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


Index Terms

Computer Science      Information Sciences

Keywords

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