

软件技术与数据库

自动提取含字母词语的领域新术语的研究

姜韶华, 党延忠

(大连理工大学系统工程研究所, 大连 116024)

收稿日期 修回日期 网络版发布日期 2007-1-8 接受日期

摘要 新术语的提取是中文信息处理领域的一个重要研究课题。针对现有提取方法的不足和很多专业术语表现为字母词语的特点, 该文提出了一种综合统计技术和规则筛选的方法: 基于长串优先和串频统计的思路进行文本切分, 得到共现字符串, 利用词语搭配规则进行过滤, 经过领域词典及评价函数的筛选, 提取出领域新术语。该方法可发现包含字母词语、专业术语等未登录词在内的频率大于等于2的任意长度的专指语义串、短语和词。实验表明了该方法的有效性及其新术语的准确率分布特征。

关键词 [专指语义串](#) [长串优先](#) [字母词语](#) [中文信息处理](#)

分类号

DOI:

对应的英文版文章: [2007-02-016](#)

通讯作者:

作者个人主页: 姜韶华; 党延忠

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF \(112KB\)](#)
- ▶ [\[HTML全文\] \(0KB\)](#)
- ▶ [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“专指语义串”的 相关文章](#)
- ▶ 本文作者相关文章
 - [姜韶华](#)
 - [党延忠](#)