

论文

一种基于知网的句子相似度计算方法

程传鹏,吴志刚

(中原工学院计算机学院, 河南 郑州 450007)

摘要:

句子相似度是衡量文档相似度的基础,在自然语言处理领域中有着非常重要的作用。目前的句子相似度计算方法忽略了句子的结构对相似度的影响。本文在分析已有研究工作的基础上,提出了一种改进的句子相似度计算方法。依据知网对“实体概念”的描述,构造出义原的语义层次树,由各个义原在树中的相对位置,计算出义原之间的相似度。对三种义原加权求和得到词语之间的语义相似度。综合句子的表层相似度和句子的词语语义以及词语的相对位置关系,得到句子的整体相似度。实验表明,在同等的测试条件下,本文所提出的句子相似度计算方法在相似度比较上更符合人的直观感觉。

关键词: 句子相似度 知网 表层相似度 语义偏移量

A Method of Sentence Similarity Computing Based on Hownet

CHENG Chuan peng,WU Zhi gang

(School of Computer Science,Zhongyuan Institute of Technology,Zhengzhou 450007,China)

Abstract:

Sentence similarity is the basis of document similarity, and sentence similarity computing plays an important role in the field of natural language processing. The current methods of sentence similarity computing neglect the influence of sentence structure. On the basis of the interrelated research, this paper proposes an improved method of similarity comparison. The semantic tree of sememe is constructed according to the description of entity conception in the Hownet, the semantic similarity of sememe is computed based on the relative positions in the sememe tree. Calculating of sentence similarity is based on surface similarity and semantic similarity. Under the same test conditions, the experiments show that the proposed method is much closer to the people's comprehension to the meanings of the sentences.

Keywords: sentence similarity; hownet; surface similarity; semantic offset similarity

收稿日期 2011-07-23 修回日期 2011-10-08 网络版发布日期 2012-02-25

DOI:

基金项目:

通讯作者:

作者简介:

作者Email:

参考文献:

本刊中的类似文章

1. 张鼎兴[1] 徐明[1] 高俊文[2] 刘爱心[3].一种多属性目标监测的无线感知网络覆盖算法[J]. 计算机工程与科学, 2008,30(4): 98-100
2. 孙■伟, 邢长征.关于中文文档复制检测算法的改进[J]. 计算机工程与科学, 2010,32(8): 101-103

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(444KB)
- ▶ [HTML全文]
- ▶ 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

本文关键词相关文章

- ▶ 句子相似度
- ▶ 知网
- ▶ 表层相似度
- ▶ 语义偏移量

本文作者相关文章

PubMed