

Laporan Tugas Akhir

**Perolehan Informasi Lintas Bahasa Inggris-Indonesia  
Berdasarkan Korpus yang Sebanding Menggunakan  
Teknik *Phi Square* dan Koefisien Korelasi Pearson**

**Aprilia  
1204000122**



**Fakultas Ilmu Komputer  
Universitas Indonesia  
Depok  
2008**

## HALAMAN PERSETUJUAN

Judul Tugas Akhir : Perolehan Informasi Lintas Bahasa Inggris-Indonesia  
Berdasarkan Korpus yang Sebanding Menggunakan Teknik  
*Phi Square* dan Koefisien Korelasi Pearson

Nama : Aprilia

NPM : 1204000122

Laporan Tugas Akhir ini telah diperiksa dan disetujui.

Depok, 4 Agustus 2008

Dra. Mirna Adriani PhD.

Pembimbing Tugas Akhir

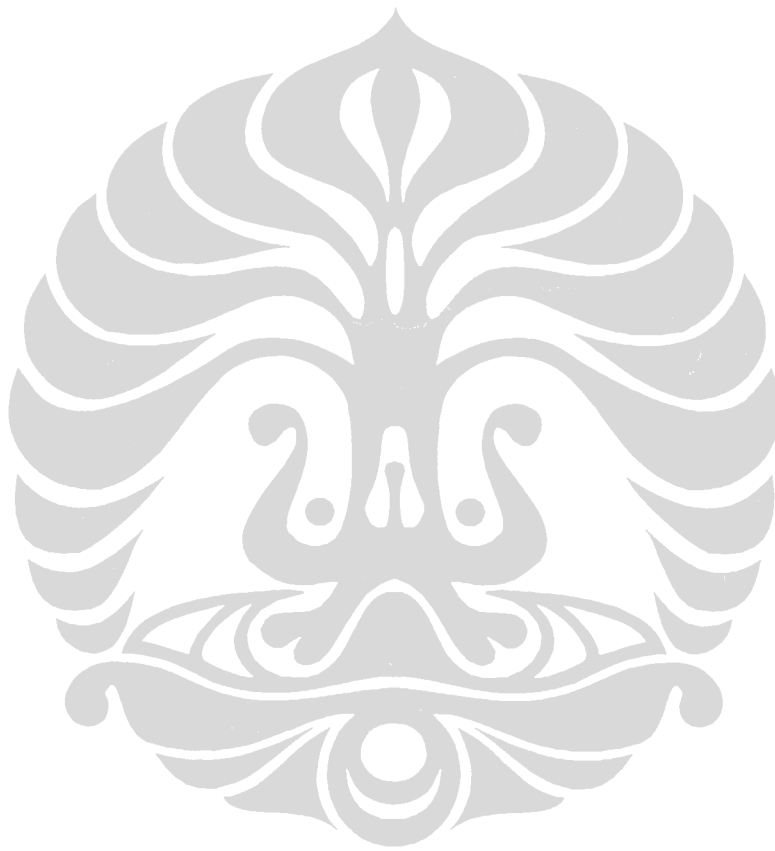
## KATA PENGANTAR

Puji syukur ke hadirat Tuhan yang maha kuasa, atas penyertaan dan kekuatan yang dianugerahkan pada penulis dalam pengerjaan tugas akhir ini. Tugas akhir ini dapat terlaksana dan terselesaikan karena bantuan dari berbagai pihak. Untuk itu, penulis ingin mengucapkan terima kasih atas dukungan berbagai pihak berikut ini:

1. Keluarga besar penulis, terutama untuk mama dan kakak-kakak penulis, yang selalu memberikan dukungan semangat dan doa kepada penulis dalam pelaksanaan tugas akhir ini.
2. Ibu Mirna Adriani, selaku pembimbing penulis dalam pengerjaan tugas akhir ini, yang setia memberi pengarahan, kritik, motivasi, dan semangat kepada penulis dalam pelaksanaan tugas akhir ini.
3. Pak Ade Azurat, selaku pembimbing akademik penulis, atas perhatian, arahan, dan bimbingan yang diberikan pada penulis selama masa perkuliahan di Fasilkom UI.
4. Bapak Hisar Maruli Manurung, sebagai salah satu dosen dan pendengar setia presentasi di Lab Perolehan Informasi, atas tambahan ilmu dan saran-saran yang dibagikan dalam pertemuan-pertemuan di Lab Perolehan Informasi.
5. Yemima dan Evan, yang memberikan semangat dan perhatian dalam pelaksanaan TA penulis.
6. Mba Sandra Sari, yang pernah mengerjakan topik yang sama, atas bantuan yang diberikan pada penulis dalam pengerjaan tugas akhir ini.
7. Teman-teman seperjuangan di Lab Perolehan Informasi, Ame, Koro, Rora, Tuti, Eliza, Desmond, Iki, Joji, Koro, Ananda, Wisnu, Ibrahim, Franky, Arfan, Rahmad, Femphy, dan yang lainnya, atas diskusi-diskusi, dukungan, dan kebersamaan dalam berjuang mengerjakan tugas akhir ini.
8. Semua teman-teman S1 Fasilkom UI angkatan 2004, untuk kebersamaan dan berbagai masa perjuangan yang kita alami bersama selama menjalani dunia perkuliahan di Fasilkom UI.

9. Seluruh dosen, staf, dan karyawan Fakultas Ilmu Komputer Universitas Indonesia. Terima kasih atas segala bantuan dan kemudahan yang diberikan kepada penulis selama menjalani studi di Fasilkom UI.

Penulis berharap tugas akhir ini dapat bermanfaat bagi penelitian selanjutnya. Penulis sadar bahwa dalam tugas akhir ini masih banyak terdapat kesalahan dan kekurangan. Oleh karena itu, kritik dan saran yang membangun akan diterima penulis demi penelitian yang lebih baik di masa mendatang.



Depok, Juli 2008

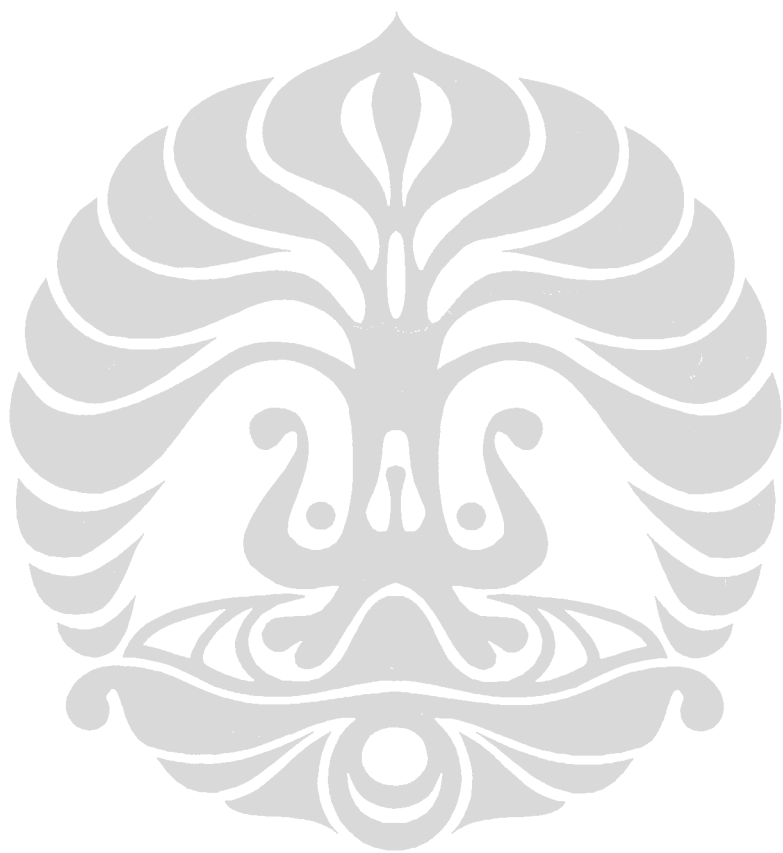
Aprilia

# DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
KATA PENGANTAR .....	iii
ABSTRAK .....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
BAB 1    PENDAHULUAN.....	1
1.1    Latar Belakang.....	1
1.2    Perumusan Masalah.....	3
1.3    Tujuan.....	4
1.4    Ruang Lingkup.....	5
1.5    Metodologi Penelitian.....	5
1.6    Sistematika Penulisan .....	6
BAB 2    LANDASAN TEORI.....	7
2.1    Perolehan Informasi ( <i>Information Retrieval</i> ).....	7
2.1.1    Operasi/Transformasi Teks .....	8
2.1.2    Pembuatan Indeks dan Pencarian Dokumen .....	10
2.1.3    Evaluasi Kinerja Perolehan Teks .....	11
2.2    Perolehan Informasi Lintas Bahasa ( <i>Cross Language Information Retrieval</i> ).....	13
2.2.1    Teknik-teknik yang Digunakan dalam Perolehan Informasi Lintas Bahasa .....	13
2.2.2    Penelitian dalam Perolehan Informasi Lintas Bahasa.....	15

2.2.3	Perolehan Informasi Lintas Bahasa Menggunakan Korpus yang Sebanding .....	19
BAB 3	EKSPERIMEN .....	25
3.1	Pencarian Korpus yang Sebanding .....	25
3.1.1	Pencarian Dokumen .....	25
3.1.2	Mencocokkan Dokumen menjadi Korpus yang Sebanding .....	26
3.2	Operasi Teks .....	32
3.3	Pembuatan Daftar Kata Inggris-Indonesia.....	33
3.3.1	<i>Phi Square</i> .....	33
3.3.2	Koefisien Korelasi Pearson.....	36
3.4	Evaluasi.....	37
3.4.1	Evaluasi Perolehan Dokumen .....	37
3.4.2	Evaluasi Ketepatan Terjemahan Kata .....	40
BAB 4	HASIL EKSPERIMEN DAN ANALISIS.....	41
4.1	Hasil Eksperimen Pembuatan Daftar Kata .....	41
4.1.1	Hasil Eksperimen dengan Menggunakan <i>Phi Square</i> .....	41
4.1.2	Hasil Eksperimen dengan Menggunakan Koefisien Korelasi Pearson .....	42
4.2	Analisis Kinerja Perolehan Dokumen.....	43
4.3	Evaluasi Ketepatan Penerjemahan Kata .....	50
BAB 5	PENUTUP .....	54
5.1	Kesimpulan .....	54
5.2	Saran .....	56
	DAFTAR PUSTAKA .....	57
	LAMPIRAN A  DAFTAR <i>STOPWORDS</i> BAHASA INGGRIS.....	60
	LAMPIRAN B  DAFTAR <i>STOPWORDS</i> BAHASA INDONESIA .....	63

LAMPIRAN C	CONTOH HASIL LUCENE .....	66
LAMPIRAN D	CONTOH HASIL TREC-EVAL.....	67



## DAFTAR TABEL

Tabel 2.1 Contoh petunjuk blok dimana kata berada dalam <i>inverted indeks</i> .....	10
Tabel 2.2 <i>Contingency</i> .....	21
Tabel 3.1 Pasangan dokumen bahasa Indonesia-Inggris di situs web Bank Danamon .....	29
Tabel 3.2 Daftar kata unik di tiap segmen pada sebuah dokumen Inggris .....	34
Tabel 3.3 Daftar kata unik di tiap segmen pada sebuah dokumen Indonesia .....	35
Tabel 3.4 Daftar Frekuensi kata yang sama .....	36
Tabel 3.5 Kueri dalam bahasa Indonesia dan Inggris .....	37
Tabel 3.6 Kueri hasil terjemahan Transtool.....	38
Tabel 3.7 Kueri hasil terjemahan kamus dwi-bahasa.....	39
Tabel 4.1 Daftar kata Inggris-Indonesia dengan penghitungan <i>phi square</i> .....	42
Tabel 4.2 Daftar kata Inggris-Indonesia dengan penghitungan korelasi Pearson .	43
Tabel 4.3 Kueri hasil terjemahan <i>phi square</i> .....	44
Tabel 4.4 Kueri hasil terjemahan dengan koefisien korelasi Pearson.....	44
Tabel 4.5 Istilah-istilah untuk teknik terjemahan kueri .....	45
Tabel 4.6 Perbandingan <i>precision</i> rata-rata antara hasil eksperimen dengan mono .....	46
Tabel 4.7 Perbandingan <i>precision</i> rata-rata antara hasil eksperimen dengan trtool .....	48
Tabel 4.8 Perbandingan <i>precision</i> rata-rata antara hasil eksperimen dengan kamus .....	50
Tabel 4.9 Jumlah kata hasil terjemahan phi yang sama dengan kamus.....	51
Tabel 4.10 Jumlah kata hasil terjemahan pearson yang sama dengan kamus.....	52
Tabel 4.11 Daftar frasa Inggris dan terjemahannya yang tepat .....	53
Tabel 4.12 Jumlah frasa yang diterjemahkan dengan tepat .....	53



## DAFTAR GAMBAR

Gambar 1.1 Data statistik pengguna internet berdasarkan bahasa.....	2
Gambar 2.1 Proses pembagian dokumen menjadi segmen .....	21
Gambar 3.1 Isi direktori ‘en’ dan ‘id’ dari situs web BI.....	27
Gambar 3.2 Isi direktori ‘www.bankmandiri.co.id’ dan ‘english’ .....	28
Gambar 4.1 Grafik perbandingan <i>precision</i> antara hasil eksperimen dengan mono .....	46
Gambar 4.2 Grafik perbandingan <i>precision</i> antara eksperimen dengan trtool .....	48
Gambar 4.3 Grafik perbandingan <i>precision</i> antara eksperimen dengan kamus dwi- bahasa.....	49

