

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Salah satu syarat untuk menyelesaikan studi di Universitas Islam Negeri Sutan Syarif Kasim Riau (Uin Suska) yaitu mahasiswa wajib melaksanakan Tugas Akhir atau Skripsi. Berdasarkan hasil kuisisioner (Lampiran A), 88% mahasiswa mengalami kesulitan untuk menemukan ide Tugas Akhir. 54% mahasiswa meminta saran kepada dosen dan 46% mahasiswa membaca literatur untuk mendapatkan ide Tugas Akhir. Ketika meminta saran ke dosen untuk ide Tugas Akhir, mahasiswa diharuskan untuk membaca beberapa literatur, untuk didiskusikan dengan dosen untuk mendapatkan ide Tugas Akhir yang cocok.

Mahasiswa mendapatkan ide Tugas Akhir dengan melakukan penelusuran literatur atau pustaka yang berhubungan dengan Tugas Akhir. Penelusuran pustaka merupakan langkah pertama untuk mengumpulkan informasi yang relevan bagi penelitian. Penelusuran pustaka berguna untuk menghindari duplikasi dari pelaksanaan penelitian. Tinjauan literatur sangat penting perannya dalam membuat suatu tulisan atau karangan ilmiah. Tinjauan literatur dapat memberikan ide dan tujuan tentang topik penelitian yang akan dilakukan (Zainal A. Hasubuan, 2007)

Literatur yang sering digunakan mahasiswa untuk Tugas Akhir yaitu Jurnal Elektronik atau e-Jurnal. Jurnal versi elektronik atau e-Jurnal lebih banyak memiliki nilai lebih dari segi kemutakhiran, penyimpanan, pemanfaatannya, serta dari waktu penelusuran yang cepat (Rusydi, 2014). E-Jurnal didapatkan dengan melakukan pencarian pada *search engine*. E-Jurnal yang banyak terdapat pada internet dapat menyulitkan *user* untuk menemukan dokumen yang relevan. Berdasarkan hasil Kuisisioner (Lampiran A), 95% mahasiswa tidak mendapatkan hasil pencarian yang diinginkan dan 71% mahasiswa hanya memperhatikan hasil pencarian pada halaman pertama. *User* bisa saja melewatkan hasil yang relevan. Kebanyakan e-Jurnal pada hasil pencarian, tidak dikelompokkan sehingga mempersulit mahasiswa untuk mendapatkan e-Jurnal yang tepat dan sesuai dengan penelitian yang diteliti. Hasil pencarian paling awal belum tentu relevan sehingga dibutuhkan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pengelompokan e-Jurnal untuk mendapatkan hasil yang relevan dan sesuai dengan jurnal elektronik yang dibutuhkan.

Konsep umum *search engine* adalah saat pengguna mencari *query*, daftar dokumen yang relevan akan ditampilkan, pengguna harus memindai daftar ini dari atas ke bawah untuk mendapatkan hasil yang diinginkan. Pada umumnya hasil dari sistem pencarian menampilkan daftar panjang dari dokumen. Daftar panjang dokumen yang terbentuk dari hasil pencarian terkadang membingungkan dan menyulitkan *user* untuk menemukan dokumen yang tepat dan tentunya membutuhkan waktu yang lama (Handojo, Wibowo, & Santo, 2011).

Hasil pencarian dikelompokkan dalam beberapa label *cluster* dengan menerapkan *Clustering*. Pengguna memindai semua label *cluster* dan kemudian memilih *cluster* dan hanya memindai hasil pencarian pada *cluster* tersebut. Keseluruhan waktu yang dibutuhkan untuk memindai keseluruhan hasil pencarian akan berkurang (Jejurkar, 2016). Efektivitas *Information Retrieval* dapat ditingkatkan dengan penerapan dokumen *Clustering*. Penerapan *Clustering* ini mengikuti *cluster-hypothesis* yaitu dokumen yang relevan akan cenderung berada pada *cluster* yang sama.(C. J. Van Rijsbergen, 1979)

*Information Retrieval* (IR) adalah menemukan materi (biasanya dokumen) dari sifat tidak terstruktur yang memenuhi kebutuhan informasi dari dalam koleksi yang besar (Manning, Raghavan, & Schütze, 2009). Mayoritas sistem *information retrieval* memberikan daftar panjang peringkat web, kebanyakan dari hasil tersebut tidak relevan. *Precision* yang rendah dari mesin pencari web ditambah dengan presentasi daftar peringkat yang panjang membuat sulit bagi pengguna untuk menemukan informasi yang dicari. Solusi untuk meningkatkan kualitas pencarian adalah pengelompokan hasil pencarian. Sistem akan menampilkan kepada pengguna daftar panjang hasil yang dibagi menjadi beberapa kelompok yang terkait dengan sub-topik (Nirkhi & Hande, 2008).

Berdasarkan permasalahan di atas maka dibangun mesin pencari *clustering* yang dapat menampilkan hasil pencarian yang relevan dengan pengelompokan (*clustering*) hasil pencarian. Penelitian ini menggunakan metode *Vector Space Model* untuk sistem IR dan metode *Suffix Tree Clustering* untuk *clustering* hasil

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

pencarian. Menurut Amin (2012) Solusi untuk mengatasi masalah sistem temu kembali informasi dengan hasil *recall* tinggi dan tingkat akurasi rendah adalah dengan membuat sistem temu kembali informasi menggunakan metode *VSM*. Metode *VSM* dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada *document-matching*. Berdasarkan penelitian Irmawati (2017) dengan judul Sistem Temu Kembali Informasi Pada Dokumen dengan Metode *Space Vector Model* dan pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* menghasilkan sistem temu kembali dengan kinerja yang cukup baik dengan rata-rata nilai *Recall* hampir 100% dan rata-rata nilai *precision* sekitar 73.6%, sehingga 73.6% dokumen *online* yang berhasil ditemu-kembalikan relevan dengan *keyword* yang diberikan. *TF-IDF* merupakan algoritma yang efisien untuk mencocokkan kata pada *query* ke dokumen yang relevan dengan *query* tersebut, *TF-IDF* dapat mengembalikan dokumen yang sangat relevan dengan *query*. (Ramos, 1999). Algoritma *Stemming* yang digunakan yaitu *Porter Stemmer*. Kelebihan *Porter Stemmer* yaitu prosesnya yang cepat dan sederhana (Porter, 1980).

Menurut Ilic et al. (2014) algoritma *Suffix Tree Clustering (STC)* yaitu algoritma klusterisasi yang memiliki waktu linier dalam pengelompokan dokumen hasil pencarian dalam bentuk grup-grup atau kluster berdasarkan kata atau frase yang terdapat dalam dokumen yang ada. Berdasarkan penelitian Handojo, dkk (2011) yang berjudul *Clustering Search Engine At Petra Christian University Library Using Suffix Tree Clustering*, penelitian ini membangun *clustering search engine* pada perpustakaan universitas menghasilkan bahwa *cluster* yang dihasilkan oleh sistem cukup baik, dalam arti untuk mengklasifikasikan dokumen yang berkorelasi satu sama lain atau memiliki klasifikasi yang sama.

Berdasarkan penjelasan yang telah diuraikan sebelumnya, maka akan dibangun Sistem Temu Kembali Informasi atau *Information Retrieval (IR)* dengan *clustering* hasil pencarian menggunakan metode *Vector Space Model* dan metode *Suffix Tree Clustering*. Sistem ini diharapkan dapat membantu mahasiswa sebagai pengguna untuk mendapatkan hasil pencarian e-Jurnal yang relevan dengan waktu yang singkat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, rumusan masalah yang dapat diambil yaitu “Bagaimana mengimplementasikan *Vector Space Model* dan *Suffix Tree Clustering* pada sistem temu kembali informasi dan *clustering e-Jurnal*?”

## 1.3 Batasan Masalah

Dalam sebuah penelitian dibutuhkan batasan masalah agar penelitian tidak menyimpang. Batasan masalah dalam penelitian ini yaitu:

1. Jurnal yang digunakan jurnal berbahasa Inggris
2. Metode *Vector Space Model* menggunakan pembobotan TF-IDF (*Term Frequency Inverse Document Frequency*)
3. Algoritma *stemming* yang digunakan yaitu algoritma *Porter Stemmer*.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah dapat dibangun Sistem Temu Kembali Informasi dan *clustering e-Jurnal* dengan Metode *Vector Space Model* dan *Suffix Tree Clustering*, yang dapat menampilkan hasil pencarian yang sudah terkelompokkan yang relevan sehingga dapat membantu *user* untuk mendapatkan e-Jurnal yang dibutuhkan.

## 1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini diatur sedemikian rupa sehingga segala kebutuhan yang dipergunakan di dalam penelitian dapat dipahami dengan mudah. Sistematika penulisan laporan Tugas Akhir ini adalah sebagai berikut:

### BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan laporan.

### BAB II LANDASAN TEORI

Bab ini menjelaskan dasar-dasar teori yang digunakan sebagai landasan dalam pembuatan Tugas Akhir seperti: Pengertian Sistem Temu Kembali Informasi (*Information Retrieval*),

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Komponen Sistem Temu Kembali Informasi (*Information Retrieval*), *Porter Stemmer*, *Vector Space Model*, *Term Frequency Inverse Document Frequency*, *Clustering*, *Suffix Tree*, *Suffix Tree Clustering*, Pengujian *Black Box*, Akurasi Sistem Temu Kembali Informasi (*Information Retrieval*), *Cluster Label Quality* dan Penelitian Terkait.

### **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan metodologi yang digunakan dalam penelitian yaitu Pendahuluan, Analisa, Perancangan, Implementasi, Pengujian dan Penutup

### **BAB IV ANALISA DAN PERANCANGAN**

Bab ini menjelaskan analisa dan perancangan yang dilakukan pada penelitian yaitu Analisis Masalah, Analisis Sistem, Analisis Proses *Information Retrieval*, Pengelompokan Hasil Pencarian, Perancangan Sistem dan Perancangan Antarmuka

### **BAB V IMPLEMENTASI DAN PENGUJIAN**

Bab ini menjelaskan implementasi dan pengujian yang dilakukan pada penelitian yaitu Implementasi, Implementasi *Interface* Sistem, Pengujian Kemampuan Sistem, Pengujian *Black Box* dan Kesimpulan Pengujian Sistem

### **BAB VI PENUTUP**

Penutup merupakan kesimpulan dari penelitian serta saran untuk penelitian selanjutnya.