



[首页](#)
[本刊简介](#)
[编委会](#)
[收录情况](#)
[投稿指南](#)
[网上订阅](#)
[广告指南](#)
[兵工学会](#)
[联系我们](#)

文章详情

稿件标题: 基于Timed PageRank的聚焦爬虫优化研究

稿件作者: 李东 1, 王虎强 2

栏目名称: 基础理论与应用研究

关键词: 传统网络爬虫; PageRank算法; 聚焦爬虫; Timed PageRank改进算法

文章摘要: 传统的基于PageRank算法的网络爬虫在抓取网页时由于只考虑了网页的超链接, 势必会使爬虫结果覆盖面广、冗余度高, 聚焦爬虫由于其可以有效地过滤与主题无关的链接, 只保留有用的链接并将其加入到待抓取的URL队列, 因此能够有效地降低爬虫冗余; 在分析PageRank算法的基础上, 将网页的时间维数和页面的内容相关度融于其中, 提出了基于Timed PageRank的改进算法, 并将该算法应用于聚焦爬虫过程中, 实践证明该算法能够有效地提高爬虫页面相关度及检索结果的查全率和查准率。

收录刊物: 2015年01期

稿件基金:

引用本文格式: 李东, 王虎强. 基于Timed PageRank的聚焦爬虫优化研究 [J]. 四川兵工学报, 2015(1):141-144.
LI Dong, WANG Hu qiang. Optimization Research on Focused Crawler Based on Improved Timed PageRank Algorithm [J] Journal of Sichuan Ordnance, 2015(1):141-144.

浏览次数: 450

下载次数: 459

[免费下载全文](#)
[下载PDF阅读器](#)

地址: 重庆市巴南区红光大道69号 重庆理工大学明德楼6楼614室 邮编: 400054
 电话: 023-68852703 传真: 023-68852703 邮箱: bqzbgcxb@126.com
 您是第 **1833663** 位访问者
[前台管理](#) [工作入口](#)