Abstract

The world wide web is a system of interlinked hypertext documents contained on the Internet, these web page may contain text, images, videos, and other multimedia and navigate between them by using hyperlink. About million new pages go online each day. It is impossible for major search engines to update their collections to meet such rapid growth. Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents. To address the above problems, domain-specific search engines were introduced, which keep their Web collections for one or several related domains. Focused crawlers were used by the domain-specific search engines to selectively retrieve Web pages relevant to particular domains to build special Web collections, which have smaller size and provide search results with high precision. In this paper, we are introducing a global focused crawling approach which is beneficial in extracting more relevant data.
References

- Jin Xu Yingping Huang Gregory Madey ?A RESEARCH SUPPORT SYSTEM FRAMEWORK FOR WEB DATA MINING?.
- Arvind Arasu,Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke. Sriram Raghavan &quot;Searching The Web&quot;,. Computer Science Department, Stanford University.
- Ying Wang, Wanli Zuo, Tao Peng, Fengling He &quot;Domain-Specific Deep Web Sources Discovery&quot; Fourth International Conference on Natural Computation.
- Zhaoqiong Gao, Yajun Du, Liangzhong Yi, Qiangqiang Peng, Yuekui Yan &quot;Incrementally Updating Concept Context Graph(CCG) for Focused Web Crawling Based on FCA&quot;; 2009.

Index Terms

Computer Science

Emerging Trends in Technology

Keywords
Www  Web Mining  Crawling  Focused Crawling.