Abstract

With the emergence of the web, lots of research efforts are made in the area of Web Mining. This paper proposes an automatic approach for automatic topic identification from the web pages. The contribution of this research is in the approach of automatic topic identification of web pages that can provide better results. The topic of the web documents is identified through ontological approach. Keywords are extracted from the basic HTML tags and co-occurrence of words in the text instead of calculating the frequency of each term exits in a web page. Domain ontology is developed to map topics of the documents. Keywords are mapped to the ontology with a Levenshtein Edit Distance to extract topic of the web page. The result could give benefit to the search engines for faster tagging of web pages.

References

Conference on Artificial Intelligence.
- Jain, S. and Pareek, J. 2010. Automatic Topic(s) Identification from Learning Material:

**Index Terms**

- Computer Science
- Information Sciences

**Keywords**

- DOM
- Word co-occurrence
- Ontology
- Topic Identification.