

数据挖掘

一种基于XQuery的优化Web信息抽取方法

支宗良¹;陈少飞²

河北省烟草专卖局¹

收稿日期 2007-8-10 修回日期 2007-10-11 网络版发布日期 2008-1-1 接受日期

摘要 由于缺乏对页面特征适应性的分析, 现有的典型系统难以保障抽取规则的健壮性。提出一种优化的Web信息抽取方法, 该方法引入了相互关联的三层规则, 在分析页面特征适应性的基础上, 从准确率和召回率两方面出发提出了抽取规则的优化算法, 并用标准XQuery表达复杂对象抽取规则。实验证明, 该方法有效地增强了抽取规则的健壮性及可用性。

关键词 [信息抽取](#) [XPath](#) [XQuery](#) [规则优化](#)

分类号

DOI:

对应的英文版文章: [A7074276](#)

通讯作者:

陈少飞 chenshaofei@yaho.com.cn, chen_shaof@sohu.com

作者个人主页: 支宗良 陈少飞

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF \(611KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“信息抽取”的 相关文章](#)
- ▶ 本文作者相关文章
 - [支宗良](#)
 - [陈少飞](#)