BAB I
PENDAHULUAN

1.1 Latar Belakang

Di tengah berlangsungnya era digital yang masif dan dinamis, teknologi telah menjadi bagian integral dalam setiap aspek kehidupan manusia(Syaputri et al., 2025). Transformasi digital secara fundamental telah mengubah cara kita bekerja, belajar, berkomunikasi, hingga menjalankan bisnis.(Aryani et al., 2025). Dalam konteks ini, teknologi tidak lagi sekadar berfungsi sebagai alat bantu, melainkan telah menjadi tulang punggung bagi produktivitas dan efisiensi aktivitas harian. Salah satu bentuk implementasi teknologi yang paling adaptif dalam mendukung transformasi digital adalah sistem informasi berbasis *website*. Sistem informasi merupakan sekumpulan prosedur atau aktivitas yang mengolah data menjadi informasi yang berguna, sehingga dapat membantu individu dalam membuat keputusan yang tepat.(Afrillia et al., 2022).

Website kini menjadi kanal utama dalam penyebaran informasi yang cepat, luas, dan interaktif, serta memainkan peran sentral dalam mendukung pengambilan keputusan berbasis data. Merujuk pada *Directory of Open Access Scholarly Resources* (*ROAD*) yang disediakan oleh *ISSN International Centre* sebagai sumber acuan dalam mengidentifikasi jurnal *open access* (*OA*) aktif per Mei 2024. Dari total 62.701 jurnal *OA* yang tercantum dalam *ROAD*, sebanyak 6.157 jurnal telah terindeks oleh *Web of Science* (*WoS*), 7.351 jurnal oleh Scopus, dan 34.217 jurnal oleh *OpenAlex*.(Maddi et al., 2025).

Seiring dengan laju pertumbuhan informasi, terutama dalam ranah keilmuan, muncul tantangan baru dalam proses pencarian dan pemanfaatan literatur ilmiah. Ribuan jurnal ilmiah diterbitkan setiap tahunnya oleh berbagai institusi dan publisher di seluruh dunia. Sayangnya, sebagian besar peneliti masih mengakses artikel secara manual melalui portal-portal terpisah seperti *Scopus*, *Semantic Scholar*, dan *Google Scholar*, yang tentunya membutuhkan waktu dan usaha ekstra untuk menemukan artikel yang benar-benar relevan.(Ofori et al., 2024).

Kondisi tersebut mengindikasikan adanya kesenjangan (*gap*) dalam penyediaan platform tunggal yang mampu mengintegrasikan berbagai sumber jurnal ilmiah secara otomatis, cepat, dan cerdas. Beberapa platform pencarian seperti *Semantic Scholar* dan *ResearchGate* memang telah hadir sebagai solusi digital, namun umumnya hanya mengandalkan satu sumber utama atau tidak mengungkapkan secara transparan mekanisme kerja dari algoritma rekomendasinya. Belum banyak sistem yang dirancang secara eksplisit untuk mendukung pencarian literatur berbasis *multi*-sumber dan menyertakan algoritma rekomendasi terbuka seperti *text mining*.(Niculescu et al., 2023).

Dari sisi penelitian, gap riset juga tampak jelas, terutama dalam hal minimnya implementasi sistem agregator jurnal ilmiah yang terintegrasi dengan lebih dari satu sumber (*multi-API*) dan mengadopsi metode rekomendasi terbuka yang dapat diukur dan dievaluasi. Sebagian besar studi terdahulu cenderung terbatas pada integrasi satu *API* (misalnya hanya *Semantic Scholar*), atau hanya berfokus pada pengumpulan data tanpa mengembangkan fitur rekomendasi cerdas yang berbasis isi artikel secara tekstual.

Beberapa contoh sistem yang telah dikembangkan antara lain *Semantic Scholar*, yang menyediakan fitur rekomendasi berbasis *AI* namun bersifat tertutup dan terbatas pada satu sumber data. *Mendeley* menawarkan saran referensi berbasis koleksi pustaka pengguna, namun pendekatannya bersifat personal dan tidak berbasis *text mining* secara terbuka. Sementara *ResearchRabbit* menghadirkan visualisasi hubungan antar artikel dan penulis dalam bentuk grafik interaktif, namun hanya mendukung jurnal dari basis data tertentu seperti *PubMed*, sehingga tidak dapat diandalkan sebagai solusi komprehensif.

Menurut (S et al., 2023) *text mining* merupakan teknik yang digunakan untuk mengekstraksi informasi dan pengetahuan yang tersembunyi dari sekumpulan dokumen teks. Melalui proses pengolahan teks, data yang awalnya tidak terstruktur dapat diubah menjadi informasi yang lebih mudah dianalisis dan dimanfaatkan untuk mendukung pengambilan keputusan.

Menanggapi berbagai keterbatasan tersebut, penelitian ini mengusulkan solusi inovatif berupa perancangan dan implementasi *website agregator* jurnal ilmiah berbasis *multi-API*. Sistem ini dirancang untuk mengintegrasikan beberapa sumber jurnal ilmiah terbuka seperti *Semantic Scholar*, *DOAJ*, dan *OpenAlex*, guna menyediakan cakupan data yang lebih luas, terkini, dan beragam. Lebih dari sekadar penggabungan sumber, sistem ini juga mengimplementasikan algoritma rekomendasi artikel berbasis *text mining* menggunakan pendekatan *Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity*. Algoritma ini dipilih karena kemampuannya dalam mengukur kemiripan semantik antar dokumen berdasarkan kata kunci, sehingga mampu memberikan rekomendasi artikel yang lebih relevan dan kontekstual.

Pengguna dapat dengan mudah melakukan pencarian artikel ilmiah berdasarkan kata kunci tertentu, dan sistem secara otomatis akan menyarankan artikel-artikel serupa berdasarkan isi teks atau *keyword* yang digunakan. Sebagai bagian dari proses validasi, sistem yang telah dirancang dan diimplementasikan akan dievaluasi menggunakan metode *Black Box Testing* (*BBT*) guna untuk mengukur tingkat keberhasilan implementasi.

Dengan pendekatan tersebut, penelitian ini diharapkan dapat menjawab kebutuhan dunia akademik dalam hal efisiensi pencarian literatur ilmiah dan mendukung proses riset seperti *SLR*, pemetaan literatur, hingga penemuan *state of the art* secara lebih efektif. Selain itu, sistem ini juga sejalan dengan *tren open access* dan pemanfaatan teknologi kecerdasan buatan dalam pengembangan sistem informasi berbasis web.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah untuk penelitian ini:

1. Bagaimana merancang dan membangun sistem agregator jurnal ilmiah yang terintegrasi dengan beberapa *API* (*multi API*)?
2. Bagaimana mengimplementasikan fitur rekomendasi artikel menggunakan pendekatan *text mining* (*TF-IDF* dan *Cosine Similarity*)?
3. Bagaimana tingkat keberhasilan sistem yang dibangun?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Merancang sistem agregator jurnal ilmiah yang mampu terintegrasi dengan beberapa *Application Programming Interface* (*API*) dari berbagai sumber artikel ilmiah seperti *Semantic Scholar*, *DOAJ*, dan *OpenAlex*, sehingga memungkinkan pengguna mengakses literatur ilmiah dari berbagai sumber dalam satu platform terpadu.
2. Menguji dan mengevaluasi tingkat fungsionalitas sistem serta tingkat keberhasilan sistem yang telah dibangun melalui pendekatan *Black Box Testing*(*BBT*).
3. Mengembangkan dan mengintegrasikan fitur rekomendasi artikel ilmiah otomatis ke dalam sistem agregator dengan menggunakan pendekatan *text mining*, khususnya metode *Term Frequency–Inverse Document Frequency* (*TF-IDF)* dan *Cosine Similarity*, guna menyajikan artikel-artikel yang relevan berdasarkan input teks atau kata kunci dari pengguna.

1.4 Manfaat Penelitian

Terdapat beberapa manfaat pada penelitian ini dari manfaat teoritis maupun praktis:

1. Memberikan kontribusi pada pengembangan ilmu di bidang sistem informasi, *text mining*, dan *information retrieval*
2. Menjadi referensi penerapan metode *TF-IDF* dan *Cosine Similarity* dalam sistem rekomendasi artikel ilmiah.
3. Bagi Mahasiswa dan Peneliti, dapat mempermudah pencarian dan rekomendasi jurnal dari berbagai sumber dalam satu *platform.*

1.5 Ruang Lingkup dan Batasan Penelitian

1. Penelitian ini hanya menggunakan tiga sumber data *API* terbuka, yaitu *Semantic Scholar*, *OpenAlex*, dan *DOAJ*, yang memungkinkan pengambilan metadata artikel jurnal secara otomatis dan legal tanpa biaya lisensi.
2. Proses rekomendasi artikel dikembangkan menggunakan metode text mining sederhana, yaitu *TF-IDF* (*Term Frequency–Inverse Document Frequency*) untuk ekstraksi bobot kata penting dan *Cosine Similarity* untuk menghitung kemiripan antar artikel.
3. Pengujian sistem dilakukan melalui pengujian fungsionalitas untuk memastikan bahwa seluruh fitur dapat dijalankan dengan benar.
4. Penelitian ini tidak mencakup pengujian performa teknis, seperti kecepatan *respons API*, skalabilitas *server*, atau beban sistem, karena fokus utamanya adalah pada implementasi fitur pencarian dan rekomendasi berbasis konten artikel.
5. Penelitian ini hanya mengambil maksimal 600 jurnal dari masing masing *API* dari 200 *OpenAlex*, 200 *Semantic Scholar*, dan 200 dari *DOAJ*, Jika jurnal tidak sampai 600 maka hanya akan muncul sebanyak yang tersedia saja.
6. Penelitian ini tidak melibatkan pengguna, hanya melakukan pengujian fungsionalitas.
7. Penelitian ini tidak berfokus terhadap tampilan website tetapi berfokus pada implementasi 3 *API*.
8. Penelitian ini menggunakan kata kunci *Artificial Intelligence* untuk menghitung persentase keberhasilan sistem, namun ada beberapa hasil juga untuk kata kunci lain yang ditampilkan dalam bentuk grafik.