

PENGGUNAAN METODE COSINE SIMILARITY DAN TF-IDF UNTUK KLASIFIKASI JUDUL SEMINAR PROPOSAL PADA FAKULTAS TEKNIK UNIVERSITAS JABAL GHAFUR

Nilawati¹, Husaini² Junaidi Salat³

Teknik Informatika, Universitas Jabal Ghafur, Sigli

e-mail: Nilahusein220999@gmail.com

ABSTRACT

The title classification system aims to help users to Group A student seminar title document into a Category. The number of Student Research title documents requires an application system that can group these documents according to their respective categories. The purpose of this study is how to build a seminar title classification system by utilizing the approach through the text of the title and description of the seminar with the TF-IDF Method to give weight to the frequency of the relationship on the occurrence of a word (term) in the calculation of the merger of title and description of the seminar. And the Cosine Similarity algorithm is used as a comparison method to know how much similarity between the two documents. This research is conducted by implementing the text mining method with cosine similarity algorithm and TF-IDF weighting so that it is expected to classify data automatically, quickly, and accurately. Using dummy data as training data used in this study, totaling 100 seminars that have been carried out with several categories of different fields of science. And for its implementation using the PHP programming language, and with the help of data resources the use of text mining that is already in PHP which is useful for the text preprocessing stage, and then will continue with the implementation process of TF-IDF Motode and Cosine Similarity algorithm.

Keywords : Cosine Similarity, classification of titles, TF-IDF, Website, Seminar

PENDAHULUAN

Kegiatan yang umum dilakukan oleh sebuah instansi adalah kegiatan pengarsipan suatu dokumen, baik dokumen dalam bentuk fisik maupun elektronik. Kegiatan pengarsipan melibatkan dokumen dengan jumlah yang cukup besar, sehingga diperlukan suatu metode yang praktis dan efisien dalam pengelolaanya. Salah satu metode yang digunakan dalam pengelolaan dokumen adalah pengklasteran atau pengklasifikasian dokumen.

Pengklasifikasian dokumen dilakukan dalam upaya memisahkan atau mengelompokkan dokumen berdasarkan ciri-ciri atau kategori tertentu. Salah satu metode yang dapat digunakan adalah dengan pengklasifikasian secara otomatis dengan text mining. Banyak metode text mining yang digunakan dalam mengklasifikasikan dokumen atau teks, salah satunya adalah algoritma cosine similarity.

Pengklasifikasian teks sangat dibutuhkan dalam berbagai macam aplikasi, terutama aplikasi yang jumlah dokumennya bertambah dengan cepat. Ada dua cara dalam penggolongan teks, yaitu text clustering dan klasifikasi teks. Text clustering berhubungan dengan menemukan sebuah struktur kelompok yang belum kelihatan (tak terpandu atau unsupervised) dari sekumpulan dokumen. Sedangkan pengklasifikasian teks dapat dianggap sebagai proses untuk membentuk golongan golongan (kelas-kelas) dari dokumen berdasarkan pada kelas kelompok yang sudah diketahui sebelumnya (terpandu atau supervised).

Fakultas Teknik Universitas Jabal Ghafur pada proses pelaksanaan seminar proposal skripsi Mahasiswa masih melakukan pengelompokan judul skripsi menggunakan model manual yaitu dengan aplikasi MS. Excell. Penerapan metode cosine similarity dan TF-IDF dapat diterapkan untuk membantu staf SBAK fakultas Teknik.

Berdasarkan permasalahan di atas Penulis tertarik untuk melakukan suatu penelitian tentang “Penggunaan Metode Cosine Similarity Dan TF-IDF Untuk Klasifikasi Judul Seminar Proposal Pada Fakultas Teknik Universitas Jabal Ghafur”.

METODE

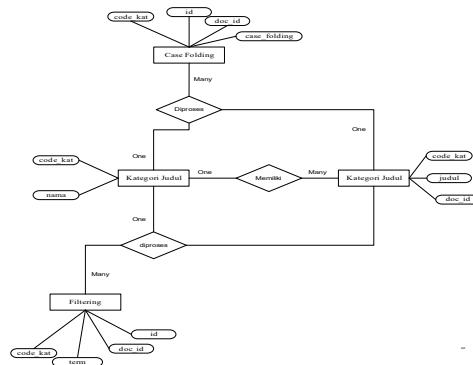
Metode yang digunakan dalam penelitian ini :

1. *Studi Literatur* (Studi Perpustakaan)
2. Observasi
3. Pengumpulan Data
4. Perancangan dan Membangun Aplikasi

HASIL DAN PEMBAHASAN

Entity Relationship Diagram (ERD)

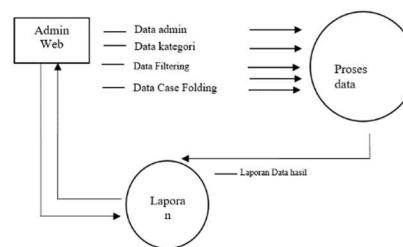
Entity Relationship Diagram (ERD) adalah alat pemodelan data utama dan akan membantu mengorganisasi data dalam suatu proyek ke dalam entitas-entitas dan menentukan hubungan antar entitas, dapat dijelaskan sebagai berikut:



Gambar 3.1 Rancangan ERD

Perancangan Proses Menggunakan Data Flow Diagram (DFD)

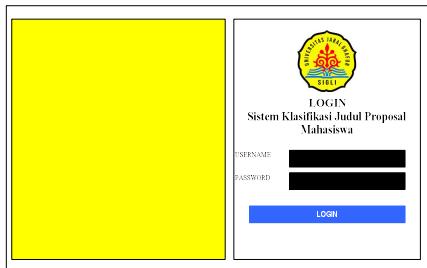
Data Flow Diagram (DFD) level 0 adalah merupakan penjabaran dari konteks diagram.



Gambar 3.2 Data Flow Diagram Level 0

Rancangan Halaman Login User

Halaman login merupakan proses authentifikasi user admin yang akan menggunakan aplikasi.



Gambar 3.3 Rancangan Halaman Login

Rancangan Halaman Dashboard

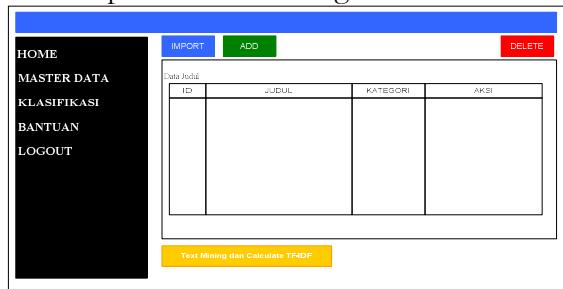
Halaman dashboard adalah halaman utama yang muncul ketika aplikasi pertama kali muncul setelah proses login berhasil.



Gambar 3.4 Rancangan Halaman Dashboard

Rancangan Halaman Data Training

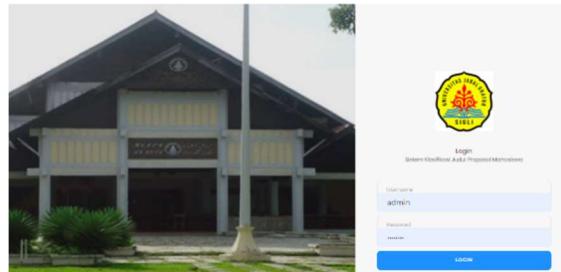
Merupakan rancangan halaman untuk proses data training.



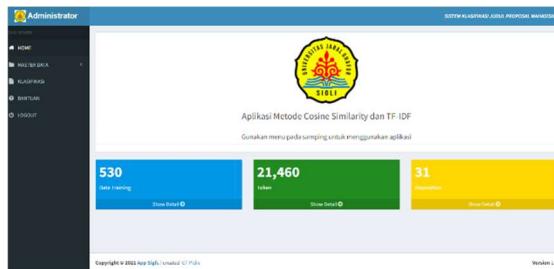
Gambar 3.5 Rancangan Halaman Data Training

Tampilan Halaman Login

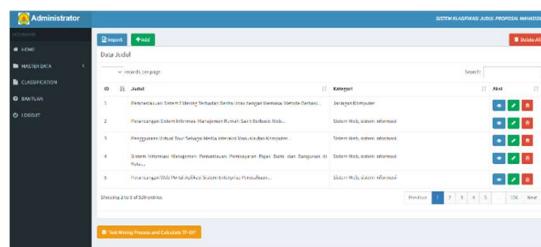
Menampilkan form login user untuk input username dan password.

**Gambar 3.6 Halaman Login User****Tampilan Halaman Dashboard**

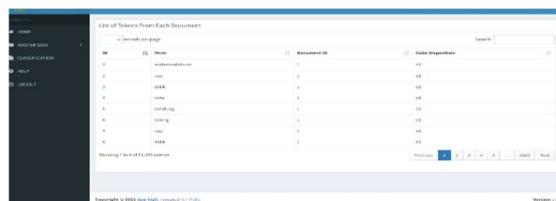
Merupakan halaman awal yang muncul setelah proses login user berhasil dilakukan.

**Gambar 3.7 Halaman Dashboard****Tampilan Halaman Data Training**

Merupakan halaman yang menampilkan data dokumen judul yang akan dilakukan klasifikasi. Untuk melakukan inputan data, dapat dilakukan melalui menu import dan add.

**Gambar 3.8 Halaman Kategori Aduan****Tampilan Halaman Data Token**

Menampilkan data token untuk pemrosesan data kata yang menjadi acuan untuk pengolah data dokumen.



Gambar 3.9 Halaman Data Token

Tampilan Halaman Weight Value

Menampilkan halaman aplikasi untuk data weight value yang telah dilakukan pemrosesan berdasarkan data dokumen.

Weight TF-IDF Value List From Each Document						
	Records per page			Searches		
MASTER DATA	ID	Term	Document ID	Count	Weight TF-IDF	
CLASSIFICATION	10000	repah	500	1	0.3462	
FILED	10000	sojar	500	1	0.3058	
DOCUMENT	10000	Kutu	500	1	0.3462	
	10001	name	500	2	0.2146	
	10001	re	500	1	0.7515	
	10001	motion	500	1	0.3778	
	10002	Ernakulam	500	2	0.3239	
	10003	Brahma	500	1	0.5512	

Gambar 3.10 Halaman Weight Value

Tampilan Halaman Basic World

Merupakan halaman aplikasi yang menampilkan kumpulan kata dasar atau disebut basic word.

Basic Word Use	
<input type="button" value="New"/>	<input type="button" value="Search/Find"/>
<input type="button" value="Edit"/>	<input type="button" value="Delete"/>
<input type="button" value="Print"/>	<input type="button" value="Print"/>
<input type="button" value="Logout"/>	<input type="button" value="Logout"/>
ID	Basic Word
24801	apple
24802	art
24803	babies
24804	calculator
24805	cheese
24806	church
24807	coffee
24808	city
24809	computer
24810	country
24811	dog
24812	egg
24813	fire
24814	flower
24815	fridge
24816	guitar
24817	house
24818	ice cream
24819	light
24820	lunch
24821	map
24822	money
24823	ocean
24824	orange
24825	paper
24826	phone
24827	post office
24828	rain
24829	school
24830	sheep
24831	shoe
24832	spoon
24833	sun
24834	teeth
24835	tree
24836	water
24837	waterfall
24838	yellow

Gambar 3.11 Halaman Kata Dasar

Tampilan Halaman Kategori Dokumen

Menampilkan daftar kategori dokumen untuk judul skripsi mahasiswa.

Gambar 3.12 Halaman Kategori Dokumen

Tampilan Halaman Klasifikasi

Merupakan halaman proses klasifikasi suatu dokumen skripsi. Proses klasifikasi akan terjadi secara otomatis untuk suatu judul yang diinputkan kedalam aplikasi.

**Gambar 3.13 Halaman Proses Klasifikasi**

KESIMPULAN

Berdasarkan hasil implementasi dan pembahasan aplikasi, dapat diambil beberapa kesimpulan diantaranya:

1. Dari sistem rekomendasi yang sudah dibuat, sudah bisa berjalan dan sesuai dengan apa yang diprediksi, seperti contoh pada tahapan testing salah satu judul yang akan diklasifikasikan dengan data kategori yang ada di database, judul proposal tentang sistem pendukung keputusan, kemudian dibandingkan dengan kategori judul yang ada di database, sesuai dengan apa yang diprediksi dengan melakukan pendekatan melalui teks atau kata.
2. Menggunakan Bahasa pemrograman PHP dan database mysql dapat membangun satu sistem aplikasi untuk proses klasifikasi data dengan proses text processing.
3. Metode cosine similarity dapat mengklasifikasikan judul proposal seminar mahasiswa dengan mudah dan cepat sesuai dengan kategori judul yang ditentukan

SARAN

Penelitian ini masih butuh banyak perbaikan di sistem klasifikasi judul proposal, bagi usernya agar menjadi lebih relevan dari sistem yang sudah dibuat, beberapa saran untuk penelitian kedepannya yaitu:

1. Membandingkan dengan penggunaan metode klasifikasi dokumen lainnya untuk melihat kemampuan dari masing-masing metode yang digunakan.
2. Penambahan fitur klasifikasi dokumen dengan banyak data bukan hanya satu judul pada aplikasi web yang dibangun
3. Membuat hak akses multi user untuk penggunaan aplikasi dengan tujuan memudahkan manajemen user aplikasi.

DAFTAR PUSTAKA

Agung Wahana Khusnul Khuluqiyah Tacbir Hendro Pudjiantoro (2019), Klasifikasi Data Pengaduan Masyarakat Pada Laman Pesduk Cimahi Menggunakan Rocchio

Ari Aulia Hakim Alva Erwin Kho I Eng Maulahikmah (2018), Automated Document Classification for News Article in Bahasa Indonesia based on Term Frequency Inverse Document Frequency (TFIDF) Approach

Ajiprayoga, H. P. Pemanfaatan Metode Cosine Similarity Dalam Menentukan Kemiripan Iklan Pada Situs Jual Beli Online. 2015.

Arifin, M. F., & Fitrianah, D. Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada. 2018.

Bening Herwijayanti Dian Eka Ratnawati Lailil Muflikhah, Klasifikasi Berita Online dengan menggunakan Pembobotan TFIDF dan Cosine Similarity, 2018.

Kaplan, R.M. (1995). A Methode for Tokenizing Text. Palo Alto Research Center (Festscript in The Honor of Prof. Kimmo Koskenniemi's 60 th Anniversary).

Sulistyo, W. & Sarno, R. (2008), Auto Matching Antar Dokumen dengan Metode Cosine Measure, Seminar Nasional Teknologi Informasi dan Komunikasi, Indonesia.

P. D. Ramadhan, "Analisis Pengelolaan Pengaduan Masyarakat," Jurnal Administrasi Publik, 2011.

Prasetya, C. S. D. Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. 2017.

Oktora, R., & Susanty, W. Perancangan Aplikasi E-Commerce Dengan Sistem Rekomendasi Item-Based Collaborative Filtering. 2012.

Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. Penerapan Metode Term Frequency Inverse Document

Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web. 2018.

Wahyuni, R. T., Prastiyanto, D., & Supraptono, E. Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. 2017.

Dian Oktaviani, S. H. Rancang Bangun Portal Seminar Nasional. 2018.

Sugiyamta. Sistem Deteksi Kemiripan Dokumen Dengan Algoritma Cosine Similarity Dan Single Pass Clustering. 2015.

Rismayani. Sistem Rekomendasi Pencarian Jodoh Syariah Menggunakan. 2018.

Kesuma, H. W. A. Penerapan Metode Tf-Idf Dan Cosine Similarity Dalam Aplikasi Kitab Undang- Undang Hukum Dagang. 2016.

Imbar, V., Radian. Adelia, Ayub, M., dan Rehatta, A. 2014. Implementasi Cosine Similarity dan Algoritma Smith Waterman untuk Mendeteksi Kemiripan Teks. Jurnal Informatika Volume 10,Nomor 1.

Sugiyamta. 2015. Sistem Deteksi Kemiripan Dokumen dengan Algoritma Cosine Similarity dan Single Pass Clustering. Jurnal Informatika Volume 7, Nomor 2.

- Susandi, D. dan Sholahudin, U. 2016. Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. *Jurnal Teknologi Informasi* Volume 3, Nomor 1.
- Nurdiana, O., Jumadi., dan Nursantika, D. 2016. Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemahan Al-Qur'an dalam Bahasa Indonesia. *Jurnal Online Informatika* Volume 1, Nomor 1.
- Kurniawan, A. Solihin, F., dan Hastarita, F. 2014. Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan Menggunakan Cosine Similarity. *Jurnal SimanteC* Volume 4, Nomor 2.
- Nurjanah, M. Hamdani. dan Astuti, I. Fitri. 2013. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Jurnal Informatika* Volume 8, Nomor 3.
- Ye, J. 2014. Vector Similarity Measures of Simplified Neutrosophic Sets and Their Application in Multicriteria Decision Making. *Internasional Journal of Fuzzy Systems* Volume 16, Nomor 2