



UNIVERSITAS INDONESIA

**PENERJEMAHAN DOKUMEN INGGRIS-INDONESIA
MENGUNAKAN MESIN PENERJEMAH STATISTIK
DENGAN *WORD REORDERING* DAN *PHRASE REORDERING***

SKRIPSI

**Hansel Tanuwijaya
120500041X**

**FAKULTAS ILMU KOMPUTER
PROGRAM ILMU KOMPUTER
DEPOK
JULI 2009**



UNIVERSITAS INDONESIA

**PENERJEMAHAN DOKUMEN INGGRIS-INDONESIA
MENGUNAKAN MESIN PENERJEMAH STATISTIK
DENGAN *WORD REORDERING* DAN *PHRASE REORDERING***

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Ilmu Komputer**

**Hansel Tanuwijaya
120500041X**

**FAKULTAS ILMU KOMPUTER
PROGRAM ILMU KOMPUTER
DEPOK
JULI 2009**

HALAMAN PERNYATAAN ORISINALITAS

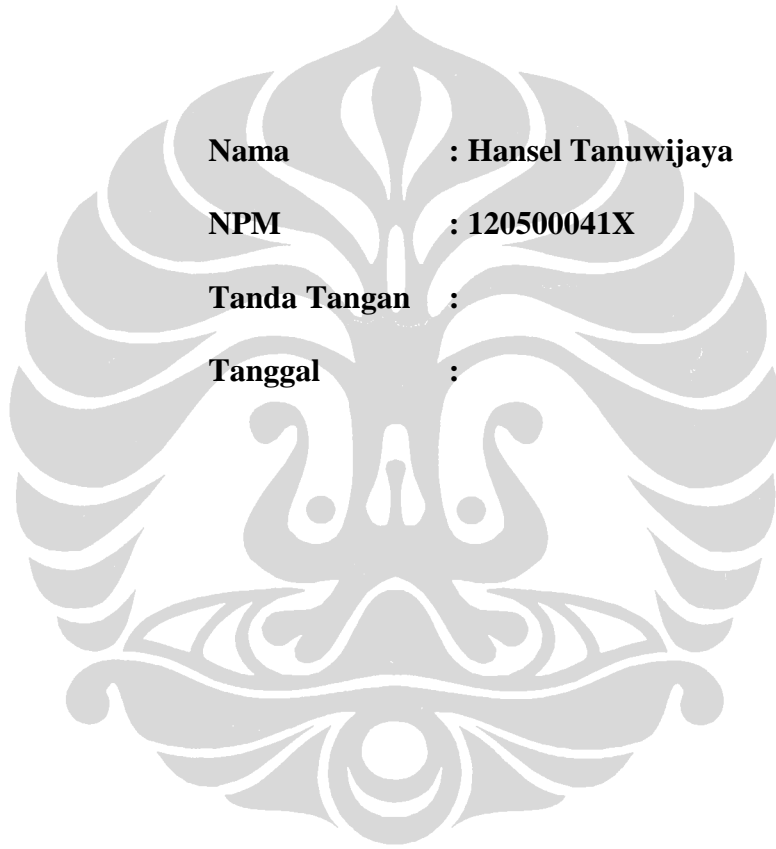
**Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Hansel Tanuwijaya

NPM : 120500041X

Tanda Tangan :

Tanggal :



HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :
Nama : Hansel Tanuwijaya
NPM : 120500041X
Program Studi : S1 Reguler Ilmu Komputer
Judul Skripsi : Penerjemahan dokumen Inggris-Indonesia
menggunakan mesin penerjemah statistik dengan
word reordering dan *phrase reordering*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi S1 Reguler Ilmu Komputer Fakultas Ilmu Komputer Universitas Indonesia

DEWAN PENGUJI

Pembimbing : Dr. Hisar Maruli Manurung (.....)
Penguji : Dr. Achmad Nizar Hidayanto (.....)
Penguji : Dr. Mirna Adriani (.....)

Ditetapkan di :

Tanggal :

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul "Penerjemahan dokumen Inggris-Indonesia menggunakan mesin penerjemah statistik dengan *word reordering* dan *phrase reordering*".

Pada kesempatan ini pula, penulis ingin mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, bantuan, dan semangat kepada penulis sehingga penyusunan laporan tugas akhir ini berjalan dengan baik. Penulis ingin mengucapkan terima kasih kepada:

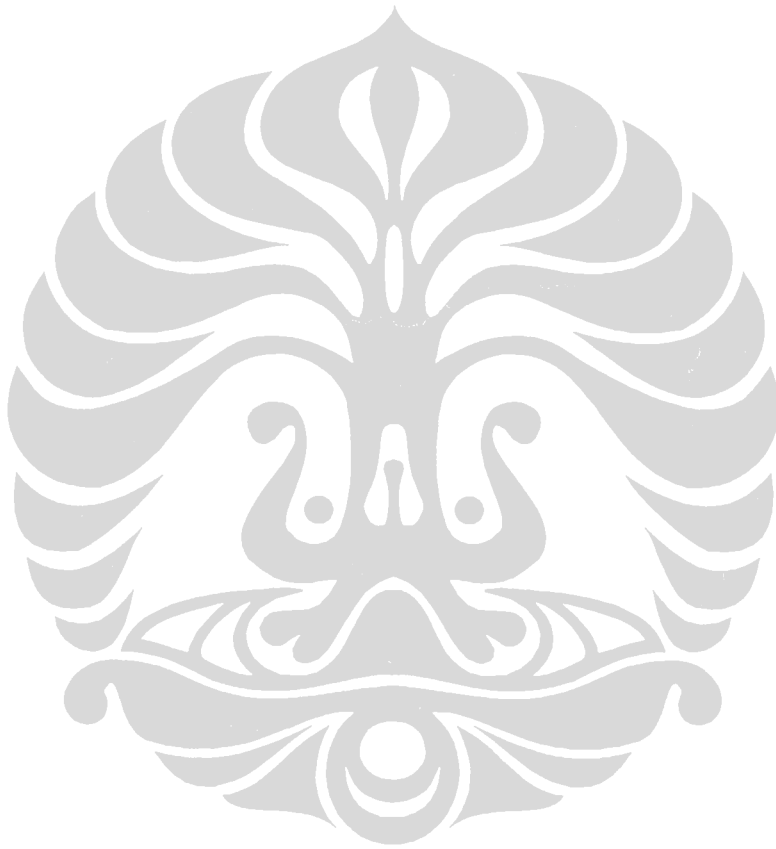
1. Orang tua penulis yang selalu memberi dukungan, semangat, dan motivasi kepada penulis dalam menyelesaikan tugas akhir ini.
2. Bapak Hisar Maruli Manurung, selaku dosen pembimbing tugas akhir penulis, yang memberikan bimbingan, nasihat dan perhatian sepanjang pelaksanaan tugas akhir.
3. Bapak Achmad Nizar Hidayanto dan Ibu Mirna Adriani, selaku dosen penguji tugas akhir penulis.
4. Bapak Rahmat M. Samik-Ibrahim, selaku pembimbing akademis penulis.
5. Amalia, Aurora, Baskoro, Bayu, Berna, Clara, Darwin, Desmond, Metti, Refly, Suryanto, Tedy, Vinky, Yonathan, dan rekan-rekan anggota Lab Perolehan Informasi lainnya.
6. Deniz Yuret, Ergun Bicici, Franky, Mark Przybocki dan Woro Sudaryanti yang telah bersedia menjawab pertanyaan dan memberikan bantuan seputar mesin penerjemah statistik.
7. Semua pihak yang belum disebutkan di atas, yang secara langsung maupun tidak langsung telah memberikan kontribusi dan bantuannya

atas kelancaran dan kesuksesan pelaksanaan tugas akhir ini hingga penyusunan laporan ini.

Saya sebagai penulis laporan ini dengan kerendahan hati memohon saran dan kritik dari berbagai pihak atas penulisan laporan. Dengan begitu penulis menerima setiap masukan yang akan menjadi suatu pelajaran untuk perbaikan ke depan. Akhir kata, penulis berharap semoga laporan tugas akhir ini dapat memberi manfaat kepada para pembaca.

Depok, Juli 2009

Penulis



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Hansel Tanuwijaya
NPM : 120500041X
Program Studi : S1 Reguler Ilmu Komputer
Departemen : 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 Dokumen Inggris-Indonesia Menggunakan Mesin Penerjemah
Statistik dengan *Word Reordering* dan *Phrase Reordering*

beserta perangkat yang ada. 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 saja sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di :

Tanggal :

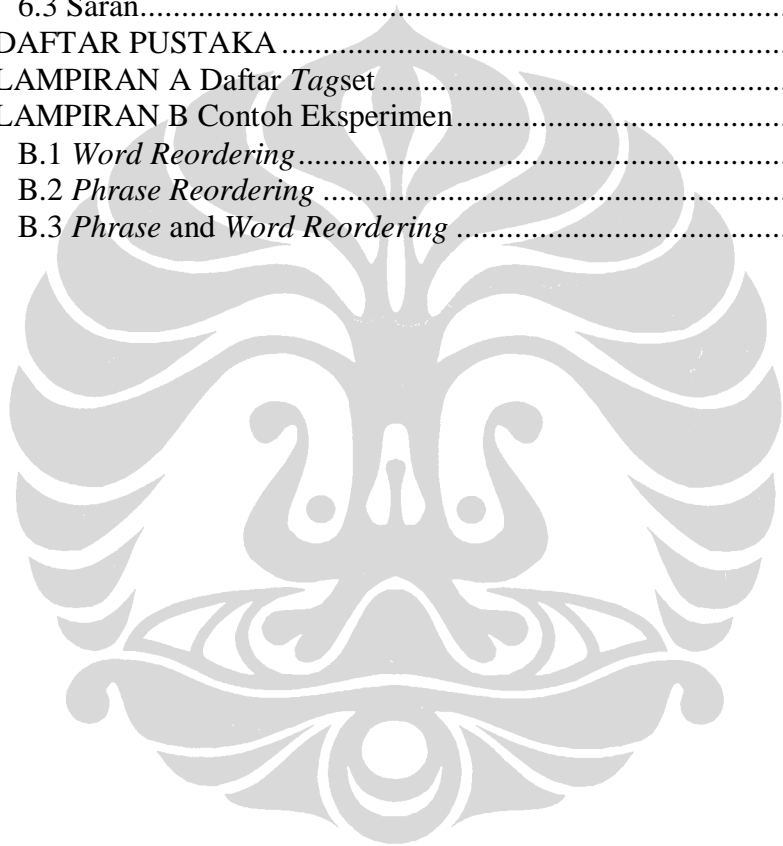
Yang menyatakan

(Hansel Tanuwijaya)

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL | i |
| 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 GAMBAR | xii |
| DAFTAR TABEL | xiv |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Tujuan Penelitian | 2 |
| 1.3 Perumusan dan Pembatasan Masalah | 3 |
| 1.4 Metode dan Tools Pengembangan | 3 |
| 1.5 Sistematika Penulisan | 3 |
| BAB 2 LANDASAN TEORI | 6 |
| 2.1 Sejarah Machine Translation | 6 |
| 2.2 Statistical Machine Translation | 8 |
| 2.2.1 Language Model | 9 |
| 2.2.1 Translation Model | 10 |
| 2.2.3 Decoder | 12 |
| 2.3 Automatic Evaluation | 14 |
| 2.4 <i>Word reordering</i> dan <i>phrase reordering</i> | 21 |
| 2.5 <i>POS Tagging</i> | 23 |
| 2.6 <i>Parsing</i> | 23 |
| BAB 3 ARSITEKTUR DAN PERANCANGAN SISTEM | 26 |
| 3.1 Arsitektur Sistem | 26 |
| 3.1.1 Arsitektur Sistem Normal (Baseline) | 26 |
| 3.1.2 Arsitektur Sistem <i>POS Tagging (Word Reordering)</i> | 27 |
| 3.1.3 Arsitektur Sistem <i>Parsing (Phrase Reordering)</i> | 28 |
| 3.1.4 Arsitektur Sistem <i>Parsing</i> dan <i>POS Tagging (Mixed)</i> | 29 |
| 3.2 Korpus yang Digunakan | 30 |
| 3.3 K-Fold Cross Validation | 31 |
| 3.4 Rancangan <i>Word Reordering</i> | 32 |
| 3.5 Rancangan <i>Phrase Reordering</i> | 35 |
| BAB 4 IMPLEMENTASI | 40 |
| 4.1 Persiapan Awal Dokumen | 40 |
| 4.2 Stanford <i>POS Tagger</i> | 40 |
| 4.3 <i>Word Reordering</i> | 42 |
| 4.4 Cleaning Dokumen | 44 |
| 4.5 Stanford Parser | 45 |
| 4.6 <i>Phrase Reordering</i> | 46 |
| 4.7 MOSES, SRILM, GIZA++ | 50 |
| 4.8 Evaluation Tools | 53 |

| | |
|--|-----|
| BAB 5 HASIL IMPLEMENTASI DAN UJI COBA | 57 |
| 5.1 Hasil dan Analisis Eksperimen Arsitektur Baseline | 57 |
| 5.2 Hasil dan Analisis Eksperimen Arsitektur Sistem POS <i>Tagging</i> | 58 |
| 5.2.1 Analisis Kesalahan Eksperimen Arsitektur Sistem POS <i>Tagging</i> | 65 |
| 5.3 Hasil dan Analisis Eksperimen Arsitektur Sistem <i>Parsing</i> | 68 |
| 5.3.1 Analisis Kesalahan Eksperimen Arsitektur Sistem <i>Parsing</i> | 75 |
| 5.4 Hasil dan Analisis Eksperimen Arsitektur Sistem <i>Mixed</i> | 76 |
| 5.4.1 Analisis Kesalahan Eksperimen Arsitektur Sistem <i>Mixed</i> | 81 |
| 5.5 Rangkuman Analisis Eksperimen | 82 |
| BAB 6 PENUTUP | 88 |
| 6.1 Kesimpulan | 88 |
| 6.2 Keterbatasan dan Hambatan | 89 |
| 6.3 Saran | 90 |
| DAFTAR PUSTAKA | 91 |
| LAMPIRAN A Daftar <i>Tagset</i> | 94 |
| LAMPIRAN B Contoh Eksperimen | 96 |
| B.1 <i>Word Reordering</i> | 96 |
| B.2 <i>Phrase Reordering</i> | 103 |
| B.3 <i>Phrase and Word Reordering</i> | 110 |



DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Komponen Statistical Machine Translation | 8 |
| Gambar 2.2 Ilustrasi Penerjemahan dengan <i>Word-Based Translation Model</i> | 10 |
| Gambar 2.3 Ilustrasi Penerjemahan dengan <i>Phrase-Based Translation Model</i> | 11 |
| Gambar 2.4 Parse Tree..... | 24 |
| Gambar 3.1 Arsitektur Baseline | 26 |
| Gambar 3.2 Arsitektur Sistem POS <i>Tagging</i> | 27 |
| Gambar 3.3 Arsitektur Sistem <i>Parsing</i> | 28 |
| Gambar 3.4 Arsitektur <i>Mixed</i> | 29 |
| Gambar 3.5 Ilustrasi 6-Fold Cross Validation | 31 |
| Gambar 4.1 Pseudocode Persiapan Awal Dokumen | 40 |
| Gambar 4.2 Stanford POS <i>Tagger</i> | 41 |
| Gambar 4.3 Pseudocode <i>Word Reordering</i> | 42 |
| Gambar 4.4 Ilustrasi JJx-NNx pada Kalimat "Rudi kicked red ball." | 44 |
| Gambar 4.5 Pseudocode Cleaning Dokumen..... | 44 |
| Gambar 4.6 Stanford Parser | 45 |
| Gambar 4.7 Pseudocode <i>Phrase Reordering</i> | 47 |
| Gambar 4.8 Parse Tree Kalimat "no matter what the cost." | 48 |
| Gambar 4.9 Ilustrasi dari BFSSearchandCollect <i>Phrase</i> | 48 |
| Gambar 4.10 Parse Tree Kalimat "the cost what no matter." | 49 |
| Gambar 4.11 Script Training pada MOSES..... | 51 |
| Gambar 4.12 Contoh menjalankan testing secara otomatis | 53 |
| Gambar 4.13 Struktur Source File | 54 |
| Gambar 4.14 Pseudocode Mengubah File Menjadi Source File | 54 |
| Gambar 4.15 Struktur Test File | 55 |
| Gambar 4.16 Pseudocode Mengubah File Menjadi Test File | 55 |
| Gambar 4.17 Struktur Reference File | 55 |

| | |
|--|----|
| Gambar 4.18 Pseudocode Mengubah File Menjadi Reference File | 56 |
| Gambar 5.1 Grafik Perbandingan Hasil Eksperimen <i>Word Reordering</i> Pada Korpus Bible | 63 |
| Gambar 5.2 Grafik Perbandingan Hasil Eksperimen <i>Word Reordering</i> Pada Korpus Novel | 63 |
| Gambar 5.3 Contoh Kesalahan <i>Tagging</i> Stanford POS <i>Tagger</i> | 66 |
| Gambar 5.4 Grafik Perbandingan Hasil Eksperimen <i>Phrase Reordering</i> Pada Korpus Bible | 73 |
| Gambar 5.5 Grafik Perbandingan Hasil Eksperimen <i>Phrase Reordering</i> Pada Korpus Novel | 73 |
| Gambar 5.6 Contoh Kesalahan <i>Parsing</i> Stanford POS <i>Tagger</i> | 75 |
| Gambar 5.7 Grafik Perbandingan Hasil Eksperimen <i>Mixed</i> dengan Hasil Eksperimen Lainnya Pada Korpus Bible | 80 |
| Gambar 5.8 Grafik Perbandingan Hasil Eksperimen <i>Mixed</i> dengan Hasil Eksperimen Lainnya Pada Korpus Novel | 80 |



DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Contoh Korpus Paralel | 7 |
| Tabel 2.2 Contoh Hasil Terjemahan dan Rujukan | 16 |
| Tabel 3.1 Contoh JJx-NNx <i>Reordering</i> | 33 |
| Tabel 3.2 Contoh PRP\$-NNx <i>Reordering</i> | 34 |
| Tabel 3.3 Contoh NNx-NNx <i>Reordering</i> | 34 |
| Tabel 3.4 Contoh ADJP-NP <i>Reordering</i> | 36 |
| Tabel 3.5 Contoh NP-NP <i>Reordering</i> | 37 |
| Tabel 3.6 Contoh ADJP-NNx <i>Reordering</i> | 38 |
| Tabel 5.1 Hasil Eksperimen Arsitektur Baseline Korpus Bible | 57 |
| Tabel 5.2 Hasil Eksperimen Arsitektur Baseline Korpus Novel..... | 58 |
| Tabel 5.3 Jumlah <i>Word Reordering</i> pada Korpus Bible dan Novel..... | 58 |
| Tabel 5.4 Hasil Eksperimen Arsitektur Sistem POS <i>Tagging</i> Korpus Bible dengan BLEU | 59 |
| Tabel 5.5 Hasil Eksperimen Arsitektur Sistem POS <i>Tagging</i> Korpus Bible dengan NIST..... | 59 |
| Tabel 5.6. Hasil Eksperimen Arsitektur Sistem POS <i>Tagging</i> Korpus Novel dengan BLEU | 60 |
| Tabel 5.7 Hasil Eksperimen Arsitektur Sistem POS <i>Tagging</i> Korpus Novel dengan NIST | 60 |
| Tabel 5.8 Perbandingan <i>Word Reordering</i> Pada Korpus Bible | 61 |
| Tabel 5.9. Perbandingan <i>Word Reordering</i> Pada Korpus Novel | 62 |
| Tabel 5.10 Rancangan Aturan <i>Word Reordering</i> yang Memberikan Peningkatan65 | |
| Tabel 5.11 Jumlah <i>Phrase Reordering</i> pada Korpus Bible dan Novel..... | 68 |
| Tabel 5.12 Hasil Eksperimen Arsitektur Sistem <i>Parsing</i> Korpus Bible dengan BLEU | 69 |
| Tabel 5.13 Hasil Eksperimen Arsitektur Sistem <i>Parsing</i> Korpus Bible dengan NIST..... | 69 |
| Tabel 5.14 Hasil Eksperimen Arsitektur Sistem <i>Parsing</i> Korpus Novel dengan BLEU | 70 |

| | |
|---|----|
| Tabel 5.15 Hasil Eksperimen Arsitektur Sistem <i>Parsing</i> Korpus Novel dengan NIST..... | 70 |
| Tabel 5.16 Perbandingan <i>Phrase Reordering</i> Pada Korpus Bible | 71 |
| Tabel 5.17 Perbandingan <i>Phrase Reordering</i> Pada Korpus Novel | 72 |
| Tabel 5.18 Jumlah <i>Phrase Reordering</i> dan <i>Word Reordering</i> pada Korpus Bible dan Novel | 77 |
| Tabel 5.19 Hasil Eksperimen Arsitektur Sistem <i>Mixed</i> Korpus Bible | 77 |
| Tabel 5.20 Hasil Eksperimen Arsitektur Sistem <i>Mixed</i> Korpus Novel..... | 77 |
| Tabel 5.21 Perbandingan <i>Mixed</i> Pada Korpus Bible | 78 |
| Tabel 5.22 Perbandingan <i>Mixed</i> Pada Korpus Novel..... | 79 |
| Tabel 5.23 Hasil Eksperimen Nilai BLEU Bahasa Inggris - Jerman | 84 |
| Tabel 5.24 Hasil Eksperimen Nilai BLEU Bahasa Spanyol - Inggris | 84 |
| Tabel 5.25 Hasil Eksperimen Nilai BLEU dan NIST Bahasa Inggris - India..... | 85 |
| Tabel 5.26 Rangkuman Hasil Eksperimen Nilai BLEU dan NIST pada Penelitian | 86 |
| Table A.1 Daftar <i>Tag</i> pada Stanford POS <i>Tagger</i> | 94 |