A Model for Personalized Keyword Extraction from Web Pages using Segmentation

Abstract

The World Wide Web caters to the needs of billions of users in heterogeneous groups. Each user accessing the World Wide Web might have his/her own specific interest and would expect the web to respond to the specific requirements. The process of making the web to react in a customized manner is achieved through personalization. This paper proposes a novel model for extracting keywords from a web page with personalization being incorporated into it. The keyword extraction problem is approached with the help of web page segmentation which facilitates in making the problem simpler and solving it effectively. The proposed model is implemented as a prototype and the experiments conducted on it empirically validate the model’s efficiency.

References

Intelligence, 114(1-2), 257–281.
- Ping-I Chen, Shi-Jen Lin, Word AdHoc Network: Using Google Core Distance to extract the most relevant information, Knowledge-Based Systems, Volume 24, Issue 3, April 2011, Pages 393-405, ISSN 0950-7051.
- Yahoo! Content Analysis Service http://developer. yahoo. com/search/content/V2/contentAnalysis. html

Index Terms
Computer Science Information Retrieval

Keywords
Keyword Extraction  Web Page Segmentation  Web Personalization