

知识工程中的知识库、本体与专家系统

魏圆圆¹, 钱平², 王儒敬¹, 王雪¹

¹(中国科学院合肥智能机械研究所, 合肥 230031)

²(中国农业科学院农业信息研究所, 北京 100081)

摘要: 随着语义 Web 思想的兴起, 对专家系统的互操作性和共享性也提出了更高的要求, 对新型知识表示方式和新型知识系统的研究是大势所趋. 本体作为共享概念模型的明确的形式化规范说明, 为不同系统之间的互操作提供了基本的框架, 是解决共享和互操作问题的有效的方法. 一些初学者对知识库、本体以及专家系统的概念产生了各种疑惑, 根据笔者的理解与实践, 对知识工程中的本体、知识库以及专家系统做出较系统的比较分析, 对这几个术语做一个澄清.

关键词: 知识库; 本体; 专家系统; 知识表示

Knowledge Base, Ontology and Expert System in Knowledge Engineering

WEI Yuan-Yuan¹, QIAN Ping², WANG Ru-Jing¹, WANG Xue¹

¹(Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China)

²(Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: With the development of the Semantic Web, the interoperability and sharing of expert system were put forward more advanced requests. The researches for new knowledge representation and knowledge base systems are important trend. As a formal, explicit specification of a shared conceptualization, ontologies provide frames for interoperability between different systems and can solve the sharing and interoperability problem effectively. There were a variety of doubts in some beginners' mind with the concepts of knowledge base, ontology and expert system. Based on the author's understanding and practice, the systematic comparative analysis between knowledge base, ontology and expert system in the domain of knowledge engineering were proposed in this paper, in order to make these terms less confused and more clearly comprehensible.

Key words: knowledge base; ontology; expert system; knowledge representation

作为构建语义 Web 的基础, 本体论的研究及其应用是目前国内外非常关注的研究热点, 成为一个新的研究领域. 近年来, 本体的开发逐渐从人工智能实验室走向领域专家的桌面上. 本体的应用从网站的分类系统(比如 Yahoo!)到网络产品销售的分门别类(比如 Amazon.com), 在互联网上的应用逐渐普及^[1]. 某些学者将本体看作是构造知识库的一种途径, 另一些学者认为本体是知识库的重要组成部分^[2,3], 此外还有专家将本体视为在不同平台间进行互操作处理的关键技术.

上世纪 90 年代, 随着本体定义的提出, 知识工程领域在建立知识库的方法上也产生了一种革命性思想, 即有关本体工程和构建本体知识库的思想^[2]. 同时, 随着并行与分布式处理、语义 Web 等新思想与技术的引入, 对专家系统的互操作性和共享性也提出了更高的要求, 对新型知识表示方式和新型知识系统的研究是大势所趋. 本体作为共享概念模型的明确的形式化规范说明, 为不同系统之间的互操作提供了基本的框架, 是解决共享和互操作问题的有效的方法. 因此, 当初研究专家系统的专家学者都对本体产生了兴趣,

基金项目: 国家自然科学基金(31171456); 中国科学院知识创新工程青年人才领域专项前沿项目

收稿时间: 2012-02-09; 收到修改稿时间: 2012-03-25

其中不乏有人提出类似“知识库和本体有什么关系?”、“本体与专家系统的区别?”、“本体能否取代专家系统?”等疑惑。本文作者根据自己的理解与实践,对知识工程领域中的本体、知识库以及专家系统做出较系统的分析。

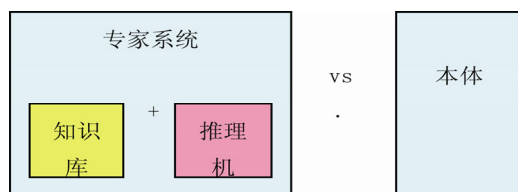


图 1 知识库、专家系统与本体

1 知识库与本体

知识库和本体的异同从描述对象和表现形式两个方面进行衡量:

1) 描述的对象

根据 MBA 智库百科中知识库的定义,知识库(Knowledge Base)是知识工程中结构化,易操作,易利用,全面有组织的知识集群,是针对某一(或某些)领域问题求解的需要,采用某种(或若干)知识表示方式在计算机存储器中存储、组织、管理和使用的互相联系的知识片集合。这些知识片包括与领域相关的理论知识、事实数据,由专家经验得到的启发式知识,如某领域内有关的定义、定理和运算法则以及常识性知识等。在人工智能界,最早给出本体定义的是 Neches 等人,他们将本体定义为“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范这些词汇外延规则”^[4]。被普遍接受的是 Studer 所给出的定义:“本体是共享概念模型的明确形式化规范说明(An ontology is a formal, explicit specification of a shared conceptualization)”^[5]。本体的定义说明,本体侧重于对领域内概念以及关系的描述,而知识库更侧重于对领域知识的表示,组织和存储。但他们都是对一个具体或是抽象领域中包含的知识的定义、表示和组织,都以机器可读的形式表达,可供机器处理。

2) 表现形式

知识库的构建需要某种知识表示方法的支持,而本体本身就是一种特殊的知识表示方法。知识库的知识可以是描述性知识或过程性知识,而本体的本质决定了它在表示所有类型知识上是有局限性的。所以说,本体是一种特殊的知识库^[6]。墨西哥的 María Auxilio

Medina Nieto 认为:本体可用于构建知识库^[7],美国斯坦福大学 Noy 和 McGuinness 认为:本体连同概念的个体实例集合组成一个知识库^[1]。

2 本体与专家系统

本体和专家系统的区别从以下五个方面进行衡量:

1) 描述的对象

本体是对共享概念模型的规范说明,即其概念在某个特定领域是公认的,是面向特定领域的概念模型。

专家系统是能够在特定的领域内模仿人类专家思维来求解复杂问题的计算机智能系统^[9]。知识库和推理机是典型专家系统的核心组成部分。通过对人类专家的问题求解能力的建模,采用人工智能中的知识表示和知识推理技术来模拟通常由专家才能解决的复杂问题,达到具有与专家同等解决问题能力的水平。专家系统知识库描述特定领域专家知识模型,只描述相关事物表面上的联系,知识库中的知识是简单事实的罗列或面向问题求解的过程知识,更多的是命题信息。所以本体和专家系统是为不同的应用目的而产生的。

2) 知识表示的方法和深度

本体有 5 个建模元语,即:类(classes),关系(relations),函数(functions),公理(axioms)和实例(instances)^[8],它们可以严格、规范、准确地刻画所描述的对象。

专家系统知识表示深度不如本体,知识表示方法较灵活,开发者可以根据知识的特点自定义知识表示范式,不同开发者对系统和知识库的理解不同,就会出现不同的知识表示方法,没有固定的建模原语,对建模也没有特殊要求。

3) 构建工具

专家系统知识库一般是领域专家通过知识录入编辑工具进行知识库的构建。十五期间,国家 863 计划支持了多家科研院所进行农业智能信息系统开发工具的研究,知识表示手段上存在很大的差异,所以也没有大家较公认的知识建模方法,至今没有出现较权威的专家系统开发工具。系统开发者熟悉知识表示,但不懂领域知识,领域专家对编写一定格式的知识库脚本也很难胜任。专家系统知识库的构建需要领域专家、和系统开发人员的密切合作。同时,由于无法彻底消除不同专家对领域知识理解上的差异,使得最终构建的知识库的可用性受到影响。

本体的构建是对客观世界的领域知识的建模, 利用本体可以用一种全新的方式将领域知识组织起来. 构建过程也必须有专家的参与, 知识获取更为严格、困难, 知识建模更为规范. 本体的构建有现成的方法可以借鉴, 也有很多成熟的本体开发工具可以使用^[10], 随着从事本体研究的人越来越多, 会出现大量的已经构建好、并得到领域专家公认的本体, 可以直接利用.

4) 推理机制

专家系统的推理建立在知识库的基础之上, 需要设计者在知识表示的基础上自己构建推理机, 推理过程面向求解问题的知识事实和过程知识, 知识表示的好坏, 直接影响到了推理的结果和效率. 推理过程是问题求解的过程.

本体知识库也是一种与其描述语言相关的知识库, 目前已经出现多种本体表示语言, 由于都采用 XML 作为其底层表示语言, 这些语言之间可以很容易地相互转换, 其中几种本体语言都存在设计良好的推理机^[11], 因此, 可以利用这些现成的推理机. 本体推理功能主要包括: 本体检查和隐含知识的挖掘^[12,13].

5) 知识的共享和重用

由于目前大部分专家系统的知识库和推理机是针对特定的系统设计使用的, 难以与其他相关领域的专家系统实现共享和重用, 较难达到推理和事实的分离. 这使得花费大量的时间和精力构建的专家系统的知识库不能得到更好的应用, 造成了资源的极大浪费. 知识表示的差异性, 也很大程度上制约了专家系统的共享性和重用性.

本体的构建是针对共享概念的明确规范说明, 所以说本体从最初的设计开始, 就是为共享的目的而构建的, 可以很容易地进行知识重用; 另外, 只要使用标准的表示语言格式对本体进行存储, 就可以进行本体的交换, 也就可以使用现有的本体推理机.

下表总结了本体和专家系统的比较结果, 对上面的分析进行补充.

表 1 本体和专家系统的比较

比较项	本体	专家系统
环境基础	语义 Web	计算机
应用时间	20 世纪末出现	20 世纪 90 年代盛行
知识类型	侧重概念及关系的描述	侧重问题的求解知识

规范化语言	以 W3C 推荐 OWL 为代表	自定义
推理方法	解析概念体系与概念间的关系	过程求解
推理机制	基于描述逻辑的推理	基于规则的推理和模型计算
性质	规范化知识服务	知识服务
逻辑基础	描述逻辑	命题逻辑

3 知识表示的表示观

知识表示是知识工程中的重要研究内容, 王珏等学者收集并分析了人工智能学界对知识表示问题讨论的各种观点, 将知识表示的观点主要归纳为认识论 (Epistemology)、本体论 (Ontology) 及知识工程 (Knowledge Engineering) 三类表示观^[14].

(1) 认识论表示观: 表示是对自然世界的描述, 知识表示自身并不体现任何智能行为, 其唯一的作用就是携带知识. 知识不完备性、知识不一致性、知识不确定性都是认识论表示观所讨论的内容.

(2) 本体论表示观: 表示是对自然世界的一种近似, 它规定了看待自然世界的方式, 即一个约定的集合. 表示只是描述了在这个世界中, 观察者当前所关心的那部分, 其它部分则被忽略. 与认识论表示观“表示的唯一功能是携带知识”的观点相反, 本体论表示观注重的是内容, 而不是语言形式. 但本体论表示观又与知识工程表示观不同, 所注重的“内容”不是某些特定领域的特殊的专家知识, 而是自然世界中的那些具有普遍意义的“一般知识”.

(3) 知识工程表示观: 表示是对自然世界描述的计算机模型, 它应该满足计算机这一实体的具体限制, 因此, 表示可以理解为一类数据结构及在其上的一组操作. 对知识的内容更强调与领域相关的、适合于该领域的、来自领域专家经验的知识. 这两个特点说明此类表示观更强调其工程实现, 而不甚关注对其行为的科学解释.

专家系统作为知识工程的典型应用, 遵循知识工程表示观, 很多情况下也涉及认识论表示观. Lenat 等人关于 CYC 系统的研究, 明确提出将“本体论表示观”作为人工智能研究的方法论^[15], 该观点得到 Davis 等人的支持^[16]. CYC 系统是美国 MCC (Microelectronics and Computer Technology Corporation) 公司的研究项目, 其目的是通过本体开发大规模常识知识库, 为常识推

理(CommonSense reasoning)提供基础。

4 结语

九十年代初,美国国防高级研究计划署(DARPA)、国家自然科学基金(NFS)等四个政府部门联合提出了知识共享计划(Knowledge Sharing Effort),以研究一种便于知识库和知识系统共享和重用的途径,包括斯坦福大学、AT&T在内的十多个研究中心参加了这个计划,而对本体的研究是这一计划的重要组成部分。其中高性能知识库系统项目(HPKB, High Performance Knowledge Base)的核心就是本体知识库。HPKB的研究目标是使知识库系统的开发者能迅速地(几个月内)建立一个容量的知识库。HPKB采用本体作为基本知识库的知识表示工具,使知识库具有良好的表达能力和可重用性,该项目的重要成果之一是开发了OKBC(Open Knowledge Base Connectivity)^[17]。TOVE(TOronto Virtual Enterprise)系统是加拿大多伦多大学的研究项目,其目的是企业建模,构造企业本体,可通过演绎推理实现对企业相关信息的问题求解^[18]。KACTUS是欧洲ESPRIT项目,通过构造支持产品知识重用的本体,支持计算机集成制造方法和知识工程方法的集成^[19]。

可以认为:基于本体的知识管理和知识服务是在语义Web时代对专家系统的继承和发展,是适应现阶段语义网和云计算环境发展的必然趋势。如何将本体和知识库系统结合起来,一方面发挥本体的优势,有效实现知识的规范表达以及知识的共享和重用;另一方面利用知识系统实现问题求解任务目标,以达到两者优势互补,将是接下来需要我们深入研究的问题。

参考文献

- Noy NF, McGuinness DL. Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
- Guarino N. Formal Ontology and Information Systems. Proc. of FOIS' 98. Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, 1998. 3-15.
- Mizoguchi R. Tutorial on ontological engineering-Part I: Introduction to Ontological Engineering. New Generation Computing, Ohmsha, Ltd and Springer-Verlay, 2003, 21(4): 365-384.
- Neches R, Fikes R, Finin T, Gruber TR, et al. Enabling technology for knowledge sharing. AI Magazine, 1991, 12(3):36-56.
- Studer R, Benjamins VR, Fened D. Knowledge engineering: principles and methods. Data & Knowledge Engineering, 1998, 25(1):161-197.
- Guarino N. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In: Paziienza MT, ed. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. Springer-Verlag, 1997: 139-170.
- Auxilio M, Nieto M. An Overview of Ontologies. From STAR Lab's research efforts, 2003, http://www.starlab.vub.ac.be/teaching/ontologies_overview.pdf
- Gomez-perez A, Benjamins VR. Overview of knowledge sharing and reuse components: ontologies and Problem Solving Methods. Proc. of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods(KRR5). Stockholm, Sweden, August 2, 1999.
- 史忠植,王文杰.人工智能.北京:国防工业出版社,2007.
- 钱平,郑业鲁.农业本体论研究及应用.北京:中国农业科学技术出版社,2006.
- 徐德智,汪智勇,王斌.当前主要本体推理工具的比较分析与研究.现代图书情报技术,2006,12:12-15.
- 潘超,古辉.本体推理机及应用.计算机系统应用,2010,19(9):163-167.
- 高琦,陈华钧.互联网 Ontology 语言和推理的比较和分析.计算机应用和软件,2004,21(10):73-76.
- 王珏,袁小红,石纯一等.关于知识表示的讨论.计算机学报,1995,18(3):212-224.
- Lenat DB, Guha RV. Building large knowledge-based systems: Representation and inference in the Cycproject. Addison-Wesley Pub. Co. 1990.
- Davis R, Shrobe H, Szolvits P. What is a Knowledge Representation. AI Magazine, 1993, 14(1):17-33.
- Chaudhri VK, Farquhar A, Fikes R, Karp PD, Rice JP. Open Knowledge Base Connectivity 2.0.3, Technical Report, <http://www.ai.sri.com/okbc/okbc-2-0-3.pdf>. 1998.
- Gruninger M, FOX MS. The Logic of Enterprise Modelling, Modelling and Methodologies for Enterprise Integration. Bernus P & Nemes L, eds. Cornwall, Great Britain:Chapman and Hall. 1996.
- Schreiber G, Wielinga B, Jansweijer W. The KACTUS view on the 'o' word. Workshop on Basic Ontological Issue in Knowledge Sharing. International Joint Conference on Artificial Intelligence. 1995.