Abstract

Over the last decade, clichéd information age has justly arrived. Moreover, the evolution of the Internet into the Global Information Infrastructure, together with the massive popularity of the Web, has also enabled the ordinary citizen to become not just a consumer of information, but also a part of it. In order to make user trouble free, it is required to save his/her time and effort. So some way is needed to give the relevant information to the user in a quick way and also enables to manage the whole lot of data without troublesome. Through this paper, the authors have used tf-idf (term frequency inverse document frequency approach) technique along with the concept of web mining to attain the required solution. Web mining is the application of data mining techniques that aims in discovering the patterns from the Web. Among its different ways, like Web usage mining, Web content mining and Web structure mining, here, efforts are only being made in the field of web content mining. In this work, a windows application is developed which act as a data analysis tool. This application is using the API of Bing search engine. The proposed algorithm is applied on the snippets (short description provided below each search result) of web search results to find those web pages that contains maximum number of query words. Moreover, it also aims at managing the information more easily on client's machine by using simple grouping technique.
References

- Bamshad Mobasher, "Data Mining for Web Personalization", Center for Web Intelligence School of Computer Science, Telecommunication, and Information Systems DePaul University, Chicago, Illinois, USA
- M. Angelaccio, B. Buttarazzi, M. Patrignanelli, 2007, "Graph Use to Visualize Web Search Results: MyWish 3.0", 11th International Conference Information Visualization (IV’07), © 2007IEEE
- Raymond Kosala, Hendrik Blockeel, 2000, "Web Mining Research: A Survey".
In ACM SIGKDD
- Y. Wang, M. Kitsuregawa, "Link-based Clustering of Web Search Results"; In Proceedings of The Second International Conference on Web-Age Information Management.

**Index Terms**

Computer Science

Information Sciences

**Keywords**

Term frequency-inverse document frequency (tf-idf)  static clustering  Mining methods and algorithms

Information Retrieval