

KLASIFIKASI DOKUMEN SKRIPSI TEKNIK INFORMATIKA DENGAN MENGGUNAKAN ALGORITMA K-MEDOID

Abstraksi

Skripsi merupakan syarat lulusnya seorang mahasiswa yang mengambil jenjang Strata 1. Oleh karena itu, seiring bertambahnya lulusan, maka bertambah pula dokumen skripsi. Dengan jumlah dokumen yang sangat besar, maka akan lebih mudah mencari suatu dokumen apabila dokumen-dokumen tersebut telah diorganisir dan dikelompokkan atau diklasifikasi sesuai dengan kategorinya. Skripsi ini bertujuan untuk membuat suatu aplikasi yang mampu mengklasifikasikan dokumen skripsi menggunakan system temu balik informasi dan dikombinasikan dengan algoritma *clustering K-Medoid*.

Proses clustering dokumen ini diawali dengan proses *parsing*, *stemming*, dan *indeksing*. Algoritma pembobotan yang digunakan adalah algoritma TF-IDF. Secara sederhana, algoritma TF-IDF ini berkaitan dengan frekuensi kemunculan kata dalam satu dokumen dan dokumen yang lain. Dari *indeks* yang dihasilkan dibentuklah *matrik* yang merupakan representasi dari *vector space model* yang komponennya berupa dokumen dan *term*. *CosSim* merupakan algoritma yang digunakan untuk mengukur jarak kedekatan antar dokumen.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

Untuk melakukan klasifikasi pengelompokan dokumen, digunakan algoritma *clustering* K-Medoid. Medoid merupakan obyek yang mempunyai nilai total jarak minimal terhadap obyek – obyek lain dalam satu *cluster* dan menjadi median dari *cluster* tersebut dan terletak paling tengah.

Pada penelitian ini dilakukan 2 jenis pengujian yaitu pengujian uji presisi dan pengujian validasi cluster. Jenis pengukuran untuk menguji validitas cluster adalah indeks *Dunn*. Dalam mengukur validasi cluster, hasil dari indeks *dunn* ini akan dikombinasikan dengan penghitungan kesamaan keyword atau term dokumen tertinggi medoid di tiap cluster. Dengan data sebanyak 200, untuk pengujian uji presisi yang dilakukan oleh 3 orang penguji didapatkan hasil presisi 62%, 36%, dan 44%. Sedangkan hasil dari validasi cluster dengan menggunakan indeks *dunn* yang dikombinasikan dengan penghitungan keyword, didapat hasil jumlah *k* yang ideal yaitu 7 dengan jumlah pasang *keyword* dengan bobot tertinggi antar medoidnya adalah 0 (nol) dan jumlah hasil perhitungan indeks *dunn* adalah 0.11044510560478325

CLASSIFICATION OF INFORMATIC ENGINEERING FINAL PROJECT DOCUMENT USING K-MEDOID ALGORITHM

Abstract

This thesis is a requirement of graduation of a student who took Strata 1. Therefore, the concomitant increase in graduates, it also increased thesis document. With a very large number of documents, it will be easier to find a document if these documents have been organized and grouped or classified according to the category. This thesis aims to create an application that's able to classify a thesis document using *information retrieval* and combined with clustering algorithm K-Medoid

This document clustering process begins with the parsing, stemming and indexing process. Weighting algorithm that used is TF-IDF algorithm. Put simply, this algorithm is related with frequency of occurrence of the word in one document and another documents. The result from index processing can be used to build a Matrix representative of a vector space model. The components of this matrix is documents and terms. *CosSim* is an algorithm that is used to measure the proximity between documents.

To perform classification of documents, used K-Medoid clustering algorithm. Medoid is an object which has the minimum total distance value to the other objects in one cluster. And become the median from this cluster, and located at the center of cluster.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

This thesis use 2 type of testing . That is validation testing and precession testing. The type of measurements to test the validity of the cluster is *dunn* index. In the cluster validation measure, the result of *dunn* index will be combined with the similarity calculation of the highest medoid document keyword or term in each cluster. With as much as 200 documents , for testing precession test performed by 3 examiners, the result is 62% ,36% and 44%. While the result of cluster validation using *Dunn's* index that is combined with with the keyword calculation, the result for the ideal *k* is 7 with the pairs of keyword with the highest weight among the medoid is 0 (zero) and the result for *dunn* index calculation is 0.11044510560478325