

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Ketersediaan informasi yang semakin banyak menjadikan ringkasan sebagai kebutuhan yang sangat penting (Mulyana, 2010). Menurut (Hovy, 2001) Ringkasan merupakan teks yang singkat dan padat yang dapat dianggap pengganti dari keseluruhan dokumen karena tetap mempertahankan kandungan informasi penting yang dimiliki dokumen oleh sumbernya. Peringkasan teks otomatis (*automatic text summarization*) adalah pembuatan bentuk yang lebih singkat dari suatu teks dengan memanfaatkan aplikasi yang dijalankan dan dioperasikan pada computer (Nugraha, 2008).

Dengan adanya ringkasan maka pembaca dapat dengan cepat dan mudah memahami intisari dari dokumen tersebut. Oleh karena itu, sistem peringkasan dokumen diperlukan untuk membantu mengurangi waktu membaca keseluruhan isi berita dengan hanya membaca hasil ringkasannya, sehingga memudahkan dalam mencari informasi berita tersebut (Mustaqhfiri, 2011).

Menurut (Jezek & Steinberger 2007) Kriteria peringkasan teks terbagi dua bagian yaitu peringkasan teks berdasarkan ekstraksi dan abstraksi. Teknik ekstraksi merupakan suatu teknik untuk menyalin unit-unit teks yang paling penting atau paling informatif dari teks sumber menjadi ringkasan, sedangkan teknik abstraksi adalah mengambil intisari dari teks sumber kemudian membuat ringkasan dengan menciptakan kalimat-kalimat baru yang merepresentasikan intisari teks sumber dalam bentuk berbeda (Jezek & Steinberger 2007). Pada penelitian ini sistem peringkasan dokumen akan dibuat menggunakan teknik ekstraksi.

Sejumlah penelitian telah dilakukan dalam membangun sistem peringkasan dokumen otomatis diantaranya penelitian tentang *Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia* oleh Aristoteles (2013), *penerapan Terms Frequency –Inverse Document Frequency*

*pada sistem peringkasan teks otomatis dokumen tunggal berbahasa Indonesia oleh Iyan Mulyana, dkk (2010), Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance oleh Muchammad Mustaqhfi, dkk (2011).*

Pada penelitian Aristoteles (2013) dilakukan penelitian tentang bagaimana meringkas dokumen tunggal berbahasa Indonesia yang berjenis *file* teks dengan menggunakan algoritma genetika. Tujuan penelitian ini adalah melakukan penentuan tingkat kepentingan atau pembobotan dari sebelas fitur teks untuk meringkas dokumen. Terdapat sebelas fitur teks yang diterapkan pada penelitian ini, yaitu posisi kalimat, *positive keyword*, *negative keyword*, kemiripan antar kalimat, kalimat menyerupai judul, kalimat yang mengandung nama entiti, kalimat yang mengandung data numerik, koneksi antar-kalimat, penjumlahan bobot antar-kalimat, dan kalimat semantic. Hasil ringkasan diuji dengan menggunakan *F-measure*, *Precision*, *Recall*. Dalam penelitian ini *compression* (ukuran ringkasan) yang dilakukan sebesar 10%, 20% dan 30%. Hasil penelitian yang diperoleh bahwa algoritma genetika dapat digunakan untuk mencari tingkat kepentingan yang optimal dari tiap fitur teks. Dengan nilai akurasi 47.46% pada *compression* 30%, Sedangkan hasil tidak optimal pada *compression* 10%.

Pada penelitian Mulyan Iyan, dkk (2008) dilakukan penelitian tentang bagaimana penerapan *Term Frequency –Inverse Document Frequency* pada sistem peringkasan dokumen tunggal. Pada penelitian ini, peringkasan teks otomatis yang di kembangkan merupakan sistem peringkasan dengan inputan berupa *single* dokumen dan secara otomatis menghasilkan ringkasan (*summary*). Tahap - tahap peringkasan teks nya yaitu : melakukan *text preprocessing* yang mana hanya dilakukan pemecahan kalimat dan *tokenizing* kata, kemudian pembobotan TF-IDF, menghitung bobot(W), lalu Proses Pengurutan dari W. tiga kalimat W yang memiliki bobot tertinggi dijadikan ringkasan. Hasil penelitian yang diperoleh bahwa metode TF-IDF dapat digunakan untuk meringkas *single document* dan memiliki tingkat akurasi 61% pada *compression* 50%. Metode TF-IDF lebih besar tingkat akurasinya dibandingkan algoritma genetika.

Pada penelitian Mustaqhfi, dkk (2011) dilakukan penelitian tentang bagaimana meringkas dokumen menggunakan metode *Maximum Marginal Relevance*. Teknologi peringkasan teks otomatis menawarkan solusi untuk membantu pencarian isi berita berupa deskripsi singkat (*summary*). Penelitian diawali dengan lima tahap *text preprocessing*: pemecahan kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses selanjutnya menghitung bobot TF-IDF, bobot *query relevance* dan bobot *similarity*. Ringkasan dihasilkan dari ekstraksi kalimat dengan menggunakan metode *maximum marginal relevance*. Metode ekstraksi *maximum marginal relevance* merupakan metode yang digunakan untuk mengurangi redundansi dalam peringkasan kalimat. Hasil perhitungan evaluasi diurutkan berdasarkan nilai *recall*, *precision* dan *f-measure* dari *persentase* yang tertinggi ke urutan terendah. Hasil penelitian yang diperoleh bahwa metode MMR dapat digunakan untuk meringkas *single document* dan memiliki tingkat akurasi 70% pada *compression* 60%. Namun *query* yang dimasukkan tidak menggambarkan isi, sehingga kalimat yang terambil tidak sesuai urutan kalimat yang baik.

Dari penelitian - penelitian sebelumnya telah berhasil membangun aplikasi peringkasan dokumen. Namun hasil yang didapat belum mencapai akurasi yang tinggi. Untuk meningkatkan akurasi sistem, peneliti mencoba menggabungkan metode TF-IDF dan MMR kemudian menambahkan *query expansion* pada dokumen. *Query expansion* merupakan suatu teknik dengan menambahkan *keyword* baru kedalam *query* awal sehingga meningkatkan performansi pencarian. Yang mana *keyword* tersebut merupakan Top-n dari perhitungan Algoritma TF-IDF. Bobot kata akan dihitung dan *diranking*. Kata yang memiliki bobot tertinggi maka akan dijadikan sebagai *keyword*. Untuk peringkasan *small document*, seperti pada berita (*news*), menggunakan nilai parameter  $=0.7$  atau  $=0.8$  pada perhitungan MMR, karena akan menghasilkan ringkasan yang baik (Jade Goldstein, 2008).

Dengan adanya penelitian “*Automatic document summarization* menggunakan metode MMR dengan Top- n *TF-IDF token query expansion* pada dokumen pendek” ini, diharapkan mampu meringkas dokumen, membentuk *query expansion* dari suatu dokumen dan meningkatkan akurasi sistem.

## 1.2 Rumusan Masalah

Dari latar belakang di atas dapat di ambil suatu rumusan masalah yaitu, bagaimana mengetahui tingkat akurasi sistem dan menampilkan *query expansion*.

## 1.3 Batasan Masalah

Desain dan pengembangan sistem peringkasan teks ini memiliki batasan atau ruang lingkup yang harus dikerjakan dengan cakupan sebagai berikut :

- 1 Penelitian ini menggunakan bahasa indonesia sebagai inputan.
- 2 Input dokumen bisa secara manual dan link website.
- 3 nilai koefisien  $= 0,7$  pada perhitungan MMR. Karena akan menghasilkan ringkasan yang baik (Goldstein, 2008).
- 4 Dokumen yang di inputkan adalah dokumen pendek seperti Artikel dan Berita.
- 5 *Keyword* yang muncul adalah 5 kata yang memiliki bobot tertinggi pada perhitungan TF-IDF.

## 1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai oleh penulis dari penelitian dan penyusunan tugas akhir ini adalah untuk mengetahui tingkat akurasi sistem dengan menggabungkan 2 metode yaitu metode TF-IDF dan MMR kemudian menampilkan *query expansion* pada dokumen tersebut.

## 1.5 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian :

## Bab I. Pendahuluan

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan.

## Bab II. Landasan Teori

Bab ini menjelaskan tentang teori-teori yang berhubungan dengan penelitian yang terdiri dari penjelasan mengenai *Automatic Text Summarization*, tipe evaluasi, hipotesa, riset penelitian sebelumnya, *Text Preprocessing*, TF-IDF, dan Algoritma *Maximum Marginal Relevance*.

## Bab III. Metodologi Penelitian

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian, yaitu identifikasi masalah, merumuskan masalah, *study literatur*, analisa sistem, perancangan sistem, implementasi, dan pengujian.

## Bab IV. Analisa Dan Perancangan Sistem

Berisi pembahasan mengenai analisa sistem meliputi analisa sistem dan perancangan pada aplikasi peringkasan dokumen.

## Bab V. Implementasi

Dalam bab ini membahas tentang implementasi aplikasi peringkasan dokumen, serta pengujian dan evaluasi.

## Bab VI. Penutup

Dalam bab ini akan dijelaskan mengenai beberapa kesimpulan yang didapatkan dari pembahasan pembuatan aplikasi tek otomatis disertai beberapa saran sebagai hasil akhir dari penelitian yang telah dilakukan.

