
**PENERAPAN COSINE SIMILARITY SEBAGAI METODE
PENGUKURAN SIMILARITY INDEX PADA
SISTEM PENGAJUAN JUDUL SKRIPSI
STKIP PGRI SUMATERA BARAT**

Anggri Yulio Pernanda¹, Muhammad Hakiki²
STKIP PGRI Sumatera Barat¹
STKIP Muhammadiyah Muara Bungo²
E-mail: anggriyulio@gmail.com¹, qiqi.lubis7@gmail.com²

Abstrak

Proses pengajuan judul skripsi akan melewati tahap validasi oleh tim validator. Namun seiring bertambah banyaknya judul yang pernah diajukan maka akan menyulitkan validator untuk mengetahui apakah judul tersebut sudah pernah diajukan sebelumnya atau tidak. Diperlukan sebuah sistem yang dapat melakukan proses temu balik informasi terhadap basis data judul skripsi serta mampu menghitung tingkat kemiripan judul tersebut. Pada penelitian ini dilakukan pengembangan sistem dengan menerapkan metode *Cosine Similarity* untuk mendapatkan nilai kemiripan dari judul yang diajukan terhadap judul skripsi yang terdapat pada basis data. Hasil dari penelitian ini berupa sistem yang dapat digunakan sebagai alat bantu temu balik informasi. Informasi yang dihasilkan dari proses tersebut berupa *similarity index* (tingkat kemiripan) judul terhadap basis data judul skripsi, yang menjadi alat bantu dalam melakukan proses validasi judul skripsi di STKIP PGRI Sumatera Barat.

Kata Kunci: Cosine Similarity; Information Retrieval; Submission System.

Abstract

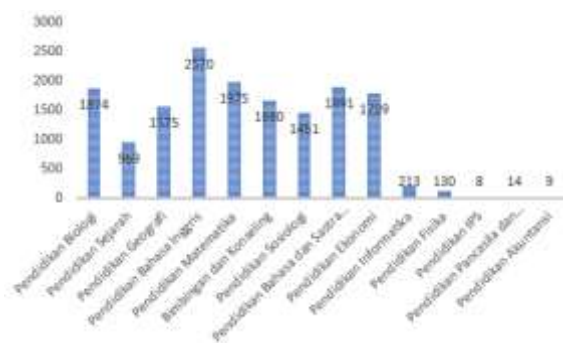
The process of submitting a thesis title will pass the validation stage by the validator team. However, as the number of titles that have been submitted increases, it will be difficult for the validator to know whether the title has been submitted before or not. We need a system that can process information retrieval against the thesis title database and is able to calculate the degree of similarity of the title. In this study, a system development was carried out by applying the Cosine Similarity method to obtain the similarity value of the proposed title to the thesis title contained in the database. The result of this research is a system that can be used as a tool for information retrieval. The information generated from this process is the degree of similarity index of the title of thesis database, which becomes a tool in conducting the title of thesis validation process in STKIP PGRI West Sumatra

Keywords: *Cosine Similarity; Information Retrieval; Submission System.*

PENDAHULUAN

Fenomena plagiarisme dalam penulisan tugas akhir bukanlah hal baru pada dunia pendidikan tinggi di Indonesia khususnya dalam kegiatan penulisan tugas akhir. Kecenderungan mahasiswa mencari judul-judul penelitian yang telah diteliti sebelumnya, tidak menutup kemungkinan tingkat isi dari skripsi tersebut juga sama.

Proses seleksi pengajuan judul skripsi di STKIP PGRI Sumatera Barat dilakukan oleh tim seleksi judul dengan melihat kesesuaian topik penelitian serta menimbang apakah judul tersebut belum pernah diteliti sebelumnya pada basis data judul skripsi secara manual.



Gambar 1. Jumlah judul skripsi
(Sumber: SIT STKIP PGRI Sumatera Barat)

Seiring waktu dengan bertambah banyaknya mahasiswa yang melakukan penelitian tugas akhir mengakibatkan sulitnya tim seleksi judul untuk melakukan validasi apakah judul tersebut telah diajukan oleh mahasiswa terdahulu, sehingga hal tersebut memberikan pengaruh terhadap tim seleksi dalam mengambil keputusan untuk menerima ataupun menolak ajuan judul tersebut.

Sistem temu balik informasi (*information retrieval*) merupakan proses di mana

kumpulan data direpresentasikan, disimpan, dan dicari untuk tujuan penemuan pengetahuan sebagai tanggapan atas permintaan pengguna (*query*) (Hanumanthappa & Nagalavi, 2018). Pada kasus pengajuan judul skripsi sistem ini dapat dimanfaatkan untuk menemukan kembali (*retrieval*) judul skripsi mahasiswa sebelumnya yang tersimpan pada basis data. Selanjutnya dalam mengukur tingkat kemiripan judul skripsi yang diajukan mahasiswa terhadap hasil proses temu balik informasi sebelumnya dilakukan perhitungan *cosine similarity*. *Cosine Similarity* merupakan metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek, dengan menghitung nilai cosinus setiap *term* pada dua dokumen (Rahutomo et al., 2012). Kelebihan dari algoritma *cosine similarity* adalah tidak terpengaruh pada panjang pendeknya suatu dokumen dan memiliki tingkat akurasi yang tinggi (Riyani et al., 2019).

Tujuan dari penelitian ini adalah menghasilkan sebuah sistem yang dapat menangan proses pengajuan judul skripsi secara online dan melakukan perhitungan *Cosine Similarity* sebagai metode pengukuran *Similarity Indexs* dalam membantu tim seleksi judul dalam mengambil keputusan untuk menerima ataupun menolak ajuan judul tersebut.

METODE PENELITIAN

Metode penelitian adalah tata cara yang lebih terperinci mengenai tahap-tahap melakukan sebuah penelitian (Hakiki, M., Fadli, R., Putra, Y., & Pertiwi, I. (2021).

Pada penelitian ini metode yang digunakan terdiri dari: pendekatan penelitian kerangka kerja (*framework*), dan kebutuhan sistem.

A. Pendekatan Penelitian

Pendekatan penelitian dalam pengembangan sistem ini menggunakan metode *Rapid Application Development* (RAD). RAD memungkinkan tim pengembangan untuk membuat "sistem yang berfungsi utuh" dalam periode waktu yang sangat singkat, jika ruang lingkup dan kebutuhan sistem sudah cukup jelas (Mishra & Dubey, 2013). Metode RAD mempunyai tiga tahapan utama dalam pengembangan sistem yaitu sebagai berikut (Fauzi & Harli, 2017):



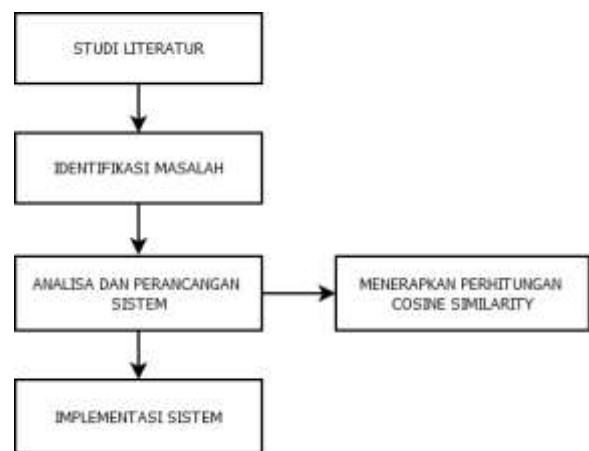
Gambar 2. Siklus RAD

1. Perencanaan Syarat-syarat
Pada tahap ini dilakukan analisa kebutuhan dari sistem yang akan dikembangkan, sehingga sistem yang dikembangkan dapat menutup semua kelemahan yang ditemukan selama ini.
2. Desain Workshop
Pada tahapan ini pengembangan melakukan perancangan dan membangun sistem sesuai dengan analisa kebutuhan yang sudah dikumpulkan sebelumnya.
3. Implementasi
Tahapan ini pengembangan dan user melakukan implementasi dan

pengujian sistem yang dikembangkan untuk dapat digunakan secara luas.

B. Kerangka Kerja Penelitian

Untuk membantu penyusunan penelitian ini, disusun kerangka kerja yang jelas. Kerangka kerja merupakan Langkah-langkah untuk menyelesaikan masalah-masalah yang dijelaskan pada gambar berikut:



Gambar 2: Kerangka Kerja

Berdasarkan kerangka kerja diatas, maka dapat diurai pembahasan masing-masing tahapan sebagai berikut:

1. Studi Literatur
Pada tahapan ini dilakukan pencarian landasan teori terkait permasalahan yang ditemukan pada proses pengajuan judul skripsi STKIP PGRI Sumatera Barat. Tujuannya adalah memperkuat analisis permasalahan serta menjadi dasar dalam melakukan penelitian.
2. Identifikasi Masalah
Pada tahap ini dilakukan identifikasi masalah terhadap objek penelitian. Pada proses pengajuan skripsi di STKIP PGRI Sumatera Barat, salah satu indikator penilaian apakah penelitian tersebut layak untuk dilanjutkan atau tidak adalah dengan melihat unsur

kebaruan atau apakah judul tersebut sudah pernah diajukan sebelumnya, hal tersebut menimbulkan masalah karena jumlah judul skripsi yang telah diajukan sebelumnya sangat banyak.

3. Analisa dan Perancangan Sistem
Setelah mendapatkan masalah yang ditemukan, selanjutnya dilakukan analisa masalah tersebut dan cara penyelesaiannya. Berdasarkan permasalahan yang telah diidentifikasi, penulis memberi solusi untuk mengembangkan sistem temu balik informasi (*information retrieval*) dan menerapkan perhitungan *cosine similarity* untuk mendapatkan tingkat *similarity index* terhadap judul yang telah diajukan sebelumnya.
4. Implementasi Sistem
Setelah menganalisa dan merancangan sistem, pada tahap ini dilakukan implementasi sistem agar dapat digunakan secara luas. Sistem yang diimplementasikan menggunakan framework laravel. laravel merupakan framework yang mempunyai skalabilitas yang kuat, sehingga dapat meningkatkan efisiensi pengembangan (He, 2015). Pada sistem yang dikembangkan juga diterapkan proses perhitungan *cosine similarity* pada judul skripsi yang diajukan. Tahap perhitungan *consine similarity* adalah dengan terlebih dahulu menghitung nilai dari judul yang diajukan. Nilai dari judul yang diajukan dapat dihitung dengan persamaan berikut:

$$\vec{d} = (w_{do}, w_{do}, \dots, w_{dk},) \quad (1)$$

Sama halnya seperti judul yang diajukan, nilai setiap judul skripsi pada basis data dihitung dengan persamaan berikut:

$$\vec{q} = (w_{qo}, w_{qo}, \dots, w_{qk},) \quad (2)$$

Dimana w_{di} dan w_{qi} ($0 \leq i \leq k$) adalah nilai yang mengindikasikan kemunculan setiap *term* pada sebuah dokumen (judul skripsi).

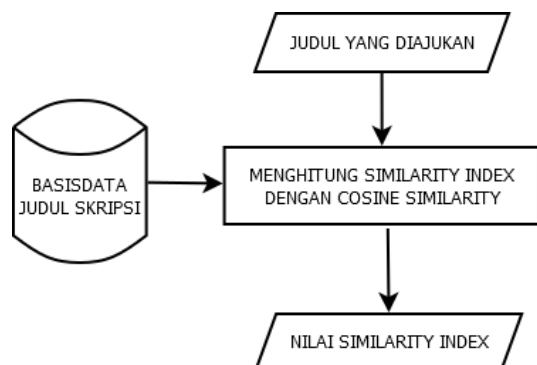
Berdasarkan nilai yang didapat sebelumnya, maka dilakukan perhitungan *similarity* terhadap kedua nilai tersebut dengan persamaan berikut:

$$\text{Sim}(\vec{q}, \vec{d}) = \frac{\sum_k^t = 1(w_{qk} \cdot w_{dk})}{\sqrt{\sum_k^t = 1(w_{qk})^2} \cdot \sqrt{\sum_k^t = 1(w_{dk})^2}} \quad (3)$$

C. Hasil dan Pembahasan

1. Perancangan Sistem

Berdasarkan perancangan sistem yang dibuat tahapan untuk mendapatkan nilai *Similarity Index* dapat dilihat pada Gambar 3.

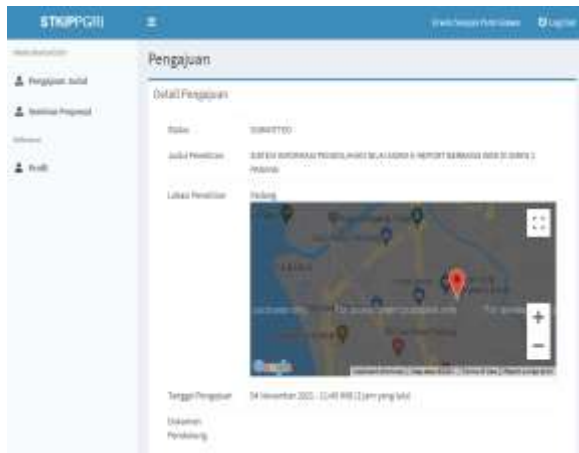


Gambar 3: Tahapan Perhitungan *Similarity Index*

Berdasarkan gambar diatas, maka dapat dijelaskan setiap tahapan sebagai berikut:

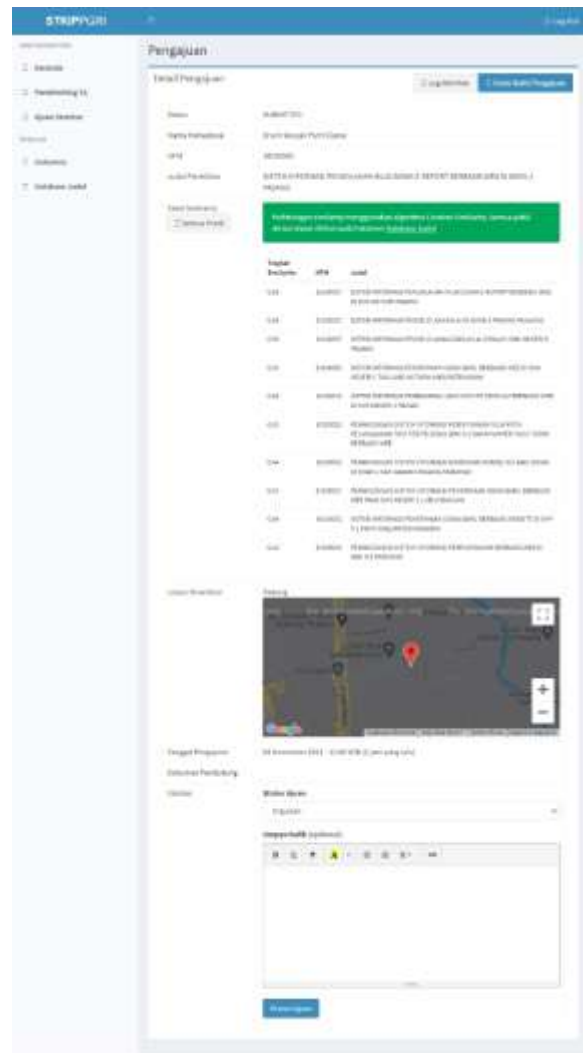
a. Pengajuan Judul

Pada tahapan ini mahasiswa melakukan pengajuan judul melalui sistem yang telah dikembangkan. Tampilan sistem pengejuan judul dapat dilihat pada Gambar 4.



Gambar 4. Halaman Pengajuan Judul Skripsi

- b. Menghitung Similarity Index
Pada tahap ini sistem akan menghitung nilai similarity index dari judul yang diajukan terhadap setiap judul yang ada pada basis data. Hasil perhitungan yang didapat nantinya akan ditampilkan dalam bentuk tabel yang diurut berdasarkan *similarity index* tertinggi.
- c. Nilai Similarity Index
Pada tahapan ini ditampilkan tabel perhitungan similarity yang dapat dilihat oleh tim seleksi judul pada halaman validasi, sebagai alat bantu untuk menentukan apakah judul tersebut layak untuk diterima atau tidak. Halaman hasil perhitungan dapat dilihat pada gambar berikut:



Gambar 5. Halaman Validasi

2. Pengujian Sistem

Pada bagian ini dilakukan pengujian perhitungan *similarity index* judul yang skripsi yang akan diajukan. Pengujian dilakukan dengan menggunakan sistem yang telah diimplementasikan secara langsung, hasil dari pengujian ini adalah mendapatkan nilai *similarity index* terhadap judul tersebut. Judul yang akan dilakukan pengujian adalah “SISTEM INFORMASI PENGELOLAAN NILAI SISWA E-REPORT BERBASIS WEB DI SMKN 1 PADANG”.

Dari hasil pengujian, didapatkan nilai *similarity index* judul yang diuji coba terhadap 10 judul yang mempunyai nilai kemiripan tertinggi pada basis data judul skripsi yang dapat dilihat pada Tabel 1.

Tabel 1. Nilai *similarity index*

Similarity Index	Judul
0.82	Sistem Informasi Pengolahan Nilai Siswa E-Report Berbasis Web Di SMK An-Nur Padang
0.55	Sistem Informasi Pengelolaan Nilai Di SMK N 2 Padang Panjang
0.50	Sistem Informasi Penerimaan Siswa Baru Berbasis Web Di SMA Negeri 1 Tanjung Mutiara Kabupaten Agam
0.50	Sistem Informasi Pengelolaan Data Nilai Siswa Di SMK Negeri 5 Padang
0.46	Sistem Informasi Pembayaran Uang Komite Sekolah Berbasis Web Di SMK Negeri 1 Painan
0.45	Perancangan Sistem Informasi Perhitungan Nilai Poin Pelanggaran Tata Tertib Siswa SMK N 1 Ranah Ampek Hulu Tapan Berbasis Web
0.44	Sistem Informasi Penerimaan Siswa Baru Berbasis Website Di SMP N 1 Panti Kabupaten Pasaman
0.44	Perancangan Sistem Informasi Penerimaan Siswa Baru Berbasis Web Pada Smk Negeri 1 Lubuk Basung
0.44	Perancangan Sistem Informasi Bimbingan Konseling Bagi Siswa Di SMAN 1 Nan Sabaris Padang Pariaman

Berdasarkan tabel diatas dapat dijelaskan bahwa Judul yang diajukan mempunyai tingkat *similarity* (kemiripan) yang tinggi dengan “SISTEM INFORMASI PENGOLAHAN NILAI SISWA E-REPORT BERBASIS WEB DI SMK AN-NUR PADANG” dimana tingkat kemiripan

judul tersebut mencapai 82%.

KESIMPULAN

Dari hasil perancangan, implementasi serta pengujian dari sistem dapat diambil kesimpulan bahwa sistem yang dikembangkan dapat digunakan sebagai alat bantu temu balik informasi. Informasi yang dihasilkan dari proses tersebut berupa nilai *similarity* (tingkat kemiripan) judul terhadap basis data judul skripsi, yang dapat menjadi alat bantu dalam melakukan proses validasi pengajuan judul skripsi di STKIP PGRI Sumatera Barat.

DAFTAR PUSTAKA

- Fauzi, A., & Harli, E. (2017). *Peningkatan Kualitas Pelayanan Melalui CRM dengan Metode*. 1(1), 76–81.
- Hakiki, M., Fadli, R., Putra, Y., & Pertiwi, I. (2021). PERANCANGAN SISTEM INFORMASI MANAJEMEN BERBASIS SEKOLAH SMA NEGERI 1 MUARA BUNGO. *Jurnal Muara Pendidikan*, 6(1), 50-57.
- Hanumanthappa, M., & Nagalavi, D. T. (2018). A survey on Information Retrieval System for Online Newspapers. *International Journal of Engineering Research & Technology (IJERT)*, 2(02), 52–56.
- He, R. Y. (2015). Design and Implementation of Web Based on Laravel Framework. *Proceedings of the*

2014 *International Conference on Computer Science and Electronic Technology*, 6(Iccset 2014), 301–304.
<https://doi.org/10.2991/iccset-14.2015.66>

Mishra, A., & Dubey, D. (2013). *A Comparative Study of Different Software Development Life Cycle Models in Different Scenarios*. 1(5).

Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012). Semantic Cosine Similarity. *Semantic Scholar*, 2(4), 4–5.

Riyani, A., Zidny, M., & Burhanuddin, A. (2019). *Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen*. 2(1), 23–27.