

论文

共享片段自动检测模型

石磊¹,程刚运²

- 1. 郑州大学南校区信息工程学院
- 2. 郑州大学信息工程学院

摘要:

提出了一种可扩展的共享片段自动检测算法,介绍了共享片段自动检测模型的结构和实现方法。该模型可在线运行自动检测算法,将流行度高的网页集合作为检测对象。实验和分析表明,该模型在保证准确率的前提下,能检测出网页集合的最大共享片段集合,提高检测效率。

关键词: 网页 片段 检测模型 Web page fragment detection model

Automatic detection of shared fragments model

Abstract:

A model of automatic online detection of shared fragments was described including its architecture and implementation schemes. Besides, a scalable algorithm for detecting shared fragments was proposed. The online detection algorithm could run automatically in the model. The model took the Web pages that were more popular than others as detection objects, and could get a set of maximal shared fragments from a large collection of Web pages. Experimental results show that it not only provides veracity of detection but also improves the efficiency.

Keywords:

收稿日期 2008-11-11 修回日期 2008-12-31 网络版发布日期 2009-06-09

DOI:

基金项目:

国家自然科学基金(编号60472044);河南省信息网络重点实验室开放基金项目(2006);无

通讯作者:程刚运

作者简介:

参考文献:

本刊中的类似文章

- 1. 郭晨娟 李战怀.基于概念的网页相似度处理算法研究[J]. 计算机应用, 2006,26(12): 3030-3032
- 2. 丛磊.Pagerank二维线性收敛方法[J]. 计算机应用, 2006,26(6): 1415-1417
- 3. 邓健爽; 郑启伦; 彭宏.基于提取网站层次结构的网页分类方法[J]. 计算机应用, 2006,26(5): 1134-1136
- 4. 王婷 陈性元 张斌 包义保 夏春涛.基于GAA-API的Web网页细粒度访问控制方法研究[J]. 计算机应用, 2007,27(5): 1274-1276
- 5. 雷阳 华继学 李成海.基于线性加权法的远程网络评估方法[J]. 计算机应用, 2007,27(10): 2420-2422
- 6. 魏丽霞 郑家恒.基于网页文本结构的网页去重[J]. 计算机应用, 2007,27(11): 2854-2856

扩展功能

本文信息

- Supporting info
- PDF(584KB)
- [HTML全文]
- 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

本文关键词相关文章

- 网页
- 片段
- 检测模型
- Web page
- fragment
- detection model

本文作者相关文章

- 石磊
- 程刚运

PubMed

- Article by Dan,I
- Article by Cheng,G.Y

7. 魏文国 谢桂园. 自适应最优搜索算法的网络蜘蛛的设计与实现[J]. 计算机应用, 2007,27(11): 2857-2859

8. 陈锦言 孙济洲 张亚平. 基于傅立叶变换的网页去重算法[J]. 计算机应用, 2008,28(4): 948-950

文章评论 (请注意: 本站实行文责自负, 请不要发表与学术无关的内容! 评论内容不代表本站观点.)

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text"/> 7078

Copyright 2008 by 计算机应用