

数据库与信息处理

基于启发式查询词选择算法的Hidden Web获取研究

姚全珠, 杨增辉, 张楠, 田元

西安理工大学 计算机学院, 西安 710048

收稿日期 修回日期 网络版发布日期 2007-11-19 接受日期

摘要 Hidden Web因为其隐蔽性而难以直接抓取, 因此成为信息检索研究的一个新领域。提出了一种获取Hidden Web信息的方法, 讨论了实现的关键技术。通过设计提出的启发式查询词选择算法, 提高了抓取的效率。实验证明了该模型和算法的有效性。

关键词 [信息检索](#) [Hidden Web](#) [爬虫](#) [启发式算法](#)

分类号

Research on crawling Hidden Web based on heuristic query selection algorithm

YAO Quan-zhu, YANG Zeng-hui, ZHANG Nan, TIAN Yuan

School of Computer Science, Xi'an University of Technology, Xi'an 710048, China

Abstract

Because of the hidden feature, Hidden Web is hard to crawl. It becomes a new direction in the field of information retrieval. In this paper a new method of Hidden Web information retrieval is proposed. It presents a generic operational model of the Hidden Web information retrieval and describes the key techniques. It introduces a new heuristic query selection algorithm which designed by this paper. Based on this technique, the crawling is more efficient. Experiments show the effectiveness of both the model and the algorithm.

Key words [information retrieval](#) [Hidden Web](#) [crawler](#) [heuristic algorithm](#)

DOI:

通讯作者 姚全珠 autumn_stone@126.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(603KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ 本刊中 [包含“信息检索”的相关文章](#)

▶ 本文作者相关文章

· [姚全珠](#)

· [杨增辉](#)

· [张楠](#)

· [田元](#)