BAB I Pendahuluan

BAB I berisi alasan dan tujuan yang ingin dicapai pada penelitian ini. Penulis memaparkan beberapa poin penting diantaranya seperti Latar Belakang. Dengan adanya bab ini, diharapkan pembaca memahami beberapa poin lainnya seperti Perumusan Masalah, Batasan Masalah, Tujuan, Hipotesis serta Metodologi Penyelesaian Masalah dari penelitian ini.

1.1 Latar Belakang

I-GRACIAS adalah sistem informasi akademik dan non akademik Universitas Telkom. Gabungan aplikasi akademik dan non akademik membuat I-GRACIAS kini memiliki 74 aplikasi. Proses pencarian aplikasi yang ada di I-GRACIAS dilakukan dengan cara pencocokkan *keyword* dengan nama aplikasi melalui *query*. Supaya pencarian kita berhasil, maka *keyword* yang dimasukkan harus sama atau setidaknya berisi suku kata yang sama dengan nama aplikasi. Apabila ada *user* baru dan *user* tersebut tidak tahu nama aplikasi yang ingin dicari, dengan kata lain *keyword* yang dimasukkan berbeda dengan nama aplikasi maka pencarian tidak akan berhasil [1]. Tentunya hal ini akan membuat pengguna aplikasi merasa bingung untuk menemukan aplikasi yang diinginkan.

Pencarian tersebut bisa dilakukan dengan pendekatan *semantic similarity* yang dihitung menggunakan WS4J. *Semantic similarity* adalah pengukuran antar-kata dengan POS yang sama berdasarkan keterhubungan semantik [1] dari Wordnet yang merupakan bagian dari Thesaurus. Metode pengukuran *Semantic similarity* dikelompokkan menjadi beberapa kelas antara lain adalah *Path based*, *Information Content based*, *Gloss based* dan *Vector based* [2]. Pemilihan *Path-based* pada penelitian ini dikarenakan pengukuran Omiotis harus berbasis *Path method* untuk mengukur *semantic similarity* [1].

WS4J merupakan *library* berbasis java yang menyediakan beberapa metode perhitungan *semantic relatedness*¹. Beberapa varian perhitungan *semantic relatedness* yang tersedia pada WS4J dengan *path-based* antara lain Hirst and St-Onge (HSO) and Wu and Palmer (WUP) [2]. Diantara jenis pengukuran tersebut, penulis memilih WUP sebagai perhitungan *similarity*, hal ini dikarenakan setiap *pairs id* WUP hampir selalu menghasilkan nilai yang lebih besar jika dibandingkan dengan HSO pada percobaan *dataset Semeval 2014 task 3*.

Pengukuran Omiotis adalah perluasan dari pengukuran *semantic relatedness*. Keutamaan pengukuran Omiotis dibandingkan dengan STASIS dan *LSA-based approach* adalah kemampuannya untuk menghitung kemiripan antar-*text* yang pendek seperti

¹<https://code.google.com/archive/p/ws4j/>

kalimat. Selain itu, kelebihan lain dari Omiotis adalah tidak tergantung pada Wordnet [1]. Maka dari itu pada pengukuran ini, sistem akan melakukan perhitungan TF-IDF dengan *harmonical mean* untuk mendapatkan nilai *lexical relevance*. Nilai *lexical relevance* ini digunakan untuk mendapatkan nilai Omiotis.

Dalam penelitian ini menggunakan pendekatan *semantic similarity* dan perhitungan Omiotis untuk mengukur kemiripan *keyword* dengan beberapa atribut aplikasi (nama aplikasi dan deskripsi aplikasi) yang akan digunakan untuk proses pencarian. Perhitungan Omiotis akan melibatkan perhitungan *semantic similarity*. Dengan demikian, penggunaan perhitungan Omiotis bisa membantu pencarian aplikasi yang ada di I-GRACIAS.

1.2 Perumusan Masalah

Perumusan masalah dari tugas akhir ini adalah sebagai berikut.

1. Bagaimana pengaruh metode *semantic similarity* untuk pengukuran Omiotis ?
2. Bagaimana pengaruh *lexical relevance* terhadap pengukuran Omiotis ?
3. Bagaimana mengevaluasi Omiotis menggunakan dataset Semeval 2014 *task 3* dan dataset I-GRACIAS ?
4. Bagaimana menerapkan *Semantic Similarity based on word thesaurus* menggunakan pengukuran Omiotis pada data I-GRACIAS ?

1.3 Batasan Masalah

Batasan masalah dari tugas akhir ini adalah sebagai berikut.

1. Bahasa yang digunakan adalah Bahasa Inggris.
2. Proses *lemmatization* menggunakan *library* Stanford Core NLP.
3. Perhitungan *semantic similarity* menggunakan *library* WS4J berdasarkan metode WUP.
4. Pencarian berdasarkan nama aplikasi dan deskripsi aplikasi.
5. Kamus kata diperoleh dari Wordnet.
6. Pengukuran evaluasi menggunakan Pearson *correlation*.

1.4 Tujuan

Tujuan dari tugas akhir ini adalah sebagai berikut.

1. Mengetahui perbedaan metode *semantic similarity* untuk pengukuran Omiotis.
2. Mengetahui pengaruh *lexical relevance* terhadap pengukuran Omiotis.
3. Mengevaluasi Omiotis menggunakan dataset Semeval 2014 *task 3* dan dataset I-GRACIAS.
4. Menerapkan *Semantic Similarity based on word thesaurus* menggunakan pengukuran Omiotis pada data I-GRACIAS.

1.5 Hipotesis

Berdasarkan studi literatur yang berkaitan dengan *Semantic Text Similarity (STS)*, maka hipotesis dari tugas akhir ini adalah sebagai berikut.

1. Semakin banyak kata yang sama muncul di dua dokumen maka nilai kemiripan akan semakin besar.
2. Kata yang tidak tersedia di Wordnet tidak akan berpengaruh meskipun memiliki nilai *lexical relevance* yang besar.
3. Pengaruh *lexical relavance* adalah sebagai nilai komplemen untuk nilai kemiripan dokumen.
4. *Similarity term a* dan *b* sama dengan *Similarity b* dan *a*.

1.6 Metodologi Penyelesaian Masalah

Metodologi dari tugas akhir ini adalah sebagai berikut.

1. Identifikasi masalah, tahap ini dilakukan dengan cara mengidentifikasi topik *semantic relatedness* dan berbagai aspek yang berkaitan, diskusi langsung dengan dosen, mengumpulkan informasi melalui media internet, dan diskusi antar-teman kelompok keahlian.
2. Studi literatur dan pengumpulan data, tahap ini adalah tahap mencari sumber yang bisa dijadikan acuan. Sumbernya bisa dari buku, *journal/paper*, konsultasi dengan dosen, dan teman kelompok keahlian. Selain itu, tahap ini merupakan tahap pengumpulan data penelitian.
3. Proposal, yakni proses pembuatan proposal berdasarkan fenomena masalah yang sudah teridentifikasi dan hasil studi literatur.

4. Perancangan sistem, tahap perancangan alur kerja sistem dan desain sistem untuk memudahkan proses implementasi.
5. Implementasi, tahap ini dilakukan jika semua data terkumpul dengan lengkap.
6. Pengujian, tahap ini adalah tahap evaluasi sistem yang dilakukan melalui uji perhitungan Pearson *correlation* (r) dari perhitungan Omiotis dengan perhitungan yang didapat dari manusia
7. Analisis, tahap ini adalah tahapan analisis terhadap hasil pengujian.
8. Pelaporan, tahap ini dilakukan dengan tujuan semua aktivitas, percobaan, dan pengujian agar bisa terdokumentasi sehingga bisa digunakan untuk bahan pengembangan selanjutnya.
9. Publikasi adalah tahap untuk mempublikasikan hasil tugas akhir ini ke perpustakaan Universitas Telkom.