

论文与报告

中文人称名词短语单复数自动识别

郎君, 秦兵, 刘挺, 李正华, 李生

1. 哈尔滨工业大学信息检索研究室 哈尔滨 150001

收稿日期 2007-5-9 修回日期 2007-9-20 网络版发布日期 接受日期

摘要

名词短语的单复数信息在共指消解中是必不可少的特征. 与英语不同, 中文属于汉藏语系, 名词本身不能明显体现单复数信息, 需要借助其所在的名词短语来进行体现. 本文在自动内容抽取 (Automatic content extraction, ACE) 语料上抽取得到人称名词短语的单复数信息, 分别采用了基于规则和机器学习的方法来进行人称名词短语的单复数自动识别. 基于规则的方法, 在一些知识资源的基础上定义了规则模板库, 每条规则采用槽和槽值的方法来进行体现; 机器学习方法采用最大熵模型组合考察了词形、词性、词义、数量关系等特征. 两种方法分别达到了 48.24\%和87.48\%的正确率. 实验结果显示, 基于规则的方法能够保证精确率而不能保证召回率, 机器学习的方法可以更好地完成单复数信息的识别任务.

关键词 [人称名词短语](#) [单复数](#) [机器学习](#)

分类号 [TP391](#)

Number Type Recognition of Chinese Personal Noun Phrase

LANG Jun, QIN Bing, LIU Ting, LI Zheng-Hua, LI Sheng

1. Information Retrieval Laboratory, Harbin Institute of Technology, Harbin 150001

扩展功能
本文信息
▶ Supporting info
▶ PDF(1177KB)
▶ [HTML全文](OKB)
▶ 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
▶ 复制索引
▶ Email Alert
相关信息
▶ 本刊中 包含“人称名词短语”的相关文章
▶ 本文作者相关文章
· 郎君
· 秦兵
· 刘挺
· 李正华
· 李生