

基于主体网格的个体驱动自主学习系统

王茂光^{1,2,3}, 管红杰², 史忠植¹

(1. 中国科学院计算技术研究所智能信息处理重点实验室, 北京 100080; 2. 中国矿业大学计算机学院, 徐州 221008;
3. 中国科学院研究生院, 北京 100039)

摘要: 网上学习系统大多只是教材的一种电子化, 没有实现用户的个性化教学。该文提出建立基于个体驱动的智能知识导航图。在学习过程中, 用户可定制个性化学习内容, 主体网格平台可通过 Agent 动态搜集个性化信息, 在参考已建立的知识点本体后, 基于用户行为观察的方式创建个性化学习档案, 并利用主体网格智能平台 AGrIP 的多 Agent 协作提供个性化的、智能自主学习系统, 为用户提供更好的知识服务。

关键词: 主体网格; 个体驱动; 知识导航; 自主学习

Ontology-driven Autonomic Learning System Based on Agent Grid

WANG Mao-guang^{1,2,3}, GUAN Hong-jie², SHI Zhong-zhi¹

(1. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080;
2. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221008;
3. Graduate University of Chinese Academy of Sciences, Beijing 100039)

【Abstract】 The instruction system on the Web lacks the personalized teaching due to the simple digitalization. This paper proposes an ontology-driven intelligent knowledge navigation map. In the personalized, intelligent and autonomic learning system, users can customize the study content directly, and the agent grid platform can collect personalized information and build personalized study documents based on the users' behaviors observations and evaluations, referring to the built knowledge ontology. It makes use of agent grid intelligent platform AGrIP to build an autonomic learning system and provides better knowledge services.

【Key words】 agent grid; ontology-driven; knowledge navigation; autonomic learning

1 概述

网络技术的迅速发展改变了人们的学习方式。但目前大多数网上学习系统提供的教学内容相对比较分散, 缺乏一种有效的方式将知识点很好地组织在一起, 无法满足用户个性化自主学习的要求。

本体作为一种能在语义和知识层次上描述信息系统的概念模型建模工具, 正得到越来越多的研究和应用。本体是共享概念模型明确的形式化规范说明^[1]。本体的目标是获取、描述和表示相关领域的知识, 提供对该领域知识的共同理解, 确定该领域内共同认可的词汇, 并从不同层次的形式化模式上给出这些词汇和词汇间相互关系的明确定义。本体语言是一种用来描述领域结构的语言, 即对领域本体进行刻画。这些结构可以采用一定的概念(类)和属性(关系)来描述, 而本体就由这些结构所表示的公理所组成。一般的本体语言可以建立在描述逻辑的基础之上, 如由 DARPA 组织提出的 DAML+OIL 以及 W3C 提出的 OWL 等。本体也可用来关联网页上的信息、相关知识结构及推理规则, 即在网页中嵌入本体。本文就是利用本体对课程内容进行合理的组织。

主体 Agent 一般是指处于特定环境下具有自主性、能动性、社交能力、反应能力等特征的硬件或基于软件的计算实体。主体通常被看成是一个意图系统, 它有自己的心智态度, 这些态度可分为 2 类: 信息类态度(如知识、信念)和能动性态度(如愿望、意图、规划等)^[2]。主体的研究在国际上受到高度重视, Barbara Hayes-Roth 在 IJCAI'95 的特约报告中指出:

智能的 Agent 既是人工智能最初的目标, 也是人工智能最终的目标。

基于主体的系统具有许多优点: 一个主体具有它所在领域的大量知识以及自身的能力和自标; 主体可以通过观察、与其他主体的通信以及对自己行为的意识而不断扩充其知识; 主体能运用其知识和能力来推理, 进行理性的规划, 并执行相应的动作^[3]。由于主体具有自治、社会交互能力, 具备感知外界环境、调整自身行为等简单对象没有的智能特性, 因此成为构建健壮、可扩展的软件系统的很有前途的技术。

主体网格本质上是充分利用多 Agent 技术建立一种自治的网格系统, 它有效地将自主性、智能性融入到网格中, 让具有智能的计算机程序在动态、开放、分布、异构和不确定的环境中协作, 为人们提供各种自主的智能服务。

本文提出了利用主体网格平台建立个体驱动的个性化知识组织方式, 利用主体网格多 Agent 协作建立个性化、智能、自主的学习系统, 以提供更好的知识服务。

基金项目: 国家“863”计划基金资助项目(2001AA113121, 2003AA115220); 国家“973”计划基金资助项目(2003CB317004); 国家自然科学基金资助项目(90604017); 北京市自然科学基金资助项目(4052025)

作者简介: 王茂光(1974-), 男, 副教授、博士研究生, 主研方向: 多主体系统, 网格计算, 软件工程, 网上教学; 管红杰, 讲师、硕士; 史忠植, 研究员、博士生导师

收稿日期: 2007-05-10 **E-mail:** wangmg@ics.ict.ac.cn

2 本体驱动的知识导航

自主学习系统提供了一个以学生为中心的网上学习环境,实现个性化学习,自主学习系统应能按照每个学生的知识背景及所要学习的内容动态地组织知识材料,以满足自主、个性化学习的需要。本文的方法是建立层次化的知识本体,采用知识点作为课程的最基本单元,整个课程是由知识点组成的网状结构;利用本体建立知识点及其关系,从而可以对其进行推理。

2.1 层次化知识本体

本体Ontology可以形式化为 $Ontology = \langle Concept, Relation, Function, Axiom, Instance \rangle$ 。其中,概念Concept可以指任何事物,如描述、功能、行为等,通常构成一个分类层次;关系Relation代表了在领域中概念间的交互作用,形式上定义为 n 维笛卡儿积的子集: $R: C_1 \times C_2 \times \dots \times C_n$,基本的关系有4种: part-of, kind-of, instance-of和attribute-of;函数Function是一类特殊的关系,形式化为 $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$,即前 $n-1$ 个元素可以决定第 n 个元素;公理Axiom代表永真断言,是定义在“概念”和“属性”的限定和准则;实例Instance是指属于某概念类的基本元素,从语义上分析,实例表示的是对象,而概念表示的是对象的集合,关系对应于对象元组的集合^[4-5]。

本文采用OWL定义层次化知识本体,每个学习者被视为一个主体Agent,根据教学大纲,Agent有必须达到的学习目标,综合评估每个学习主体所处状态和需求,以定位于系统某个预定义的学习目标,系统提供的知识导航^[6]都是由本体驱动面向该目标来进行的,如图1所示。

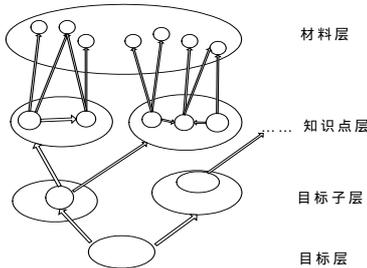


图1 基于本体的知识网络层次

(1)目标本体,定义各门课程及每门课程章节的学习目标及其关系。根据所学的课程、章、节定义知识点的相关性。一个学习目标是知识点实体的集合,而中间的子目标知识根据某种策略人为地把相关的知识点聚集起来,因此是虚拟层。子目标的存在使知识体系结构更清晰,有利于在自主学习时选择学习子目标,方便理解、记忆,并使知识导航功能更灵活。系统提供可视化的树型结构,把目标-子目标-知识点连接起来,表示出整个学习目标的知识体系。

(2)知识点本体,是彼此独立而又具有一定相关性的知识点组成,是该模型结构的核心层,是一个有向无环图。每个知识点定义一个概念性知识范畴,并确定它的认知度参数,表示各知识点在整个课程体系中的重要性。各知识点间的关联参数被记录在本体库中,其中的参数有知识点的前驱知识点、后继知识点及知识点之间的关联度。

(3)领域本体,主要存放构成知识点的相关内容,每个知识点可关联到若干领域本体,若干相关度大的知识点的集合构成子目标,而若干子目标组成一个学习目标,这样就把一门课程的知识通过本体间的关系划分为4层。

2.2 本体驱动的知识点导航

从基于本体的知识网络层次中的知识点本体中可获得各知识点的难度系数、类型、知识点之间的关联度、考核习题及其与知识点间的关联度等。在学习过程中,根据每个学生的学习模型,在动态学习的过程中建立个性化学习档案。

系统通过测试学生对知识的掌握情况,结合知识点给出的各种参数,进行评估,通过知识点本体为学生提供下一步应该学习哪一个知识点的建议:在推进学习进度时给出后继知识点;在回顾已学得知识时给出先驱知识点。算法如下:

```
If start_study{
Then next_study_state(study_content, awake)
And next_link_state(observation, awake)
And next_link_state(observation_result,awake)
And previous_link_state(study_state, review)
And current_link_state(study_state, bad)
And evaluation() }
```

根据基于本体的知识网络层次图和知识点导航原理构造出知识点网络结构图,根据该图为学生提供可视化学习导航的功能,知识点网络结构图把一门课程中各知识点层次及其相关性显式地表示出来,并标识出该知识点在整个知识体系中的位置,使学生可以从整体上把握自己的学习过程。

2.3 个性化学习档案的建立

实现个性化自主学习系统的难点是在学习过程中,学习主体Agent对知识的认知是动态变化的。本文采用Agent技术,在动态学习的过程中用已建立的层次化知识本体作为参考本体,自动地建立个性化学习档案。

首先,在建立课件时用本体注释Web文档和应答查询,将语义信息编码到网页中描述学习内容。然后通过Agent动态搜集学生学习个性化信息,建立用户学习个性档案,一种办法是显式地让用户提供知识点、重难点等,存储成用户学习个性档案,该方法简单易行、容易实施,缺点是加重了用户负担,而且用户学习目标会随时间动态变化。另一种办法是在参考已建立的知识本体基础上,基于用户行为观察的方式创建个性化学习档案。通过记录用户网上学习的行为来建立用户学习个性档案,记录的信息包括学习内容、访问网页的频率、时间、网页是否被标记收藏等^[7-8]。用户访问的网页被自动分类成各个概念,这些概念包含在参考本体中,分类结果被积累起来。这使得参考本体中的概念得到了以用户浏览的相关信息量为基础的权重。

用已建立的知识本体作为参考本体,定义相关课程的概念及其关系等,如数据结构中的树内容,可以用UML描述,如图2所示。

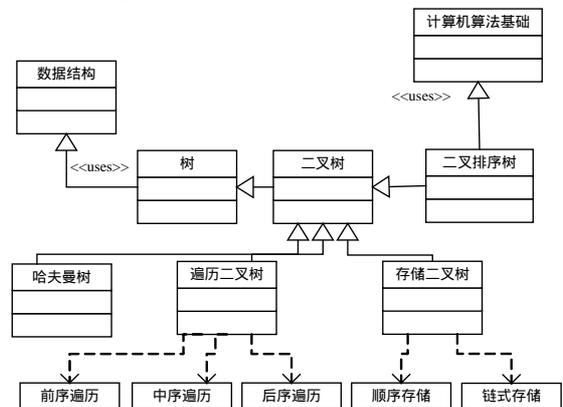


图2 数据结构树本体

在本体使用过程中,可以对其推论并进一步得到学生整

个知识体系的掌握情况,还可以与其他课程的相关内容(如计算机算法课中的排序算法等)建立联系。

当用户访问时,从用户Web浏览器缓存搜集来的网页周期性地被分类到恰当的知识点参考本体中。对每个被访问的页面来讲,不失一般性,设文档 k 向量为 $d_k=\{d_{1k}, d_{2k}, \dots, d_{nk}\}$,文档 k 参考的知识点本体 j 概念向量为 $c_j=\{c_{1j}, c_{2j}, \dots, c_{nj}\}$, d_k 与 c_j 的相似度用如下方法计算:

$$\text{similarity}(d_k, c_j) = \frac{\sum_{i=1}^n (d_{ik} \cdot c_{ij})}{\sqrt{\sum_{i=1}^n d_{ik}^2 \cdot \sum_{i=1}^n c_{ij}^2}}$$

其中, d_{ik} 为文档 k 中第 i 个词汇的权重; c_{ij} 为概念 j 中第 i 个词汇的权重; n 为文档集和 D 中特有词汇的数量。

有最高相似度值的概念被认为是与学习的内容和目标有关的。在 Agent 自动根据用户访问建立个性学习档案的过程中,一开始,用户个性学习档案中的所有本体概念的权都是 0。随着越来越多的网页被分类,分类器得出的权重逐渐积累起来。被访问文档的概念权重不断增加,高权重的概念代表用户学习的内容更关注。

在用户学习浏览的过程中,进一步考虑网页的长度和访问的时间。直观上讲,如果用户长时间停在某网页上,他们对此学习内容的关注应该增加。但是,如果网页很长,那么时间因素的影响可能应该下降,因为时间长可能是由于信息量大。用如下公式结合时间和网页长度调整相似度:

$$\text{simi}(d_k, c_j) = \text{timelength} \cdot \text{similarity}(d_k, c_j)$$

$$\text{timelength} = \ln \frac{\text{time}}{\text{length}}$$

整个智能导航学习过程如图 3 所示。

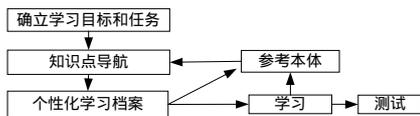


图 3 智能导航学习过程

3 基于主体网格的自主学习系统

3.1 主体网格平台 AGrIP

主体 Agent 技术集软件、通信、分布系统的技术于一体,能封装私有特征,具有灵活的协调能力和更好的网络适应性,具有比对象粒度更粗、抽象级别更高的特征,更符合人的认知过程。本文在主体网格的基础上实现了主体网格平台 AGrIP,以有效实现网格平台中异构资源的共享、多主体的协同工作等,主体网格平台 AGrIP 平台(<http://www.intsci.ac.cn>)在 2006 年通过了专家组的鉴定,达到国内领先水平。平台结构如图 4 所示。

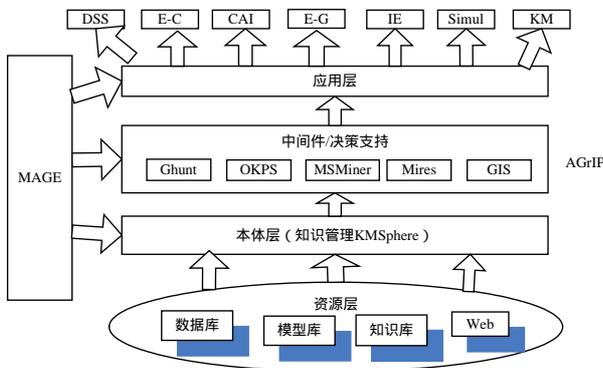


图 4 AGrIP 平台结构

主体网格平台 AGrIP 由多主体平台 MAGE(Multi-Agent

Environment)、中间件层(本体中间件层、工具软件层)和应用层组成。底层核心平台使用多主体平台 MAGE 集成异构数据源和软件,把这些工具包有效地组合成为一个智能信息平台。多主体环境 MAGE 是基于智能主体和多主体技术,利用最新的面向主体的软件方法学和软件工程的基本思想,开发的面向主体的软件分析、设计、实现和部署的集成开发和运行环境。在该环境中,可以引导用户进行面向主体的建模和设计,为用户提供一种面向主体的软件开发和集成模式。

本体中间件层 KMSphere(<http://www.intsci.ac.cn>)提供基于本体的语义信息理解和推理,中间件工具软件层包括智能信息搜索工具 GHunt、多策略数据挖掘工具 MSMiner、专家系统工具 OKPS 等为高层应用提供具体领域的开发。应用层提供了各种智能服务,如辅助决策、协同商务、自主学习等。

3.2 基于主体网格平台的自主学习系统

设自主学习模型 $ST=\langle Ag, I, V, A, P, O, F \rangle$,其中, $Ag=\{a_1, a_2, \dots, a_n\}$ ($n \geq 1$)为同时参与学习的学生主体,对于单个学生, $n=1$,为其特例; $I=\{I_1, I_2, \dots, I_m\}$, m 为学习任务总数,每个任务与学习目标相关; $V=\{V_{11}, V_{12}, \dots, V_{1m}\}$, $I \in [1, m]$, V_{ij} 代表学习任务 I_i 的各项考核指标,如分数、课程设计等; $P=\{p_1, p_2, \dots, p_m\}$ 为学习任务的偏序集合,定义了学习中的次序关系; $O=\{O_1, O_2, \dots, O_m\}$ 为学习目标的参考本体; F 为评价函数,判断学生的学习曲线。

基于 AGrIP 的自主学习系统如图 5 所示。

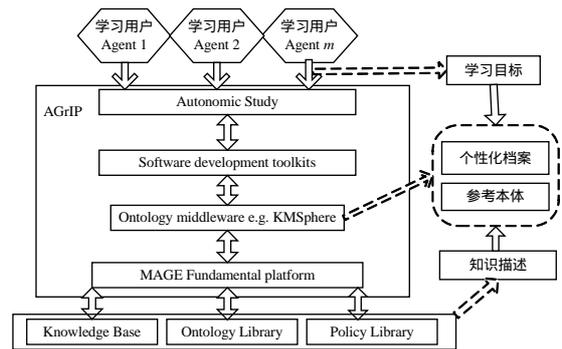


图 5 基于主体网格的自主学习系统

主体网格平台 AGrIP 建立在多 Agent 平台 MAGE 的基础上,提供基于主体的分布式网络计算环境,设计的 Agent 完全符合 FIPA 规范标准。

AGrIP 的工作流程如下:(1)学习 Agent 通过用户界面知识智能导航图选择所学的内容。(2)系统自动通过软件工具层对数据资源进行处理。(3)通过本体中间件 KMSphere 提供基于本体的语义信息的理解和推理,挖掘出有用的信息。(4)建立个性化学习档案,AGrIP 会将该个性化档案存储到 XML Profile 文件中,在参考本体的基础上,进一步对其进行推理并动态调整所学内容。其中,协调 Agent 会分析用户学习过程中的问题,并与参考本体比较,若发现某个知识点掌握不好,提醒前驱和后继节点显示并掌握;信息 Agent 负责学习内容的存取。(5)对单个学生和班级的学习效果进行评价,判断单个学生的学习曲线和班级学习成绩是否符合正态分布。

自主学习以用户个性化学习档案为先导来进行学习过程中的动态提示和推理,以用户特定的需求以及在一段时间内的学习目标为衡量标准来筛选信息。用户界面提供友好的智

(下转第 34 页)