

# 基于STC的中文文本聚类算法

王国强[1];郑海清[2];牛军钰[2]

( [1]上海市杨浦区业余大学, 上海200092; [2]复旦大学计算机科学与工程系, 上海200433 )

中图分类号: TP391

**摘要:** 提出了一种文档聚类方法, 对用户的检索结果中类似的文档进行聚类, 提供目录结构, 辅助用户浏览检索结果. 首先分析了现有的文本聚类方法, 讨论了它们的优势和不足, 然后提出了基于后缀树的中文文本聚类算法, 并详细描述了该算法的原理和构造使用过程, 及在算法实现的过程中遇到的关键问题及解决方案.

**关键词:** 后缀树;;文本聚类;;文本处理

 [阅读文章\(pdf\)](#)

关闭本页