

主题图和本体在知识化金融培训系统构建中的应用研究*

□ 朱锐 李颖 / 中国科学技术信息研究所 北京 100038

孙卫 / 科学出版社 北京 100717

摘要: 针对金融行业培训系统课程资源的多样性、资源间缺乏有效组织而造成的学习资源难以重用、不能满足个性化用户学习需求的现状, 该研究提出了一种基于知识工程技术的实用化解决方案。文章首先对比分析了主题图和本体的技术和应用特点; 其次, 融合和利用它们在资源的结构化、知识化描述等架构方面的不同优势, 设计了基于主题图和本体的金融知识体系框架。文章还分别采用主题图和本体最通用的开源工具 Ontopia 和 Protégé, 初步构建了金融领域知识网络的原型, 以示范说明主题图和本体在整个金融知识体系结构中的应用, 为后期大规模开发知识化金融培训系统奠定了基础。最后, 简要阐述本研究的结论、相关评价和今后的研发工作。

关键词: 主题图, 本体, 知识组织, 语义Web, 语义技术, 金融培训, e-Learning

DOI: 10.3772/j.issn.1673-2286.2012.04.004

1 背景

随着金融行业竞争越来越激烈, 几乎所有的金融企业都会针对不同岗位的企业新老员工进行培训, 来提升他们的专业技能和整体素质, 从而提高企业自身的竞争力。然而, 传统的培训系统在建设时, 由于没能实施有效的知识化组织, 学习资源会在不断的使用过程中, 变得越来越庞杂无序, 使得金融培训系统对资源的综合利用效率很低。教师很难利用这些非结构化资源来灵活、高效地重组和设计新的培训

课程, 学员也不能根据自身的知识能力和学习风格有效检索、获取所需的课程资源来学习。具体来说, 目前的金融培训系统主要存在以下两个较为突出的问题:

1) 系统缺乏对金融培训资源的知识化组织。对于一个较大的银行或者金融企业而言, 系统通常都拥有十分丰富的培训资源, 包括各类课程、试题以及海量的网络资料。然而, 由于这些资源分布十分分散, 资源间又没有潜在的语义关联, 极大地影响了这些资源在学习平台上的共享和利用。要解

决这一问题, 需要系统地对这些学习资源进行知识化组织, 建立起系统、结构化的课程体系, 并开发相应的导航和检索工具, 提高资源的利用率。

2) 系统不能满足用户的个性化培训的需求。对员工而言, 每个人自身的知识能力和学习风格各不相同, 而对于数字资源而言, 最大的问题在于用户很难查找自己所需的资源内容^[1]。要向学员提供个性化培训, 就要建立一套科学的培训系统, 为不同层次的学员提供个性化的学习课程。一方面, 需

* 基金项目: “十二五科技支撑计划——科技知识组织体系共享服务平台建设”资金支持(编号: 2011BAH10B03-2); 中国科学技术信息研究所“汉语科技词系统建设与应用工程”重点工作支持。

要对现有的资源进行加工,降低课程资源的信息粒度(information granularity),同时对小粒度的资源模块进行重组,以达到资源再利用的目标;另一方面,需要为系统建立一套基于资源内容的主题导航,通过导航来正确指引不同的学员浏览和学习不同的知识模块。

在金融培训资源里,最有价值也是最重要的部分是概念或者知识。概念和概念间的关系是学习对象资源的原子结构,能否对概念以及概念间的语义关系进行准确有效的描述和组织,是影响学员从海量信息中发现和获取自己所需知识的主要因素之一,同时也是对学习资源进行重用来构建知识化金融培训系统的关键。

主题图和本体作为两种国际标准化知识模型,能抽象和描述应用领域的概念知识以及概念间的语义关系,随着高度结构化的知识库在知识化系统中的出现,对于领域应用而言,它们变得十分重要。本文试图吸收综合主题图和本体的优势,深入分析金融领域的特点和金融培训系统的问题,设计基于主题图和本体的知识化金融知识培训系统,希望以此解决传统信息系统存在的问题。

2 主题图和本体的概要及优势对比分析

2.1 主题图和本体概要

2.1.1 主题图

主题图(Topic Maps, TMs),是一种用于描述信息资源知识结构的元层(Meta-Layer)方式^[2],也是一种关联知识结构^[3]。TMs由主

题(Topic)、关联(Association)、资源指引(Occurrence)组成,通过XTM(XML Topic Maps)语言来描述概念与信息资源之间的关系^[4]。从知识组织结构来看,TMs分为信息层(information layer)和知识层(knowledge layer)两部分,信息层是所有物理资源的集合,资源层是主题和关联的集合,两层结构通过资源指引进行链接。从功能上看,TMs可以表示知识概念间的相互联系,也可以定位某一知识概念所对应的资源位置。主题图技术是一种基于主题的知识描述和知识导航技术,同时是一种十分灵活的知识模型。

2.1.2 本体

本体(Ontology),通常被称为领域模型(Domain Model)或概念模型(Conceptual Model),是对共享概念模型形式化的、明确的、规范的说明^[5]。作为语义网一种重要的表示概念的模型,从知识组织角度上说,本体解决的是如何将相关领域的共有知识形式化、概念化,实现对领域知识的推理,同时使计算机能够理解。从知识共享的角度上看,本体提供了一个通用概念定义集合,是在各种知识系统间表达和交流某些主题知识的共同语言^[6]。

2.2 主题图和本体的优势对比分析

主题图和本体都可作为信息、知识的底层架构技术,都能用于组织较高层次的知识抽象,两种技术联系紧密,但它们有着各自不同的技术标准和概念模型,在各自

的发展中表现出不同的特点和差别。本体优势在于捕获领域概念知识,描述概念和概念间的关系,产生基于逻辑的知识推理,然而,领域本体的构建和维护成本耗费巨大,而且需要建立复杂的本体映射才能提供良好的知识导航,从这个角度上看,本体如果不进行轻量化处理,就很难在实际企业级的知识系统中得以应用。主题图的优势在于可构建以主题为中心的概念关联模型,从海量的Web信息中集成与主题相关的信息,通过信息层(information layer)和知识层的无缝耦合向用户提供友好的知识分类和导航。目前,在实际应用中,这两种技术在不断地趋于融合。表1分别就两种技术的发展特点、相关标准、领域应用以及各自的优缺点等方面进行对比分析。

3 基于主题图和本体的金融知识体系架构设计

鉴于主题图和本体两种技术的特点,本研究认为:要实现金融培训资源的结构化组织、向学员提供个性化知识导航服务,可以采用其中一种技术或者结合两种技术构造混合的信息模型。并且,要真正发挥其在金融领域中的应用,就需要充分考虑金融培训系统的资源特点,通过结合采用主题图和本体在概念语义描述和知识导航方面的优点,辅以各种知识工具,来组织和结构化金融培训资源,这是一种可行的解决方案。图1设计了一种基于两种技术的金融知识体系架构,主要包含有三层:对象资源层、知识组织层、语义化导航层。

对象资源层:学员培训或者自主学习所使用的各种学习资源,也

表1 主题图与本体的对比分析

	主题图 (Topic Maps)	本体 (Ontology)
相关概念	主题 (topic)、关联 (association)、资源指引 (occurrences)、范围 (scope)、主题同一性 (subject identity)、导航 (navigator)、分面 (faceted)。	陈述或命题 (statement)、公理 (axiom)、概念或类 (concept、class)、属性 (property)、函数 (function)、实例 (instance)、层次结构 (hierarchy)、个体 (individual)。
目的	最初是为了对日益激增的海量信息进行管理。随着发展, 主题图在不断进行功能补充和丰富。	捕获领域知识, 提供对领域知识的共同理解, 确定共同认可的词汇, 并从不同层次的形式化模式上给出这些词汇 (术语) 和词汇之间相互关系的明确定义。
技术基础	元数据 (Metadata)、XML、XTM、资源描述框架 (RDF)。	元数据、XML、资源描述框架 (RDF) 和 OWL (OWL Lite、OWL DL 和 OWL Full)。
相关标准	ISO/IEC 13250 1-7、ISO 18048、ISO 19756 (包括主题图数据模型、主题图规范标准、主题图压缩语法、主题图图形模型、主题图查询语言、主题图约束语言等) ^[2] 。	ISO/IEC 13763 -3、ISO/IEC 24707、ISO/IEC FDIS 19763-3、ISO ISO/TS 15926-8 (Web Ontology Language, 即 OWL、SPARQL、RDFs 等)、SUMO、IEEE1074、WSMO、SWSF、OWL-S ^[7] 。
技术特点	主题 (Topics) 与关联 (Associations) 的集合; 主题间的关联关系可自由定义; 可链接到有关主题的信息资源; 具有唯一识别主题的机制, 比如 URI; 可设定有效的范围 (Scope); 具有合并功能; 容易具体化。	本质是从对象世界中分离出基本概念, 形成概念的集合。概念间包括上下位关系 (is-a 关系) 和其他关联关系。即本体抽出概念和关系的定义, 对语义限制公理化。提供十分丰富的描述概念间关系的关系实体, 其逻辑描述语言 OWL 让本体具有一定的推理能力。
应用	语义门户、e-Learning、电子政务、业务流程建模、分类管理、元数据管理、信息集成、企业知识化管理、知识导航、电子出版等。	本体论工程 (如何创建特定领域本体; 通用本体的创建)、本体表示、转换、集成等, 主要用于智能检索、人工智能、软件工程、电子商务、语义门户构建 ^[8] 。
优势	可以利用 XTM 构建, 灵活性高, 较 Ontology 简单, 构建维护主题图成本相对较低, 产出投入比大。利用主题来描述知识; 很好地建立主题与资源的关系, 具有一定的知识导航功能; 构建与维护主题图的成本较小。	可以利用 OWL 构建, 同时还需要相关领域专家的人工参与, 构建与维护领域本体耗费成本巨大, 产出投入比很小。有一定的逻辑推理功能, 同时由于其层级概念关系, 方便人们整体上快速了解该领域的知识以及各个知识之间的关系。
局限	目前, 主题图主要利用 XTM 构建, 不特别强调推理功能, 虽有概念关联关系, 但是对于描述关系的实体很少。	本体构建和维护成本巨大, 目前重点在概念与概念间关系的描述, 重点不是为系统提供知识导航 (需要建立复杂的本体映射)。

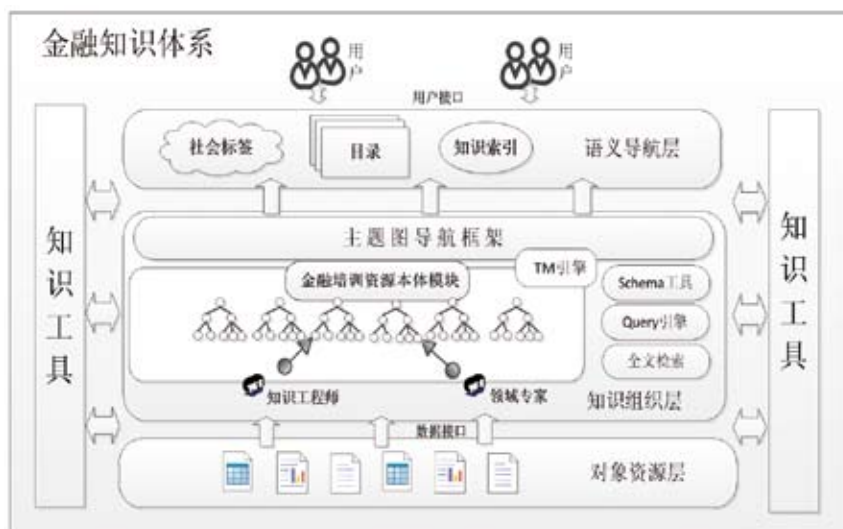


图1 基于主题图和本体的金融知识体系

是系统需要知识加工的对象资源基础。内容主要涉及金融相关的产品、企业流程、文本课程、相关理论书籍、金融法律法规和规范文件、行业研究报告、课程试题等类型; 格式涉及有文本、图表、公式以及图像等等。

知识组织层: 领域专家和知识工程师对基础对象资源进行抽取和模块化处理, 抽象、列举其中的领域概念以及概念间的关系, 建立金融培训资源本体。并在资源本体的基础上融入主题图引擎、Schema 工具、主题图导航、Query引擎、全

文检索等模块,其主要作用是对底层分散的基础资源进行结构化,对知识模块的深度组织和加工,构建金融知识系统的元数据以及领域本体,并结合主题图技术和相关共同组成知识组织层,来为最终系统服务提供知识库、知识分类和检索基础支持。

语义导航层:主要包括用户检索系统、标签云、网站主导航和子导航系统、分面引导目录等等,该层需要利用主题图来搭建最终的金融知识主题关联结构,使得最终学员用户可以通过接口来明晰相关的知识结构,同时能根据学员学习的进度和需求,以个性化的方式从不同的知识库中确认、定位和提取相关的知识内容资源。

这些层级体系主要利用知识工具来对资源形成模块进行加工组织、概念描述和知识导航。这些工具既包含自动分类工具、知识模块加工工具、词表管理工具、内容管理工具和自动标签等工具来使用人工定义的规则或者模式匹配算法自动为文件配置元数据(相当于根据分类法把文件指定到某些类别下,主要可用于自动分类、自动索引、自动标注、聚类),也包括搜索引擎和分析软件。

4 主题图和本体在金融知识体系结构中的应用探索

随着与主题图和本体相关标准和技术的发展,目前,两种技术都有相应的工具来进行编辑、管理和维护。主题图方面,比较流行的工具有AJAX Topic Map、TM4J TopicMap、OKS Samplers(包含Omnigator、Ontopoly和Vizigator)、Wandora、Wordpress Topic Maps

(WP2TM),功能涉及主题图的编辑、管理、合并、可视化等等^[9]。本体方面,目前比较常见的工具主要有Protégé、Swoop、OntoEdit、OILED、WebODE、Altova SemanticWorks、KAON、Ontolingua、PoolParty^[10,11]等等,其功能涉及本体的编辑、映射、合并、可视化分析以及本体建模、推理等。相比其他工具而言,OKS Samplers和Protégé都是根据Windows应用程序风格进行设计,开发人员更容易进行导航和管理操作。因此,本文将采用这两种工具和相关插件,用以对比分析两种技术在知识组织和可视化方面的差别。

4.1 主题图在金融体系架构中的应用探索

OKS Samplers采用Ontopia主题图引擎,并通过它提供API来开发基于XTM标准的应用^[9],它能很好地支持以主题为中心、为海量金融培训资源创建知识化门户。利用OKS Samplers来构建金融培训资源主题图,不仅可以利用Ontopoly将金融培训资源相关的主题类型层级结构和主题关联转化为金融培训资源主题图本体,还可以结合主题类型和关联类型对主题创建相关主题和实例,最后形成完整的一个金融培训资源主题图。图2利用Omnigator显示了金融培训资源主题图中以金融市场为中心的主题类型层级结构、关联类型等,而该主题类型的具体资源通过内部资源指引(Internal Occurrences)进行描述。

这种以主题为中心的知识组织方法能将所有与金融培训资源的关联节点可视化显示出来,同时,用户还可以根据自己的兴趣对中心主题

的不断变换选择而进行重新调整和排序,这样既易于学员对于整体知识的把握,同时也可以根据具体的主题的关联来明晰相关概念主题的知识网络,从这个角度来看,主题图在金融e-Learning知识培训体系中可以作为一个知识图来为学员提供知识脉络的导航。

同时,在金融体系框架中,以主题图技术思想为主的知识系统能为金融企业e-Learning培训带来以下几点优势:(1)激发学员学习主题相关的知识;(2)易于对整体知识概念框架的理解;(3)学员较容易发现有趣的学习资源内容;(4)可以对具体学习效果进行可视化评估;(5)学员可以自主设计与自身知识结构相符的学习途径;(6)通过主题图的可视化导航,容易对与主题相关的整体知识结构的把握;(7)能够很容易地发现和确定必需的基础知识和相关的扩展知识。

OKS Samplers的Omnigator浏览器可以对任意的主题图进行显示,如图3是香港交易所相关的部分知识和资源的可视化显示。学员在e-Learning学习过程中,当浏览到与“证券交易所”主题时,系统就会将该主题相关的所有主题以图形方式显示出来,并能揭示出该主题与其他主题间的语义关联。例如,利用主题图的“股票市场”和“地区”、“产品”建立的关联关系,学员可以从主题图的图导航中很快捷地发现与证券交易所相关的知识以及资源,比如,世界上存在证券交易所以及其所经营的产品业务、所在地方以及每个证券交易所目前的指数等等,如果点击那个地区,则会出现与该地区相关连的任何资源,如与“香港”相关的金融机构、香港资金市场、外汇市场、黄金市场



图2 金融市场方面知识概念结构在ontopia上的开发界面

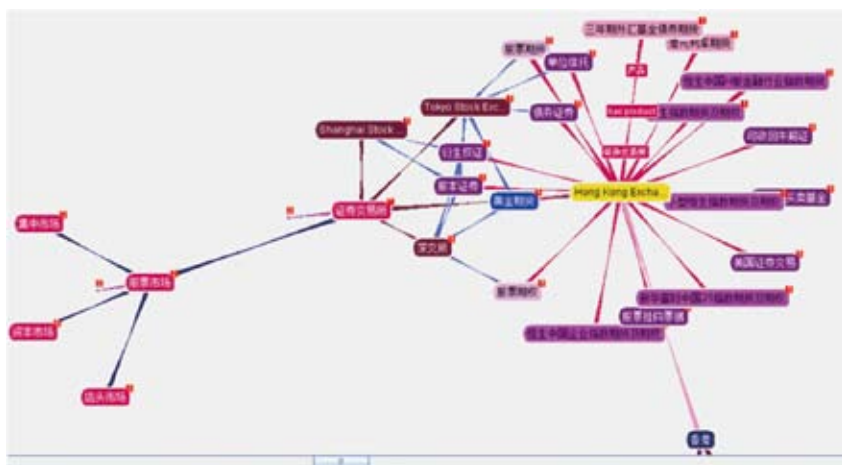


图3 Omnigator上以“香港交易所”为中心的主题图图形化显示

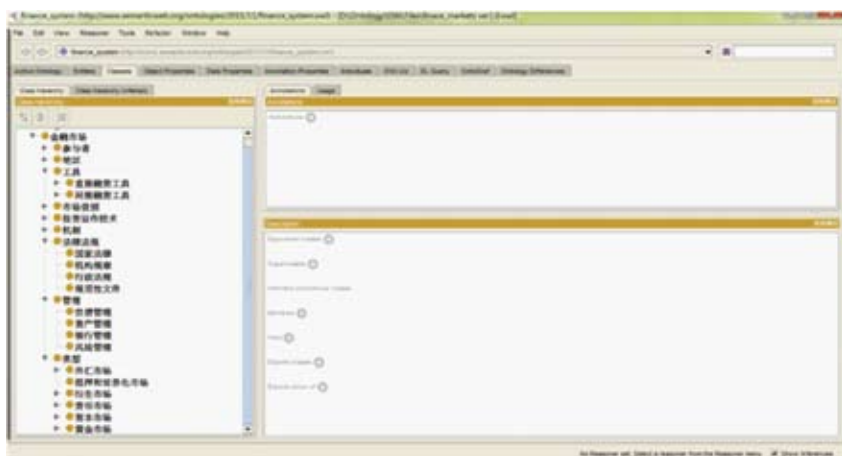


图4 Protégé 4.2上金融市场知识概念结构的可视化显示

以及金融监管等主题的知识关系网络和资源。

4.2 本体在整个金融体系结构中的应用探索

Protégé是斯坦福大学利用java语言开发的一个开源本体工具，它集成了本体编辑和知识库编辑，同时Web上还有许多以该软件为基础开发的插件，它以OKBC模型为基础，具有开放的接口，提供大量的插件，支持包括XML、EDF(S)、OWL、OIL、DAML、DAML+OIL等形式的本体表示语言，具有不错的可扩展性，用户可以很方便地使用它来进行本体和知识库的建设。

图4是一个利用Protégé构建的金融培训本体的开发工具主界面，在金融培训资源本体的构建中，开发人员根据金融培训资源本体中的重要术语、概念建立本体框架，利用树形方式创建类和相关的概念层级结构，并对类的属性槽（Slot）以及取值类型进行定义，最后对领域本体编码、形式化，形成规范化的概念模型。

Protégé也可以通过插件的形式来为构建好的概念层级结构提供图形化显示。在图5中，左侧显示的是金融市场领域的树形概念层级结构，右侧显示的是通过插件OntoGraf对应生成的概念图形结构。可以看出，金融培训资源本体的构建符合软件开发过程，不依赖于具体的软件而存在，其形成领域通用模型具有极高的可重用性，方便在其之上进行开发和利用。同时，本体在依照领域专家的描述建立后也并不是一成不变，它需要根据领域研究的不断深入而进行反复调整，这个过程需要投入大量的资源。

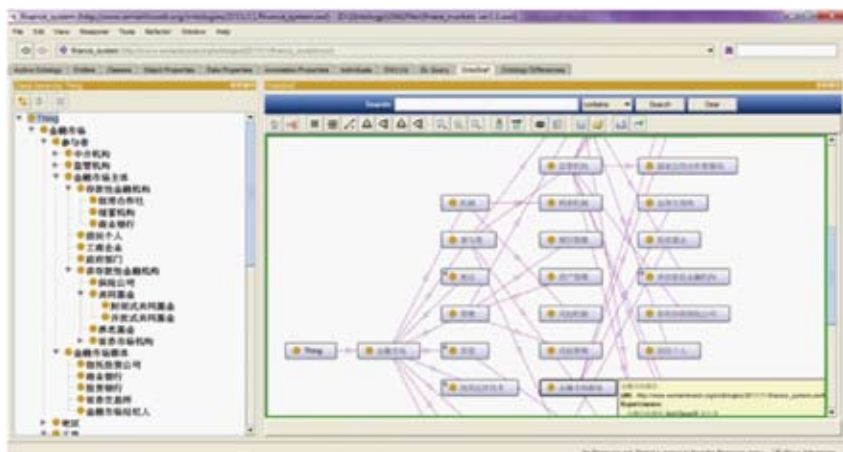


图5 金融市场方面知识概念结构在Protégé 4.2上的概念层级结构的显示

5 结论

本体是知识化金融培训系统的知识组织框架,主题图的建立以及构建系统智能化检索都需要以本体为基本依据。系统通过本体来描述金融培训领域的概念模型,可以使底层资源对象更容易被机器理解和共享,有助于为学员自主学习提供相关概念的知识层次结构,提供查询扩展,增强知识的语义检索查询。

主题图通过主题、关联、资源指引建立对象资源的结构化表示模型,同时资源层和知识层的结合对资源中蕴含的知识结构进行知识索引和导航,定义好的主题概念网络不但能用来描述和组织金融培训资源,还可以揭示培训资源的内容特征,从而实现以主题为中心聚合信息与知识,为学员个性化学习提供主题相关的知识网络和资源指引。

实际企业级应用中,通过综合考虑资源描述的灵活性、构建与维护的成本投入、用户的知识导航效果等因素,企业一般更愿意采纳主题图或融合这两种技术,为知识化系统提供轻量级领域本体,并以此为基础生成灵活的主题图模型,将传统的“以课程为中心”的课程模式转变为“以主题为中心”的个性化学习模式。

6 评价与未来工作

主题图和本体两种技术在实际应用的效果可以从建设投入成本、对概念的知识描述与知识导航三个方面进行考虑:建设投入成本上,在实际应用中,领域本体建设需要许多领域专家消耗大量的时间和资源,成本投入巨大,领域本体建设完后,仍需不断对其进行维护和管理,主题图的建立没有本体严格,

投入较之要少;从对概念的知识描述效果来看,本体较主题图的概念形式化程度更大,能更明确层级化概念表示和整体知识结构,主题图更灵活,而且通过知识层和资源层以及层之间的关联的建立揭示主题、资源间的内容特征;知识导航方面,主题图较本体能提供效果更好的概念导航以及资源指引,可视化显示的方式也更加灵活。

本文对基于主题图和本体的金融知识系统研究还处于实验原型状态,存在着许多不足,本文虽然参考两种标准建立了金融培训资源主题图和本体,但基本是手工建立,工作量较大;同时,本文在评价两种技术的过程中没有建立一套完整的系统评价体系。这些有必要在未来工作中进行深入的探讨。例如,在主题图本体建设方面,可以尝试利用知识抽取、主题聚类等方法对现有的实验原型进行扩展,实现半自动化甚至自动化方式来创建;在实际系统应用方面,也需要结合主题图和本体这两种技术,尝试建立一套完整通用的知识组织指标评价方法和体系。

作为本研究未来的后续工作,将结合前期研究和原型实验取得的成果,开发实用性知识化金融培训系统,以满足金融领域培训资源的有效重用和个性化获取需求,并探索该方法论在其他领域的扩展应用。

参考文献

- [1] PEPPER S. Understanding Topic Maps: Towards a Subject-Centric Revolution [OL]. [2012-02-10]. <http://www.topicmaps.com/tm2008/pepper2.pdf>.
- [2] 李颖,练霞,朱礼军.主题图的概况及应用研究(上篇)[J].数字图书馆论坛,2010(5):20-27.
- [3] PEPPER S. The TAO of Topic Maps: Finding the Way in the Age of Infoglut [OL]. <http://www.ontopia.net/topicmaps/materials/tao.html>.
- [4] ISO/IEC 13250:Topic maps information technology, document description and processing languages [S/OL]. <http://www1.y12.doe.gov/capabilities/sgml/sc34/>

document/0322.htm.

[5] BORST W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse [D]. University of Twente, Enschede, 1997.

[6] 李善平, 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.

[7] 李景. 本体技术标准化综述研究[J]. 现代图书情报技术, 2007(8): 12-17.

[8] 甘建侯, 姜跃, 夏侯明. 本体方法及其应用[M]. 北京: 科学出版社, 2011.

[9] [OL]. [2012-02-10]. <http://www.topicmapslab.de/projects>.

[10] MALIK S K, et al. Developing an University Ontology in Education Domain using Protégé for Semantic Web [J]. International Journal of Engineering Science and Technology [J]. 2010, 2(9): 4673-4681.

[11] [OL]. [2012-02-10]. http://techwiki.openstructs.org/index.php/Ontology_Tools.

作者简介

朱锐, 中国科学技术信息研究所情报学全日制在读硕士。研究方向: 知识导航、主题图等。E-mail: zhur@istic.ac.cn

李颖, 信息系统专业博士。相关研究课题: 语义知识组织、基于主题的知识组织技术的应用等。E-mail: liying@istic.ac.cn

孙卫, 中国科学技术信息研究所, 计算机高级工程师, 中国计算机学会高级会员, 北京通信学会理事。目前主要在知识组织、知识挖掘、知识处理技术上进行研究及教学, 在科学出版社(中国科技出版传媒股份有限公司)从事数字出版的相关工作。E-mail: sunwei0125@gmail.com

Application Research of Topic Maps and Ontology in Building an Knowledgeable Financial Training System

Zhu Rui, Li Ying / Institute of Scientific and Technical Information of China, Beijing, 100038

Sun Wei / Science Press, Beijing, 100717

Abstract: The diversity and lack of effective organization of resources of financial training systems has made themselves difficult to reuse, and can not meet the learning needs of individual users. This study presents a practical solution based on knowledge engineering. First, this article analyse technical and application characteristics of Ontology and TopicMaps. Second, it designs an financial knowledge system framework based on Ontology and TopicMaps. In order to demonstrate and illustrate the role of the two techniques in the whole financial system, it also initially builds financial domain knowledge network structure by Protégé and Ontopia, which provides good guidance for developing knowledgeable financial training system. Finally, it briefly describes the conclusions, evaluation and future related R&D work of this study.

Keywords: TopicMaps, Ontology, Knowledge organization, Semantic web, Semantic technology, Financial training, e-Learning

(收稿日期: 2012-03-15)