Abstract

Information retrieval is a science of gathering information from unstructured data, the online information source i. e., www. WWW contains data of heterogeneous types and of high dimension. Retrieving information from such database is a tedious work. Many researches are going on, to find a best optimal solution. A search engine is the tool for retrieving information from www. The internet helps the user to get the required information from www. A search engine respond to the user-need by answering their query, contains: Database, Web crawler, and Ranking algorithm. The optimality of the search engine is based on the ranking algorithm. The rank list is prepared based on the relevancy score. In this work we propose to use a novel algorithm, Multi-type Feature Co-selection for Clustering (MFCC) to the search engine as an alternative for the ranking algorithm. MFCC has proved its efficiency in clustering the heterogeneous web documentation.

References

Index Terms

Computer Science

Algorithms

Keywords

Information Retrieval  Search Engine  MFCC algorithm  Feature selection