现代图书情报技术 2007 (3): 43-45 ISSN: 1003-3513 CN: 11-2856/G2

知识组织与知识管理

文本分类C#实现*

刘华[1,2]
暨南大学华文学院/海外华语研究中心[1]

收稿日期 2007-1-29 修回日期 2007-2-12 网络版发布日期 2007-3-26 接受日期

摘要 设计并实现一个基于向量空间模型和简单贝叶斯的文本分类系统，系统采用层级多标签的分类策略。详细介绍词语切分统计、终分类器值计算、层级小类校正和兼类判断四个子系统模块。基于向量空间模型分类的第一级大类和层级小类的微平均分别为89.7%和77.8%，简单贝叶斯分别为67.6%和66.5%。

Abstract Based on Vector Space Model(VSM) and Na ve-Bayes(NB), completed a multilayer and multi-classification text categorization system. Introduce detailedly four modules: words' segmentation and frequency statistics, calculating between classifications' and document, emendating the veracity of parent-class by emendation of subclass, judging whether document has multi-classification and multi-label. Text representation based on Vector Space Model has 89.7% MicroF1 of parent- category, 77.8% of sub- category; text representation based on Na ve-Bayes has 67.6% MicroF1 of parent- category, 66.5% of sub- category.

关键词 文本分类 向量空间模型 简单贝叶斯

Key words Text categorization;Vector space model;Na&ive-Bayes

分类号 TP93

DOI:

通讯作者:
刘华 liuhua0461@sina.com;liuhua7586@blcu.edu.cn
作者个人主页: 刘华

扩展功能
本文信息
Supporting info
PDF(0KB)
[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“文本分类”的 相关文章
本文作者相关文章
• 刘华
•
•