

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dengan adanya kebutuhan dalam mencari keterkaitan informasi berupa teks yang dibaca dengan referensi teks lain yang ada seperti artikel berita dan sejenisnya, maka metode pengukuran *Semantic Relatedness* yang baik sangat dibutuhkan supaya hasil pencarian sebuah referensi teks lain yang saling terkait dapat menghasilkan sebuah hasil yang tepat. Selama ini pencarian referensi teks yang terkait kebanyakan memanfaatkan kesamaan kategori, padahal parameter kategori masih terlalu luas untuk menentukan keterkaitan antar teks. Karena teks dengan kategori yang sama belum tentu membahas hal yang sama juga. Misal sebuah teks/artikel berita dengan kategori teknologi, bisa saja artikel tersebut membahas tentang *smartphone*, *finger print*, dan beberapa pembahasan yang sebenarnya tidak terkait. Dari fakta tersebut perlu dilakukan pengukuran keterkaitan teks agar pencarian referensi teks lain yang saling terkait bisa menghasilkan hasil yang tepat.

Atas dasar itulah tugas akhir kali ini akan dilakukan pengujian dalam mengukur nilai keterkaitan antar teks menggunakan sebuah metode *Salient Semantic Analysis* dengan *TextRank for Keyword Extraction* sebagai *preprocessing*. Sebelum melakukan pengukuran *Semantic Relatedness* tahap awal adalah melakukan *preprocessing* dengan *TextRank for Keyword Extraction* untuk mendapatkan *keyword* dari sebuah teks. Setelah dilakukan *preprocessing* tahap berikutnya adalah menjalankan metode *Salient Semantic Analysis*. Dalam metode *Salient Semantic Analysis* pengukuran *Semantic Relatedness* dilakukan dengan membangun *Semantic Profile* dari *Salient Encyclopedic Features* [1]. Secara umum metode *Salient Semantic Analysis* terdiri dari 2 langkah [1]. Langkah pertama kita memanfaatkan sebuah corpus sebagai *Knowledge Based* yang bersumber dari *Wikipedia* untuk menghasilkan daftar nilai kata [1]. Langkah kedua adalah mulai mengukur *Semantic Relatedness* dari sebuah kata/teks [1]. Diharapkan dengan dilakukannya tugas akhir ini maka dapat diketahui hasil pengukuran keterkaitan antar teks menggunakan metode *Salient Semantic Analysis* dengan *TextRank for Keyword Extraction* sebagai *preprocessing*.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, maka dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana menggunakan *TextRank for Keyword Extraction* sebagai *preprocessing* dalam menentukan *keyword* sebuah teks.
2. Bagaimana menggunakan metode *Salient Semantic Analysis* dalam mengukur *Semantic Relatedness* dari sebuah kata/teks.
3. Bagaimana hasil pengukuran keterkaitan antar teks menggunakan metode *Salient Semantic Analysis*.

## 1.3 Tujuan

Tujuan yang ingin dicapai dari pengerjaan Tugas Akhir ini adalah untuk mengetahui bagaimana hasil pengukuran keterkaitan antar teks menggunakan metode *Salient Semantic Analysis* dengan *TextRank for Keyword Extraction* sebagai *preprocessing*.

## 1.4 Batasan Masalah

Adapun batasan masalah dari permasalahan ini adalah :

1. *Corpus* yang digunakan mengambil *knowledge based* yang bersumber dari *wikipedia*.
2. Jumlah maksimal artikel *wikipedia* yang dibangkitkan untuk membangun *Corpus* adalah 5 artikel.
3. Jumlah *list keyword* yang dipilih saat proses *preprocessing* adalah 20.
4. Teks yang digunakan adalah sebuah teks yang menggunakan bahasa inggris.
5. Pengujian nilai korelasi pada *word relatedness* menggunakan dataset *WordSim353*.
6. Pengujian ketepatan nilai keterkaitan antar teks menggunakan pembandingan nilai berdasarkan manusia.

## 1.5 Metodologi Penelitian

Metodologi yang dilakukan dalam menyelesaikan tugas akhir ini adalah :

### 1. Tahap Studi Literature

Aktifitas pada tahap ini adalah mempelajari tentang *Text Mining*, *Text Relatedness*, *Semantic Relatedness*, *Salient Semantic Analysis* yang didapat dari berbagai sumber dan referensi.

### 2. Analisis dan Perancangan Sistem

Aktifitas pada tahap ini adalah melakukan analisis terhadap studi literature yang ada. Dan melakukan perancangan sistem yang tepat.

### 3. *Preprocessing* dan pengumpulan data

Aktifitas pada tahap ini adalah melakukan pengumpulan data berupa kumpulan artikel/teks yang akan digunakan dalam penelitian ini. Setelah itu melakukan proses *preprocessing* dengan menggunakan metode *TextRank for Keyword Extraction [2]* untuk mendapatkan intisari dari sebuah teks.

### 4. Implementasi Metode

Aktifitas pada tahap ini adalah implementasi metode yang sudah dirancang untuk melaksanakan penelitian ini. Yaitu melakukan analisis efektifitas pengukuran keterkaitan antar teks menggunakan metode *Salient Semantic Analysis*.

### 5. Analisis Hasil

Aktifitas pada tahap ini adalah melakukan analisis dari hasil pengujian yang sudah dilakukan pada implementasi metode.

### 6. Pembuatan Laporan

Aktifitas pada tahap ini adalah menyusun laporan akhir dan pengumpulan dokumentasi berdasarkan implementasi dan hasil penelitian.