

基于主题学习的伙伴推荐算法^①

张 杰¹, 林木辉²

¹(福建师范大学 协和学院, 福州 350007)

²(福建师范大学 教育学院, 福州 350007)

摘要: 学习伙伴是开放虚拟学习社区的重要资源, 学习伙伴中的助学者可以帮助普通学习者克服学习障碍, 相互提高沟通交流能力. 在构建出基于本体的知识库后, 综合学习者的兴趣、认知和热心度特征, 提出了一种基于主题学习的伙伴推荐算法. 实验结果表明, 学习者在提供学习经历后, 算法可以计算出它与其他学习者之间的伙伴评分, 评分差异较好反映了真实学习环境中的最佳学习伙伴经验, 有效地提高了开放虚拟学习社区构建的个性化和智能化, 从而提升开放学习社区中学习者的学习效率和效果.

关键词: 主题学习; 虚拟学习社区; 学习伙伴; 助学者

Learning Partners Recommendation Algorithm Based on Thematic Study

ZHANG Jie¹, LIN Mu-Hui²

¹(Concord College, Fujian Normal University, Fuzhou 350007, China)

²(Education College, Fujian Normal University, Fuzhou 350007, China)

Abstract: Learning partners are important resources in opening virtual studying community. The tutors who come from the learning partners can help the general learner out of learning disabilities, and improve their ability of communication each other. After constructing a Knowledge Database based on ontology, we record the learner characters about interest, grade of cognize, degree of enthusiasm. According to these factors, we bring forward a learning partner recommendation algorithm using Thematic Study. The result of experiment indicates that the algorithm can make different partner grade on the foundation of the learner's study experience. These grade can reflect the reality situation. It is significant for opening virtual studying community to improve its personalization and intellectualization, consequently to advance study efficiency.

Key words: thematic study; virtual studying community; learning partner; tutor

当前, 科学技术发展日新月异, 不断学习和更新知识俨然已是个人的立世之本、生存之道. 随着计算机和网络技术发展, 虚拟学习社区是人们进行终身学习的一种重要的方式, 它不受时空限制的开放学习方式受到广大学习者的欢迎. 在传统的学校教育中, 学习活动往往以班级的方式开展, 学习者不但有经验丰富老师进行现场辅导, 而且集体学习使得学习者容易得到他人的认同, 同学之间交流合作也有效地促进学习者水平的提高. 在开放的虚拟学习社区中, 海量的学习资源往往让学习者感到困惑和迷茫, 造成迷航现象; 同时,

由于教师、学习者和学习集体在学习期间处于分离状态, 当在学习过程中遇到障碍难以解决时, 学习者常感孤独和无助, 导致厌学, 学习热度和效率也会持续降低. 在这种状况下, 学习者往往渴望有“专家”、“伴”或“榜样”能够与之进行互动交流, 以期能顺利完成学习目标.

显然, 通过在虚拟学习平台中突出对学习者的学习过程的引导, 推荐最佳学习伙伴, 并能自动地将相近者组织成一个具有共同学习兴趣和目标的共同体, 可以增强虚拟学习社区中各组成要素的耦合

^①基金项目:福建省教育厅 A 类(JA12086);福建省科技厅自然科学基金(2011J01343)

收稿时间: 2013-12-08;收到修改稿时间: 2014-01-02

与连接,提高学习者的学习效果和粘度。

1 国内外研究现状

目前,国内外对虚拟学习社区的探索大多局限于概念、学习策略和学习模式等理论方面,而对于实现层面的关键构建技术研究相对较少。例如 Hope N. Tillman 提出了虚拟学习社区的定义和特征等问题^[1];华南师范大学的赵建华教授和李克东教授阐述了基于协作方式的教学设计^[2];北京师范大学黄怀荣教授探讨了 CSCL 的理论和方法^[3];首都师范大学的王陆教授对虚拟学习社区中的社会网络结构与网络教育效果之间的因果关系做了深入分析^[4],这些分析主要是在离线状态进行,学习伙伴推荐需要动态实时计算。虚拟学习社区中的协同就是实现学习者之间的交互,而从本质上讲,交互是由资源构成的,资源是构成交互的资源,离开交互的资源是毫无意义的^[5]。在一个开放的虚拟学习社区中,资源可以分为两类:人力资源和非人力资源。非人力资源主要包括与学习内容相关的教学媒体,这个方面,国内已有一些研究:吉林大学的刘志勇提出了一种基于语义网的个性化学习资源推送算法^[6];华中师范大学的吴彦文利用混合算法来自动推荐学习资源^[7]。这些算法推荐的都是非人力资源的教学媒体。对于另一类重要资源:人力资源即学习伙伴的推荐研究却相对较少。通过与有相似学习背景的助学者沟通和交流,不仅让学习者在专业学习上能够获得帮助与指导,增强信心,缓解孤独和无助;而且在与人交往过程中,也培养和锻炼了交流与合作能力。可见,助学者的作用是不言而喻的。本文在构建一个基于本体知识库和特征向量模型(VSM)的基础上,描述出学习者的个性化特征模型,设计了一种有效的学习伙伴推荐算法,提高开放型虚拟学习社区的智能化和人性化,促进学习效果的提升。

2 相关概念阐述

丁兴富指出:学习者、助学者和学习资源是远程教育的教学三要素,助学者不必一定是学校老师^[8]。虚拟学习社区中如何加强和突出这三者的自动连结,是体现智能化和人性化的重要标志。

1)相似学习伙伴

相似学习伙伴是指一群在开放虚拟学习社区中的用户,他们内部之间有着相似的学习行为、主题或兴趣点,而在领域认知程度上可能存在差异。相似学习伙伴

表明学习者之间具有相近的学习背景和兴趣特征,实质上是社区学习网络中对用户类别的划分。

2)助学者概念

助学者是开放虚拟学习社区中一种特殊的角色,该角色在某些学习主题中具备较强的认知能力和专家特征,他们既是学习者,也是辅助者,是社区中分享经验,助人为乐的精神领袖。

3 学习者特征表示

学习者模型是记录学习者情况的一种数据结构,是对学习者信息的抽象描述和表示。它根据学习者学习过程中的学习行为和练习测试结果,分析出学习者的知识水平、认知能力、学习动机、学习风格和学习历史变迁等信息的特征^[9,10]。由于本研究的重点是为平台中的学习者推荐最佳学习伙伴,学习者模型中应当突出描述的是学习者的认知状态、热心度及个性化兴趣偏好信息。

学习行为的发生离不开学习媒体,以学习者之间交互的媒体为载体,将学习行为中发生的媒体交互映射到语义层中的本体概念,学习者之间的相似性问题就转换为计算概念之间的相似性是问题。运用本体来构建教学媒体组织和管理中的语义层,因而学习者个性化特征表示也采用本体方式。

3.1 基于本体的知识库构建

知识库作为学习媒体的组织和管理核心,其构建可以分为两个部分来进行:知识空间和信息空间^[11,12]。知识空间是语义层,信息空间中学习媒体通过映射到知识空间获得语义,知识空间利用本体来建模。起源于哲学的本体是对领域知识的概念化、形式化、明确化的描述^[13]。概念是对领域中可以标识的事物描述的集合,本体就是用特定领域的概念,以及这些概念之间的关系,来描述概念之间的语义。概念之间的关系通常是一种多对多关系,把概念当着结点,关系看成边,组成本体图。在具体应用中,可以结合实际需要,进行适当的约束和简化。在此,本体中的概念与学习主题对应,将领域知识按照一定粒度抽象成概念(即学习主题),其关系主要是派生和包含关系,而学习媒体则可隶属于某些概念,组成学习媒体组织结构图如图1所示:

其中 C_i 表示本体中的概念, LR_{kj} 代表与某些概念相关的学习媒体。

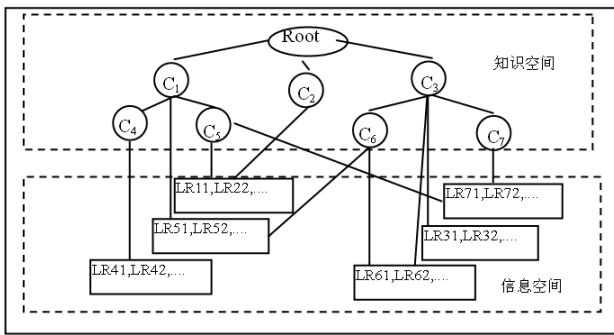


图 1 知识库中的学习媒体组织结构图

3.2 基于本体的 VSM 表示学习者个性化特征

向量空间模型(VSM)在文本分类中具有非常广泛的应用,其基本思路是利用一个关键词组成的向量近似描述文档的内容,由于关键词具有随机性,造成向量的空间维度一般较高,且关键词之间被认为是线性无关的。

在虚拟学习社区中,由于学习者的兴趣和行为大多围绕学习媒体进行的,因此根据其交互的媒体来分析和理解学习者的兴趣和认知能力显然是一种有效的策略。学习者对某一个学习媒体感兴趣,其隐藏的行为是对某一个知识点感兴趣,利用知识库中本体概念对学习媒体的标注作用,可以将学习者的特征描述从信息空间层迁移到知识空间中的语义层,即领域本体层。本体中的概念存在关系且数目较少,可以很好地克服 VSM 模型中向量维度较高,且线性无关的缺点。

在一个开放的虚拟学习社区中,学习者往往要经过注册才能成为正式成员,其主要学习行为有浏览资源,下载资源,上传资源,提问,评论,答疑,练习测试等。通过对这些学习行为分析和测试反馈,不然得出他们的学习特征。比如:学习者浏览、下载、提问、评论、答疑、测试行为都一定程度上反映出对所隶属的学习主题的兴趣和关注;学习者提问、评论、答疑、测试行为能反映出其对所属的学习主题认知能力;学习者上传、评论、答疑行为反映了学习者的热心度。

定义 1. 学习者的学习行为组成学习信息对象集合 $SB = \{b_i | 0 \leq i \leq M\}$, M 表示对象总数; 每个 b_i 可以用一个向量 $v = \{(c_1, a_1, v), (c_2, a_2, v), \dots, (c_k, a_k, v)\}$, c_k 表示本体中的概念, a_k 表示学习行为, v 表示学习行为的分值; 所有学习信息对象被量化为三个维度: 关注度、认知能力、热心度, 构成特征向量 $V = \{(c_1, i_1, d_1, e_1), (c_2, i_2, d_2, e_2), \dots, (c_k, i_k, d_k, e_k)\}$, k 表示该学习者感兴趣的学习主题

数, c_k 表示第 k 个主题概念; i_k 表示对 c_k 的关注度, 层级分为 {初步, 一般、专注}; d_k 表示对 c_k 的认知能力, 划分为 {不了解, 了解, 理解, 应用} 4 个级别; e_k 表示对 c_k 的热心度, 也划分为 5 类, 值分别为 {0, 0.25, 0.5, 0.75, 1}.

4 学习伙伴推荐算法

4.1 基于本体的主题概念相关性

一般而言, 从图论的角度, 领域本体概念之间关系是一种多对多关系, 表示的是一个图结构. 在虚拟学习社区中, 知识库的组织和核心本体概念所对应的知识点是主要是派生-继承关系和整体-部分关系, 在此可以将其看成是一种特殊的图结构---主题概念树。

定义 2. 若 C_X 和 C_Y 为概念树中任意的两个主题概念, 则它们之间的最小边数, 称为语义距离, 记作 $D(C_X, C_Y)$.

不难看出, 两个概念的语义距离越小, 其相关度越大; 反之, 则相关度越大. 且 $D(C_X, C_Y)$ 满足: ①非负性: $D(C, C_Y) \geq 0$; ②同一性: $D(C_X, C_X) = 0$; ③对称性: $D(C_X, C_Y) = D(C_Y, C_X)$.

在本体概念树中, 若两个主题概念离根越远, 表示语义越具体, 则两个主题概念越相近。

定义 3. $Depth(C)$ 为概念树中主题概念 C 的深度, $Depth(root) = 1$; $Depth(C) = Depth(CParent) + 1$, 其中 $root$ 表示根主题概念, $CParent$ 为主题概念 C 的父概念; $Depth(T)$ 为概念树的深度, 其值为最大主题概念深度。

定义 4. $CA(C_X, C_Y)$ 表示主题概念 C_X 和 C_Y 共同祖先集合; $NearestCA(C_X, C_Y)$ 为它们的最近共同祖先。

显然, 同样语义距离的 2 个主题概念, 它们的最近共同祖先的深度越大, 说明它们之间的重合度越好, 相关度应该增加. 依次原理定义如下两个概念之间的相关度:

$$RD(c_x, c_y) = \frac{\alpha + 2Depth(NearestCA(c_x, c_y))}{\alpha + D(c_x, c_y) + 2Depth(NearestCA(c_x, c_y))} = \frac{\alpha + 2Depth(NearestCA(c_x, c_y))}{\alpha + Depth(c_x) + Depth(c_y)} \quad \text{公式 (1)}$$

其中, C_X 和 C_Y 为概念树中任意的两个主题概念, α 为调整因子。

4.2 最佳学习伙伴发现准则

学习伙伴发现的算法思想: 虚拟学习社区是真实学习社群的数字化延伸, 在构建上可以借鉴真实学习环境中交互有效发生的经验: 首先, 学习者之间一定要有相近的学习背景和学习主题; 其次, 助学者必须是该

领域中的专家, 在认知水平上具备较高的造诣; 最后, 助学者具备热心度.

设有任意的 2 个学习者 x、y, 其对应的用户特征向量空间分别为 $V_x = \{ (c_{x1}, i_{x1}, d_{x1}, e_{x1}), (c_{x2}, i_{x2}, d_{x2}, e_{x2}), \dots, (c_{xm}, i_{xm}, d_{xm}, e_{xm}) \}$ 和 $V_y = \{ (c_{y1}, i_{y1}, d_{y1}, e_{y1}), (c_{y2}, i_{y2}, d_{y2}, e_{y2}), \dots, (c_{yn}, i_{yn}, d_{yn}, e_{yn}) \}$, $\bar{c}_x \in V_x$ 是学习者 x 的兴趣学习主题, 现要在虚拟学习社区中寻找评分值最高的 n 个学习者成为 x 学习主题的最佳学习伙伴(即助学者), 从而产生推荐.

定义学习者 x,y 知识背景相似性 BSim(x,y)如下:

$$BSim(x,y) = \frac{\sum_{c_x \in V_x, c_y \in V_y} \max_{RD(c_x, c_y)} RD(c_x, c_y)}{m} \quad \text{公式 (2)}$$

定义学习用户 x 关于学习主题对任一学习者 y 的伙伴评分 Partner(x,y)计算如下:

$$Partner(x,y, \bar{c}_x) = \max_{c_y \in V_y} e^{BSim(x,y)+c_{y1}} \times RD(\bar{c}_x, c_{y1}) \times d_{y1} \quad \text{公式 (3)}$$

分析公式(3)不难发现, 当学习者 x 与学习者 y 的学习经历越相似, 系数因子会越大, 否则会变小; 学习者 y 对学习主题的认知度和热心度越大, 也会让评分公式的值升高, 反之, 则值会降低. 显然, 公式(3)反映了现实学习环境交互有效发生的经验.

4.3 算法描述

基于上述学习伙伴发现准则, 在开放虚拟学习社区中可以在线实时为学习者关注的学习主题推荐 Top-n 助学者.

算法输入: 目标学习者 x 及其关注的学习主题 \bar{c}_x

算法的数据结构:

- ① 基于主题概念树的知识库
- ② 学习者特征向量表

算法的输出: 目标学习者 x 关于学习主题 \bar{c}_x 的 n 个助学伙伴

算法的执行步骤:

算法可以分为两部分执行: 离线计算主题概念相关度和在线实时推荐.

① 离线计算:

基于主题概念树根据公式(1)算出树中任意两个主题概念之间的相关度, 为公式(2)和公式(3)的实时运算做准备. 由于本体知识库的构建一般在系统设计初期就开发完成, 后期的管理维护相对稳定, 所以离线执

行, 可以减低实时计算的时间.

② 在线实时推荐:

- a) 基于学习者特性向量表, 利用公式(2)和公式(3)算出目标学习者 x 与其他用户之间在学习主题 \bar{c}_x 上的伙伴评分.
- b) 对得到的评分值做降序排列
- c) 选取 Top-n 个作为目标学习用户 x 的学习伙伴.

5 实验分析

为了验证所提出算法的有效性和进一步改进算法, 我们设计具备一般性且简化的软件系统虚拟学习社区. 实验结果表明, 所提出的学习者模型能够有效地表示学习者的个性化特征, 所提出的学习者相似性计算较好地区分出相近的学习伙伴.

5.1 算法实验步骤描述

1) 建立基于本体的软件系统知识库

咨询领域专家后, 受篇幅所限, 建立一个简化的软件系统领域本体, 为虚拟学习社区中的学习者和学习资源标注提供统一、标准的知识背景. 生成的知识空间结构如图 2 所示(具体应用中可以构建更加详细的概念关系). 将概念树中的结点按照层次遍历的顺序对其进行编号, 具体编号对应的概念名如表:

表 1 概念关系图中编号与概念名的对应列表

编号	概念名	编号	概念名	编号	概念名
1	软件系统	12	C 语言	23	DIRECTOR
2	基础软件	13	C#语言	24	数据类型
3	应用软件	14	JAVA 语言	25	运算表达式
4	操作系统	15	SQL SERVER	26	C 语句结构
5	程序语言	16	ORACLE	27	数组
6	关系数据库	17	MS-OFFICE	28	函数
7	办公软件	18	金山 WPS	29	指针
8	图像软件	19	PHOTOSHOP	30	文件
9	多媒体软件	20	AUTOCAD	31	Word
10	Windows	21	3DMAX	32	Excel
11	Linux	22	FLASH	33	PPT

2) 本体树中的主题概念相关度

根据公式(1), 设置调整因子, 可得到本体中两个概念之间的相关度. 计算得到部分概念的相关度列举如下表:

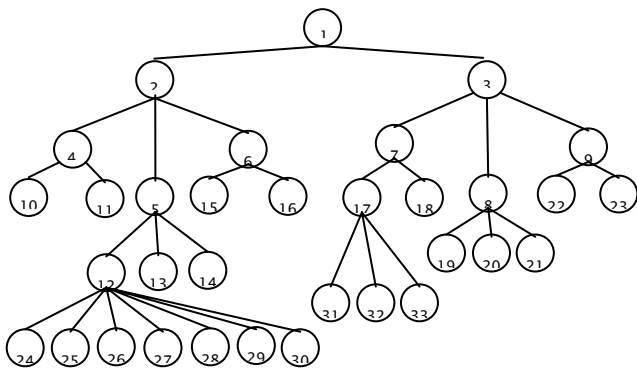


图 2 知识空间中的概念关系

表 2 概念相关度

相关概念	值	相关概念	值
RD (PHOTOSHOP, PHOTOSHOP)	1	RD (WORD, WORD)	1
RD(Windows, Linux)	0.7649	RD(Linux, Windows)	0.7649
RD(MS-OFFICE, 金山 WPS)	0.7647	RD(PHOTOSHOP, 3DMAX)	0.7647
RD(数组, 函数)	0.8095	RD(数组, 指针)	0.8095
RD(基础软件, 应用软件)	0.5556	RD(数据类型, DIRECTOR)	0.2632

从表 2 中的可以看出, 概念相关度值具有几个特点: ①概念自身的相关度为 1; ②概念之间相关度是对称的; ③当概念之间距离相等时, 相关度值会随深度的增加而增大。

3) 学习伙伴推荐

设经过一段时间学习之后, 学习者的学习行为信息描述为表 3 的特性向量表。

表 3 学习者特征向量表

学习者	学习记录
Learner1	(Windwos,1,2,0.5)(Word,1,3,0.25)
Learner2	(Linux,2,4,0.25)(Word,1,5,0.75)
Learner3	(Windows,1,4,0.25) (c#语言,1,5,0.25)
Learner4	(Windwos,2,5,0.5) (Word,1,5,0.25) (Excel,1,2,1)
Learner5	(Windwos,2,4,0.75) (Word,1,4,0.75) (PPT,1,3,1)
Learner6	(PHOTOSHOP,1,1,0)(FLASH,1,1,0) (PPT,1,3,0.5)
Learner7	(PHOTOSHOP,2,5,0.5) (autocad,1,3,0.5)(3DMAX,1,2,0.5)
Learner8	(PHOTOSHOP,2,5,0.5) (c#语言,1,5,1) (SQL SERVER,1,4,1)

根据公式(2)和公式(3)计算:

① Learner1 想学习 Word 时, 对其他学习者的伙伴评分。

② Leraner6 想学习 Photoshop 时, 对其他学习伙伴的评分

表 4 Leraner1 与其他学习者的学习经历相似度及

Word 主题学习伙伴评分

	BSim	Partner
Learner2	0.8824	15.5095
Learner3	0.6316	2.4246
Learner4	1	11.892
Learner5	1	13.4544
Learner6	0.3685	5.0345
Learner7	0.3839	4.3708
Learner8	0.5016	4.3708

表 5 Learner6 与其他学习者的学习经历相似度及

Photoshop 主题学习伙伴评分

	BSim	Partner
Learner1	0.5856	4.2268
Learner2	0.5856	5.975
Learner3	0.2838	2.1288
Learner4	0.5856	4.782
Learner5	0.5856	4.782
Learner7	0.8246	12.523
Learner8	0.6677	11.2325

令 $n=3$, 可得向 Learner1 推荐的助学者是(Learner2, Learner5, Learner4), 而向 Learner6 推荐的学习伙伴是(Learner7, Learner8, Learner2)。

从表 4 中可以看出, Learner2 与 Learner1 的学习经历较为接近, 该用户在学习 Word 的认知程度达到 5, 热心度达到 0.75, 表明其对 Word 的掌握程度很高, 而且很乐于助人, 所以评分最高; 而 Learner4, Learner5 虽然和 Learner1 的经历相似度为 1, 但由于认知程度或热心度低于 Learner2, 因此评分都比其低。另外, 对于 Learner6 的学习伙伴评分中, 虽然 Learner7, Learner8 对于 Photoshop 认知能力和热心度相当, 但是由于 Learner7 与 Learner 的学习经历更为接近, 所以评分更高, 更优先得到推荐。

由此可见, 从表 4、表 5 中得到的评分差异, 反映了虚拟学习社区中在助学者发现经验上与真实学习环境是一致的, 表明所提出的学习伙伴推荐算法能促进学习者之间有效交互的发生, 从而提高开放虚拟学习社区的智能化和个性化, 提升学习者的学习效率和效

果。

5.2 算法时间复杂度分析

算法的执行可分为两部分: 离线执行和实时执行。

由于本体概念树在系统设计时就基本稳定, 公式(1)的运算可以完全离线进行, 在实时计算时可以直接读取。

对于实时计算部分, 假定虚拟学习社区中的用户数为 n , 本体概念树中的主题概念数为 m ; 分析公式(2), 计算某个学习者与另一个学习者之间的学习经历相似度最坏情况下的时间复杂度为 $O(m \times m)$, 而计算学习者与其他用户的推荐评分则为 $O(n \times m \times m)$ 。

但仔细分析后不难发现, 由于学习者学习的知识点一般都相对有限, 可以认定知识点个数为常量 C , 因此, 实际运行的算法复杂度一般为 $O(n \times C \times C)$ 。

参考文献

- 1 Tillman NH. Virtual community building using internet tools <http://www.Hopetillman.com/i100/vc.htm>. [2004-10-16].
- 2 赵建华, 李克东. 信息技术环境下基于协作学习的教学设计. 电化教育研究, 2000(4):7-13.
- 3 黄怀荣. CSCL 的理论与方法. 电化教育研究, 1999(6):25-30.
- 4 王陆. 虚拟学习社区的社会网络结构研究[学位论文]. 西安: 陕西师范大学, 2009.
- 5 张立国. 虚拟学习社区交互研究[学位论文]. 西安: 陕西师范大学, 2008.
- 6 刘志勇, 刘磊, 刘萍萍, 杨帆, 贾冰. 一种基于语义网的个性化学习资源推荐算法. 吉林大学学报(增刊), 2009(9):391-395.
- 7 吴彦文, 黄珍. 混合算法用于学习资源的自动推荐. 计算机应用与软件, 2008(12):202-204.
- 8 丁兴富. 远程学习圈. 构建远程教学与远程学习的基础理论. 中国远程教育, 2001(7):10-14.
- 9 高毅, 申瑞民, 杨帆, 等. 基于开放 E-Learning 平台的学生行为分析中心. 计算机工程, 2004(15):86-87.
- 10 谢忠新, 王林泉, 葛元. 智能教学系统中认知型学生模型的建立. 计算机工程与应用, 2005(3):229-232.
- 11 林木辉, 张杰, 汪星一. 一种基于本体的学科知识库构建方法. 闽江学院学报, 2008(5).
- 12 林木辉, 张杰, 包正委. 智能教学系统中基于本体的知识表示及推送研究. 福建师范大学学报(自然科学版), 2009(1).
- 13 Fensel. Ontologies: silver bullet for knowledge management and electronic commerce. Berlin. Springer-Verlag. 2001.