

LAPORAN TAHUNAN
PENELITIAN FUNDAMENTAL



**EFEKTIVITAS ALGORITMA SIMILARITAS SEMANTIK
BERBASIS JEJARING KATA UNTUK MENGUKUR KEMIRIPAN
KALIMAT BAHASA INDONESIA**

Tahun ke-1 dari rencana 2 tahun

Husni Thamrin, MT, Ph.D (0618077001)
Dra. Atiqa Sabardila, M.Hum (0621066401)

UNIVERSITAS MUHAMMADIYAH SURAKARTA
OKTOBER 2014

HALAMAN PENGESAHAN

Judul Kegiatan : Efektivitas Algoritma Similaritas Semantik Berbasis Jejaring Kata dalam Mengukur Kemiripan Kalimat Bahasa Indonesia

Peneliti / Pelaksana

Nama Lengkap : HUSNI THAMRIN
NIDN : 0618077001
Jabatan Fungsional :
Program Studi : Teknik Informatika
Nomor HP : 081218822041
Surel (e-mail) : husni.thamrin@ums.ac.id

Anggota Peneliti (1)

Nama Lengkap : DRA. ATIQA SABARDILA M. HUM.
NIDN : 0621066401
Perguruan Tinggi : Universitas Muhammadiyah Surakarta

Institusi Mitra (jika ada)

Nama Institusi Mitra :
Alamat :
Penanggung Jawab :
Tahun Pelaksanaan : Tahun ke 1 dari rencana 2 tahun
Biaya Tahun Berjalan : Rp. 45.000.000,00
Biaya Keseluruhan : Rp. 114.949.600,00

Mengetahui
Wakil Dekan bid. Akademik



(Dr. Nurgiyatna, M.Eng)
NIP/NIK 881

Surakarta, 6 - 11 - 2014,
Ketua Peneliti,

(HUSNI THAMRIN)
NIP/NIK706

Menyetujui,
Kepala LPPM UMS



(Dr. Agus Ulinuha)
NIP/NIK 656

RINGKASAN

Algoritma similaritas semantik adalah urutan langkah untuk menghitung kemiripan kalimat atas dasar makna/konsep kata yang menyusun kalimat. Algoritma similaritas banyak diterapkan misalnya untuk *software* deteksi plagiasi, pengelompokan dokumen, dan *software* penerjemah. Penerapan algoritma similaritas semantik pada teks bahasa Indonesia belum banyak menunjukkan hasil antara lain karena belum adanya jejaring kata bahasa Indonesia. Penelitian ini mengupayakan penyusunan jejaring kata bahasa Indonesia dan melakukan studi “*Efektivitas Algoritma Similaritas Semantik Berbasis Jejaring Kata dalam Mengukur Kemiripan Kalimat Bahasa Indonesia.*”

Tujuan dari program penelitian ini pada tahun pertama ada dua. Pertama adalah membuat sistem basis data kamus bahasa Indonesia yang menyediakan informasi tentang kata termasuk dalam konteks jejaring kata melalui web service. Kedua adalah membuat aplikasi yang menerapkan algoritma similaritas semantik berbasis jejaring kata atau kamus untuk mengukur similaritas kata, frase atau kalimat dalam sebuah kalimat bahasa Indonesia.

Penelitian pada tahun pertama menghasilkan kamus *online* bahasa Indonesia yang dapat memberi informasi tentang kata, definisi kata, dan sinonim. Kamus online semacam ini sudah tersedia. Kamus yang dihasilkan dalam penelitian ini memberikan informasi tambahan yang tidak ada pada kamus lain yaitu informasi tentang hiponim dan meronim. Informasi tentang sinonim dan hiponim merupakan dasar pembentukan jejaring kata. Pengetahuan tentang kata, sinonim dan hiponim digunakan dalam penelitian ini sebagai ukuran keterkaitan antar kata. Keterkaitan antar kata digunakan dalam algoritma similaritas untuk mengukur kemiripan antar kalimat. Penggunaan pengetahuan tentang sinonim dan hiponim terbukti meningkatkan skor kemiripan antara kalimat yang mengandung kata berbeda namun mempunyai keterkaitan sebagai sinonim atau hiponim.

Pengujian algoritma dilakukan lebih lanjut dengan menerapkan algoritman pada proses pengelompokan teks (*text clustering*). Objek pengelompokan teks diambil dari ratusan kalimat pendek yang diperoleh dari analisis SWOT (*strength weakness opportunity threat*) sebuah lembaga. Pengujian menunjukkan bahwa klustering berdasarkan kemiripan kalimat yang memanfaatkan pengetahuan sinonim dan hiponim tidak lebih baik dibanding klustering dengan perhitungan kemiripan kosinus biasa.

Kesimpulan penelitian ini adalah bahwa skor kemiripan kalimat dapat diperbaiki jika algoritma memperhatikan keterkaitan makna antar kata. Pemanfaatan algoritma dalam proses klustering tidak memperbaiki hasil pengelompokan kalimat. Tahap berikutnya yang perlu dilakukan adalah membuat kumpulan dokumen (korpus) sebagai alat uji standar dari observasi tentang similaritas, menguji algoritma pada proses klasifikasi dokumen, dan menguji proses klustering pada dokumen dengan jumlah kata yang bervariasi.

Kata kunci: similaritas, similaritas semantik, sinonim, hiponim, bahasa Indonesia

PRAKATA

Puji Syukur kami panjatkan kehadiran Allah Subhanahu wa Ta'ala dengan selesainya penelitian tahun pertama hibah fundamental berjudul: **“EFEKTIVITAS ALGORITMA SIMILARITAS SEMANTIK BERBASIS JEJARING KATA UNTUK MENGUKUR KEMIRIPAN KALIMAT BAHASA INDONESIA.”**

Hingga tahap ini proses penelitian tidak akan berlangsung lancar tanpa bantuan berbagai pihak. Oleh karena itu penulis hendak mengucapkan terima kasih yang besar kepada:

1. Saudara Fauzan Natsir yang banyak membantu dalam proses penghimpunan data kamus dan entri data kamus hiponim dan meronim.
2. Bapak Dr. Nurgiyatna, M.Sc yang mengizinkan penulis menggunakan kalimat-kalimat dalam analisis SWOT di lembaga yang dipimpinnya sebagai bahan analisis similaritas.
3. Tim pengembang website kateglo.com yang menjadikan data kamus besar bahasa Indonesia tersedia secara online sehingga dapat digunakan dalam penelitian terkait kata-kata bahasa Indonesia.
4. Direktorat Jenderal Pendidikan Tinggi, Kemdiknas yang melalui Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Muhammadiyah Surakarta yang telah menyediakan dana dalam skema penelitian hibah fundamental sehingga penelitian ini dapat berjalan.

Kami telah berusaha melaksanakan kegiatan penelitian ini sebaik mungkin, namun kami menyadari bahwa tidak ada yang sempurna dari hasil karya manusia. Kritik dan saran sangat kami butuhkan untuk kebaikan dan penyempurnaan pelaksanaan penelitian ini selanjutnya.

Surakarta, 6 Nopember 2014

Peneliti

DAFTAR ISI

HALAMAN SAMBUNG.....	i
HALAMAN PENGESAHAN.....	ii
RINGKASAN.....	iii
PRAKATA.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	vi
DAFTAR LAMPIRAN.....	vii
BAB 1. PENDAHULUAN.....	1
BAB 2. TINJAUAN PUSTAKA.....	3
2.1. Penelitian yang Mendahului.....	3
2.2. Peta Jalan Penelitian.....	5
BAB 3. TUJUAN DAN MANFAAT PENELITIAN.....	8
3.1. Tujuan Penelitian.....	8
3.2. Manfaat Penelitian.....	8
BAB 4. METODE PENELITIAN.....	10
4.1. Perhitungan similaritas.....	13
4.2. Data uji.....	15
4.3. Penilaian Kinerja.....	16
BAB 5. HASIL PENELITIAN.....	17
5.1 Website Kamus.....	17
5.2. Konsep basis data kamus.....	19
5.3. Implementasi Jejaring Kata dalam Program Komputer.....	19
5.4. Data pengujian.....	23
5.5. Kinerja algoritma berbasis elemen jejaring kata.....	24
BAB 6. RENCANA TAHAPAN BERIKUTNYA.....	27
BAB 7. KESIMPULAN DAN SARAN.....	28
7.1. Kesimpulan.....	28
7.2. Saran.....	28
DAFTAR PUSTAKA.....	29

DAFTAR GAMBAR

Gambar 1. Peta jalan penelitian yang mengawali penelitian yang sedang diusulkan.....	7
Gambar 2. Rangkaian Kegiatan Penelitian "Efektivitas Algoritma Similaritas Semantik Berbasis Jejaring Kata untuk Mengukur Kemiripan Kalimat Bahasa Indonesia".	12
Gambar 3. Tampilan kamus online untuk kata "rumah". Pada kamus terdapat informasi tentang hiponim dan meronim dari kata yang dicari.....	17
Gambar 4. Data dalam format JSON yang menyajikan informasi tentang kata "bunga" sebagaimana ditampilkan melalui sebuah web browser.....	18
Gambar 5. Class diagram dari basis data kamus.....	19

DAFTAR LAMPIRAN

Lampiran 1. Instrumen penelitian.....	32
Lampiran 2. Personalia Tenaga Peneliti.....	32
Lampiran 3. Publikasi.....	32
Lampiran 4. Kluster dengan similaritas kosinus murni.....	45
Lampiran 5. Kluster dengan metode similaritas kosinus yang dimodifikasi.....	50
Lampiran 6. Produk Penelitian Lainnya.....	55