World Wide Web is considered the most valuable place for Information Retrieval and Knowledge Discovery. While retrieving information through user queries, a search engine results in a large and unmanageable collection of documents. Web mining tools are used to classify, cluster and order the documents so that users can easily navigate through the search results and find the desired information content. A more efficient way to organize the documents can be a combination of clustering and ranking, where clustering can group the documents and ranking can be applied for ordering the pages within each cluster. Based on this approach, in this paper, a mechanism is being proposed that provides ordered results in the form of clusters in accordance with user’s query. An efficient page ranking method is also proposed that orders the results according to both the relevancy and the importance of documents. This approach helps user to restrict his search to some top documents in particular clusters of his interest.
Reference

- Jaroslav Pokorny, Jozef Smizansky, “Page Content Rank: An approach to the Web Content Mining”.
- http://pr.efactory.de/e-pagerank-algorithm.shtml
Index Terms

Computer Science  Information Retrieval

Key words

Document Clustering

PageRank

Web Mining

Weighted PageRank

World Wide Web