

ABSTRAK

Nama : Indri Nurandini

Jurusan : Matematika

Judul : *Clustering* Hadist menggunakan Algoritma *COATES*

Seiring dengan perkembangan teknologi, jumlah data teks pun semakin banyak. Walaupun jumlah data teks sangat banyak, tetapi data teks yang diolah untuk menjadi sebuah informasi yang memiliki pengetahuan masih sedikit. Oleh karena itu, berkembang teknik *text mining*. *Text mining* adalah proses ekstraksi pola berupa pengetahuan dari sebagian besar jumlah data teks, data teks dapat berupa *paper*, berita, Al-quran, dan hadist. *Clustering* hadist merupakan salah satu penerapan dalam proses *text mining* yang menggunakan hadist sebagai data teks untuk dikelompokkan. *Clustering* merupakan proses pengumpulan data kedalam kelas-kelas data yang serupa tanpa penentuan kategorisasi sebelumnya. Salah satu perkembangan dalam proses *clustering* adalah pemanfaatan *informasi sampingan* dari sebuah data yang berupa *metadata* dengan menggunakan Algoritma *Content and Atribut tambahan Based on Text Clustering (COATES)*. *Metadata* adalah informasi yang ditanam pada *file* yang isinya berupa penjelasan tentang *file* tersebut. Adapun *metadata* dari sebuah hadist yang digunakan dalam penelitian ini adalah “*sanad*” dari “*matan*” hadist pada kitab Al-Muwwata yang dimanfaatkan sebagai *informasi sampingan* dalam proses *clustering* menggunakan Algoritma *COATES* untuk mengetahui pengaruh “*sanad*” dalam proses *clustering*. Metode penentuan *centroid* dengan menggunakan *neighbors and link* menghasilkan data lebih merata di setiap *clusternya*. Lalu, hasil dari inialisasi *cluster* tersebut dievaluasi dengan menggunakan metode *silhoutte coefficient* dengan nilai yang diperoleh dari nilai s_i masing-masing data dengan nilai rata-rata *silhoutte coefficient* adalah 0.20877.

Kata Kunci: *Text Mining, Clustering, Hadist, Informasi Sampingan, Metadata, Algoritma COATES, Algoritma k-means, Neighbors and Link, Silhoutte Coefficient.*

ABSTRACT

Name : **Indri Nurandini**

Majors : **Mathematics**

Title : ***Clustering of Hadist using COATES Algorithm***

Along with the development of technology, the amount of text data is increasing. Although the amount of text data is very large, but the text data that is processed to become information that has little knowledge. Therefore, developing text mining techniques. Text mining is the process of extracting patterns in the form of knowledge from most of the amount of text data, text data can be in the form of paper, news, Quran, and hadith. Hadith clustering is one application in the text mining process that uses hadith as text data to be grouped. Clustering is the process of collecting data into similar data classes without prior categorization determination. One of the developments in the clustering process is the use of side information from a data in the form of metadata using additional Content and Attribute Algorithms Based on Text Clustering (COATES). Metadata is information that is planted in a file whose contents are an explanation of the file. The metadata of a hadith used in this study is "sanad" from the "matan" hadith in the book Al-Muwwata which is used as a side information in the clustering process using the COATES Algorithm to determine the effect of "sanad" in the clustering process. The method of determining centroid by using neighbors and links produces more even data in each cluster. Then, the results of the cluster initialization are evaluated using the silhouette coefficient method with the value obtained from the s_i value of each data with the average value of silhouette coefficient is 0.20877.

Keywords: *Text Mining, Clustering, Hadith, Side Information, Metadata, COATES Algorithm, k-means Algorithm, Neighbors and Link, Silhoutte Coefficient.*

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG