



UNIVERSITAS INDONESIA

**PENGLASIFIKASIAN ARTIKEL BERITA BERBAHASA
INDONESIA SECARA OTOMATIS MENGGUNAKAN
ONTOLOGI**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar S.Kom

Prajna Wira Basnur

1205007244

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

DEPOK

JULI, 2009

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Prajna Wira Basnur
NPM : 1205007244
Program Studi : Ilmu Komputer
Judul Skripsi : Pengklasifikasian Artikel Berita Berbahasa Indonesia
Secara Otomatis Menggunakan Ontologi

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 : Dana I. Sensuse, Ph. D (.....)

Penguji : Heri Kurniawan, M. Kom (.....)

Penguji : Rizal Fathoni Aji, M. Kom (.....)

Ditetapkan di : Fakultas Ilmu Komputer

Tanggal : 24 Juli 2009

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya akhirnya penulis dapat menyelesaikan tugas akhir ini dengan baik. Laporan tugas akhir ini disusun berdasarkan hasil studi penulis dalam mengerjakan tugas akhir. Penulis menyadari bahwa tugas akhir ini tidak dapat terlaksana dengan baik tanpa kerja sama dan dukungan dari berbagai pihak yang telah membimbing, membantu, memberikan saran, kritik dan motivasi. Untuk itu, penulis mengucapkan terima kasih yang tulus kepada:

1. Orang tua, adik, dan seluruh anggota keluarga lainnya yang selalu memberi dukungan, semangat, dan motivasi kepada penulis dalam menyelesaikan tugas akhir ini. gǎn'ēn dàjiā de zhùgào, jīngshén, quàngào, hé zhīchí.
2. Bapak Dana Indra Sensuse selaku dosen pembimbing tugas akhir. Terima kasih atas kesabarannya, bantuannya, dan masukannya selama pengerjaan dan penulisan laporan tugas akhir ini.
3. Bapak Rizal Fathoni Aji, M. Kom, yang telah hadir ketika sidang penulis dan menguji serta memberi masukan dalam perbaikan laporan penulis.
4. Bapak Heri Kurniawan, M. Kom, yang telah hadir ketika sidang penulis dan menguji serta memberi masukan dalam perbaikan laporan penulis.
5. Ibu Kasiyah M Junus selaku dosen pembimbing akademis. Terima kasih atas doa, dukungan, dan motivasi kepada penulis selama empat tahun ini.
6. Adit, Alex, Ananta, Armando, Bambang, Bayu, Bernadia, Clara, Darwin, Evi, Hansel, Haryadi, Kukuh, Mursal, Nur, Purwanto, Ranu, Refly, Risnal, Rizal, Teddy, Tiara, Vinky, dan Yenni. Gǎnxiè dàjiā de zhīchí hé bāngzhù.
7. Arigayo, Hendra, Hendro, Isna, Jessi, Maria, Nudia, Pretti, Suci, Syamsiah, Yessi, dan Yuriandi. Gǎnxiè dàjiā de zhùdǎo hé zhīchí.
8. Rozi gēge. gǎn'ēn nǐ de shǒutí jìsuànjī jiè, zhùdǎo, zhīchí, quàngào, hé bāngzhù, dànsì wǒ bù tīng nǐ de quàngào.

9. Teman-teman Kelas Mandarin 2 LBI Fakultas Ilmu Budaya Universitas Indonesia. xièxie wǒ de hěn rènào péngyǒu.
10. Teman-teman ekstrakurikuler Bulutangkis. xièxie wǒ de yǔmáoqiú wán péngyǒu, wǒ hé nǐmen hěn xǐhuan yīqǐ wán yǔmáoqiú.
11. Teman-teman Fakultas Ilmu Komputer Universitas Indonesia, khususnya angkatan 2005.
12. Semua staf pengajar, administrasi, dan keluarga besar Fakultas Ilmu Komputer Universitas Indonesia.
13. Bapak dan Ibu kos. Terima kasih atas dukungan, motivasi, dan doanya.
14. Pemerintah Kota Dumai dan Dinas Pendidikan Kota Dumai. Terima kasih atas dukungan dan bantuannya selama empat tahun ini.
15. Seluruh pihak yang baik secara langsung dan tidak langsung membantu dalam pengerjaan penelitian dan laporan tugas akhir ini.

Penulis menyadari bahwa dalam melakukan penelitian ini masih mempunyai banyak kekurangan dan keterbatasan. Oleh karena itu, dengan segala kerendahan hati penulis bersedia menerima setiap saran dan masukan yang membangun untuk laporan tugas akhir ini. Semoga penyusunan laporan ini dapat memberikan manfaat bagi para pembaca semua.

Depok, Juli 2009

Prajna Wira Basnur

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

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

Nama : Prajna Wira Basnur
NPM : 1205007244
Program Studi : Ilmu Komputer
Fakultas : Ilmu Komputer
Jenis karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :
Pengklasifikasian Artikel Berita Berbahasa Indonesia Secara Otomatis Menggunakan Ontologi

berserta 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 : 24 Juli 2009
Yang menyatakan

(Prajna Wira Basnur)

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR PERSAMAAN.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan.....	2
1.3 Tujuan	2
1.4 Ruang Lingkup.....	3
1.5 Metodologi Penelitian.....	3
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	5
2.1 Perolehan Informasi	5
2.2 Klasifikasi Dokumen	6
2.3 Naïve Bayes.....	8
2.4 Knowledge Engineering	11
2.5 Ontologi	11
2.6 Persiapan Dokumen Teks	15
2.6.1 Case Folding	15

2.6.2 Tokenisasi	15
2.6.3 Pembuangan Stopwords	15
2.6.4 Pemotongan Imbuhan	16
2.7 Pembobotan Kata	16
2.7.1 Binary Weighting	17
2.7.2 Term Frequency	17
2.7.3 Term Frequency – Inverse Document Frequency	17
2.8 Evaluasi	18
2.9 Penelitian dalam Bidang Klasifikasi Dokumen Teks	20
BAB 3 PERANCANGAN	23
3.1 Data	23
3.2 Persiapan Dokumen Teks	23
3.2.1 Case Folding	24
3.2.2 Tokenisasi	25
3.2.3 Pembuangan Stopwords	26
3.2.4 Pemotongan Imbuhan	27
3.3 Pembobotan Kata	27
3.4 Pemodelan Ontologi	31
3.5 Metode Klasifikasi Dokumen Teks	32
3.5.1 Naïve Bayes	33
3.5.2 Ontologi	35
BAB 4 IMPLEMENTASI	37
4.1 Persiapan Dokumen	37
4.1.4 Case Folding	37
4.1.4 Tokenisasi	38
4.1.3 Pembuangan Stopwords	39
4.1.4 Pemotongan Imbuhan	40
4.2 Pembobotan Kata	41
4.3 Pemodelan Ontologi	43
4.4 Klasifikasi Dokumen Teks	49

4.4.1 Naïve Bayes	49
4.4.2 Ontologi.....	54
BAB 5 HASIL DAN PEMBAHASAN	60
5.1 Hasil Klasifikasi Dokumen.....	60
5.2 Hasil Klasifikasi dari Aspek Penggunaan Stopwords dan Stemming	58
5.3 Hasil Klasifikasi dari Aspek Penggunaan Stemming	60
5.4 Hasil Klasifikasi dari Aspek Penggunaan Stopwords.....	62
5.5 Hasil Klasifikasi dari Aspek tanpa Penggunaan Stopwords dan Stemming	65
5.6 Hasil Klasifikasi dari Aspek Jumlah Kategori	67
5.7 Rangkuman Hasil Eksperimen.....	69
BAB 6 KESIMPULAN DAN SARAN	71
6.1 Kesimpulan.....	71
6.2 Kendala.....	71
6.3 Saran	72
DAFTAR PUSTAKA	73

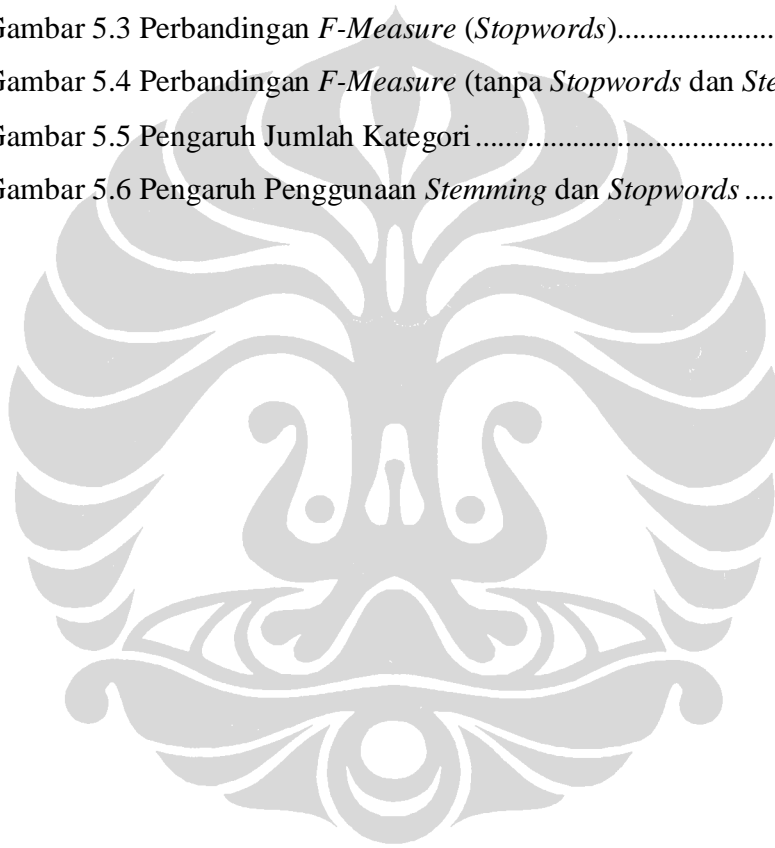
DAFTAR TABEL

Tabel 2.1 Matriks Keputusan untuk Menghitung Akurasi Klasifikasi Dokumen Teks	18
Tabel 3.1 Daftar Kategori dan Jumlah Dokumen Teks	23
Tabel 3.2 <i>Term Documents Matrix</i> ($Dok_j \times TF_{jk}$).....	28
Tabel 3.3 <i>Term Documents Matrix</i> untuk Dokumen 1 dan Dokumen 2.....	29
Tabel 3.4 DF_k dan IDF_k untuk Dokumen 1 dan Dokumen 2.....	29
Tabel 3.5 Nilai $TF-IDF_{1k}$	30
Tabel 3.6 Nilai $TF-IDF_{2k}$	30
Tabel 5.1 Variabel Eksperimen	57
Tabel 5.2 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stopwords</i> dan <i>Stemming</i> (Ontologi).....	59
Tabel 5.3 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stopwords</i> dan <i>Stemming</i> (Naïve Bayes).....	59
Tabel 5.4 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stemming</i> (Ontologi)	61
Tabel 5.5 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stemming</i> (Naïve Bayes).....	61
Tabel 5.6 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stopwords</i> (Ontologi).....	63
Tabel 5.7 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> menggunakan <i>Stopwords</i> (Naïve Bayes).....	64
Tabel 5.8 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> tanpa menggunakan <i>Stopwords</i> dan <i>Stemming</i> (Ontologi).....	66
Tabel 5.9 Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-Measure</i> tanpa menggunakan <i>Stopwords</i> dan <i>Stemming</i> (Naïve Bayes).....	66
Tabel 5.10 Ringkasan Nilai <i>F-Measure</i>	69

DAFTAR GAMBAR

Gambar 3.1 Contoh Dokumen Masukan 1	24
Gambar 3.2 Contoh Dokumen Masukan 2	24
Gambar 3.3 Tahap <i>Case Folding</i> Dokumen 1	24
Gambar 3.4 Tahap <i>Case Folding</i> Dokumen 2	24
Gambar 3.5 Tahap Tokenisasi Dokumen 1	25
Gambar 3.6 Tahap Tokenisasi Dokumen 2	26
Gambar 3.7 Tahap Pembuangan <i>Stopwords</i> Dokumen 1	26
Gambar 3.8 Tahap Pembuangan <i>Stopwords</i> Dokumen 2	26
Gambar 3.9 Tahap Pembuangan Imbuhan Dokumen 1	27
Gambar 3.10 Tahap Pembuangan Imbuhan Dokumen 2	27
Gambar 3.11 Perancangan Klasifikasi Dokumen Teks (Naïve Bayes)	35
Gambar 3.12 Perancangan Klasifikasi Dokumen Teks (Ontologi)	36
Gambar 4.1 <i>Pseudocode</i> Proses <i>Case Folding</i>	38
Gambar 4.2 <i>Pseudocode</i> Proses Tokenisasi	39
Gambar 4.3 <i>Pseudocode</i> untuk Proses Pembuangan <i>Stopwords</i>	40
Gambar 4.4 <i>Pseudocode</i> Proses Pemotongan Imbuhan	41
Gambar 4.5 <i>Pseudocode</i> Pembobotan Kata	43
Gambar 4.6 Representasi Ontologi Olahraga	43
Gambar 4.7 Representasi Subkelas Masing-masing Kategori	44
Gambar 4.8 Contoh Representasi <i>Instances</i> pada Kelas “Bulutangkis”	45
Gambar 4.9 Contoh Representasi <i>Instances</i> pada Kelas “Basket”	46
Gambar 4.10 Contoh Representasi <i>Instances</i> pada Kelas “Otomotif”	47
Gambar 4.11 Contoh Representasi <i>Instances</i> pada Kelas “Sepakbola”	47
Gambar 4.12 Contoh Representasi <i>Instances</i> pada Kelas “Tenis”	48
Gambar 4.13 Contoh Tampilan PROTEGE Versi 3.3.1	49

Gambar 4.14 Format Berkas ARFF	50
Gambar 4.15 <i>Pseudocode</i> Klasifikasi Dokumen Teks menggunakan Naïve Bayes...	51
Gambar 4.16 Hasil keluaran klasifikasi dokumen teks menggunakan Naïve Bayes ..	54
Gambar 4.17 <i>Pseudocode</i> Klasifikasi Dokumen menggunakan Ontologi	56
Gambar 4.18 Hasil keluaran klasifikasi dokumen teks menggunakan Ontologi	59
Gambar 5.1 Perbandingan <i>F-Measure</i> (<i>Stopwords</i> dan <i>Stemming</i>).....	60
Gambar 5.2 Perbandingan <i>F-Measure</i> (<i>Stemming</i>).....	62
Gambar 5.3 Perbandingan <i>F-Measure</i> (<i>Stopwords</i>).....	65
Gambar 5.4 Perbandingan <i>F-Measure</i> (tanpa <i>Stopwords</i> dan <i>Stemming</i>).....	67
Gambar 5.5 Pengaruh Jumlah Kategori	68
Gambar 5.6 Pengaruh Penggunaan <i>Stemming</i> dan <i>Stopwords</i>	70



DAFTAR PERSAMAAN

Persamaan (2.1) Probabilitas kategori c jika diketahui dokumen d	9
Persamaan (2.2) Probabilitas kategori c irisan dokumen d.....	9
Persamaan (2.3) Probabilitas kategori c jika diketahui dokumen d	9
Persamaan (2.4) Probabilitas kategori c jika diketahui dokumen d dengan perhitungan hasil perkalian dari probabilitas kemunculan semua kata pada dokumen d	9
Persamaan (2.5) probabilitas kata w jika diketahui kategori c dengan <i>laplacian smoothing</i>	10
Persamaan (2.6) probabilitas kategori c.....	10
Persamaan (2.7) Penentuan kategori hasil klasifikasi.....	10
Persamaan (2.8) Rumus Binary Weighting	17
Persamaan (2.9) Rumus Term Frequency.....	17
Persamaan (2.10) <i>Rumus Term Frequency – Inverse Document Frequency</i>	18
Persamaan (2.11) Rumus <i>Recall</i>	19
Persamaan (2.12) Rumus <i>Precision</i>	19
Persamaan (2.13) Rumus <i>F-Measure</i>	19
Persamaan (3.1) Rumus Penghitungan Bobot kata TF-IDF _{jk} (x _{jk})	28
Persamaan (3.2) Rumus IDF _k	28
Persamaan (3.3) Rumus Menghitung Nilai Kemiripan antara sebuah <i>node</i> dan dokumen.....	35

DAFTAR LAMPIRAN

LAMPIRAN 1 DAFTAR STOPWORDS	76
LAMPIRAN 2 CONTOH ARTIKEL BERITA.....	80
LAMPIRAN 3 ONTOLOGI DALAM FORMAT RDF/OWL	85

