Abstract

Web documents contain information that include image, video. The retrieved information are oriented towards the user's input for search. But the popular search engines like google uses algorithms for prioritizing and retrieving the results. However the information accessed also depends on the cookies as part of the local system. In real requirements based on the category, type and interest of users it is required to provide contents. This paper has a complete survey of different approaches/algorithms that are in existence for analyzing the web documents and providing to the user. The paper has a complete study of the different performance parameters that can be used for analyzing the results. The paper also draws
conclusions for selecting the appropriate approach based on different scenarios and different user input criteria provided by users.

References

- Skabar Andrew, Abdalgader, Khaled, "Clustering Sentence-Level Text Using a Novel Fuzzy Relational Clustering Algorithm"; IEEE Transactions on Knowledge and Data Engineering, Vol. 25, PP. 62-75, 2013.
- Cui Hang, Wen Ji-Rong, Nie Jian-Yun Y., Ma Wei-Ying Y., "Query expansion by mining user logs"; IEEE Transactions on Knowledge and Data Engineering, Vol. 15, PP. 829 - 839, 2003.
- Bollegala D., Weir D., Carroll J., "Cross-Domain Sentiment Classification using a
Sentiment Sensitive Thesaurus, IEEE Transactions on Knowledge and Data Engineering, 2012

- Andrea Paola Barraza, Angela Carrillo-Ramos, "Basic requirements to keep in mind for an ideal agent-based web information retrieval system in ubiquitous environments", Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services (IIWAS'10), 2010.

Index Terms
Computer Science
Web Based Application

Keywords
Data Model  Cluster  Threshold  Filtering  Classification  Agent  Learning