

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Buku sebagai sumber informasi konvensional, mempertahankan relevansinya hingga saat ini. Perpustakaan, sebagai wadah utama bagi koleksi buku, memainkan peran integral dalam menentukan maju atau mundurnya suatu institusi, terutama dalam konteks institusi pendidikan. Posisinya yang krusial tercermin dalam Undang-Undang No.43 Tahun 2007 Tentang Perpustakaan, yang menggambarkan perpustakaan sebagai institusi profesional yang mengelola koleksi, karya tulis, karya cetak, dan karya rekam dengan sistem baku untuk memenuhi berbagai kebutuhan, termasuk pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka(UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 43 TAHUN 2007 TENTANG PERPUSTAKAAN, 2007).

Fenomena mengelola koleksi pustaka yang sangat melimpah menjadi suatu tantangan yang dihadapi oleh perpustakaan institusi pendidikan khususnya Perpustakaan Universitas Udayana. Jenis pustaka yang dikelola tidak hanya buku saja namun juga meliputi: jurnal, *prosiding*, ensiklopedia, simposium, laporan penelitian, majalah, *e-book*, atlas, almanak, audio visual (CD-ROM dan DVD-ROM), skripsi, tesis dan disertasi. Koleksi skripsi merupakan salah satu pustaka penting karena jumlahnya yang bertambah dengan cepat tiap tahun dan sering dijadikan bahan referensi penelitian oleh mahasiswa. Anggota perpustakaan khususnya mahasiswa mungkin kesulitan menavigasi dan menemukan skripsi yang sesuai dengan minat atau kebutuhan mereka dalam koleksi yang begitu melimpah.

Selain itu, mahasiswa memiliki waktu terbatas untuk mencari skripsi yang relevan dengan topik studi atau penelitian mereka. Sistem perpustakaan Universitas Udayana masih menggunakan metode kueri *SQL* berbasis perintah “*WHERE LIKE*” untuk mencari skripsi berdasarkan judul pustaka skripsi. Pendekatan ini, meskipun sederhana dan mudah diimplementasikan, memiliki keterbatasan yang signifikan, antara lain:

1. Keterbatasan Relevansi: Metode pencarian pustaka menggunakan kueri *SQL* “*WHERE LIKE*” hanya melakukan pencocokan pola secara literal sehingga seringkali menghasilkan banyak dokumen yang kurang relevan karena tidak memperhitungkan konteks dari judul skripsi;
2. Skalabilitas terbatas untuk dataset besar: Pada koleksi pustaka yang besar, penggunaan “*WHERE LIKE*” cenderung memerlukan sumber data komputasi yang tinggi karena pencarian yang dilakukan adalah *full table scan*. Hal ini menyebabkan kinerja pencarian menurun secara signifikan.

Pustakawan, sebagai penjaga dan pengelola informasi, dihadapkan pada tugas berat untuk menyediakan informasi yang sesuai dengan minat dan kebutuhan anggota perpustakaan. Hal ini dapat mengurangi efektivitas perpustakaan sebagai penyedia informasi yang relevan. Oleh karena itu, perlu adanya inovasi dalam pengelolaan informasi perpustakaan.

Inovasi dalam pengelolaan informasi perpustakaan berbasis rekomendasi diharapkan mampu menavigasi koleksi pustaka yang masif dengan cepat dan akurat. Sistem informasi perpustakaan tidak hanya mengandalkan pencocokan pola literal seperti metode pencarian kueri *SQL* “*WHERE LIKE*”, tetapi juga

mengintegrasikan metode modern sistem temu kembali. Dengan demikian, sistem akan menampilkan hasil yang relevan sesuai dengan kebutuhan anggota perpustakaan, dimana setiap kueri yang dimasukkan akan menghasilkan dokumen yang benar-benar sesuai dengan topik penelitian. Anggota perpustakaan, khususnya mahasiswa, tidak lagi harus menghabiskan banyak waktu mencari referensi, karena sistem akan secara otomatis mengurutkan dokumen berdasarkan relevansi semantik dan konteks internal, meningkatkan efisiensi dan efektivitas dalam pemenuhan informasi. Sistem ini telah sukses diimplementasikan di berbagai bidang sebagai strategi untuk mengelola informasi yang masif dan memberikan rekomendasi yang sesuai dengan preferensi anggota perpustakaan(Haruna et al., 2020; Shang et al., 2024; Sharma et al., 2020; Sofia University "St. Kliment Ohridski", Bulgaria & Berbatova, 2019). Sistem ini juga diharapkan dapat mengurangi beban kerja pustakawan, sehingga mereka dapat lebih fokus pada pengelolaan dan penyediaan informasi strategis lainnya.

Dalam konteks ekstraksi fitur teks untuk sistem rekomendasi skripsi, penggunaan metode *information retrieval* tradisional seperti BM25 masih memiliki keterbatasan dalam memahami konteks semantik dan mengidentifikasi term-term sinonim atau term-term dengan makna yang mirip (El Mahdaouy et al., 2018; Mekontchou et al., 2023). Metode BM25, meskipun efektif dalam banyak kasus, tidak selalu mampu menangkap makna kontekstual dari term-term dalam teks secara menyeluruh. Metode *query expansion* dengan pendekatan *local analysis* mampu mengatasi permasalahan ini. Penggunaan metode ini juga didasari pada (Azad & Deepak, 2019): pertama, kueri yang dimasukkan oleh pengguna mungkin

terlalu pendek untuk mengekspresikan apa yang ingin dicari oleh pengguna. Kedua, meskipun pengguna mengetahui apa yang dicari, pengguna tidak tahu cara merumuskan kueri yang tepat. Ketiga, pengguna seringkali tidak yakin tentang apa yang dicari hingga pengguna melihat hasilnya.

Selain itu, dalam penulisan paragraf bahasa Indonesia, struktur paragraf seringkali bersifat deduktif, di mana konteks atau inti paragraf berada di awal paragraf, atau induktif, di mana konteks atau inti paragraf berada di akhir paragraf. Posisi term dalam suatu paragraf menjadi penting untuk menentukan konteks dari paragraf tersebut. Oleh karena itu, meskipun term ekspansi memiliki *similarity* yang tinggi, posisi term dalam paragraf juga berpengaruh dalam menentukan kesamaan konteks.

Penelitian ini akan mengeksplorasi metode-metode modern seperti Word2Vec untuk meningkatkan pemahaman konteks semantik dalam proses *query expansion* dan penentuan kesamaan dokumen. Selain itu, penelitian ini akan memperhitungkan posisi term atau kata kunci dalam paragraf untuk menentukan bobot relevansi yang lebih akurat. Dengan demikian, diharapkan sistem rekomendasi skripsi yang dihasilkan dapat memberikan rekomendasi yang lebih baik dan relevan, membantu mahasiswa dalam menemukan skripsi yang sesuai dengan topik studi atau penelitian mereka dengan lebih efisien.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka identifikasi masalah yang dijadikan bahan penelitian yaitu sebagai berikut:

1. Sistem Perpustakaan Universitas Udayana masih menerapkan sistem pencarian konvensional dengan perintah “WHERE LIKE” untuk mencari skripsi berdasarkan judul.
2. Metode TF-IDF cendrung memberikan bobot yang lebih tinggi pada dokumen yang panjang hanya berdasarkan frekuensi term, tanpa mempertimbangkan relevansi konteks sebenarnya.
3. Metode BM25 masih memiliki keterbatasan dalam memahami konteks semantik serta mengintegrasikan posisi term pada dokumen. Mengingat posisi term memberikan indikasi pentingnya konteks utama dokumen, terutama dokumen bahasa Indonesia.

### **1.3 Batasan Masalah**

Pembatasan masalah digunakan untuk menghindari adanya penyimpangan dan kejelasan pada hasil studi yang akan diteliti. Hal ini dilakukan agar penelitian ini terlebih terarah dan memudahkan dalam pembahasan sehingga tujuan penelitian akan tercapai. Adapun batasan pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data primer didapatkan dari aplikasi SIMAK (Sistem Informasi Akademik) Universitas Udayana;
2. Perancangan sistem rekomendasi berbasis aplikasi website;
3. Penggunaan metode Word2Vec untuk menghasilkan ekspansi kueri;
4. Penggunaan BM25 untuk perhitungan relevansi dokumen;
5. Sistem rekomendasi akan dijadikan fitur dalam Aplikasi E-Perpustakaan Universitas Udayana;

6. Fokus pada sistem rekomendasi untuk skripsi, tidak termasuk jenis dokumen akademis lainnya;
7. Korpus dokumen dan proses preprocessing difokuskan pada term bahasa Indonesia;
8. Proses tokenisasi belum menangani term dengan tanda baca khusus atau simbol.

#### **1.4 Rumusan Masalah**

Berdasarkan latar belakang di atas, maka dapat dirumuskan butir-butir rumusan masalah sebagai berikut:

1. Bagaimana prosedur dalam mengembangkan metode rekomendasi skripsi berbasis BM25 dengan bobot posisi kata adaptif?
2. Bagaimana performansi metode yang dikembangkan pada berbagai skenario pengujian?

#### **1.5 Tujuan Penelitian**

Dari rumusan masalah yang telah dipaparkan di atas tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Mengembangkan prosedur untuk membangun metode rekomendasi skripsi berbasis BM25 dengan bobot posisi kata adaptif.
2. Menguji dan menganalisis performa metode yang diusulkan pada berbagai skenario pengujian menggunakan metrik evaluasi *MAP*, *Precision*, *Recall* dan *F1-Score*.

## **1.6 Manfaat Penelitian**

Penelitian ini diharapkan mampu memberikan manfaat secara akademis maupun secara praktis. Berikut ini merupakan manfaat dari penelitian ini.

### 1. Manfaat akademis

Penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam bidang *Information Retrieval* dan *Document Retrieval*. Kontribusi yang ditawarkan berupa modifikasi BM25 melalui penambahan bobot posisi kata adaptif. Selain itu terdapat kolaborasi metode dengan *query expansion local analysis* menggunakan Word2Vec. Hasil penelitian dapat dijadikan referensi bagi akademisi dan peneliti yang tertarik untuk mengkaji pengaruh posisi term terhadap relevansi dokumen, khususnya pada dokumen berbahasa Indonesia.

### 2. Manfaat praktis

Metode yang dikembangkan dapat diimplementasikan sebagai fitur rekomendasi skripsi, sehingga memudahkan pustakawan dalam menawarkan skripsi yang bisa dijadikan rekomendasi mahasiswa dalam melakukan penelitian.