

网络与通信(Network and communications)

网络爬虫效率瓶颈的分析与解决方案

尹江¹; 尹治本¹; 黄洪¹

西南交通大学¹

收稿日期 2007-11-12 修回日期 2008-1-7 网络版发布日期 2008-5-7 接受日期

摘要 网络爬虫的效率, 直接关系到搜索引擎系统为用户提供的服务质量。如何设计高效、快速的网络爬虫, 成为目前网络爬虫研究的热点。要提高网络爬虫的爬行效率, 除了需要改进网络爬虫的爬行策略之外, 还需要优化网络爬虫自身的设计, 改进网络爬虫自身的结构, 消除效率瓶颈。通过对网络爬虫结构、应用环境以及用户要求的分析, 提出一个通用网络爬虫的改进设计方案, 并通过实验得到较好的测试结果。

关键词 [爬行策略](#) [套接字](#) [多线程](#) [网络爬虫](#)

分类号

DOI:

对应的英文版文章: [A7116180](#)

通讯作者:

尹江 j_yeen@163.com; johnny_yeen@yahoo.com.cn

作者个人主页: 尹江 尹治本 黄洪

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF \(959KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“爬行策略”的 相关文章](#)

▶ 本文作者相关文章

· [尹江](#)

· [尹治本](#)

· [黄洪](#)