

信息检索技术

基于Lucene的中文字典分词模块的设计与实现

向晖¹; 郭一平²; 王亮^{1,3}

华中科技大学控制科学与工程系¹

华中科技大学图书馆²

收稿日期 2006-5-19 修回日期 网络版发布日期 2006-8-15 接受日期

摘要 在分析Lucene语言分析器结构的基础上, 采用基于中文字典的正向最大匹配分词算法, 设计实现能够在基于Lucene的搜索引擎中处理中文信息的中文分词模块。

Abstract This paper introduces the construction of language analyzer in Lucene, designs and implements Chinese words segmentation module which uses forwards maximum match algorithm (FMM). This module can disposes Chinese information well and efficiently in the search engine based on Lucene.

关键词 [搜索引擎](#) [Lucene](#) [中文分词](#) [正向最大匹配算法](#)

Key words Search engine; Lucene; Chinese words segmentation; Forwards Maximum match algorithm

分类号 [G254](#)

DOI:

通讯作者:

向晖 xcaids@126.com

作者个人主页: 向晖 郭一平 王亮

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“搜索引擎”的 相关文章](#)
- ▶ 本文作者相关文章
 - [向晖](#)
 - [郭一平](#)
 - [王亮](#)
 -