

个性化本体的知识查询研究^{*}

韩永国^{1,2} 孙世新¹

(电子科技大学计算机科学与工程学院 成都 610054)¹

(西南科技大学计算机科学与技术学院 绵阳 621010)²

摘要 在分布式系统中,既有共享概念的领域本体,又存在个体的知识组织个性,因此个性化本体的研究具有重要的现实意义。本文基于开放系统假设,定义个体的外现类及个体之间的相识关系,使用 Web 本体语言表示个性化本体的类描述和类公理。提出一个基于消解原理的个性化本体网络的知识查询递归算法,该算法不依赖注册服务器和分布式哈希表,具有较强的动态适应性。以一个(教)师(学)生本体网络的知识查询为例,讨论了个性化本体网络及其知识查询算法的实现。

关键词 个性化本体,知识查询,消解原理,知识工程,人工智能

Study on Knowledge Query in Personalized Ontologies

HAN Yong-Guo^{1,2} SUN Shi-Xin¹

(Computer Science and Engineering College, University of Electronic Science and Technology of China, Chengdu 610054)¹

(Computer Science and Technology College, Southwest University of Science and Technology, Mianyang 621010)²

Abstract There is not only domain ontology of shared concepts but also personalized organization of knowledge in distributed system. Study on personalized ontology is of practical significance. Based on Open World Assumption of distributed system, extensional classes of peer and acquaintance relations between peers are defined. Class description and class axioms of personalized ontology are expressed by Web Ontology Language. A recursive algorithm of querying knowledge in personalized ontology network is proposed, with exploitation of principle of resolution. The algorithm has intrinsic adaptability to dynamic situation as it is not dependent on central server or distributed hash table. A peer-to-peer ontology network of teachers and students is given and the knowledge querying in it is illustrated.

Keywords Personalized ontology, Knowledge querying, Principle of resolution, Knowledge engineering, Artificial intelligence

1 引言

本体(Ontology)是可共享概念模型的形式化规格说明,其基本组成元素包含概念、概念的属性、概念的关系、公理,以及属性和关系的约束(也可以看成公理)。对于一个特定领域的概念(称为领域本体),要求概念准确(Conceptualization)、能够共享(Share),同时要求明确(Explicit)、形式化(Formal),需要领域专家、计算机科学家、软件工程师等的广泛协作完成,工作十分艰巨。但本体作为一种能在语义和知识层次上描述信息系统的概念建模工具,自提出以来,引起了大学和科研机构的普遍关注,并在许多领域得到广泛应用。在知识管理和利用方面,利用本体模型实现语义 Web^[1]、医学知识库^[2]等。在分布式协同中,也广泛使用本体方法,例如 MAS^[3]和工作流^[4]。在网格研究中,利用本体方法建立知识网络体系结构^[5],并应用本体模型组织人类语言及其机器翻译,如 WordNet^[6]、SENSUS^[7]。

在分布式系统中实现领域本体时,一般采用注册服务器^[8]、分布式哈希表^[9]等方法。在现实的分布式系统中,既有共享概念的领域本体,又存在个体的知识组织个性,因此个性化本体的研究也受到极大关注^[10]。

本文基于“开放系统假设”,定义个体的“外现类”及个体之间的相识关系,使用 OWL DL^[11]表示类描述和类公理。提出一个基于消解原理的个性化本体网络的知识查询递归算法,并以一个(教)师(学)生本体网络的知识查询为例,实现了其个性化本体及其知识查询算法。

2 个性化本体网络

2.1 个性化本体

设一个师生网络是一个 P2P 网络 P2PNet, $P2PNet = \{P_i\}_{i \in 1, \dots, n}$, 对于每个教师和学生实体 P_i , 按自己的原则定义本体 P_i , 本体 P_i 由一组类描述及其类之间的关系构成。类描述包括全类(universal class)、空类(empty class)、原子类(atomic class), 以及原子类的合取(conjunction)、析取(disjunction)和否定(negation)。类描述及其 OWL DL^[11] (Web 本体语言)表示如表 1。

设原子类的类名是唯一的,用符号 $P_i:A$ 表示实体 P_i 的本体中的原子类 A 。实体 P_i 的词汇集(vocabulary)是其原子类名称的集合。类之间的关系包括:等价(equivalence)、包含(inclusion)、不相交(disjointness)。从而,类公理(class axioms)如表 2。

^{*} 国家高技术研究发展计划 863 项目(2003AA116060)、四川省科技厅攻关计划(05GG07-006)。韩永国 博士生,副教授,主要研究领域为网络分布计算、知识工程、协同技术;孙世新 教授,博士生导师。

表1 类描述及其 OWL 表示

类描述	OWL 表示
全类(\top)	Thing
空类(\perp)	Nothing
原子类	classID
合取(\cup)	intersectionOf
析取(\cap)	unionOf
否定(\neg)	complmentOf

表2 类公理

类公理	OWL 描述
(模态)全部(\equiv)	Class($P; A$ complete C)
(模态)部分(\subseteq)	Class($P; A$ partial C)
等价(\equiv)	equivalentClass
包含($C_1 \subseteq C_2$)	rdfs:subClassOf
不相交($C_1 \cap C_2 \equiv \perp$)	disjointWith

在实际系统中,一个实体往往只知道其自身或相识邻居的结构、性质和行为,无法获得或预期系统其他部分的结构、性质和行为(或代价很高),这称为“开放世界假设”OWA (Open World Assumption)。反之,如果实体可以获得或预期整个系统的结构、性质和行为,称为“封闭世界假设”CWA (Closed World Assumption)。

在开放世界假设下,一个实体自身的组织一般采用面向对象的思想,即只允许外部实体访问自己的部分信息。在 P2PNet 中,以原子性“外现类”定义存储在本地并允许外部实体访问的数据,记为 $P:ViewA$,其 OWL 表示为 $rdfs:subClassOf$ 。

2.2 个性化本体网络

在 P2PNet 中,一个新实体加入网络时,声明自身与相识本体之间的映射。如果一个实体需要查询知识,则它使用自身本体进行查询,并将查询提交给相识实体。如果被讯问实体可以回答,返回结果实例,否则通过映射传递到其相识实体。应答在实体之间传递时,可以根据映射进行转换。为此,实体之间的映射关系极为重要。

在 $P2PNet = \{P_i\}_{i \in 1 \dots n}$ 中,设实体 P_i 的词汇集是 V_{P_i} , 令:

$$V = \bigcup_{i \in 1 \dots n} V_{P_i}$$

相识关系(映射)定义为:

$$ACQ = \{(v, P_i, P_j) \mid (v, P_i, P_j) \subseteq V \times P \times P, v \in P_i \cap P_j, i \neq j\}$$

(v, P_i, P_j) 表示实体 P_i 和 P_j 相识,并共享变量(词汇) v 。

3 知识协同查询算法

设三元组 (l, P, c) ,其中 P 是实体, l 是文字,表示词汇集 V 中的原子变量及其否定; c 是子句,由若干文字的析取构成。算法从某个初始实体三元组 (l_0, P_0, c_0) 开始,执行时生成三元组历史序列 $h = \{(l_0, P_0, c_0), \dots, (l_n, P_n, c_n)\}$ 。如果查询子句 c_0 不可分解,则 $l_0 = c_0, l_1 = c_0$;如果可以分解,则分解为 $\neg l_1 \vee c_0$,其中 l_1 传递到相识实体 P_1 。

对于一个实体 P 的子句 c ,令 $S(c)$ 表示 P 与相识实体的共享变量的析取, $L(c)$ 表示非共享变量的析取。 $S(c) = \square$ (空子句)表示 c 中不包含共享变量。

在实体内部,采用消解原理 (principle of resolution) 算

法,它是由 J. A. Robinson 于 1965 年首先提出,是一个相当有效的谓词逻辑机械化推理方法。对于一个文字 q ,用 $Res(q, P)$ 表示 $P \cup \{q\}$ 的消解式。

由开放世界假设,定义实体 P 的兴趣变量集合 $TV_P = \{v \mid (v, P, P_j) \in ACQ, \text{且 } v \in TV_{P_j}\}$,将查询结果分为本地查询结果 LR 和其他查询结果 RR,LR 只与被直接请求的实体有关。最终结果由 LR 和 RR 组合获得。组合方式按文字 l 或子句 c 分类组合,分别记为 $classify_l()$ 或 $classify_c()$ 。算法如图 1。

```

RecursiveQuerying (q, P)
{
    return Querying(q, {P}, NIL)
}

Querying (q, {P}, h)
{
    if (  $\exists P \in SP$  且  $q \in P$  ) or (  $\forall P \in SP$  (  $q, P, \neg \in h$  ) ) return NIL
    else if (  $q, \neg \in h$  ) return {  $\square$  }
    else {  $\forall P \in SP$  Loc(P)  $\leftarrow$  {  $q$  }  $\cup$  Res (  $q, P$  ) }
    if (  $\exists P \in SP$  且  $\square \in Loc(P)$  ) return {  $\square$  }
    else {  $\forall P \in SP$  Loc(P)  $\leftarrow$  {  $c \in Loc(P) \mid L(c) \in TV_P$  } }
    if (  $\forall P \in SP$  且  $\forall c \in Loc(P)$   $S(c) = \square$  ) return  $\bigcup_{P \in SP} Loc(P)$ 
    else {
        LR  $\leftarrow$   $\bigcup_{P \in SP} \{ c \in Loc(P) \mid S(c) \in TV_P \}$ 
        foreach  $P \in SP$  且  $c \in Loc(P)$  且  $S(c) \neq \square$ 
        {
            if (  $\neg q \vee c \in P$  )  $P \leftarrow P \setminus \{ \neg q \vee c \}$ 
            foreach  $l \in S(c)$  Ans (  $l$  )  $\leftarrow$  Querying (  $l, ACQ(LP), \{(q, P, c)\}$  ,  $h$  )
            RR  $\leftarrow$  (  $classify_l(Ans(l))$  )  $\cup$   $classify_c(L(c))$ 
            LR  $\leftarrow$  LR  $\cup$  RR
        }
        return LR
    }
}

```

图1 知识协同查询算法

4 一个师生个性化本体网络的实现

4.1 描述

以教师 Alice、Bob 和学生 Chris、Dora 本体组成的网络为例(图 2)。教师 Alice 是分布式计算专家,她把自己的知识组织为 Alice:G,其中课堂教授知识为 Alice:R, Alice:R \subseteq Alice:G。Alice:R 分为三个互不相交的层次: Alice:B、Alice:M、Alice:D,分别对应于本科、硕士、博士。

$$Alice:R \equiv Alice:B \cup Alice:M \cup Alice:D$$

$$Alice:B \cap Alice:M \equiv \perp, Alice:B \cap Alice:D \equiv \perp, Alice:M \cap Alice:D \equiv \perp$$

Alice 的网络知识记为 Alice:N,操作系统知识记为 Alice:O。Alice:O 分为并发 Alice:C、调度 Alice:S、存储管理 Alice:V。

$$Alice:C \subseteq Alice:O, Alice:S \subseteq Alice:O, Alice:V \subseteq Alice:O$$

Alice 存储在本地并允许访问的知识表示为:

$$Alice:ViewM \subseteq Alice:M, Alice:ViewC \subseteq Alice:C, Alice:ViewS \subseteq Alice:S, Alice:ViewV \subseteq Alice:V, Alice:ViewN \subseteq Alice:N$$

Bob 是一名操作系统教师,他把自己的知识描述为 Bob:U,其中关于 UNIX 方面的知识为 Bob:Q。Bob 允许存储在他本地的所有知识均可以被访问,即

$$Bob:ViewU \subseteq Bob:U, Bob:ViewQ \subseteq Bob:Q$$

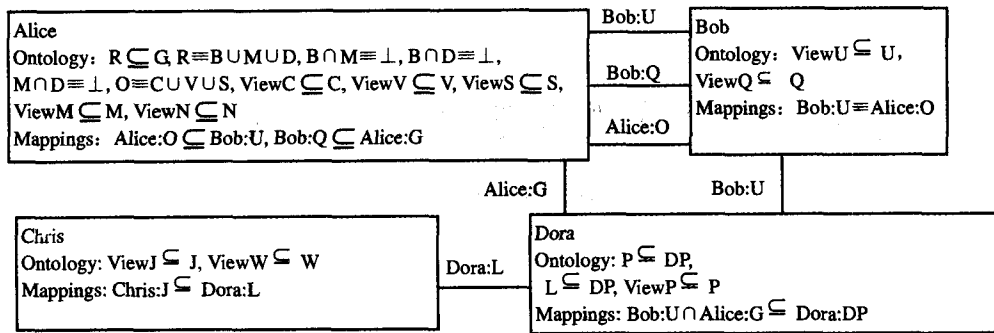


图2 师生 P2P 网络

学生 Chris 把自己存储在本地的 Java 程序设计知识组织为 Chris:J, Web 设计知识组织为 Chris:W, 且均可以被访问。即

$\text{Chris:View}J \subseteq \text{Chris:J}, \text{Chris:View}W \subseteq \text{Chris:W}$

学生 Dora 是分布式处理方向的研究生, 把自己的知识组织为 Dora:DP, 包括算法分析知识 Dora:P 和程序设计知识 Dora:L, 其中只有算法分析知识可以被访问, 即

$\text{Dora:View}P \subseteq \text{Dora:P}$

Alice, Bob, Chris 和 Dora 相互了解部分: Alice 非常欣赏 Bob 在操作系统方面的才能, 认为自己在操作系统的知识 Bob 均掌握, 即 $\text{Alice}:O \subseteq \text{Bob}:U$, 并且认为 Bob 关于 UNIX 的知识很广泛, 即 $\text{Bob}:Q \subseteq \text{Alice}:G$ 。

Bob 知道自己的操作系统知识与 Alice 的操作系统知识相同, 即 $\text{Bob}:U \equiv \text{Alice}:O$ 。

Chris 知道自己的 Java 程序设计知识是 Dora 程序设计知识的一部分, 即 $\text{Chris:J} \subseteq \text{Dora:L}$ 。

Dora 期望获得 Bob 关于分布式操作系统方面的知识, 即 $\text{Bob}:U \cap \text{Alice}:G \subseteq \text{Dora:DP}$ 。

4.2 推理

假设一个用户通过实体 Dora 查询分布式处理的知识 (Dora:DP), 实体之间的协同推理如下。

(1.1) 根据 $\text{Dora:P} \subseteq \text{Dora:DP}$ 和 $\text{Dora:View}P \subseteq \text{Dora:P}$, 实体 Dora 给出本地结果 $\text{Dora:View}P$ 。

(1.2) 由于 $\text{Dora:L} \subseteq \text{Dora:DP}$, 且 Dora 与 Chris 共享 Dora:L , 查询扩展到相识本体 Chris。在 Chris 中, 根据映射 $\text{Chris:J} \subseteq \text{Dora:L}$, 以及 $\text{Chris:View}J \subseteq \text{Chris:J}$, 相识本体 Chris 给出结果 $\text{Chris:View}J$ 。

(1.3) 根据映射 $\text{Bob}:U \cap \text{Alice}:G \subseteq \text{Dora:DP}$, 且 Dora 与 Alice 共享 $\text{Alice}:G$, 与 Bob 共享 $\text{Bob}:U$, 查询扩展到相识本体 Alice 和 Bob。Alice 和 Bob 分别进行并行查询。

(2.1) 在 Bob 中, 由于 $\text{Bob:View}U \subseteq \text{Bob}:U$, Bob 向 Dora 返回 $\text{Bob:View}U$ 。

(2.2) 在 Bob 中, 由于存在映射 $\text{Bob}:U \equiv \text{Alice}:O$, Bob 将查询扩展到 Alice, Alice 向 Bob 返回结果 $\text{Alice:View}C \cup \text{Alice:View}V \cup \text{Alice:View}S$ 。Bob 也将该结果返回给 Dora。

(3.1) 在 Alice 中, 使用 $\text{Alice:R} \subseteq \text{Alice:G}$, $\text{Alice:R} \equiv \text{Alice:BU} \cup \text{Alice:MU} \cup \text{Alice:D}$, 以及 $\text{Alice:View}M \subseteq \text{Alice:M}$, 因此 Alice 向 Dora 返回 $\text{Alice:View}M$ 。

(3.2) 在 Alice 中, 由于映射 $\text{Bob}:Q \subseteq \text{Alice:G}$, Alice 从 Bob 处获得 $\text{Bob:View}Q$ 。

(4.1) Dora 从 Bob 处获得的两个结果 (2.1) 和 (2.2), 与从 Alice 处获得的两个结果 (3.1) 和 (3.2), 组合获得如下四个结果并发给用户 ($\text{Bob}:U \cap \text{Alice}:G \subseteq \text{Dora:DP}$):

$\text{Alice:View}M \cap \text{Bob:View}U,$
 $\text{Alice:View}M \cap (\text{Alice:View}C \cup \text{Alice:View}V \cup \text{Alice:View}S),$
 $\text{Bob:View}Q \cap \text{Bob:View}U,$
 $\text{Bob:View}Q \cap (\text{Alice:View}C \cup \text{Alice:View}V \cup \text{Alice:View}S)。$

结论 当前本体研究的热点包括语义匹配^[12]、自动概念学习^[13]、本体集成^[14]、本体与规则的集成^[15]等。本文基于“开放系统假设”, 符合现实分布式系统的特征。同时, 本文的方法符合 W3C 的 OWL DL 标准, 对于本体语义匹配、本体集成等有较大的参考意义。由于本文的算法不依赖注册服务器和分布式哈希表等技术, 本质上具有较强的动态适应性。另一方面, 本文实现的师生网络, 对于知识教育系统的实现, 具有借鉴意义。

参考文献

- Berners-Lee T, et al. The semantic web. Scientific American, 2001, 284: 35~43
- Leroy G, Chen H. Meeting Medical Terminology Needs-The Ontology-Enhanced Medical Concept Mapper. IEEE Transactions on Information Technology in Biomedicine, 2001, 5(4): 261~270
- Keleberda I, et al. Personalized Distance Learning Based on Multiagent Ontological System. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04): 2004, 777~779
- Cannataro M, et al. Integrating Ontology and Workflow in PROTEUS, a Grid-Based Problem Solving Environment for Bioinformatics. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), Vol, 2, 2004, 90~94
- Cannataro M, Talia D. the knowledge grid. Communication of ACM, 2003, 46(1): 89~93
- <http://wordnet.princeton.edu/>
- <http://www.isi.edu/natural-language/resources/sensus.html>
- Halevy A Y, et al. Schema mediation in peer data management systems. In: Proceedings of ICDE'03, International Conference on Data Engineering, 2003, 505~518
- Abiteboul S, et al. Constructing and querying peer-to-peer warehouses of xml resources. In: Proceedings of the 2nd International VLDB Workshop on Semantic Web and Databases, 2004
- Adjiman P, et al. SomeWhere in the Semantic Web. In: Third International Workshop on Principles and Practice of Semantic Web Reasoning, 2005, LCNS 3703, 1~16
- W3C Recommendation 10 February 2004. OWL Web Ontology Language Guide. <http://www.w3.org/TR/owl-guide/>
- Castano S, et al. From Surface to Intensive Matching of Semantic Web Ontologies. In: Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA'04), 2004, 140~144
- Missikoff M, et al. Integrated Approach to Web Ontology Learning and Engineering. In: IEEE Computer, 2002, 60~63
- Lambrix P, Tan He. A Framework for Aligning Ontologies. In: Third International Workshop on Principles and Practice of Semantic Web Reasoning, 2005, LCNS 3703, 17~31
- Rosati R. Semantic and Computational Advantages of the Safe Integration of Ontologies and Rules. In: Third International Workshop on Principles and Practice of Semantic Web Reasoning, 2005, LCNS 3703, 50~64