

基于本体的个性化智能答疑系统的设计^①

Design of Ontology – Based Personalized Intelligent Answering System

程南清 (宁波广播电视大学信息与教学资源中心 浙江宁波 315016)

摘要: 针对目前智能答疑系统存在着没有充分考虑学生个体差异性的现象,本文提出了一个基于本体的个性化智能答疑系统,利用本体构建专业领域知识库,使系统实现在语义层次上检索学生的提问,系统同时通过网络教学系统获取学生的学习行为特征,并以此为条件过滤系统检索的结果,以向学生提供适合其学习的答案集。经实验表明,本系统能提供较强的个性化智能答疑服务。

关键词: 本体 个性化智能答疑 知识库 学习特征模型 推理

1 前言

答疑系统是网络教学系统重要的组成部分,它能够一定程度上跨越时、空的局限性及时解答学生的疑难问题,消除学生的学习障碍,加强远程师生的交流。随着信息技术和人工智能技术的不断发展,网络教学中出现了更方便、有效的智能答疑系统,智能答疑系统在分词算法、问题模式匹配和答案库设计等方面做了大量的研究,极大的提高了系统的智能性和响应速度,降低了教学人员的工作压力。但当前智能答疑系统还存在着以下现象:答疑系统独立于教学管理系统(LMS),没有充分考虑学生在线学习的行为特征和认知水平等个体学习的差异性,向所有的学生提供相同的内容和呈现方式的答案集。

针对上述问题,本文提出一个基于本体的个性化智能答疑系统模型。系统采用本体(Ontology)构建领域本体知识库,既方便实现领域知识的层次结构和语义的表示,使系统可在概念语义层次上理解处理和检索学生的提问;又可有效地实现与网络学习管理系统(LMS)共享知识库,将网络教学中的传道与解惑环节相结合,方便获取学生在线学习的特征和认知水平,使系统有效地向其提供适合其学习的问题答案集和学习建议。

2 基于本体的个性化智能答疑系统模型

为了帮助学生获取适合其个性特征和认知水平的

问题答案,以便更好的解决网络学习中的障碍,本文提出的基于本体的个性化智能答疑系统原型,如图1所示,主要功能模块包括:知识库、学生特征提取、问题处理、问题查询和问题解答个性化呈现等部分组成。

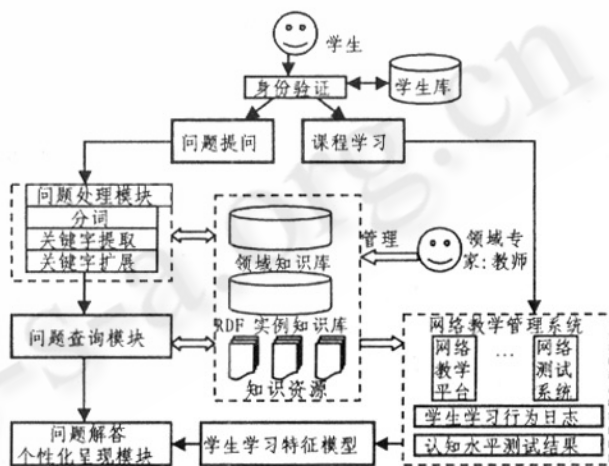


图1 基于本体的个性化智能答疑系统模型

2.1 知识库

知识库是系统的基础。知识库由知识资源、领域知识库两部分构成。知识资源是回答学生提问的知识文档,由领域专家事先写入本地知识库,并在系统运行和反馈过程中不断充实。领域知识库由领域本体库和RDF实例知识库两部分组成。领域本体则是领域内共同认可的词汇和词汇之间的相互关系,是机器理解和交流信息的基础。RDF实例知识库则是知识文档的语

① 基金项目:宁波市2007年教育科学规划研究课题

义索引库,它支持系统对知识文档进行语义检索,建立语义索引必须借助领域本体对知识文档进行标引。

2.2 学生学习特征提取模块

学生学习特征提取模块主要功能是通过分析网络教学管理系统(LMS)收集的学生在线学习行为信息,而获取学生的在线学习行为特征和认知水平。学生学习特征通过学习特征模型进行表示,随着学生的学习,模型内的学生学习特征数据也将动态地修改。系统不提供匿名服务,学生注册后,系统将为学生建立学习特征模型,学生方可学习、提问。

2.3 问题处理模块

学生在学习过程中随时运用自然语言进行提问,系统首先将学生的问题提交给问题处理模块。问题处理的第一步是分词,分词是机器对自然语言表达的问题文本进行关键词拆分,系统以专业领域内已加入到本体知识库的概念为词表,进行匹配(反向、正向)学生自然语言提问的文本,从中提取出能代表学生提问意图的关键词集合,然后再利用领域本体对关键词进行同义、近义扩展,并由系统构造查询表达式。这样就可将本应由学生完成的构造查询表达式的过程,转变为由系统自动实现,减轻学生检索负担,以实现答疑系统较高的智能性。

2.4 问题查询模块

根据问题处理阶段的结果,问题查询模块应用本体概念关系的推理函数,进行语义检索,在知识库中搜索与问题相关的答案,并按相关规则实时的返回结果集。如果没有检索到答案,则将问题自动转发给领域专家进行人工答疑,领域专家进行相关问题答案的组织,并对本体知识库进行人工维护(增加、删除或修改等),以异步的方式向学生提供答案。

2.5 问题解答个性化呈现模块

问题解答个性化呈现模块根据学生的学习特征模型,利用语义过滤函数进行过滤查询模块提供的答案集,将适合学生认知水平的答案以其喜欢的形式(如:媒体类型)提供给学生。

3 主要技术的实现

基于本体的个性化智能答疑系统涉及两个方面的关键技术:课程知识库本体的构建和基于本体的个性化推理。

3.1 知识库的本体构建

本体的概念来源于哲学,主要是研究客观事物存在的本质。T. Gruber 认为,本体是对概念化对象的明确表示和描述。Nicola Guarino 把概念化对象 C 定义为一个有序三元组 $C = \langle D, W, R \rangle$,其中: D 是一个领域; W 是该领域中相关事务状态的集合; R 是领域空间 $\langle D, W \rangle$ 上概念关系的集合^[1]。本体可以形式化规范的说明概念模型,并可通过概念之间的关系描述概念的语义,是一种有效表现概念层次结构和语义的模型。

建构基于本体的知识库,是利用领域本体进行全面表达领域知识,使知识库能以机器可读、可理解的形式实现对深层知识和隐性知识的描述,知识库的建构是答疑系统进行智能信息处理的基础。一般而言,课程知识体系是由知识点的集合与知识点之间逻辑关系的集合组成的,知识点是根据教学大纲,对课程进行分解后的原子单元,而传统的教材以章节知识点的层次结构形式组织课程知识体系,这种简单的线性关系,不能清晰地表示知识点的各种语义关系(如语义蕴含关系和外延关系),更不能在问题检索时实现对概念语义的理解。因此,领域知识库本体的构建实质就是研究单个概念(知识点对象)的属性特征和各个概念(知识点对象)之间的相互关系,使用本体技术将这些概念(知识点)及其相互关系形式化地表示并存储于计算机中。为了方便阐述,相关描述具体如下:

定义 1 领域本体 (domain ontology) 为一个三元组, $O = (\text{Meta_data}, C, R)$, 其中 Meta_data 定义领域本体的基本元数据,包括领域本体标识号、名称、主题、领域专家(如:课程教师)等元信息; C 为领域本体范围内的概念集合; R 为领域本体内概念之间的关系集合。C、R 具体定义如下:

定义 2 领域本体 O 概念间的关系 $R = (\text{SR}, \text{PR}, \text{FR}, \text{CR}, \text{IR}, \text{Before}, \text{After})$ ^[2], 其中:

SR (Synonymy Relation) 为概念间的同义关系,表示一个概念与另一个概念所表示的意义相同,如: SR (操作系统, OS) 表示“操作系统”与“OS”同义,因此,通过同义关系可实现概念语义的扩展。

PR (Partner Relation) 为概念间的类属关系,表示一个概念与它的子类概念之间的关系,可形成概念之间的逻辑层次分类结构。如: PR (DOS, OS) 表示

DOS 是 OS 的子概念, OS 是 DOS 的父概念。PR 关系具有传递性 ($PR(C1, C2) \wedge PR(C2, C3) \rightarrow PR(C1, C3)$)、属性继承 ($PR(C1, C2) \wedge FR(C2, F) \rightarrow FR(C1, F)$)、实例归属 ($PR(C1, C2) \wedge IR(i, C1) \rightarrow IR(i, C2)$) 的性质。通过 SR 关系可以充分地描述出知识库的知识层次结构。

FR (Feather Relation) 为特征关系, 指某一概念与其特征属性之间的关系, 如: $FR(OS, Media)$ 表示“OS”概念拥有的媒体类型的属性。

CR (Compose Relation) 为合成关系, 即: 一个概念与其各分概念之间的整体和部分的关系, 如: $CR(MS Word, MS Office)$ 表示 MS Word 是 MS Office 的成员。通常父概念可以由其所有以 FR 关系和 CR 关系为连接弧的子概念来代替, 来实现相应语义扩展。

IR (Instance Relation) 为实例关系, 即: 对于概念 C 及其实例集 S, 当 $e \in S$, 则 e 与 C 的关系就是实例关系。IR 关系具有属性继承的性质, 即: $IR(e, C) \wedge FR(C, F) \rightarrow FR(e, F)$ 。实例关系体现了概念与个体之间的关系, 为知识的存储和维护提供了规范。

Before、After 用来表示概念出现的时间顺序先后或因果关系, Before 表示一个概念在另一个概念之前出现, After 表示一个概念在另一个概念之后出现, 这两个关系具有传递性和逆关系, 可表现出所有概念之间的前序或后序关系。

定义 3 概念 $C = (Meta_Data, Learn_Step, Media_Cat, Diffs, *relation)^{[3]}$ 。其中 Meta_Data 包含概念的标识号、名称、关键字、所属本体等基本元数据。Learn_Step 为学习概念 C 需要的学习过程集, 一个完整的学习过程至少包括说明、内容、练习 (或测验) 和评估等 4 个阶段^[4]。Media_Cat 为概念 C 知识文档表现的媒体类型集, 知识文档可由文本、演示文稿、图片、动画、视频、音频等 9 类^[5] 媒体表现形式。Diffs 为概念所涉及的教学内容的难度集, 如: 难、一般和易等。*relation 为概念 C 的指针集, 包含该概念与其它概念各类关系的指针, 如: 指向同义概念的指针、指向下层概念的指针、指向上层概念的指针、指向具体实例概念的指针等。

根据上述定义, 知识库利用本体进行知识表示, 实现了规范、清晰的定义领域内的概念和概念之间的关

系, 体现出领域内的知识层次关系和知识结构体系, 实现将一个传统的知识库转化为一个用本体描述的语义网络, 将传统的知识点之间逻辑关系的推理转化为基于本体概念关系的语义网络的推理, 提高了概念查询、推理的速度和准确度。本文以《大学计算机基础》课程为例, 构建了部分领域本体, 图 2 是其中一部分, 领域概念之间的弧代表语义关系, 本体被表示成一个网状的结构。

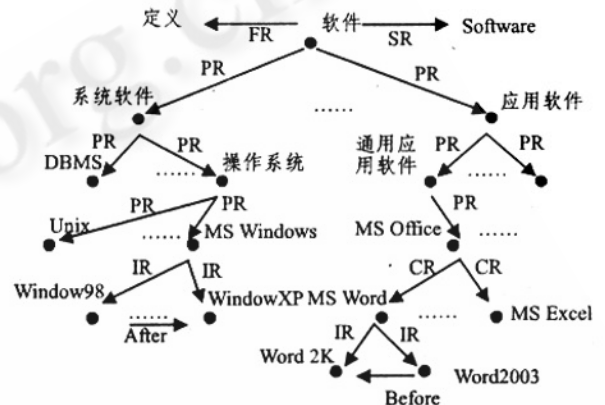


图 2 《大学计算机基础》课程部分领域本体语义网络图

智能答疑系统给出的答案集是一组领域知识文档序列, 这些知识文档被机器可读、可理解是答疑系统进行智能信息处理的基础。因此, 需要建立知识文档语义索引库, 具体定义如下:

定义 4 知识文档的语义索引 $RDFs = (Doc_Base, C, Properties)$, 其中 Doc_Base 为每一个知识文档的基本元信息, 如: 文档的 URI、文件名等; C 表示该知识文档所描述的概念, 根据定义 3, 本体中每个概念都有众多不同属性的知识文档描述; Properties 是一个三元组 $(Media, Diff, Learn_Step)$, 表示该知识文档的属性值, 如: 媒体类型、难度水平和适合的学习阶段。

知识文档通过语义索引后, 通过属性三元组, 使机器方便地根据学生的学习情况 (如: 学习进度、知识水平和喜好媒体类型) 提供不同答案集, 实现答疑的智能化、个性化。

3.2 基于本体的个性化推理

为了实现关键词的语义扩展检索, 向学生提供符合其学习特征的个性化问题答案集, 本文定义了学生学习模型和对本体操作的推理函数、过滤函数, 系统将学生学习模型中的学习特征转化为知识库本体语义网

络的语义环境,以过滤检索信息,实现向学生提供其感兴趣的信息。

3.2.1 学生学习模型的建构^[6]

学生学习模型是一种面向网络学习的用户模型,它是对网络学习环境中的某个或某类学生的抽象,描述学生的认知水平、个性特征等多方面的信息。学生学习模型是实现个性化答疑的前提和基础。根据系统设计需要,将学生学习模型描述如下:

$Student_Model = (Base_Info, Media_Info, Level_Info, Learn_Step, Other_Info)$

其中 $Base_Info$ 指学生基本信息,如:姓名、学号、专业、年级、班级、通信方式等。 $Media_Info$ 指学生喜欢的媒体类型,如:文本、演示文稿、图片、动画、视频、音频等。 $Level_Info$ 指学生的认知水平,是一个二元组 $(C_i, Level)$, $level$ 表示对概念 C_i 的认知水平,如:掌握、理解、略知、不知等。 $Learn_Step$ 是一个二元组 $(C_i, Step)$,指学生学习概念 C_i 的进度。 $Other_Info$ 指学生其它的学习信息。随着学生的学习,其学习模型数据也随之动态更新。

学生学习模型的初始值通过表单询问方式而获取。随着学生在线学习的进行,系统根据网络教学管理系统(LMS)中学习模块和测试模块收集的学生点击的知识文档序列(类 $Web\ Log$)信息和测试信息,结合知识文档的属性,经适当的统计分析而获取学生的关键学习特征,如:学生偏爱的媒体类型、知识点的学习进度和掌握情况等。

3.2.2 推理函数和个性化过滤函数

本体的推理就是分析学生提交的问题并抽取核心关键词,根据语义分析结果,查找核心概念和问题在本体库中的所在位置,并检索出问题的答案,然后根据学生的学习特征过滤这些答案,将适合其特征的答案提供给学生。为了实现问题关键词的语义检索,本文定义 6 个语义推理函数和 1 个过滤函数,由这些函数实现个性化检索功能:

(1) 推理函数

函数 $Fun_SR(C)$ 是基于 SR 关系的函数,可获取概念 C 的所有同义的本体概念(同义的词汇),如: $Fun_SR(Windows\ 2000) = \{Windows\ 2K, Win\ 2000, Win\ 2K, \dots\}$ 。

函数 $Fun_CR(C)$ 是基于 CR 关系的函数,可获取组

成概念 C 的各部分概念,如: $Fun_CR(MS\ Office) = \{MS\ Word, MS\ Excel, MS\ Access, MS\ Powerpoint, \dots\}$ 。

函数 $Fun_FR(C)$ 是基于 FR 关系的函数,用于获取概念 C 的属性值,可以细化学生针对某一概念 C 的提问,比如: $Fun_FR(OS) = \{定义, 发展史, 分类, \dots\}$, 则“OS 的定义”与“OS 的分类”是针对“OS”概念的不同问题。

上述三个函数可以将问题关键词进行语义扩展和转换,以提高检索的全面性和精确性。

函数 $Fun_PR(C)$ 是基于 PR 关系的函数,可从本体 O 中求出概念 C 的所有子概念以及其父概念。利用此函数可以快速的检索到概念 C 在语义网络中位置及相关的概念。

函数 $Fun_Before(C)$ 、 $Fun_After(C)$, 是基于 $Before$ 、 $After$ 关系的函数,可获取概念 C 的前序或后序概念,可以将检索的结果按照先后顺序进行排序。

函数 $Fun_IR(C)$ 是基于 IR 关系的函数,可从本体 O 中求出概念 C 的所有实例,可以为学生提供相关学习参考资料。

(2) 过滤函数

根据定义 4,概念 C 拥有众多不同属性的知识文档。为了机器能够根据学生不同的学习行为特征,而提供适合其特征的知识文档,本文还定义了一个过滤函数:

$Fun_Filter(C, Stu_Cha(stu_id))$, 其中 C 为被检索的概念 C ; $Stu_Cha(stu_id)$ 为学生 (stu_id 为学生标识号)学习行为特征,是一个三元组: $(learn_step, Media, Diff)$ 。如图 3,过滤函数将推理函数获得的概念集进行个性化过滤,将每个概念中不相关的知识文档排除掉。

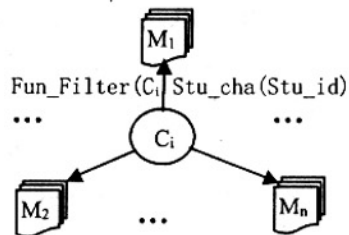


图 3 过滤函数过滤概念 (C_i) 的知识文档

4 实验结果与性能分析

本文以面向网络教育的本科基础课《大学计算机基础》为实验对象,具体实验方法为:注册一个学生用

户 A, 选取 3 个关键词为提问关键词, 测试条件假设: (1) A 初始特征为偏爱文本类知识文档; (2) 同一关键词进行 2 次查询, 第 1 次是 A 注册后, 完成初始假设特征设置后进行, 第 2 次是完成第 1 次查询后进入网络教学系统学习, 并要求 A 浏览视频和 PPT 类知识文档, 然后进行查询, 则具体实验的相关数据如表 1 所示。

从表 1 中实验结果数据可以看出:

随着学生在线学习行为的变化, 系统向其提供的答案也发生变化。由于每个学生的学习行为和认知水平是不同的, 因此, 本系统可实现向学生提供个性化的答疑服务。

表 1 部分实验结果数据

| 检索词 | 相关记录数 | | | 第 1 次 | | | 第 2 次 | | |
|------|-------|-----|----|--------|-----|----|--------|-----|----|
| | | | | 检索出的数量 | | | 检索出的数量 | | |
| | 文本 | PPT | 视频 | 文本 | PPT | 视频 | 文本 | PPT | 视频 |
| 操作系统 | 17 | 3 | 2 | 27 | 0 | 0 | 7 | 5 | 3 |
| 网络 | 23 | 5 | 3 | 43 | 0 | 0 | 17 | 8 | 3 |
| 病毒 | 16 | 2 | 1 | 32 | 0 | 0 | 11 | 6 | 6 |

5 结束语

本文构建了一个基于本体的个性化答疑系统, 简

介了系统模型及其各个组成部分, 同时重点阐述了领域知识库的本体构建, 以及系统利用本体函数的推理向学生提供个性化的答案集。经实验表明, 本系统能有效地根据学生学习行为和特征, 利用本体的语义关系, 向学生提供不同的表现形式、深度的个性化答案集, 系统具有较强的针对性和导向性。

参考文献

- Guarino N, Formal ontology and information systems [A]. Formal Ontology in Information System [C], Trento: TOS Press, 1998: 6-8.
- 张正, 左春, 王裕国. 基于语义扩展的 Web 服务发现方法[J]. 通信学报, 2007, (1): 57-63.
- 程南清. 基于学习对象的个性化网络教学资源的研究与构建[J]. 中国电化教育, 2007, (9): 56-58.
- 加涅. 学习的条件和教学论[M]. 皮连生译, 上海: 华东师范大学出版社, 1999. 69-84.
- 《教育资源建设技术规范(征求意见稿)》, <http://www.edu.cn/html/keyanfz/doc/d.DOC>.
- 吴文彦, 吴郑红. 基于学习者个性模型的智能答疑平台的设计[J]. 电化教育研究, 2005, (6): 64-66.