

· 知识组织 知识发现 知识服务 ·

领域本体可视化构建研究^{*}

李光达¹ 常 春² 张峻峰¹ 郑怀国¹ 谭翠萍¹

(1. 北京市农林科学院农业科技信息研究所 北京 100097;

2. 中国科学技术信息研究所 北京 100038)

摘 要 为了使本体构建可以在多人参与下达成知识共识,将可视化技术引入本体构建的过程中,开发可视化本体构建平台,形成了一套特有的领域本体构建方法。文章介绍了可视化的相关概念,以及本体构建的常用方法;在领域本体构建实践的基础上,归纳出了领域本体可视化构建的基本流程和步骤;最后讨论了本方法的优势及未来发展趋势。

关键词 可视化 领域本体 本体构建 本体构建方法

中图分类号 G254.2

文献标识码 A

文章编号 1002-1965(2013)09-0171-04

Research on Visualization Method of Domain Ontology Building

Li Guangda¹ Chang Chun² Zhang Junfeng¹ Zheng Huaiguo¹ Tan Cuiping¹

(1. Institute of Agricultural Sci-tech Information, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097;

2. Institute of Scientific & Technical Information of China, Beijing 100038)

Abstract To reach a knowledge consensus on domain ontology, the visualization method is used in domain ontology building, a visualization system of ontology building is developed, and a unique method of domain ontology building is summarized. In this paper, concepts of visualization and common methods of ontology building are introduced. Then, based on the practice of domain ontology building, processes and procedures of its visualization method are summarized. Finally, the advantages and trends of the method are discussed.

Key words visualization domain ontology ontology building ontology building method

0 引 言

领域本体承载着特定领域的相关知识,是领域知识组织的语义化工具。领域本体构建作为一项知识组织基础工程,具有构建时间长、参与人数多、步骤复杂等特点。目前,领域本体构建方法研究虽然已经取得一定的成果,但是还存在诸多问题需要解决,例如,指导方法过于抽象、构建周期过长、缺少具体构建实例等。为了达成领域知识共识,使领域本体构建直观化、具体化、简单化,本研究将可视化技术和方法应用于领域本体构建的过程中,研究可视化的领域本体构建系统,探索领域本体可视化构建的基本流程和步骤,通过图形化和网络化技术提升领域本体的构建效率。

1 可视化

可视化是为了形成一个精神上的模型或影像^[1]。可视化应用于抽象的数量和关系,以从数据中获取洞察^[2]。可视化分为科学可视化和信息可视化。科学可视化是计算机图形学的一个重要领域;而信息可视化则包含了运筹学及相关学科的很多内容,更侧重于抽象信息的可视化,包括访问的结果以及数据各部分之间的关系,用于指导和加速查找的过程。

在可视化领域,相关技术不断涌现,单从维度上可分为二维可视化技术、三维可视化技术和多维可视化技术,这些技术在深入研究的同时也开始应用于各个领域。可视化技术可以丰富信息和知识的表示方式,

收稿日期:2013-05-21

修回日期:2013-07-10

基金项目:农业科技成果转化资金项目“U农蔬菜通”研发与应用(编号:2010GB2A000013)。

作者简介:李光达(1983-),男,硕士,助理研究员,研究方向:信息组织、Ontology、信息管理、信息技术等;常 春(1966-),男,博士,研究员,研究方向:信息管理、本体论、叙词表;张峻峰(1970-),女,研究馆员,研究方向:农业信息化;郑怀国(1967-),女,副研究馆员,研究方向:农业信息管理;谭翠萍(1980-),女,硕士,助理研究员,研究方向:农业信息管理。

展示文字难以描述的复杂关系和结构;同时,可视化技术有利于促进知识的理解与吸收,使不同的用户达成共识^[3]。因此,可视化技术能够提供直观的交流环境,促进人们之间互相交流和沟通,满足领域知识组织的需要,可以用于支撑本体构建的实现。

2 本体及本体构建

Ontology 是概念模型的明确的规范说明^[4]。更为直观的描述是本体是描述概念和概念间关系的一种模型^[5]。从本体的定义可以发现,本体构建是收集相关概念,以及发现和构建概念之间关系的过程。本体构建是本体领域的基础,对于本体的应用起到至关重要的作用。然而,由于本体工程学相对不成熟,每个研究小组都用自己的方法构建本体^[6]。这些方法包括 IDEF5、骨架法、TOVE 法、METHONTOLOGY、KAC-TUS、SENSUS、七步法等^[7]。面向特定应用的本体构建可以综合使用多种方法,同时也可以根据专业领域、学科背景或研究课题制定适合自身的方法。本文所提出的领域本体可视化构建方法就是在农业领域本体构建的实践中总结和研究出的一套方法。该方法将可视化技术引入到本体构建中来,开发在线支持多人协作和支持构建过程可视化的领域本体构建平台。与以往的本体构建方法相比,本方法的直观性和可操作性更强,方法具体考虑到本体构建的阶段性和人员分工,在符合总体构建任务的前提下,最大化满足不同本体构建人员的个性化需求。可视化本体构建平台全面辅助本体构建的每个阶段,使原本枯燥的本体构建工作变得简单、直观和易操作。

3 领域本体可视化构建

领域本体可视化构建将可视化技术引入本体构建,利用焦点加上下文技术、关联更新技术、滤镜技术等可视化技术和多种编程语言在 .NET 平台上开发了领域本体可视化构建平台,实现了本体构建的可视化录入、可视化分析、可视化关系构建、可视化管理等功能。在本体构建的过程充分利用可视化方法直观展示信息,通过图或表的形式反映概念、实例及其关系,尽力消除本体构建过程中的模糊信息,使本体构建人员之间达成共识、协同工作,提高领域本体构建的效率和准确性。在系统支持的基础上,总结出本方法的基本构建流程,如图1所示。

首先,依据本体的用途确定本体的范围,然后判断当前是否已经存在该领域本体,如果存在考虑是否直接加以改造复用,如果不存在,就需要着手本体构建。本体构建首先需要获得概念,因此需要判断是否有可用的知识库作为概念获取来源,以便于将其直接导入

本体构建系统,如果存在类似的可用知识库,就可以通过一定的规则和算法将知识库直接导入本体构建系统,然后利用概念抽取算法自动获取概念和实例,如果没有可用的知识库,就需要依据一定纸质资料人工录入概念和实例。其次,在概念集合的基础上,利用可视化方法建立概念间的关系,最后,进行本体评价和检验,完成本体构建。参考该流程,领域本体可视化构建的具体步骤如下:

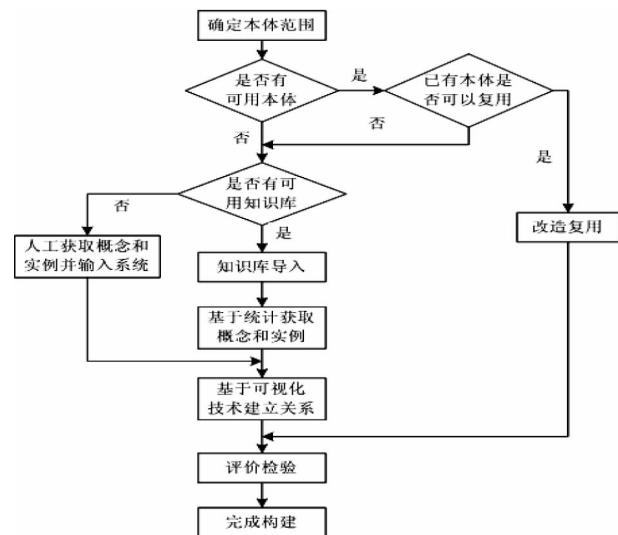


图1 领域本体可视化构建流程

3.1 确定本体范围 根据领域本体的应用目的,进行需求分析和信息收集。划定领域本体的范围。划定本体的范围是领域本体构建的初步内容,但是这一步直接关系到本体的适用性以及本体的构建规模和成本。是各种本体构建方法中必不可少的一步。

3.2 考察本体构建条件 本体的主要作用就是解决知识的共享和重用问题。在确定领域本体的范围后,首先需要判断在该领域中是否已经有可用的本体,如果有可用的本体,考虑是否加以复用。如果没有可用的本体,再考虑本体构建,但是在设计和建立领域本体之前也需要考虑是否具有充分的资料和准备,即判断是否有可供本体吸收的知识库,以便为本体构建提供充分语料,简化本体构建的难度。如果存在相关词表、语料库等相关知识库,借助计算机、统计分析、可视化等技术可以实现一部分自动化构建,提高本体构建的效率。如果没有相关基础,就需要衡量本体构建的难度是与本体构建重要性之间的关系,从而决定是否动手从零开始构建。因此,考察本体构建条件是本体构建前非常重要的一个环节。

3.3 获取概念和实例 本体是由概念(类)、实例(个体)和属性三个元素组成的。在三个组成元素中,概念是本体的骨架,实例是对概念的具体化,概念和实例是本体中的实体元素,而属性是描述性元素,是对概念进行描述和连接,概念之间的关系是通过属性加以

体现。在领域本体可视化构建方法中首先要确定本体的骨架,即首先列举出领域相关的概念。在以往的研究中已有研究者介绍了概念的多种获取途径,例如词表、词典、书籍、论文、网络等途径^[8]。在本方法中,根据有无可直接导入的知识库进行划分。如果存在电子版可用的知识库,借助一定的工具和算法可直接将知识库导入本体构建系统,根据知识库的特点通过预设概念提取算法,自动在导入的知识库中抽取概念,目前关于自动概念抽取的算法已经出现了大量的研究成果,将算法适当的修正便可应用于本系统中;如果不存在电子版可用的知识库或语料库,就需要人工手工录入概念,概念的来源依然可以参考一些较为成熟的词表和分类法,在手工录入的基础上逐步扩展概念集合,尽量全面的获取领域概念。

在本体元素输入构建系统后,采用可视化的方法按照统一标准,用图形表示本体的元素信息。例如利用圆圈表示概念,概念出现的频率越高,圆圈就越大;利用菱形表示实例;利用连接线表示关系,在线上注释关系名称等。

3.4 概念关系可视化建立 本体与一般知识组织方式的主要区别在于本体能够表达语义关系,因此,在本体构建过程中,非常重要的一步是发现和构建概念间关系。在概念集合的基础上,通过属性将概念连接起来,从而建立起语义关系。但是在概念关系建立的过程中,往往存在着关系模糊、缺乏共识、难于发现等问题。为了快速发现概念关系,同时实现多人参与、协同工作,达成领域知识共识,项目组开发了领域本体可视化构建平台,利用二维图形和表格辅助本体构建人员建立概念关系。在平台上进行的可视化操作包括:整体概念布局分析、局部概念可视化分析、概念关系直观构建、概念关系可视化协同构建等。

3.4.1 整体概念布局分析。概念之间的关系分为强关联和弱关联,对于本体构建来说,强关联是需要重点发现的关系,弱关联把整个领域知识连接成一张密度不等的网。为了使构建者可以在整体上把握概念分布,首先借助专利地图中的可视化方法,发现领域概念间的强关联。通过概念布局图将领域内部的概念分为关系更为密切的几个概念簇,然后以概念簇为单位进行关系构建。以导入语料库为例,在语料库中提取的领域概念,基于概念共现统计,绘制出本领域的概念分布情况,如图2所示,该领域概念大致分布在三个关系密切的簇内,簇内的概念关系即为发现的强关联,而

连接这三个簇的概念关系即为弱关联。

3.4.2 局部概念可视化分析。在整体概念布局分析的基础上,通过滑动圈选的方式,可人为的圈定某一个或几个概念簇作为重点研究对象,利用局部放大或关系表格等可视化方式,确立选定范围内概念间的关系。例如,在图2中圈定了某一个簇,通过查看两两概念间的共现次数来辅助本体构建人员具体确定两两概念间关系。在导入的语料库中,系统通过共现计算,统计出概念之间的共现次数,并且由低到高自动排序,从而将最有可能存在关系的概念显示在最前面。如图3,在导入的语料中,“蔬菜”和“黄瓜”共现次数最高为246次,构建人员在可视化分析的基础上,通过人为判断,可赋予“蔬菜”与“黄瓜”从属关系。

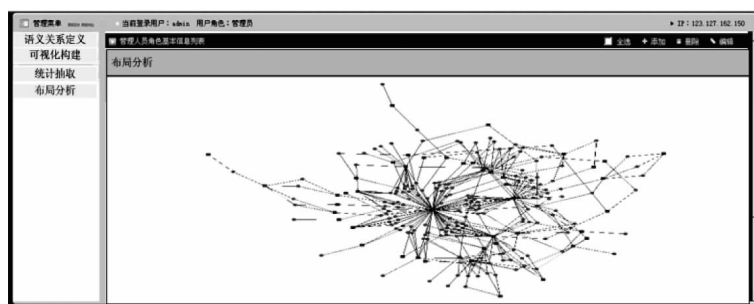


图2 整体概念布局图



图3 局部概念统计分析

为了快速确立概念关系,该可视化本体构建平台支持概念关系的批量赋予。例如,将“蔬菜”与“黄瓜”、“蔬菜”与“番茄”、“蔬菜”与“辣椒”等具有相似关系的两两概念进行批量勾选,然后再选择“从属关系”就可以上述两两概念间都赋予从属关系,提高关系建立的效率。

3.4.3 概念关系直观构建。当本体构建过程中没有语料库可作为概念关系确立的参考时,就需要人为的构建概念关系。首先,通过导入或人工录入的方式将领域概念集合输入本体构建系统,然后,利用概念关系的可视化构建功能人为的发现和建立概念间的关系。在系统中,首先圈定相关概念,被圈定范围内的概念以平铺的方式都显示在一个页面上,当选定一种关系后,通过连接线的方式将存在该关系的概念两两连接起来,系统前台进行可视化连接操作,系统后台自动

记录概念间的关系,如图 4 所示。通过可视化的方式,提高了关系建立的速度,使本体构建工作更为直观、简便,同时充满趣味。

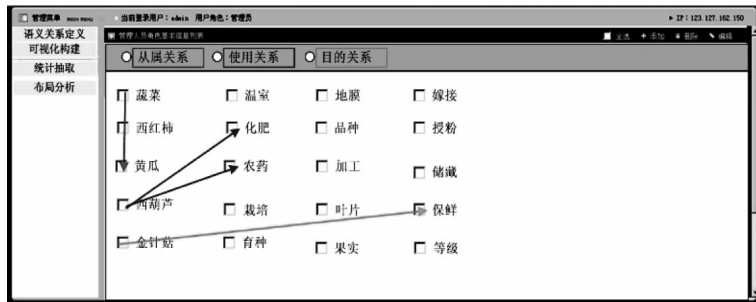


图 4 概念关系直观构建

3.4.4 概念关系可视化协同构建。本体是对领域知识的共同理解,因此专家参与和多人协同构建是本体构建所必须的。本体构建工作一般总是由一组人合作进行的,也正因为本体构建工作的合作特性,需要相关技术来为分布各地的小组成员共同构建本体提供便利。但是,目前多数本体构建工具只是单机工作,不支持在线多人协同工作和专家审阅修改。为了提高本体的准确性和可信度,本平台开发了可视化的本体协同构建功能,在系统中为本体构建者设置角色信息,根据角色的不同,设定不同的操作权限和可视化界面,从而满足不同角色的个性化需求。在本体构建过程中,常见的角色有管理人员、构建人员、审定人员等。

管理人员,关注的主要是本体的整体架构、进度、准确性等,因此,利用可视化技术为管理人员呈现出本体的整体化信息,同时支持他们深入细节进行查看和修改,还具有角色管理和权限分配功能,为其他本体建设人员分配角色和权限。如图 5 所示,可视化的本体构建进度管理功能。

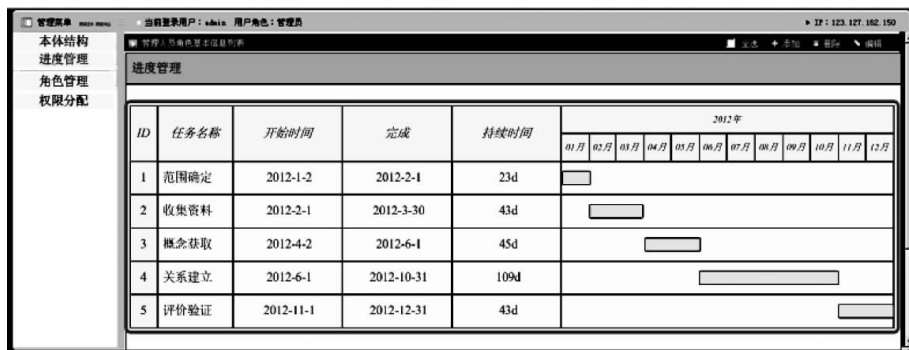


图 5 本体构建进度管理图

构建人员,往往关注的是具体构建任务,因此重点为他们显示局部的具体信息,提供局部可视化放大或统计图表支持,辅助构建人员完成具体的构建任务,如图 6 所示。

审定人员,多是指本领域的专家,他们在本体构建中的主要任务是对本体知识的准确性、专业性、权威性进行指导、审定、修正,因此审定人员需要全面、细致的

查看本体中的每个细节。因此,利用可视化技术为审定人员在一个界面内显示出尽可能详细的信息,支持信息的分屏、分段显示,辅助审定人员有步骤分阶段的完成本体的校验工作。

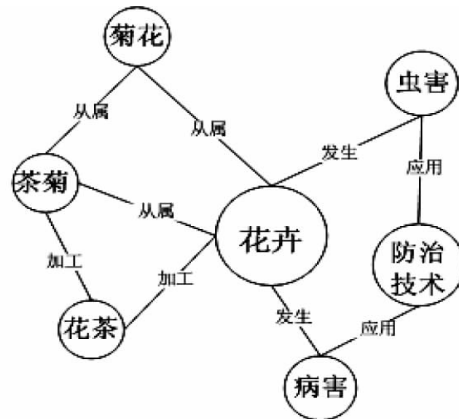


图 6 本体局部细节图

利用可视化的本体构建平台,支持多角色多用户的协同工作,在充分考虑用户特点的前提下,把用户的个人目标统一到本体构建的总目标下,充分发挥个人的专长和优势,共同完成领域本体构建任务。

3.5 本体评价和验证 经过前面几个步骤,已建立了一个初步的本体。但是由于本体构建人员有可能不是领域的专家,对于领域知识的理解存在一定的局限,因此初步建成的本体可能存在一定的错误、漏洞,这就需要进行本体评价和验证。在该过程中,邀请领域专家参与,从本体的正确性、一致性、可扩展性、有效性、本体的规模及描述能力等方面进行评价和验证^[9]。因此直观、全面、准确的展示信息是评价和验证的基础,在该过程中,可以利用可视化技术形成整个本体的网络模型,同时支持局部信息放大和浏览。通过直观视觉,达成知识共识。

4 总结与讨论

领域本体可视化构建方法相对于其他本体构建方法更为直观,通过直观的视觉信息,使本体构建者对领域知识达成共识。通过可视化技术使本体构建不

再是枯燥、抽象的文字工程,而是使本体构建的整个过程都实现了图形化呈现,基于系统统计和可视化图表辅助本体构建人员完成本体构建工作,利用图形标识和可视化动态连接等方法大大简化本体构建工作。基于该方法适合开展多人协作,多角色配合的大型领域本体构建。

(下转第 127 页)

第二,患者个人隐私问题。由于公众的办事信息具备隐私性,因此,用户将顾虑业务数据和个人身份信息收录于云计算平台。

第三,网络资源问题。由于云计算服务是基于互联网,所有的应用数据需基于远程网络传输,数据传输量激增,对带宽以及延迟性等指标提出了很高的要求,而现有的网络对于满足云计算实施和落实还有一定的距离。

面对上述的问题,平台在下阶段应着手解决上述平台中存在的缺陷。

a. 对于异构系统,将系统的各种服务功能通过 SOAP 协议封装成标准的 Web Services 服务进行信息交换,实现服务注册、服务接口、服务访问一体化的信息交换服务管理。

b. 对于公众而言,出于保证用户的安全考虑,给用户提供独占的私有资源以保证云的可信性。

c. 通过在端口和传统防火墙的地址拦截来智能地检测、分类和控制应用程序带宽。如为了限制过大的文件传输,系统可以制定相应策略,通过文件预定的大小限制来识别和限制 FTP 和 P2P 文件传输。

6 结语与展望

随着云计算的迅速发展,其便捷的管理与灵活的设计将为电子政务信息化建设带来了新的发展机遇,为解决政务信息化建设中信息资源的综合开发问题提供了崭新的思路。基于云计算的区域信息融合平台其成本低廉并且易于扩展,同时,对基层电子政务单位技术人员要求低,适合我国当前电子政务信息发展,适用于我国正在进行的政府机关体制改革。借助云计算的优点,构建区域政务信息融合平台既提升我国政务信息化水平,同时又创新政务服务模式,对公众提高政务

服务质量具有重要的意义。

目前,政务信息融合平台只是实现了对科技政务领域的信息资源的整合,为用户提供的信息具有粗放性与单一性的特点,并不能满足对信息资源知识性的服务需求。在现阶段建立的政务信息资源池的基础上,借助各类信息技术,对信息资源进行数据挖掘并形成知识服务模型将是下阶段平台建设的主要任务。

参考文献

- [1] 乔建忠. 基于业务关联的政务信息资源分类系统的研究与实现[J]. 现代图书情报技术, 2010(9): 28-36
- [2] Bram Klievink, Marijn Janssen. Realizing Joined-up Government - Dynamic Capabilities and Stage Models for Transformation[J]. Government Information Quarterly, 2009, 27(11): 275-284
- [3] 商晓帆. 电子政务中的信息资源整合研究[J]. 情报杂志, 2008(11): 145-147
- [4] 张玉涛, 夏立新. 基于主题图的电子政务信息资源整合模型研究[J]. 情报杂志, 2009, 28(7): 161-165
- [5] 夏立新, 叶飞. 利用主题图实现城市圈电子政务系统数据整合模型的构建[J]. 现代图书情报技术, 2009(11): 17-22
- [6] 鲁俊杰, 侯卫真. 面向信息资源整合的电子政务云平台构建研究[J]. 图书馆学研究, 2012(13): 36-40
- [7] Ian Foster, Zhao Yong, Ioan Raicu, et al. Cloud Computing and Grid Computing: 360-degree Compared[C]. Grid Computing Environments Workshop, 2008: 1-10
- [8] Malden A Vouk. Cloud Computing - issues, Research and Implementations[J]. Journal of Computing and Information Technology, 2008(4): 235-246
- [9] Luis M Vaquero, Luis Rodero-Merino, Juan Caceres, et al. A break in the Clouds: Towards a Cloud Definition[C]. Acm SIGCOMM Computer Communication Review, 2008
- [10] 冯登国, 张敏, 张妍, 等. 云计算安全研究[J]. 现代情报, 2011, 31(4): 71-83

(责编: 刘影梅)

(上接第174页)

在本研究中主要讨论了基于二维可视化技术的本体构建方法。随着计算机图形技术的发展, 三维和多维可视化技术逐渐成熟, 并且已经初步应用于各领域。因此, 今后可以考虑将相关技术和方法应用在本体构建过程中, 建设三维或多维展示的本体构建平台, 使本体构建和展示更为直观、形象、生动。

参考文献

- [1] Robert Spence. Information Visualization: Design for Interaction[M]. Englewood Cliffs: Prentice Hall, 2007
- [2] E H Chi. A Framework for Visualizing Information[M]. [S L]. Kluwer Academic Publishers, 2002
- [3] Schnotz W, Kurschner C. External and Internal Representations in the Acquisition and Use of Knowledge: Visualization Effects

on Mental Model Construction[J]. Instructional Science, 2008, 36(3): 175-190

- [4] 黄烟波, 张红宇, 李建华, 等. 本体映射方法研究[J]. 计算机工程与应用, 2005(18): 27-29, 33
- [5] 张志刚. 领域本体构建方法的研究与应用[D]. 大连: 大连海事大学, 2008
- [6] 韩婕, 向阳. 本体构建研究综述[J]. 计算机应用与软件, 2007, 24(9): 21-23
- [7] 杜文华. 本体构建方法比较研究[J]. 情报杂志, 2005(10): 24-25
- [8] 李光达, 常春. 构建本体时获取概念方法研究[J]. 情报科学, 2009(05): 713-716, 722
- [9] 刘琳娜, 薛建武, 汪小梅. 领域本体构建方法的研究[J]. 情报杂志, 2007(4): 14-16

(责编: 贺小利)