

开发研究与设计技术

基于Linux的网络爬虫系统

王 锋, 王 伟, 张 璟, 罗作民

(西安理工大学计算机科学与工程学院, 西安 710048)

收稿日期 修回日期 网络版发布日期 接受日期

摘要 针对目前影响爬虫程序效率的诸多关键因素, 在研究爬虫程序内部运行机理的基础上, 进行架构优化, 改进爬虫程序中的相关算法。在Linux网络环境下, 通过对实现的爬虫程序运行进行检测, 反馈出该解决方案和改进之处具有可行性, 提高了页面抓取的效率和爬虫程序的整体性能。

关键词 [网络爬虫](#); [URL调度](#); [DNS解析](#); [哈希算法](#)

分类号 [N945.15](#)

DOI:

通讯作者:

作者个人主页: [王 锋](#); [王 伟](#); [张 璟](#); [罗作民](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF](#) (320KB)
- ▶ [\[HTML全文\]](#) (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“网络爬虫; URL调度; DNS解析; 哈希算法”的 相关文章](#)
- ▶ [本文作者相关文章](#)