

Manuscript ID : 00000-60444

International Journal of Computer Engineering and Technology

Volume 4, Issue 4, July-August 2013, Pages 163-169, Page Count - 7



Source ID : 00000005

## PRIORITY BASED FOCUSED WEB CRAWLER

Jaytrilok Choudhary <sup>(1)</sup> Devshri Roy <sup>(2)</sup>

<sup>(1)</sup> Department of Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India.

<sup>(2)</sup> Department of Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India.

### Abstract

*The World Wide Web (WWW) has billions of documents and these documents are attached to each other using hyperlinks. Web crawler is a heart of Search engine that gathers these documents from WWW. Maximum documents present on WWW are dynamic and changes periodically. Hence, Crawler needs to refresh these documents to update database of search engine. In this paper, we have proposed a priority based focused web crawling algorithm. The web pages corresponding to URL (Uniform Resource Locator) are downloaded from web and calculated the relativity score of downloaded page with focus word. We store URL and its relativity score with focus word in priority queue instead of normal queue. So, every time priority queue returns maximum Score URL to crawl next. The overall performance gain over simple crawler is 87% and over focused crawling is 24%.*

### Author Keywords

Priority, focus word, web pages, downloader, search engine

### Index Keywords

World Wide Web, Web crawler, Data structure, Crawling algorithm

**ISSN Print:** 0976-6367

**Source Type:** Journals

**Publication Language:** English

**Abbreviated Journal Title:** IJCET

**Publisher Name:** IAEME Publication

**Major Subject:** Physical Sciences

**Subject area:** Software Engineering

**ISSN Online:** 0976-6375

**Document Type:** Journal Article

**DOI:**

**Access Type:** Open Access

**Resource Licence:** CC BY-NC

**Subject Area classification:** Computer Science

**Source:** SCOPEDATABASE

### Reference

#### References (11)

1. N. Singhal, A. Dixit, and A.K. Sharma  
Design of a Priority Based Frequency Regulated Incremental Crawler

(2010) *International Journal of Computer Applications*, Volume 1, Issue 1, Page No 42-47,

- 
2. Ah Chung Tsoi, et al  
A Simple Focused Crawler  
  
(2003) WWW2003 ACM,  
Article Link: <https://cs.brynmawr.edu/Courses/cs380/fall2006/www12DanielePoster.pdf>
  3. S.Ganesh, M.Jayaraj, V.Kalyan, S. Murthy, and G.Aghila  
Ontology-based Web crawler  
  
(2004) *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004*,  
DOI: <https://doi.org/10.1109/ITCC.2004.1286658>  
Article Link: <https://ieeexplore.ieee.org/document/1286658>
  4. D. Mukhopadhyay, A. Biswas, and S. Sinha  
A New Approach to Design Domain Specific Ontology Based Web Crawler  
  
(2010) *Proceedings of 10th IEEE International conference on Information Technology*,
  5. X. Chen, and X. Zhang  
HAWK: A Focused Crawler with Content and Link Analysis  
  
(2008) *2008 IEEE International Conference on e-Business Engineering*,  
DOI: <https://doi.org/10.1109/ICEBE.2008.46>  
Article Link: <https://ieeexplore.ieee.org/document/4690687>
  6. D. Hati, B. Sahoo, and A Kumar  
Adaptive Focused Crawling Based on Link Analysis  
  
(2010) *2010 2nd International Conference on Education Technology and Computer*,  
DOI: <https://doi.org/10.1109/ICETC.2010.5529641>  
Article Link: <https://ieeexplore.ieee.org/abstract/document/5529641>
  7. S.Thenmalar, and T. V. Geetha  
Concept based Focused Crawling using Ontology  
  
(2011) *International Journal of Computer Applications, Volume 26, Issue 7, Page No 29-32*,
  8. S. Chakrabarti, M. van den Berg, and B. Dom  
Focused crawling: a new approach to topic-specific Web resource discovery  
  
(1999) *Proceedings of the 8th International WWW Conference*,
  9. A. Suganthi, G.S.Sumithra, J.Hindusha, A.Gayathri and S.Girija  
Semantic Web Services and its Challenges  
  
(2010) *International Journal of Computer Engineering and Technology, Volume 1, Issue 2, Page No 26 - 37*,  
Article Link: [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJCET/VOLUME\\_1\\_ISSUE\\_2/IJCET\\_01\\_02\\_003.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_1_ISSUE_2/IJCET_01_02_003.pdf)
  10. Alamelu Mangai J, Santhosh Kumar V and Sugumaran V  
Recent Research in Web Page Classification - A Review  
  
(2010) *International Journal of Computer Engineering and Technology, Volume 1, Issue 1, Page No 112 - 122*,  
Article Link: [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJCET/VOLUME\\_1\\_ISSUE\\_1/IJCET\\_01\\_01\\_011.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_1_ISSUE_1/IJCET_01_01_011.pdf)
  11. Houda El Bouhissi, Mimoun Malki and Djamila Berramdane  
Applying Semantic Web Services

**Scope Database Link:** <https://sdbindex.com/documents/00000005/00000-60444.pdf>

**Article Link:** [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJCET/VOLUME\\_4\\_ISSUE\\_4/IJCET\\_04\\_04\\_015.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_4_ISSUE_4/IJCET_04_04_015.pdf)

*(2013) International Journal of Computer Engineering and Technology, Volume 4, Issue 2, Page No 108 - 113,*

*Article Link:* [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJCET/VOLUME\\_4\\_ISSUE\\_2/IJCET\\_04\\_02\\_012.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_4_ISSUE_2/IJCET_04_02_012.pdf)

---

---

## About Scope Database

What is Scope Database

Content Coverage Guide

Scope Database Blog

Content Coverage API

Scope Database App

© Copyright 2021 Scope Database, All rights reserved.

## Customer Service

Help

Scope Database Key Persons

Contact us