

**PERANCANGAN DAN IMPLEMENTASI SISTEM DETEKSI
PLAGIARISME PROYEK AKHIR DENGAN METODE *COSINE
SIMILARITY***

*Design and Implementation of a Final Project Plagiarism Detection System Using
Cosine Similarity Methode*

PROPOSAL PROYEK AKHIR

Diajukan sebagai syarat untuk mengambil Mata Kuliah Proyek Akhir

oleh :

RIVAL FAUZI

6705184127



**D3 TEKNOLOGI TELEKOMUNIKASI
FAKULTAS ILMU TERAPAN
UNIVERSITAS TELKOM
2021**

LEMBAR PENGESAHAN

Proposal Proyek Akhir dengan judul :

PERANCANGAN DAN IMPLEMENTASI SISTEM DETEKSI PLAGIARISME
PROYEK AKHIR DENGAN METODE *COSINE SIMILARITY*

*Design and Implementation of a Final Project Plagiarism Detection System Using Cosine
Similarity Methode*

oleh :

RIVAL FAUZI

6705184127

Telah diperiksa dan disetujui untuk diajukan sebagai syarat mengambil
Mata Kuliah Proyek Akhir
pada Program Studi D3 Teknologi telekomunikasi Universitas Telkom

Bandung, 15 Maret 2021

Menyetujui,

Pembimbing I

Pembimbing II

Muhammad Iqbal, S.T., M.T.

NIP. 10840012

Tita Haryanti, S.T., M.T.

NIP. 20950009

ABSTRAK

Plagiarisme merupakan sebuah tindakan untuk mengambil ide, mengambil hasil riset, mengakuisisi hasil riset, dan meringkas suatu tulisan tanpa menyebutkan sumbernya baik disengaja maupun tidak disengaja. Metode *cosine similarity* merupakan salah satu metode yang dapat digunakan untuk menghitung nilai kemiripan antar dokumen dengan membandingkan dokumen yang diupload pada basisdata sehingga diperoleh presentase nilai kemiripan antar dokumen. *External plagiarism detection (EPD)* digunakan sebagai pendekatan dalam membandingkan isi artikel. Perancangan dan pengimplementasian sistem akan dilakukan dalam beberapa tahapan dengan harapan sistem dapat bekerja dengan optimal dan dapat mendeteksi kemiripan teks secara akurat.

kata kunci : plagiarisme, *cosine similarity*, *EPD*

DAFTAR ISI

LEMBAR PENGESAHAN.....	i
ABSTRAK	ii
DAFTAR ISI	iii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan dan Manfaat	2
1.3 Rumusan Masalah	2
1.4 Batasan Masalah	2
1.5 Metodologi	3
BAB II DASAR TEORI.....	4
2.1 Plagiarisme.....	4
2.2 <i>Cosine Similarity</i>	5
2.3 <i>External Plagiarism Detection</i>	5
BAB III MODEL SISTEM	6
3.1 Blok Diagram Sistem.....	6
3.2 Tahapan Perancangan	7
3.3 Perancangan	8
BAB IV BENTUK KELUARAN YANG DIHARAPKAN	9
4.1 Keluaran yang Diharapkan	9
4.2 Jadwal Pelaksanaan.....	9
DAFTAR PUSTAKA.....	10

BAB I

PENDAHULUAN

1.1 Latar Belakang

Plagiarisme tidak diperkenankan dalam dunia akademik. Selain itu, plagiarisme bertentangan dengan sifat jujur yang dibutuhkan di dunia ilmiah maupun akademik. Tanpa sifat jujur, ilmu pengetahuan tidak berkembang seperti sekarang. Akan tetapi, masih ada orang yang bergerak di bidang ilmu pengetahuan, baik sebagai peneliti maupun akademisi, yang melakukan tindakan plagiarisme. Hal ini dipicu salah satunya karena kurangnya kemampuan menulis artikel ilmiah.

Menentukan plagiarisme dalam sebuah tulisan tidaklah mudah. Dokumen pembandingan sangatlah banyak. Apalagi di era Internet dewasa ini. Artikel banyak dituliskan di dunia maya. Untuk itu, dibutuhkan aplikasi yang dapat membantu mendeteksi potensi plagiarisme sebuah tulisan.

Irfan Pahlevi dan *Moch. Arif Bijaksana* (2016) [1] pernah mengembangkan aplikasi web yang dapat menghitung tingkat kemiripan suatu abstrak tugas akhir mahasiswa dengan abstrak tugas akhir lainnya menggunakan metode *cosine similarity*, ini membuktikan bahwa metode *cosine similarity* dapat digunakan untuk melakukan pendeteksian plagiarisme. *Cosine similarity* juga pernah digunakan oleh *Gunawan dkk.* (2018) [2] untuk mencari hubungan antar text dari dua dokumen. Selain itu dalam penelitian yang telah dilakukan *Tomáš Foltýnek dkk.* (2020) [3] telah dijabarkan beberapa poin harapan untuk pengembangan aplikasi deteksi plagiarisme, salah satunya adalah pendeteksian penggunaan kata yang memiliki arti yang sama atau sinonim kata.

Beberapa situs *online* telah mengembangkan sistem deteksinya masing-masing dari sistem *open source* hingga yang berbayar. Situs smallseotools.com [4] memiliki alat yang digunakan untuk mendeteksi plagiarisme, dengan berbagai jenis dokumen inputan yang dapat digunakan secara gratis, namun pendeteksian dibatasi untuk 1000 kata apabila tanpa *login* dan 5000 kata dengan *login*, apabila melebihi batas kata akan terdapat *error*. Situs unicheck.com [5] juga memiliki sistem deteksi plagiarisme dengan keluaran berupa persentase dan *hyperlink* menuju sumber yang

dianggap asli, namun diperlukan pembayaran uang untuk biaya berlangganan. Layanan pada duplichecker.com [6] dapat melakukan pendeteksian dengan inputan situs *online* lain dengan meletakkan *hyperlink* situs yang akan di cek, namun penggunaan layanan secara penuh diharuskan untuk membayar biaya langganan. Kesamaan dari berbagai situs *online* pendeteksi plagiarisme adalah hanya menjangkau dokumen yang terbuka di internet.

Proyek Akhir ini merancang sebuah sistem deteksi plagiarisme proyek akhir berbasis website. Sistem ini kemudian akan diterapkan pada basisdata proyek akhir D3 teknologi Telekomunikasi sehingga dibatasi untuk basisdata lokal. Artinya, dokumen pembanding yang akan dijadikan acuan adalah dokumen yang ada di basisdata aplikasi dan bukan dokumen yang ada di Internet.

1.2 Tujuan dan Manfaat

Adapun tujuan dari Proyek Akhir ini, sebagai berikut:

1. Dapat merancang sistem deteksi plagiarisme menggunakan metode *cosine similarity*.
2. Dapat mengimplementasikan sistem pada basisdata lokal yang dijadikan acuan.

1.3 Rumusan Masalah

Adapun rumusan masalah dari Proyek Akhir ini, sebagai berikut:

1. Bagaimana merancang sistem deteksi plagiarisme menggunakan metode *cosine similarity*?
2. Bagaimana mengimplementasikan sistem pada basisdata lokal yang menjadi acuan?

1.4 Batasan Masalah

Dalam Proyek Akhir ini, dilakukan pembatasan masalah sebagai berikut:

1. Perancangan dan implementasi sistem dilakukan menggunakan metode *cosine similarity*.
2. Basisdata yang digunakan adalah basisdata lokal dari aplikasi.
3. Sistem dirancang dan diimplementasikan berbasis website.

1.5 Metodologi

Metodologi pada penelitian ini, sebagai berikut:

1. Studi Literatur

Hal yang dilakukan adalah mencari informasi dan pendalaman materi-materi yang terkait melalui referensi yang tersedia di berbagai sumber.

2. Perancangan dan Implementasi

Melakukan implementasi sistem sesuai dengan parameter dan model yang telah dirancang.

3. Analisa Sistem

Mengamati hasil dari sistem yang dirancang dan menyimpulkannya.

4. Penarikan Kesimpulan

Menarik kesimpulan dari seluruh tahapan yang telah dilakukan dengan masukkan dari dosen pembimbing maka dapat diambil kesimpulan dari hasil yang telah dilakukan.

BAB II

DASAR TEORI

2.1 Plagiarisme

Menurut Kamus Besar Bahasa Indonesia (KBBI) plagiat merupakan pengambilan karangan (pendapat dan sebagainya) orang lain dan menjadikannya seolah-olah karangan (pendapat dan sebagainya) sendiri, misalnya menerbitkan karya tulis orang lain atas nama dirinya sendiri [7]. Sedangkan plagiarisme sendiri merupakan tindakan plagiat yang melanggar hak cipta [8]. Dalam peraturan Menteri Pendidikan Nasional nomer 17 tahun 2010 bab I pasal satu dinyatakan bahwa plagiat adalah perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memadai [9]. Selain itu peraturan mengenai sanksi untuk tindak plagiarisme ditulis dalam undang undang nomor 20 tahun 2003 pasal 25 ayat 2 yang menyatakan bahwa Lulusan perguruan tinggi yang karya ilmiahnya digunakan untuk memperoleh gelar akademik, profesi, atau vokasi terbukti merupakan jiplakan dicabut gelarnya [10].

Seseorang dikatakan plagiarisme apabila mengambil gagasan, mengambil hasil riset, mengakuisisi hasil riset, dan meringkas suatu tulisan tanpa menyebutkan sumbernya. Terdapat beberapa kategori plagiarisme, yaitu *word by word plagiarism* yang berarti mengutip bagian karya seseorang lebih dari 10 halaman. *Word switch plagiarism* yaitu plagiarisme yang dilakukan dengan mengubah kalimat dengan kata-kata dan paragraf baru tetapi masih meniru gaya tulisan pengarang asli. *Metaphor plagiarism* yaitu menggunakan tulisan orang lain untuk memperjelas karya sendiri. *Idea plagiarism* merupakan tindakan plagiarisme dengan mengambil ide atau gagasan orang lain. Dan kategori yang terakhir yaitu *self plagiarism* yang berarti mengutip karya orang lain secara identik tanpa mencantumkan nama pengarang asli kemudian mengirim ke berbagai jurnal untuk di publikasikan (Wibowo 2012) [11].

2.2 *Cosine Similarity*

Cosine similarity merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran [12]. *Cosine Similarity* akan melakukan penghitungan tingkat kemiripan (*similarity*) dengan membandingkan antara dua objek atau lebih. Hasil dari perbandingan tersebut berupa sudut cosinus x dengan 0 sebagai nilai terkecil yang berarti tidak terdapat kemiripan dan 1 sebagai nilai terbesar yang berarti memiliki nilai kemiripan.

Uji tingkat kemiripan teks dilakukan dengan rumus :

$$\frac{\sum_{n=1}^j (nA \times nB)}{\sqrt{\sum_{n=1}^j (nA)^2} \times \sqrt{\sum_{n=1}^j (nB)^2}}$$

Dengan $j = |A \cap B|$, nA = jumlah kemunculan kata indeks ke-n dari daftar kata pada kalimat A, nB = jumlah kemunculan kata indeks ke-n dari daftar kata pada kalimat B.

2.3 *External Plagiarism Detection*

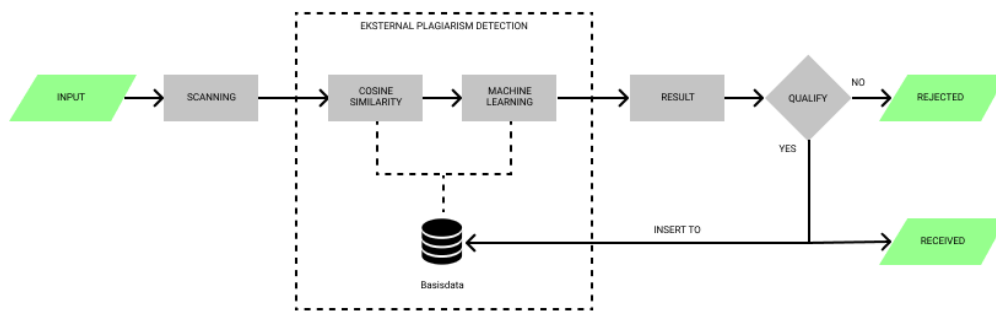
External Plagiarism Detection atau deteksi plagiarisme eksternal merupakan pendekatan dimana sistem membandingkan dokumen yang diduga plagiat dengan dokumen di repositori untuk menemukan bagian plagiasinya [13]. Dengan kata lain sistem akan membandingkan sebuah dokumen yang dicurigai plagiat (dokumen *input*) dengan seluruh dokumen pada sebuah koleksi dokumen yang diasumsikan sebagai dokumen original. Dengan memodelkan dokumen dan menentukan batas kesamaan dua buah dokumen dianggap plagiat, sistem pendeteksi akan mencari seluruh dokumen yang memiliki kesamaan teks dengan dokumen input melebihi batas tersebut.

BAB III

MODEL SISTEM

3.1 Blok Diagram Sistem

Pada bab ini akan dijelaskan mengenai perancangan aplikasi sistem deteksi plagiarisme berbasis web dengan menggunakan metode *cosine similarity*, diagram alur perancangan sistem, dan analisa kebutuhan sistem.



Gambar 3 1 Model Sistem Perancangan Aplikasi

Adapun model sistem yang telah dibuat dapat dilihat pada Gambar 3 1. dengan penjelasan sebagai berikut:

1. *Input*

Di awal *user* diminta untuk *input* dokumen dengan tipe dokumen yang telah ditentukan sebelumnya.

2. *Scanning*

Setelah dokumen telah diinput sistem akan mulai bekerja dengan melakukan *scanning* dan mengkonversi berbagai tipe dokumen input menjadi dokumen yang dapat dibaca oleh sistem.

3. *Penilaian*

Tahap selanjutnya akan dilakukan penilaian pada dokumen dengan pendekatan *External Plagiarism Detection* yang menjadi beberapa bagian yaitu *Cosine similarity*, merupakan metode untuk menghitung nilai kesamaan, *Machine Learning*, bertujuan untuk mengidentifikasi penggunaan kata yang memiliki arti yang sama, dan ada basisdata, dimana

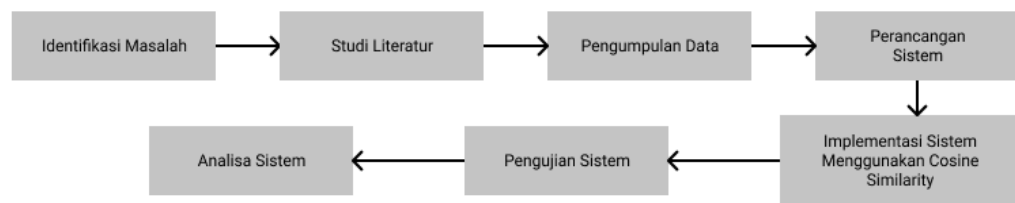
basisdata yang digunakan merupakan basisdata lokal yang berarti pengecekan hanya akan dilakukan pada berkas-berkas yang terdapat pada basisdata aplikasi.

4. *Result*

Hasil yang didapatkan terdiri dari dua bagian, pertama dalam bentuk nilai persentase seberapa besar kemiripan dokumen inputan dengan dokumen yang ada di basisdata, dan yang kedua bagian mana saja yang dianggap memiliki indikasi plagiat. Setelah didapat nilai persentase sistem akan melakukan kualifikasi dimana jika nilai persentase melebihi batas *grade* yang telah ditentukan maka sistem akan memberikan peringatan bahwa dokumen ditolak, namun jika nilai persentase dibawah batas *grade* yang telah ditentukan sistem akan menampilkan peringatan diterima dan melakukan penyimpanan dokumen kedalam basisdata, sehingga dokumen yang memiliki nilai persentase dibawah batas *grade* akan dianggap sebagai dokumen yang original.

3.2 Tahapan Perancangan

Adapun tahapan perancangan pada penelitian ini direpresentasikan ke dalam diagram pada Gambar 3.2



Gambar 3.2 Tahap Perancangan

Penelitian dimulai dari identifikasi masalah. Tahap selanjutnya yaitu studi literatur yaitu mengumpulkan teori pendukung penelitian. Dilanjutkan pada tahap pengumpulan data. Tahap selanjutnya adalah perancangan sistem yaitu untuk memahami alur sistem yang akan dibuat dan mengimplementasikan metode cosine similarity untuk menghitung kemiripan teks dalam proses deteksi plagiarisme. Setelah melakukan perhitungan nilai kemiripan dokumen, akan dilanjutkan tahap pengujian sistem dengan memperhatikan kebenaran sistem dengan metode yang

telah digunakan. Dari semua tahap diatas, dapat dilanjutkan dengan melakukan analisis dari perhitungan metode serta sistem yang telah dibuat.

3.3 Perancangan

Pada Proyek Akhir ini akan dirancang sistem deteksi plagiarisme dengan menggunakan metode *cosine similarity* berbasis website, dalam pengembangannya akan digunakan beberapa alat bantu sebagai berikut :

1. *Bootstrap*

Bootstrap adalah sebuah *framework open-source* khusus *front end*, yang nantinya akan digunakan untuk membangun tampilan dari website.

2. *Node JS*

Node JS adalah perangkat lunak yang didesain untuk mengembangkan sistem *back-end* dari aplikasi dan ditulis dalam sintaks bahasa pemrograman *JavaScript*.

3. *MySQL*

MySQL merupakan sistem manajemen basisdata dimana basisdata yang akan digunakan adalah basisdata lokal yang artinya pengecekan hanya akan dilakukan pada berkas berkas yang terdapat pada basisdata aplikasi.

BAB IV

4.1 Keluaran yang Diharapkan

Keluaran yang diharapkan dari implementasi sistem deteksi plagiarisme menggunakan metode *cosine similarity* adalah sebagai berikut :

- Dapat mengimplementasikan metode *cosine similarity* pada sistem deteksi plagiarisme dengan pendekatan *EPD*.
- Aplikasi berbasis web yang mempunyai kemampuan mendeteksi kesamaan teks secara akurat.

4.2 Jadwal Pelaksanaan

Adapun jadwal pengerjaan Proyek Akhir bisa dilihat pada Tabel 4. 1 sebagai berikut :

Tabel 4. 1 Jadwal Pelaksanaan

[illegible]

DAFTAR PUSTAKA

- [1] I. Pahlevi and M. A. Bijaksana, "Perhitungan Kemiripan Dokumen Bahasa Indonesia Menggunakan Metode Cosine Similarity (Studi Kasus : Abstrak Tugas Akhir Fakultas Informatika Universitas Telkom)," Open Library Telkom University, Bandung, 2016.
- [2] D. Gunawan, C. A. Sembiring and M. A. Budiman, "The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents," IOP, 2018.
- [3] T. Foltýnek, D. Dlabolová, A. A. Naumeca, S. Razi, J. Kravjar, . L. Kamzola, J. G. Dib, Ö. Çelik and D. W. Wulff, "Testing of support tools for plagiarism detection," International Journal of Educational Technology in Higher Education, 2020.
- [4] smallseotools.com, <https://smallseotools.com/plagiarism-checker/>, Online: Diakses pada tanggal 25 Maret 2021, 2021.
- [5] unichack.com, <https://unichack.com/>, Online: Diakses pada tanggal 25 Maret 2021, 2021.
- [6] duplichecker.com, <https://www.duplichecker.com/>, Online: Diakses pada tanggal 25 Maret 2021, 2021.
- [7] KBBI, Kamus Besar Bahasa Indonesia (KBBI), Online: Diakses pada tanggal 15 maret 2021, 2021.
- [8] KBBI, Kamus Besar Bahasa Indonesia (KBBI), Online: Diakses pada tanggal 15 maret 2021, 2021.
- [9] Republik Indonesia, Peraturan Menteri Pendidikan Nasional nomer 17 tahun 2010 tentang pencegahan dan penanggulangan plagiat di perguruan tinggi, Jakarta, 2010.
- [10] Republik Indonesia, Undang Undang No. 20 tahun 2003 tentang sistem pendidikan nasional, Jakarta, 2003.
- [11] R. P. Pratama, APLIKASI DETEKSI PLAGIARISME MENGGUNAKAN METODE COSINE SIMILARITY, Malang, 2018.
- [12] R. T. Wahyuni, D. Prastiyanto and E. Suprpto, Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi, Semarang: Jurnal Teknik Elektro, 2017.
- [13] S. P. Gunawan, L. D. Krisnawati and A. R. Chrismanto, "Analisis Fitur Stilometri dan Strategi Segmentasi pada Sistem Deteksi Plagiasi Intrinsik Teks," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, pp. 988-997, 2020.