

20届机检会论文选登

BBS中文新词语自动挖掘*

吕学强¹;黄河¹;李渝勤¹;施水才^{1;2}

北京信息科技大学中文信息处理研究中心¹

收稿日期 2006-10-13 修回日期 2006-11-6 网络版发布日期 2007-1-24 接受日期

摘要 针对从BBS文本中自动挖掘新词语的问题,提出一种结合统计和规则的简单易行的方法,采用中文分词、频数统计、词性过滤、词语碎片组合等关键技术。据此方法开发的系统可以自动挖掘不限长度、不限领域、不限类别的与上下文无关的任意新词语。

Abstract A simple method using statistics and rule is presented for mining Chinese new words in BBS texts automatically, in which we use such key technologies as Chinese segmentation, frequency statistics, speech pattern filter and a series of operations on word fragments. A system developed in this method can mine random context-insensitive new words in any length and in any field, of any kind.

关键词 [自动挖掘](#) [新词语](#) [统计](#) [规则](#)

Key words Auto-mining; New word; Statistics; Rule

分类号 [TP391](#)

DOI:

通讯作者:

吕学强 lv.xueqiang@trs.com.cn; studystrong@sohu.com

作者个人主页: 吕学强 黄河 李渝勤 施水才

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF](#) (409KB)

▶ [\[HTML全文\]](#) (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“自动挖掘”的 相关文章](#)

▶ 本文作者相关文章

• [吕学强](#)

• [黄河](#)

• [李渝勤](#)

• [施水才](#)

•