ABSTRACT

Web document search is used for finding user knowledge patterns based on the query search. In day to day development, the speed and accuracy is most important for all resources to attain proficiency in search. Web directory holds huge amount of information archives and data of dissimilar fact things based on the similar cases. So relevance of information is crucial fact in the search to retrieve the data. Development begins from the data mining and moves to the information search world. The importance given to the document, mining similarity measure improves the analysis to retrieve the best case results. Various researches concentrate the document mining semantic similarity measure and search case algorithms to improve the accuracy. But the performance is not achieved due to poor novelty and dissimilar facts of high dimensionality in cluster case analysis. Grouping the essential detail based on the relevance of document mining had the accuracy in various clustering strategy. Web domain strategy considers only the index search to match case the document retrieval from the data source. They do not formulate the search case matching of the mining measures. The inaccurate information extraction does not give optimized result for search case similarity.

The knowledge based management improves the result in document search using semantic search by extracting patterns relevance on text document mining. The efficient management resource of search case similarity uses lexical analysis to extract the information in document clustering. Various available document clustering techniques like hierarchical, multilevel, k-means, Hybrid are the formalized techniques which do not provide effective document content mining. Initially to propose a lexical content semantic search using vantage point of indexing approach (LCSS-VP) in web document clustering for effective web document search.
Specifically, to give a best solution based on the Latent semantic indexing (LSI) for effective search document mining, the relation between the documents in the datasets are analyzed. A vantage point based similarity measure with LSI method is proposed to improve the search case semantic measure among all the web documents. Addition to improve the accuracy of cluster indexing in search case to implement a linear propagation based web document clustering and semantic search using time space (LPSTS-RD) Rayleigh distribution. The predominant words are identified by the Self Organizing Pattern, and they are clustered with the semantic similarity by applying Rayleigh distribution. This improves the document clustering in web search repository. Using vantage point Latent Semantic Indexing the clustered predominant words of all the documents are indexed. Thereby, information retrieval process is enhanced. The results shows a profound improvement in accuracy of information retrieval process when compared to the existing methods.