

- >> 首页
- >> 被收录信息
- >> 投稿须知
- >> 模板下载
- >> 信息发布
- >> 常见问题及解答
- >> 合作单位
- >> 产品介绍
- >> 编委会/董事会
- >> 关于我们
- >> 网上订阅
- >> 友情链接

#### 友情链接

- >> 中国期刊网
- >> 万方数据资源库
- >> 台湾中文电子期刊
- >> 四川省计算应用研究中心
- >> 维普资讯网

## 基于GNP算法的分布式爬虫调度策略\*

GNP-based scheduling strategy for distributed crawling

摘要点击: 30 全文下载: 13

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

中文关键词: [分布式爬虫](#) [任务调度](#) [负载均衡](#) [网络测量](#) [全局网络定位](#)

英文关键词: [distributed crawling](#) [scheduling strategies](#) [load balancing](#) [network measurement](#) [GNP\(global network positioning\)](#)

基金项目: 国家“973”重点基础研究发展计划资助项目(G2005CB321806); 国家自然科学基金资助项目(60703014); 高等学校博士学科点专项科研基金资助课题(20070213044); 哈尔滨工业大学优秀青年教师培养计划(HITQNJ.S.2007.034)

作者

单位

[刘爽<sup>1</sup>](#), [姜春祥<sup>2</sup>](#), [张伟哲<sup>1</sup>](#) [\(1.哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001; 2.国家计算机网络应急技术处理协调中心 黑龙江分中心, 哈尔滨 150001; 3.国家计算机网络应急技术处理协调中心, 北京 100029\)](#)

中文摘要:

针对分布式搜索引擎的任务调度及负载均衡问题, 提出了基于GNP算法的分布式爬虫调度策略和负载均衡的方法。利用网络距离预估取代大规模的网络距离测量, 不仅提高了系统的响应速度, 还减少了系统对广域网造成的压力。通过在广域网上部署爬虫节点, 构建分布式搜索引擎, 应用该调度策略进行实验, 验证了系统性能有较大提高。

英文摘要:

In order to solve task scheduling and load balancing problems of distributed search engines, this paper proposed a GNP-based scheduling strategy for distributed crawling and a load balancing method. Adopted internet distance estimating mechanism as a replacement for large-scale network distance measurement, which not only improved response time of the system, but also reduced WAN pressure caused by the system. Through deploying crawling nodes at WANs, built a distributed search engine, and implemented several scheduling strategies. The online experiment shows great improvement in system's performance.

您是第2827010位访问者

主办单位: 四川省计算机研究院 单位地址: 成都市武侯区成科西路3号

服务热线: 028-85249567 传真: 028-85210177 邮编: 610041 Email: arocmag@163.com

蜀ICP备05005319号 本系统由北京勤云科技发展有限公司设计