

文章编号: 1673- 9469(2011) 01- 0110- 03

基于 ontology 的智能答疑系统的研究

李立¹, 管燕²

(1. 安庆广播电视大学, 安徽 安庆 246003; 2. 连云港高等师范专科学校, 江苏 连云港 222006)

摘要: 答疑系统是现代远程教育中一个重要的组成部分。针对目前智能答疑系统在知识共享和问题求解精确性方面的不足, 本文提出一种基于 ontology 的智能答疑系统模型, 以大学计算机信息技术教程课程为例, 构建课程知识本体, 并在基于本体的推理和答案匹配模式方面进行了尝试。

关键词: 本体; 智能答疑; 知识库

中图分类号: TP391

文献标识码: A

The design of intelligent question answering system based on ontology

LI Li¹, GUAN Yan²

(1. Anqing Radio and Television University, Anhui Anqing 246003, China;

2. Lianyungang Teacher's College, Jiangsu Lianyungang 222006, China)

Abstract: The question answer system is the important part of the Distant Education. This paper proposes for an intelligent question answer system to improve performance of knowledge share and answer accuracy in intelligent question answering system, which is realized on the basis of the ontology. Taking a course in university information technology' for an example, the authors construct the semantic network of course ontology. Attempts on the reasoning based on ontology and answer pattern matching are conducted.

Key words: ontology; intelligent question answer; repository

课程答疑系统能够在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, 加强远程教育中师生的交流。目前国内具有智能性的网上答疑系统比较少, 大多数采用电子邮件、留言板、聊天室、语音会议系统等基于网络的人际交互方式实现人工答疑。同时也出现了一些结合基于数据库的自动答疑系统, 但这些系统一般智能性不强, 准确性不高, 答疑很难做到广泛的知识共享。设计一个反馈及时、答疑智能、知识共享、回答准确的网络答疑系统, 解决实践中所提出的问题, 已经成为网络教育中要解决的一个重要课题。

本文提出的基于 ontology 的智能答疑系统模型, 以《大学计算机信息技术》课程为例, 构建该课程的部分领域本体, 在自然语言提问的语义理解上, 深入到领域内部结合本体与自然语言处理技术, 设计基于 ontology 的智能答疑系统。

1 ontology 答疑系统模型

基于 ontology 的智能答疑系统模型, 用户用自然语言提出问题, 先在 FAQ 库中匹配答案。如果有, 将结果返回。否则, 进入知识库进行检索, 该模块主要包括问句分析、查询产生和知识查询三个组件, 通过这三个组件与领域本体库和语义知识库进行交互, 从中提取答案, 反馈给用户。如果没有将返回, 答案不满意, 则等待人工答疑。

2 系统的主要技术实现

2.1 构建领域本体

领域本体是专业性的本体, 可以帮助教学专家建立人性化的知识答疑管理系统与方法, 提高教学水平, 使答疑系统中的知识和人类用自然方式表达的信息可以相互转换, 使答疑系统更加符

合学习者的需要^[1]。目前构建本体还没有公认的建设标准, 通常都按照具体领域的应用需求, 确定自己的原则、标准和定义。现在较为流行的有两种主要的构建领域本体的方法: (1) 在领域专家的帮助下用本体描述语言将本体描述出来。(2) 从结构化的数据或文本中学习和抽取(发现)概念, 构建领域本体。

在本系统中, 将这两种方法相结合, 首先采用基于文本解析和语义关联分析的方法从文本中抽取和学习概念和概念间关系, 得到本体的原型, 即本体学习; 领域专家或用户可以通过本体编辑器修改本体原型以达到满意的程度。具体步骤如下:

合理界定本体的领域和范围。课程领域本体可依据已有的教学大纲较准确地确定其范围。如《大学计算机信息技术教程》, 内容主要包括计算机组成原理、计算机软件、计算机网络、数字媒体、信息系统等。

确定本体中的类和实例。类是具有类似性质的所有个体的抽象, 个体称为类的实例。

根据继承关系建立类之间的层次结构。具体的方法有三种: 自顶向下法、自向上法、综合法。

给各个类定义合适的槽。槽是类和实例的描述符。用以描述对象的一种属性, 或与其他对象的关系。

建立公理库。在定义了以上各元素之后, 用公理来解释这些元素并表示它们之间的关联和约束。

描述类中的实例。同一个类中的实例尽管使用相同的槽, 但是槽值各不相同, 根据描述的实例来最终确定槽值。

以 Protégé _ 3. 2 作为表示本体内部概念关系的绘图工具。

在整个过程中, 最重要的是确定核心概念及其关系。要进行领域本体的构建首先要确定领域中的关键概念。确定上位概念、较为显著的概念和常用概念, 以建立核心概念集, 进而确定本体中的类。

整理出核心概念后, 要进一步明确这些核心概念的定义和他们之间的关系。本系统对《大学计算机信息技术教程》的 218 个常见问题进行分析, 发现这些问题中的概念主要存在以下 8 种关系。因此在遵循本体建立准则^[2]的基础上, 采用自顶向下法, 采用 8 种概念关系, 按章构建课程知识本体。概念由 11 元属性描述^[3]: $C = (n, f, pa, ki, in, at, sy, cop, af, be, conv)$ 。

其中, n 表示概念 C 的名称; f 表示概念 C 的标识码; pa 表示概念 C 的部分概念; ki 表示概念 C 的类属概念; in 表示概念 C 的实例概念; at 表示概

念 C 的属性概念; sy 表示概念 C 的同义概念; cop 表示与概念 C 比较; af 表示概念 C 的后序概念; be 表示概念 C 的前序概念; $conv$ 表示概念 C 转换成另一概念。

作为领域知识的一种模型, 领域本体需要不断的改进才能接近人对客观世界的认识。没有哪一个时刻的本体是完善的。因此, 本体库需要在系统的使用过程中不断地加以完善和维护。

2. 2 问题模式化

在智能答疑系统中, 首先需要对学生所提出的问题进行分析处理。对于同一个问题, 学生可能有多种提问方式。因此, 需要对这种一个问题多种提问方式的情况进行分析, 以便能够得到一个统一的表示方式, 即问题模式。

在本系统中由于使用本体, 涉及关系为类、属性、实例以及它们之间关系, 通过问题模式化可以提供对以下三类问题的识别支持: (1) 求类或实例的描述问题。如: “什么是存储器?”。(2) 求类或实例的某些特定属性值问题。如: “存储器具有哪些性能?”。(3) 求类、实例及属性之间关系问题。如: “内存和外存有何区别?”。

针对以上三种问题模式, 本文使用一个查询确定算法。

模式一: 问句中只映射一个实体 A , 其类型为 $ClassA$ 。则得出需要查询的是关于 A 的描述(DESCRIBE), 故产生输出查询三元组(DESCRIBE? x) 和 $\langle ? x, rdf: typeOf, OEIQAS: ClassA \rangle$ 。

模式二: 问句中映射的实体 A 的类型为 $ClassA$, B 的类型为 $propertyB$ 。根据 RDF 语义模型, 则丢失需要查询的元素为 $Class$ 类型实例, 不妨设置为“? $xclass$ ”。若 $PropertyB$ 的定义域为 $ClassA$, 则输出查询三元组 $\langle OEIQAS: A, OEIQAS: B? xclass \rangle$ 。若 B 的值域为 $ClassA$, 则输出查询三元组 $\langle ? xclass, OEIQAS: B, OEIQAS: A \rangle$ 。否则输出为 null。

模式三: 问句中映射的实体 A 的类型为 $ClassA$, B 的类型为 $ClassB$ 。则丢失需要查询的元素为 $Property$ 类型实例, 不妨设置为“? $xproperty$ ”。若 $ClassA$ 有属性 $PropertyA$ 的值域是 $ClassB$, 则产生的查询三元组为 $\langle OEIQAS: A, ? xproperty, OEIQAS: B \rangle$; 若 $ClassB$ 有属性 $PropertyB$ 的值域是 $ClassA$, 则产生的查询三元组为 $\langle OEIQAS: B, ? xproperty, OEIQAS: A \rangle$; 若上述情况都满足, 则产生查询三元组集 $\langle OEIQAS: A, ? xproperty, OEIQAS: B \rangle$ 、 $\langle OEIQAS: B, ? xproperty, OEIQAS: A \rangle$, 关系为 OPTIONAL。

2.3 基于本体的知识推理

本体的推理就是分析用户提交的问题并抽取核心的关键词,特别是疑问词,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,根据本体库中的信息,到知识库中提取问题的知识,并转换为易于理解的表达形式,发布给用户。对于基于 ontology 的智能答疑系统来说,应该充分发挥本体的推理能力,做到对知识的推理与扩展,让系统更加智能^[4]。本系统利用本体属性的特征,在产生词语三元组时定义了如下推理规则扩充语义教育资源陈述:

1) 领域本体中有属性 $P_1 \text{ inverseOf } P_2$, P_i 是 P_1 的实例, P_n 是 P_2 的实例; ; 对于知识 $\text{Statement} \langle S_i, P_i, O_i \rangle$ 成立, 则新知识 $\text{Statement} \langle O_i, P_n, S_i \rangle$ 也成立。

2) 领域本体中有 P_i 是属性 P 的实例, 而 P 又被声明为 $\text{TransitiveProperty}$ 。对于知识 $\text{Statement} \langle S_i, P_i, O_i \rangle$ 、 $\langle O_i, P_i, O_n \rangle$ 成立, 则新知识 $\text{statement} \langle S_i, P_i, O_n \rangle$ 也成立。

3) 领域本体中有 P_i 是属性 P 的实例, 而 P 又被声明为 SymmetricProperty 。对于知识 $\text{Statement} \langle S_i, P_i, O_i \rangle$ 成立, 则新知识 $\text{Statement} \langle O_i, P_i, S_i \rangle$ 也成立。其中, inverseOf 、 $\text{TransitiveProperty}$ 和 SymmetricProperty 分别表示 OWL 本体的翻转、传递和对称属性特征。

2.4 答案匹配

本系统通过 3 种途径匹配用户答案:

将用户提问的标准问句在 FAQ 库中搜索, 如果有相同的, 直接把 FAQ 库中这个问题对应的答案返回。

如果搜索后发现 FAQ 中没有相同的问题, 本系统借助关键词序列, 不仅计算用户问题与 FAQ 中问题的相似度, 还计算用户问题与 FAQ 中答案的相似度, 这样做突破了目前常用的方法, 即计算用户问句与 FAQ 库中问句的语义相似度, 返回相似性较高的问题所对应的答案。问题与答案的表示可以采用常用的向量空间模型, 问题与答案的相似度计算可以计算向量的相似度。本系统设定一个相似度的阈值为 α , 将相似度大于 α 的答案按相似度的大小反馈给用户。

如果应用前两种途径都匹配不到的答案的,

则等待人工处理, 并将处理后的问题、答案添加到 FAQ 库中。

3 实验结果与性能分析

为了验证本文提出的基于 ontology 的智能答疑系统的有效性, 对《大学计算机信息技术教程》第六章内容分别构建了课程的知识本体, 针对该课程的 80 个随机提问, 测试结果如表 1。

表 1 实验结果

Tab. 1 Experimental results

实验总题目数: 80 个
直接回答题目数: 57 个 直接回答准确率: 71.25%
人机交互回答题目数: 14 个 间接回答准确率: 17.5%

从实验的数据来看, 总体还是比较满意的, 总体准确率达 88.75%。表明本文提出的基于 ontology 的智能答疑系统切实可行且有效。该系统在直接回答的准确率上还不是很高, 是因为构建的领域本体不够丰富, 很大程度上影响着答疑系统的准确率。系统的本体知识库还没有经过多次的实践迭代。本体工程的开发应该是一个反复迭代的过程, 在初始版本建立后, 还要反复地将其应用到实际系统中或者与领域专家进行评价和排错, 直至满意。

4 结束语

利用领域本体提供的知识和丰富的语义信息, 使该系统的准确率更高, 可以得到较满意的结果。

参考文献:

- [1] 史月飞, 张小艳, 邵刚. Ontology 理论及其在智能答疑系统中的应用[J]. 科技信息(学术研究版), 2007, 229(17): 192-193.
- [2] DAVID EICHMANN. The RBSE crawler- balancing effective search against web load[A]. In Proceedings of the First International World Wide Web Conference[C]. Geneva, Switzerland, 1994. 113-120
- [3] 管燕, 仲兆满. 基于课程知识本体的 FAQ 库自动生成方法研究[J]. 中国远程教育, 2010(07): 68-72.
- [4] 张争万. 基于本体论的网络教育智能答疑系统模型研究[J]. 软件导刊, 2008(12): 35-36.

(责任编辑 刘存英)