BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Dengan berkembangnya teknologi informasi mengakibatkan teknologi internet berkembang semakin pesat, hal ini berdampak pada penggunaan internet. Tujuan dari penggunaan internet adalah untuk mendapatkan informasi dengan cepat dan akurat. Dengan pertambahan informasi ini berbanding lurus dengan jumlah dokumen yang ada didunia internet, salah satu contohnya adalah berita. Dokumen berita merupakan kumpulan informasi tentang banyak peristiwa penting yang terjadi dan terbaru secara berkala. Untuk memahami isi teks berita melalui ringkasan teks memerlukan waktu yang lebih singkat dibandingkan membaca seluruh isi teks. Oleh karena itu peringkasan teks menjadi hal yang sangat penting untuk dilakukan. Dengan adanya ringkasan, maka dapat menghemat waktu pembaca juga dapat menghindari pembacaan teks yang tidak relevan dengan informasi yang diharapkan oleh pembaca, terutama ketika sangat banyak informasi yang tersedia diinternet.

Peringkasan teks otomatis merupakan solusi yang tepat untuk mengatasi permasalahan tersebut. Peringkasan teks otomatis adalah suatu proses pembentukan yang lebih singkat dari sebuah teks dengan memanfaatkan aplikasi yang dijalanakan dan dioperasikan pada komputer. Penelitian tentang peringkasan teks otomatis telah dilakukan beberapa kali dengan menerapkan berbagai metode. Salah satu penelitian peringkasan teks otomatis yang ada [1], yaitu dengan menerapkan algoritma tf-idf. Kalimat yang memiliki bobot skor tertinggi langsung dijadikan sebagai kalimat ringkasan. Tetapi hal ini dirasa kurang baik jika hasil ringkasan tidak relevan dengan judul teks. Penelitian tersebut selanjutnya menggabungkan teknik kemiripan antar judul dengan kalimat yang disebut *query relevance* dengan algoritma *cosine similarity* dan menghasilkan nilai akurasi sebesar 77% untuk precission, 60% untuk recall, dan 66% untuk f-measure. Adapun penelitian [2], yaitu kompresi kalimat menggunakan hidden markov model untuk meringkas sebuah kalimat berbahasa Inggris yang pada susunan kalimatnya terdapat kata yang

dirasa tidak terlalu penting. Hasil penelitian tersebut [2], menghasilkan ringkasan yang lebih pendek namun tidak mengubah makna dari kalimat tersebut. Disebutkan pada artikel yang berjudul "Teori Dasar Hidden Markov Model" mengatakan metode HMM memiliki dua keuntungan yaitu pertama, model ini sangat kaya dalam strukur metematika sehingga dapat digunakan sebagai teori dasar dibanyak bidang aplikasi. Kedua, jika model ini digunakan dengan benar, ia dapat bekerja dengan efektif dan efisien untuk implementasi beberapa aplikasi penting [3]. Pada penelitian [2] menyebutkan metode HMM lebih tahan terhadap gangguan (noise) dan tidak bergantung pada bahasa tertentu terhadap bahasa untuk kompresi kalimat.

Berdasarkan penelitian peringkasan teks otomatis sebelumnya, penelitian kali ini akan dilakukan dengan mengimplementasikan algoritma tf-idf dan cosine similarity pada metode hidden markov model. Perbedaan pada penelitian [2] yaitu terletak pada hasil ringkasan, dimana pada penelitian kali ini akan meringkas sebuah teks dokumen untuk mencari dan menggali kalimat penting yang tersembunyi pada teks dokumen. Hasil ringkasan nantinya diharapkan dapat memberikan informasi yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang masalah yang telah dijelaskan sebelumnya, maka perumusan masalah yang ada adalah bagaimana cara mengimplementasi *hidden markov model* untuk peringkasan teks otomatis.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk membangun aplikasi peringkasan teks menggunakan metode *hidden markov model*. Tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui akurasi hasil ringkasan dengan menggunakan *hidden markov model.*

1.4 Batasan Masalah

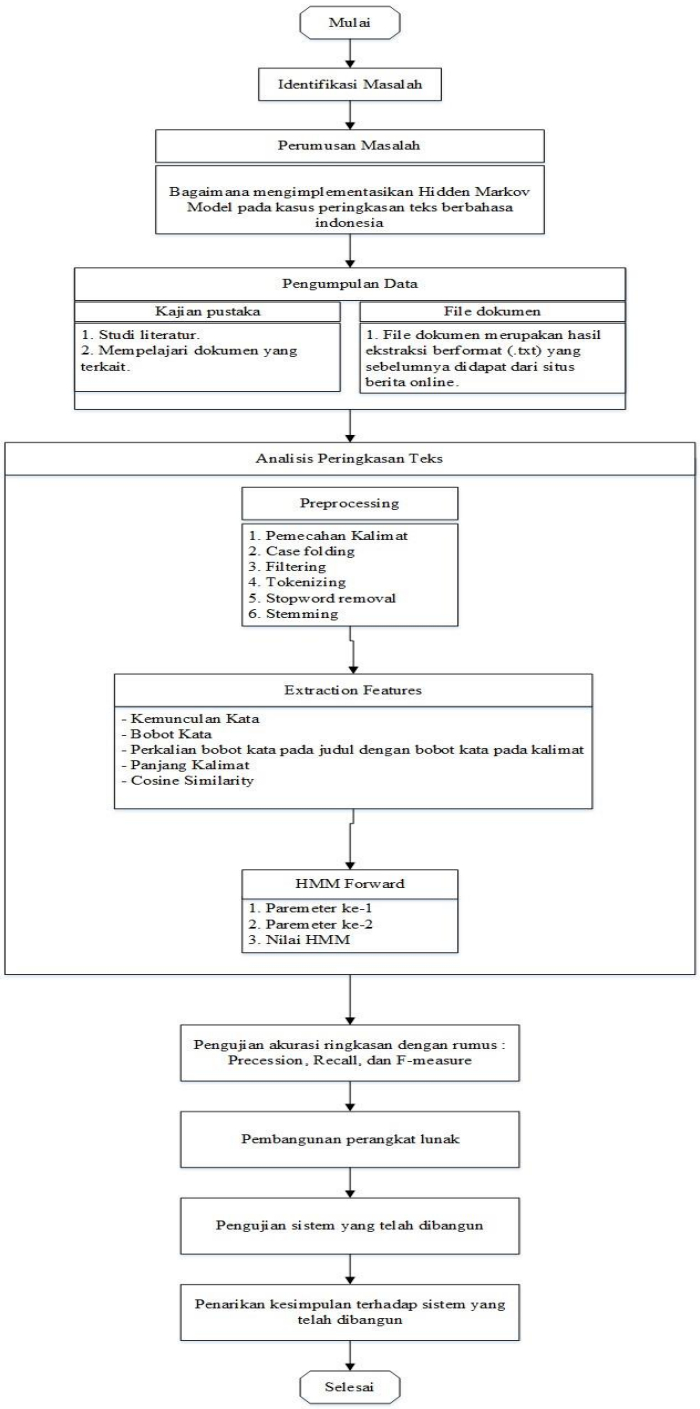
Dengan mempertimbangkan kompleksitas yang mungkin ada pada suatu ringksan, maka pada Tugas Akhir ini terdapat beberapa batasan masalah, yaitu :

1. Jenis teks dokumen yang digunakan adalah dokumen tunggal (artikel berita).

2. Teks dokumen yang diringkas merupakan teks berbahasa Indonesia.
3. Teks dokemen yang digunakan diambil dari media berita *online.*
4. Format *file* dokumen yang digunakan adalah *txt* yang sebelumnya telah diekstrak.
5. Peringkasan teks dilakukan secara ***offline***.
6. Ringkasan berupa hasil ekstraksi kalimat-kalimat dari dokumen asli.
7. Sistem tidak dapat melakukan proses pembelajaran.
8. Bahasa pemrograman yang digunakan adalah *JAVA* dan berbasis *Object Oriented Programming.*

1.5 Metodologi Penelitian

Metode penelitian ini menggunakan metode eksperimental. Metode eksperimental merupakan penelitian yang memanipulasi atau mengontrol situasi alamiah dengan cara membuat kondisi buatan (*artificial condition*) yang dilakukan oleh peneliti atau penelitian yang dilakukan dengan mengadakan manipulasi terhadap objek penelitian, serta adanya kontrol yang disengaja terhadap objek penelitian. Alur penelitian dapat dilihat pada gambar 1.1.



Gambar 1. 1 Alur Penelitian

Berdasarkan alur penelitian pada gambar 1.1 dapat dijelaskan sebagai berikut :

1. **Identifikasi Masalah**

   Tahapan pertama dalam penelitian adalah mengidentifikasi masalah. Penelitian mengidentifikasi masalah atau pun kendala pada proses peringkasan teks, analisis kebutuhan algoritma.

2. **Perumusan Masalah**

   Tahapan kedua adalah perumusan masalah. Peneliti merumuskan masalah bagaimana mengimplementasi *Hidden Markov Model* pada kasus peringkasan teks.

3. **Pengumpulan Data**

   Tahap ketiga adalah pengumpulan data. Metode pengumpulan data yang dilakukan pada penelitian ini yaitu :

   a. **Kajian Pustaka**

      Kajian pustaka dilakukan dengan mempelajari jurnal, *paper,* buku dan situs *oneline* yang terkait dengan topik peringkasan teks, tahapan *preprocessing,* pembobotan dengan TF-IDF, pembobotan *Cosine Similarity, Hidden Markov Model,* dan pengujian akurasi menggunakan rumus *precession, recall,* dan *f-measure.*

   b. ***File* Dokumen**

      File dokumen diambil dari beberapa situs berita *online* yang sudah familiar, *file* tersebut kemudian diolah kebentuk format *file* (.txt).

4. **Analisis Peringkasan Teks**

   Tahapan ini berupa analisis peringkasan teks. Adapun analisis yang dilakukan dalam penelitian ini yaitu :

   a. ***Preprocessing***

      Preprocessing merupakan tahapan untuk mempersiapkan teks menjadi data yang akan diolah pada tahapan berikutnya. *Preprocessing* pada penelitian ini terdiri dari beberapa tahapan yaitu proses pemecahan kalimat, *case folding, filtering, tokenizing, stopword removal,* dan *stemming.*

b. ***Extraction Feature***

Merupakan proses pengolahan kalimat menggunakan beberapa fitur sehingga menghasilkan nilai-nilai yang nantinya akan digunakan untuk meringkas dokumen. *Extraction features* untuk penelitian ini meliputi kemunculan kata, bobot kata, dan *cosine similarity.*

c. **HMM *Forward***

Merupakan tahapan terakhir. Pada tahapan ini hasil dari pembobotan *tf-idf* , *cosine similarity* akan dimasukkan pada notasi hmm kemudian dicari nilai parameter ke-1 dan parameter ke-2. Kemudian dilakukan perhitungan untuk mencar nilai probabilitas menggunakan algoritma *forward.* Penentuan kalimat ringkasan dilakukan berdasarkan nilai probabilitas tertinggi dari hasil algoritma *forward.*

5. **Pengujian Akurasi Ringkasan**

Tahapan ini merupakan pengujian akurasi hasil ringkasan menggunakan rumus *precession, recall,* dan *f-measure.*

6. **Pembangunan Perangkat Lunak**

Metode pembangunan perngkat lunak yang digunakan adalah model *prototype.* Tahapan proses model *prototype* adalah sebagai berikut :

a. Pengumpulan Kebutuhan

Tahapan ini adalah mengumpulkan data dengan cara mengumpulkan jurnal, *paper,* artikel, buku dan bacaan-bacaan yang berkaitan dengan penelitian ini.

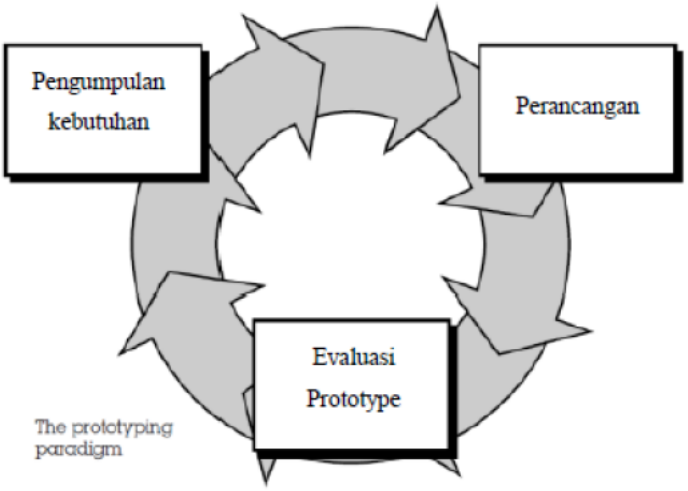
b. Perancangan

Yaitu menganalisis masalah yang terbagi menjadi analisis sistem, analisis data masukan, analisis metode, analisis kebutuhan fungsional, analisis kebutuhan non fungsional, perancangan struktur menu, perancangan antarmuka, dan jaringan semantik.

c. Evaluasi *Prototype*

Yaitu melakukan pengujian dengan melihat akurasi dari hasil penelitian terhadap *software* yang dibangun.

Gambar model *prototype* dapat dilihat pada gambar 1.2 sebagai berikut:



Gambar 1. 2 Model *Prototype*

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian skripsi ini disusun untuk memberikan gambaran umum mengenai penelitian yang dilakukan. Sistematika penulisannya sebagai berikut :

Bab 1 Pendahuluan

Bab ini berisi latar belakang, perumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian, dan sistematika penulisan.

Bab 2 Tinjauan Pustaka

Bab ini menjelaskan landasan teori yang berisi teori-teori yang berkaitan dengan implementasi *hidden markov model* untuk peringkasan teks otomatis. Teori-teori tersebut terdiri dari teori peringkasan teks, proses peringkasan teks, kata, kalimat, paragraf, berita, algoritma peringkasan teks, algoritma TF-IDF, konsep *Cosine Similarity, Hidden Markov Model,* algoritma *forward,* evaluasi peringkasan teks, pemrograman berorientasi objek, *Unified Modelling Language* (UML), dan perangkat lunak pendukung.

Bab 3 Analisis dan Perancangan

Bab ini menjelaskan analisis yang berkaitan dengan sistem, yaitu meliputi analisis masalah yang menjelaskan permasalahan-permasalahan yang terdapat pada sistem, analisis tada masukan, analisis metode, analisis kebutuhan non fungsional, analisis kebutuhan fungsional, perancangan struktur menu, perancangan antarmuka, dan jaringan semantic, serta penggunaan *tools* seperti Visio 2013 untuk membangun *use case diagram, activity diagram, class diagram,* dan *sequence diagram.*

Bab 4 Implementasi dan Pengujian

Bab ini menjelaskan implementasi dari analisis dan perancangan sistem sebelumnya kedalam bentuk bahasa pemrograman berbasis desktop yaitu *JAVA.* Hasil implementasi selanjutna akan dilakukan pengujian, mengingat bahw sistem yang dapat digunakan dengan baik harus melalui tahapan pengujian. Pengujian yang dilakukan adalah menguji akurasi hasil ringkasan teks dengan hasil ringksan manual dan menguji sistem peringkasan teks otomatis kepada pengguna.

Bab 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dari keseluruhan isi laporan skripsi dan saran yang diuraikan dengan tujuan agar adanya pengembangan sistem peringkasan teks otomatis dengna mengimplementasikan metode yang lebih baik untuk hasil akurasinya.