



UNIVERSITAS INDONESIA

**PENERJEMAHAN TEKS INDONESIA-INGGRIS OTOMATIS
MENGUNAKAN MESIN PENERJEMAH STATISTIK BERDASARKAN
FRASE DENGAN KOLEKSI DOKUMEN MENGGUNAKAN *PART-OF-
SPEECH TAGGING* DAN LEMA**

SKRIPSI

Metti Zakaria W.

1205000568

**PROGRAM : ILMU KOMPUTER
FAKULTAS : ILMU KOMPUTER**

**DEPOK
JULI, 2009**



UNIVERSITAS INDONESIA

**PENERJEMAHAN TEKS INDONESIA-INGGRIS OTOMATIS
MENGUNAKAN MESIN PENERJEMAH STATISTIK BERDASARKAN
FRASE DENGAN KOLEKSI DOKUMEN MENGGUNAKAN *PART-OF-
SPEECH TAGGING* DAN LEMA**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar S.Kom

Metti Zakaria W.

1205000568

PROGRAM : ILMU KOMPUTER

FAKULTAS : ILMU KOMPUTER

DEPOK

JULI, 2009

HALAMAN PERNYATAAN ORISINALITAS

**Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Metti Zakaria W.

NPM : 1205000568

Tanda Tangan :

Tanggal : 17 Juli 2009

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Metti Zakaria W.
NPM : 1205000568
Program Studi : Ilmu Komputer
Judul Skripsi : Penerjemahan Teks Indonesia-Inggris Otomatis Menggunakan Mesin Penerjemah Statistik Berdasarkan Frase dengan Koleksi Dokumen Menggunakan *Part-of-Speech Tagging* dan Lema

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana S.Kom pada program studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia

DEWAN PENGUJI

Pembimbing : Dra. Mirna Adriani Ph.D. ()

Penguji : Dina Chahyati S.Kom., M.Kom ()

Penguji : Dr. Hisar Maruli Manurung ()

Ditetapkan di : Fakultas Ilmu Komputer

Tanggal : 17 Juli 2009

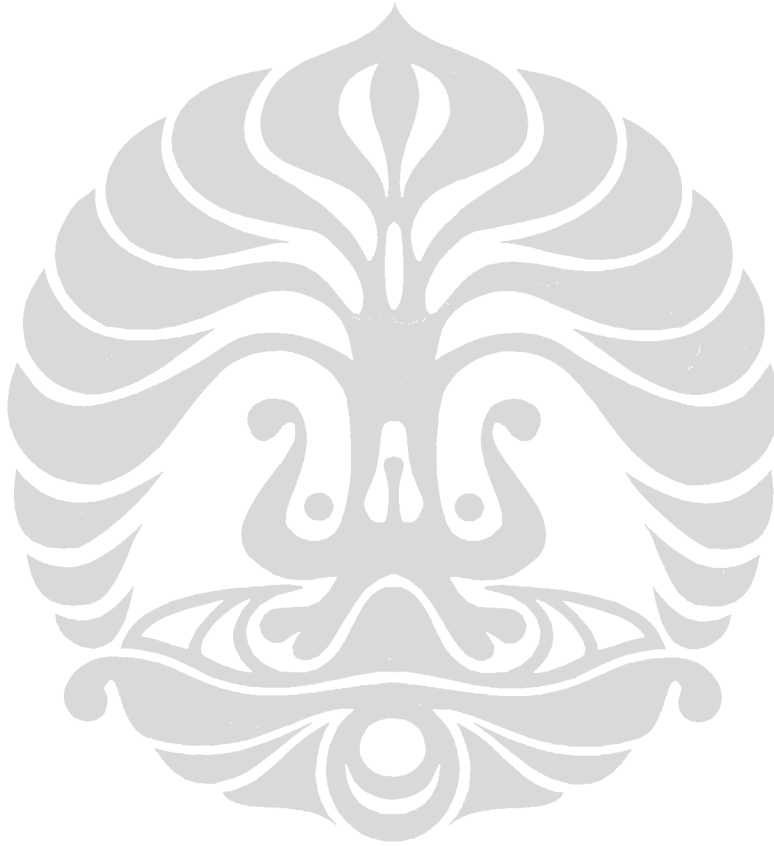
KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan karunia-Nya penulis dapat menyelesaikan tugas akhir dalam jangka waktu yang telah ditentukan dan menghasilkan laporan tugas akhir ini. Selama penulis melakukan tugas akhir, penulis mendapatkan bantuan dari berbagai pihak yang sangat berarti bagi penulis. Oleh sebab itu, penulis hendak menyampaikan ungkapan terima kasih kepada berbagai pihak sebagai berikut:

1. Orang tua, kakak dan adik serta anggota keluarga lainnya yang selalu mendoakan, memberi dukungan serta motivasi baik moral maupun material dan semangat pantang menyerah kepada penulis selama melakukan tugas akhir.
2. Ibu Mirna Adriani selaku dosen pembimbing tugas akhir.
3. Bapak Suryana Setiawan selaku pembimbing akademis.
4. Rekan-rekan yang ada di Lab *IR*, yang telah membantu, mendukung dan menemani penulis selama pengerjaan tugas akhir.
5. Anak-anak Fasilkom 2008, yaitu Frisca dan Yessica Sari yang selalu memberikan perhatian serta semangat kepada penulis.
6. Anak-anak KMBUI, diantaranya adalah Melissa FISIP angkatan 2005 yang juga selalu mendoakan serta menyemangati penulis untuk dapat menyelesaikan tugas akhir dengan baik dan tepat pada waktunya.
7. Rekan-rekan mahasiswa pelaksana tugas akhir yang sama-sama dibimbing oleh Ibu Mirna Adriani, yang telah membantu penulis dalam menyelesaikan tugas akhir.
8. Anak-anak angkatan 2005, yang telah menemani penulis selama 4 tahun di Fakultas Ilmu Komputer telah memberikan kenangan-kenangan yang indah selama 4 tahun tersebut.

9. Semua pihak yang tidak bisa penulis sebutkan di sini satu per satu yang telah ikut membantu penulis dalam menyelesaikan tugas akhir ini dengan baik dan tepat waktu.

Penulis sangat sadar bahwa dalam melakukan penelitian ini banyak kekurangan dan kesalahan yang telah penulis lakukan. Oleh karena itu, dengan segala kerendahan hati penulis bersedia menerima setiap saran dan kritik yang membangun. Semoga laporan tugas akhir ini dapat bermanfaat bagi semua pembaca di masa yang akan datang.



Depok, Juli 2009

Metti Zakaria W.

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Metti Zakaria W.

NPM : 1205000568

Program Studi : Ilmu Komputer

Fakultas : Ilmu Komputer

Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

Penerjemahan Teks Indonesia-Inggris Otomatis Menggunakan Mesin Penerjemah Statistik Berdasarkan Frase dengan Koleksi Dokumen Menggunakan *Part-of-Speech Tagging* dan Lema

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 17 Juli 2009

Yang menyatakan

(Metti Zakaria W.)

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	II
HALAMAN PENGESAHAN	III
KATA PENGANTAR.....	IV
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	VI
ABSTRAK.....	VII
ABSTRACT	VIII
DAFTAR ISI.....	IX
DAFTAR TABEL	XII
DAFTAR GAMBAR.....	XIII
DAFTAR LAMPIRAN	XV
BAB 1 PENDAHULUAN.....	1
1.1 LATAR BELAKANG	1
1.2 PERMASALAHAN	4
1.3 TUJUAN	4
1.4 RUANG LINGKUP.....	5
1.5 METODOLOGI PENELITIAN.....	5
1.6 SISTEMATIKA PENULISAN.....	6
BAB 2 LANDASAN TEORI.....	8
2.1 MESIN PENERJEMAH.....	8
2.1.1 Mesin Penerjemah berdasarkan Aturan	9
2.1.2 Mesin Penerjemah berdasarkan Contoh	11
2.2 MESIN PENERJEMAH STATISTIK.....	12
2.3 MESIN PENERJEMAH STATISTIK BERDASARKAN FRASE.....	14
2.3.1 Model Penerjemahan berdasarkan Frase	14
2.3.2 Model Penerjemahan yang Difaktorkan	17

Universitas Indonesia

2.3.2.1	Lematisasi.....	20
2.3.2.2	<i>Part-of-Speech Tagging</i>	21
2.3.3	Model Bahasa	25
2.3.4	<i>Decoding</i>	27
2.3.4.1	Pilihan Terjemahan	29
2.3.4.2	Algoritma Inti	30
2.3.4.3	Rekombinasi Hipotesa	32
2.3.4.4	<i>Beam Search</i>	32
2.4	EVALUASI MESIN PENERJEMAH STATISTIK	36
2.4.1	N-Gram <i>Precision</i>	36
2.4.2	N-Gram <i>Recall</i>	38
2.5	BEBERAPA PENELITIAN SEBELUMNYA	40
2.5.1	Penelitian Koehn dan Hoang.....	40
2.5.2	Penelitian Marine Carpuat	42
2.5.3	Penelitian David Chiang	42
2.6	FOKUS PENELITIAN	43
BAB 3	EKSPERIMEN.....	44
3.1	DATA.....	44
3.1.1	Korpus Dwibahasa Indonesia-Inggris.....	44
3.1.2	Daftar Kata Bahasa Indonesia yang sudah Diberikan POS Tag	48
3.1.3	Kamus Besar Bahasa Indonesia	49
3.2	APLIKASI YANG DIGUNAKAN.....	50
3.2.1	ABBYY FineReader.....	50
3.2.2	NLProcessor	51
3.2.3	Moses	53
3.2.4	mteval.....	54
3.2.5	Program Tambahan.....	54
3.2.5.1	Program Penambahan POS Tag pada Korpus Pelatihan.....	55
3.2.5.2	Program Penambahan Lema pada Korpus Pelatihan	57
3.3	SKENARIO EKSPERIMEN	60
3.3.1	Eksperimen Penggunaan Korpus Jenis 1	63
3.3.2	Eksperimen Penggunaan Korpus Jenis 2	66

3.3.3	Eksperimen Penggunaan Korpus Jenis 3	70
3.3.4	Eksperimen Penggunaan Korpus Jenis 4	73
BAB 4 HASIL EKSPERIMEN DAN ANALISIS		78
4.1	HASIL EKSPERIMEN.....	78
4.1.1	Hasil Eksperimen Penggunaan Korpus Jenis 1	79
4.1.2	Hasil Eksperimen Penggunaan Korpus Jenis 2	82
4.1.3	Hasil Eksperimen Penggunaan Korpus Jenis 3	85
4.1.4	Hasil Eksperimen Penggunaan Korpus Jenis 4	89
4.2	ANALISIS HASIL EKSPERIMEN.....	92
4.2.1	Analisis Hasil Eksperimen Korpus Jenis 1	93
4.2.2	Analisis Hasil Eksperimen Korpus Jenis 2	100
4.2.3	Analisis Hasil Eksperimen Korpus Jenis 3	106
4.2.4	Analisis Hasil Eksperimen Korpus Jenis 4	111
4.2.5	Analisis Semua Jenis Korpus	116
BAB 5 PENUTUP		119
5.1	KESIMPULAN	119
5.2	SARAN.....	120
DAFTAR PUSTAKA.....		123

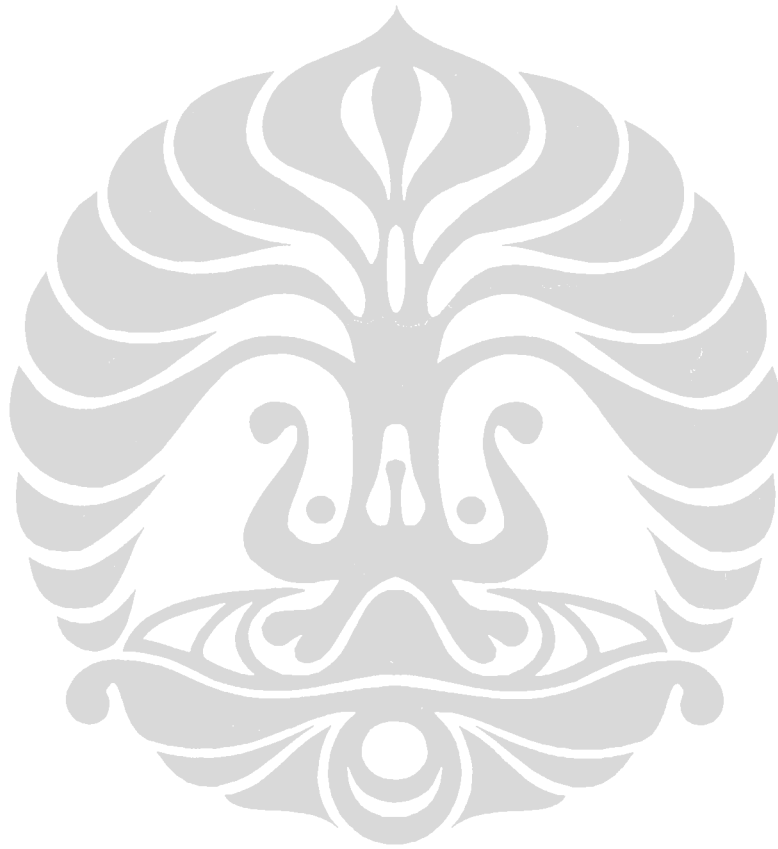
DAFTAR TABEL

Tabel 2.1 Contoh Korpus Dwibahasa Paralel	11
Tabel 2.2 <i>Tag-set</i> Penn Treebank yang Dimodifikasi (Kategori <i>Open Class</i>).....	22
Tabel 2.3 <i>Tag-set</i> Penn Treebank yang Dimodifikasi (Kategori <i>Closed Class</i>) ...	23
Tabel 2.4 <i>Tag-set</i> Bahasa Indonesia (Kategori <i>Open</i> dan <i>Close Class</i>)	24
Tabel 3.1 Korpus Dwibahasa Indonesia-Inggris yang Digunakan	47
Tabel 3.2 Contoh Daftar Acuan Kata Bahasa Indonesia dan POS Tag-nya	48
Tabel 3.3 Contoh Isi KBBI yang Dimodifikasi	49
Tabel 3.4 Jumlah Korpus Dwibahasa Pelatihan dan Pengujian yang Digunakan .	62
Tabel 4.1 Hasil Evaluasi Korpus Jenis 1 (Berita).....	79
Tabel 4.2 Hasil Evaluasi Korpus Jenis 1 (Kitab Suci).....	80
Tabel 4.3 Hasil Evaluasi Korpus Jenis 1 (Novel).....	80
Tabel 4.4 Hasil Evaluasi Korpus Jenis 1 (Percakapan)	81
Tabel 4.5 Hasil Evaluasi Korpus Jenis 1 (Gabungan)	82
Tabel 4.6 Hasil Evaluasi Korpus Jenis 2 (Berita).....	82
Tabel 4.7 Hasil Evaluasi Korpus Jenis 2 (Kitab Suci).....	83
Tabel 4.8 Hasil Evaluasi Korpus Jenis 2 (Novel).....	84
Tabel 4.9 Hasil Evaluasi Korpus Jenis 2 (Percakapan)	84
Tabel 4.10 Hasil Evaluasi Korpus Jenis 2 (Gabungan)	85
Tabel 4.11 Hasil Evaluasi Korpus Jenis 3 (Berita).....	86
Tabel 4.12 Hasil Evaluasi Korpus Jenis 3 (Kitab Suci).....	86
Tabel 4.13 Hasil Evaluasi Korpus Jenis 3 (Novel).....	87
Tabel 4.14 Hasil Evaluasi Korpus Jenis 3 (Percakapan)	88
Tabel 4.15 Hasil Evaluasi Korpus Jenis 3 (Gabungan)	89
Tabel 4.16 Hasil Evaluasi Korpus Jenis 4 (Berita).....	89
Tabel 4.17 Hasil Evaluasi Korpus Jenis 4 (Kitab Suci).....	90
Tabel 4.18 Hasil Evaluasi Korpus Jenis 4 (Novel).....	91
Tabel 4.19 Hasil Evaluasi Korpus Jenis 4 (Percakapan)	91
Tabel 4.20 Hasil Evaluasi Korpus Jenis 4 (Gabungan)	92

DAFTAR GAMBAR

Gambar 2.1 Penerjemahan pada Mesin Penerjemah berdasarkan Aturan	10
Gambar 2.2 Proses Penerjemahan pada MPS	14
Gambar 2.3 <i>Phrase-based Translation</i>	15
Gambar 2.4 Pilihan Terjemahan	29
Gambar 2.5 Proses Generasi Kalimat <i>Output</i> dengan Algoritma <i>Beam Search</i> ...	30
Gambar 2.6 Estimasi <i>Future Cost</i> pada <i>Beam Search</i>	33
Gambar 2.7 <i>Pseudo-code</i> Algoritma <i>Beam Search</i>	34
Gambar 2.8 <i>Stacks</i> Hipotesa	35
Gambar 3.1 ABBYY FineReader	51
Gambar 3.2 Contoh <i>Output</i> NLProcessor	52
Gambar 3.3 Contoh Korpus Pelatihan Inggris dengan POS Tag	52
Gambar 3.4 Contoh Korpus Pelatihan Indonesia dengan POS Tag	56
Gambar 3.5 Contoh Korpus Pelatihan Indonesia dengan Lema	59
Gambar 3.6 Contoh Korpus Pelatihan Bahasa dengan Lema dan POS Tag	59
Gambar 3.7 Contoh Korpus Pelatihan Inggris dengan Lema	59
Gambar 3.8 Contoh Korpus Pelatihan Inggris dengan Lema dan POS Tag	60
Gambar 3.9 Proses Eksperimen yang Dilakukan	61
Gambar 3.10 <i>Script</i> Pelatihan Korpus Jenis 1	64
Gambar 3.11 Cara Menjalankan Berkas <code>moses.ini</code>	66
Gambar 3.12 <i>Script</i> Pelatihan Korpus Jenis 2	68
Gambar 3.13 <i>Script</i> Pelatihan Korpus Jenis 3	71
Gambar 3.14 <i>Script</i> Pelatihan Korpus Jenis 4	75
Gambar 4.1 Contoh Hasil Eksperimen Tahap Pengujian	78
Gambar 4.2 BLEU <i>Score</i> Percakapan Korpus Jenis 1 (Semua Model Bahasa)	94
Gambar 4.3 BLEU <i>Score</i> Novel Korpus Jenis 1 (Semua Model Bahasa)	96
Gambar 4.4 BLEU <i>Score</i> Kitab Suci Korpus Jenis 1 (Semua Model Bahasa)	97
Gambar 4.5 BLEU <i>Score</i> Percakapan Korpus Jenis 2 (Semua Model Bahasa) ...	101
Gambar 4.6 BLEU <i>Score</i> Novel Korpus Jenis 2 (Semua Model Bahasa)	103
Gambar 4.7 BLEU <i>Score</i> Kitab Suci Korpus Jenis 2 (Semua Model Bahasa)	104
Gambar 4.8 BLEU <i>Score</i> Percakapan Korpus Jenis 3 (Semua Model Bahasa) ...	107

Gambar 4.9 BLEU <i>Score</i> Novel Korpus Jenis 3 (Semua Model Bahasa).....	108
Gambar 4.10 BLEU <i>Score</i> Kitab Suci Korpus Jenis 3 (Semua Model Bahasa)...	110
Gambar 4.11 BLEU <i>Score</i> Percakapan Korpus Jenis 4 (Semua Model Bahasa) .	112
Gambar 4.12 BLEU <i>Score</i> Novel Korpus Jenis 4 (Semua Model Bahasa).....	113
Gambar 4.13 BLEU <i>Score</i> Kitab Suci Korpus Jenis 4 (Semua Model Bahasa)...	115
Gambar 4.14 BLEU <i>Score</i> Tertinggi Setiap Kategori pada Semua Jenis Korpus	116



DAFTAR LAMPIRAN

Lampiran 1: Contoh Korpus Pelatihan	128
Lampiran 2: Contoh Korpus Pengujian	147
Lampiran 3: Contoh Hasil Eksperimen Tahap Pengujian (Hasil Terjemahan)...	148
Lampiran 4: Contoh Kamus Besar Bahasa Indonesia (KBBI).....	154
Lampiran 5: Contoh Daftar Acuan Kata Indonesia yang telah Diberi POS Tag.	156
Lampiran 6: Contoh Hasil <i>Output</i> mteval.....	158

