



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Skripsi adalah istilah di Indonesia yang menyebut karya tulis ilmiah mahasiswa sarjana, yang berisikan hasil penelitian yang membahas suatu permasalahan di bidang ilmu tertentu [1]. Sebelum membuat skripsi mahasiswa harus mengusulkan sebuah judul, tetapi judul yang diusulkan oleh seorang mahasiswa terkadang ditolak dikarenakan judul yang diusulkan sudah ada atau mirip dengan judul yang telah ada. Karena seiring berjalannya waktu, judul skripsi yang telah diterima oleh fakultas akan terus bertambah. Oleh karena itu diperlukan sebuah sistem yang dapat mencari kemiripan judul skripsi. Penelitian ini berfokus untuk merancang sebuah aplikasi yang dapat mencari kemiripan judul skripsi, tidak seperti Turnitin.com yang dapat mencari kemiripan dokumen secara keseluruhan.

Teknologi sekarang mempermudah banyak kegiatan manusia, dimulai dari hal yang sederhana menjadi yang lebih rumit. Perkembangan teknologi terjadi karena seseorang menggunakan kemampuan dan kecerdasan akalinya untuk menyelesaikan setiap masalah yang dihadapinya [2]. Perubahan tersebut ada yang terjadi secara perlahan, ada pula yang terjadi secara drastis dikarenakan pertentangan antara satu ilmu pengetahuan dengan ilmu pengetahuan yang baru [3].

Salah satu hasil perkembangan itu adalah otomatisasi proses program yang menjalankan tugas manusia. *Robotic Process Automation* atau RPA, dapat diartikan sebagai otomatisasi tugas layanan yang sebelumnya dilakukan oleh manusia [4]. Salah satu contoh menerapkan RPA adalah *Web Scraping*. *Web Scraping* dapat mempunyai arti sebagai proses mengekstraksi dan menggabungkan *content* dari Web dengan cara yang sistematis. Dalam proses seperti itu, perangkat lunak yang dikenal sebagai robot berbasis Web, yang dapat meniru interaksi *browsing* antara manusia dan *Web server*. Secara satu-per-satu, robot tersebut akan mengakses dan menarik data seperlunya [5].

Berdasarkan masalah yang dijelaskan di atas, RPA dapat membantu mahasiswa mencari jalan keluar. Dengan menggunakan fitur *Web Scrapping*, penulis dapat mengumpulkan judul skripsi yang telah diterima dan data lainnya, untuk mencari kemiripan dengan judul yang telah diminta. Data yang diambil dapat langsung di simpan ke dalam sebuah dokumen yang akan dapat digunakan untuk proses lain-

nya.

Penelitian akan menggunakan *Natural Language Processing* atau NLP, linguistik komputasi yang berkaitan dengan interaksi antara komputer dan bahasa alami manusia, untuk mengolah bahasa yang telah dikumpulkan, dan dibantu oleh TF-IDF dan *Cosine Similarity*, yang dapat menghitung seberapa sering sebuah kata muncul dalam sebuah kalimat. Dengan menggunakan ini, kita bisa mendapatkan keakuratan algoritme dan dapat dengan mudah membandingkan akurasi yang diperlukan oleh proyek kita [6].

TF-IDF akan digunakan untuk mengubah teks dan kalimat judul skripsi dan mengubahnya menjadi nilai numeral yang dipakai untuk perbandingan. Metode ini akan menentukan bobot, ukuran yang mengevaluasi pentingnya istilah (atau kata) dalam kumpulan dokumen. Teks yang penting akan meningkat secara proporsional dengan jumlah kemunculan dalam dokumen [7].

Metode yang digunakan untuk menghitung nilai TF-IDF adalah *Cosine Similarity*. Pengukuran kesamaan adalah beban komputasi utama dalam tugas pemahaman dokumen dan *Cosine Similarity* adalah salah satu cara mencari kesamaan teks yang paling populer [8]. Hasil yang dikeluarkan akan menunjukkan seberapa mirip sebuah judul skripsi yang dicari.

Penelitian yang terkait mengenai TF-IDF dan *Cosine Similarity* telah dilakukan beberapa kali, salah satunya adalah penelitian oleh Wahyuni [9], yang telah menggunakan TF-IDF dan *Cosine Similarity* untuk mengelompokkan dokumen skripsi yang menghasilkan nilai akurasi sebesar 98% dengan menggunakan 50 jenis laporan skripsi. Selanjutnya Iriananda[10], yang menggunakan TF-IDF dan *Cosine Similarity* untuk mengklasifikasikan laporan pengaduan, yang menghasilkan nilai akurasi sebesar 84% dengan menggunakan *dataset* sebanyak 200 *data*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka rumusan masalah yang dalam penelitian ini adalah, sebagai berikut:

- Bagaimana cara menerapkan TF-IDF dan *Cosine Similarity* untuk mencari kemiripan judul skripsi?
- Bagaimana kecepatan dan keakuratan eksekusi TF-IDF dan *Cosine Similarity* dalam melakukan pengecekan kemiripan judul skripsi jika dibandingkan dengan situs *Knowledge Center* Universitas Multimedia Nusantara?

1.3 Batasan Permasalahan

Untuk menghindari perluasan masalah, maka terdapat batasan masalah dalam melakukan penelitian ini, yaitu:

- Data yang digunakan diambil dari kc.umn.ac.id jurusan Informatika
- Judul skripsi yang diambil memiliki rentang waktu dari tahun 2016 sampai dengan 2021
- Data yang diambil berupa Judul Skripsi, Nama Penulis, Tahun publikasi, dan *abstract*
- *Website* dibentuk menggunakan *Flask*
- Server yang digunakan berbentuk *localhost* untuk aplikasi dan *database*
- Corpus bahasa indonesia digunakan yang diambil dari library *nlTK*
- Corpus tambahan yang digunakan adalah : implementasi, rancang, bangun, sistem, penerapan, algoritma, algoritme, aplikasi, pengaruh, metode, penggunaan, pengaruh, analisis, pengembangan, perancangan, klasifikasi, peran.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini dapat dijabarkan menjadi:

- Mengembangkan aplikasi berbasis web yang dapat mencari kemiripan judul skripsi dengan menggunakan TF-IDF dan *Cosine Similarity*.
- Mencari kecepatan dan performa TF-IDF dan *Cosine Similarity* dalam mencari kemiripan judul ketika dibandingkan dengan situs *Knowledge Center Universitas Multimedia Nusantara*

1.5 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah aplikasi yang dibangun diharapkan dapat membantu pengguna untuk mempermudah mencari tahu apakah judul skripsi yang diusulkan mempunyai kemiripan dengan judul skripsi yang telah ada.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika laporan.
- Bab 2 LANDASAN TEORI
Bab ini menjelaskan tentang teori-teori dasar yang berkaitan dengan laporan ini, diantaranya adalah pengertian dari *Natural Language Processing*, *IF-IDF*, *Cosine Similarity*, *Robotic Process Automation* dan *UIPath*.
- Bab 3 METODOLOGI PENELITIAN
Bab ini berisikan tentang metode penelitian yang dilakukan dan diagram penelitian yang menggunakan *flowchart*.
- Bab 4 HASIL DAN DISKUSI
Bab ini menjelaskan tentang hasil implementasi penelitian dengan *source code*, *test scenario*, hasil yang diperoleh, serta evaluasi yang didapat berdasarkan hasil tersebut.
- Bab 5 KESIMPULAN DAN SARAN
Bab ini berisikan kesimpulan dari penelitian dan saran untuk pengembangan yang bersangkutan dengan penelitian ini.

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A