

1. Pendahuluan

1.1 Latar Belakang

Kebutuhan manusia akan informasi di masa teknologi semakin meningkat, seiring dengan meningkatnya kecerdasan manusia secara keseluruhan. Semakin meningkatnya kecerdasan manusia, semakin membutuhkan informasi secara cepat dan akurat sesuai dengan kebutuhannya saat itu. Dalam upaya untuk memenuhi kebutuhan tersebut, telah hadir berbagai media untuk mendapatkan informasi dan dapat diakses dimana saja yaitu dengan internet melalui *platform* komputer, laptop, tablet, dan *smartphone*. Namun bagaimana jika terdapat ratusan ribu informasi yang berbeda, sedangkan informasi yang dibutuhkan spesifik. Untuk itu dikembangkanlah cara untuk mengatasi permasalahan ini, yaitu *Information Retrieval*.

Information Retrieval adalah proses pencarian informasi dari sekumpulan dokumen yang sesuai dengan informasi yang dibutuhkan. Proses yang terjadi dalam *Information Retrieval* pada dasarnya untuk memenuhi keinginan pengguna akan informasi dengan cara menentukan dokumen dalam sekumpulan dokumen yang harus ditemubalikkan dan bentuk *query* merupakan representasi dari informasi yang diinginkan pengguna[2]. Salah satu metode yang dapat mengatasi ini adalah LSI (*Latent Semantic Indexing*). Metode ini mempresentasikan term dari dokumen menjadi vektor yang berbentuk *term* dokumen matriks sehingga dapat menghubungkan antara *term* dengan dokumen yang dicari walaupun *query* yang dimasukan tidak terdapat dalam dokumen yang dicari. Dari hasil penelitian sebelumnya menyatakan bahwa LSI menggunakan metode QR *Decomposition* dengan teknik *gram schmid* memiliki nilai *recall* dan *precesion* yang baik akan tetapi matriks ortogonalnya yang tidak stabil yang mengakibatkan akurasiya berkurang[13]. Oleh karena itu, untuk mendekomposisi matriks tersebut penulis akan menggunakan metode QR *decomposition* dengan transformasi *Householder*. Dekomposisi QR (juga disebut QR faktorisasi) dari matriks adalah dekomposisi matriks A menjadi produk $A = QR$. Dekomposisi QR adalah salah satu alat dasar

dalam komputasi statistika dan digunakan untuk mencari matriks *orthogonal* Q dan matriks segitiga atas R.

1.2 Rumusan masalah

Berdasarkan latar belakang, dapat dirumuskan masalah sebagai berikut :

1. Bagaimana mengimplementasikan *Latent Semantic Indexing* dalam *Information Retrieval*.
2. Bagaimana membangun sistem LSI menggunakan *QR Decomposition* dengan transformasi *Householder* dalam menemukan dokumen yang relevan terhadap sebuah *query* yang dimasukkan oleh user dengan dataset yang ada.
3. Bagaimana menganalisis performansi *Recall* dan *Precision* dan waktu proses pencarian dokumen yang *relevan*.

1.3 Tujuan

Tujuan yang ingin dicapai dalam Tugas Akhir ini adalah :

1. Mengimplementasikan *Latent Semantic Indexing* pada *Information Retrieval*.
2. Menganalisis kemampuan LSI dengan metode *QR Decomposition* dengan Transformasi *Householder* dalam pencarian dokumen yang relevan walaupun dalam dokumen tidak mengandung *term* dari *query*.
3. Menganalisis performansi *Recall* dan *Precision* dan waktu proses dari metode dalam menemukan dokumen relevan dengan *query*.

1.4 Batasan Masalah

Batasan permasalahan yang terdapat pada Tugas Akhir ini adalah :

1. Koleksi dataset yang digunakan *cranfield collection* berbahasa inggris yang diakses dari situs http://ir.dcs.gla.ac.uk/resources/test_collections/cran/ dalam format *text*.
2. Melakukan proses *stop list* pada tahap *indexing* untuk menghilangkan kata yang tidak memiliki makna dan jumlah kemunculannya sangat sering terjadi pada setiap dokumen.
3. Tidak melakukan proses *stemming* supaya dapat meningkatkan relevansi terhadap *Latent Semantic Indexing* dalam *Information Retrieval*.

4. Melakukan skema pembobotan Tf-Idf sebelum melakukan dekomposisi matriks.
5. Menggunakan Transformasi *Householder* dalam *QR Decomposition*.
6. Menggunakan *cosine similarity* untuk perbandingan kemiripan dokumen yang ada dalam dataset dengan *query*.
7. Menggunakan parameter *recall* dan *precision* untuk mengukur akurasi sistem.

1.5 Metodologi Penyelesaian Masalah

Pada Tugas Akhir ini penulis menerapkan beberapa metode untuk menyelesaikan masalah sebagai berikut:

1. Studi literatur

Mempelajari konsep-konsep *Latent Semantic Indexing*, *QR Decomposition*, yang terdapat pada beberapa sumber literatur. Sumber literatur dapat berupa buku teks, *paper*, *website*, blog, jurnal, dan dokumen teknis sistem terdahulu.

2. Koleksi data

Menggunakan data *cranfield collection* berbahasa inggris yang diakses dari situs http://ir.dcs.gla.ac.uk/resources/test_collections/cran/ dalam format *text*.

3. Perancangan Sistem

Merancang sistem menggunakan konsep *Latent Semantic Indexing* dengan metode *QR Decomposition* dengan menentukan terlebih dahulu parameter-parameter yang diuji, deskripsi sistem, spesifikasiperangkat lunak sesuai untu kebutuhan pengujian metode.

4. Implementasi

Mengimplementasikan *Latent Semantic Indexing* menggunakan metode *QR Decomposition* dengan Transformasi *Householder* dalam menemukan dokumen yang relevan, yaitu *cosine similarity*.

5. Pengujian dan Analisis

Penulis akan menguji dan menganalisis *cosine similarity* metode *QR Decomposition* dengan Transformasi *Householder* dalam LSI dan mengukur akurasi dalam parameter *recall* dan *precision*.

