

·理论园地·

领域本体的构建方法研究

张文秀 (南京审计学院审计系 江苏南京 211815)

朱庆华 (南京大学国家信息资源管理南京研究基地 江苏南京 210093)

摘 要: 领域本体构建是一项复杂的系统工程。从国内外本体和领域本体研究出发,比较分析了典型的领域本体构建方法,提出循环式领域本体构建法,并使用该方法构建软件缺陷领域本体。

关键词: 本体 领域本体 本体构建 软件缺陷

中图分类号: B016

文献标识码: A

文章编号: 1003-6938(2011)01-0016-04

Research on Construction Methods of Domain Ontology

Zhang Wenxiu (Department of Audit, Nanjing Audit University, Nanjing, Jiangsu, 211815)

Zhu Qinghua (National Center for Information Resource Management, Nanjing University, Nanjing, Jiangsu, 210093)

Abstract: The construction of domain ontology is a complicated system project. On the basis of researches on ontology and domain ontology at home and abroad, it compares the typical ontology construction methods. And, it proposes a comprehensive ontology construction method. Moreover, it uses the method and constructs software defect domain ontology.

Key words: ontology; domain ontology; construction method; software defect

CLC number: B016

Document code: A

Article ID: 1003-6938(2011)01-0016-04

1 引言

本体(Ontology)起源于哲学,是对世界上客观存在物的系统描述。近年来,人们将本体的概念引入人工智能、知识工程和图书情报领域,用以解决信息提取、知识概念表示和知识组织体系方面的有关问题。利用本体思想从不同角度对信息集合进行标引,表示信息内容与知识组织体系之间的链接关系,可以将本体与信息系统进行链接,从而使用户在使用信息的过程中更加便捷地浏览和理解相关概念和资源,还可以利用本体中的语义关系及推理规则集合进行推理,从而实现基于本体的智能分析和知识组织,并通过智能分析来预测知识增长点。同时,由于本体描述信息的语义,并采用一定的编码语言让计算机可以读懂,更加有利于实现智能检索和查询。可以说,本体是机器自动推理和智能化高级信息服务的基础。^[1]

本体构建是一项复杂的系统工程,需要领域专家和知识工作人员以及系统分析人员等各方人员按照一定的方法、采用适当的工具协作完成。由于目前本体构建大都针对某一特定问题域,因此,领域本体的构建也代表了本体构建的思想和方法。

2 本体与领域本体研究现状

由于应用领域的不同,对本体研究的侧重点也有所不同。涉及特定学科领域的本体,被称为领域本体(Domain Ontology)。领域本体的目标是捕获相关的领域知识,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型上给出这些概念和概念之间相互关系的明确定义,提供该领域中发生的活动以及该领域的主要理论和基本原理等。

近两年来,在语义网描述语义的触动下,越来越多的描述各种各样领域的本体产生了。例如,Plinius ontology是关于陶瓷物质化学成分的本体,而 Chemical-Elements(化学元素)是关于化学元素周期表的本体。^[2]国外著名的知识本体还有 WordNet、SENSUS、CYC、SUMO 等。我国近几年也出现了如知网、国共合作历史领域本体等。对特定领域的本体研究和开发目前已涉及的领域包括企业本体、医学概念本体、经济学本体、花卉学本体、酶催化生物学本体、陶瓷材料机械属性本体等。

随着领域本体渐渐成为研究热点,构建和进化领域本体已经成为研究的核心和关键。从构建方式上看,现在的本体构建大多是手工构造的。目前,大量的个人和学术

团体正专注于本体自动构建工具的研究。本体工具主要有本体的编辑管理工具、本体解析工具和推理机。^[3]最常见的两个本体编辑工具是 Protégé 和 KAON, 本体解析工具是 Jena, 推理机是 Racer。尽管手工构造也有一些规范的方法和可用的工具^[4]然而, 构建本体的工作还基本停留在人工或半人工阶段, 国内外为数不少的本体构建工具和方法体系还没有一个完全成熟的, 因此仍难以进行大范围的本体构建。而且, 已有的领域本体研究主要是对本体的编制与构建进行尝试与实验, 较少研究相应的编制规范与标准。目前, 需要领域专家参与是本体构建方法中的瓶颈, 如何通过知识挖掘手段自动获取本体是当前, 也是今后一段时期的重要研究领域。

领域本体的应用研究也有广泛的内容。如: 中国科学院致力于研究形式化本体在领域知识的复用和共享中的基础和作用、基于专业领域知识复用的虚拟领域本体的构建。^[5]此外, 由于本体明确地表达概念及其之间的关系, 并且具有推理能力, 因此, 利用本体可以实现自动推理和智能化高级信息服务。在情报学领域, 梁战平提出在本体信息环境下进行信息获取和智能分析的情报学研究模式。^[6]基于本体的信息研究与分析将成为一个新方向。

从整体来看, 本体研究目前仍处于理论研究日趋成熟、应用研究相对滞后的阶段。单纯从技术角度描述本体的较多, 理论联系实践并在实际系统中应用的领域本体则比较少。^[7]

3 领域本体构建方法

本体的构建是对概念本身以及概念与概念之间的关系进行形式化描述, 多是面向特定领域。出于对不同学科领域和具体工程的不同考虑, 领域本体构建的过程各不相同。现行的本体构建方法都不是经权威标准化机构认证的方法。

比较成功的本体构建项目大多借鉴软件工程方法, 产生了一系列诸如面向对象思想(基于 UML)、原型化思想等的本体构建方法, 并从系统需求分析出发, 明确需求、规范文档、实时评价等方面规范领域本体的构建,^[8]再根据项目自身的特点和专家经验进行。

3.1 本体构建方法的比较

当前典型的本体构建方法都是从具体的本体构建项目中总结获得的, 最早的本体构建方法总结出现在 1995 年, 是根据企业本体(Enterprise Ontology)以及 TOVE 项目本体的实际开发过程获得的经验总结。此后, 陆续出现了一些新的本体构建方法, 如 METHONTOLOGY、骨架法、KACTUS 工程法、SENSUS 法、IDEF-5 方法、七步法等。这些方法都由一个总体流程和各步的操作规则构成。

本研究通过文献调查法分析七种本体构建方法的基本步骤(见表 1), 选取生命周期、所采用的相关技术、方法的细节、方法的特点以及本体的应用领域这几个方面进行比较(见表 2)。^[9]

表 1 七种本体构建方法的过程比较

名称	规划	考虑复用	分析	表示	评价与优化
TOVE 法	有	无	有	有	有优化
METHONTOLOGY 法	有	无	有	有	有评价
骨架法	有	无	有	有	有评价
KACTUS 法	有	无	有	有	无
SENSUS 法	无	无	有	有	无
IDEF5 法	有	无	有	有	优化与验证
七步法	有	有	有	有	无

表 2 七种本体构建方法的综合比较

名称	生命周期	相关技术	方法细节	特点	应用领域
TOVE 法	非真正生命周期	不确定	少	基于问题	企业
METHONTOLOGY 法	有	有, 不全	详细	生命周期	化学、(Onto) 2Agent
骨架法	非真正生命周期	不确定	少	指导方针	企业
KACTUS 法	没有	不确定	很少	知识复用	多用途复杂技术系统知识建模
SENSUS 法	没有	不确定	一般	启发式	电子科学、军事等
IDEF5 法	没有	有, 不全	详细	结构化	企业
七步法	非真正生命周期	有	详细	系统	医学等

通过表 1、表 2 的比较分析可以看出, 七种本体构建方法中比较完整、成熟的是七步法和 METHONTOLOGY 法, 而 IDEF5 法、SENSUS 法、TOVE 法、骨架法和 KACTUS 法则一般。尽管如此, 每种方法体系都有它的特点和与适用的领域, 即使是不成熟的方法也有它的优点。

除了上述七种典型的本体构建方法外, 还有很多本体研究学者在本体实际开发中提出了适合本专业领域需求的本体构建方法, 例如 Staab 等提出的 On- To- Knowledge 法。^[10]此外, 我国学者(如李景^[11]、董慧^[12]、刘柏嵩^[13]等)在借鉴国外本体构建方法的基础上, 根据中文本体构建的实际情况, 也提出了一些有影响的本体构建方法。

总的来说, 由于不同领域的概念具有不同特点, 本体构建的理论和方法目前尚无统一的标准, 因而难以在不同领域本体的构建中保持一致。对于任何专业领域, 都不存在某一种唯一适合的途径或模式, 可能存在好几种方法都可行。最佳的办法依赖于领域本体的构建者所采用

的应用软件以及可以预见的扩展功能。

3.2 领域本体构建方法的创新

在选择本体构建方法时应根据实际情况采用最适当的方法,或者综合多种方法的优点,对现有构建方法进行提升和优化,从而提出更适合特定领域的构建方法或者更具有广泛适用性的构建方法。本研究没有简单地选用某一种方法,而是基于以上对七种典型本体构建方法的比较研究,选择以斯坦福大学医学院所提出的成熟度较高的“七步法”的思路为基础,同时综合 Methontology 法、IDEF5 法具有评价与优化的优点。具体说来,本研究对“七步法”进行了如下修改:

(1)将第五步——定义类的属性和第六步定义属性的分面(Facets)合并。因为一个属性可能由多个“分面”组成,一个属性的“分面”,就是属性取值的类型(Value Type)、容许的取值(Allowed Values)、取值个数(Cardinality,集的势、基数)和有关属性取值的其他特征。在定义类的属性时也需要定义属性的分面,只有把属性的各个分面都定义了才能认为属性定义好了。

(2)将第七步——创建实例在领域本构建过程中省略。实例体现在特定领域的具体应用中,在领域本体的一般构建过程中可省略创建实例这一步骤,而是将创建实例放到领域本体的实际应用中,使本体的构建方法更具一般性。

(3)本体构建过程中应该对所初步构建起来的本体及时进行总结与评价,发现不足,进行改进。因此,借鉴了 Methontology 法和 IDEF5 法等具有的本体评价这一步骤。而且,考虑到本体构建及完善应该是一个不断循环往复、螺旋式上升的过程,因而,提出了由六大步骤组成一个循环的领域本体构建方法,可称之为“循环法”(见图1)。

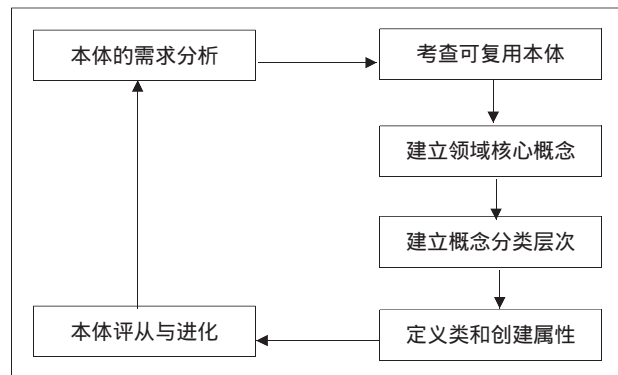


图1 循环法领域本体构建流程

4 软件缺陷领域本体构建

本研究采用图1所示的循环法领域本体构建流程构建软件缺陷信息管理领域本体,以验证该方法的可行性和科学性。具体步骤如下:

4.1 本体的需求分析

正如同七步法中第一步的要求,本体构建首先必须确定构建目标、范围和场景等。

本研究所面向的领域是软件缺陷信息管理,其领域本体的构建目标是建立面向应用的领域本体,使缺陷信息更趋结构化,以便于获取知识,更便于进行多种多样的信息分析以实施管理。

构建的范围是软件缺陷信息管理中涉及到的缺陷信息及缺陷管理流程。

场景有两个:①软件组织中测试工程师、质量工程师、软件工程师和项目经理共同参与的软件缺陷管理过程;②软件组织的中高层管理者为了管理和决策而希望得到经过分析处理的缺陷信息。也就是说,本研究是针对软件缺陷信息管理中涉及到的缺陷信息以及软件缺陷管理流程,提取出一套核心概念集及其属性,并确定它们之间的相互关系,建立可以应用的本体,再进行基于本体的软件缺陷信息分析。

4.2 考查可复用本体

本体最大的特点之一是共享和重用,考查复用现有本体的可能性是保证本体的这一特性得以实现的一个重要方式,通过复用现有本体也可以提高本体构建的效率。本体同时还具有开放性,可以随时更新,也是复用现有本体的体系结构基础。

国内外对本体的研究还处于实验阶段,比较缺乏成熟、实用的领域本体。现有的领域本体主要是企业本体、医学概念本体、经济学本体等,而对于软件缺陷信息管理这个领域目前还没有发现与之相似、可以重用的本体。

4.3 建立领域核心概念集

这一步是要确定软件缺陷信息管理领域本体所涉及的概念,这就要列出所有领域内的重要术语,收集所有有用的和潜在有用的领域概念、语义、属性、实例等,再经过整理、提炼,从而建立概念汇总表。建立领域核心概念集之前,应首先通过收集领域信息充分了解领域知识,信息来源可以包括领域专家、专业书籍、网络等。在建立领域核心概念集时,可以借鉴已有本体的结构和体系,从软件缺陷管理领域中分解、提取、去重、归并所涉及到的主要概念,并力求概念集的完整性,从而建立概念汇总表。

4.4 建立概念分层结构

知识工作者和领域专家协同工作,梳理概念汇总表中的每个概念,主要采用自底向上的方法,从一个个底层概念的定义开始,一一明确每个概念的含义。分析概念之间可能存在的种种关系,这些关系包括显性关系和隐性关系。最后,采用综合法进行归纳和演绎,确定领域概念之间的关系,并用明确的方式记录出来。

软件缺陷信息管理领域中核心概念集是在 IBM 的软件缺陷分类^[14]基础上建立起来的,软件缺陷分类已经体现了概念的分层结构(见图 2)。

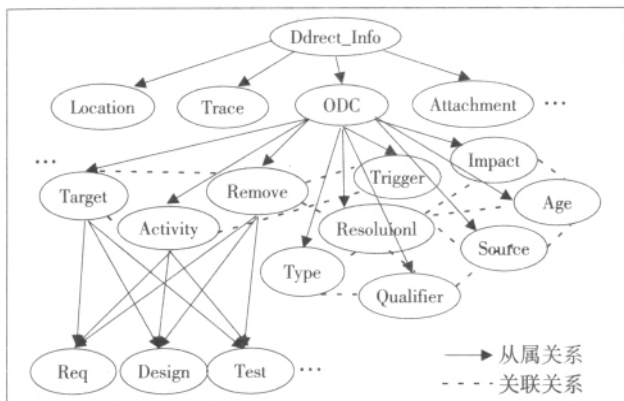


图2 软件缺陷领域概念集中的概念及其关系(部分)

随着本体构建的推进和发展,还可能对所建立的概念集中的概念及其关系进行修改、添加和删减。

4.5 定义类和创建属性

概念汇总表经过分层确定其从属结构,定义了类,剩余的概念将会是类的属性。每个属性都有它的属性名,都确定它所描述的类。这一定义类和创建属性的过程就是本体的表示,即采用建模工具和语言定义类和创建属性来建立本体。在众多本体建模工具和语言中选择适当的本体建模工具和语言来表示本体内部概念关系的工具,将已得到的领域概念集和概念之间的相互关系形式化地表现出来。本研究选取 Protégé 为本体建模工具,采用 W3C 推荐的 OWL 作为本体描述语言,所构建的类和属性如图 3 所示。

4.6 本体的评价与进化

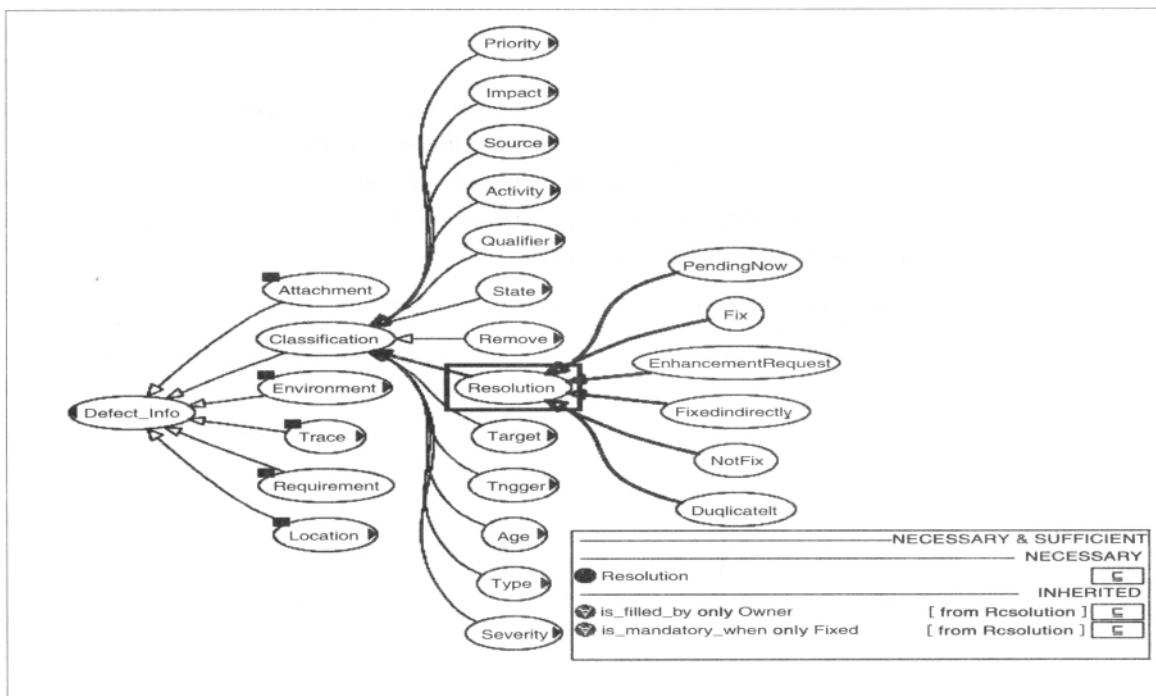


图3 OWLviz 中的软件缺陷领域本体(部分)

目前还没有一致认可的本体确认和评价的标准,对本体的评估涉及是否满足需求分析阶段所设定的目标,涉及如何正确构建本体,涉及本体及其定义内容的清晰性、一致性、完整性、可扩展性以及灵活性。本体初步构建起来以后,可以据此进行评价和改进。领域本体也是具有生命周期的,在本体初步构建好之后将是长期的本体操作阶段、维护阶段,其间还要持续地进行本体的完善工作,这是因为本体的构建不是一蹴而就的,而是需要不断的改进。

以上是本研究所采用的本体构建流程的六个步骤,其中,作为起始的本体构建初始步骤包括分析本体构建

需求与考查可复用本体,建立领域核心概念集、建立概念分层结构、定义类和创建属性是整个流程中的核心步骤,最后是对所构建本体的评价和改进,实现了本体创建和提升的循环过程。

5 结语

领域包含大量的概念和对象,要迅速建立一个全面的领域本体是不现实的。可行的做法是领域专家和知识工作者根据软件工程的原型化方法建立一个原型化的领域本体,然后在使用的过程中逐渐迭代,采用螺旋上升的方式逐步完善已有的本体模型。以上只 (下转第 40 页)

型更为高尚。在人类阅读需求的金字塔上,生存型比消遣型、消遣型比发展型、发展型比研究型的阅读需求来得更为迫切,覆盖面也更为广泛而已。

5.3 四种阅读类型对于图书馆的意义

我国图书馆体系主要由公共图书馆、高校图书馆、科研院所图书馆、企事业单位图书馆和私立图书馆构成。由于各级各类图书馆各具特色,且所承担的职能各不相同,作为图书馆,应当根据本馆所处层次地位、承担的基本职能和社会对于图书馆的基本定位来确定图书馆应当重点满足哪一层次的阅读需求。

我国的公共图书馆是由“国家图书馆——省级图书馆——市级图书馆——县级图书馆——县级以上社区及乡镇图书馆”组成的公共图书馆体系。国家图书馆和省级图书馆属于大型公共图书馆,主要承担文献收藏和保护的职责,提供文献信息及参考咨询服务,开展图书馆学理论与图书馆事业发展研究,指导下级图书馆业务工作,履行有关文化交流职能。以满足发展型和研究型阅读需要为主。市县级及以下基层公共图书馆主要承担向公众提供,以满足群众的生存型、消遣型阅读需求为主,兼顾发展型阅读需求。

高校图书馆和科研院所图书馆则主要是为高校及科研院所的教学科研活动服务,提供满足本单位成员所需

的各种信息资源和服务。因此,高校和科研院所图书馆主要以满足发展型和研究型阅读需求为主。

企事业单位图书馆主要是为本单位生产活动提供相应的技术支持,主要以搜集本行业相关的各种信息资料并提供情报服务解决企业生产过程中所遇到的各种问题,同时提供企业开展技术创新所需要的各种资料,以满足企业成员的生存型和研究型阅读为主。同时,企事业单位也应当适当考虑到职工的业余生活,为其提供一定的消遣性阅读材料,使其工作娱乐兼顾,生活张弛有度。

参考文献：

- [1] 孙素英. 学生阅读能力的发展特点研究[J]. 语文建设, 2009 (12).
- [2][5] 汤松梅. 读者阅读需求类型初探[J]. 长春大学学报, 2007, 17(3).
- [3][4] 李群. 读者的阅读类型和读者阅读统计的分析[J]. 重庆图情通讯, 2002 (9).
- [6] 国金荣. 研究读者类型 全方位为读者服务[J]. 鸡西大学学报, 2009 (1).
- [7][8][9] 傅金祥. 阅读能力及阅读类型浅析[J]. 阅读与写作, 2006 (3).

作者简介：徐菊,女,江苏建康职业学院图书馆助理馆员。

(上接第 19 页) 是领域本体构建的六个主要步骤,之后还有长期的本体完善与操作阶段,其中有大量工作如本体自动优化等有待研究和实践。

参考文献：

- [1] 张玉峰等. 基于 Semantic Web 的个性化网络导航机制[J]. 情报学报, 2005 (24): 438-444.
- [2] M.F. López, et al. Building a Chemical Ontology Using Methonotology and the Ontology Design Environment[J]. IEEE Intelligent System. Jan./ Feb. 1999 (1): 37-46.
- [3] 陈谷川, 陈豫. 语义网知识组织系统的研究与构架[J]. 现代图书情报技术, 2006, 135(4): 24-28.
- [4] Y. Ding, S. Foo. Ontology research and development: part1—a review of ontology generation [J]. Information Science, 2002, 28(2): 234-260.
- [5] 陈刚等. 基于领域知识重用的虚拟领域本体构造[J]. 软件学报, 2003(3): 350-355.
- [6] 梁战平. 开创情报学的未来——争论的焦点问题研究[J]. 情报学报, 2007, 26(1): 14-19.
- [7] 杜小勇等. 学科领域本体的构建与进化——以经济学领域本体为例[J]. 现代图书情报技术, 2007 (3): 7-12.
- [8] 袁媛. 领域本体建设的方法论和工具研究[D]. 北京: 中国人民大学硕士学位论文, 2004.

- [9] 李景. 本体理论在文献检索系统中的应用研究[M]. 北京: 北京图书馆出版社, 2005: 111.
- [10] S. Staab, etc. Knowledge processes and ontologies[J]. IEEE Intelligent Systems, Special Issue on Knowledge Management, 2001, 16(1): 26-34.
- [11] 李景等. 构建领域本体的方法[J]. 计算机与农业, 2003, (7): 7-10.
- [12] 董慧等. 基于本体的数字图书馆检索模型研究(III)——历史领域资源本体构建[J]. 情报学报, 2006, 25(5): 564-574.
- [13] 刘柏嵩. 面向数字图书馆的本体学习研究[J]. 大学图书馆学报, 2006 (6): 30-34, 38.
- [14] IBM Research Center for Software Engineering. Orthogonal Defect Classification[EB/OL]. [2007-04-12] <http://www.research.ibm.com/softeng/ODC/ODC.HTM>.
- [15] 李景, 孟连生. 构建知识本体方法体系的比较研究[J]. 现代图书情报技术, 2004 (7): 17-22.

作者简介：张文秀(1975-) ,女,博士,南京审计学院审计系讲师,研究方向:信息系统审计;朱庆华(1963-) ,男,博士,南京大学国家信息资源管理研究基地教授,研究方向:信息资源管理。