INTISARI

BUG REPORT CLUSTERING MENGGUNAKAN *AGGLOMERATIVE HIERARCHICAL CLUSTERING* DAN *WEIGHTED JACCARD COEFFICIENT*

Krisnawan Hartanto
19/448711/PPA/05794

Clustering menggunakan *AHC* dengan perhitungan jarak *Jaccard coefficient* hanya mempertimbangkan kumpulan kata-kata yang sama antara dua dokumen tapi tidak mempertimbangkan apakah suatu kumpulan kata-kata itu penting atau tidak. Penelitian terdahulu menambahkan algoritma *Inverse Document Frequency (IDF)* pada *Jaccard Coeffcient* untuk menghitung tingkat kepentingan kelompok kata.

Penelitian ini bertujuan meningkatkan performa *clustering bug report* menggunakan *AHC* dengan menambahkan *Inverse Document Frequency* pada perhitungan *Jaccard Coefficient.* Penelitian dilakukan pada *dataset* dengan 651 baris *bug*. *Clustering* dilakukan dengan model *unigram, bigram, trigram,* dan *quadrigram*.

Hasil penelitian menunjukan bahwa penambahan *IDF* pada perhitungan *Jaccard Coefficient* meningkatkan nilai *silhouette* pada *AHC* untuk parameter *bigram* sebesar 13.13%, *trigram* sebesar 5.64%, dan *quadrigram* sebesar 0.45%. Dari hasil penelitian, kombinasi *Jaccard Coefficient* dengan *Inverse Document Frequency* unggul pada tiga dari empat parameter pengujian.

ABSTRACT

Clustering using AHC with the calculation of the distance Jaccard coefficient only considers the same set of words between two documents but does not consider whether a set of words is important or not. Previous research added the Inverse Document Frequency (IDF) algorithm to Jaccard Coeffcient to calculate the importance of word groups.

This study aims to improve the performance of clustering bug reports using AHC by adding Inverse Document Frequency to the calculation of the Jaccard Coefficient. The study was conducted on a dataset with 651 bug lines. Clustering is done with unigram, bigram, trigram, and quadrigram models.

The results showed that the addition of IDF to the Jaccard Coefficient calculation increased the silhouette value on AHC for bigram parameters by 13.13%, trigrams by 5.64%, and quadrigrams by 0.45%. From the research results, the combination of Jaccard Coefficient with Inverse Document Frequency excels in three of the four test parameters.

Keyword: *Cluster, Agglomerative Hierarchical Clustering, AHC, IDF, bug repor*