数据库与信息处理

基于词平台汉字编码的自动标引研究

焦 慧，刘 迁，贾惠波

1.清华大学 精密仪器与机械学系，北京 100084
2.精密测试技术及仪器国家重点实验室，北京 100084

摘要　　自动标引是基于内容检索的关键技术之一。目前国内的汉语自动标引研究主要集中于汉语自动分词这个前期处理问题上。提出了一种基于词平台的汉字编码方法，建立了一种新的中文计算机文档表达格式，使词成为最小的信息单位，汉语分析无需再进行自动分词，可直接进行自动标引，从而提高自动标引的效率和质量。

关键词　　自动标引　词平台　汉字编码　自动分词

分类号

Auto-indexing based on Chinese characters coding on words platform

JIAO Hui，LIU Qian，JIA Hui-bo

1.Department of Precision Instruments and Mechanology，Tsinghua University，Beijing 100084，China
2.State Key Laboratory of Precision Measurement Technology and Instruments，Beijing 100084，China

Abstract

Auto-indexing is one of the key techniques of information retrieval based on contents.At present the research on Chinese auto-indexing mainly focuses on automatic segmentation which is a predisposal problem.This paper presents a kind of Chinese characters coding method on words platform，and establishes a new Chinese text format in computer which makes words the smallest information unit.Based on this method，auto-indexing does not rely on segmentation as before.Thereby the efficiency and quality of auto-indexing would be improved.

Key words　 auto-indexing　 words platform　 Chinese characters coding　 automatic segmentation

DOI:

通讯作者　焦 慧 E-mail：jiaoh04@mails.tsinghua.edu.cn

扩展功能

本文信息

- Supporting info
- PDF(861KB)
- [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"自动标引"的相关文章
- 本文作者相关文章
- · 焦 慧
- · 刘 迁
- · 贾惠波