Abstract

The crawled web pages should be organized in a fashion where they are more understandable to machine, for producing the results which are meaningful and relevant. The set of web pages can be categorized into different contextual sense if the crawler has the technique to understand their meaning and the domain identification. The contextual relevance of the web documents can be known, if the frequent occurring patterns of the keywords in the web page are identified. This can be achieved through data mining technique for generating frequent patterns, using FP-Growth. It will help in deducing the set of keywords of the documents and this knowledge is added in the knowledge store which will further facilitate in the building the ontology for the crawled web pages and organizing them and thus increasing the rank of the document.
Extraction of Contextual Relevance of the Web Document using F-P Growth


- Jiawei Han and Micheline Kamber, ”Data Mining Concepts and Techniques”; Elsevier publication, Second edition, 2007.


Index Terms

Computer Science

Information Systems

Keywords

Context ontology frequent patterns relevance FP-growth