Abstract

The World-Wide Web is a huge network of billions of workstations and this network contains billions of web pages containing information on a wide variety of topics. There are a lot of topics discussed by people, opinions and suggestions shared on various social networking sites that the users are interested in. Low precision and low recall still exists in the current search engines. So a search engine that is effective and one that applies Web mining technology has become very important. A discussion on the various technologies used to implement a search engine and its techniques like indexing and searching on the world wide web is done in this paper. The authors propose to describe the method to create a search engine by using JSoup and Apache Lucene API in the paper.

References

2014, p. 1150, American Medical Informatics Association.

one online healthcare forums using hidden markov model,” BMC medical informatics and decision
making, vol. 14, no. 1, p. 91.


for xml,” in Proceedings of the 29th international conference on Very large data bases-Volume
29, pp. 45–56, VLDB Endowment.


search on internet tabular information extraction for answering queries,” in Proceedings of the
ninth international conference on Information and knowledge management, pp. 243–249, ACM.

searching, reusing and evolving distributed ontologies,” in Proceedings of the 12th international
conference on World Wide Web, pp. 439–448, ACM.

accessibility of web”.


Index Terms

Computer Science Information Sciences

Keywords

Web, crawler, searching, indexing, JSoup, Apache Lucene.