



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dengan cepatnya perkembangan teknologi saat ini, kemudahan dalam mengakses informasi menjadi tidak terbatas. Salah satu masalah yang dapat ditimbulkan dari hal ini adalah *information overload* atau kelebihan informasi. Istilah *information overload* itu sendiri pertama kali diciptakan oleh seorang profesor ilmu politik bernama Bertram Gross dan dituliskan pertama kali dalam bukunya yang berjudul *The Managing of Organizations* pada tahun 1964. Kemudian, istilah tersebut dipopulerkan oleh Alvin Toffler dalam bukunya yang berjudul *Future Shock* pada tahun 1970 (Franganillo, 2018). Gross mendefinisikan *information overload* sebagai sebuah kondisi dimana jumlah *input* atau masukan ke dalam sebuah sistem, melampaui kapasitas pemrosesan informasi sistem tersebut (Gross, 1964). Para pengambil keputusan (*decision makers*) mempunyai kemampuan kognitif terbatas dalam memproses suatu informasi, hasilnya, ketika *information overload* terjadi, kualitas keputusan yang diambil akan menurun (Gross, 1964).

Selain itu, menurut salah satu wakil presiden perusahaan komputer besar IBM, Larry Bowden, efek dari kelebihan informasi ini memengaruhi efektifitas dalam bekerja karena banyaknya informasi yang tidak relevan dan tidak sesuai dengan tujuan yang ingin dicapai (Spira, 2011). Netcraft mencatat dalam survei yang dilakukan pada bulan Oktober 2017, jumlah respon situs yang diterima adalah 1.815.237.491 situs dan 6.886.362 *web-facing computers*. Jumlah ini

mendapat peningkatan sebesar 10,2 juta situs dan 88.300 komputer dari bulan sebelumnya (Silvia dkk., 2014).

Dengan banyaknya respon yang diterima tersebut, maka jumlah informasi yang tidak relevan untuk suatu tujuanpun meningkat dengan cepat. Solusi yang dapat digunakan untuk mendapatkan informasi dengan tepat dan dalam waktu yang singkat adalah dengan menggunakan *Automatic Text Summarization*. *Automatic text summarization* adalah sebuah proses untuk menciptakan suatu dokumen atau teks yang berisi tentang informasi penting dan maksud utama dari sumber informasi secara otomatis (Brownlee, 2017).

Terdapat dua pendekatan yang ada pada *automatic text summarization*, yaitu *extractive summarization* dan *abstractive summarization* (Hahn dan Mani, 2000). *Extractive summarization* adalah pendekatan yang dilakukan dengan mengambil informasi penting dari teks atau naskah sumber tanpa ada perubahan pada struktur kalimat. Sehingga, hasil yang didapatkan dari pendekatan ini adalah potongan-potongan informasi penting yang ditambahkan menjadi suatu ringkasan. *Abstraction summarization* adalah pendekatan yang dilakukan dengan mengambil informasi penting dan kemudian dilakukan parafrase atau pengungkapan kembali suatu konsep dengan cara lain dalam bahasa yang sama. Hasil yang didapatkan dari pendekatan ini adalah sebuah ringkasan baru dengan kalimat yang berbeda, namun dengan maksud atau tujuan yang sama dengan naskah atau teks asli.

Penelitian mengenai *automatic text summarization* ini sudah pernah dilakukan oleh Silvia, Pitri Rukmana, dan Vivi Regina Aprilia pada jurnal yang berjudul “Aplikasi Automatic Text Summarizer” dengan menggunakan metode *Genetical Algorithm Based* dan *Lexical Chains* (Silvia dkk., 2014). Simpulan yang

disampaikan adalah algoritma tersebut dapat menghasilkan ringkasan *extractive* yang mewakili informasi penting dari suatu dokumen teks tunggal berbahasa Indonesia dengan cepat. Selain itu, penelitian terhadap *automatic text summarization* ini juga pernah dilakukan oleh Dhimas Anjar Prabowo, Muhammad Fhadli, Mochammad Ainun Najib, Handika Agus Fauzi, dan Imam Cholissodin pada tahun 2015 dan dalam jurnal yang berjudul “TF-IDF-Enhanced Genetic Algorithm untuk Extractive Automatic Text Summarization” (Prabowo dkk., 2015). Penelitian ini menggunakan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) dan digabungkan dengan algoritma *Enhanced Genetic Algorithm* yang merupakan pengembangan dari *Genetic Algorithm* sehingga didapatkan hasil yang lebih informatif. Simpulan yang disampaikan adalah dengan kedua algoritma tersebut, maka ringkasan yang dihasilkan menjadi lebih relevan dengan informasi yang ingin disampaikan pada teks asli. Dengan hanya menggunakan algoritma TF-IDF, informasi yang didapatkan hanya dengan melakukan penghitungan dari nilai tertinggi suatu kalimat. Kemudian dengan adanya algoritma tambahan *Enhanced Genetic Algorithm*, ringkasan yang dihasilkan lebih mudah dipahami, sehingga menyerupai *abstractive summarization*.

Kemudian, penelitian terhadap *Automatic Text Summarization* yang dilakukan oleh Gede Aditra Pradnyana dan Komang Ari Mogi pada tahun 2014 dalam jurnal yang berjudul “Implementasi Automated Text Summarization untuk Dokumen Tunggal Berbahasa Indonesia dengan Menggunakan Graph-Based Summarization dan Algoritma Genetika” (Pradnyana dan Mogi, 2014). Pada jurnal tersebut, penggunaan *Graph-Based Summarization* adalah untuk mencari jalur terpendek atau mengetahui persamaan antar kalimat yang telah dilakukan

pembobotan dengan menggunakan metode TF-IDF. Simpulan yang disampaikan adalah nilai generasi atau hasil dari Algoritma Genetika dan populasi atau penyatuan kalimat hasil ekstrasi berbading lurus dengan waktu proses dan penggunaan sumber daya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka masalah yang dirumuskan adalah bagaimana merancang dan membangun aplikasi *automatic text summarization* dengan menggunakan metode *Maximum Marginal Relevance* berbasis *web*.

## 1.3 Batasan Masalah

Berdasarkan rencana pelaksanaan penelitian, maka dicantumkan batasan-batasan masalah sebagai berikut.

1. Sumber naskah atau dokumen teks yang digunakan adalah berbahasa Indonesia, sehingga semua variabel yang digunakan dalam penelitian ini adalah bahasa Indonesia.
2. Penilaian atau pembobotan kalimat yang akan dijadikan sumber informasi adalah menggunakan metode *Term Frequency*.
3. Dokumen yang akan digunakan sebagai sumber ringkasan berekstensi .doc atau .docx, dan .txt.
4. Panjang kata maksimum dalam suatu dokumen yang dijadikan sumber ringkasan adalah 500 kata.

#### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah sebelumnya, tujuan dari penelitian ini adalah merancang dan membangun aplikasi *automatic text summarization* dengan menggunakan metode *Maximum Marginal Relevance* berbasis *web*.

#### 1.5 Manfaat Penelitian

Penelitian ini bermanfaat untuk mendapatkan informasi penting dan relevan dari sumber informasi yang tersedia. Sehingga dapat meningkatkan efektifitas dalam bekerja tanpa membuang sumber daya yang berlebih untuk mencari maksud dari sebuah dokumen.

#### 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari lima bab, yaitu sebagai berikut.

##### 1. BAB I PENDAHULUAN

Bab ini terdiri atas latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

##### 2. BAB II LANDASAN TEORI

Bab ini menjelaskan landasan teori dari penelitian yang dilakukan, seperti *text summarization*, Porter Stemming, dan Algoritma *Maximum Marginal Relevance*.

##### 3. BAB III METODOLOGI DAN PERANCANGAN SISTEM

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

Bab ini berisikan metodologi penelitian yang digunakan dan perancangan sistem. Perancangan sistem yang dimaksud, yaitu perancangan Data Flow Diagram, Flowchart, dan rancangan antar muka.

4. **BAB IV IMPLEMENTASI DAN UJI COBA**

Bab ini berisi hasil implementasi sistem dan hasil dari uji coba sistem yang pada sistem peringkas berita.

5. **BAB V KESIMPULAN DAN SARAN**

Bab ini terdiri atas simpulan dari hasil pengujian sistem dan saran untuk pengembangan sistem di kemudian hari.

