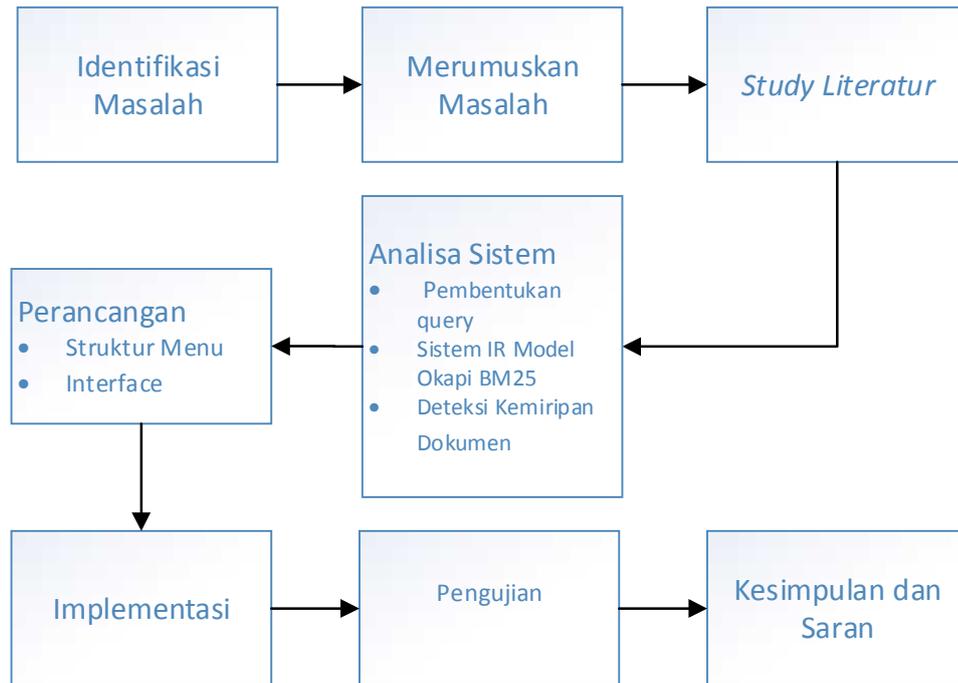


## BAB III

### METODOLOGI PENELITIAN

Tahapan penelitian yang dilakukan dalam penelitian adalah sebagai berikut.



Gambar 3.1 Tahapan Penelitian

#### 3.1 Identifikasi Masalah

Pada tahapan ini dilakukan identifikasi permasalahan bahwa ketika dokumen yang memiliki isi teks yang panjang dan seluruh isi teks dokumen tersebut akan dijadikan sebagai *query* pencarian, hal ini tentu akan berpengaruh terhadap hasil pencarian. Untuk itu perlu adanya algoritma yang dapat melakukan penyeleksian terhadap isi dokumen, mencari bagian mana yang cocok dan tepat mewakili isi dokumen untuk dijadikan *query* pencarian.

#### 3.2 Merumuskan Masalah

Dari tahapan identifikasi masalah dapat dirumuskan, bagaimana membentuk *query* dari dokumen yang panjang yang diduga melakukan

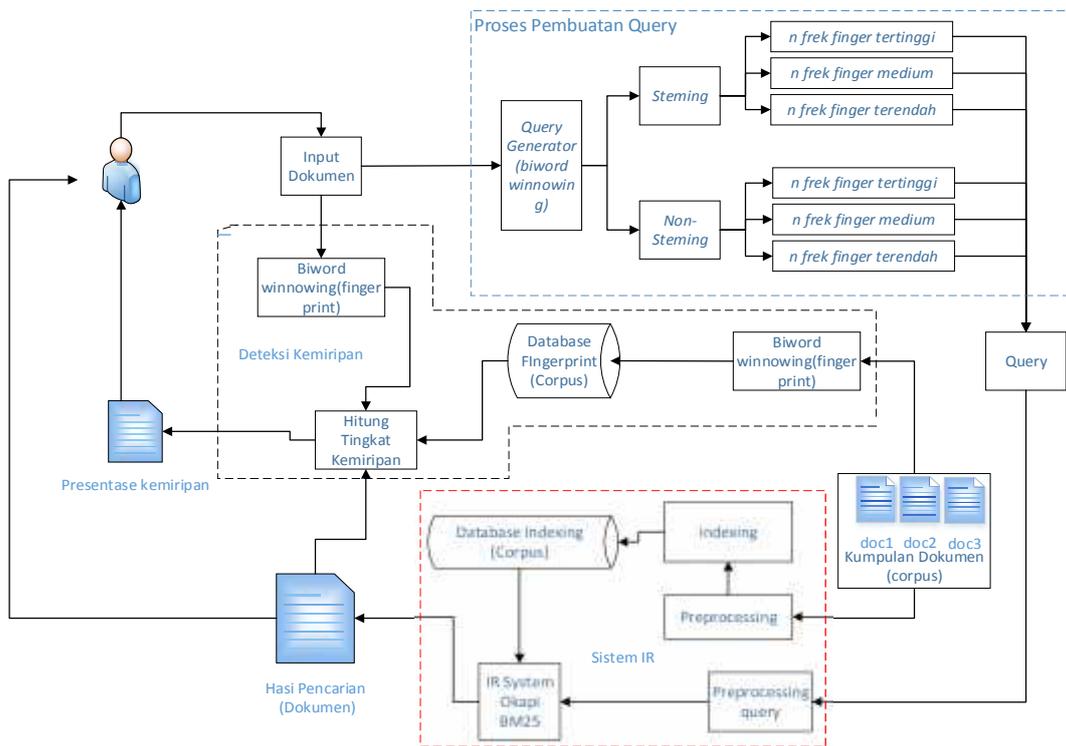
plagiarisme dengan menggunakan fingerprint algoritma *biword winnowing* yang kemudian *query* tersebut akan dicari dokumen sumber menggunakan algoritma Okapi BM25.

### 3.3 Study Literatur

Pada tahapan ini akan dilakukan pengumpulan informasi dan referensi yang berhubungan dengan masalah yang diteliti pada tugas akhir ini melalui jurnal ilmiah dan buku-buku.

### 3.4 Analisa Aplikasi

Pada tahapan ini akan dijelaskan secara rinci tentang proses sistem deteksi sumber plagiarisme yang akan dibangun sehingga mempermudah pemahaman terhadap sistem. Pada tahap ini terbagi menjadi beberapa proses :



Gambar 3.2 Analisa Sistem *Source Detection*

1. Analisa pembentukan *query* dari dokumen yang dicurigai plagiarisme menggunakan *fingerprint* dengan algoritma *biword winnowing*. Ada dua jenis *query* yang akan dibentuk, yakni
  - a. *Query* dari *fingerprint biword winnowing* dengan *stemming*. Pada tahap ini akan dilakukan :
    1. Melakukan pembersihan teks.
    2. Melakukan pemotongan teks serta melakukan *filter stopwords* dan *stemming*, kemudian dibentuk *biword* yang dienkripsi dengan menggunakan *MD5*.
    3. Menghitung nilai *hash*.
    4. Membentuk *window* dengan ukuran *w*
    5. Pemilihan beberapa nilai *hash* menjadi *document fingerprinting*. Kemudian melakukan perangkungan frekuensi *fingerprint*.
    6. Membentuk masing-masing *query* dengan menggunakan *n fingerprint* dengan frekuensi tertinggi, *n fingerprint* dengan frekuensi tengah, *n fingerprint* dengan frekuensi rendah.
  - b. *Query* dari *fingerprint biword winnowing* tanpa *stemming*. Pada tahap ini akan dilakukan :
    1. Melakukan pembersihan teks.
    2. Melakukan pemotongan teks serta melakukan *filter stopwords*, kemudian dibentuk *biword* yang dienkripsi dengan menggunakan *md5*.
    3. Menghitung nilai *hash*
    4. Membentuk *window* dengan ukuran *w*
    5. Pemilihan beberapa nilai *hash* menjadi *document fingerprinting*. Kemudian melakukan perangkungan frekuensi *fingerprint*.
    6. Membentuk masing-masing *query* dengan menggunakan *n fingerprint* dengan frekuensi tertinggi, *n fingerprint* dengan frekuensi tengah, *n fingerprint* dengan frekuensi rendah.
2. Analisa Sistem IR Okapi BM25. Tahapan ini terbagi menjadi 3, yakni

- a. Pembangunan *index* (pembuatan *inverted index*) dan pembobotan kata sebelum diterapkannya model Okapi BM25. Pada tahap ini akan dilakukan :
    1. Mengumpulkan dokumen yang akan di-*index* (dikenal dengan nama *corpus* atau koleksi dokumen).
    2. Pemisahan rangkaian kata (*tokenization*).
    3. Melakukan *linguistic preprocessing* : *stopword* dan *stemming*
    4. Indexing.
    5. Pembobotan kata
  - b. Preprocessing *query*. Tahapan yang dilakukan
    1. Proses tokenisasi terhadap *query*
    2. Melakukan *linguistic preprocessing* : *stopword* dan *stemming*
    3. Pembobotan kata pada *query*
  - c. Menghitung relevansi antara *query* dengan dokumen menggunakan model Okapi BM25. Dalam perhitungan menggunakan model Okapi BM25 menggunakan persamaan 2.6. Menurut Bintana (2012), dengan menggunakan persamaan 2.6 sudah dapat memperoleh dokumen yang relevan.
3. Analisa Deteksi kemiripan dokumen dengan Algoritma *Biword Winnowing*.
    - a. Pembuatan *database fingerprint* semua dokumen sumber dengan menggunakan algoritma *biword winnowing* yang akan diterapkan pada saat pengujian. Pada tahap ini akan dilakukan :
      1. Melakukan pembersihan teks.
      2. Melakukan pemotongan teks menjadi *biword* yang kemudian dienkripsi menggunakan *MD5*
      3. Menghitung nilai *hash*
      4. Membentuk *window* dengan ukuran *w*
      5. Pemilihan beberapa nilai *hash* menjadi *document fingerprinting*.
      6. Pembuatan *database fingerprint*
    - b. Pembentukan *fingerprint* dokumen yang diduga plagiarisme dengan algoritma *biword winnowing*. Pada tahap ini akan dilakukan :

1. Melakukan pembersihan teks.
  2. Melakukan pemotongan teks menjadi *biword* yang kemudian dienkripsi menggunakan *MD5*
  3. Menghitung nilai *hash*
  4. Membentuk *window* dengan ukuran *w*
  5. Pemilihan beberapa nilai *hash* menjadi *document fingerprinting*.
- c. Menghitung tingkat kemiripan (*similarity*) dokumen dengan menggunakan persamaan *jaccard coefficient*

### 3.5 Perancangan Aplikasi

Pada tahap ini dilakukan perancangan aplikasi. Pada tahap ini akan dilakukan :

- a. Perancangan *flat* Teks  
Merancang penyimpanan data dalam *flat* teks (*plain text*)
- b. Perancangan Struktur Menu  
Merancang menu-menu pada aplikasi yang memiliki fungsi masing-masing sesuai tujuan.
- c. Perancangan *interface* aplikasi.  
Merancang atau mendesain tampilan antar muka aplikasi dengan pengguna. Tampilan yang dibangun dapat memberikan gambaran umum mengenai aplikasi yang akan dibuat.

### 3.6 Implementasi

Pada proses implementasi ini akan dilakukan pembuatan modul yang telah dirancang dan dianalisa selanjutnya diimplementasikan pada bahasa pemrograman dan dilakukan pengujian untuk mengetahui tingkat keberhasilan aplikasi yang telah ada. Berikut adalah spesifikasi lingkungan implementasi perangkat keras dan perangkat lunak :

1. Perangkat keras
  - Processor : *Intel(R) Core(TM) i5 CPU M 430 @2.5GHz*
  - Memori (RAM) : 4,00 GB

Harddisk : 640 GB

## 2. Perangkat Lunak

Sistem Operasi : *Windows 8 Enterprise* 64 bit

Bahasa Pemrograman : PHP

*Tools* Perancangan : Notepad++

*Web Browser* : Google Chrome / CoolNovo

## 3.7 Pengujian

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 hasil pembuatan *query* dari dokumen yang diduga plagiarisme terhadap hasil pencarian dokumen sumber. Pengujian ini dilakukan untuk mengetahui jenis *query* mana yang menghasilkan dokumen sumber yang relevan. Pengujian ini dilakukan dengan memperhatikan urutan nilai *jaccard coefficient* yang diperoleh. Jika nilai *jaccard coefficient*-nya yang diperoleh terurut menurun (besar ke kecil) maka dapat dikatakan jenis *query* tersebut terbaik, karena sesuai dengan urutan dokumen sumber yang diperoleh, yakni dokumen sumber pertama memiliki nilai *jaccard coefficient* yang lebih tinggi dari dokumen sumber kedua, dokumen sumber kedua memiliki nilai *jaccard coefficient* yang lebih tinggi dari dokumen ketiga dan seterusnya.
2. Pengujian kemiripan dokumen uji terhadap dokumen sumber yang diperoleh dari proses pencarian dengan model Okapi BM25. Pengujian ini bertujuan untuk melihat apakah dokumen sumber yang diperoleh benar memiliki kemiripan terhadap dokumen uji. Pengujian ini dilakukan dengan memperhatikan adanya kesamaan isi antara dokumen uji dengan dokumen sumber yang ditunjukkan oleh sistem.

### **3.8 Kesimpulan dan Saran**

Tahapan ini berisi tentang kesimpulan yang dapat diambil dari penelitian yang telah dilakukan. Pada bagian saran berisi saran-saran yang penulis berikan untuk mengembangkan aplikasi agar ke depannya menjadi lebih baik.