

学习元平台的语义技术架构及其应用*

□杨现民 余胜泉

摘要：语义 Web 强调对网络资源附加可供机器理解和处理的语义信息，使人 与电脑之间更好地协同工作，同时也使资源的大规模重用和自动化处理成为可能。近年来，随着语义 Web 技术的不断发展，各种 e-Learning 平台（适应性学习系统、智能导师系统、语义化知识社区等）纷纷尝试引入语义技术，为用户提供更加个性化和智能化的学习服务。学习元平台（LCS）基于语义化组织的学习资源，应用语义 Web 体系框架中的本体技术和推理技术，构建了包含 Ontology API 层、语义层、标注层和推理层的语义技术框架。依托该框架，LCS 分别在学习资源的有序进化控制、学习资源的动态语义关联、学习资源的动态语义聚合、学习资源的适应性推荐、学习资源的语义化检索以及社会认知网络的动态构建等方面进行探索性研究，实现从资源的语义化组织到具有更强智能性和个性化的高级应用服务。这些应用实践表明，将语义技术在 e-Learning 系统中推广应用，需要积极引入外部开放、成熟的本体和更多高级应用服务，集成具有较强针对性和运行效率的推理机，在人工标注和自动标注相结合的基础上，关注资源和用户的语义化表征，重视情境模型的设计和情境本体的构建。

关键词：学习元；学习元平台；e-Learning 平台；语义 Web；语义技术框架

中图分类号：G434 **文献标识码：**A **文章编号：**1009-5195(2014)01-0089-11 doi:10.3969/j.issn.1009-5195.2014.01.013

***基金项目：**江苏省高校哲学社会科学基金项目“服务终身教育的泛在学习环境研究”（2013SJB880033）；国家自然科学基金项目“泛在学习的资源组织模型及其关键技术研究”（61073100）。

作者简介：杨现民，博士，硕士生导师，江苏师范大学教育研究院（江苏徐州 221116）；余胜泉，教授，博士生导师，北京师范大学教育技术学院（北京 100875）。

一、引言

语义 Web 强调对网络资源附加可供机器理解和处理的语义信息，可以使人与电脑之间更好地协同工作，同时也使资源的大规模重用和自动化处理成为可能，是实现未来智能型网络的数据基础。近年来，随着语义 Web 技术的不断发展，各种语义 Web 应用项目开始不断涌现，如 DBpedia、FOAF、SIMILE、Linking Open Data (LOD)、GoPubMed 等。Google、Yahoo、Microsoft 等 IT 商业巨头也开始纷纷进军语义 Web 领域，开始出现一些较为成熟的语义 Web 产品，比如 Freebase、Google Base、Twine、Semantic Mediawiki、Powerset、SearchMonkey 等。这些语义 Web 产品的面世使得语义 Web 不再神秘，越来越融入人们的日常生活。

由于语义 Web 技术在知识表示、信息共享、智能推理上的优势，引起了 e-Learning 领域研究者的广泛关注。各种 e-Learning 平台（适应性学习系统、智能导师系统、语义化知识社区等）纷纷尝试引入语义技术来为用户提供更加个性化和智能化的学习服务。学习元平台（Learning Cell System, LCS）（杨现民等，2013a）是为泛在学习环境设计开发的一种开放知识社区（<http://lcell.bnu.edu.cn>），包括学习元、知识

群、知识云、学习社区、个人空间、学习工具等六大功能模块。LCS 以学习元为基本的资源单元，学习元（余胜泉等，2009）是一种语义化组织的学习资源，多个学习元可以聚合成知识群。LCS 在系统设计初期便充分考虑了语义技术的优势和应用模式，经过一年多的设计、开发和不断优化，LCS 在语义技术架构和应用上做了有价值的探索。本文重点分析学习元平台中语义技术的整体架构及其应用，最后从语义技术的角度对 e-Learning 系统的发展提出了几点启示，期望能对语义 Web 技术支持下的 e-Learning 系统设计、开发和优化提供一定的借鉴与参考。

二、语义技术在 e-Learning 领域的应用现状

目前，语义 Web 技术中的本体技术和推理技术在 e-Learning 领域的应用最为广泛，主要集中在自适应系统开发、学习资源管理与共享、适应性资源配送和个性化学习内容推荐、语义检索、智能答疑等方面。

基于本体设计和开发新型的 e-Learning 系统已经成为国内外研究的热点。曹乐静等（2005）利用本体和 Web 服务技术构建了一个四层架构（资源层、语义层、集成层、服务层）的开放适应性 e-Learning 系统，利用本体来描述学习资源的语义，通过 Web 服务支持个性化学习和系统间的资源共享。刘卫红等

(2006)提出了基于本体的e-Learning系统层次结构模型,该系统模型共分为五层结构,自下而上依次为数据层、元数据层、本体层、知识层和服务层。基于本体的e-Learning系统与一般e-Learning系统在架构上的主要区别是在数据层之上增加了本体层,通过构建不同类别的本体来对资源进行一致性、规范化的语义标注。另外,除了基于本体的新型e-Learning系统的开发外,国外学者Dos Reis Mota等(2010)开发了一个智能代理,可以将基于本体和e-Learning标准的适应性学习系统与当前国际流行的Moodle平台进行整合,辅助实现更加有意义的个性化学习。如何将传统e-Learning系统与基于本体的适应性e-Learning系统进行快速、无缝整合,充分发挥二者的优势,将成为e-Learning的重要发展方向。

学习资源的设计与共享一直是e-Learning领域的重要研究方向。基于本体的教育资源具有权威性、规范性、可共享性等特点。(刘革平等,2009)基于本体的资源组织与管理方面,已经出现了大量的研究。邵国平等(2008)在分析现有e-Learning在资源管理方面存在的资源缺乏关联性、检索效率低、描述标准不一致等问题的基础上,认为通过学习课程本体建模、智能搜索代理和个性化内容导航等关键技术可以促进e-Learning资源管理。李艳燕博士(2005)基于语义Web的思想,提出了一个基于语义的学习资源组织模型。Nilsson等(2002)对基于本体技术的元数据方案和传统的基于静态词汇表的元数据方案进行了对比研究,建议用RDF代替XML来对元数据进行标注,使e-Learning中所用到的元数据更为灵活和动态,并降低多个基于不同元数据的系统互操作和共享的难度。Jovanović等(2007)认为现有的对学习内容的描述规范(如LOM)和对学习活动的描述规范(如IMS-LD)不足以包含高级学习服务所需要的全部信息,提出了一个针对学习对象(Learning Object)的本体描述框架,来描述学习对象的上下文信息。

上面提到应用本体技术对学习资源的内容和结构进行语义描述,其实,除了对资源进行本体化的描述外,还可以对学习者的上下文环境等进行语义建模,从而实现e-Learning资源的适应性推荐和用户的个性化学习。Bouzeghoub等(2006)利用基于本体的领域模型、学习者模型、学习对象模型来描述一个适应性的学习系统。黄海江等(2007)研究了基于本体的学习内容个性化推荐,指出个性化推荐设计的一般思路为:(1)建立用户个性化的兴趣模型——用户概貌(User Profile)、兴趣模板或兴趣剖面;(2)根据用户概貌对新信息进行过滤,把用户可能感兴趣

的信息自动推送给用户;(3)根据用户的反馈,对兴趣模型进行修正。Zeng等(2009)认为用户知识需求的获取对于适应性学习系统来说非常重要,他们基于课程本体研究了如何利用问答历史记录和阅读历史记录来获取用户的知识学习需求,并通过两组实验验证了此方法的有效性。陈和平等(2007)设计了一个基于Ontology和JENA推理机制的个性化e-Learning系统模型,并阐述了推理技术在个性化e-Learning系统中的应用。Diaconescu等(2008)通过语义Web和JENA Rules技术的结合来增强Moodle平台的功能,研究了如何在当前Moodle课程资源的基础上使用JENA规则进行内容推理,实现课程内容的适应性推荐。孔德华等(2006)提出了一种新的基于XML的适应性e-Learning系统模型。该模型采用了贝叶斯网络的推理机制来建立学生学习模型,以处理学生学习过程中产生的众多不确定因素,解决学习内容的动态组织、学习策略和学习诊断。

语义检索是语义Web的重要应用领域之一,也是目前信息检索领域的热门研究课题之一。通过对语义标注进行匹配和推理,基于语义的检索系统能有效地提高检索的精确性和覆盖率。Tane等(2003)设计了一种从含有语义标注的学习材料中自动寻找和组织所需课件资源的工具。它借助本体来反映用户的需求,然后通过对需求的语义分析,在语义Web中抓取合适的资源,并组织成有意义的课件。李艳燕博士(2005)提出了一种查找事实或语义相关资源的语义搜索算法,利用学习对象的语义信息对用户请求进行自动语义标注,并确定用户的搜索目标,再采用语义关联路径的启发式规则来指导搜索,为用户提供丰富的语义互联资源。Vega-Gorgojo等(2005)对网格环境下学习服务的语义检索进行了研究。他们摒弃传统的服务发现机制(服务索引、UDDI),采用本体描述学习服务,实现了学习服务的语义检索。服务提供者可以使用本体词汇描述服务,教育工作者可以使用领域概念进行学习服务搜索。

智能答疑和学习评估方面,也有一些研究成果。刘艳等(2009)建立了基于知识关系的覆盖型贝叶斯网络学习评估模型,并实现了基于贝叶斯网络学习评估模型的e-Learning原型系统。该系统能够对学生的知识掌握水平进行评估,并根据学生的评估结果,为学生提供个性化的导学建议。姜少峰等(2003)研究了贝叶斯网络推理在远程答疑专家系统中的应用,重点分析了贝叶斯网络推理与决策的方法及知识表示与知识库设计技术。郑耿忠(2008)将基于范例的推理(CBR)引入到答疑系统的设计中,研究了基于CBR

的智能答疑系统范例库的构建方法,对BP神经网络和范例匹配算法在CBR范例库检索中的应用进行了分析。其研究表明,基于CBR的推理能有效地提高答疑系统的效率和准确性,进一步提高答疑系统的智能性。

总的来说,语义Web技术的发展逐步成熟,正在从实验室慢慢走向商用。本体和推理作为语义Web体系架构的核心技术,引起了e-Learning领域研究者的极大关注。国内外众多研究机构和学者开始借助本体和推理技术来解决当前e-Learning领域存在的资源重复建设、检索效率低、个性化支持不足等问题。除此之外,语义Web技术的应用还涉及e-Learning课程构建、(李好等,2010)e-Learning测试与评估、(郭成栋等,2006)学习组织记忆、(Abel et al., 2004)学习设计(Sicilia et al., 2011)等。语义Web技术在e-Learning领域的应用,尤其是在资源组织和共享方面的应用思路为学习元平台的语义技术架构提供了指导。

三、学习元平台中的语义技术架构

学习元平台整体语义架构(见图1)的核心,是应用了语义Web体系框架中的本体技术和推理技术。最底层是Ontology API层。LCS使用Java版的JENA本体操作API构建本体编辑与管理环境,进行本体相关操作,包括概念与属性的添加/编辑/删除/查询、本体模型创建/读取、本体导入/导出等。本体和语义词典共同组成语义层,是学习资源语义信息标注与提取的基础。LCS中的本体架构由知识本体、用户本体和情境本体组成。语义词典主要集成了哈工大扩展版的同义词词林(Che et al., 2010)和中文WordNet。语义层之上是标注层。语义标注信息的获取与存储是LCS开展更多上层智能应用(如资源语义聚合、资源语义检索、适应性资源推荐、社会认知网络构建等)的数据基础。LCS中的语义标注信息,一方面通过学习资源的半自动化语义标注程序产生;另一方面通过学习资源的语义基因提取程序产生。在语义标注层之上是推理层,通过编辑推理规则,借助JENA内置的推理机,可以完成各种推理应用,如资源关系推理、人际关系推理等。

1. 本体组织架构

本体的构建是实现泛在学习环境智能化、支持个性化学习、实现学习资源适应性推荐的数据基础。LCS包括三类本体,分别是知识本体(Knowledge Ontology)、用户本体(User Ontology)和情境本体(Context Ontology)。每类本体的构建遵循W3C的OWL规范,本体的查询使用W3C推荐的SPARQL语

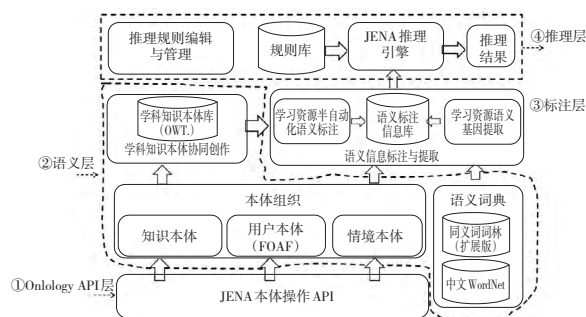


图1 学习元平台整体语义框架

言。其中,知识本体是指面向不同学科领域的教学内容本体;用户本体是对LCS中用户基本信息及其用户关系的规范性说明;情境本体是对学习者的学习上下文环境进行的规范性说明,包括时间、地点、学习设备等信息(见图2)。

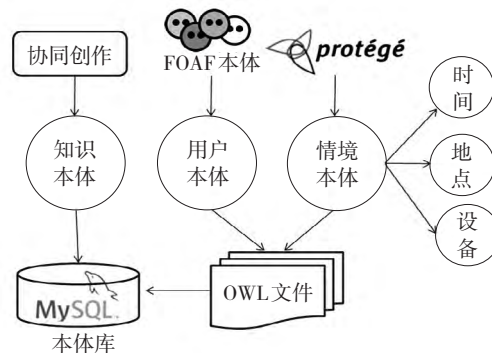


图2 本体组织架构

由于知识本体所属学科领域的多样性和复杂性,LCS采用协同创作的本体构建模式,鼓励普通用户和学科专家通过协作来动态构建学科知识本体。LCS中的用户本体直接引入了当前国际上比较流行的FOAF本体(<http://www.foaf-project.org/>)。情境本体采用Protégé工具进行构建,直接导出OWL文件,然后通过LCS的本体导入接口将OWL文件中的内容存储到MySQL的本体数据表中。

2. 语义词典集成

除了引入本体技术外,LCS还集成了哈工大扩展版的同义词词林和中文WordNet作为语义词典。同义词词林共包含77343条词语,通过为每条词语进行特定格式的编码来表征词语之间的同义和同类关系。WordNet(<http://wordnet.princeton.edu/>)是由Princeton大学的心理学家、语言学家和计算机工程师联合设计的一种基于认知语言学的英语词典。它不光把单词以字母顺序排列,而且按照单词的意义组成一个“单词的网络”,其框架的合理性已被词汇语义学界和计算词典学界所公认。(于江生等,2002)东南大学在英文WordNet词典的原理基础上实现了中文版的WordNet(以下简称C-WordNet),包含约118000个中文词和

115400个同义词集。C-WordNet 中词汇概念间的语义关系主要包括上下位、同义、反义、整体和部分、蕴含、属性、致使等不同的语义关系。(吴思颖等, 2010)

语义词典主要提供语义词汇及其语义关系服务, 可用于学习资源的语义标注、语义检索、动态关联与动态聚合, 也可用于相似用户的识别与聚类。为了将一些新的词汇及时添加进来, LCS 还提供了语义词典编辑功能。管理员可以按照同义词词林和中文 WordNet 的编码规则动态扩充语义词典。

3. 学科知识本体协同创作

本体协同开发机制, 可以改变采用传统本体构建方法构建本体时参与人数较少, 构建的本体较为片面、难以推广的缺陷。(张媛等, 2007) Tudorache 等 (2008) 分析了协同本体开发的具体需求, 并开发了 Collaborative Protégé 工具来支持多人协同本体开发。Sarraiya 等 (2010) 采用定性信息搜集的方法研究协同本体构建。张媛等 (2007) 提出了社会化本体协同开发模型, 并对社会化本体协同开发机制的原理、权限、体系模型、专家组兴趣型社群形成以及相关算法等进行了研究。

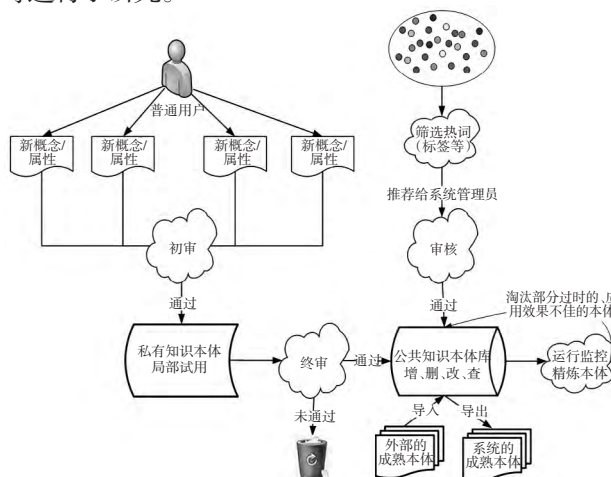


图3 学科知识本体的协同创作

随着语义 Web 的快速发展, 本体的重要性越来越突出, 需求量越来越大, 单靠个人或组织的力量无法实现本体的有效开发和演化发展, 本体的协同创作与进化已成为本体领域的重要研究内容。基于协同创作与进化的指导思想, LCS 基于 JENA 框架开发了在线知识本体的协同创作与管理环境, 通过开放本体的创建权限, 允许任何用户参与本体创作, 从下到上构建学科知识本体, 并在系统运行过程中, 实时监控、评价知识本体的应用效果。一方面, 采用本体精炼技术自动将部分过时的、不合格的本体淘汰掉; 另一方面, 不断将趋于稳定的、被普遍接受的知识本体纳入到系统本体中, 逐步形成各个学科领域以群体智慧为

基础的、可进化的知识本体。当前, LCS 中知识本体的产生途径主要有三种: 一是由普通用户创建, 二是由系统管理员将外部成熟的本体直接从后台导入, 三是通过系统自动挖掘 (如一些热门标签、应用较广的知识群内的私有本体等)。知识本体的类型包括两类, 分别是私有本体 (只能在局部范围内试用) 和公共本体 (可在系统全局中正式应用)。

4. 学习资源半自动化语义标注

通过语义标注可以将 Web 的状态从机器可读提高到机器可理解, 这是发展和实现语义 Web 的基础。语义标注是用本体来描述网页中的概念或概念实例, 具体实现是给网页上添加语义信息。(陈星光等, 2009) 传统的语义标注方法主要分为手动标注、半自动标注和自动标注三类。(张玉芳等, 2010) 手动标注方法大多集中在建立标注及分享标注工具上; 半自动语义标注是利用词汇语义分析对网页进行标注; 自动语义标注主要利用机器学习的方法, 从统计学的角度为网页建立标注。这三种标注方式各有优劣: 手动标注准确性高, 但耗时耗力; 自动标注省时省力, 但准确性不高; 半自动方式一定程度上可以节省人力, 但仍需要人工干预, 方可保证标注的质量。近些年, 语义标注的范围逐步扩展, 正逐步从文本标注扩展到图片标注、(Su et al., 2011) 视频标注 (Garca-Barriocanal et al., 2011) 等多媒体语义标注。

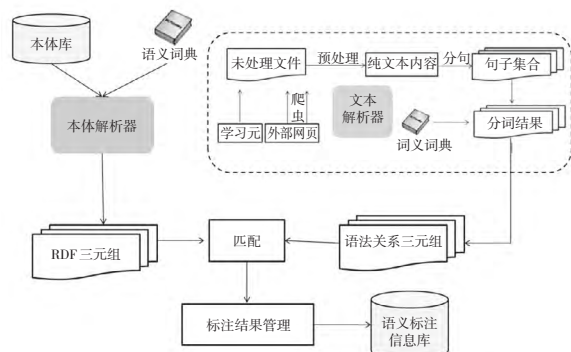


图4 基于领域本体的学习资源半自动化标注

基于领域本体的语义标注可以理解为将自然语言中的句子转换成本体中 RDF 陈述的过程, 也就是将人可理解的语言转换成计算机可理解的语言的过程。图4描述了 LCS 中学习资源半自动化语义标注的技术路线。本体解析器的目的是解析领域本体, 生成一系列的 RDF 三元组。这里可以借助惠普公司的 JENA API 进行领域本体的解析, 解析之后的结果是一系列的 RDF 三元组, 形如: (类, 属性, 实例)。文本分析器的目的是借助自然语言处理技术, 生成文档中句子的语法关系三元组, 形如: (主语, 谓语, 宾语)。得到 RDF 三元组和语法关系三元组之后, 将两

者进行匹配。对于匹配成功的语法关系三元组进行语义标注；匹配不成功的，可以根据不成功的情况进行本体的扩展或者不进行语义标注。

5. 学习资源语义基因提取

语义基因是指能够反映资源内容所要表达含义的基本信息单元，形式上表现为基于本体描述的带有权重的概念集合（包括核心概念以及概念间的关系）。语义基因不是简单的关键词集合，而是资源背后所隐藏的语义概念网络。形象地来说，语义基因就好比一棵大树的“树根”，控制着大树的性状和生长方向。

语义基因在形式上表现为基于本体描述的带有权重的概念集合（包括核心概念以及概念间的关系）。语义基因可以被形式化地表示为有序三元组，即 $SG = \langle CS, WS, RS \rangle$ 。其中CS是核心概念集合，WS是概念项的权重集合，RS为核心概念间的关系集。

提取学习资源语义基因的前提条件是领域本体库的建立。语义基因本质上是基于本体的资源内容特征项，即用标准化的本体数据来表征资源的核心内容。LCS中关于语义基因的设置主要有两种方式：一种是手动设置，即让资源的创建者手动添加语义基因，从领域本体库中选择能够准确表征资源内容的本体类，并赋予不同的权重；二是自动提取，即通过程序自动从资源的文本内容中提炼出核心的语义特征项（概念）及关系，并通过一定的规则为每个语义特征项赋予不同的权重。学习资源语义基因提取的总体技术框架如图5所示。

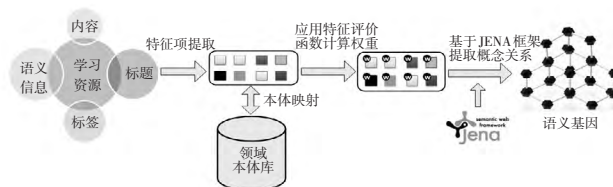


图5 学习资源语义基因提取总体技术框架

首先，将资源的标题、标签、内容和语义标注信息作为学习资源语义基因提取的四种数据来源，并根据各自的重要性赋予不同的权重；接下来，结合领域本体库，借助Web数据挖掘领域较为成熟的文本特征项提取技术，从资源中提取出一系列的特征词（核心概念），并将这些特征词映射到本体，存放到CS集合中；然后，通过预先设定好的特征评价函数为每个特征项赋予不同的权重值，将这些权重值放到WS集合中；最后，通过JENA框架将这些特征词在领域本体库中存在的语义关系以三元组的形式提取出来放到RS集合中。

6. 推理规则与推理机应用

JENA自身提供了基于规则的推理机（Rule-Based

Reasoner）。该推理机可以应用于RDFS和OWL本体中，支持基于RDF有向图的规则推理，提供前向、后向和混合推理模式。规则集合被绑定到数据模型中，形成新的推理数据模型，用于用户查询操作，返回的查询结果不仅包含原始数据集中的三元组，还包括使用推理规则得出的数据。

JENA的每条规则都采用产生式表示，“ $\rightarrow$ ”左侧部分表示推理的条件，右侧部分表示推理的结果，条件项和结果项都采用RDF三元组（Subject, Predicate, Object）的形式描述。一个简单规则如下：

String rule6="[(?x lc:isPartof ?y) $\rightarrow$ (?y lc:hasPart ?x)]"

该规则表达的含义是：如果X是Y的一部分，那么Y包含X。

LCS中基于JENA框架开发了在线的推理规则编辑与管理功能。系统管理员可以根据业务需求动态添加、编辑、删除推理规则，也可以通过“启用”和“禁止”功能控制推理规则的活跃状态。若规则被禁用，JENA推理机将不会加载该规则。



图6 推理机应用的基本思路

LCS中应用JENA推理机的基本思路如图6所示：按照JENA规定的规则语法要求编写推理规则；创建推理机实例；将推理规则（集）加载到推理机中；应用推理机对本体数据模型进行推理；生成推理后的新本体数据模型；编写SPARQL查询语言提取需要的三元组数据。LCS中的推理功能既可以用于资源关联关系的计算，也可以用于社会认知网络的构建，还可以用于学习路径和资源的推荐。LCS中开发了推理规则的编辑与管理环境，可以根据需要动态生成新的推理规则或淘汰陈旧的、失效的推理规则。

四、学习元平台中的语义技术应用

语义技术架构的最终目的是构建更加智能化的应用服务。LCS中依托上述语义技术的整体架构，分别在学习资源进化的智能控制、学习资源的动态语义关联、学习资源的动态语义聚合、学习资源的适应性推荐、学习资源的语义化检索以及社会认知网络的动态构建等方面进行了探索性研究。

1. 学习资源进化的智能控制

Web2.0时代人人可以生产、消费、传播资源。开放环境下用户群体的复杂性和生产的自由化直接导致数字资源的爆炸式增长和无序进化。学习资源进化过程的智能控制，对于促进开放环境下学习资源的有序建设和进化发展具有重要意义。LCS作为一种开放的

知识社区，允许任何用户开放创作学习资源，协同编辑资源内容。因此，LCS同样面临维基、百科等所无法避免的“信任”问题。为了解决此问题，LCS提出一种有效的资源进化智能控制方法。（杨现民等，2013b）

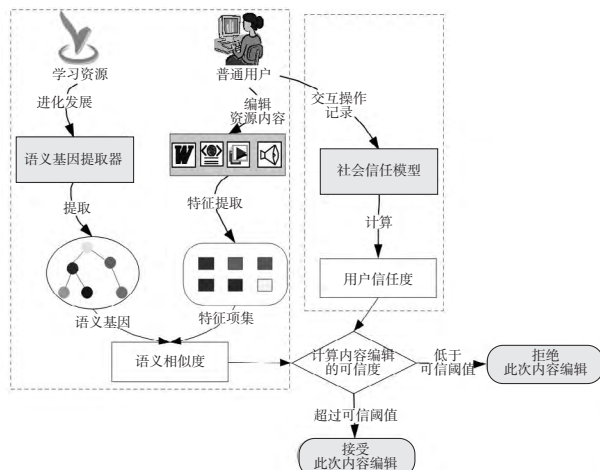


图7 学习资源进化中的智能控制

图7描述了基于语义基因和社会信任模型的内容进化智能控制技术的总体框架。其中，社会信任模型是参照现实社会中的信任关系构建的一套可计算的信任评估技术，可用于评价代理、用户、资源等任何参与网络交互实体的信任度。上述技术框架的核心是综合两方面的信息进行内容编辑的可信度计算：一方面，应用新添加内容的特征信息和当前资源的语义基因进行语义相似度计算；另一方面，基于用户的交互操作数据，应用社会信任模型计算用户的信任度。设定内容编辑可以被接受的可信度阈值，如果此次内容编辑的可信度超过阈值，则自动接受此次内容编辑结果；否则，自动拒绝。资源进化智能控制功能模块已经在LCS中运行近一年，运行效果良好，大大减轻了资源管理者的内容审核负担，促进了LCS中学习资源的快速、有序进化。

2. 学习资源的动态语义关联

当前的e-Learning资源普遍缺乏内在知识逻辑的关联性，因为资源之间的联系是通过一般的超链接形成的外在关联，基于HTML的数据组织或资源库中的分类不能体现数据内在的语义联系。为了实现学习资源

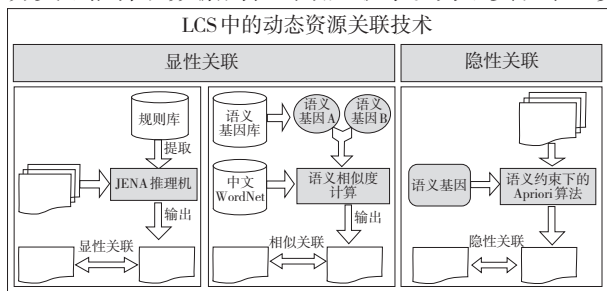


图8 学习资源的动态语义关联

源间的动态语义关联，LCS构建了如图8所示的技术框架。

学习资源间的关联主要包括两种类型，一种是显性关联，另一种是隐性关联。（杨现民等，2013d）显性关联是从语义出发，基于系统已有的关系类型建立的资源关联，易被用户观察和识别；隐性关联是从语义上难以通过人工发现，但可以通过数据挖掘技术识别出来的潜在资源关联。显性关联的建立采用了基于规则的推理技术和基于语义基因的相似关系计算技术，隐性关联的建立主要采用了基于语义约束的关联规则挖掘技术。为了清晰展现学习资源个体间的语义关系，LCS开发了如图9所示的可视化资源关系网。

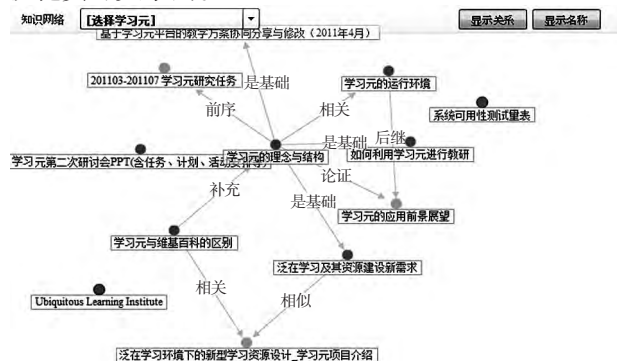


图9 LCS中资源语义关联的可视化展现

3. 学习资源的动态语义聚合

学习资源的动态语义聚合不是简单地将多个学习资源组成一个资源集，而是通过技术手段实现具有内在逻辑关系的学习资源之间有意义的结合。资源聚合的价值和意义集中体现在两个方面：一是可以实现多个小粒度的具有内在逻辑关系的资源单元集中呈现，通过资源的内聚，减轻学习者“机械性”检索资源的负担；二是可以将碎片化、零散性的知识组织成更加完整的知识单元，有助于学习者系统、全面地进行知识建构。LCS中根据资源聚合结构的不同，将学习元的聚合划分为主题资源圈和有序知识链两种形态，每种形态具有各自的结构特点和适用情境。LCS中学习

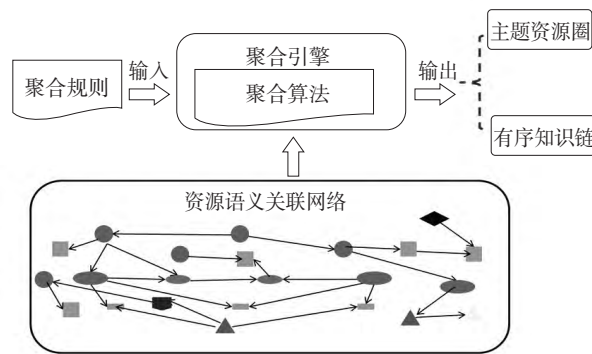


图10 学习资源的动态语义聚合

资源的动态语义聚合技术框架如图 10 所示。

LCS将具有相同主题、语义上高度相关的多个资源聚合成主题资源圈，按照平行结构进行集中呈现。这种资源聚合形态的特点是资源间不存在上下位的层次关系，同属于某一主题，采用平行的列表方式进行呈现。主题资源圈聚合的基本实现思路是：采用BFS（Breadth First Search）在有向资源关系图中寻找具有相似关系的资源节点，依据相似关系衰减函数计算两两节点之间的关联程度，将满足最低阈值要求、高度相似内聚的节点自动聚合在一起，最终生成若干个主题资源圈。图 11 显示了主题资源圈的聚合效果。

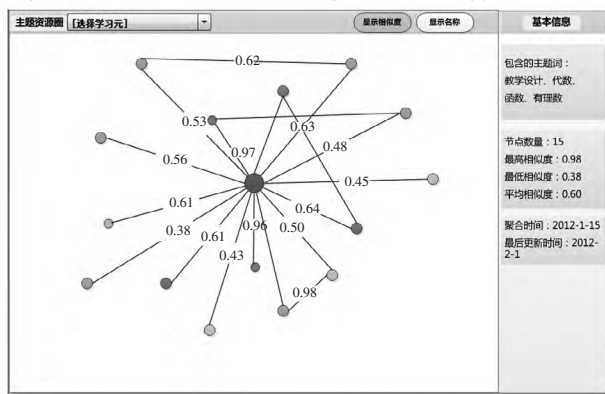


图 11 LCS中主题资源圈的聚合效果

LCS将多个具有前后序关系的资源聚合成有序知识链，链条上的知识点具有显性的前序、后继关系。也就是说，按照正常的学习流程，需要先学完前一个知识点，方可进入到下一个知识点的学习。有序知识链实际上起到一个知识导航的作用，减少网络学习中的“迷航”。有序知识链聚合的基本实现思路是：从整个资源语义关系图中提取出表示有序关系（前驱、后继、是基础）的所有资源节点和边，组成知识序列有向图 G；深度优先遍历（Depth First Search, DFS）G，将所有知识路径找出来，生成若干条有序知识链。图 12 显示了有序知识链的聚合效果。

4. 学习资源的语义化检索

学习元的一个重要特色是，它是经过语义标注的

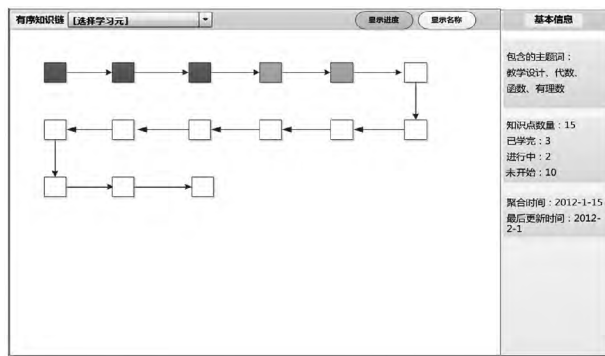


图 12 LCS中有序知识链的聚合效果

学习资源，其资源文件紧紧围绕内部的领域知识本体展开。因此，LCS可以提供比基于简单文本匹配的检索更为强大的语义检索功能。完整的学习元语义检索的流程（程昱，2009）如图 13 所示。

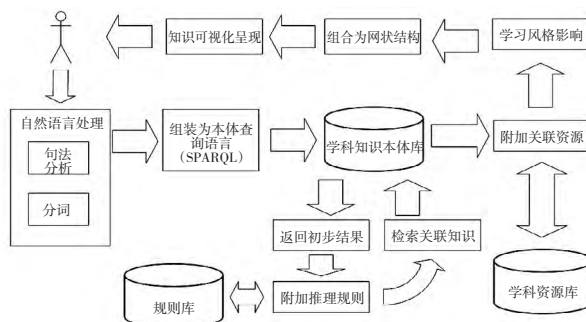


图 13 学习资源的语义化检索流程

目前LCS提供了基于关键词的初级检索和基于复合知识属性的高级检索两种检索功能。基于关键词的语义检索相对于文本匹配的关键词检索，优势主要体现在查全率方面。检索结果不仅包括标题、标签等文本中包含检索词的学习资源，而且包含那些主题词与检索词在语义上等价或近似的学习资源。例如用户检索“数字化学习”，检索结果不仅会包括含有“数字化学习”主题词的内容，而且会返回包含“e-Learning”、“数位学习”的内容，并提示一些相关度较高的主题词，如“移动学习”、“混合式学习”、“远程学习”、“网络教学”等，并依据其语义关系形成一个动态的网状图，供学习者进一步了解和检索。

两个知识点是否具有语义上等价的关系，是通过学习元本体模型中的内置属性“别名”和内置关系“等价”来判断的。当用户声明两个知识点是“等价关系”，或一个知识点的名称与另一个知识点的“别名”相同，我们就认为他们具备语义上的等价关系。当一个知识点作为搜索的候选结果呈现时，与其具备语义上等价关系的知识点也会并入到搜索的候选结果列表。

这些语义关系的构建，一方面通过用户手动填写知识点的属性和语义关系来完成。例如为创建的知识点的“别名”属性添上关键词“嬴政”；当用户检索“嬴政”的时候，“秦始皇”也能作为匹配的结果呈现。另一方面，通过基于本体的推理规则，能发现一些用户没有直接声明的语义关系，例如：知识点 x 与 y 等价，y 与 z 等价，知识点 x 有别名 a、b，而等价属性具有传递性质，则借助推理规则可推出知识点 y、z 也有别名 a、b。上述规则用简化的规则描述语言表示如下：

规则1: 等价 $(?x, ?y) \wedge$ 等价 $(?y, ?z) \rightarrow$ 等价 $(?x, ?z)$

规则2: 别名 $(?x, ?a) \wedge$ 等价 $(?x, ?y) \rightarrow$ 别名 $(?y, ?a)$

类似的, 可以通过设定推理规则, 结合用户自定义的知识类型和属性, 实现更复杂的推理逻辑, 例如亲属关系的推理:

父亲 $(?x, ?y) \wedge$ 哥哥 $(?x, ?z) \rightarrow$ 叔叔 $(?z, ?y)$

LCS通过学习元的知识本体模型, 为学习资源提供了一种可由用户定制的动态元数据描述, 使得用户对不同知识类型, 能够输入不同的检索条件, 过滤一些在文本上匹配但是语义上不符合查询者要求的结果, 实现较为精准的查询。例如对“历史人物”类型知识的检索, 其检索条件如下图所示:

属性名	查询内容	组合条件
出生日期		请选择
去世日期		请选择
朝代		请选择
父亲		请选择
母亲		请选择
性别		请选择
前序知识点		请选择
后继知识点		请选择

图14 复合条件的语义检索

当检索别的知识类型如“历史事件”时, 检索条件的属性项又会发生变化, 动态的调整为“历史事件”这个知识类型包含的属性。这种基于知识本体的动态元数据检索, 比LOM、Dublin Core等静态元数据标准更具灵活性和扩展性, 能根据不同的应用领域、用户的具体需要而随时调整, 有助于学习者更准确地找到他们所需要的学习资源。此外, 借助用户模型中的学习风格(包括设备偏好、感觉通道偏好、活动类型偏好), 当用户开启基于用户模型调整的功能时, 可以实现对检索结果的再次排序, 将用户偏爱的资源类型优先呈现给用户。

5. 学习资源的适应性推荐

学习元是一种特殊的结构化资源, 它不仅包括学习内容, 还包括学习活动、人际资源等。故向用户推荐个性化的学习元也就是将个性化的学习内容、学习活动和人际资源同时推荐给用户, 方便用户在学习所需知识的同时参与相应的活动, 既加深对知识的了解, 又通过与相关用户的交互消除学习时的孤独感, 吸取他人经验, 促进社会性学习。此外, 泛在学习环境下, 用户可以使用各种学习设备进行无所不在的学习。(杨现民等, 2013c) 由于设备物理特征的差异, 资源推荐时还需要考虑同一份资源在不同设备上的自适应呈现问题。LCS中学习资源适应性推荐的技术框架(陈敏等, 2011) 如图15所示。

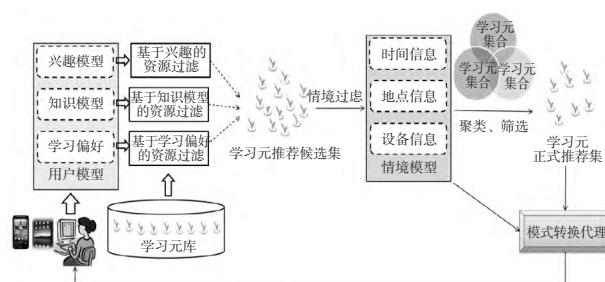


图15 学习资源的适应性推荐

LCS中学习资源适应性推荐的整体思路是: 首先, 从用户兴趣、已有知识和学习偏好三个角度出发, 从学习元库中获得对应的推荐结果作为推荐资源候选集; 然后, 结合情境模型中的时间、地点、设备等信息, 过滤一部分不合适的资源集, 比如iPad无法播放Flash课件, 则自动将包含Flash的学习元过滤掉; 接着, 利用综合聚类算法, 将推荐资源候选集进行归类、筛选, 得到最符合用户需求的推荐结果列表; 最后, 根据情境模型中的设备特征信息, 通过资源格式转换代理, 在用户使用的各种终端设备上适应性呈现资源。

用户模型和情境模型是上述技术框架的核心部分。其中, 用户模型由兴趣模型、知识模型和学习偏好模式组成。基于兴趣的资源过滤, 主要通过计算用户之间的兴趣相似度, 将兴趣高度相似用户群体间收藏、订阅和频繁访问的资源作为推荐对象。基于知识模型的资源过滤, 主要通过用户对用户的学习行为和结果进行评价分析, 结合学科知识本体, 识别用户的知识结构缺陷, 进而推荐那些能够促进用户知识结构完善的资源。基于学习偏好的资源过滤, 主要结合用户的学习风格自动筛选符合用户偏好的学习资源。情境模型由时间、地点和设备三类信息组成, 时间信息通过读取服务器的系统时间获得, 地理位置信息通过读取GPS数据获得, 设备信息通过调用系统相关API直接获取。

6. 社会认知网络的动态构建

开放环境下的学习资源除了可以作为独立完整的学习单元存在外, 还可以作为学习者认知网络联通的中介点。也就是说, 学习相同或相似主题学习内容的学习者还可以透过学习资源实现社会认知网络的构建, 这与联通主义学习观所倡导的“联结和再造”价值取向是一致的。随着学习者之间的不断交互, 会逐渐形成一个具有相同学习兴趣和爱好、交往频繁的认知网络。与Social Network界定的一般的交际网络不同, 社会认知网络是由知识和人共同构成的网络, 是在人与知识的深度互动过程中构建起来的。学习者通过社会认知网络不仅能获取所需要的物化资源, 还能

找到相应的人力资源，如通过某一个学习内容，可以快速定位到这个内容领域最权威的专家或适合的学习伙伴等。（余胜泉等，2011）图 16 描述了 LCS 中社会认知网络动态构建的技术实现框架。

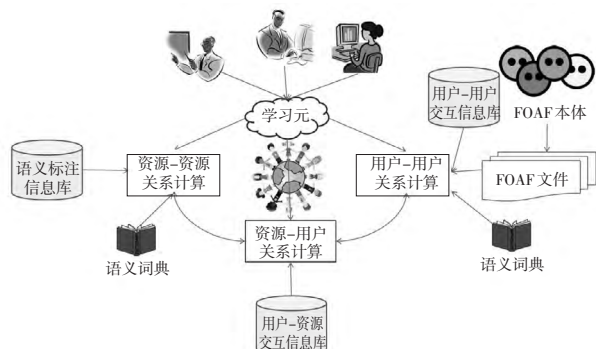


图16 社会认知网络的动态构建

LCS中存在用户与资源的交互(浏览、编辑、创建、评论、批注等)、用户与用户的交互(添加好友、邀请协作、发短消息等),这些交互信息将分别存入用户-资源交互信息库和用户-用户交互信息库。每个资源实体通过半自动化的语义标注程序将获得语义标注信息,这些信息将存入语义标注信息库。每个用户实体的信息都基于FOAF本体进行规范化描述,并自动生成对应的FOAF文件。

LCS中的社会认知网络主要通过三种关系的计算而生成,分别是资源与资源的关系(如相似、相关、前序、等价等)、用户与资源的关系(如参与、贡献等)以及用户之间的人际关系(如协作、好友、竞争、师徒等)。资源之间的语义关系利用语义标注信息库和语义词典进行推理、计算和动态更新;用户与资源之间的关系通过用户-资源交互信息库中的数据进行挖掘,动态建立和更新用户与资源之间的关系。用户之间人际关系的计算一方面利用用户-用户交互信息库中的数据,通过设定相应的规则,比如用户A如果频繁访问用户B的个人空间,则可认为用户A是

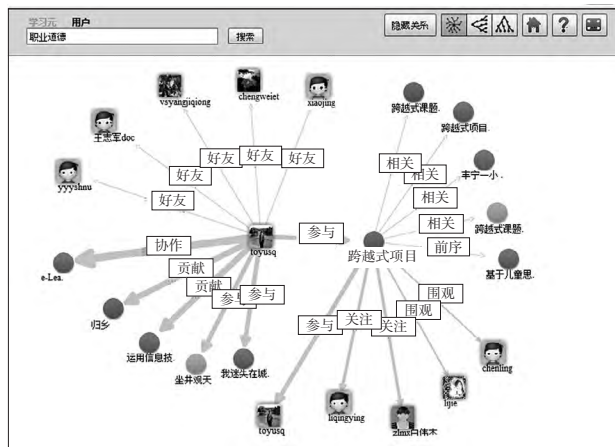


图17 LCS中的社会认知网络

用户B的关注者；另一方面，对用户FOAF文件中的相关字段进行计算处理，比如比较 Organization 字段，若相同，则视为同事；或者借助语义词典对用户FOAF中的Interest字段进行相似度计算，判断二者是否兴趣相投。图 17 显示了 LCS 中的社会认知网络。

五、总结与启示

LCS中的语义技术框架已经初步形成,一些应用也已在语义技术架构之上进行探索,并取得一定的成效。随着语义Web技术的不断发展,越来越多的e-Learning支撑系统将采用语义技术架构,实现从资源的语义化组织到具有更强智能性和个性化的高级应用服务(如适应性资源推送、个性化学习、智能导航等)的开发。语义技术将使软件系统具备更好地理解用户学习需求的能力,有助于大大提升数字化环境下的学习体验。笔者认为,将语义技术在e-Learning系统中推广应用需要注意以下几点:

第一, 积极引入外部开放、较成熟的本体, 比如 FOAF 本体、Freebase 中的开放本体、学科领域本体 (如 Gene 本体) 等。本体构建是一个复杂的过程, 耗时耗力。如果外部已有相关的成熟本体, 可以直接导入到系统本体库, 而无需重复开发, 突出本体的共享价值。此外, 还要重视教育领域学科知识本体的建设, 为面向学科的智能应用开发提供本体数据支持。

第二, 纯手动或完全自动化的语义标注不具有很强的应用实践价值。在e-Learning系统开发或升级改造时, 比较理想的方法是采用人工标注与自动标注相结合的方式, 实现对数字学习资源较为可靠和快捷的语义标注。

第三, JENA内置的推理引擎,在大数据量处理时,难以保障推理效率。因此,建议集成具有较强针对性和运行效率/效果好的推理机,如 Pellet、Sesame、Jess等。

第四, 语义技术的应用方面需要拓展思路, 除了在资源的智能控制、资源语义关联与聚合、语义检索、适应性推荐、认知网络构建等方面“用武之地”外, 还可以结合现实需求引入更多语义技术的高级应用服务, 如资源社区潜在专家的挖掘、个人知识地图与学习路径的自动生成、学习资源的动态聚类等等。

第五,除了重视资源的语义标注外,还要特别关注用户信息的语义化表征。Web2.0时代的到来,越发强化人际关系在促进社会化学习中的重要作用。从支持学习的角度出发,如何动态构建人与人之间的社会认知网络,是未来e-Learning系统发展需要重点关注的问题之一。

第六,重视情境模型的设计与情境本体的构建。泛在学习是一种高度情境化的学习方式,e-Learning系统通过提供情境感知功能,将获取更多有价值、有助于促进适应性学习的情境信息。同时,在情境模型和情境本体的支持下,e-Learning系统真正升级为适应性学习系统。

参考文献:

- [1]曹乐静,刘晓强(2005).基于本体和Web服务的适应性e-Learning系统[J].计算机系统应用,(4):16-23.
- [2]朝乐门,张勇,邢春晓(2011).DBpedia及其典型应用[J].现代图书情报技术,(3):80-87.
- [3]陈和平,郭晶晶,吴怀宇等(2007).基于Ontology和Jena的个性化e-Learning系统研究[J].武汉理工大学学报(交通科学与工程版),(6):1049-1052.
- [4]陈敏,余胜泉,杨现民等(2011).泛在学习的内容个性化推荐模型设计——以“学习元”平台为例[J].现代教育技术,(6):13-18.
- [5]陈星光,张文通,汪霞(2009).基于领域本体的自动化语义标注方法的研究[J].科学技术与工程,(8):2215-2218.
- [6]程罡(2009).泛在学习环境下的学习资源共享模型——学习元的体系结构和运行环境研究[D].北京:北京师范大学.
- [7]郭成栋,杨贯中,唐金鹏等(2006).e-Learning中基于对象本体的测试与评估[J].计算机工程,(24):72-74.
- [8]黄海江,杨贯中(2007).基于本体的学习内容个性化推荐[J].科学技术与工程,(14):3394-3398.
- [9]姜少峰,朱群雄(2003).Bayesian推理在远程答疑专家系统中的应用[J].北京化工大学学报(自然科学版),(6):95-98.
- [10]孔德华,王锁柱(2006).基于XML的自适应e-Learning系统模型的研究[J].山西师范大学学报(自然科学版),(2):20-23.
- [11]李好,杨贯中(2010).基于本体的e-Learning课程构建[J].计算机工程与设计,(4):881-884.
- [12]李艳燕(2005).基于语义的学习资源管理及利用[D].北京:中国科学院研究生院(计算技术研究所).
- [13]刘艳,张锐(2009).基于贝叶斯网络的学习评估模型及其在e-Learning系统的应用[J].滁州学院学报,(4):49-55.
- [14]刘革平,赵端花(2009).基于形式化本体的数字化学习资源共享技术研究[J].西南师范大学学报(自然科学版),(6):204-207.
- [15]刘卫红,吴江(2006).本体在e-Learning系统中的应用研究[J].计算机应用研究,(4):63-67.
- [16]邵国平,余盛爱,郭莉(2008).语义Web对e-Learning中资源管理的促进[J].江苏广播电视大学学报,(5):23-26.
- [17]吴思颖,吴扬扬(2010).基于中文WordNet的中英文词语相似度计算[J].郑州大学学报(理学版),(2):66-69.
- [18]杨现民,程罡,余胜泉(2013a).学习元平台的设计开发及其应用场景分析[J].电化教育研究,(3):55-61.
- [19]杨现民,余胜泉(2013b).开放环境下学习资源内容进化的智能控制研究[J].电化教育研究,(9):83-88.
- [20]杨现民,余胜泉(2013c).生态学视角下的泛在学习环境设计[J].教育研究,(3):103-110.
- [21]杨现民,余胜泉,张芳(2013d).学习资源动态语义关联的设计与实现[J].中国电化教育,(1):70-75.
- [22]于江生,俞士汶(2002).中文概念词典的结构[J].中文信息学报,(4):12-19.
- [23]余胜泉,陈敏(2011).泛在学习资源建设的特征与趋势——以学习元资源模型为例[J].现代远程教育研究,(6):14-22.
- [24]余胜泉,杨现民,程罡(2009).泛在学习环境中的学习资源设计与共享——“学习元”的理念与结构[J].开放教育研究,(1):47-53.
- [25]张龙昌,刘志哈,王攀等(2010).基于FOAF的分布式移动SNS应用[J].电信科学,(5):88-92.
- [26]张玉芳,艾东梅,黄涛等(2010).结合编辑距离和Google距离的语义标注方法[J].计算机应用研究,(2):555-562.
- [27]张媛,孙新(2007).SOBM中社会化本体协同开发机制研究[A].第4届信息系统及其应用学术会议(WISA)论文集[C].北京.
- [28]张云璐,刘咏宁,谢铭等(2011).基于FOAF演化博弈的网络资源可信度判别[J].计算机研究与发展,(S3):334-342.
- [29]郑耿忠(2008).基于范例推理的智能答疑的研究与实现[J].微计算机信息,(12):273-275.
- [30]Abel, M.-H., Benayache, A. & Lenne, D. et al.(2004). Ontology-Based Organizational Memory for e-Learning[J]. Educational Technology & Society, 7 (4):98-111.
- [31]Bouzeghoub, A., Defude, B. & Duitama, J. F. et al. (2006). A Knowledge-Based Approach to Describe an Adapt Learning Objects[J]. International Journal on E-Learning, 5(1):95-102.
- [32]Che, W., Li, Z. & Liu, T.(2010)LTP: A Chinese Language Technology Platform[A]. Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations[C]. Beijing:13-16.
- [33]Diaconescu, I.-M., Lukichev, S. & Giurca, A. (2008). Semantic Web and Rule Reasoning inside of E-Learning Systems[DB/OL].http://link.springer.com/chapter/10.1007/978-3-540-74930-1_26#page-1.
- [34]Dos Reis Mota, J. & Fernandes, M. A.(2010). Adaptivity and Interoperability in e-Learning Using Ontologies[DB/OL].http://link.springer.com/chapter/10.1007/978-3-642-16952-6_60#page-1.
- [35]Garca-Barriocanal,E., Sicilia,M.-A.& Snchez-Alonso,S. et al.(2011). Semantic Annotation of Video Fragments as Learning Objects: A Case Study with Youtube Videos and the Gene Ontology[J].Interactive Learning Environments,19(1):25-44.
- [36]Jovanović, J., Gašević, D. & Knight, C., et al.(2007). Ontologies for Effective Use of Context in e-Learning Settings[J]. Educational Technology & Society, 10 (3):47-59.
- [37]Nilsson, M., Palmér, M. & Naeve, A. (2002). Semantic Web Metadata for e-Learning—Some Architectural Guidelines

[A].The Proceedings of 11th W3C 2002 Conference[C].

[38]Sarraiya, J., Jardim-Gonçalves, R. & Gaspar, T. et al. (2010).Collaborative Ontology Building Using Qualitative Information Collection Methods[A].The 5th IEEE Conference. of Intelligent Systems[C]:61-66.

[39]Sicilia,M.-Á.,Lytras, M. D. & Sánchez-Alonso, S. et al. (2011).Modeling Instructional-Design Theories with Ontologies: Using Methods to Check Generate and Search Learning Designs [J].Computers in Human Behavior, 27(4):1389-1398.

[40]Su, J.-H., Chou, C.-L. & Lin, C.-Y. et al.(2011). Effective Semantic Annotation by Image-to-Concept Distribution Model[J]. IEEE Transactions on Multimedia, 13(3): 530-538.

[41]Tane, J., Schmitz, C. & Stumme, G. et al. (2003).The Courseware Watchdog: An Ontology-Based Tool for Finding and Organizing Learning Material[A]. Klaus, D. & Lutz W. Mobiles Lernen und Forschen-Beitrag der Fachtagung an der Universität[M].

Kassel University Press:93-104.

[42]Tudorache, T., Noy, N. F. & Tu, S. et al.(2008). Supporting Collaborative Ontology Development in Protégé [A]. Proceedings of the 7th International Conference on The Semantic Web[C]:17-32.

[43]Vega-Gorgojo,G., Bote-Lorenzo,M. L. &Gomez-Sanchez,E. et al.(2005). Semantic search of Learning Services in aGrid-Based Collaborative System[A].Fifth IEEE International Symposium on Cluster Computing and the Grid (CCGrid'05)[C]: 19-26.

[44]Zeng, Q., Zhao, Z.& Liang, Y. (2009). Course Ontology-Based User's Knowledge Requirement Acquisition From Behaviors Within E-Learning Systems[J]. Computers & Education, 53(3): 809-818.

收稿日期 2013-10-18

责任编辑 汪 燕

The Architecture and Applications of Semantic Technologies in Learning Cell System

Yang Xianmin, Yu Shengquan

Abstract: The Semantic Web places emphasis on the attached semantic information to Internet resources that can be read and processed by machine, so as to improve coordination between human and computer, and make it possible for resources to be reused and automatically processed in a large scale. With the advance of semantic web technology, it is introduced into various e-Learning platforms such as adaptive learning system, intelligent tutoring system, and semantic knowledge communities, in order to provide more personalized and intelligent learning services. Based on the semantically organized resources, by applying the ontology and reasoning technology of the semantic web system, the Learning Cell System (LCS) builds a semantic framework that contains the ontology API layer, the semantic layer, the annotation layer and the reasoning layer. Through this framework, LCS performs exploratory research in the fields of resource evolution control, dynamic semantic association and aggregation, adaptive recommendation of resources, semantic search and dynamic construction of social cognition network. It can also realize the semantic organization of resources and highly intelligent advanced application services. These practices suggest that to spread the application of semantic technology in e-learning systems, open and mature ontology and more advanced application services need to be introduced. And we should also integrate more targeted and efficient inference engines into the system, pay attention to semantic characterization of resources and users as well as the design of context models and construction of context ontology.

Keywords: Learning Cell; Learning Cell System; E-Learning Platform; Semantic Web; Semantic Technology Framework