

# 基于 FCA 的本体原型系统的设计与实现

张瑞玲<sup>1</sup>,徐红升<sup>1</sup>,沈夏炯<sup>2</sup>

ZHANG Rui-ling<sup>1</sup>,XU Hong-sheng<sup>1</sup>,SHEN Xia-jiong<sup>2</sup>

1.洛阳师范学院 信息技术学院,河南 洛阳 471022

2.河南大学 数据与知识工程研究所,河南 开封 475001

1.Academy of Information Technology, Luoyang Normal University, Luoyang, Henan 471022, China

2.Institute of Data and Knowledge Engineering, Henan University, Kaifeng, Henan 475001, China

E-mail: ruilingzhang@163.com

**ZHANG Rui-ling,XU Hong-sheng,SHEN Xia-jiong.Design and realization of ontology prototype system based on FCA. Computer Engineering and Applications,2008,44(19):80-83.**

**Abstract:** After analyzing key problems in ontology prototypal system,ontology prototypal system based on FCA (Formal Concept Analysis) is presented to make up these defects,and elucidates that main feature of the system is ontology building,merging and three-dimensional visualization display.This system proposes the determination of sub-concept lattice in order to reduce display.By examples,all of connotative ontology concepts were created from field,a superior-subordinate relationship was formed among concepts,and comparability among non-level class concepts and relativity between concepts and objects were gave by combine interrelated probability mode,thus inquiring efficiency among ontology was improved.

**Key words:** ontology;FCA(Formal Concept Analysis);prototype system;ontology merging;concept lattice;three-dimensional visualization display

**摘 要:**经过分析本体原型系统中存在的关键问题,提出基于 FCA 的本体原型系统来弥补缺陷,并阐明了系统的主要特色是采用 FCA 技术进行本体的构建、合并、三维可视化展现,其中为了在本体展现中简化概念格的显示,提出子概念格的判定方法。最后通过实例,从领域中生成了所有隐含的本体概念,形成了概念之间的上下级关系,结合相关性几率模式给出了非层次级概念之间的相似性以及概念与对象之间的相关度,从而提高了用户在本体中的查询效率。

**关键词:**本体;FCA;原型系统;本体合并;概念格;三维可视化

**DOI:**10.3778/j.issn.1002-8331.2008.19.023 **文章编号:**1002-8331(2008)19-0080-04 **文献标识码:**A **中图分类号:**TP18

Formal Concept Analysis(FCA)即形式概念分析是应用数学的一个分支<sup>[1]</sup>,它来源于哲学领域对概念的理解。概念格作为形式概念分析中核心的数据结构,从外延和内涵两方面对概念进行符号形式化描述,实现计算机可以理解的语义信息。

本体的概念在 10 年前被引入计算机科学中用以形式化表达知识,本体的定义由 Gruber<sup>[2]</sup>在 1993 年提出“本体是概念模型的明确规范说明”,1997 年 W.N.Borst<sup>[3]</sup>给出了另外一个定义“本体是共享概念模型的明确的形式化规范说明”,Studer<sup>[4]</sup>在上述两个定义的基础上提出“本体是共享概念模型的明确的形式化规范说明”。

目前已存在的本体原型系统存在以下不足:(1)对于本体的展现均采用二维形式,缺乏从多角度立体查看本体的能力;(2)对于复杂的本体,展现图也相应地庞大,这将淹没有用的信息;(3)本体概念间的相似度分析粗略。本文提出的基于 FCA

的本体原型系统能弥补这些缺陷,它不仅能为研究者提供方便的研究平台,也能为新方法、新技术的开发提供实验平台。同时,结合新的研究内容,它能很快地发展成一个 FCA 应用的原型系统,具有理论和应用相结合的特点。目前,对 FCA 软件工具的研究与开发主要集中在国外<sup>[5-8]</sup>,内容主要以集成 FCA 理论的基本算法、提供流行工具的接口为主<sup>[9]</sup>。

## 1 系统分析与设计

### 1.1 设计思想

基于 FCA 的本体原型系统首先让用户进行初始本体的录入与编辑,然后将初始本体转化成单值形式背景,再利用 Godin<sup>[10]</sup>造格算法来构造概念格。对于构造好的格,将每个形式概念用唯一的属性来标示,这样把概念格转换成本体。由于知识的共享与重用,需要将同领域中的两个本体进行合并,将构

**基金项目:**国家自然科学基金(the National Natural Science Foundation of China under Grant No.60575035);河南省重点科技攻关项目(No.072102310088);

河南省教育厅自然科学研究资助项目(No.2006520010)。

**作者简介:**张瑞玲(1964-),女,副教授,主要研究领域为:概念格、本体及粗糙集;徐红升(1979-),男,助教,主要研究领域为形式概念分析、知识发现和本体;沈夏炯(1963-),男,博士,副教授,研究方向是软件工程、知识发现。

**收稿日期:**2008-01-07 **修回日期:**2008-04-08

建本体的形式背景进行合并处理,最后针对合并后的形式背景,运用 Godin 造格算法进行构造,形成合并后的概念格,然后转换成对应的本体。

对于展现本体,一是基于简化概念格的三维可视化形式来展现,主要针对子概念格的可视化互操作,识别出格中的子概念格,并将其折叠以简化概念格的显示;二是结合概念间的相关性几率模式<sup>[1]</sup>,给出本体中非层次概念之间的相似关系以及属性概念与对象之间的相关度。

## 1.2 系统主要功能

该系统开发环境为 Windows XP 操作系统、后台数据库为 MS SQL Server,前台开发工具为 MS Visual Studio.Net 开发套件。系统主要实现的功能有:

(1)初始本体的编辑:用户可以按照领域内类(Classes)和属性(Slots)的对应关系,以及相应属性的值(Facets)来编辑初始的本体。

(2)初始本体的转换:将初始本体转化成由类和属性组成的二维表,表中的数值代表的是相应属性的值,并将其转化成单值形式背景。

(3)本体的构造:该模块包括概念格的构造和概念格到本体的转换。

(4)本体的合并:根据概念格的合并理论,将两个形式背景进行相应的处理,形成合并后的形式背景。

(5)本体的展现:一种是基于简化概念格的三维可视化形式;另外一种是对具有非层次关系概念之间的相似性以及属性概念与对象之间的相关度进行描述。

## 2 主要内容

### 2.1 子概念格的判定

**定义 1** 设  $(A, \leq)$  为偏序集,对于  $B \subseteq A$ ,如有  $a, b \in A$ ,且对任意  $x \in B$ ,都满足  $x \leq a$ ,则称  $a$  为子集  $B$  的上界。同理,若对任意  $x \in B$ ,都满足  $b \leq x$ ,则称  $b$  为子集  $B$  的下界。

**定义 2** 设  $(A, \leq)$  为偏序集,  $B \subseteq A$ ,  $a$  为  $B$  的任一上界,若对  $B$  的所有上界  $y$  均有  $a \leq y$ ,则称  $a$  为  $B$  的最小上界(上确界 supremum),记为  $\sup(B)$ 。同样,若  $b$  为  $B$  的任一下界,若对  $B$  的所有下界  $z$  均有  $z \leq b$ ,则称  $b$  为  $B$  的最大下界(下确界 infimum),记为  $\inf(B)$ 。

**定义 3** 若  $(M, \leq)$  是一个偏序集,  $a, b, c, d$  是  $M$  中的元素且  $b < c$ 。则集合  $[b, c] := \{x \in M | b \leq x \leq c\}$  称为间隔(interval),集合  $[a] := \{x \in M | x \leq a\}$  称为主理想(principal ideal),集合  $[d] := \{x \in M | x \geq d\}$  称为主滤子(principal filter)。并且,  $a < b \Leftrightarrow a < b$  且  $[a, b] = \{a, b\}$ 。

**定理 1** 已知一个完全格  $(V, \leq)$ ,假设格中任意两个元素  $u, l$ ,满足  $l < u$ ,  $u$  的主要理想  $C = [u]$ ,  $l$  的主要滤子  $D = [l]$ ,令集合  $M = C \cap D$ ,  $A = D - \{v \in V | l < v\}$ ,  $B = C - \{v \in V | v < l\}$  如果  $M = A = B$ ,那么格  $V$  中必存在一个间隔  $\{v \in V | l \leq v \leq u\}$ 。

**证明** 显然在完全格  $(V, \leq)$  中,求集合  $\{v \in V | l < v\}$  的补集为  $E = C \cup \{v \in V | \text{不存在 } u \leq v \text{ 或 } v \leq u\}$ ,所以  $A = D - \{v \in V | l < v\} = D \cap E$ ,如果  $M = A$ ,那么  $C \cap D = D \cap E = D \cap (C \cup \{v \in V | \text{不存在 } u \leq v \text{ 或 } v \leq u\})$ ,根据集合的  $\cup$  操作符合分配规律,  $C \cap D = (D \cap C) \cup (D \cap \{v \in V | \text{不存在 } u \leq v \text{ 或 } v \leq u\})$ ,可以得到,  $D \cap \{v \in V | \text{不存在 } u \leq v \text{ 或 } v \leq u\} = \phi$ ,即  $\{v \in V | l \leq v\} \cap \{v \in V | \text{不存在 } u \leq v \text{ 或 } v \leq u\} = \phi$ ,由这个结果可以知道,格  $V$  中元素  $l$  的所有父元素(包括元素  $l$ )集合中不包含与元素  $u$  无偏序关系的元素,即得结论 1:  $l$  的所有父元素都与  $u$  有偏序关系,得到结论 1:  $\{v \in V | l \leq v, v \leq u \text{ 或 } u \leq v\}$ ;

同理,如果  $M = B$  可以推出  $u$  的所有子元素都与  $l$  有偏序关系,得到结论 2:  $\{v \in V | l \leq v, v \leq l \text{ 或 } l \leq v\}$ ;

综上所述,根据结论 1 和结论 2,格  $V$  中的任意元素  $v$  如果  $l \leq v$  并且  $v \leq u$ ,那么  $v$  必有  $l \leq v \leq u$ ,所以格  $V$  中必存在一个间隔  $\{v \in V | l \leq v \leq u\}$ ,定理得证。

**定理 2** 已知一个完全格  $(V, \leq)$ ,如果存在一个间隔  $W \{v \in V | l \leq v \leq u\}$ ,那么该间隔  $W$  就是格  $V$  的完全子格。

**证明** 根据定义 1,2,对于一个完全格  $(V, \leq)$ ,显然  $W \subseteq V$ ,对于任意  $w \in W$ ,如果  $V$  中存在某一元素  $s$ ,都满足  $s \leq w$ ,那么  $s$  为集合  $W$  的下界。由  $l \leq w \leq u$  知,满足  $s \leq l \leq w$ ,所以在集合  $W$  的所有下界中最大的元素为  $l$  即  $\bigwedge W = l$ ;同理,在集合  $W$  的所有上界中最小的元素为  $u$  即  $\bigvee W = u$ ,如果  $T \subseteq W$ ,则有  $T \subseteq V$ ,同理,  $\bigvee T = \bigvee W = u$ ,并且  $\bigwedge T = \bigwedge W = l$ ,显然  $u \in W, l \in W$ ,可见,  $T \subseteq W \Rightarrow \bigvee T \in W$ ,并且  $T \subseteq W \Rightarrow \bigwedge T \in W$ ,根据定理 2 可知,该间隔  $W \{v \in V | l \leq v \leq u\}$  就是格  $V$  的完全子格,定理得证。

子概念格的判定算法:  
输入:概念格  $L(K)$   
输出:子概念格  $L(K)_{\text{child}}$

(1)BEGIN  
(2)FOR 每一个格节点  $upper \in L(k)$   
DO  
(3) FOR  $low \in L(k)$  并且至少存在  $u \in L(k)$  使  $low \leq u \leq upper$   
DO

(4) 寻找节点  $low$  所有父节点得到集合  $lowfather(low)$ ,包括节点  $low$ ;

(5) 寻找节点  $low$  所有子节点得到集合  $lowchild(low)$ ,包括节点  $low$ ;

(6) 寻找节点  $upper$  的所有父节点得到集合  $upperfather(upper)$ ,包括节点  $upper$ ;

(7) 寻找节点  $upper$  的所有子节点得到集合  $upperchild(upper)$ ,包括节点  $upper$ ;

(8)  $W = upperchild(upper) \cap lowfather(low)$ ;

(9)  $lowfather(low)$  减去  $upperfather(upper)$ ,在差中添加上节点  $upper$  得到集合  $A$ ;

(10)  $upperchild(upper)$  减去  $lowchild(low)$ ,在差中添加上节点  $low$ ,得到集合  $B$ ;

(11)IF  $W$  等于  $A$  THEN

(12) IF  $W$  等于  $B$  THEN

(13)  $L(K)_{\text{child}} = W$ ; //保存间隔为以后折叠作准备

(14) return  $L(K)_{\text{child}}$ ;

(15) EDNIF

(16) ENDIF

(17) ENDFOR

(18) ENDFOR

(19) END

### 2.2 本体的合并

本体合并(Ontology merging)指将领域内已经存在的几个本体并在一起,消除重叠的和不协调的部分。把不同的本体合并为一个具有更合理概念体系和更强的知识表达能力的本体。本体的合并模块如图 1 所示。

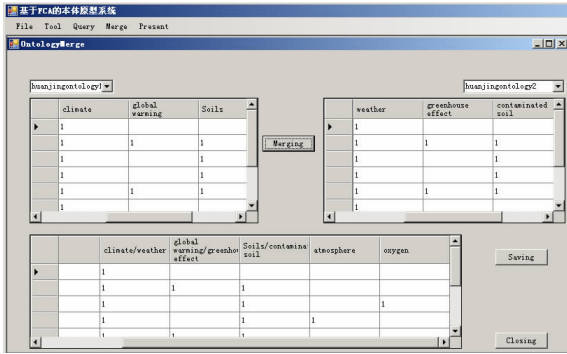


图 1 本体的合并

据构建本体的领域不同有下面两种情况: 同域本体的合并、交叉领域本体的合并。

#### (1)同域本体的合并

如果两个合并的本体来自同一个领域,则它们是内涵一致的形式背景。其合并会有以下三种情况:在两个形式背景中可能存在 is-a 关系的两个本体概念,例如两个概念 m、n,存在 m is-a n,则在并置后的形式背景中只出现 n。即:如果(g,m) ∈ I 并且 m is-a n 意味着(g;n) ∈ I;如果两个形式背景中存在两个名称不同但意义相同的本体概念如 a、b,那么可以在并置后的形式背景中以“ab”来表示这个概念,意思就是 a 或者 b;如果两个形式背景中存在两个名称、含义均不同,但它们从领域文档中的分布一样如 p、q,那么在并置后的形式背景中以 p(q)来表示,意思是 p 且 q。

#### (2)交叉领域本体合并

如果合并的本体来自不同的领域,这样的形式背景是内涵和外延都有交集的形式背景。对这样的两个形式背景在进行合并时,要同时采取并置和叠置处理,并且保证合并后的形式背景中对象和属性的唯一性。

对于本体合并的评价一直都是人们讨论的问题,该原型系统把基于概念格合并原理所合并出的最终本体与 Noy 和 Musen 提出的基于标准信息抽取技术所合并的本体<sup>[2]</sup>,进行分析比较,从而评估合并的效果。

### 2.3 本体的三维可视化

本体的可视化给人们提供了直观的分析与观察知识单元的内在联系的方法,该系统采用组件技术自动生成三维布局图,并设置相应的属性、方法提供对该组件的操作,由用户根据情况进行各种变换以及调节每个结点在其所在层中的布局,并将调整结果随时保存,本体的三维可视化展现如图 2 所示。

系统根据其对应的概念格展现,将本体在格中表现出来。关于概念格的展现,本系统采用基于三维布局的概念格展现方式<sup>[3]</sup>。对于复杂的格结构,为了简化概念格的三维显示,本系统针对子概念格的可视化互操作,提出关于子概念格的判定定理以及折叠与打开算法。

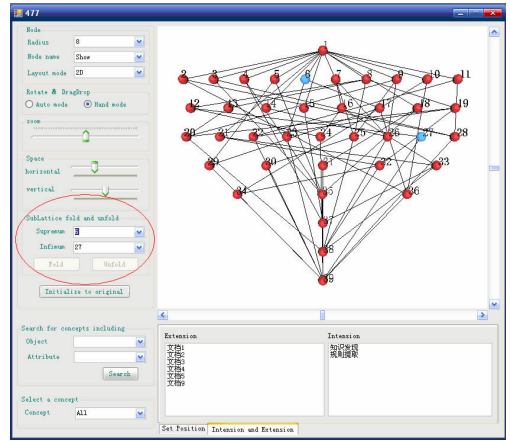


图 2 本体的三维可视化展现

### 3 应用分析

本系统结合概念间的相关性几率模式<sup>[4]</sup>,根据以下计算公式来表达本体中具有非层次关系概念之间的相关性及属性概念与对象之间的相关度,计算公式如下:

$$f_{jk} = relevancy(T_j, T_k) = \frac{\sum_{i=1}^n d_{ijk}}{\sum_{i=1}^n d_{ij}} \times WeightingFactor(T_k) \quad (1)$$

$$d_{ijk} = tf_{ijk} \times lg\left(\frac{N}{df_{jk}} \times w_j\right) \quad (2)$$

$$d_{ij} = tf_{ij} \times lg\left(\frac{N}{df_j} \times w_j\right) \quad (3)$$

$$WeightingFactor(T_k) = \frac{lg \frac{N}{df_k}}{lg N} \quad (4)$$

公式(1)描述两个词汇之间的关联程度,而且关联程度具有方向性,以不同词汇为中心所计算出来的相关性不相同,公式(1)可拆分为式(2)、式(3)、式(4)三个子公式。对公式中的变量说明如下:N 表示文件的总数,w<sub>j</sub> 表示不存在词汇 j 的文件次数。在式公(2)中,d<sub>ijk</sub> 由词 T<sub>k</sub> 及 T<sub>j</sub> 所同时出现的次数及不存在某文件中的文件次数来决定,tf<sub>ijk</sub> 表示词汇 j 与 k 共同出现在文件 i 的次数,df<sub>jk</sub> 则表示词汇 j 与 k 同时出现的文件总数;在式公(3)中,d<sub>ij</sub> 由词汇 j 对于文件 i 的相关性决定,tf<sub>ij</sub> 即文件 i 中词汇 j 出现的次数,df<sub>j</sub> 即所有存在词汇 j 的文件总数。当两个词汇具有较高的关联度时,在同一份文件中它们出现的次数也应该很高且集中在某些特定的文件之中;在公式公(4)中,WeightingFactor(T<sub>k</sub>)表示词 T<sub>k</sub> 对文件的特殊性,词汇 T<sub>k</sub> 越普遍时,则 WeightingFactor(T<sub>k</sub>)的值就会越小。

本体的相关性展现如图 3,图中给出了本体中的所有概念,当用户选择其中一个概念时,将显示该概念的父、子概念,相关概念的权重,以及与此概念相关的文档。在图 3 中,当用户选择了概念“知识发现”后,在父概念列表中显示“规则提取”,子概念列表显示“信息检索”,同时与“知识发现”相关的概念有:“关联规则”、“概念”、“Web 服务”、“本体”、“服务组合”等,从权重关系中可以看出概念“关联规则”相关度最大为 0.234 965,表明与概念“知识发现”具有非层次关系的概念中,最相关是“关联规则”。另外我们也可以查询到与“知识发现”相关的文件有:“文档 3”、“文档 1”、“文档 5”等,其中与“文档 3”的相关性

权重最大为 36.164 78。由此,该系统不仅给出与查询概念有上下级关系的概念,还可以为用户推荐最相关的文档,并且可以让用户针对最相关的概念继续寻找,这样明显提高了本体的查找效率。



图3 本体概念的相关性展现图

#### 4 结束语

该系统的主要特色是采用 FCA 技术进行本体的构建、三维可视化展现与合并,并结合概念相关性几率模式来表达本体,从而有效地弥补了已存在本体原型系统的不足,为本体的开发提供一种工具。另外,通过实例,说明了由本系统所构建的本体在查询方面的优势,使用户所查询得到的结果具有语义上的相关性,从而提高了用户在本体中的查询效率。

总之,采用 FCA 技术为本体的构建、合并提供了一种统一的应用技术。在今后的 FCA 研究中,增进 FCA 在知识表达过程中的表达能力是其中的一个研究方向;在理论和应用上加强 FCA 与本体的结合是今后的研究努力方向。

#### 参考文献:

- [1] Ganter B,Wille R.Formal concept analysis:mathematical foundations[M].Berlin:Springer Verlag,1999.
- [2] Gruber T.R.A translation approach to portable ontology specifica-

tion KSL[R].1993.

- [3] Borst W N.Construction of engineering ontologies for knowledge sharing and reuse[D].The Netherlands:University of Twente,1997.
- [4] Studer R,Benjamins V R,Fensel D.knowledge engineer principles and methods[J].Data and Knowledge Engineering,1998,25(1/2):161-97.
- [5] Duquenne V,Chabert C,Cherfouh,et al.Structuration of phenotypes/genotypes through galois lattices and implications[C]//Nguifo E,Liquière M,Duquenne V.CLKDD'01,Concept Lattices-based Theory,Methods and Tools for Knowledge Discovery in Databases,2001,42:21-32.
- [6] Burmeister P.Formal concept analysis with conImp:introduction to the basic features TU-Darmstadt[R].Germany,Darmstadt University,1996.
- [7] Correia J H,Kaiser T.B.A mathematical model for TOSCANA-systems:conceptual data systems[C]//Proceedings of Second International Conference on Formal Concept Analysis,Sydney,Australia,2004:39-46.
- [8] Becker II P.Numerical analysis in conceptual systems with toscana[J]//Proceedings of Second International Conference on Formal Concept Analysis,ICFCA 2004,Sydney,Australia,2004:96-103.
- [9] Tilley T.Tool support for FCA [C]//Proceedings of Second International Conference on Formal Concept Analysis,Sydney,Australia,2004:104-111.
- [10] Godin R,Missaoui K,Alaoui H.Incremental concept formation algorithms based on Galois(concept) lattices[J].Computational Intelligence,1995,11(2):246-267.
- [11] 徐红生,申夏炯,刘宗田.基于形式概念的语义网本体的构建与展现[J].计算机科学,2007,32(2):171-173.
- [12] Noy N F,Musen M A.Prompt:algorithm and tool for automated ontology merging and alignment[C]//Proc AAAI'00,2000:450-455.
- [13] 马骏,沈夏炯,刘宗田.基于三维空间的概念格自动布局研究[J].计算机科学,2006,33(5):244-246.
- [14] Kang S H,Huh W,Lee S,et al.Automatic classification of WWW documents using a neural network[C]//Intl Conf on Production Research,Bangkok,Thailand,2000.

(上接 71 页)

- [4] Sanders W H,Meyer J F.A unified approach for specifying measures of performance,dependability,and performability[C]//Avizienis A,Laprie J.Dependable Computing for Critical Applications,Dependable Computing and Fault-Tolerant Systems.Berlin:Springer-Verlag,1991,4:215-237.
- [5] PERFORM performability engineering research group,möbius user manual,Version 1.8.0-draft[R/OL].USA:University of Illinois,2005.

<http://www.mobius.uiuc.edu/download.html>.

- [6] Sanders W H.Integrated frameworks for multi-level and multi-formalism modeling[C]//Proceedings of the 8th International Workshop on Petri Nets and Performance Models,Zaragoza,Spain,1999.
- [7] Stillman A J.Model composition within the mobius modeling framework[D].USA:University of Illinois,1999.
- [8] Braam P J.The lustre storage architecture[EB/OL].(2004-11).<http://www.clusterfs.com>.