信息检索技术

利用词语共现进行Ontology的概念获取

耿骞,耿崇

北京师范大学管理学院

摘要　作为大规模的语义知识资源库，Ontology在信息处理中具有重要的作用。但是，如何有效地构建Ontology却是一个重要的问题。对于自动构建Ontology的过程来说，首要的问题就是如何获取领域概念。本文尝试了一种利用词语共现获取领域概念的方法，用于支持领域Ontology的构建。该方法首先通过人工领域分析，获得起始领域概念，然后利用起始概念从语料库中抽取共现的概念，从而获取相关的概念知识。同时，本文以1998年1月份的人民日报语料库为语料，针对外交和体育两个领域，尝试从中提取相关的概念，从而检验利用词语共现获取领域概念的实际效果。

Abstract As a large semantic knowledge resource, Ontology plays an important role in information processing. However, how to construct an effective Ontology is an important problem to its application. The first issue in automatic Ontology creation is domain concepts acquisition. In this article we experiment on a method to obtain domain concepts which are based on lexical co occurrence (and then to support the automatic Ontology construction). The first step of this method is to obtain the primary starting concepts by manual analysis, and then to extract relative co-occurrence concepts from the corpus. Based on the corpus of People's Daily, January 1998, the article focuses especially on the fields of sports and diplomacy. We extract the relative concepts, to examine the practical results of co-occurrence-based domain concepts acquisition.

关键词　词语共现,概念提取,本体构建

Key words Lexical co-occurrence,Concept extraction,Ontology construction

分类号　G250 7
DOI:

通讯作者:
耿崇 g123ch@163.com
作者个人主页: 耿骞;耿崇

扩展功能

本文信息
- Supporting info
- PDF(0KB)
- [HTML全文](0KB)
- 参考文献[PDF]
- 参考文献

服务与反馈
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- 引用本文
- Email Alert
- 文章反馈
- 浏览反馈信息

相关信息
- 本刊中 包含"词语共现,概念提取,本体构建"的 相关文章

本文作者相关文章
- 耿骞
- 耿崇