

数据库、信号与信息处理

汉语文本自动分词算法的研究

何国斌, 赵晶璐

西南大学 计算机与信息科学学院, 重庆 400715

收稿日期 2008-8-5 修回日期 2008-10-28 网络版发布日期 2010-1-28 接受日期

摘要 分析了中文分词词典的机制, 提出了一种改进的整词分词字典结构, 并针对机械分词算法的特点, 将其与概率算法相结合, 探讨了一种中文自动分词概率算法。采用哈希及二分法对词典进行分词匹配。实验表明, 该算法具有较高的分词效率和准确率, 对于消歧义词也有较好的性能。

关键词 [自动分词](#) [分词算法](#) [字典](#) [歧义切分](#)

分类号 [TP391](#)

Research on algorithm of Chinese word automatic segmentation

HE Guo-bin, ZHAO Jing-lu

College of Computer and Information Science, Southwest University, Chongqing 400715, China

Abstract

Chinese segmentation mechanism is analyzed. An improved structure of segmentation dictionary is presented, and in view of the characteristics of the mechanical Chinese word segmentation, combined with probabilistic algorithm, a Chinese Word Automatic Segmentation probabilistic algorithm is discussed. Hashing and binary search is used to segmentation match. Experiment indicates that the algorithm can greatly improve the speed of Chinese segmentation and precision, and strengthen the processing of dispelling ambiguity.

Key words [automatic segmentation](#) [segmentation algorithm](#) [dictionary](#) [ambiguity segmentation](#)

DOI: 10.3778/j.issn.1002-8331.2010.03.037

通讯作者 何国斌 hgg07123@163.com

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(574KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“自动分词”的 相关文章](#)
- ▶ 本文作者相关文章

- [何国斌](#)
- [赵晶璐](#)