ABSTRACT

In classical world, internet things of data be most widely used for various organization. During the usage of data makes relevant information needs for searching documents. So searching contents needs multiple meaningful information (meta data) for this reason data mining tasks are intently used for text mining. Text clustering methods can be used to structure large sets of text or hypertext documents that are make easy to extract relevant information. The well-known methods of text clustering, need not really address the special problems of text clustering, because meaningful information have various resource with high dimensionality of the data, very large size of the databases and understandability of the cluster description. The World Wide Web continues to grow at an amazing speed. On the other hand, there is also a quickly growing number of text and hypertext documents managed in organizational intranets, representing the accumulated knowledge of organizations that becomes more and more important for their success in today’s information society. Due to the huge size, high dynamics, and large diversity of the web data and of organizational intranets, it has become a very challenging task to find the truly relevant content for some user or purpose.

Due to analyze of text statement for various matching conditions in semantic process makes suitable prompt for meaningful information.

In first approach, to implement An Graph Based Sentence Level Semantic Linkage Weighing Model For Efficient Text Clustering. In this method a graph based sentence level semantic linkage weighting (GSSWM) approach for clustering text documents is discussed. Each statement of the document is considered as a node of graph and for each node the method computes semantic similarity for each sentence of document towards the
available classes. The method maintains number of semantic concepts and their linkage with the terms under the class as well as the terms of other semantic classes. The semantic linkage represents the interior and exterior relations the node has with the semantic graph. Based on computed sentence level semantic linkage weight, the text document is clustered and the method produces efficient clustering than other methods in all the factors of text clustering.

In second approach, to implement an Non-Class Element Based Iterative Text Clustering Algorithm for Improved Clustering Accuracy Using Semantic Ontology. In this method, a non-class element based iterative clustering approach is discussed. First the method extracts the terms of the class using the preprocessing algorithm. The preprocessing algorithm extract the terms and for each term the method performs stemming and tagging. Finally the method selects a subset of terms from the entire term set. For each term identified, the method computes the semantic bound measure using the semantic ontology. Then the method computes the semantic closeness measure for the document. Based on computed semantic closeness measure, the method selects the class for the document. Then the method identifies the non class elements from the document and for each class, the method computes non class weight. Based on computed semantic closeness and non class weight, the method computes the true weight for each class. Finally a class is selected based on computed true weight.

Finally, to implement a Graph Based Relational Depthness Estimation Based Text Clustering Using Semantic Ontology. The method uses the semantic ontology which contains the collection of classes and their properties. The properties and the classes are arranged in different hierarchical level. Using the semantic ontology the method generates the semantic graph and for the input text document the method extract the terms using the preprocessing algorithms. Then for each term from the term set, the method computes the
depthness estimation for each class. Based on the depthness estimated, the method selects the class of the document.