Abstract

Now a day's classification of document is an important area for research, as large amount of electronic documents are available in form of unstructured, semi structured and structured information. Document classification will be applicable for World Wide Web, electronic book sites, online forums, electronic mails, online blogs, digital libraries and online government repositories. So it is necessary to organize the information and proper categorization and knowledge discovery is also important. This paper focused on the existing literature and explored the techniques for automatic documents classification i.e. documents representation, knowledge extraction and classification. In this paper author propose an algorithm and architecture for automatic document collection.

References

- S. Gopal, Y. Yang. Multilabel classification with meta-level features. ACM SIGIR

**Index Terms**

Computer Science
Information Sciences

**Keywords**

Text mining  Web mining  Automatic document classification