

软件技术与数据库

基于SPAM-FPT的WebLog访问序列模式挖掘

朱 莉1, 应吉康1, 卜忠飞2

(1. 华东师范大学信息学院计算中心, 上海 200062; 2. 扬州市教育局, 扬州 225002)

收稿日期 修回日期 网络版发布日期 2007-9-4 接受日期

摘要

WebLog访问序列模式挖掘将数据挖掘中的序列模式技术应用于Web服务器上的日志文件, 以此来改善Web的信息服务, 而在对海量的数据挖掘时, 系统资源开销很大。该文结合SPAM、PrefixSpan的思想, 提出一个新的算法——SPAM-FPT, 该算法通过建立First_Positon_Table, 避免了SPAM中的“与操作”、“连接操作”以及PrefixSpan中大量的“投影数据库”的建立, 可以快捷地挖掘数据库中所有“频繁子序列”。

关键词 [序列模式挖掘](#) [WebLog](#) [频繁子序列](#) [SPAM-FPT](#)

分类号 [TP391](#)

DOI:

通讯作者:

作者个人主页: [朱 莉1](#); [应吉康1](#); [卜忠飞2](#)

扩展功能
本文信息
▶ Supporting info
▶ PDF (362KB)
▶ [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“序列模式挖掘”的相关文章
▶ 本文作者相关文章
• 朱 莉1, 应吉康1, 卜忠飞2