

博士论文

基于隐马尔可夫模型的Web信息抽取

刘亚清, 陈 荣

(大连海事大学信息科学技术学院, 大连 116026)

收稿日期 修回日期 网络版发布日期 接受日期

摘要 针对Web信息抽取领域中存在的“项缺失”和“项无序”问题, 提出一种基于隐马尔可夫模型的Web信息抽取方法。将Web文档解析为一棵扩展的DOM树, 映射待抽取的信息项为状态, 映射待抽取的信息项在扩展DOM树中的路径为词汇, 使用归纳算法构造隐马尔可夫模型。实验结果证明该方法可以获得更好的抽取性能。

关键词 [信息抽取; 隐马尔可夫模型; 扩展DOM树](#)

分类号 [TP393](#)

DOI:

通讯作者:

作者个人主页:

刘亚清;陈 荣

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(175KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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; 隐马尔可夫模型; 扩展DOM树”的 相关文章](#)

▶ [本文作者相关文章](#)