



UNIVERSITAS INDONESIA

**PENERJEMAHAN TEKS INDONESIA – JEPANG OTOMATIS
MENGUNAKAN MESIN PENERJEMAH STATISTIK BERDASARKAN
FRASE**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

BERNADIA PUSPASARI

1205000223

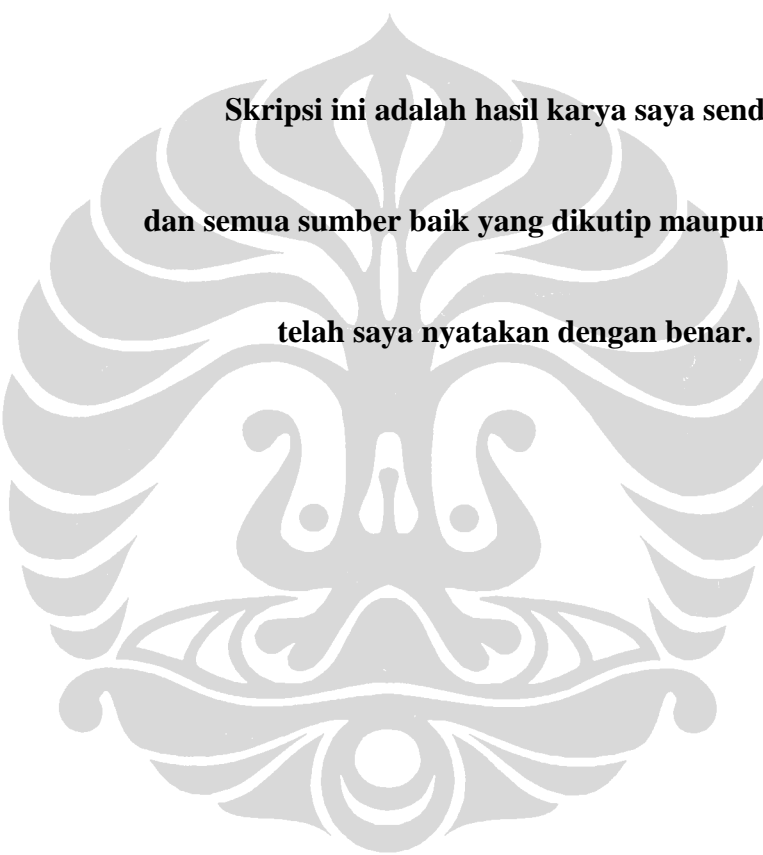
FAKULTAS ILMU KOMPUTER

PROGRAM 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 : Bernadia Puspasari

NPM : 1205000223

Tanda Tangan :

Tanggal : 21 Juli 2009

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Bernadia Puspasari

NPM : 1205000223

Program Studi : Ilmu Komputer

Judul Skripsi : Penerjemahan Teks Indonesia – Jepang Otomatis
Menggunakan Mesin Penerjemah Statistik Berdasarkan Frase

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia

DEWAN PENGUJI

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

Penguji : Dr. Hisar Maruli Manurung (.....)

Penguji : Adila Alfa Krisnadhi, M. Sc. (.....)

Ditetapkan di : Fakultas Ilmu Komputer

Tanggal : 21 Juli 2009

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan tugas akhir ini dalam waktu yang telah ditentukan. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan hingga pada penulisan laporan tugas akhir ini, sangatlah sulit bagi penulis untuk menyelesaikan tugas akhir ini. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada berbagai pihak sebagai berikut:

1. Orang tua serta kakak penulis yang senantiasa mendoakan serta memberikan dukungan moral dan material selama pengerjaan tugas akhir ini.
2. Ibu Mirna Adriani, selaku dosen pembimbing tugas akhir, atas segala bimbingan dan perhatiannya selama pengerjaan tugas akhir ini.
3. Bapak Hisar Maruli Manurung dan Bapak Adila Alfa Krisnadhi, selaku dosen penguji tugas akhir, atas segala saran dan kritik untuk laporan tugas akhir ini.
4. Bapak Johny Moningka dan Ibu Dina Cahyati, selaku dosen pembimbing akademis, atas segala bimbingan serta nasihat yang telah diberikan selama penulis menjalani masa studi di Fakultas Ilmu Komputer ini.
5. Naomi P. Adhi, Phillip L., dan Syafril B., yang telah membantu penulis mengumpulkan korpus paralel untuk pengerjaan tugas akhir ini, serta Hugo R. yang telah membantu proses *scanning* dokumen-dokumen tercetak.
6. Armando Y., Clara V., Darwin C., Deni L. Hakim, Metti Z. W., Sugianto A., Refly H., Bayu D. T., Hansel T., Suryanto A., Teddy, Vinky H., dan seluruh rekan-rekan yang berada di lab *Information Retrieval* yang telah membantu serta memberikan semangat selama pengerjaan tugas akhir
7. Bambang Adhi P., Evi D. J., Mursal R., Yohanes I., Yenni N., Tiara M. P., dan Rizki M., yang telah memberikan semangat kepada penulis.

8. Clara V., Dyta A., Adiyana Eka W., Naomi P. Adhi, dan Fachmy M. F., yang telah menjadi teman baik penulis serta selalu memberikan semangat kepada penulis dalam pengerjaan tugas akhir ini.
9. Rekan-rekan Fasilkom UI angkatan 2005, yang telah menemani penulis selama 4 tahun di Fasilkom serta memberikan kenangan indah dalam kebersamaan kita.
10. Anak-anak KUKSA CSUI yang selalu menemani penulis dalam suka dan duka sebagai satu keluarga.
11. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu penulis dalam menyelesaikan tugas akhir ini dengan baik.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam pengerjaan tugas akhir ini. Oleh karena itu, penulis sangat terbuka untuk setiap saran dan kritik yang membangun, untuk pengembangan tugas akhir ini selanjutnya. Semoga laporan tugas akhir ini dapat bermanfaat bagi pembacanya di masa yang akan datang.

Depok, Juli 2009

Bernadia Puspasari

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Bernadia Puspasari

NPM : 1205000223

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 – Jepang Otomatis Menggunakan Mesin Penerjemah Statistik Berdasarkan Frase

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 : 21 Juli 2009

Yang menyatakan

(Bernadia Puspasari)

vi

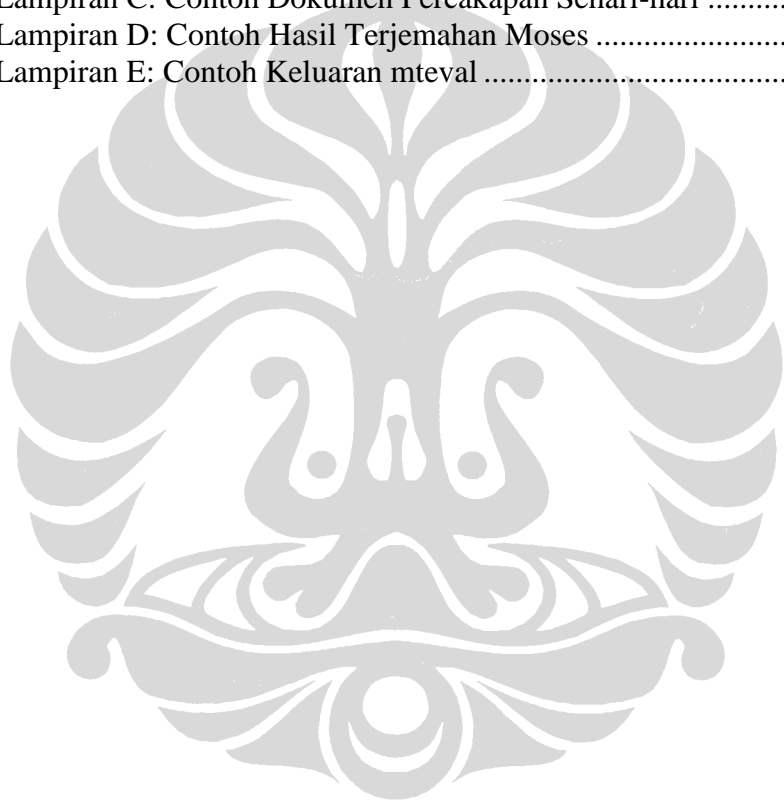
Universitas Indonesia

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
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan	4
1.3 Tujuan Penelitian	4
1.4 Ruang Lingkup.....	5
1.5 Metodologi Penelitian.....	5
1.6 Sistematika Penulisan Laporan	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	10
2.2 Mesin Penerjemah Statistik.....	11
2.2.1 Model Bahasa.....	13
2.2.2 Model Penerjemahan berdasarkan Frase.....	14
2.2.3 Model Penerjemahan yang Difaktorkan	16
2.2.4 Penyejajaran Kata	18
2.2.5 <i>Decoding</i>	20
2.2.5.1 Pilihan Terjemahan	22
2.2.5.2 Algoritma Inti.....	23
2.2.5.3 Hipotesis Rekombinasi	24
2.2.5.4 <i>Beam Search</i>	24
2.3 Analisis Morfologi untuk Bahasa Jepang	25
2.4 Evaluasi Mesin Penerjemah Statistik.....	27
2.5 Penelitian Sebelumnya.....	30
2.5.1 Penelitian Philipp Koehn dan Hieu Hoang	31
2.5.2 Penelitian Ondrej Bojar.....	32
2.5.3 Penelitian Erik A. Chapla.....	32
2.6 Fokus Penelitian	33
BAB 3 EKSPERIMEN	34
3.1 Data	34
3.1.1 Korpus Dwibahasa Paralel Indonesia - Jepang.....	34

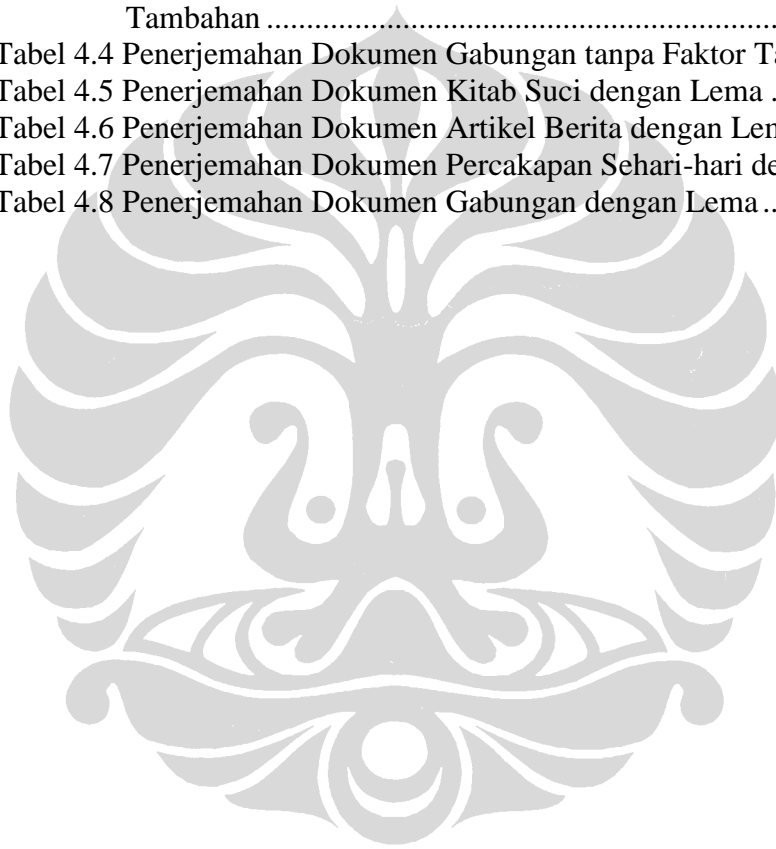
3.1.2	Kamus Besar Bahasa Indonesia	37
3.2	Aplikasi yang Digunakan	38
3.2.1	FreeOCR 2.6	38
3.2.2	Juman	39
3.2.3	Kakasi	41
3.2.4	Moses	43
3.2.5	mteval.....	44
3.2.6	Program Tambahan	44
3.3	Skenario Eksperimen	45
3.3.1	Eksperimen Menggunakan Korpus tanpa Faktor Tambahan.....	47
3.3.1.1	Eksperimen dengan Menggunakan Model Bahasa 1-gram	47
3.3.1.2	Eksperimen dengan Menggunakan Model Bahasa 3-gram	50
3.3.1.3	Eksperimen dengan Menggunakan Model Bahasa 5-gram	52
3.3.2	Eksperimen Menggunakan Korpus dengan Faktor Tambahan Lema.....	53
3.3.2.1	Eksperimen dengan Menggunakan Model Bahasa 1-gram	53
3.3.2.2	Eksperimen dengan Menggunakan Model Bahasa 3-gram	57
3.3.2.3	Eksperimen dengan Menggunakan Model Bahasa 5-gram	59
BAB 4 HASIL EKSPERIMEN DAN ANALISIS		61
4.1	Hasil Eksperimen	61
4.1.1	Hasil Eksperimen Menggunakan Korpus tanpa Faktor Tambahan	61
4.1.1.1	Hasil Eksperimen dengan Jenis Koleksi Dokumen Kitab Suci.....	62
4.1.1.2	Hasil Eksperimen dengan Jenis Koleksi Dokumen Artikel Berita.....	62
4.1.1.3	Hasil Eksperimen dengan Jenis Koleksi Dokumen Percakapan Sehari-hari.....	63
4.1.1.4	Hasil Eksperimen dengan Jenis Koleksi Dokumen Gabungan	64
4.1.2	Hasil Eksperimen Menggunakan Korpus dengan Faktor Tambahan Lema	65
4.1.2.1	Hasil Eksperimen dengan Jenis Koleksi Dokumen Kitab Suci.....	65
4.1.2.2	Hasil Eksperimen dengan Jenis Koleksi Dokumen Artikel Berita.....	66
4.1.2.3	Hasil Eksperimen dengan Jenis Koleksi Dokumen Percakapan Sehari-hari.....	67
4.1.2.4	Hasil Eksperimen dengan Jenis Koleksi Dokumen Gabungan	67

4.2	Analisis Hasil Eksperimen.....	68
4.2.1.	Analisis Hasil Eksperimen Menggunakan Korpus Pelatihan tanpa Faktor Tambahan.....	68
4.2.2	Analisis Hasil Eksperimen Menggunakan Korpus Pelatihan dengan Lema.....	76
BAB 5 PENUTUP		83
5.1	Kesimpulan	83
5.2	Saran	84
DAFTAR PUSTAKA		86
Lampiran A: Contoh Dokumen Kitab Suci.....		90
Lampiran B: Contoh Dokumen Artikel Berita.....		92
Lampiran C: Contoh Dokumen Percakapan Sehari-hari		93
Lampiran D: Contoh Hasil Terjemahan Moses		94
Lampiran E: Contoh Keluaran mteval		95



DAFTAR TABEL

Tabel 2.1 Contoh Kalimat Masukan pada Mesin Penerjemah Berdasarkan Contoh	10
Tabel 3.1 Tabel Korpus Dwibahasa Paralel yang Digunakan	37
Tabel 3.2 Pasangan Kata dan Lema	38
Tabel 3.3 Tabel Perubahan Hasil Konversi	43
Tabel 3.4 Jumlah Kalimat Dalam Korpus Indonesia – Jepang Kanji	47
Tabel 4.1 Penerjemahan Dokumen Kitab Suci tanpa Faktor Tambahan	62
Tabel 4.2 Penerjemahan Dokumen Artikel Berita tanpa Faktor Tambahan	63
Tabel 4.3 Penerjemahan Dokumen Percakapan Sehari-hari tanpa Faktor Tambahan	64
Tabel 4.4 Penerjemahan Dokumen Gabungan tanpa Faktor Tambahan.....	64
Tabel 4.5 Penerjemahan Dokumen Kitab Suci dengan Lema	65
Tabel 4.6 Penerjemahan Dokumen Artikel Berita dengan Lema	66
Tabel 4.7 Penerjemahan Dokumen Percakapan Sehari-hari dengan Lema	67
Tabel 4.8 Penerjemahan Dokumen Gabungan dengan Lema	68



DAFTAR GAMBAR

Gambar 2.1 Model Dasar Sistem Mesin Penerjemah Statistik	12
Gambar 2.2 Penerjemahan berdasarkan Frase	14
Gambar 2.3 Kumpulan Pilihan Terjemahan [Koehn, 2006]	23
Gambar 2.4 Algoritma <i>Beam Search</i> untuk Menghasilkan Kalimat Terjemahan [Koehn, 2006]	23
Gambar 2.5 <i>Pseudocode Beam Search</i> [Koehn, 2006]	25
Gambar 2.6 Contoh Hasil Analisis Morfologi Bahasa Jepang [Kudo, 2004] “telah diolah kembali”	26
Gambar 2.7 Contoh Hasil Terjemahan Mesin Penerjemah Statistik dengan Salah Satu Kata <i>Overgenerate</i>	29
Gambar 3.1 Contoh Tampilan FreeOCR 2.6	39
Gambar 3.2 Menjalankan Juman dari <i>Command Line</i>	40
Gambar 3.3 Contoh Keluaran Juman	41
Gambar 3.4 Perintah Untuk Menghasilkan Keluaran Dalam Bentuk Romaji	41
Gambar 3.5 Contoh Berkas Masukan untuk Kakasi	42
Gambar 3.6 Contoh Berkas Hasil Konversi Menggunakan Kakasi	42
Gambar 3.7 Contoh Keluaran Program Penambahan Lema	45
Gambar 3.8 Alur Pelaksanaan Eksperimen	46
Gambar 3.9 Contoh <i>Script</i> Pelatihan Korpus tanpa Faktor dengan Model Bahasa 1-gram	49
Gambar 3.10 <i>Script</i> Menjalankan <i>moses.ini</i>	50
Gambar 3.11 Contoh <i>Script</i> Pelatihan Korpus tanpa Faktor dengan Model Bahasa 3-gram	51
Gambar 3.12 Contoh <i>Script</i> Pelatihan Korpus tanpa Faktor dengan Model Bahasa 5-gram	52
Gambar 3.13 Contoh <i>Script</i> Pelatihan Korpus dengan Faktor Tambahan Lema dan menggunakan Model Bahasa 1-gram	55
Gambar 3.14 Contoh <i>Script</i> Pelatihan Korpus dengan Faktor Tambahan Lema dan menggunakan Model Bahasa 3-gram	58
Gambar 3.15 Contoh <i>Script</i> Pelatihan Korpus dengan Faktor Tambahan Lema dan menggunakan Model Bahasa 5-gram	59
Gambar 4.1 Nilai Akurasi Penerjemahan Dokumen Kitab Suci (dalam bentuk Kanji) tanpa Faktor Tambahan	69
Gambar 4.2 Nilai Akurasi Penerjemahan Dokumen Kitab Suci (dalam bentuk Romaji) tanpa Faktor Tambahan	69
Gambar 4.3 Nilai Akurasi Penerjemahan Dokumen Artikel Berita (dalam bentuk Kanji) tanpa Faktor Tambahan	70
Gambar 4.4 Nilai Akurasi Penerjemahan Dokumen Artikel Berita (dalam bentuk Romaji) tanpa Faktor Tambahan	71
Gambar 4.5 Nilai Akurasi Penerjemahan Dokumen Percakapan Sehari- hari (dalam bentuk Kanji) tanpa Faktor Tambahan	72
Gambar 4.6 Nilai Akurasi Penerjemahan Dokumen Percakapan Sehari- hari (dalam bentuk Romaji) tanpa Faktor Tambahan	73

Gambar 4.7 Nilai Akurasi Penerjemahan Dokumen Gabungan (dalam bentuk Kanji) tanpa Faktor Tambahan.....	75
Gambar 4.8 Nilai Akurasi Penerjemahan Dokumen Gabungan (dalam bentuk Romaji) tanpa Faktor Tambahan.....	75
Gambar 4.9 Nilai Akurasi Penerjemahan Dokumen Kitab Suci (dalam bentuk Kanji) dengan Faktor Lema	76
Gambar 4.10 Nilai Akurasi Penerjemahan Dokumen Kitab Suci (dalam bentuk Romaji) dengan Faktor Lema.....	77
Gambar 4.11 Nilai Akurasi Penerjemahan Dokumen Artikel Berita (dalam bentuk Kanji) dengan Faktor Lema	78
Gambar 4.12 Nilai Akurasi Penerjemahan Dokumen Artikel Berita (dalam bentuk Romaji) dengan Faktor Lema.....	79
Gambar 4.13 Nilai Akurasi Penerjemahan Dokumen Percakapan Sehari-hari (dalam bentuk Kanji) dengan Faktor Lema	79
Gambar 4.14 Nilai Akurasi Penerjemahan Dokumen Percakapan Sehari-hari (dalam bentuk Romaji) dengan Faktor Lema	80
Gambar 4.15 Nilai Akurasi Penerjemahan Dokumen Gabungan (dalam bentuk Kanji) dengan Faktor Lema	81
Gambar 4.16 Nilai Akurasi Penerjemahan Dokumen Gabungan (dalam bentuk Romaji) dengan Faktor Lema.....	81

