

开发研究与设计技术

面向垂直搜索引擎的Web站点划分方案

李学凯, 许 笑, 孙春奇, 张伟哲, 李 斌

(哈尔滨工业大学计算机学院, 哈尔滨 150001)

收稿日期 修回日期 网络版发布日期 接受日期

摘要 分析传统搜索引擎分配任务的方式及存在的问题, 根据垂直搜索引擎的特点, 提出一种比传统方法粒度更细的任务分配方式——网站划分。该分配方式将较大规模的网站切分为若干较小规模的子集, 并将子集交给若干爬虫节点并行抓取, 以加快爬虫系统的整体获取速率, 作为对传统方法的有效优化。将网站划分算法应用于样本数据集, 验证其有效性。

关键词 [垂直搜索引擎](#); [任务分配](#); [网站划分](#); [爬虫](#)

分类号 [N945](#)

DOI:

通讯作者:

作者个人主页: [李学凯](#); [许 笑](#); [孙春奇](#); [张伟哲](#); [李 斌](#)

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF \(253KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“垂直搜索引擎; 任务分配; 网站划分; 爬虫” 的相关文章](#)

▶ [本文作者相关文章](#)