

1. Pendahuluan

1.1 Latar belakang

Perkembangan teknologi dan ilmu pengetahuan mengakibatkan ketersediaan informasi meningkat, terutama informasi yang bersifat dinamis yang dapat diakses secara *online*. Informasi *online* dapat berupa artikel berita, dokumen-dokumen, video klip, film, musik, dan lain sebagainya. Kebutuhan user akan informasi yang berupa artikel berita menyebabkan user menghabiskan waktu yang cukup lama untuk membaca keseluruhan isi artikel berita tersebut. Oleh karena itu, diperlukan adanya solusi agar user tidak perlu membaca keseluruhan isi artikel berita akan tetapi user tetap mendapatkan informasi yang singkat dan padat yang merepresentasikan isi artikel berita tersebut. Solusinya adalah sebuah sistem yang mampu meringkas artikel berita yaitu, sistem peringkasan teks otomatis (*automatic text summarization*). Peringkasan untuk artikel berita sesuai diterapkan pada dokumen tunggal (*single document*) karena prosesnya yang cepat.

Summary atau ringkasan didefinisikan sebagai sebuah teks yang dihasilkan dari satu atau lebih teks sumber yang mengandung informasi signifikan/penting dari teks sumber dan panjangnya tidak lebih dari setengah teks sumber[5]. Ringkasan dapat diklasifikasikan dalam dua kategori, yaitu ekstraksi dan abstraksi. Ringkasan ekstraksi merupakan sejumlah kalimat atau frase yang terpilih dari teks sumber dengan nilai tertinggi dan diletakkan bersama pada sebuah teks baru yang lebih pendek tanpa mengubah isi teks. Sedangkan ringkasan abstraksi menggunakan metode linguistik untuk memeriksa dan menginterpretasikan teks. Sebagian besar sistem peringkasan teks otomatis menggunakan metode ekstraksi untuk menghasilkan ringkasan. Peringkasan teks otomatis bekerja optimal pada dokumen-dokumen yang terstruktur dengan baik seperti pada sejumlah artikel, berita, laporan, dan karya ilmiah.[17]

Graph-based summarization algorithm atau peringkasan teks berbasis graf merupakan suatu metode peringkasan teks yang *language independent* yang dapat menghasilkan ringkasan ekstraktif. Graf tekstual merupakan teks sumber yang direpresentasikan menjadi sebuah graf. Pembangunan graf tekstual menggunakan konsep *similarity* antar unit teks. *Vertex* pada graf tekstual dapat berupa unit-unit teks seperti kata-kata, kalimat-kalimat, atau paragraf-paragraf dalam teks. *Edge* dalam graf tekstual menunjukkan keterhubungan antar *vertex*. Keterhubungan antar *vertex* dapat berupa *similarity* antar kalimat ataupun hubungan leksikal atau gramatikal antar kata/frasa.[13]

Dewasa ini sudah banyak *tools* untuk peringkasan teks, tetapi belum banyak yang membahas kualitas ringkasan teks. Kualitas ringkasan teks sangat penting karena melihat tingginya tingkat kepentingan suatu dokumen. Dimana ketika menghadapi banyak dokumen yang harus dibaca dalam waktu singkat, ringkasan teks yang relevan dengan dokumen sangat diperlukan.[17] Teori *fuzzy set* dapat merepresentasikan dan menangani masalah ketidakpastian yang dalam hal ini dapat berarti keraguan, ketidaktepatan, kekuranglengkapan informasi, dan kebenaran yang bersifat sebagian[18]. Dalam tugas akhir ini, penulis menggunakan *fuzzy system* untuk melakukan ekstraksi kalimat-kalimat penting dalam peringkasan teks.

Dalam tugas akhir ini, penulis akan menganalisis kualitas ringkasan antara penggabungan linier *Graph-Based Summarization Algorithm* dibandingkan dengan *Graph-Based Summarization Algorithm* dengan *fuzzy system* dilihat dari *precision*, *recall*, dan *f-measure* menggunakan *ROUGE Evaluation toolkit*. Sehingga dari analisis perbandingan kedua metode ini akan diketahui kualitas ringkasan yang lebih baik.

1.2 Perumusan masalah

Permasalahan yang menjadi objek penelitian pada tugas akhir ini adalah

1. Bagaimana menerapkan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System* terhadap dokumen tunggal berbahasa Indonesia, khususnya artikel berita?
2. Bagaimana menganalisa pengaruh parameter-parameter *fuzzy* untuk ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System*?
3. Bagaimana kualitas ringkasan teks antara penggabungan linier *Graph-Based Summarization Algorithm* dibandingkan dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy System*?

Dalam tugas akhir ini, peringkasan teks dilakukan dengan batasan masalah sebagai berikut:

1. Teks sumber yang akan diringkaskan adalah artikel berita berbahasa Indonesia yang diperoleh dari *website* berita atau koran yang berupa *softcopy* dokumen teks.
2. Peringkasan teks dilakukan secara *offline*.
3. Peringkasan teks yang dilakukan adalah peringkasan dokumen tunggal.
4. Hasil peringkasan berupa ekstraksi dari teks sumber.
5. Tidak melakukan proses *stemming* terhadap teks masukan.
6. Metode yang digunakan adalah *Fuzzy Based Method*.
7. Evaluasi dilakukan dengan membandingkan *content overlap* antara hasil peringkasan otomatis dan ringkasan referensi dengan menggunakan *ROUGE (Recall-Oriented Understudy for Gisting Evaluation) evaluation toolkit* [3].

1.3 Tujuan

Tujuan tugas akhir ini adalah:

1. Menerapkan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* terhadap dokumen tunggal berbahasa Indonesia, khususnya artikel berita.
2. Menganalisa pengaruh parameter-parameter *fuzzy* untuk ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system*.
3. Membandingkan kualitas ringkasan teks antara penggabungan linier *Graph-Based Summarization Algorithm* dengan penggabungan *Graph-Based Summarization Algorithm* dengan *Fuzzy system*. Kualitas ringkasan teks dilihat dengan *ROUGE evaluation toolkit*.

1.4 Metodologi penyelesaian masalah

Metodologi yang dilakukan untuk menyelesaikan permasalahan dengan menggunakan langkah-langkah sebagai berikut:

1. Studi literatur dan pengumpulan data

Mengumpulkan bahan-bahan referensi yang menunjang proses penelitian, yaitu yang berhubungan dengan peringkasan teks dan *fuzzy system*. Pengumpulan data - data yang diperlukan untuk mendukung penyelesaian masalah, seperti teks berita untuk digunakan sebagai input.

2. Analisis dan perancangan sistem

Melakukan analisis dan perancangan terhadap sistem yang dibangun, menganalisis metode yang akan digunakan untuk menyelesaikan permasalahan, termasuk menentukan bahasa pemrograman yang digunakan, arsitektur, fungsionalitas, dan antarmuka sistem. Input sistem berupa dokumen teks artikel berita. Output dari sistem berupa teks hasil ringkasan artikel berita.

3. Implementasi dan pembangunan sistem

a. Membangun sistem peringkasan teks otomatis dengan mengimplementasikan *Graph-Based Summarization Algorithm* berdasarkan analisis dan perancangan yang dibuat, dengan tahapan:

- Membangun sistem untuk mencari skor ringkasan teks dengan *Textrank*.
- Membangun sistem untuk mencari skor ringkasan teks dengan *Similarity with title*.
- Membangun sistem untuk mencari skor ringkasan teks dengan penggabungan linier skor *TextRank* dan *similarity with title*.

b. Membangun sistem peringkasan teks otomatis dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* berdasarkan analisis dan perancangan yang dibuat, dengan tahapan:

- Membangun sistem *fuzzy* untuk mencari skor gabungan antara *Textrank* dan *Similarity with title*.

4. Pengujian dan analisis

Pengujian dan analisis dilakukan oleh penulis tanpa melibatkan pihak lain, meliputi:

- a. Pengujian untuk mengetahui kualitas ringkasan teks dengan penggabungan linier *Graph-Based Summarization Algorithm*.
- b. Pengujian untuk mengetahui kualitas ringkasan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system*.
- c. Analisis pengaruh parameter *fuzzy* terhadap hasil ringkasan.
- d. Analisis perbandingan kualitas ringkasan teks antara teks dengan *Graph-Based Summarization Algorithm* dengan teks dengan penggabungan *Graph-Based Summarization Algorithm* dengan *fuzzy system* berdasarkan hasil pengujian *precision* dan *recall* dengan *ROUGE evaluation toolkit*.

5. Pengambilan kesimpulan dan saran, serta penyusunan laporan Tugas Akhir.