

## 1. Pendahuluan

### 1.1 Latar Belakang

Untuk mengetahui keterkaitan antar pasangan kata maka perlu dilakukan pengukuran nilai *Semantic Relatedness*. Contoh nya pasangan kata *football* dan *soccer* memiliki nilai *semantic relatedness* berdasarkan *human rater* sebesar 9.03[2], lalu pasangan kata *stock* dan *jaguar* memiliki nilai *semantic relatedness* berdasarkan *human rater* sebesar 0.92 [3], yang artinya pada pasangan kata *football* dan *soccer* memiliki keterkaitan yang tinggi, sedangkan pada pasangan kata *stock* dan *jaguar* memiliki nilai keterkaitan yang rendah. Pengukuran *Semantic Relatedness* pada saat ini digunakan untuk mencari keterkaitan antar pasangan kata, lalu digunakan pada pasangan teks yang bertujuan untuk mendefinisikan seberapa terkait antar pasangan teks. Teks yang dimaksud dapat berupa web artikel, koran, dan dokumen lainnya. Dengan mengetahui relasi antar pasangan teks dapat memudahkan pembaca untuk memperoleh informasi yang terkait tanpa harus mencari di seluruh dokumen yang ada.

*Semantic Relatedness* merupakan suatu nilai kedekatan dari suatu pasang kata yang di dapatkan dari proses perhitungan untuk mencari keterkaitan antara dua kata yang memperhitungkan berbagai faktor yang berkaitan dengan makna atau konsep yang disampaikan oleh kata tersebut dan mencari hubungan dari antara pasangan kata-kata. Ukuran keterkaitan antar kata antar kata diukur dari leksikal dan semantik nya. *Wikipedia Link Based Measure* merupakan teknik yang digunakan untuk mendapatkan nilai dari pengukuran *semantic relatedness*. WLM menggunakan Wikipedia sebagai kamus atau *world knowledge* pada suatu kata. Metode ini menggunakan *hyperlink* yang terdapat pada Wikipedia untuk perhitungan *semantic relatedness*. *Semantic relatedness* dari suatu pasang kata dapat diperoleh dari relasi antar link pada artikel terkait [2].

Dalam tugas akhir ini akan mengimplementasikan dan menganalisis pengukuran *Semantic Relatedness* dengan menggunakan *Wikipedia Link Based Measure*. Pengukuran *Semantic Relatedness* antar pasangan kata yang berupa judul artikel

pada wikipedia dengan menggunakan *WLM* memberikan kemudahan dalam proses nya karena dataset tersedia dalam ukuran yang tidak terlalu besar. Pengukuran *Relatedness* dihitung dengan menggunakan dua pengukuran yaitu *TFxIDF Inspired* dan *Normalized Google Distance*. Pada pengukuran menggunakan *TFxIDF Inspired* data yang dibutuhkan yaitu berupa daftar relasi link pada *article of interest* terhadap artikel lain (*link out*). Sedangkan pada pengukuran *Normalized Google Distance* data yang dibutuhkan yaitu berupa daftar relasi link yang merujuk kepada *article of interest* (*link in*).

## 1.2 Perumusan Masalah

Berdasarkan dari latar belakang yang telah dijelaskan, dapat disimpulkan beberapa perumusan masalah dalam penelitian tugas akhir ini yaitu :

1. Bagaimana cara membuat data yang berisi relasi antar artikel pada Wikipedia?
2. Bagaimana cara memperoleh *link in* dan *link out* pada artikel Wikipedia?
3. Bagaimana cara menghitung *Semantic Relatedness* antar dua artikel pada Wikipedia dengan menggunakan pengukuran *TF.IDF* dan *Normalized Google Similarity*?
4. Bagai mana cara untuk mengevaluasi sistem ?

## 1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Dataset yang digunakan hanya berupa list *link page* dari seluruh konten di *English Wikipedia* tahun 2008
2. Penelitian ini hanya sebatas mencari nilai *Semantic Relatedness* antar dua inputan kata yang berupa judul artikel pada wikipedia.
3. Evaluasi sistem dilakukan dengan membandingkan nilai *Semantic Relatedness* dari sistem dengan dataset *WS-353* (353 pasang kata) dan sistem *Wikipedia Miner*.

## 1.4 Tujuan

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan pengukuran *word relatedness* dengan menggunakan Wikipedia-Link dengan dua pengukuran *semantic relatedness* yaitu *TF.IDF Inspired* dan *Normalized Google Distance*.
2. Menganalisa pengaruh data link yang dihasilkan dari sistem terhadap nilai *semantic relatedness* dengan dua pengukuran *TF.IDF Inspired* dan *Normalized Google Distance*.
3. Menganalisa hasil pengujian sistem dari nilai *semantic relatedness* yang didapatkan.

## 1.5 Metode Penelitian

Adapun tahap-tahap metodologi untuk menyelesaikan masalah pada penelitian ini, antara lain:

1. Studi Literatur  
Studi literatur yang dilakukan adalah mempelajari konsep dan mengkaji materi yang berkaitan dengan penelitian Tugas Akhir seperti pengukuran *Semantic Relatedness*, *TF.IDF Inspired*, *Normalized Google Distance* dan materi lain nya yang mendukung.
2. Pengumpulan data  
Pengumpulan data diperoleh dari Wikipedia yaitu berupa link-page yang diolah dengan menggunakan *Software* Matlab yang menghasilkan daftar relasi link antar artikel[4].
3. Pembangunan Sistem  
Sistem akan memproses inputan berupa satu pasang kata (artikel dalam wikipedia), lalu mencari *Link Out* dan *Link In* dari masing-masing kata. Setelah mendapatkan data *Link Out* dan *Link In* nilai *Semantic Relatedness* pada pasangan kata dapat dihitung dan menjadi output dari sistem ini. *Software* yang digunakan adalah Netbeans IDE 7.2.1 dengan bahasa pemrograman Java.

#### 4. Implementasi

Mengimplementasikan dua pengukuran *TF.IDF Inspired* dan *Normalized Google Distance* untuk menghitung nilai *Semantic Relatedness* pada pasangan Artikel di Wikipedia.

#### 5. Pengujian

Hasil nilai *Semantic Relatedness* yang diperoleh akan dikorelasikan terhadap data set WS-353 dan Sistem Wikipedia Miner.

#### 6. Analisa Hasil Pengujian

Hasil dari pengujian akan di analisa apa yang mempengaruhi baik/buruk nya hasil dari evaluasi sistem.

#### 7. Pembuatan Laporan

Setelah hasil pengujian dan analisis, akan dibuat laporan data hasil running secara menyeluruh dari penelitian Tugas Akhir ini.