

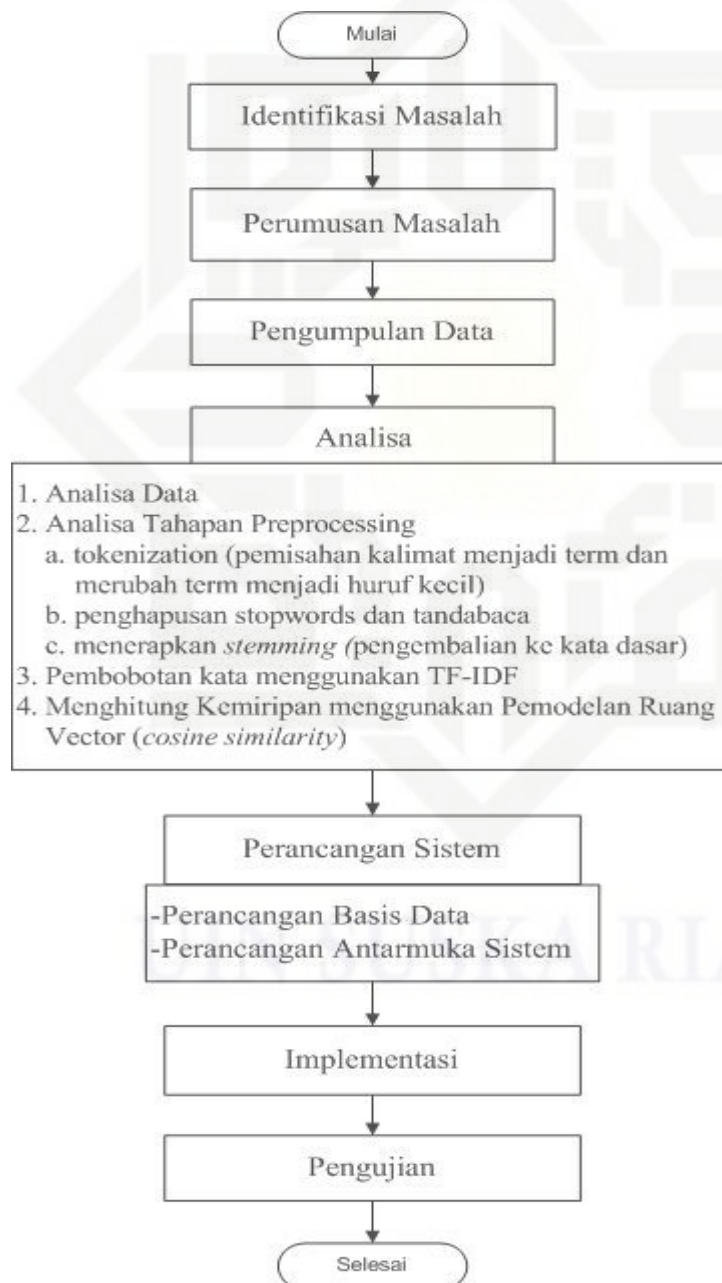
Hak Cipta Dilindungi Undang-Undang

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.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian menjelaskan bagaimana langkah-langkah yang akan dilakukan dalam penelitian untuk dapat menjawab perumusan masalah penelitian. Tahapan penelitian dalam tugas akhir ini dapat dilihat pada Gambar 3.1 berikut:



Gambar 3.1. Tahapan Penelitian

3.1 Identifikasi Masalah

Tahapan ini menjadi penting agar pengguna memperoleh terjemahan dalam Al-Qur'an yang relevan dengan *query* agar menemukan informasi yang dicari dengan cepat dari sekumpulan dokumen yang ada dalam *database*.

3.2 Perumusan Masalah

Dari tahapan identifikasi masalah, maka penulis mencoba merumuskan permasalahannya yaitu dapat menemukan kembali terjemahan dalam Al-Qur'an yang relevan terhadap *query* pengguna dengan membangun sebuah aplikasi yang dapat memberikan dokumen secara terurut dengan menggunakan pemodelan ruang vektor dan algoritma *Enhanced Confix Stripping (ECS)*.

3.3 Pengumpulan Data

Teknik pengumpulan data dapat didefinisikan sebagai suatu cara yang digunakan untuk memperoleh data yang dibutuhkan sebagai bahan masukan bagi penulis dalam penyusunan penelitian ini. Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari penelitian Wulandari (2013) berupa terjemahan ayat Al-Qur'an dalam bahasa Indonesia. Wulandari (2013) merujuk terjemahan Al-Qur'an bahasa Indonesia versi Departemen RI yang diakses melalui situs <http://www.qurandownload.com/> dan kemudian dipetakan kedalam 800 dokumen. Setiap dokumen berisi gabungan beberapa terjemahan ayat secara utuh. Sebagai contoh, pada Surat Al-Baqarah ayat 1 sampai dengan 20 dijadikan 3 dokumen, yaitu ayat 1-5, ayat 6-7 dan ayat 8-20 dan selanjutnya hingga ayat terakhir pada surah yang terakhir.

3.4 Analisa

Pada tahapan ini akan dijelaskan bagaimana sistem yang akan dibuat bisa mengurutkan dokumen yang terkait dengan query berdasarkan derajat kemiripan

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.

yang lebih tinggi ke derajat kemiripan yang lebih rendah. Maka dibutuhkan analisa data, analisa tahapan *preprocessing*, pembobotan kata dan analisa pemodelan ruang vektor.

3.4.1 Analisa Data

Melakukan analisa data masukan (*input*) terhadap sistem dan analisa data keluaran (*output*) oleh sistem yang diharapkan relevan dengan kebutuhan pengguna.

3.4.2 Analisa Tahapan Preprocessing

Melakukan analisa tahapan-tahapan awal yang akan dilakukan dalam pembangunan *information retrieval system* sebelum diterapkan Model Ruang Vektor dalam pengukuran kesamaan (*similarity measure*). Menurut Darmawan (2011), tahapan-tahapan tersebut adalah :

- a. Melakukan *tokenization*.

Melakukan pemisahan kalimat menjadi kata dan pengubahan huruf besar menjadi huruf kecil. Penjelasan langkah ini adalah dengan memisahkan kalimat dalam terjemahan ayat Al-Qur'an menjadi kata-kata tunggal dan merubah kata tunggal tersebut kedalam huruf kecil dan kemudian disimpan dalam variabel. Kata yang tersimpan dalam variabel akan menjadi *tokenTerm* (Darmawan, 2011).

- b. Penghapusan *stopword* dan tandabaca.

Proses penghapusan *stopwords* dilakukan untuk kata-kata yang dianggap tidak mencerminkan isi dari dokumen. Yaitu kata sambung, kata ganti, kata yang berulang. Langkahnya adalah mencocokkan tiap kata dalam *tokenTerm* dengan daftar *stopword* dan tandabaca di *database*. Kata yang tidak cocok dengan *stopword* di simpan dalam variabel (Darmawan, 2011).

Menerapkan *Stemming* (pengembalian ke kata dasar).

Proses *stemming* dilakukan agar kata (*token*) menjadi bentuk dasar. *Stemming* melakukan pembuangan bentuk awalan dan akhiran pada kata dan kemudian membandingkan kata yang ada dalam terjemahan Al-Qur'an dengan daftar kamus (kata dasar) yang telah dipersiapkan terlebih dahulu. *Stemming* yang penulis gunakan dalam penelitian ini adalah *Enhanced Confix Stripping (ECS)*. Langkah dari ECS adalah mencocokkan *term* pada elemen di *index* dengan kata dasar. Jika cocok *term* disimpan dalam *stemTerm*. Jika tidak lakukan *rulePrecedence* yaitu pembuangan awalan dan akhiran. Lakukan *reduksiAwalan* dan cocok kembali dengan kamus kata dasar, jika cocok *term* disimpan, jika tidak lakukan *reduksiAkhiran* dan lakukan pencocokan kembali dengan kamus kata dasar, jika cocok maka *term* akan disimpan jika tidak maka *term* pada elemen *index* lah yang akan disimpan (Darmawan, 2011).

3.4.3 Pembobotan Kata Menggunakan TF-IDF

Proses ini dilakukan untuk memberikan bobot pada setiap *term*. Pembobotan yang dilakukan adalah pembobotan lokal dan global yang dikenal dengan *tf-idf*. Cara kerja pembobotan ini dengan menghitung frekuensi kemunculan *term*. *Tf* yaitu menghitung *term* yang terdapat dalam dokumen, *idf* menghitung dokumen yang mengandung *term* (Darmawan, 2011).

3.4.4 Menghitung kemiripan menggunakan Pemodelan Ruang Vektor

Memanfaatkan Pemodelan Ruang Vektor yang dapat merepresentasikan *bag of words* dan dikonversi kedalam suatu model ruang vektor. Dimana model ruang vektor ini akan membandingkan antara dokumen satu dengan dokumen yang lainnya berdasarkan *query*. Menurut Cios (2007) terdapat empat fungsi untuk mengukur kemiripan (*similarity measure*) yang digunakan oleh pemodelan ruang vektor yaitu *Cosine similarity*, *Inner similarity*, *Dice similarity*, *Jaccard similarity*

Penelitian ini akan menggunakan *cosine similarity* untuk menghitung nilai cosinus sudut antar vektor. Penggunaan fungsi *cosine similarity* dipilih karena

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.

menurut Hamzah (2008), *cosine* memiliki tingkat efisiensi komputasi yang lebih baik dibanding fungsi yang lain yaitu dalam hal kinerja dan waktu komputasi.

3.4 Perancangan Sistem

Tahap perancangan sistem temu kembali informasi merupakan tahapan dalam membuat rincian sistem temu kembali informasi berdasarkan analisa agar dapat dimengerti oleh pengguna (*user*). Rancangan utama yang akan dilakukan, yaitu:

3.5.1 Perancangan Basis Data

Transformasi koleksi dokumen ke dalam basis data agar dapat digunakan oleh sistem yang menerapkan langkah-langkah dalam analisa fungsional.

3.5.2 Perancangan Antarmuka Sistem

Perancangan antarmuka sistem (*interface*) yang baik sehingga mudah digunakan (*user friendly*). Untuk mendukung pengguna memudahkan mencari informasi yang dibutuhkan.

3.5 Implementasi

Pada tahapan implementasi ini akan dilakukan pembuatan sistem yang telah dirancang dalam tahap perancangan kedalam bahasa pemrograman. Implementasi sistem akan dilakukan dengan spesifikasi perangkat sebagai berikut:

Perangkat Keras

Processor : Intel(R) Core(TM) i3-2330M CPU @ 2.20GHz

Memori (RAM) : 2 GB

Perangkat Lunak

Sistem Operasi : Windows 8 Pro 32-bit

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.

Bahasa Pemrograman	: PHP 5.4.7
Web Servers	: Apache 2.4.3
Data Base	: Mysql 5.5.27
Web Browser	: Google Chrome 54.0

3.6 Pengujian

Pengujian merupakan tahapan dimana sistem akan dijalankan. Tahap pengujian diperlukan untuk menjadi ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan. Penelitian ini melakukan pengujian terhadap kata dengan ekspansi *query* dan kata tanpa ekspansi *query* dengan menghitung *precision* dan *recall* terhadap pengujian tersebut. *Precision* adalah nilai relevansi antara dokumen dan *query* atau dengan kata lain adalah nilai akurasi dokumen berdasarkan *query*. Sedangkan *recall* adalah nilai dari jumlah dokumen yang dapat ditemukan kembali dari seluruh jumlah koleksi dokumen berdasarkan *query*.