

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

NAMA : GUNARTO SINDORO
NRP : 5108100188
DOSEN WALI : SARWOSRI, S.Kom, M.T

2. JUDUL TUGAS AKHIR

“Rancang Bangun Aplikasi Pengubah Struktur dan Susunan Kalimat dengan Menggunakan Analisa Semantik dan Pemrosesan Bahasa Alami”

3. LATAR BELAKANG

Dengan perkembangan ilmu pengetahuan dan teknologi, maka pembuatan suatu riset ilmiah atau pembuatan karya tulis ilmiah biasanya menggunakan kutipan-kutipan dari karya tulis ilmiah yang lain.

Kutipan-kutipan yang dibuat biasanya menggunakan kata-kata yang sama dan menyertakan daftar kutipan untuk menjaga kevalidan hak cipta. Untuk mempermudah pengutipan, maka perlu untuk mengubah kalimat yang dibuat pada karya tulis ilmiah dengan menggunakan struktur maupun kata-kata lain yang menghasilkan kalimat bermakna sama. Kalimat yang dibuat dengan kata-kata lain juga dapat memberikan pengertian yang lebih mendalam dari kalimat asli.

Bagi seorang peneliti, sebuah karya ilmiah bisa dikembangkan berkali-kali, dengan membuat publikasi karya ilmiah yang berbeda. Pada kasus ini, isi dari karya ilmiah yang baru mungkin mengutip sebagian dari karya ilmiah yang sudah dibuat sebelumnya, namun bila kutipan menggunakan kata-kata yang sama persis, maka akan beresiko disebut sebagai otoplagiat.

4. TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan pembuatan Tugas Akhir ini adalah untuk membuat beberapa kalimat baru dari sebuah kalimat dengan mengubah struktur dan kata-kata yang digunakan secara otomatis pada kalimat namun tidak merubah makna kalimat tersebut atau memiliki tingkat kesamaan diatas batas ambang tertentu.

5. RUMUSAN MASALAH

Permasalahan yang diangkat dalam menyelesaikan Tugas Akhir ini adalah bagaimana mengubah sebuah kalimat menjadi kalimat yang lain tanpa mengubah makna dari kalimat tersebut secara otomatis.

Detail permasalahan yang dihadapi:

- Bagaimana menggunakan analisa semantik untuk mengubah struktur kalimat dan memperoleh makna kalimat yang sama.
- Bagaimana mengubah kalimat dari aktif ke pasif dan sebaliknya namun tetap memperoleh makna kalimat yang sama.

6. DASAR TEORI

6.1. KALIMAT

Di bidang linguistik, kalimat adalah ekspresi dalam *bahasa alami*¹, dan sering didefinisikan untuk menunjukkan sebuah unit gramatikal yang terdiri dari satu atau lebih kata-kata yang umumnya berkaitan lemah secara sintaksis untuk kata-kata yang mendahului atau mengikutinya [1].

6.2. SINONIM

6.2.1. SINONIM KATA

Sinonim kata adalah kata yang berbeda namun memiliki arti yang sama atau hampir sama [2]. Sinonim itu sendiri dibagi menjadi beberapa jenis yaitu: sinonim kata, sinonim kata kerja, sinonim kata sifat, sinonim kata keterangan dan sinonim kata depan. Sinonim didefinisikan juga dengan melihat konteks dari kalimat yang ada, karena pada kalimat yang berbeda makna dari kata yang sama bisa berbeda.

¹ Bahasa alami adalah bahasa manusia, seperti bahasa inggris, bahasa indonesia, dll dan bukan bahasa pemrograman komputer seperti C, C++, Java, PHP, dll.

6.2.2. SINONIM KALIMAT

Sinonim kalimat adalah kalimat yang berbeda namun memiliki arti secara keseluruhan yang mirip. Sinonim kalimat dapat diperoleh dengan beberapa pendekatan yaitu dengan mengubah struktur kalimat maupun dengan mengubah beberapa kata di dalam kalimat dengan menggunakan sinonim, hiponim, maupun hipernim.

6.3. PART-OF-SPEECH TAGGING (POST)

Dalam linguistik, *Part-Of-Speech Tagging*² (POST) adalah sebuah proses menandai kata-kata di dalam kalimat sebagai bagian tertentu dari *part-of-speech*, didasarkan pada definisi dan konteks [3]. Pembagian kategori pada *Part-Of-Speech Tagging* dikembangkan pertama oleh Brown_Corpus³ sehingga kata-kata yang digunakan mempunyai tingkat ambiguitas yang lebih rendah [4]. Untuk mengurangi tingkat ambiguitas, *Part-Of-Speech Tagging* tidak hanya mengubah *part-of-speech* dengan kata lain, karena beberapa kata-kata dapat mewakili lebih dari satu *part-of-speech* pada waktu yang berbeda dan beberapa bagian dari *part-of-speech* yang kompleks dan tak terucapkan.

7. RINGKASAN ISI TUGAS AKHIR

Dalam tugas akhir ini, akan dibuat perancangan model penyusun ulang kalimat dengan mengenali struktur dan pola kalimat menggunakan metode Semantik dan *Natural Language Processing* (NLP). Metode semantik yang dimaksud adalah penggunaan sinonim, hipernim dan hiponim di dalam konteks kalimat. Pengolahan dengan menggunakan semantik dibutuhkan untuk mencari padanan atau persamaan kata dengan mencari sinonim dari kata tersebut, sedangkan pengolahan dengan NLP digunakan untuk mengetahui struktur penyusun kalimat sehingga bisa menyusun ulang kalimat menjadi bentuk yang berbeda.

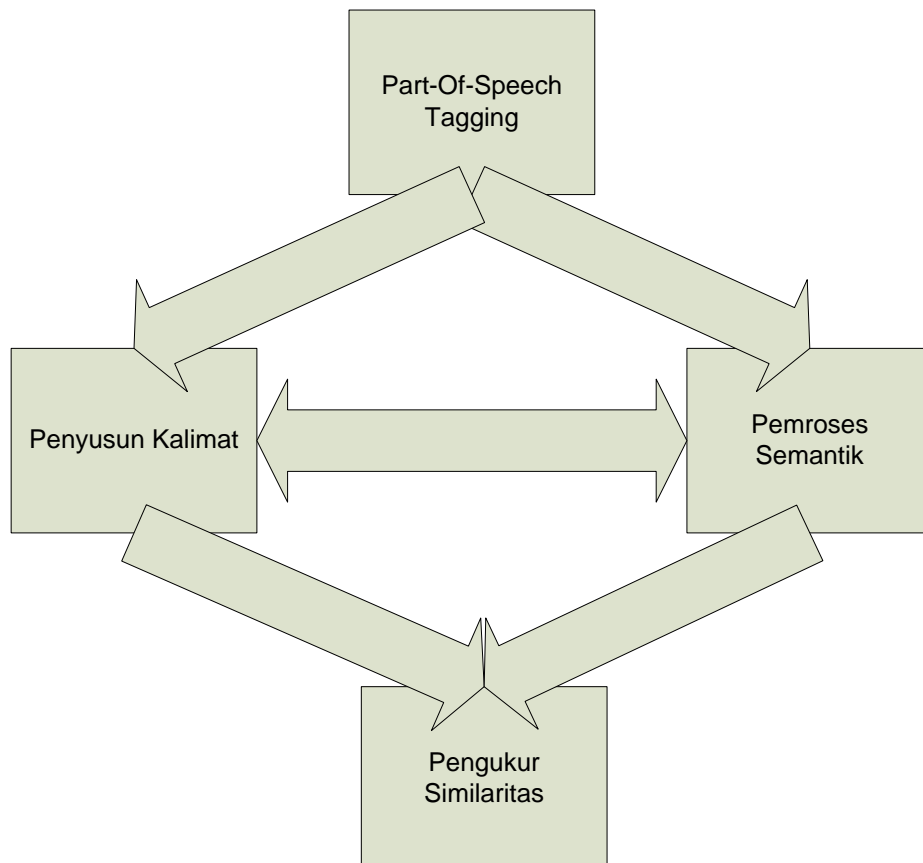
² Part-of-speech tagging adalah pemecahan struktur kalimat, menjadi bentuk subjek, predikat, objek dan bentuk yang lain.

³ Brown University Standard Corpus of Present-Day American English disingkat menjadi Brown Corpus, dibuat oleh Henry Kucera dan W. Nelson Francis pada tahun 1960 an.

Rancangan logika aplikasi ini adalah sebagai berikut:

- Kalimat dipecah strukturnya dengan menggunakan *Part-of-Speech Tagging* (POST).
- Kata benda dicari padanannya dengan metode semantik.
- Kata kerja dicari padanannya dengan metode semantik.
- Kata sifat dicari padanannya dengan metode semantik.
- Kata keterangan dicari padanannya dengan metode semantik.
- Struktur subjek dan objek ditukar bila merupakan kalimat transitif.
- Menambahkan kata depan maupun konjungsi untuk menyesuaikan struktur kalimat.
- Hitung tingkat kesamaan dari kalimat baru, bila di atas batas ambang maka masukkan di dalam daftar.

Desain Block Diagram Sistem



Input dari aplikasi berupa sebuah kalimat tunggal atau kalimat majemuk setara. Sedangkan hasil output dari aplikasi ini adalah berupa beberapa

rekomendasi kalimat yang mempunyai makna hampir sama dan tingkat kesamaan kalimat. Kalimat-kalimat yang ditampilkan adalah kalimat yang mempunyai tingkat kesamaan yang diatas batas ambang atau *threshold* tertentu.

8. BATASAN MASALAH

Sejumlah permasalahan yang dibahas dalam Tugas Akhir ini akan dibatasi ruang lingkup pembahasannya, antara lain :

- Bahasa yang didukung berupa bahasa Inggris.
- Kalimat yang dimaksud berupa kalimat sederhana, kalimat tunggal, dan kalimat majemuk setara, namun bukan majemuk bertingkat.

9. METODOLOGI

Pembuatan tugas akhir ini dilakukan dengan mengikuti metodologi sebagai berikut :

1. Studi literatur

Pada tahap ini dilakukan pengumpulan informasi yang diperlukan untuk perancangan sistem. Informasi tersebut dapat diperoleh dari literatur, *paper*, maupun buku-buku bahasa pemrograman.

Literatur-literatur yang digunakan antara lain adalah referensi dari website maupun e-book yang membahas masalah kalimat, *Part-Of-Speech Tagging*, sinonim kata, hipernim, hiponim maupun metode – metode yang lain.

Contoh dari tools rewriter yang ada di pasaran antara lain:

1. Autorewrite (<http://www.autorewrite.com/tool/>).
2. ArticleQueen: Article Rewriter & Article Spinner (<http://www.articlequeen.com/>).

Perbandingan tools-tools yang sudah ada:

No.	Nama Tools	Berbayar / Gratis / Open Source	Synonim	Struktur Kalimat	Aktif Pasif	Desktop / Online
1	Autorewrite	Berbayar	V	V	-	Desktop
2	Articlequeen	Gratis, namun harus mendaftar	V	V	-	Online

2. Analisis dan perancangan sistem

Pada tahap ini dilakukan analisis kebutuhan dan perancangan sistem untuk merumuskan solusi yang tepat dalam pembuatan aplikasi serta kemungkinan yang dapat dilakukan untuk mengimplementasikan rancangan tersebut.

Input dari sistem yang akan dibangun adalah sebuah kalimat sederhana. Hasil output dari aplikasi yang diharapkan adalah beberapa kalimat dengan struktur atau kata-kata yang berbeda namun memiliki makna yang hampir sama. Hasil output yang dibangun dibedakan berdasarkan jenis algoritma dari pencarian makna lain (sinonim, hiponim atau hipernim) dan algoritma untuk evaluasi kecocokan dari dua buah kalimat (JCn⁴, Lin⁵).

Algoritma JCn menggunakan gabungan dari metode node-based (information content) approach dengan metode edge-based (distance) approach. Metode ini memberikan bobot pada edge antara parent node p dan child node c_i dengan probabilitas kondisional $P(c_i, p)$. Perhitungan

probabilitas dihitung dengan cara
$$P(c_i|p) = \frac{P(c_i \cap p)}{P(p)} = \frac{P(c_i)}{P(p)}$$
, sehingga dapat didefinisikan *link strength* (LS) sebagai

$$LS(c_i, p) = -\log(P(c_i|p)) = IC(c_i) - IC(p)$$
. Secara

keseluruhan jarak antara dua node dapat didefinisikan dengan penjumlahan dari bobot edge-edge yang dilalui pada lintasan terpendek.

$$Dist(w_1, w_2) = \sum_{c \in \{path(c_1, c_2) - LSuper(c_1, c_2)\}} wt(c, parent(c))$$

Algoritma Lin menggunakan beberapa asumsi untuk mendefinisikan similaritas.

Asumsi pertama adalah Commonality antara A dan B diukur dengan : $I(common(A, B))$, dimana $common(A, B)$ adalah dalil yang menyatakan kesamaan dari A dan B. $I(S)$ adalah jumlah informasi tertentu yang terkandung dalam proposisi.

⁴ Lin, 1998

⁵ Jiang and Conrath, 1997

Asumsi kedua adalah perbedaan antara A dan B diukur dengan $I(\text{description}(A,B)) - I(\text{common}(A,B))$, dimana $\text{description}(A, B)$ adalah dalil yang menjelaskan apa yang A dan B deskripsikan.

Asumsi ketiga adalah similaritas antara A dan B, $\text{sim}(A, B)$ adalah sebuah fungsi yang dibuat dari commonality dan description dari A dan B, dimana $\text{sim}(A, B) = f(I(\text{common}(A, B)), I(\text{description}(A, B)))$, dan domain dari f adalah $\{ x, y \mid x \geq 0, y > 0, y \geq x \}$.

Asumsi keempat adalah similaritas antara dua buah objek yang identik adalah 1.

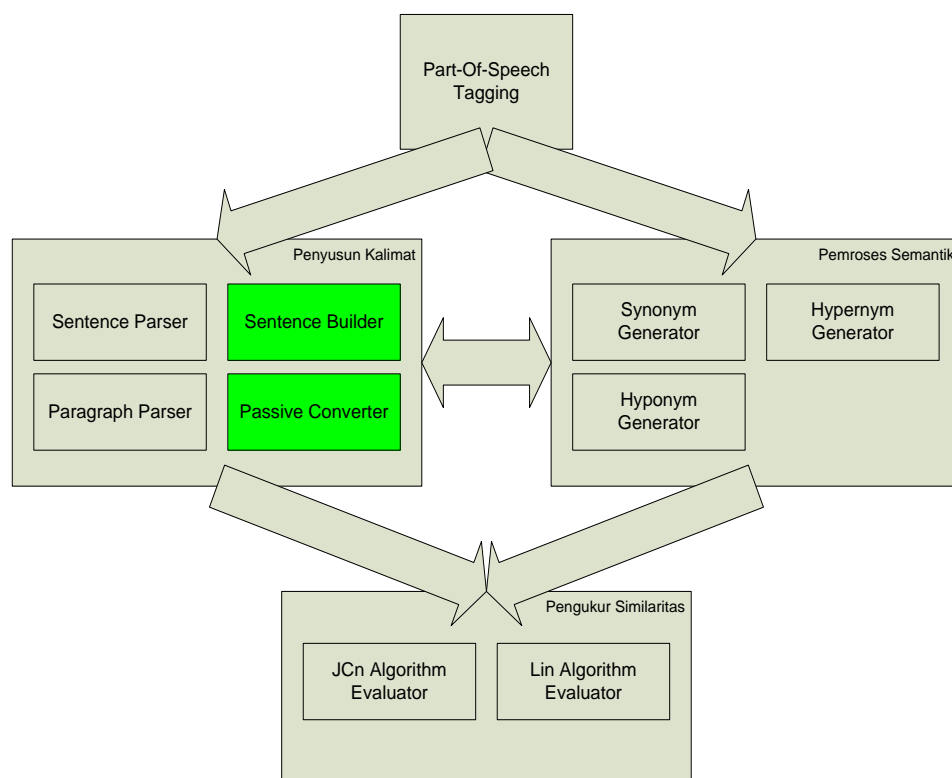
Asumsi kelima adalah $\forall y > 0, f(0, y) = 0$.

Asumsi keenam adalah
$$\forall x_1 \leq y_1, x_2 \leq y_2 : f(x_1 + x_2, y_1 + y_2) = \frac{y_1}{y_1 + y_2} f(x_1, y_1) + \frac{y_2}{y_1 + y_2} f(x_2, y_2)$$

Dari asumsi-asumsi tersebut dibuat teorema similaritas dengan

$$\text{sim}(A, B) = \frac{\log P(\text{common}(A, B))}{\log P(\text{description}(A, B))}$$

Dari analisa di atas, dibuatlah desain lengkap dari sistem sebagai berikut dengan menggunakan representasi blok diagram:



3. Implementasi

Pada tahap ini dilakukan implementasi sistem secara menyeluruh. Pengimplementasian dilakukan dengan berbekal pedoman-pedoman yang diperoleh pada tahap sebelumnya.

Implementasi dilakukan dengan bahasa pemrograman Java versi 1.7, menggunakan sistem operasi Windows 7. Hasil akhir produk berupa desktop application yang menggunakan library – library pendukung seperti Lingpipe, Wordnet, dll. Library Lingpipe digunakan untuk melakukan proses tagging terhadap *Part-Of-Speech*, sedangkan library wordnet digunakan untuk melakukan pencarian sinonim, hipernim dan hiponim pada kalimat, serta digunakan untuk melakukan evaluasi terhadap tingkat kesamaan pada kalimat mula-mula dan kalimat baru. Algoritma yang digunakan untuk melakukan evaluasi tingkat kesamaan adalah algoritma JCN, Lin dan algoritma-algoritma lain yang masih belum dapat ditentukan pada saat ini.

4. Ujicoba dan evaluasi

Pada tahap ini dilakukan ujicoba terhadap sistem yang dibuat, tujuannya untuk menemukan kesalahan-kesalahan yang mungkin terjadi serta melakukan perbaikan untuk lebih menyempurnakan sistem yang dibuat.

Evaluasi terhadap algoritma JCN dan Lin akan dilakukan untuk memperoleh hasil yang optimal terhadap penyaringan kalimat output yang dihasilkan. Evaluasi dilakukan dengan memberikan kategori – kategori pada kalimat yang digunakan untuk percobaan, akan diberikan beberapa macam kalimat yang mempunyai struktur berbeda, maupun dengan makna yang berbeda.

5. Penyusunan Buku Tugas Akhir

Pada tahap ini akan dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat.

10. DAFTAR PUSTAKA

- [1] [http://en.wikipedia.org/wiki/Sentence_\(linguistics\)](http://en.wikipedia.org/wiki/Sentence_(linguistics))
- [2] <http://en.wikipedia.org/wiki/Synonym>
- [3] http://en.wikipedia.org/wiki/Part_of_speech_tagging
- [4] http://en.wikipedia.org/wiki/Brown_Corpus#Part-of-speech_tags_used
- [5] Jay J. Jiang, David W. Conrath. 1997. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. Taiwan.
- [6] D. Lin. 1998. An Information-Theoretic Definition of Similarity. Canada.

JADWAL KEGIATAN

No.	Tahapan	Bulan				
		Ags'11	Sep'11	Okt'11	Nov'11	Des'11
1	Studi Literatur					
2	Analisis dan Desain Perangkat					
3	Implementasi Perangkat Lunak					
4	Uji Coba dan Evaluasi					
5	Penyusunan Buku Tugas Akhir					

LEMBAR PENGESAHAN

Surabaya, 22 September 2011

Menyetujui

Dosen pembimbing

Dr. Ir. Siti Rochimah, M.T

NIP. 132103631