

博士论文

中文分词和词性标注模型

刘遥峰, 王志良, 王传经

(北京科技大学信息工程学院, 北京 100083)

收稿日期 修回日期 网络版发布日期 接受日期

摘要 构造一种中文分词和词性标注的模型, 在分词阶段确定N个最佳结果作为候选集, 通过未登录词识别和词性标注, 从候选结果集中选优得到最终结果, 并基于该模型实现一个中文自动分词和词性自动标注的中文词法分析器。经不同大小训练集下的测试证明, 该分析器的分词准确率和词性标注准确率分别达到98.34%和96.07%, 证明了该方法的有效性。

关键词 [分词](#); [词性标注](#); [最短路径](#)

分类号 [TP18](#)

DOI:

通讯作者:

作者个人主页: [刘遥峰](#); [王志良](#); [王传经](#)

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF](#)(95KB)

▶ [\[HTML全文\]](#)(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“分词; 词性标注; 最短路径”的 相关文章](#)

▶ [本文作者相关文章](#)