IMPLEMENTATION OF TOPICS EXTRACTION USING HYPERGRAPH PARTITIONING

Student’s Name : Gestyana Ari Restanti
NRP : 5106100151
Major of Department : Teknik Informatika FTIf – ITS
Advisor : Diana Purwitasari S.Kom, M.Sc

Abstrak

Documents clustering and classification were methods that used by user to organize their documents collections. One of the clustering and classification methods was clustered and classified it by the topics. Sometimes user couldn't know what the discussed topics were or they want to organize it using different topics. The identification of topics that used was by the keywords. Topics extraction supposed to collect important terms from documents collection then used it to identify the topics. The important terms in the collection were mapped into graph of word vertices. We used hypergraph partitioning method to partition the graph into sub-graphs that contain keywords to identify the mostly discussed topics.

On this project, we used Fiduccia-Mattheyeses and K-Way Spectral Clustering algorithm to perform hypergraph partitioning. Fiduccia-Mattheyeses swap the nodes among the partitions iteratively until we find the satisfying result. K-way spectral clustering was algorithms that cluster points using eigenvectors of matrices derived from the data.

From the experiments, both of the algorithms could perform partitioning that resulted keywords collections. Each partition could identify the discussed topics successfully. But, k-way spectral clustering algorithm showed better result than Fiduccia-Mattheyeses with precision and recall reached 72,99% and 43,90% respectively.

Keyword: topics extraction, hypergraph partitioning, Fiduccia-Mattheyesse, spectral clustering.