

知识组织与知识管理

全二分快速自动分词算法构建

张海营<sup>1,2</sup>

襄樊学院网络中心<sup>1</sup>

收稿日期 2007-1-30 修回日期 2007-3-13 网络版发布日期 2007-4-30 接受日期

**摘要** 分析现有分词算法存在的不足,在此基础上提出一种新的分词词典,通过为分词词典建立首字Hash表和词索引表两级索引,使得该分词词典支持全二分最大匹配分词算法,利用该分词算法进行自动分词,其时间复杂度实现了大的改善。

**Abstract** This paper analyses the shortcoming of segmentation algorithm, designs a new algorithm for Chinese phrase segmentation. By building two levels index for Chinese thesaurus, we attain a highly efficient Chinese phrase segmentation thesaurus which supports hashing operation by means of the first Chinese character in a string and full binary search. Based on this thesaurus, we design a new algorithm for Chinese phrase segmentation.

**关键词** [分词算法](#) [汉语分词](#)

**Key words** Segmentation algorithm; Chinese segmentation

**分类号** [G252.7TP391](#)

**DOI:**

通讯作者:

张海营 [xfu\\_www@126.com](mailto:xfu_www@126.com)

作者个人主页: 张海营

### 扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF](#) (OKB)

▶ [\[HTML全文\]](#) (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](#)

▶ [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“分词算法”的 相关文章](#)

▶ 本文作者相关文章

• [张海营](#)

•