

文章编号: 1007- 2985(2010) 05- 0060- 03

基于背景知识的本体进化框架设计*

张子振

(安徽财经大学管理科学与工程学院, 安徽 蚌埠 233030)

摘要: 随着应用环境和用户的需求不断变化, 本体必须不断进化. 目前本体进化工作要解决的最重要的问题是如何确定新的数据信息与进化本体中已存知识之间的正确关系. 基于背景知识的本体进化框架, 利用背景知识挖掘出新旧数据信息之间的关系, 改善了整个本体进化过程.

关键词: 本体; 本体进化; 背景知识

中图分类号: TP311

文献标志码: A

本体是构成语义 Web 的重要因素, 同时由于本体的知识共享和可重用性, 本体也经常被用作先进信息系统的知识骨架. 所以本体管理人员必须及时的对本体进行更新, 以反映影响这些系统生命周期的变化, 如系统数据源的变化, 新功能的增加等, 这就是本体进化要解决的问题.

本体进化是对本体进行的适时调整以及变化一致性的传播^[1], 由于本体进化需要对变化的需求进行鉴别, 对进化本体实施相应的变化, 并对进化后的多个不同版本的本体进行管理, 所以本体进化是一项非常复杂、非常耗时的工作. 目前已有多项研究可以解决本体进化过程中的一些复杂问题. 文献[2- 3]探讨了本体进化过程中的变化表示问题和版本化过程实施的问题. 文献[4- 5]研究了如何挖掘出那些需要添加到进化本体中的潜在信息. 这些潜在信息主要是通过基于本体的信息系统数据源的变化而获得的. 笔者介绍的本体进化框架 Evolva 可以利用背景知识, 在进化本体中已经存在的知识与要添加到进化本体中的术语之间建立关系, 有利于本体进化过程中本体变化的自动实施. 先介绍了本体进化框架 Evolva 的基本构成, 接着重点介绍了本体进化过程中新旧数据信息之间关系发现的实现, 最后对本体进化的工作提出了展望.

1 本体进化框架 Evolva

本体进化框架 Evolva 完全依赖多个背景知识源支持整个本体进化过程的, 它包括 5 个基本的组成部件: 信息发现, 数据确认, 本体变化, 进化确认以及进化管理, 它的基本结构图如图 1 所示.

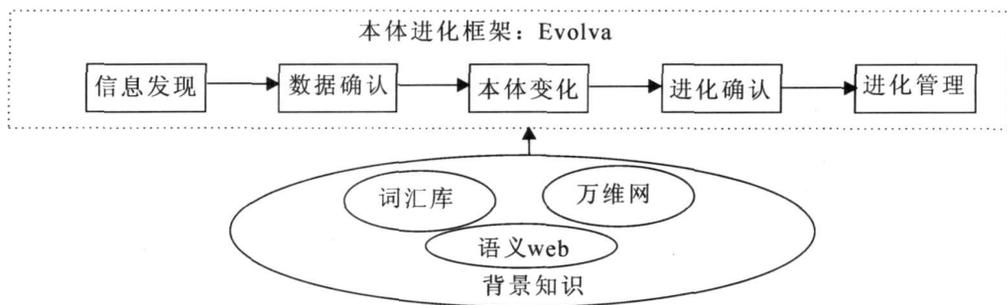


图 1 本体进化框架 Evolva 基本结构

1.1 信息发现

第 1 步工作是在与基于本体的信息系统相关的数据源中发现潜在的新信息. 数据源的格式有多种, 可以是非结构化的数据, 如文本文档, 也可以是结构化的数据, 如本体和数据库. 对于不同格式的数据, 处理方式也不同: 文本文档使用信息抽取, 本

* 收稿日期: 2010- 06- 11

作者简介: 张子振(1982-), 男, 山东聊城人, 安徽财经大学管理科学与工程学院讲师, 硕士, 主要从事语义 Web 信息检索、信息安全研究.

体学习或者实体识别技术处理; 其他外部本体则将其转换成与进化本体语言相兼容的本体; 数据库内容则转换成本体语言.

1. 2 数据确认

经过“信息发现”所发现的潜在信息要经过确认, 可以通过使用一系列的启发式规则 (如所抽取的术语的长度) 来实现数据确认, 尤其是从文本文献中发现的新信息. 对于结构化的数据 (本体和数据库) 则无需确认, 因为结构化的数据已经过明确良好的定义.

1. 3 本体变化

在所抽取的术语和进化本体中的概念之间建立正确的关系, 这些关系是通过搜索多个背景知识源鉴别出来的. 进化本体中相应的变化将被直接执行, 并记录相应的变化.

1. 4 进化确认

对本体实施变化后会引起本体的不一致性和不连贯性, 并且由于存在多个数据源, 在本体进化过程中, 也可能产生数据复制, 这样就会产生冲突知识, 对于这个问题可以利用自动推理技术来解决.

1. 5 进化管理

在本体进化以及进化本体对依赖本体的变化传播过程中, 给本体管理者一定程度的控制权, 主要是记录、跟踪本体变化, 找出并解决本体进化过程中未解决的问题.

2 关系发现的实现

2. 1 实现过程

本体进化的核心任务是将新知识集成到进化本体中, 传统的做法是由本体管理者完成. 对于集成到进化本体中的术语, 本体管理者要依据自己所掌握的领域知识确定进化本体中与新术语相关的元素, 以及它们之间的关系, 这是一项繁琐而耗时的工作. 本体进化框架 Evolvea 通过组合多个数据源 (词汇库 WordNet 语义 Web 中的在线本体以及 Web 中的文本文献), 利用其中的背景知识鉴别新术语和本体元素之间的关系. 整个关系发现的实现过程如图 2 所示.

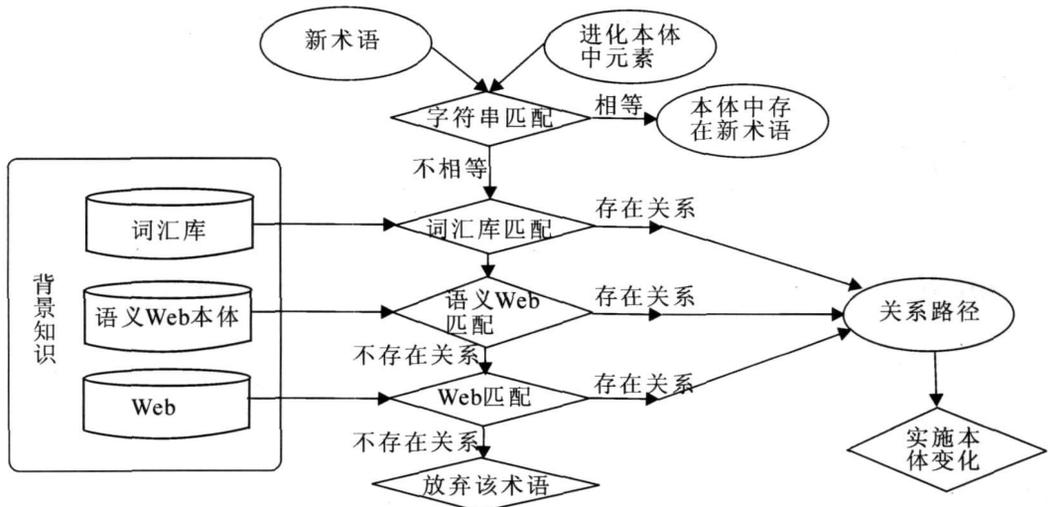


图 2 关系发现过程示意图

(1) 首先通过字符串匹配器 Jarv^[6] 检测进化本体中是否存在与新术语相同的元素, 如果存在则无需实施变化, 否则进入下一步;

(2) 在词汇库 WordNet 中搜索与新术语相关的元素, 并通过公式: $Sin(C_1, C_2) = 2N_3 / (N_1 + N_2 + 2N_3)^{17}$ 来计算二者的相似度. 若相似度比较高, 则可以通过 WordNet 推导出相应的关系; 反之, 若没有搜索到与新术语相关的任何元素, 则进入下一步.

(3) 通过关系发现引擎 Scarlet 对在线本体进行搜索, 确定新术语与其中元素的关系; 若不能找到与新术语有任何关系的元素, 则进入下一步.

(4) 通过搜索引擎应用程序接口在 Web 中搜索与新术语相关的元素, 若找不到与新术语有任何关系的元素, 则可以放弃这一新术语.

2. 2 实验数据

本体进化框架 Evolvea 研究人员从 20 篇文档中, 利用抽取算法抽取了 520 个术语, 进化本体中已存概念个数为 256 个. 通过字符串匹配器 Jarv^[6] 检测出 21 个术语在进化本体中已经存在, 有 7 个术语和进化本体中的相似度系数超过 0.92 剩下

的 492 个术语, 词汇库 WordNet 中搜索, 利用文献 [7] 中的相似度计算公式, 发现有 162 个术语与进化本体中的概念存在有关系, 所确定的部分关系实例如表 1 所示.

表 1 利用 WordNet 确定的部分关系

被抽取的术语	进化本体中的概念	关系	关系路径
合同	人	\subseteq	合同 \subseteq 代表 \subseteq 谈判者 \subseteq 联系人 \subseteq 人
企业	合作伙伴	\subseteq	企业 \subseteq 合作伙伴
儿童	人	\subseteq	儿童 \subseteq 人

3 结语

尽管本体进化框架 Evolve 可以利用外部的背景知识源实现本体进化, 但是仍有几方面工作需要进一步完善, 首先不同数据源的组合方式有待进一步研究, 目前试验中是对多个数据源进行线性组合, 接下来将考虑数据源的并行组合; 其次, 匹配过程和术语定位过程有待改善; 最后由于本体进化是一项不断持续的过程, 所以本体进化框架中的组件需要不断地优化.

参考文献:

- [1] HAASE P, STOJANOVIC L. Consistent Evolution of Owl Ontologies [C] // Proceedings of the Second European Semantic Web Conference Heraklion Greece 2005 182-197.
- [2] KLEIN M. Change Management for Distributed Ontologies [D]. Amsterdam: Vrije University, 2004.
- [3] NOY N F, CHUGH A, LIU W, et al. A Framework for Ontology Evolution in Collaborative Environments [C] // Proceedings of the 5th Int. Semantic Web Conference (ISWC-06). USA: Athens, 2006 544-558.
- [4] ALANIH, HARRIS S, ONEIL B. Windowing Ontologies Based on Application Use [C] // Proceedings of 3rd European Semantic Web Conference (ESWC-06). Montenegro Budva 2006 185-199.
- [5] BLOEHDORN S, HAASE P, SURE Y, et al. Ontology Evolution [M]. [S.l.]: John Wiley & Sons, 2006 51-70.
- [6] COHEN W W, RAVKUMAR P, FIENBERG S E. A Comparison of String Distance Metrics for Name Matching Tasks [C] // Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03). Mexico: Acapulco, 2003 73-78.
- [7] WU Z, PALMER M. Verb Semantics and Lexical Selection [C] // Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics Gaines NM, 1994 133-138.

Ontology Evolution Framework Based on Background Knowledge

ZHANG Zhi-zhen

(College of Management Science and Engineering Anhui University of Finance & Economics Bengbu 233030, Anhui China)

Abstract Ontologies need to evolve continually, for the application environment and users' application requirements are always changing. The most important issue in the ontology evolution is how to establish the right relation between the newly arising information and the existing knowledge in the ontology. An ontology evolution framework based on background knowledge is introduced, which can exhibit the relation between the newly arising information and the existing knowledge in the ontology, and improve the ontology evolution process.

Key words ontology; ontology evolution; background knowledge

(责任编辑 陈炳权)