

**PENERAPAN STRUKTUR TAKSONOMI WIKIPEDIA
PADA METODE K-NEAREST NEIGHBOR
DALAM PENGKLASIFIKASIAN ARTIKEL BLOG BERBAHASA INGGRIS**

Tugas Akhir



Oleh :
Maria Evarisma Wulandari
22074207

**Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana Yogyakarta
Tahun 2011**

**PENERAPAN STRUKTUR TAKSONOMI WIKIPEDIA
PADA METODE K-NEAREST NEIGHBOR
DALAM PENGKLASIFIKASIAN ARTIKEL BLOG BERBAHASA INGGRIS**

Tugas Akhir



**Diajukan kepada Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana sebagai salah satu syarat
dalam memperoleh gelar Sarjana Komputer**



Oleh :

Maria Evarisma Wulandari

22074207

**Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana Yogyakarta
Tahun 2011**

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul:
PENERAPAN STRUKTUR TAKSONOMI WIKIPEDIA PADA METODE K-NEAREST
NEIGHBOR DALAM PENGKLASIFIKASIAN ARTIKEL BLOG BERBAHASA INGGRIS
Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada
pendidikan sarjana Program Studi Teknik Informatika, Fakultas Teknologi Informasi Universitas
Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaaan di
lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi
manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika di kemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi
lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaaan saya.



Yogyakarta, 27 Mei 2011

(Maria Evarisma Wulandari)

22074207

INTISARI

PENERAPAN STRUKTUR TAKSONOMI WIKIPEDIA PADA METODE K-NEAREST NEIGHBOR DALAM PENGKLASIFIKASIAN ARTIKEL BLOG BERBAHASA INGGRIS

Saat ini banyak sistem klasifikasi yang diciptakan dalam rangka membantu pengguna dalam melakukan pengklasifikasian dokumen dengan menggunakan berbagai metode klasifikasi yang ada. Namun sebagian besar sistem klasifikasi tersebut hanya berbasis pada statistik semata dan sebagian besar BOW (*Bag of Words*) yang terbentuk hanyalah berdasarkan pada satuan kata yang sering tidak dapat mewakili dokumen.

Penulis membangun sebuah sistem klasifikasi dengan menggunakan metode k-Nearest Neighbor (KNN) dan memanfaatkan struktur taksonomi Wikipedia yang menekankan pada hubungan semantik dari kata yang ada di dalam dokumen. Tujuan penelitian ini adalah untuk melihat seberapa jauh peran taksonomi Wikipedia dalam meningkatkan nilai akurasi sistem. Studi kasus yang digunakan oleh penulis adalah artikel *blog* berbahasa inggris yang berupa dokumen RSS dan bersumber pada www.wordpress.com. Sistem melakukan tokenisasi berdasarkan pada frase yang didapatkan dari kumpulan topik yang terdaftar pada Wikipedia, sehingga pemecahan kata pada artikel uji berdasarkan pada kumpulan frase ini. Cara tokenisasi seperti ini membantu meningkatkan akurasi sistem dalam menentukan artikel dengan tingkat kesamaan dengan artikel yang diuji.

Sistem klasifikasi yang dibangun oleh penulis dapat menghasilkan tingkat akurasi yang tinggi dengan menggunakan nilai $k = 6$ dan seleksi fitur 50%, yaitu sebesar 98.83 %. Pemanfaatan struktur Wikipedia memberi manfaat dalam meningkatkan nilai akurasi klasifikasi artikel.

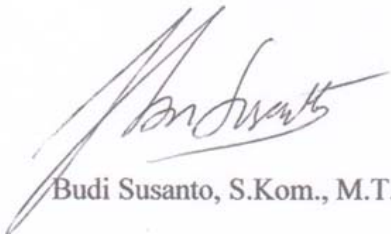
HALAMAN PERSETUJUAN

Judul : Penerapan Struktur Taksonomi Wikipedia pada Metode k-Nearest Neighbor dalam Pengklasifikasian Artikel Blog Berbahasa Inggris
Nama : Maria Evarisma Wulandari
NIM : 22074207
Mata Kuliah : Tugas Akhir
Kode : TI2126
Semester : Genap
Tahun akademik : 2010/2011


Telah diperiksa dan disetujui
Di Yogyakarta,
Pada tanggal 25 Mei 2011

Dosen Pembimbing I

Dosen Pembimbing II



Budi Susanto, S.Kom., M.T.



Willy Sudiarto Raharjo, S.Kom, M.Cs

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN STRUKTUR TAKSONOMI WIKIPEDIA

PADA METODE K-NEAREST NEIGHBOR

DALAM PENGKLASIFIKASIAN ARTIKEL BLOG BERBAHASA INGGRIS

Oleh Maria Evarisma Wulandari / 22074207

Dipertahankan di depan dewan Penguji Tugas Akhir/Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana – Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu

Syarat memperoleh gelar

Sarjana komputer

Pada tanggal

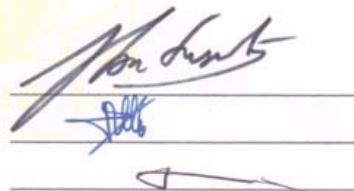
10 Juni 2011

Yogyakarta, 17 Juni 2011

Mengesahkan,

Dewan Penguji

1. Budi Susanto, S.Kom., M.T.
2. Willy Sudiarto Raharjo, S.Kom., M.Cs.
3. Ir. Sri Suwarno, M.Eng.

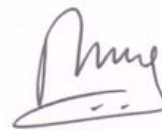


Dekan



(Drs. Wimmie Handiwidjojo, MIT)

Ketua Program Studi



(Nugroho Agus Haryono, S.Si., M.Si.)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang senantiasa melindungi dan melimpahkan rahmat serta pertolongan-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik.

Dalam proses pembuatan Tugas Akhir ini, penulis banyak mendapat bantuan dan dukungan dari berbagai pihak berupa saran, bimbingan dan doa. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom, M.T., selaku dosen pembimbing I yang telah memberikan inspirasi kepada penulis mengenai penelitian ini, serta telah meluangkan waktu untuk membimbing penulis dalam pembuatan Tugas Akhir ini.
2. Bapak Willy Sudiarto Raharjo, S.Kom, M.Cs., selaku dosen pembimbing II yang telah memberi bimbingan, saran, dan petunjuk, serta telah memberi solusi untuk masalah yang muncul selama proses penyelesaian Tugas Akhir ini.
3. Bapak dan Mama yang senantiasa mempertanyakan keadaan Tugas Akhir ini dan senantiasa memberikan doa serta dukungan hingga Tugas Akhir ini selesai.
4. Eyang tersayang yang telah dengan sabar menemani dan mendukung penulis dalam menyelesaikan Tugas Akhir ini.
5. Adik tercinta, Cecilia Desvita Ratna Elvandari dan Kanaya Ragil. Terima kasih untuk interferensi yang sudah diberikan selama ini, terkadang memang sangat membantu. Dan untuk Ibu Siti, terima kasih banyak atas dukungannya.
6. Felix Dwi Jaya, terima kasih atas dukungan, bantuan, dan waktu yang telah diluangkan untuk berdiskusi dan menemani dalam *refreshing* di sela-sela pengerjaan Tugas Akhir ini.
7. Fanny, Danny, Yoming, dan Peter yang telah mewarnai hidup perkuliahan dengan aneka derita dan hiburan.
8. Rachel dan Nandra, secara langsung maupun tidak langsung, kalian mengurangi beban pikiran selama pengerjaan Tugas Akhir ini. *Thanks, Girls!*
9. Vista, Ario, Yesi. Terima kasih.

10. Mas Roni dan segenap staf administrasi fakultas Teknologi Informasi yang selalu siap melayani pertanyaan dan permintaan bantuan dari penulis.
11. Teman-teman seperjuangan yang mungkin masih berjuang dalam menyelesaikan kewajiban perkuliahan di kampus UKDW.
12. Pihak-pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulis hingga Tugas Akhir ini dapat diselesaikan

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih kurang dari sempurna. Oleh karena itu, penulis meminta maaf apabila terdapat kesalahan baik dalam penyusunan laporan maupun kesalahan lain yang telah penulis lakukan dalam penyelesaian Tugas Akhir ini.

Akhir kata, penulis berharap semoga laporan Tugas Akhir ini dapat memberikan tambahan pengetahuan bagi pembaca dan semoga penelitian ini dapat berguna bagi kita semua.

Yogyakarta, Mei 2011

Penulis

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN TUGAS AKHIR	iii
INTISARI	iv
HALAMAN PERSETUJUAN	v
HALAMAN PENGESAHAN	vi
UCAPAN TERIMA KASIH	vii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
Bab 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	7
1.3 Batasan Masalah.....	7
1.4 Hipotesis.....	3
1.5 Tujuan Penelitian.....	3
1.6 Metode Penelitian.....	3
1.7 Sistematika Penulisan.....	4
Bab 2 TINJAUAN PUSTAKA.....	6
2.1. Tinjauan Pustaka	6
2.2. Landasan Teori	8
2.2.1. Web Blog	8
2.2.2. Wordpress	8
2.2.3. <i>RDF (Resource Description Framework)</i>	9
2.2.4. <i>RSS (Really Simple Syndication)</i>	11
2.2.5. Wikipedia dan DBpedia.....	12

2.2.6. Ontologi dan Taksonomi	13
2.2.7. OWL	14
2.2.8. Sistem Berbasis Pengetahuan	15
2.2.9. Penggalian Data (<i>Data Mining</i>).....	16
2.2.10. <i>Text Mining</i>	16
2.2.11. Klasifikasi Dokumen	18
2.2.12. Metode K-Nearest Neighbor (KNN)	18
2.2.13. Akurasi / Presisi	20
Bab 3 ANALISIS DAN PERANCANGAN SISTEM.....	21
3.1. Kebutuhan Hardware dan Software	21
3.1.1. Kebutuhan <i>Hardware</i>	21
3.1.2. Kebutuhan <i>Software</i>	21
3.2. Diagram <i>Use Case</i>	22
3.3. Konfigurasi dan Spesifikasi Sistem	24
3.3.1. Konfigurasi Sistem	24
3.3.2. Spesifikasi Sistem	25
3.4. Arsitektur Sistem.....	26
3.5. Algoritma dan <i>Flowchart</i>	27
3.5.1 <i>Parsing File</i> Basis Data Wikipedia	27
3.5.2 Alur Kerja Sistem Secara Umum	28
3.5.3. Pra-pemrosesan Teks.....	30
3.5.4. Hapus Artikel Uji.....	32
3.5.5. Klasifikasi Artikel Uji	32
3.6. Kamus Data	24
3.6.1. Tabel Taxonomy.....	35
3.6.2. Tabel Category_Label	35
3.6.3. Tabel Wiki_Topic.....	36
3.6.4. Tabel Category_Topic	36
3.6.5. Tabel Wiki_Token.....	37

3.6.6. Tabel Vector_Token.....	38
3.6.7. Tabel Vector_Doc.....	38
3.6.8. Tabel Token_Topic.....	39
3.6.9. Tabel Testing_Article.....	39
3.6.10. Tabel Testing_Token.....	40
3.6.11. Tabel Stopword.....	41
3.7. Diagram Skema.....	41
3.7.1 Diagram Skema Penyimpanan Tabel Berbasis Dbpedia.....	41
3.7.2. Diagram Skema Penyimpanan Tabel Artikel Uji.....	42
3.8. <i>Site Plan</i> dan Rancangan Antarmuka Sistem.....	43
3.8.1. Site Plan.....	43
3.8.2. Rancangan Antarmuka Sistem.....	44
3.8.2.1. Rancangan Antarmuka Halaman Awal.....	44
3.8.2.2. Rancangan Antarmuka Halaman <i>Default (Home)</i>	45
3.8.2.3. Rancangan Antarmuka Klasifikasi Artikel.....	46
3.9. Rancangan Evaluasi Sistem Klasifikasi.....	47
Bab 4 IMPLEMENTASI DAN ANALISIS SISTEM.....	48
4.1 Antarmuka Sistem.....	48
4.1.1 <i>Form Login</i>	48
4.1.2 Halaman Persiapan Basis Data.....	49
4.1.3 Halaman Daftar Artikel.....	49
4.1.4 Halaman Detil Artikel.....	51
4.1.5 <i>Form</i> Tambah Artikel Pelatihan.....	52
4.1.6 <i>Form</i> Uji Artikel Pelatihan.....	53
4.1.7 Halaman Klasifikasi.....	54
4.1.8 Fitur <i>Logout</i>	56
4.2 Implementasi Sistem.....	56
4.2.1 Pengumpulan Dokumen/Artikel.....	56
4.2.2 Pra-pemrosesan Teks.....	57

4.2.3 Klasifikasi Dokumen	59
4.3 Evaluasi Sistem	61
4.3.1 Evaluasi Akurasi Sistem	61
4.3.1.1 Pengujian pada Nilai $k = 9$	62
4.3.1.2 Pengujian pada Nilai $k = 6$	64
4.3.1.3 Pengujian pada Nilai $k = 3$	65
4.3.1 Evaluasi Urutan Token pada Artikel	68
Bab 5 KESIMPULAN DAN SARAN	70
5.1 Kesimpulan	70
5.2 Saran	70

© UKDW

DAFTAR TABEL

Tabel 3.1 Tabel Taxonomy	35
Tabel 3.2 Tabel Category_Label	35
Tabel 3.3 Tabel Wiki_Topic	36
Tabel 3.4 Tabel Category_Topic	37
Tabel 3.5 Tabel Wiki_Token	37
Tabel 3.6 Tabel Vector_Token	38
Tabel 3.7a Tabel Vector_Doc	38
Tabel 3.7b Lanjutan Tabel Vector_Doc	39
Tabel 3.8 Tabel Token Topic	39
Tabel 3.9 Tabel Testing_Article	40
Tabel 3.10 Tabel Testing_Token	40
Tabel 3.11 Tabel Stopword	41
Tabel 4.1 Tabel Hasil Artikel Teruji Benar pada Pengujian Artikel dengan $k = 9$	62
Tabel 4.2 Tabel Hasil Artikel Teruji Benar pada Pengujian Artikel dengan $k = 6$	64
Tabel 4.3 Tabel Hasil Artikel Teruji Benar pada Pengujian Artikel dengan $k = 3$	65
Tabel 4.4 Tabel Hasil Pengujian Artikel pada Urutan Token Berbeda	69



DAFTAR GAMBAR

Gambar 2.1	Contoh Graf Struktur Identifikasi dengan <i>RDF</i>	10
Gambar 2.2	Contoh struktur <i>RDF</i> dalam Bahasa <i>XML</i>	11
Gambar 2.3	Contoh Kelas milik DBpedia beserta Jumlah <i>Instance</i> dan Contoh Properti.....	14
Gambar 2.4	Contoh Struktur OWL Milik DBpedia.....	15
Gambar 2.5	Gambaran Pengklasifikasian Dokumen Menggunakan Metode KNN.....	19
Gambar 3.1	Diagram <i>Use Case</i>	23
Gambar 3.2	Arsitektur Sistem.....	27
Gambar 3.3	<i>Flowchart Parsing File</i> Basis Data Wikipedia	29
Gambar 3.4	Flowchart Alur Kerja Sistem Secara Umum.....	29
Gambar 3.5	<i>Flowchart</i> Pra-pemrosesan Teks	31
Gambar 3.6	<i>Flowchart</i> Hapus Artikel Uji.....	32
Gambar 3.7a	<i>Flowchart</i> Klasifikasi Artikel Uji Bagian 1	33
Gambar 3.7b	<i>Flowchart</i> Klasifikasi Artikel Uji Bagian 2	34
Gambar 3.8	Diagram Skema Penyimpanan Tabel Berbasis Dbpedia.....	42
Gambar 3.9	Diagram Skema Penyimpanan Data Artikel Uji	43
Gambar 3.10	<i>Site Plan</i> Sistem	44
Gambar 3.11	Rancangan Antarmuka Halaman Awal	45
Gambar 3.12	Rancangan Antarmuka Halaman <i>Home</i>	46
Gambar 3.13	Rancangan Antarmuka Klasifikasi Artikel.....	47
Gambar 4.1	Form Login.....	48
Gambar 4.2	Halaman Persiapan Basis Data.....	49
Gambar 4.3	Halaman Daftar Artikel (Pengguna).....	50
Gambar 4.4	Halaman Daftar Artikel (Admin)	51
Gambar 4.5	Halaman Detil Artikel	52
Gambar 4.6	<i>Form</i> Tambah Artikel Pelatihan.....	53
Gambar 4.7	<i>Form</i> Uji Artikel Pelatihan.....	54
Gambar 4.8	<i>Form</i> Klasifikasi.....	55

Gambar 4.9	Halaman Hasil Klasifikasi.....	56
Gambar 4.10	<i>Pseudocode</i> Pra-pemrosesan Teks hingga Tahap Stemming.....	58
Gambar 4.11	<i>Pseudocode</i> Perhitungan Bobot Token	59
Gambar 4.12	<i>Pseudocode</i> Seleksi Fitur	59
Gambar 4.13	<i>Pseudocode</i> Pembentukan Vektor Dokumen	60
Gambar 4.14	<i>Pseudocode</i> Klasifikasi Artikel Uji.....	61
Gambar 4.15	Grafik Kecenderungan Nilai Akurasi Sistem Klasifikasi pada $k = 9$	63
Gambar 4.16	Grafik Kecenderungan Nilai Akurasi Sistem Klasifikasi pada $k = 6$	65
Gambar 4.17	Grafik Kecenderungan Nilai Akurasi Sistem Klasifikasi pada $k = 3$	66
Gambar 4.18	Grafik Perbandingan Nilai Akurasi Klasifikasi Sistem pada Berbagai Variasi Nilai k dan Nilai Seleksi Fitur.....	68

© UKDW

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Saat ini sudah banyak sistem klasifikasi yang diciptakan dalam rangka membantu pengguna dalam melakukan pengklasifikasian dokumen, baik dokumen yang berbentuk *plain text*, halaman berekstensi .pdf, halaman *web*, maupun dalam bentuk yang lain dengan menggunakan berbagai metode klasifikasi yang ada. Namun sebagian besar sistem klasifikasi tersebut berbasis hanya pada statistik semata dan sebagian besar BOW (*Bag of Words*) yang terbentuk hanyalah berdasarkan pada satuan kata yang sering kali tidak dapat mewakili dokumen.

Melihat fenomena seperti yang tertulis di atas, penulis memutuskan untuk membangun sebuah sistem klasifikasi dengan memanfaatkan struktur taksonomi Wikipedia yang menekankan pada hubungan semi semantik dari kata yang ada di dalam dokumen. Taksonomi Wikipedia ini telah banyak digunakan untuk berbagai aplikasi, antara lain membantu melakukan improvisasi pada pencarian di Wikipedia, implementasi pada aplikasi bergerak dan aplikasi geografis (*mobile and geographic application*), klasifikasi dokumen dan *bookmarking (document classification and annotation bookmarking)*, dan aplikasi lainnya.^[1]

Dalam penelitian ini, penulis akan membangun sebuah sistem klasifikasi dengan menggunakan metode k-Nearest Neighbor (KNN) dengan memanfaatkan struktur taksonomi Wikipedia dalam tahap pra-pemrosesan dokumen yang diharapkan dapat meningkatkan nilai akurasi pengklasifikasian dokumen. Untuk melakukan penelitian ini, penulis menggunakan studi

^[1] Sumber <http://wiki.dbpedia.org/UseCases>

kasus berupa artikel *blog* berbahasa inggris yang berupa halaman RSS dari suatu alamat *blog* yang bersumber pada www.wordpress.com.

1.2 Perumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini adalah seberapa jauh peran struktur taksonomi Wikipedia terhadap sistem klasifikasi artikel *blog* dengan menggunakan metode KNN. Ukuran yang akan digunakan sebagai evaluasi adalah perhitungan nilai akurasi dari hasil pengklasifikasian oleh sistem.

1.3 Batasan Masalah

Demi menghindari meluasnya fokus dalam penelitian ini, maka penulis merumuskan batasan masalah sebagai berikut :

1. Sistem yang akan dibangun adalah sistem berbasis *web* (*web based*) dengan pengujian sistem pada jaringan lokal penulis.
2. Data pelatihan (*training data*) yang digunakan dalam penelitian ini adalah kumpulan abstrak dari artikel berbahasa inggris milik Wikipedia yang bersumber pada downloads.dbpedia.org/3.5.1/en/long_abstracts_en.nt.bz2.
3. Data uji yang digunakan dalam penelitian ini adalah artikel *blog* yang berupa halaman RSS dari sebuah alamat *blog* yang bersumber pada www.wordpress.com.
4. Pemilihan artikel *blog* dilakukan secara manual oleh penulis.
5. Artikel *blog* yang digunakan sebagai dokumen uji hanya artikel berbahasa inggris.
6. Kelas yang disajikan dalam sistem klasifikasi artikel *blog* ini adalah kelas *Person*, dengan sub-kelas yakni *Artist* (*Actor* dan *Musical Artist*), *Athlete* (*Basketball Player* dan *Soccer Player*), dan *Politician* (*President* dan *Senator*).
7. Sistem melakukan pembobotan dengan algoritma *TF-IDF*.

8. Proses transformasi teks meliputi penyeragaman teks menjadi huruf kecil, penghapusan angka dan karakter khusus, penghapusan *stopword* dan tahap *stemming* dengan menggunakan algoritma Porter Stemmer.

1.4 Hipotesis

Hipotesis dari penelitian ini adalah bahwa dengan menggunakan bantuan struktur taksonomi Wikipedia dengan metode KNN pada pengklasifikasian dokumen pelatihan berupa artikel Wikipedia, nilai akurasi hasil klasifikasi dapat ditingkatkan dengan prediksi nilai akurasi lebih besar dari 75%.

1.5 Tujuan Penelitian

Tujuan utama penelitian ini adalah untuk melihat seberapa jauh peran taksonomi Wikipedia dalam meningkatkan nilai akurasi dari hasil pengklasifikasian artikel *blog* dengan menggunakan metode KNN

Adapun sub tujuan dari penelitian ini adalah menghasilkan sebuah sistem yang dapat melakukan pengklasifikasian artikel *blog* dengan nilai akurasi tinggi dengan memanfaatkan struktur taksonomi Wikipedia.

1.6 Metode Penelitian

Metodologi yang digunakan dalam penelitian ini adalah :

- a. Studi pustaka

Studi pustaka dilakukan dengan mencari dan mempelajari sumber-sumber pustaka yang berkaitan dengan topik penelitian yang diambil oleh penulis, yaitu mengenai *text*

mining (proses preprocessing hingga data siap untuk digunakan sebagai masukan pada proses klasifikasi), metode KNN (konsep dan algoritma), sistem berbasis pengetahuan, dan taksonomi Wikipedia (konsep dan penggunaan). Sumber yang digunakan oleh penulis berupa buku dan sumber ilmiah di Internet yang dapat dipercaya.

b. Observasi

Observasi dilakukan dengan mempelajari struktur taksonomi Wikipedia (wiki.dbpedia.org) yang dalam penelitian ini merupakan fokus pengembangan sistem klasifikasi artikel *blog* dengan memanfaatkan struktur yang telah tersedia. Observasi juga dilakukan dengan memperhatikan penggunaan metode KNN dalam proses pengklasifikasian dokumen.

c. Konsultasi dan diskusi

Konsultasi dan diskusi dilakukan bersama dosen dan teman-teman yang memiliki pengalaman di bidang yang sama dengan bidang yang diambil oleh penulis pada penelitian ini.

1.7 Sistematika Penulisan

Penulisan laporan tugas akhir ini akan dibagi menjadi 5 (lima) bagian, yaitu :

a. Bab 1 Pendahuluan

Bab 1 memberikan gambaran umum mengenai hal yang akan diteliti oleh penulis dalam tugas akhir ini. Pendahuluan memuat latar belakang masalah, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode penelitian, dan sistematika penulisan laporan.

b. Bab 2 Tinjauan Pustaka

Bab ini terdiri dari dua sub bab, yaitu tinjauan pustaka dan landasan teori. Tinjauan pustaka memaparkan penelitian-penelitian terdahulu beserta teori yang berkaitan dengan topik penelitian yang diambil oleh penulis, sedangkan landasan teori berisi konsep-konsep yang digunakan dalam mendukung penelitian ini.

c. Bab 3 Analisis dan Perancangan Sistem

Bab 3 terdiri dari beberapa sub bab yang digunakan dalam perancangan sistem, antara lain spesifikasi sistem, arsitektur sistem, diagram *use case*, algoritma dalam membangun sistem, rancangan antarmuka sistem, dan rancangan pengujian sistem.

d. Bab 4 Implementasi dan Analisis Sistem

Bab ini membahas mengenai implementasi serta pengujian sistem yang telah dibangun oleh penulis berdasarkan pada rancangan sistem yang telah diuraikan pada Bab 3. Bab ini juga berisi hasil dari proses yang dilakukan oleh sistem dan analisis dari sistem yang telah berjalan.

e. Bab 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dari penelitian yang dilakukan oleh penulis beserta saran yang diberikan oleh penulis bagi penelitian-penelitian mendatang yang memiliki topik yang sama dengan penelitian ini.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan analisis sistem, maka diperoleh kesimpulan sebagai berikut :

1. Sistem klasifikasi menggunakan metode k-Nearest Neighbor dan memanfaatkan struktur taksonomi Wikipedia terbukti dapat meningkatkan nilai akurasi hasil klasifikasi dengan rata-rata nilai akurasi di atas 75%.
2. Sistem klasifikasi menggunakan metode KNN dan memanfaatkan struktur taksonomi Wikipedia ini dapat menghasilkan tingkat akurasi sebesar 98,83% berdasarkan pada pengujian yang telah dilakukan.
3. Nilai akurasi optimal dicapai pada penggunaan k sebesar 6 dan seleksi fitur sebesar 50%.
4. Pemecahan token berdasarkan frase-frase topik yang telah didaftarkan sebelumnya oleh sistem memberi pengaruh terhadap proses klasifikasi.
5. Urutan token pada artikle tidak mempengaruhi hasil klasifikasi.
6. Pemberian tf yang tinggi bagi token yang terdeteksi sebagai salah satu dari frase yang telah terdaftar memberi pengaruh signifikan bagi tingkat akurasi sistem.

5.2 Saran

Saran yang diajukan oleh penulis untuk pengembangan sistem demi mencapai hasil yang lebih baik dan mendapatkan temuan-temuan yang baru adalah :

1. Basis data milik Wikipedia dapat dimanfaatkan lebih jauh dalam berbagai tujuan pengembangan ilmu pengetahuan, tidak hanya terbatas pada klasifikasi artikel saja.

Oleh karena itu, basis data ini dapat dimanfaatkan demi menunjang penelitian selanjutnya.

2. Dalam mempercepat proses pencarian dan perhitungan data, maka dapat digunakan *indexing* pada basis data. Penggunaan *indexing* ini sangat membantu dalam mempersingkat waktu proses klasifikasi maupun waktu pencarian data.

© UKDW

DAFTAR PUSTAKA

- Feldman R., & Sanger J. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Han E., Karypis G., Kumar V. (2001). *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification*. Diakses 13 Februari 2010, dari http://www.google.co.id/url?sa=t&source=web&ct=res&cd=1&ved=0CAgQFjAA&url=http%3A%2F%2Fglaros.dtc.umn.edu%2Fgkhome%2Ffetch%2Fpapers%2FwknnPAKDD01.pdf&ei=w5uCS9brNIW4rAeqnIjJBw&usq=AFQjCNGCTfMEAxAuJly2KAMqSg1Kxu6fZw&sig2=SONiFUdyfwv_Evj-U8LVJQ.
- Han J. & Kamber M. (2006). *Data Mining : Concepts and Technique 2nd Edition*. San Francisco: Morgan Kaufmann Publishers.
- Manning C. D., Raghavan P., & Schütze H. (2008). *An Introduction to Information Retrieval*. New York: Cambridge University Press.
- Miah M. (2009). *Improved k-NN Algorithm for Text Classification*. Diakses 27 Februari 2010, dari <http://dbxlab.uta.edu/dbxlab/Muhammed/DMIN2009.pdf>.
- Ridok A., Furqon M.T. (2009). *Pengelompokan Dokumen Berbahasa Indonesia Menggunakan Metode K-NN*. Diakses 13 Februari 2010, dari http://matematika.brawijaya.ac.id/web/cms/index2.php?option=com_docman&task=doc_view&gid=319&Itemid=99999999.
- Wang, P., Domeniconi, C., Hu, J. (2008). *Cross Domain Text Classification Using Wikipedia*. Diakses pada tanggal 22 November 2010 dari http://www.comp.hkbu.edu.hk/~cib/2008/IIB08Nov/feature_article_%20IIBWiki.pdf.