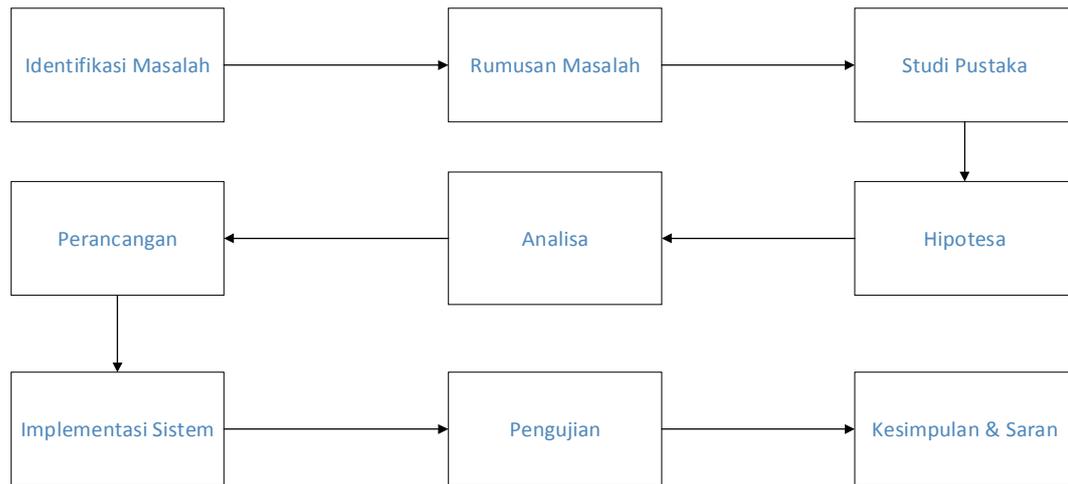


BAB III

METODOLOGI PENELITIAN

Pada penelitian tugas akhir ini ada beberapa tahapan penelitian yang akan dilakukan seperti yang terlihat pada gambar 3.1:



Gambar 3.1. Tahapan Penelitian

3.1 Identifikasi Masalah

Pada tahapan ini dilakukan identifikasi permasalahan bahwa sulitnya mendapatkan sumber referensi dari sebuah dokumen yang diduga melakukan plagiarisme terhadap dokumen lainnya serta terbatasnya melakukan pendeteksian plagiarisme dengan perbandingan 1:1.

3.2 Rumusan Masalah

Dari tahapan identifikasi masalah, dapat dirumuskan Bagaimana mendeteksi sumber dari sebuah dokumen teks untuk melakukan pendeteksian plagiarisme terhadap banyak dokumen berdasarkan *word phrasing* menggunakan model ruang vektor.

3.3 Studi Pustaka

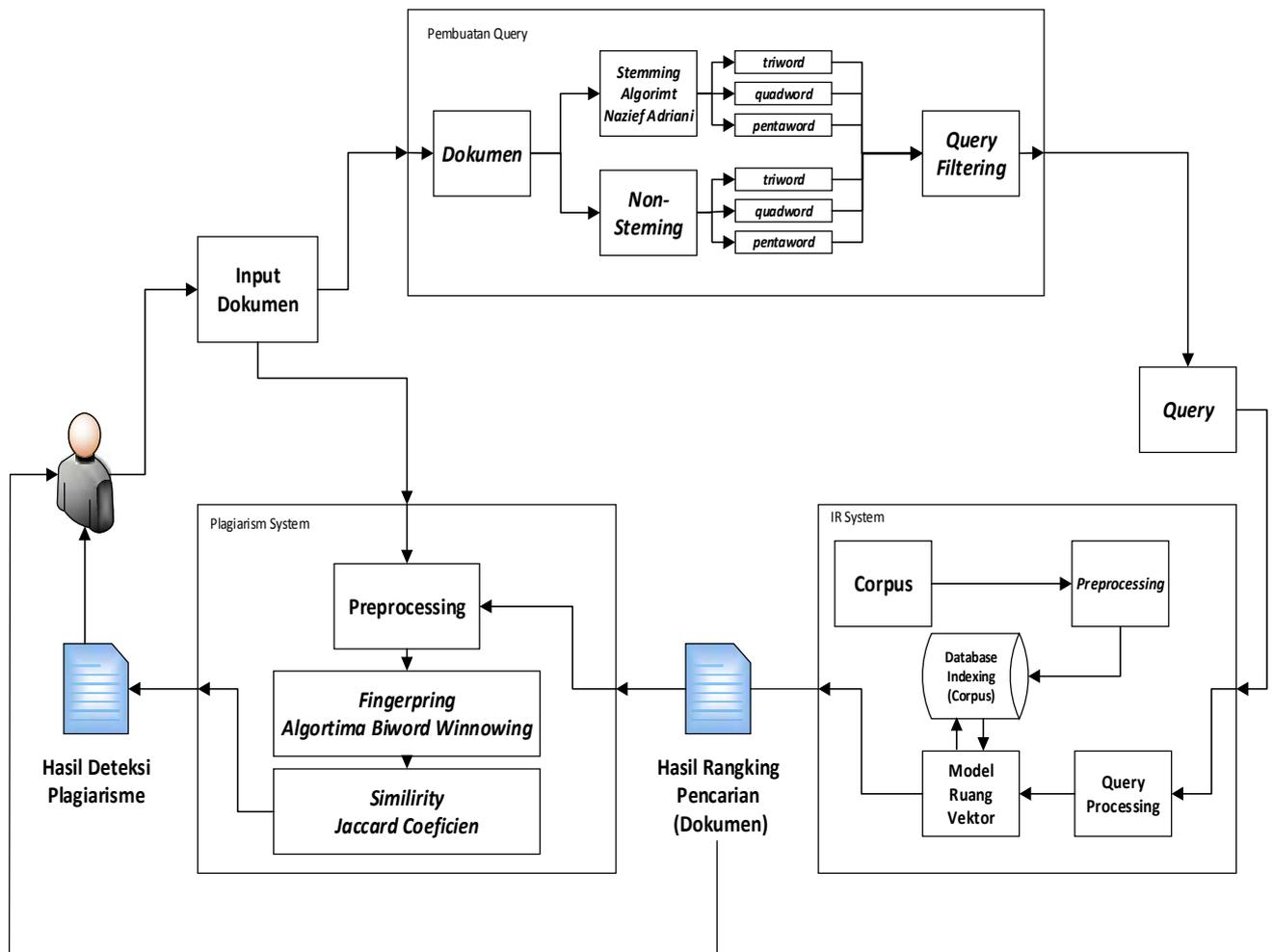
Pada tahap ini dilakukan pengumpulan data dengan cara mencari referensi-referensi terkait yang dibutuhkan untuk penelitian. Referensi tersebut dapat berupa buku-buku, jurnal-jurnal, tulisan penelitian dan juga artikel-artikel dari internet yang memiliki kaitan dengan kasus yang sedang dilakukan dalam penelitian.

3.4 Hipotesa

Metode ini dilakukan dengan memahami pembentukan *query* berdasarkan *words phrasing* dan Algoritma *Stemming* Nazief Adriani, sehingga diperoleh suatu hipotesa berupa pendekatan token berbentuk *triword*, *quadword* dan *pentaword* serta proses *stemming* dan *non-stemming*. Hipotesa awal, pendekatan token berbentuk *triword* dengan *stemming* dengan frekuensi tertinggi bekerja lebih baik dibanding token berbentuk *quadword* dan *pentaword stemming* maupun *non-stemming* dalam pembentukan *query* yang untuk menghasilkan dokumen relevan dan memiliki nilai *jaccard coefficient* yang tinggi.

3.5 Analisa

Analisa adalah tahapan yang dilakukan setelah pengumpulan data dari penelitian tugas akhir ini. Analisa berarti metode khusus yang digunakan untuk menganalisis masalah pada pendeteksian sumber untuk melakukan pendeteksian plagiarisme terhadap banyak dokumen. Analisa ini dilakukan untuk mengetahui langkah-langkah apa saja yang akan dilakukan untuk mendapatkan hasil sesuai dengan hipotesa yang diberikan.



Gambar 3.2 Analisa Kerangka Teoritis Penelitian

Berdasarkan gambar 3.2 terdapat tiga proses utama yang akan dilakukan pada penelitian *source detection* ini, diantaranya:

1. Analisa pembentukan *query* dari dokumen yang dicurigai plagiat berdasarkan *word phrasing*. Adapun *query* yang akan dibentuk, dibedakan menjadi dua yaitu :
 - a. *Query* dengan menggunakan *stemming*, pada tahap ini akan dilakukan :
 1. Pembersihan teks dokumen.
 2. Menghilangkan angka pada dokumen.

3. Melakukan pemotongan teks dokumen kedalam token tunggal, serta melakukan pembuangan *stopwords* dan *stemming*, kemudian menyusun token menjadi *triword*, *quadword* dan *pentaword*
 4. Melakukan perhitungan frekuensi kemunculan *triword*, *quadword* dan *pentaword* yang sama.
 5. Membentuk masing-masing *query* berdasarkan ranking frekuensi yaitu 5 *triword*, *quadword* dan *pentaword* dengan frekuensi tertinggi, 5 frekuensi terendah dan 5 frekuensi tengah.
- b. *Query* dengan menggunakan tanpa *stemming*, pada tahap ini akan dilakukan :
1. Pembersihan teks dokumen.
 2. Menghilangkan angka pada dokumen.
 3. Melakukan pemotongan teks dokumen kedalam token tunggal, serta melakukan pembuangan *stopwords* kemudian menyusun token menjadi *triword*, *quadword* dan *pentaword*.
 4. Melakukan perhitungan frekuensi kemunculan *triword*, *quadword* dan *pentaword* yang sama.
 5. Membentuk masing-masing *query* berdasarkan ranking frekuensi yaitu 5 *triword*, *quadword* dan *pentaword* dengan frekuensi tertinggi, 5 frekuensi terendah dan 5 frekuensi tengah.

2. Analisa *Information Retrieval System*

Secara garis besar, ada tiga tahapan yang ditangani oleh sistem ini, yaitu melakukan *preprocessing* terhadap dokumen, melakukan *preprocessing* terhadap *query* dan menerapkan metode tertentu dalam hal ini menggunakan model ruang vektor untuk menghitung kedekatan (relevansi / *similarity*) antara dokumen dan *query* hasil pembuatan di proses sebelumnya. Adapun tiga tahapan tersebut, yaitu:

a. *Preprocessing*

Melakukan analisa terhadap tahapan *preprocessing* untuk koleksi dokumen (korpus) untuk Pembangunan *index* untuk diterapkan pada model ruang vektor. Adapun tahapan pembangunan *index* sebagai berikut.

1. Mengumpulkan dokumen yang akan di-*index*
2. Penghapusan format khusus dan markup dari dalam dokumen
3. Pemisahan rangkaian kata (*tokenization*)
4. *Linguistic Preprocessing* : Penghapusan *stop words*
5. *indexing*

b. *Preprocessing Query*

Melakukan *preprocessing* terhadap *query* yang bertujuan untuk membersihkan *query* hingga pembobotan, adapun proses yang akan dilakukan pada tahap ini adalah sebagai berikut:

1. Menghilangkan tanda baca pada dokumen
2. Mengubah dokumen kebentuk huruf kecil
3. Menerapkan *stopword removal*
4. Menerapkan *stemming* (mengembalikan kata ke kata dasar)
5. Pembobotan setiap istilah *query*

Pembobotan pada *query* ini mengacu dari hasil *indexing* pada *preprocessing* dokumen.

c. Penerapan Model Ruang Vektor

Melakukan analisa terhadap persamaan model ruang vektor untuk menghitung nilai korelevanan dokumen terhadap *query* yang dibentuk dari dokumen inputan sehingga dokumen yang di rangkingkan berdasarkan nilai korelevanan yang diperoleh.

3. Analisa Deteksi kemiripan dokumen (*plagiarism system*) dengan Algoritma *biword winnowing*

a. Pembentukan *fingerprint* dokumen dengan algoritma *biword winnowing*. Pada tahap ini akan dilakukan:

1. Melakukan pembersihan teks.

2. Melakukan pemotongan teks kedalam bentuk *biword* kemudian dienkripsi menggunakan *MD5*
 3. Menghitung nilai *hash*
 4. Membentuk *window* dengan ukuran *w*
 5. Mendapatkan nilai *fingerprint*
- b. Menghitung tingkat kemiripan (*similarity*) dokumen dengan menggunakan persamaan *jaccard coefficient*

3.6 Perancangan Perangkat Lunak

Pada tahap ini akan dilakukan perancangan aplikasi, adapun Tahapan yang dilakukan adalah:

- a. Merancang *use case diagram* untuk menentukan pengguna (*actor*) yang terlibat dan proses apa saja yang bisa dilakukan (*scenario*) terhadap sistem.
- b. Merancang Dekripsi Fungsional pada sistem diantaranya *Context Diagram (CD)*, *Data Flow Diagram (DFD)* dan *Entity Relationship Diagram (ERD)*
- c. Perancangan Struktur Menu
Merancang menu-menu pada aplikasi yang memiliki fungsi masing-masing sesuai tujuan yang akan dicapai.
- d. Perancangan *interface* aplikasi.
Merancang atau mendesain tampilan antar muka aplikasi dengan pengguna. *Interface* yang akan dibangun adalah *interface input* dan *output*. Dengan demikian akan terlihat *interface* dari sistem dan dapat memberikan gambaran terhadap sistem yang akan dibangun.

3.7 Implementasi pengujian

Setelah dilakukan perancangan aplikasi, maka akan dilakukan tahap implementasi. Implementasi merupakan tahap dimana aplikasi siap untuk dioperasikan sesuai dari hasil analisis dan perancangan yang telah dilakukan, sehingga akan diketahui apakah aplikasi yang dirancang benar-benar dapat menghasilkan tujuan yang ingin dicapai.

Implementasi pengembangan aplikasi deteksi plagiarisme dokumen ini akan dibangun pada spesifikasi *hardware* dan *software* sebagai berikut:

1. Perangkat keras

Processor	: Intel® Core™ i3-2328-M CPU 2.20Ghz
Memori (RAM)	: 2,00 GB
Harddisk	: 500 GB

2. Perangkat Lunak

Sistem operasi	: <i>Windows 8 Enterprise 32-bit OS</i>
Bahasa pemrograman	: <i>Hypertext Preprocessor (PHP)</i>
DBMS	: <i>mySQL</i>
Tool	: <i>Notepad++</i>

3.8 Pengujian Sistem

Tahapan pengujian yang akan dilakukan pada sistem pendeteksian sumber plagiarisme dokumen yang dibangun bertujuan untuk mengetahui kesalahan dan kemudian memperbaikinya. Pengujian yang akan dilakukan sebagai berikut:

1. Pengujian konfigurasi pembuatan *query* yaitu proses *stemming*, pembentukan frasa kata serta pemilihan hasil frekuensi frasa kata. Hal ini bertujuan untuk mendapatkan konfigurasi terbaik yang menghasilkan dokumen yang *relevan* dengan dokumen yang di-*input* serta mendapatkan dokumen dengan *similarity jaccard coefficient* yang tinggi.
2. Pengujian Kemiripan dokumen *input* terhadap dokumen yang ada pada *corpus* sesuai dengan konfigurasi yang direkomendasikan pada pengujian pertama. Hal ini dilakukan untuk mengetahui lebih jelas tingkat akurasi aplikasi *source detection* dalam mencari sumber dokumen dan kemiripan terhadap dokumen.

3.9 Kesimpulan dan Saran

Pada tahapan ini berisi tentang kesimpulan yang dapat ditarik dari penelitian yang telah penulis lakukan, yang dimulai dari tahapan analisa hingga tahapan pengujian aplikasi. Selanjutnya pada bagian saran berisi saran-saran yang penulis berikan untuk mengembangkan aplikasi yang penulis bangun dalam

penelitian tugas akhir ini sehingga ke depannya dapat diteruskan dan menjadi lebih baik.