

论文

基于规则挖掘和Naive Bayes方法的组合型歧义字段切分

张严虎 潘璐璐 彭子平 张靖波 于中华

四川大学计算机学院 四川大学计算机学院 四川大学计算机学院 四川大学计算机学院 四川大学计算机学院

摘要:

组合型歧义字段切分是中文自动分词的难点之一。在对现有方法进行深入分析的基础上,提出了一种新的切分算法。该算法自动从训练语料中挖掘词语搭配规则和语法规则,基于这些规则和Naive Bayes模型综合决策进行组合型歧义字段切分。充分的实验表明,相对于文献中的研究结果,该算法对组合型歧义字段切分的准确率提高了大约8%。

关键词:

Resolving combinational ambiguity in Chinese word segmentation based on rule mining and Naive Bayes method

Abstract:

Combinational ambiguity is one of the most difficult problems for Chinese word segmentation. After in-depth analysis of the other algorithms in literature, the paper proposed a new segmentation algorithm. The algorithm automatically mined word collocation rules and grammar rules from training corpus, and then made integrated decisions to resolve combinational ambiguity based on the mined rules and Naive Bayes method. Extensive experiments show that the proposed algorithm obtains an accuracy increase of 8% against the related works.

Keywords:

收稿日期 2008-01-14 修回日期 2008-03-07 网络版发布日期

DOI:

基金项目:

通讯作者: 于中华

作者简介:

参考文献:

本刊中的类似文章

文章评论 (请注意:本站实行文责自负,请不要发表与学术无关的内容!评论内容不代表本站观点.)

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(616KB)
- ▶ [HTML全文]
- ▶ 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  
本文作者相关文章

- ▶ 张严虎
- ▶ 潘璐璐
- ▶ 彭子平
- ▶ 张靖波
- ▶ 于中华

PubMed

- ▶ Article by
- ▶ Article by
- ▶ Article by
- ▶ Article by
- ▶ Article by

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标	<input type="text"/>	验证码	<input type="text"/> 1851

