

流程导向的新产品开发领域本体构建方法研究^{*}

李海刚, 尹万岭

(上海交通大学 管理学院 管理信息系统系, 上海 200052)

摘 要: 分析了新产品开发领域的特点, 基于对现有本体构建方法的分析和改进, 结合基于本体的知识检索, 提出了流程导向的新产品开发领域本体构建方法。该过程分为原型展示、本体详细构建、本体进化三个阶段。重点对该方法的过程进行了详细阐述。最后, 对流程导向的新产品开发本体构建方法进行了评述。

关键词: 新产品开发; 领域本体; 流程导向; 本体构建方法

中图分类号: TP301 文献标志码: A 文章编号: 1001-3695(2008)12-3587-03

Study of ontology design methodology for process-oriented new product development

LI Hai-gang, YIN Wan-ling

(Dept. of Management Information System, College of Management, Shanghai Jiaotong University, Shanghai 200052, China)

Abstract: This paper analyzed the role of ontology in the knowledge sharing and reuse in NPD domain firstly. Then, according to the character of the NPD domain, applied the process theory and prototype theory into the ontology construction in NPD and put forward a process-oriented NPD domain ontology construction methodology finally.

Key words: new product development(NPD); domain ontology; process-oriented; ontology construction methodology

市场对企业开发的新产品要求越来越高, 而企业也投入了越来越多的人力物力以提高新产品开发的效率和质量。随着网络技术和计算机技术的发展, 新产品开发领域也呈现出一些新的问题:

a) 特定领域内的各个企业新产品开发的特点不一样。这就对企业知识管理提出了很高的要求。知识工程师在构建知识管理系统的过程中, 不仅需要了解如何构建企业知识库, 而且对新产品开发知识也要有一定的了解, 才能更好地掌握企业的需求。

b) 新产品开发是一项巨大的工程, 要求团队合作, 互相沟通。随着跨国集团和战略联盟体的出现, 企业间合作开发新产品的情况越来越普遍。不同企业的新产品研发人员由于企业文化不同、领域所属不同等原因, 沟通上可能产生问题, 而知识不能共享。

c) 企业内部存在着大量的新产品开发的历史知识, 包括结构化知识、半结构化知识和非结构化知识, 这些知识可能是新产品开发的相关文件、产品数据库、e-mail、Web, 甚至可能是工作经验。这样即不利于企业知识共享, 也不利于宝贵历史知识的重用。

随着市场竞争的日益激烈, 企业对智能化需求的逐步提高, 利用知识管理系统, 采用跨企业合作开发新产品越来越普遍, 新产品开发领域的这些问题也逐步凸现出来。所有这些问题都提出了一个共同的需求, 那就是实现知识的共享和重用。随着企业新产品开发对知识的需求量逐渐增大, 企业对知识管理的要求也越来越高。新产品开发领域的专业性以及企业知识资源存储的异构性, 阻碍了企业知识共享和重用。

为了实现企业内部与企业之间的知识及知识处理系统的共享和互操作, 不同的用户、组织和软件系统必须进行有效的通信。但是由于不同的需求及应用背景, 致使关于同一个基本概念可能会产生不同的理解, 势必会阻碍共享及互操作的进行。本体论的提出减少或消除了概念及术语的混乱, 为解决这一问题提供了一条有益的途径。

1 本体及本体的研究概述

1.1 本体的定义

Gruber 于 1993 年提出本体的定义被人们广泛接受: “An ontology is an explicit specification of a conceptualization”。Guarino 为了进一步说明本体的含义, 引入域空间(domain space) 的概念。他定义域空间: 结构 D, W 。其中: D 指某一领域; W 指该领域中所有可能状态的集合。概念化定义为有序的三元组: $C = \langle D, W, R \rangle$ 。其中 R 是域空间 D, W 上概念关系的集合。也就是说, 本体是为说明某种语言词汇表的内在意义而设计的一套逻辑公理^[1]。

1.2 本体论研究现状

McCarthy^[2] 将本体论应用到人工智能(AI) 领域, 成为人工智能研究领域的一个重要分支。在知识工程领域、数据库管理系统领域、软件工程领域均面临着对某个领域知识表达概念化的问题。为了解决这个问题, 本体论在知识表达方面的应用逐渐发展起来^[1]。

在国外, 本体论的主要研究包括本体论的起源和发展、本体论的定义、本体论的描述语言、构建方法论、本体论与语义

Web 和本体论的应用等。在国内,本体论的研究也是一个热点,有许多专家都从事这方面的研究,但是主要集中在本体的应用研究方面。国内进行本体研究的主要有三支科研力量:中国科学院计算所、数学所、自动化所的若干实验室;哈尔滨工业大学计算机系;浙江大学人工智能研究所。他们的主要研究项目有武汉大学信息可视化和知识检索研究、中国科学院计算技术研究所大规模知识系统的研究、中国科学院数学研究所常识知识库的研究、浙江大学人工智能研究所基于本体论的产品信息集成研究等^[1]。

2 现有本体构建方法的比较分析

本体构建方法研究领域还没有成熟的本体构建方法。在国内外,相对比较成熟的本体构建方法主要有以下几种: a) 骨架法(skeletal methodology) 是 Uschold 等人^[3] 在爱丁堡大学开发企业本体的过程中所总结提炼而成,用于构建企业建模过程中的知识本体。他们虽然不是第一个提出本体构建方法的人,但是他们最先提出构建本体需要在方法论的指导下进行。 b) TOVE 本体构建方法是 Gr üniger 等人^[4] 基于他们在 TOVE (Toronto virtual enterprise) 项目上本体构建的经验提出的本体构建方法。 c) KACTUS 工程本体构建方法是 Bernaras 等人^[5] 在 KACTUS 工程中提出的本体构建方法。KACTUS 项目是欧洲 ESPRIT 框架下的研发项目之一。 d) METHONTOLOGY 本体构建方法是由马德里大学理工分校人工智能图书馆提出的本体构建方法。它专用于化学领域的本体构建^[6]。 e) 七步法是由斯坦福大学的 Noy 等人^[7] 提出的本体构建领域著名的方法。他们指出,并没有任何一个本体构建方法是完全正确的。他们提出的这个方法只是他们构建本体的经验,不同的应用领域和应用的目 的,构建本体的方法并不完全相同。

对比分析这些本体构建方法,得到如表 1 所示的结论。

表 1 本体构建方法比较

方法论	评价标准						
	产生领域	适用情况	详细程度	概念识别策略	构建本体策略	工具和技术	成熟度
Gruber 本体构建原则	无	所有情况	详细	无	无	KIF	低
骨架法	企业本体构建	所有情况	不详细	综合法	完全独立	头脑风暴等	低
TOVE 法	企业流程和活动	所有情况	不详细	综合法	部分依赖	无	低
骨架—TOVE 法	企业本体	所有情况	一般	综合法	部分依赖	头脑风暴等	中
KACTUS 法	电力网络系统	从零开发构建本体	不详细	自顶向下	完全依赖	无	低
METHON-TOLOGY	图书馆	所有情况	详细	综合法	完全独立	控制活动相关方法论	最高
七步法	医学	所有情况	详细	综合法	完全独立	Prot ég é 2000	高

没有一种方法是适用于所有领域的; METHONTOLOGY 方法和七步法成熟度最高;每一种方法都是在特定应用领域的基础上提出来的,方法通用性较差。同时现有这些方法还存在下列问题:

a) 新产品开发领域有知识密集性和专业性的特点,可能会导致新产品研发人员与本体库构建知识工程师之间出现沟通不便、交流困难的问题。所以在构建本体的过程中,本文引入了原型化的思想。

b) 新产品开发活动有其固定的开发流程,流程中的各个

开发阶段都需要特定的专业知识,因此可以按照新产品开发流程的每个阶段的新产品开发知识、人员、资源进行分类,以该流程为导向构建新产品。

3 新产品开发领域本体构建方法

3.1 流程导向的新产品开发领域本体构建原则

整个新产品流程本体构建过程都需要在本体构建原则(表 2)的指导下进行,而本体构建原则也是本体评估的主要标准。本体构建的成败很大程度上取决于领域专家的参与程度。因此,本体构建过程应该得到作为新产品开发领域专家的研发人员的全程支持。新产品开发流程中,本体构建的每个阶段、每个任务的成果都需要形成文档,这样有利于基于本体知识库的构建、本体的进化和维护等。

表 2 本体构建原则

原则	本体构建原则描述
清晰性	本体的定义应该明确、客观,并易于理解
一致性	由概念、公理和函数推导出来的结论不能与现有的概念、公理和函数相矛盾
可扩展性	在不修改现有本体体系前提下容易添加新的类和属性等本体元素
冗余性	由于本体的概念,出现重叠的现象最少
本体约束最少	在满足知识管理的前提下,对本体约束条件尽可能地少
编码独立性	本体的构建不依赖于特定的本体描述语言

3.2 流程导向的新产品开发领域本体构建方法的内涵

本文基于现有本体构建方法,结合新产品开发领域特点,引入原型化思想,结合基于本体的知识检索,提出了流程导向的新产品开发领域本体构建方法,如图 1 所示。图中新产品开发流程的本体构建过程分为原型展示、本体详细构建、本体进化三个阶段。其中,项目管理活动贯穿新产品流程本体构建的全过程,它包括各个阶段任务的安排控制和完成质量等。

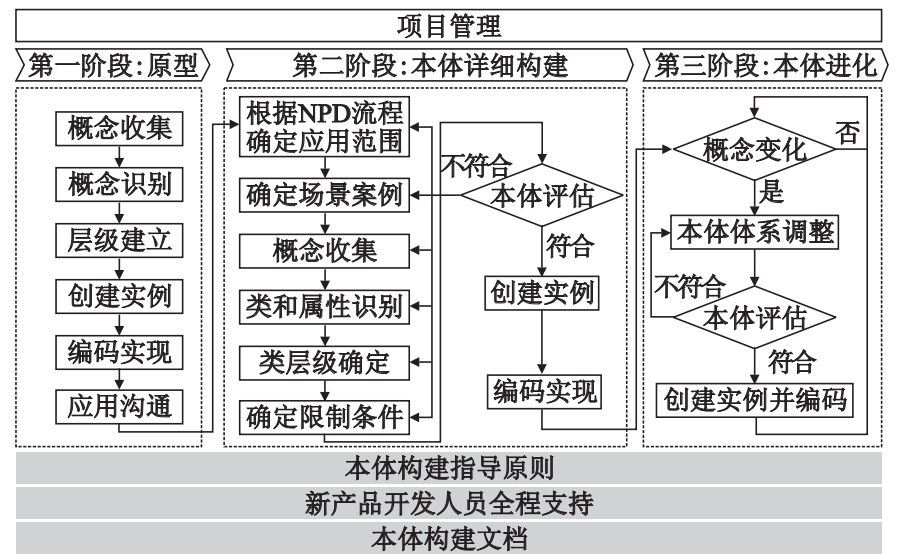


图1 流程导向的新产品开发领域本体构建方法

3.3 流程导向的新产品开发领域本体构建步骤

3.3.1 本体原型展示

第一阶段为本体原型展示阶段,目的在于创建一个简单的本体原型,方便知识工程师和新产品研发人员的沟通理解。依据上述分析,如果没有一个原型,知识工程师和新产品研发人员沟通起来就没有依据;有了本体的原型,才能使他们之间的沟通更方便,才能使新产品研发人员更好地理解本体的意义和作用,了解本体构建的大致流程,从而为第二阶段本体的详细构建打下良好的基础。该阶段具体环节如表 3 所示。

表 3 本体原型展示环节	
展示环节	具体含义
概念收集	原型展示的概念收集比较简单,并不需要收集所有的概念。知识工程师向关键的新产品开发人员介绍本体概念之后,在关键人员的帮助下,知识工程师可以根据自身对特定新产品的理解,收集新产品开发的基本概念
概念识别	对第一步收集到的所有概念进行概念识别,区分其中的类、属性、关系等
层级建立	根据区分出来的类、属性和关系,构建本体层级体系,确定类之间的继承关系,类和属性的所属关系,构建本体层级树
创建实例	根据新产品开发领域简单的、通俗的知识,构建本体的实例,进而进行简单的知识表示
编码实现	对上面构建的本体原型,选择特定的本体描述语言实现
应用沟通	依据快速构建的本体原型,给新产品研发人员展示本体的功能和基于本体的知识表示效果,使研发人员加深对本体的理解

本阶段整个过程都在本体构建原则的指导下进行,同时进行必要的文档记录,以减少下一阶段的工作量。本阶段结束后,新产品研发人员理解了本体的本质,减少了知识工程师和新产品研发人员之间沟通的障碍,于是进入下一阶段。

3.3.2 本体详细构建

第二阶段为本体详细构建阶段,目标是构建详尽的新产品开发流程本体模型。通过本阶段可创建符合新产品开发要求的本体。构建的本体库有利于新产品开发人员间的知识共享和重用,也有利于知识的检索和推理。具体实现分为以下几步:

a) 确定应用范围。由于新产品开发领域有着固定的流程,可以根据新产品开发流程确定新产品开发流程中本体的应用范围。虽然不同行业不同企业对新产品开发流程阶段的划分各不相同,但是宏观的流程是一致的。本文采用 Cooper & Kleinschmidt 提出的阶段—关卡模型(stage-gate model)作为新产品开发流程划分阶段的依据,根据新产品开发流程各个阶段划分新产品开发流程本体的应用范围,并以此作为新产品开发中知识组织的依据。

b) 确定场景和案例。在确定了新产品开发流程本体的应用范围之后,就可以根据相关的应用阶段组织本体的应用案例和应用场景。企业可将新产品开发历史数据库、案例库、文档作为基础,根据本体应用阶段组织相应的案例和场景。这些案例和场景由各个阶段新产品研发人员提出,在他们的工作中经常遇到。这些场景将作为概念抽取的重要来源。

c) 概念收集。本阶段采用两种主要的方法: (a) 头脑风暴法。由新产品研发人员在已经确定的案例和场景中提出新产品开发领域相关的各种概念。(b) 挖掘概念法。利用数据挖掘、文本挖掘和 Web 挖掘等工具对企业历史文档、数据库、知识库进行挖掘,利用现有的本体库,抽取出新产品开发领域中的关键概念。本阶段不需要区分这些概念的类和属性,只负责收集与新产品开发相关的所有概念,并按照新产品开发流程进行组织。

d) 类和属性的识别。在收集了新产品开发相关概念之后,开始区分这些概念哪些属于类,哪些属于类的属性。同时,注意分析哪些术语指的是同一个概念,即这些术语是同义词,并达成共识,用哪个术语在本体库中表示这个概念,同时记录它们之间的同义词关系。例如: 概念 “Screen ” “LCD ” “Display ”, 本体库中取 LCD 为基准概念,设置同义属性 “hasKey ”, Screen. hasKey = LCD, 而 LCD. hasKey = NULL。这样基于本体的知识检索进行同义处理时,只要检索 hasKey 属性值就可以查找到标准概念。

e) 类层级确定。类层级的确定主要有三种方式: (a) 自顶

向下,从最一般的概念开始建立本体层级体系。例如动物、狗、牧羊犬三个概念,“动物 ”的范围最大,为顶级,其次是“狗 ”,最底层是“牧羊犬 ”。(b) 自下向上,从最具体的类开始构建。例如首先建立“牧羊犬 ”类,其次才是“狗 ”,最后是“动物 ”。(c) 综合法,从中间的概念开始建立。例如首先建立“狗 ”类,然后分别向上和向下进展,建立“动物 ”和“牧羊犬 ”类。对于本体构建人员,综合法是最容易掌握的方法,也最容易被客户接受。

f) 确定限制条件。在定义了类的层级体系之后,需要定义每个类的属性限制条件,包括属性取值类型、允许值的范围、值的个数等其他特性。

g) 本体评估。对已经构建的本体体系进行评估检查,判断其是否符合本体构建的原则。如果合格,则进入下一步;否则根据检查结果,返回前面本体构建步骤,修改已构建的本体体系。

h) 创建实例。选择要创建实例的类,构建类的实例并确定实例属性的取值。

i) 编码实现。选择适当的本体描述语言描述本体。

至此,已经成功构建符合企业新产品开发流程要求的本体。该本体体系可以作为企业新产品开发中知识重用和共享的基础。随着本体在企业中的应用逐步深入,会不断出现新的知识,从而出现新的概念,因此就需要不断调整本体体系,从而开始第三阶段: 本体进化阶段。

3.3.3 本体进化阶段

第三阶段为本体进化阶段。随着本体的应用,本体体系需要不断地调整进化。这个阶段是一个漫长的循环阶段。

a) 概念变化: 随着新产品开发的进行,会出现新的知识和知识需求,此时需要判断是否需要调整本体体系,分析是否增加了新的概念,现有概念是否需要修改或删除等。如果是,则进行修改;否则等待新的知识出现再进行判断。

b) 本体体系调整: 如果需要调整本体体系,则需要根据第二阶段的相应步骤对现有本体体系进行修改。

c) 本体评估: 修改完成之后,需要对修改后的本体体系进行评估。如果不合格,则需要继续调整,直至合格后再进入下一步。

d) 创建实例并编码。修改合格之后,可以创建新增加的类的实例,然后对所作修改之处编码实现。等待新的知识和知识需求出现之后,再进行判断是否需要本体进化。

4 结束语

本文提出的本构建方法主要有两个特点: a) 引入了原型化思想,使本体构建知识工程师与新产品开发人员的沟通变得相对容易; b) 以业务流程为导向进行本体构建,如图 2 所示。

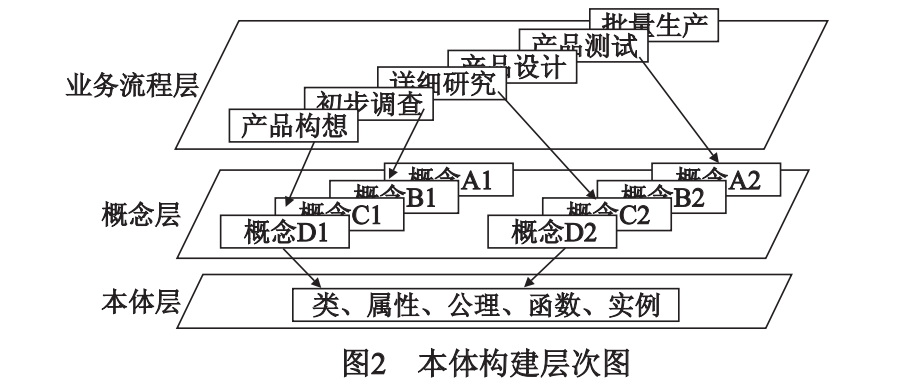


图 2 中表现了新产品开发领域由业务流程层到概念层,最后到本体层的转换过程,体现了本体构建过程中以新产品开发流程为导向的思想。在业务流程层,明确企业新产品开发流程存在的阶段,并组织相关开发人员在每个阶段 (下转第 3592 页)

阵, 将该矩阵处理后, 作为 HMM 的输入进行解码。模型的解码采用前向—后向算法。其中前/后向因子应采用新的 $e_k(s^t)$ 作相应修改, 可参考文献[10]。二级结构状态概率由同一类二级结构的状态概率相加得到, 最终的二级结构取三种结构概率和最大的。

2 预测结果

各种预测方法在 CB513 数据集中 492 条序列上进行 7-交叉验证, 其准确率如表 1 和 2 所示。

表 1 不同模型 Q_3, Q_i 及 SOV 预测准确率					%
模型	Q_3	Q_H	Q_E	Q_C	SOV
单个 BP 网络	74.46	77.02	64.19	78.81	66.51
两个 BP 网络串联	75.00	77.66	65.45	78.89	70.82
BP + 15-状态 HMM	77.21	79.23	66.27	84.49	72.52

表 2 不同模型的相关系数 C_i 的预测准确率			
模型	C_H	C_E	C_C
单个 BP 网络	0.686 0	0.584 7	0.555 4
两个 BP 网络串联	0.700 7	0.595 7	0.557 8
BP + 15-状态 HMM	0.721 9	0.614 9	0.590 8

从表 1 和 2 中可看出, 串联网路的运用确实比单个神经网络预测效果要好, 其 Q_3 准确率提高了 0.54%, SOV 提高了 4.31%, 相应的三态准确率 Q_H, Q_E, Q_C 也比单层网络提高了 0.08% ~1.26%。三态相关系数 $C_H、C_E、C_C$ 分别增加了 0.014 7、0.011、0.002 4。

神经网络与隐马尔可夫模型相结合的预测方法, 其准确率与串联神经网络相比又有所提高。其 Q_3 准确率提高了 2.21%, SOV 值提高了 1.7%。SOV 值的增加意义重大, 因为 SOV 是专门针对二级结构预测特点而提出的。详细算法请参见文献[5]。其值反映了预测方法预测出 H、E、C 片段的能力。三态准确率和相关系数也都有所提高。相关系数描述了两个事件的相关性, 所以相关系数越高, 说明预测的性能越好。

3 结束语

隐马尔可夫模型是近年来应用非常广泛的统计模型, 在蛋白质二级结构预测中, 单独使用隐马尔可夫模型的预测准确率往往不如神经网络方法理想。但在实验中, 笔者综合考虑了已知的生物学知识并对蛋白质序列进行了模式分析的基础上, 提出了比 3-状态 HMM 更加优化的 15-状态 HMM。它包含了更多的生物序列信息, 因此虽然模型简单, 但生物学意义明确; 此外, 通过改进的 HMM 算法将神经网络的输出与 HMM 有机结

合。从实验结果上看, 混合模型预测效果有所提高。

HMM 是将整条序列作为模型的输入, 它的建模特性能部分地利用蛋白质结构的远程相关性, 因为序列中较远的上游结构的潜在有用信息已经包含在隐状态中了(Markov 性的假设前提是对将来发生影响的只与当前的事实有关, 而无须考虑历史的信息。因为历史的信息都已经包含在当前事实中了)。神经网络与隐马尔可夫模型相结合的方法既能充分考虑相邻氨基酸残基间的相互影响(神经网络的特点), 也能在一定程度上照顾到序列的远程相关性(HMM 的特点), 因此带来了更高的预测准确率。

参考文献:

[1] ONTGOMERIE S, SUNDARARAJ S, GALLIN W, *et al.* Improving the accuracy of protein secondary structure prediction using structural alignment[J] . BMC Bioinformatics, 2006, 7(6) : 301-313.

[2] CROOKS G E, BRENNER S E. Protein secondary structure: entropy, correlations and prediction[J] . Bioinformatics, 2004, 20(10) : 1603-1611.

[3] CUFF J, BARTON G. Application of multiple sequence alignment profiles to improve protein secondary structure prediction[J] . Proteins, 2000, 40(3) : 502-511.

[4] JONES D T. Protein secondary structure prediction based on position-specific scoring matrix[J] . J Mol Biol, 1999, 292(2) : 195-202.

[5] ZEMLA A, VENCLOVAS C, FIDELIS K, *et al.* A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment[J] . Proteins, 1999, 34(2) : 220-223.

[6] 飞思科技产品研发中心. 神经网络理论与 MATLAB 7 实现[M] . 北京: 电子工业出版社, 2006: 44-52.

[7] 孙啸, 陆祖宏, 谢建明. 生物信息学基础[M] . 北京: 清华大学出版社, 2005: 188-194.

[8] 卢光莹, 华子千. 生物大分子晶体学基础[M] . 北京: 北京大学出版社, 2006: 115-119.

[9] HOEBEKE M, SCHBATH S. R 'MES: finding exceptional motifs [EB/OL] . (2006-12-28) [2007-07-12] . <http://genome.jouy.inra.fr/ssb/rmes/rmes3.userGuide.pdf>.

[10] MARTELLI P L, FARISELLI P, KROGH A, *et al.* A sequence-profile-based HMM for predicting and discriminating β barrel membrane proteins[J] . Bioinformatics, 2002, 18(1) : 46-53.

(上接第 3589 页) 收集新产品开发本体的初始概念; 然后依据各个阶段将初始概念进行分类, 这样有利于概念领域的确定; 辨别各个阶段的概念中的类、属性、公理、函数和实例, 构建新产品开发领域本体库。

参考文献:

[1] 李善平, 尹奇, 胡玉杰, 等. 本体论研究综述[J] . 计算机研究与发展, 2004, 41(7) : 1041-1052.

[2] McCARTHY J. Circumscription: a form of non-monotonic reasoning [J] . Artificial Intelligence, 1980, 13(5) : 27-39.

[3] USCHOLD M, KING M. Towards a methodology for building ontologies[C] //Proc of Workshop Held in Conjunction with IJCAI on Basic Ontological Issues in Knowledge Sharing. 1995.

[4] GR NINGER M, FOX M S. Methodology for the design and evaluation of ontologies[C] //Proc of Workshop on Basic Ontological Issues in Knowledge Sharing. 1995.

[5] BERNARAS A, LARESGOITI I, CORERA J. Building and reusing ontologies for electrical network applications[C] //Proc of European Conference on Artificial Intelligence. 1996: 298-302.

[6] FERN NDEZ M, G MEZ-P REZ A, JURISTO N. METHONTOLOGY: from ontological art towards ontological engineering[C] //Proc of AAAI Spring Symposium on Ontological Engineering. 1997: 33-40.

[7] NOY N F, McGUINNESS D L. Ontology development 101: a guide to creating your first ontology [EB/OL] . (2002) . http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.