盘（如意）。那么 5 行 7 列的交叉处是 X，表示对象具有的属性，基于 FCA 的知识集成模型实例生成的形式背景如表 1 所示：

表 1 基于 FCA 的知识集成模型实例的形式背景

	能计算	能奏乐	木制	金属制成	有算珠	能推理
算盘	X		X		X	
乐音算盘	X	X	X	X	X	
电子算盘	X		X	X	X	
算盘(如意)						X

（2）依据形式背景构造概念格：除了原有的 4 个对象以外，又生成了具有能计算、木制、有算珠和金属制成 4 个属性的新的概念节点，即图 2 中的 4 个空心的圆点，由基于 F C A 的知识集成模型实例生成的概念格如图 2 所示

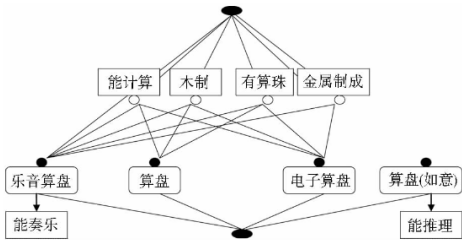


图 2 基于 FCA 的知识集成模型实例的概念格

（3）对概念格的底端节点、相关概念和顶端节点进行处理，去掉现实世界中不存在的底端节点，得到知识集成模型中的知识模型。从图 2 可以看出，从形式背景到概念格转换，最后到生成知识模型，总共构建了 9 个形式概念节点（其中一个根节点，4 个对象节点，4 个新的概念节点）。由于篇幅有限，只列出“根节点”、对象“乐音算盘”和新生成的节点“金属制的算盘”三个节点，它们分别是：①节点 1：（{概念：计算工具}，{属性：NULL}，{联系：根节点}，{公理：NULL}，{实例：乐音算盘、算盘、电子算盘、算盘（如意）}）。②节点 2：（{概念：乐音算盘}，{属性：能计算，能奏乐，木制，金属制成，有算珠}，{联系：节点 3 的子节点}，{公理：NULL}，{实例：乐音算盘}）。③节点 3：（{概念：金属制成的算盘}，{属性：金属制成的}，{联系：节点 1 的子节点，节点 2 的父节点}，{公理：NULL}，{实例：乐音算盘，电子算盘}）

#### 3.3 利用本体描述的知识模型进行知识概念创新

从图 2 可以看出，通过基于形式概念分

析的知识集成建模后得到了4个新的知识概念,分别是能计算的、木制的、有算珠的和金属制成的算盘,从而达到了知识概念创新的效果。

### 3.4 解决利用本体描述的知识模型中语义异构的问题

从图2可以看出,此实例中存在语义异构的情况,算盘和乐音算盘(如意)属于同形异义词。利用本领域的专业词汇工具书进行查找,确定知识模型中的这两个同形异义词中的算盘才是该领域(计算工具)真正需要的概念节点;然后在知识模型中保留算盘这个概念节点,并且把涵义不同的概念节点“乐音算盘(如意)”和它的属性从概念模型中删除,这样就在建模过程中减少了概念节点间语义含糊不清的问题。

### 3.5 使用OWL语言描述知识模型

把比较合理的知识模型用OWL语言描述出来:

```
<owl:Ontology rdf:
about="计算工具"/> // 本体头
部

<owl:Class rdf:ID=
"乐音算盘"> // 定义对象
<rdf:s:subClassOf>
// 乐音算盘是金属制的算盘的子节点
<owl:Class rdf:abo
ut="#金属制的算盘"/></rdf
s:subClassOf> // 结束定义
子节点
</owl:Class> // 结束定
义对象

<owl:ObjectProper
ty rdf:ID="金属制成的"> // 定
义属性
<rdf:s:domain rdf:s
ource="#算盘"/> // 用#算盘
引用算盘对象
</owl:ObjectPrope
rty> // 结束定义属性.....
</owl:Ontology> // 结束
定义本体计算工具
```

## 4 结语

在知识集成建模过程中引入形式概念分析的思想方法后,用户访问网络资源时可以获得两个主要优势:(1)可以了解已知知识概念并且产生创新知识概念,这样用户访问网络资源时可以通过产生的创新知识概念扩大自己的检索范围,提高了网络资源的查全率。(2)可以找到隐含知识概念及概念间的关系,进而对知识模型中的节点进行异构整合,消除冗余概念(即消除异形同义词),也就是可以在检索结果中消除同形异义词,提高了网络资源的查准率,同时又节省了用户排除干扰信息的时间。

### 参考文献

- [1] 王欣, 徐宝祥. 基于形式概念分析的知识集成建模研究[J]. 数据分析与知识发现, 2010, 26(4).
- [2] 刘萍, 彭小芳. 基于形式概念分析的词汇相似度计算[J]. 数据分析与知识发现, 2020, v. 4;No. 41(05):70-78

---

<sup>1</sup> 邓志鸿, 唐世渭, 张铭等。Ontology 研究综述, 北京大学学报(自然科学版), 2002, 38 730-738