

贺 炜,郭云飞,扈红超.基于状态约束的大规模正则表达式匹配算法[J].通信学报,2013,(10):183~190

基于状态约束的大规模正则表达式匹配算法

States constrain-based algorithm for largescale regular expression matching

投稿时间: 2012-07-15

DOI: 10.3969/j.issn.1000-436x.2013.10.021

中文关键词: [正则表达式](#) [状态爆炸](#) [自动机转化](#) [状态约束](#)

英文关键词: [regular expression](#) [state explosion](#) [automata convert](#) [state constrains](#)

基金项目: 国家科技支撑计划基金资助项目 (2012BAH02B03, 2012BAH02B01); 国家高技术研究发展计划 (“863” 计划) 基金资助项目 (2011AA01A103, 2011AA01A101, 2011BAH19B04)

作者

单位

[贺 炜, 郭云飞, 扈红超](#)

[国家数字交换系统工程技术研究中心, 河南 郑州 450002](#)

摘要点击次数: 194

全文下载次数: 63

中文摘要:

通过观察不确定有限自动机NFA到确定性有限自动机DFA的转化过程, 分析内存增长的原因, 提出了一种基于状态间约束关系的正则表达式匹配算法Group2-DFA。Group2-DFA通过两级分组, 利用状态间的约束关系, 将原始NFA转化为NFA和DFA的混合结构。实验表明, 在保持一定处理速率的前提下, Group2-DFA能够有效地减少内存占用。在300条规则下, Group2-DFA吞吐率能够达到1Gbps, 并且减少约75%的状态数。

英文摘要:

By analysis of state explosion in deterministic finite automata DFA, a novel algorithm Group2-DFA based on state constrains was proposed to reduce the memory usage. With the state constrains, states in NFA were classified into several groups. Group2-DFA introduces two-level classification and merges NFA and DFA together to a hybrid FA construction. The experiments show that Group2-DFA can reduce memory usage efficiently and keep high throughput with a small increase of memory reading time. With 300 regex rules, Group2-DFA can cut 75% states and achieve 1Gbps throughput.

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

关闭

版权所有: 《通信学报》

地址: 北京市丰台区成寿寺路11号邮电出版大厦8层814室 电话: 010-81055478, 81055479
81055480, 81055482 电子邮件: xuebao@ptpress.com.cn

技术支持: 北京勤云科技发展有限公司