

软件技术与数据库

基于压缩倒排文件的中文全文检索仿真系统

宋 懿, 国德峰

(上海交通大学计算机科学与工程系, 上海 200240)

收稿日期 修回日期 网络版发布日期 2008-4-24 接受日期

摘要 探讨基于压缩倒排文件的中文全文检索技术, 包括数据压缩方法、存储、检索与排名机制。借助中科院的高精度ICTCLAS中文分词系统, 采用C++/STL语言仿真实现了一个中文全文检索系统。该文列出部分关键代码, 利用搜狗实验室提供的数据进行实验。通过改进压缩算法, 系统的磁盘利用率提高了近80%。

关键词 [中文全文检索](#) [压缩倒排文件](#) [排名](#)

分类号 [TP391](#)

DOI:

通讯作者:

作者个人主页: [宋 懿](#); [国德峰](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF](#) (106KB)
- ▶ [\[HTML全文\]](#) (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“中文全文检索”的相关文章](#)
- ▶ [本文作者相关文章](#)
- ▶ [宋 懿, 国德峰](#)