

{tag}

{/tag}

IJCA Proceedings on International Conference
on Emerging Trends in Computing and Communication

© 2018 by IJCA Journal

ICETCC 2017 - Number 3

Year of Publication: 2018

Authors:

Rohini Navnathkhedkar

Madhuri Dalal

{bibtex}icetcc129.bib{/bibtex}

Abstract

As deep web grows at a very fast pace, there has been increased interest in techniques that help efficiently locate deep-web interfaces. However, due to the large volume of web resources and the dynamic nature of deep web, achieving wide coverage and high efficiency is a challenging issue. We propose a two-stage framework, for harvesting deep web interfaces. In the first stage of harvesting, performs site-based searching for center pages with the help of search engines, avoiding visiting a large number of pages. To achieve more accurate results for a focused crawl ranks websites to prioritize highly relevant ones for a given topic. In the second stage, it achieves fast in-site searching by excavating most relevant links with an

adaptive link-ranking.

Refer

ences

- Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang, Hai Jin "SmartCrawler: A Two Stage Crawler for efficiently harvesting Deep-Web interfaces"; IEEE Transactions on Services Computing Volume: 99 PP Year: 2015.
- L. Barbosa and J. Freire, "An adaptive crawler for locating hidden web entry points," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 441–450.
- . Olston and M. Najork , "Web Crawling";, Foundations and Trends in Information Retrieval, vol. 4, No. 3 ,pp. 175–246, 20.
- Y. He, D. Xin, V. Ganti, S. Rajaraman, and N. Shah, "Crawling deep web entity pages," in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 355–364.
- Barbosa and J. Freire, "Searching for hidden-web databases,";in Proc. 8th Int. Workshop Web Databases, 2005, pp. 1–6.
- Rabia and Sami, Lalitha K. , "Understanding the Deep Web"; (2010). Library Philosophy and Practice (e-journal). Paper 364. <http://digitalcommons.unl.edu/libphilprac>.

Index Terms

Computer Science

Information Sciences

Keywords

Deep Web Ranking Adaptive Learning Two-stage Crawler.