

1. Pendahuluan

1.1 Latar belakang

Kemudahan digitalisasi teks saat ini semakin meningkatkan jumlah informasi yang tersedia. Namun banyaknya ketersediaan informasi tersebut tidak berarti identifikasi dan pemilihan informasi yang dibutuhkan mudah dilakukan. Identifikasi mengenai informasi apa yang terkandung dari suatu teks dapat diperoleh dengan membaca kata kunci. Namun, banyak dokumen teks yang ada tidak disertai dengan kata kunci.

Pada referensi [3] dinyatakan bahwa kata kunci dapat dianggap sebagai ringkasan yang lebih padat dari suatu teks. Kata kunci tidak hanya berguna untuk mengetahui deskripsi suatu teks, tetapi juga dapat berperan penting untuk proses *indexing* otomatis, peringkasan teks dan meningkatkan performansi kategorisasi teks [3][6].

Mengekstrak kata kunci dari suatu teks tentu saja dapat dilakukan secara manual. Namun, jika teks tersebut merupakan teks yang cukup panjang dan tersedia dalam jumlah yang besar, proses ekstraksi kata kunci secara manual akan memakan waktu yang lama. Oleh karena itu dibutuhkan suatu pengekstrak kata kunci yang dapat mengurangi *resource* yang dibutuhkan jika ekstraksi kata kunci dilakukan secara manual.

Banyak algoritma/metode yang dapat diimplementasikan untuk mengekstrak kata kunci, misalnya algoritma genetika, *naïve bayes*, dan *rule induction*. Dalam Tugas Akhir ini, algoritma yang diimplementasikan adalah *TextRank*. *TextRank* dipilih karena algoritma ini tidak membutuhkan data latih dan bersifat *language independent*, yaitu tidak bergantung pada bahasa tertentu dan dapat digunakan di semua bahasa. *TextRank* termasuk algoritma perankingan berbasis graf yang menentukan nilai pentingnya sebuah *vertex* berdasarkan informasi global yang didapatkan dari keseluruhan graf. Graf yang dibangun merepresentasikan teks dimana *vertex* mewakili unit teks dan *edge*-nya merupakan relasi antar *vertex*.

Tugas Akhir ini mengimplementasikan *TextRank* pada ekstraksi kata kunci dengan melakukan pengujian pada pengaturan parameter *TextRank*, yaitu jenis graf, filter kata atau pemilihan kata yang direpresentasikan sebagai *vertex*, dan ukuran *window*. Jenis graf yang akan diuji adalah *undirected*, *directed forward*, dan *directed backward*. Filter kata yang digunakan dapat dibagi menjadi 3 kategori, yaitu *all open class words* atau semua kata, kata benda saja, dan kata benda yang digabung dengan kata sifat. Sedangkan ukuran *window* divariasikan antara 2 sampai 10 kata. Pengujian juga dilakukan terhadap variasi nilai *damping factor* dan *threshold* yang digunakan dalam pumus algoritma *TextRank*.

Evaluasi dilakukan dengan cara membandingkan kata kunci hasil ekstraksi sistem dengan kata kunci yang dibuat oleh manusia. Setiap teks masukan diuji terhadap kombinasi parameter *TextRank* yang diimplementasikan dalam sistem ekstraksi yang dibuat. Kata kunci yang dihasilkan kemudian dibandingkan dengan kata kunci asli dengan parameter evaluasi *recall*.

Data uji yang digunakan adalah abstraksi jurnal TA S1 fakultas Informatika ITTelkom dan abstraksi jurnal Seminar Nasional Aplikasi Teknologi Informasi (SNATI). Alasan pemilihan data tersebut adalah karena abstrak pada umumnya

selalu dilengkapi dengan kata kunci yang diberikan oleh penulisnya. Hal ini untuk memudahkan proses evaluasi sistem.

1.2 Perumusan masalah

Perumusan masalah dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana menerapkan algoritma *TextRank* pada sistem ekstraksi kata kunci teks berbahasa Indonesia.
2. Bagaimana pengaruh pengaturan parameter-parameter *TextRank* (jenis graf, jenis filter kata, dan ukuran *window*) terhadap kata kunci hasil ekstraksi sistem bila dibandingkan dengan kata kunci asli dengan parameter uji *recall*.
3. Bagaimana pengaruh penggunaan variasi nilai *damping factor* dan *threshold* terhadap *recall* sistem.

Batasan masalah dalam Tugas Akhir ini adalah :

1. Teks yang digunakan sebagai data dalam Tugas Akhir ini adalah abstrak jurnal-jurnal TA S1 Teknik Informatika dan abstrak jurnal-jurnal Seminar Nasional Aplikasi Teknologi Informasi (SNATI).
2. Format teks dalam bentuk .txt.
3. Tidak melakukan proses *stemming*.
4. Sistem yang dibangun bersifat *offline*.
5. Untuk keperluan evaluasi, kata kunci asli yang telah ditentukan penulis jurnal digunakan sebagai pembanding hasil ekstraksi sistem. Oleh karena itu kata kunci asli dianggap sebagai pilihan kata kunci yang benar.

1.3 Tujuan

Tujuan Tugas Akhir ini adalah:

1. Menerapkan algoritma *TextRank* pada sistem ekstraksi kata kunci teks berbahasa Indonesia.
2. Menganalisis pengaruh pengaturan parameter-parameter *TextRank* (jenis graf, jenis kata, dan ukuran *window*) terhadap kata kunci hasil ekstraksi sistem bila dibandingkan dengan kata kunci asli dengan parameter uji *recall*.
3. Menganalisis pengaruh penggunaan variasi nilai *damping factor* dan *threshold* terhadap *recall* sistem.

1.4 Metodologi penyelesaian masalah

Metodologi yang dilakukan untuk menyelesaikan permasalahan pada Tugas Akhir ini adalah sebagai berikut:

1. Melakukan studi literatur khususnya mengenai ekstraksi kata kunci dan *TextRank*
2. Melakukan pencarian data yang diperlukan untuk mendukung penyelesaian masalah, yaitu jurnal-jurnal TA S1 Teknik Informatika IT Telkom dan jurnal-jurnal SNATI yang berbahasa Indonesia.

3. Melakukan analisis algoritma yang akan diterapkan dalam ekstraksi kata kunci. Algoritma yang dianalisis adalah *TextRank*.
4. Melakukan analisis kebutuhan perangkat lunak dan perancangan perangkat lunak sistem ekstraksi kata kunci yang menerapkan algoritma *TextRank*.
5. Melakukan implementasi sistem sesuai dengan perancangan yang telah dilakukan.
6. Melakukan pengujian sistem dan menganalisis hasil keluaran sistem yang berupa kumpulan kata kunci. Pengujian dilakukan dengan melakukan perubahan pengaturan parameter *TextRank* dan membandingkan kata kunci hasil sistem dengan kata kunci asli dengan parameter *recall*.
7. Pembuatan laporan Tugas Akhir.