

- >> 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

基于启发式规则的网页主题信息精确定位方法*

Approach of pinpointing subject information in Web pages based on heuristic rules

摘要点击: 39 全文下载: 14

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

中文关键词: [启发式规则](#) [信息抽取](#) [主题信息定位](#) [模板化网页](#)

英文关键词: [heuristic rules](#) [information extraction](#) [subject information localization](#) [template Web pages](#)

基金项目: 国家教育部人文社会科学重点研究基地重大项目 (07JJD740063); 湖北省科技攻关项目 (2007AA101C49)

作者

[胡金柱](#), [周星](#), [舒江波](#), [熊春秀](#)

单位

[\(华中师范大学 计算机科学系, 武汉 430079\)](#)

中文摘要:

目前大部分的信息抽取方法都是针对主题信息块的提取, 没有进一步深入到各个单独主题信息的抽取。针对这一问题, 提出了一种基于启发式规则的网页主题信息精确定位方法。首先针对各个单独的主题, 分析其多方面的特征, 制定出对应的启发式规则; 然后利用不同的规则对定位主题重要度不同的这一特点, 得到启发式规则的权值矩阵; 最后利用基于启发式规则的定位算法精确定位各个主题。将该方法用于网页主题信息抽取系统中, 抽取系统能够有效地对各个单独的主题进行定位和抽取。实验结果表明, 该方法具有很好的有效性和准确性。

英文摘要:

At present, most of information extraction methods aim at the extraction of subject information block, not further penetrate into the extraction of each independent subject information. To solve this problem, this article proposed an approach of pinpointing subject information in Web pages based on heuristic rules. Firstly, for each independent subject, it analyzed its various characteristic, and formulated corresponding heuristic rules. Then, it obtained weight matrix of heuristic rules by using the feature that different rules had different importance to locate subject. Finally, according to localization algorithm of heuristic rules, it pinpointed each subject. The method has been applied to an automatic extraction system, and the experimental result shows the effectiveness and accuracy of the method.

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