Abstract

Due to increasing size of text and audio data over internet, various techniques are needed to help with the finding and extraction of very specific information relevant to a user’s task. Text mining is a variant on a field called data mining that tries to discover curious patterns from large databases. Singular value decomposition this technique is used for dimensionality reduction of large database on Apache MAHOUT Hadoop framework. In this paper, different existing Information Retrieval, Pattern Matching, Rule Generation Algorithm is reviewed. In addition for extracting questionnaires and curiosity based sentences from large database some different implementation of the algorithms is proposed. Finally the extract user required information from large unstructured database.

References

1. Mrs. Sayantani Ghosh, Mr. Sudipta Roy, and Prof. Samir K. Bandyopadhyay, 2012 A tutorial review on Text Mining Algorithms.
2. Che-Rung Lee, Ya-Fang Chang, 2013 Enhancing Accuracy and Performance of Collaborative Filtering Algorithm by Stochastic SVD and Its MapReduce Implementation
5. Haralampos Karanikas and Babis Theodoulidis Manchester, 2001 Knowledge Discovery in Text and Text Mining Software”, Centre for Research in.

Index Terms

Computer Science Information Sciences

Keywords

Apache MAHOUT; hadoop; information retrieval; pattern matching; rule generation.