

**APLIKASI SISTEM TEMU KEMBALI BERBASIS XML UNTUK MULTISITE BERITA
NASIONAL**

Tugas Akhir



Oleh

Erika

22074278

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2011

**APLIKASI SISTEM TEMU KEMBALI BERBASIS XML UNTUK MULTISITE BERITA
NASIONAL**

Tugas Akhir



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



Disusun Oleh :

Erika
22074278

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

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

APLIKASI SISTEM TEMU KEMBALI BERBASIS XML UNTUK MULTISITE BERITA NASIONAL

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan sarjana Program Studi Teknik Informatika, Fakultas Teknik Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

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

Yogyakarta, 12 September 2011



(Erika)

22074278



HALAMAN PERSETUJUAN


Judul : Aplikasi Sistem Temu Kembali Berbasis XML untuk
Multisite Berita Nasional
Nama : Erika
NIM : 22074278
Mata Kuliah : Tugas Akhir
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2011/2012

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada Tanggal

Dosen Pembimbing I

Dosen Pembimbing II


Budi Susanto, S.Kom, M.T.


Lucia Dwij Krisnawati, S.S., M.A.

HALAMAN PENGESAHAN

SKRIPSI
APLIKASI SISTEM TEMU KEMBALI BERBASIS XML UNTUK MULTISITE
BERITA NASIONAL

Oleh : Erika / 22074278

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

Yogyakarta, 21 Oktober 2011

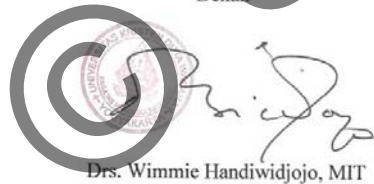
Mengesahkan,

Dewan Penguji :

1. Budi Susanto, S.Kom., M.T.
2. Lucia Dwi Krishawati, S.S., M.A.
3. Restyandito, S.Kom., M.SIS.
4. Antonius Rachmat, S.Kom., M.Cs.

Dekan

Ketua Program Studi



Drs. Wimmie Handiwidjojo, MIT



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

UCAPAN TERIMAKASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerah, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Aplikasi Sistem Temu Kembali Berbasis XML untuk Multisite Berita Nasional dengan baik dan tepat waktu.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunaannya.

Dalam menyelesaikan pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Bapak **Budi Susanto, S.Kom., M.T.**, selaku pembimbing 1, yang telah banyak memberikan ide, masukan, kritik dan saran dalam penulisan laporan dan pembuatan program Tugas Akhir ini.
2. Ibu **Lucia Dwi Krisnawati, S.S., M.A.**, selaku pembimbing 2, yang telah banyak memberikan ide, masukan, kritik dan saran selama penulisan laporan dan pembuatan program Tugas Akhir ini.
3. Keluarga tercinta Papa, Mama, cie Elly, cie Enfina, cie Erawaty, cie Elinda dan cie Elisa untuk segala kasih sayang, kesabaran, perhatian serta dukungan doa yang luar biasa yang selalu menjadi motivasi dan semangat penulis sehingga penulis selalu bersemangat dalam menyelesaikan Tugas Akhir ini.
4. Sahabat-sahabatku Jora, Juta, Willy, Lusi, Dian, Veve, Griffin, Selvi, Eric, Rudy, Martin, dan Nita yang telah memberikan dukungan dan semangat.

5. Teman-teman alumni yang telah membantu dalam pengumpulan data sehingga memperlancar penulis dalam proses penyusunan Tugas akhir ini.
6. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa Program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian. Sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu membuat program Tugas Akhir. Sekali lagi penulis mohon maaf yang sebesar-besarnya. Dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, 12 September 2011

Penulis



INTISARI

Informasi merupakan suatu kebutuhan yang menduduki prioritas tinggi dalam kehidupan manusia. Menemukan informasi yang relevan (sesuai) dengan kebutuhan tidaklah mudah. Konsep dasar sistem temu kembali (*Information Retrieval System*) merupakan sebuah konsep yang dapat dimanfaatkan untuk membangun sebuah aplikasi yang dapat membantu menemukan informasi yang relevan. Pencarian informasi saat ini tidak hanya sebatas data lokal melainkan telah memanfaatkan teknologi internet dalam pencarian informasi secara global, *web search engine*. Dimana format universal dalam *web search engine* adalah XML. Dengan diadopsinya XML sebagai standar publikasi dan pertukaran data terstruktur telah menciptakan kesempatan besar untuk IR yang lebih baik. XML merupakan bahasa penanda berbasis *text* yang diapit oleh *start tags* dan *end tags* sebagai penanda, dan *tag name* menyediakan informasi pada jenis dari *content* yang dikandung (Fuhr & Großjohann, n.d.).

Pencarian informasi berbasis XML dapat dilakukan dengan langkah-langkah berikut: Pertama, *web crawler* akan mengambil data XML dalam bentuk RSS dari situs-situs berita yang telah ditentukan. Kemudian akan dilakukan proses *parsing* terhadap data-data tersebut. Hasil *parsing* tersebut yang akan dihitung bobotnya dan dicocokkan dengan inputan *user* menggunakan VSM. Dikarenakan data merupakan digunakan XML *Retrieval* yang akan menghitung *context resemblance* (CR) dari *keyworduser* terlebih dahulu kemudian baru dihitung tingkat *similarity* menggunakan VSM dengan rumus SIMNOMERGE.

Dengan menggunakan metode VSM akan didapatkan dokumen yang relevan dengan inputan *user*. Dimana contoh inputan *user* dapat dilihat di BAB 1 ataupun di aplikasi.

Kata kunci : *vector space model, xml retrieval, information retrieval, web search engine, web crawler.*

DAFTAR ISI

PERNYATAAN KEASLIAN TUGAS AKHIR.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
UCAPAN TERIMA KASIH.....	iv
INTISARI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
BAB 1.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Hipotesis.....	3
1.5 Tujuan Penelitian.....	3
1.6 Metode Penelitian.....	3
1.7 Sistematika Penulisan.....	4
BAB 2.....	6
TINJAUAN PUSTAKA.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	7
2.2.1 Web Crawler.....	9
2.2.2 Information Retrieval.....	10
2.2.3 XML Retrieval.....	12
2.2.4 Teks Preprocessing.....	14
2.2.5 Stopwords Removal.....	14
2.2.6 Term Weighting.....	15
2.2.7 Vector Space Model (VSM) untuk XML Indexing.....	16
2.2.8 Query Parsing.....	18
BAB 3.....	20
ANALISIS DAN PERANCANGAN SISTEM.....	20

3.1 Spesifikasi Sistem	20
3.2 Rancangan Arsitektur Sistem	22
3.3 Flowchart Diagram	22
3.4 Perancangan Database	27
3.4.1 Proses Pengumpulan Data	27
3.4.2 Sistem <i>Database</i> dan Relasinya	27
3.5 Rancangan <i>User Interface</i>	28
BAB 4	31
IMPLEMENTASI DAN ANALISIS SISTEM	31
4.1 Implementasi Sistem	31
4.1.1 Proses <i>Crawling</i>	31
4.1.2 Proses <i>Indexing</i>	31
4.1.3 Proses <i>Pembobotan</i>	32
4.1.4 Proses <i>Query Parsing</i>	32
4.1.5 Proses <i>Retrieving</i>	33
4.1.6 Proses <i>Sorting</i> Dokumen	34
4.2 Implementasi Tampilan	34
4.2.1 Halaman Awal	34
4.2.2 Halaman Hasil	35
4.2.3 Halaman jika tidak ada dokumen relevan	36
4.3 Pengujian dan Analisis Sistem	37
4.3.1 Pengujian Sistem	37
BAB 5	56
KESIMPULAN DAN SARAN	56
5.1 Kesimpulan	56
5.2 Saran	56
DAFTAR PUSTAKA	57
LAMPIRAN	Lampiran A-1

DAFTAR TABEL

Tabel 4.1 Tabel Hasil Temu Kembali Informasi	37
Tabel 4.2 Tabel Dokumen untuk query “perampokan atm bank muamalat”	39
Tabel 4.3 Tabel Dokumen untuk query “zaenal arifin hoesin”	40
Tabel 4.4 Dokumen untuk query “desc: amandemen uud”	40
Tabel 4.5 Dokumen untuk query “title: aceh gempa”	41
Tabel 4.6 Dokumen untuk query “”anggota dpr doyen intervensi”” ..	42
Tabel 4.7 Tabel Dokumen query “surat palsu mk”	43
Tabel 4.8 Tabel Dokumen query “neneng jadi buronan”	44
Tabel 4.9 Tabel Dokumen query “desc: “nazaruddin ditahan””	45
Tabel 4.10 Tabel Dokumen query “title: “dpr pilih pimpinan kpk”” ..	45
Tabel 4.11 Tabel Dokumen query “mafia pemilu panggil kapolri” ...	46
Tabel 4.12 Tabel Dokumen query “genangan air di underpass kuningan”	47
Tabel 4.13 Tabel Dokumen query “arus mudik lebaran”	48
Tabel 4.14 Tabel Dokumen query “”dugaan suap pembangunan wisma atlet””	49
Tabel 4.15 Tabel Dokumen query “desc: mahasiswa binus menghilang”	50
Tabel 4.16 Tabel Dokumen query “desc: “sabar gorky””	50
Tabel 4.17 Tabel Dokumen query “title: pria tewas di hotel”	51
Tabel 4.18 Tabel Dokumen query “staf kbri paris”	52
Tabel 4.19 Tabel Dokumen query “title: “perangi koruptor””	52
Tabel 4.20 Tabel Dokumen query “”surat nazar””	53
Tabel 4.21 Tabel Dokumen query “mobil rental laris manis”	54

DAFTAR GAMBAR

Gambar 2.1. Arsitektur Web Crawler.....	10
Gambar 2.2. Hubungan antara Dokumen <i>Relevant</i> dan <i>Retrieved</i>	11
Gambar 3.1 Arsitektur Sistem	22
Gambar 3.2 Flowchart Crawler.....	23
Gambar 3.3 Flowchart Indexing	24
Gambar 3.4 Flowchart Retrieving	26
Gambar 3.5 Contoh RSS	27
Gambar 3.6 <i>Database</i> sistem Temu Kembali.....	28
Gambar 3.9 <i>User interface</i> jika berita yang dicari tidak ada.....	30
Gambar 4.1 Halaman Awal.....	35
Gambar 4.2 Halaman Hasil.....	36
Gambar 4.3 Halaman jika tidak ada dokumen relevan.....	36
Gambar 4.4 Grafik <i>Precision query</i> “perampokan atm bank muamalat” pada tiap level <i>Recall</i>	39
Gambar 4.5 Grafik <i>Precision query</i> “zaenal arifin hoesin” pada tiap level <i>Recall</i>	40
Gambar 4.6 Grafik <i>Precision query</i> “desc: amandemen uud”..... pada tiap level <i>Recall</i>	41
Gambar 4.7 Grafik <i>Precision query</i> “title: aceh gempa”..... pada tiap level <i>Recall</i>	42
Gambar 4.8 Grafik <i>Precision query</i> “”anggota dpr doyan intervensi”” pada tiap level <i>Recall</i>	42
Gambar 4.9 Grafik <i>Precision query</i> “”surat palsu mk””..... pada tiap level <i>Recall</i>	43
Gambar 4.10 Grafik <i>Precision query</i> “neneng jadi buronan”	44
pada tiap level <i>Recall</i>	44
Gambar 4.11 Grafik <i>Precision query</i> “desc:”nazaruddin ditahan””... ..	45

Gambar 4.12 Grafik <i>Precision query</i> "title:"dpr pilih pimpinan kpk"" pada tiap level <i>Recall</i>	46
Gambar 4.13 Grafik <i>Precision query</i> "mafia pemilu panggil kapolri" pada tiap level <i>Recall</i>	47
Gambar 4.14 Grafik <i>Precision query</i> "genangan air di underpass kuningan" pada tiap level <i>Recall</i>	48
Gambar 4.15 Grafik <i>Precision query</i> "arus mudik lebaran"..... pada tiap level <i>Recall</i>	49
Gambar 4.16 Grafik <i>Precision query</i> ""dugaan suap pembangunan wisma atlet"" pada tiap level <i>Recall</i>	49
Gambar 4.17 Grafik <i>Precision query</i> "desc:mahasiswi binus menghilang" pada tiap level <i>Recall</i>	50
Gambar 4.18 Grafik <i>Precision query</i> "desc:"sabar gorky"" pada tiap level <i>Recall</i>	51
Gambar 4.19 Grafik <i>Precision query</i> "title:pria tewas di hotel"..... pada tiap level <i>Recall</i>	51
Gambar 4.20 Grafik <i>Precision query</i> "staf kbri paris" pada tiap level <i>Recall</i>	52
Gambar 4.21 Grafik <i>Precision query</i> "title:"perangi koruptor"" pada tiap level <i>Recall</i>	53
Gambar 4.22 Grafik <i>Precision query</i> ""surat nazar"" pada tiap level <i>Recall</i>	53
Gambar 4.23 Grafik <i>Precision query</i> "mobil rental laris manis" pada tiap level <i>Recall</i>	54
Gambar 4.24 Grafik eleven point interpolated average precision.....	55

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Informasi merupakan suatu kebutuhan yang menduduki prioritas tinggi dalam kehidupan manusia. Menemukan informasi yang relevan (sesuai) dengan kebutuhan tidaklah mudah. Konsep dasar sistem temu kembali informasi (*Information Retrieval System*) merupakan sebuah konsep yang dapat dimanfaatkan untuk membangun sebuah aplikasi yang dapat membantu menemukan informasi yang relevan.

Pencarian informasi biasanya menggunakan *query* yang diinputkan oleh *user* kemudian dicocokkan dengan dokumen yang ada dan menghasilkan dokumen yang relevan dengan yang di-inputkan oleh *user*, hal ini biasa disebut dengan proses *Information Retrieval (IR)*. Proses IR sekarang ini bukan hanya melibatkan data terstruktur, text, audio, video, suara, dll, tapi juga telah mencakup XML (*eXtensible Markup Language*) yang telah banyak ada di *database* dan *domain IR* (Grossman & Frieder, 2004).

Dengan menggunakan Proses Temu Kembali Informasi dan metadata berbasis XML, dapat diterapkan sebuah sistem pencarian kembali berita pada beberapa situs berita nasional. Hal ini ditujukan agar para pembaca dapat mempergunakan sistem ini untuk menemukan berita lama maupun baru dari beberapa situs yang diberikan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka sistem ini dirancang untuk memberikan berita yang dicari oleh *user*. Masalah yang dibahas dalam penelitian ini adalah: Seberapa besar *presisi* dan *recall* terhadap sistem IR yang memperhatikan struktur XML sebagai pendukung metadata dengan *Vector Space Model* (VSM) terhadap dokumen berita dalam format RSS (*Really Simple Syndication*).

1.3 Batasan Masalah

Pada permasalahan ini, pembuatan program aplikasi maupun penelitian dibatasi oleh parameter-parameter sebagai berikut ini:

1. Situs-situs yang digunakan adalah sebagai berikut:
 - <http://www.kompas.com/>
 - <http://www.detiknews.com/>
 - <http://www.republika.co.id/>
2. Dokumen yang diambil berupa file XML dalam format RSS berbahasa Indonesia.
3. Data yang dipergunakan adalah *tag title*, *link*, *description*, *pubDate*, *guid* yang berada di dalam *tag item*.
4. Sistem melakukan pembobotan menggunakan algoritma TF-IDF.
5. Inputan pencarian dari *user* merupakan kata kunci, dapat berupa kalimat ataupun berupa sebuah kata, dimana inputan bukan berupa *stopwords* (kata sambung dan kata tunjuk), misalnya: dan, dari, aku, kamu, kita, dll.
 - Pada pencarian berdasarkan *tag title*, *tag desc*, dan *tag date* hanya boleh mengikutsertakan **satu tag** dalam setiap inputan *query*.
 - Pencarian dengan *tag date* dilakukan dengan format dd-mm-yyyy atau dd mm yyyy atau dd/mm/yyyy .

- Pencarian pada *tag* tanggal hanya boleh menggunakan numerik, selainnya hanya boleh menggunakan alfabet.
- Output berupa judul berita dan deskripsi berita dimana judul yang ditampilkan berupa *link* ke berita tersebut.

1.4 Hipotesis

Penggunaan data XML dalam format RSS dengan atribut *title*, *description*, dan *pubDate* dalam pencarian berita nasional dapat menghasilkan pencarian yang cukup akurat, ditinjau dari hasil *presisi* dan *recall*, dengan menggunakan metode *Vector Space Model*.

1.5 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut: Pertama, sistem dapat membantu pembaca menemukan berita lama maupun baru kapanpun. Kedua, mengukur kecepatan proses *indexing* dan *crawling*. Ketiga, mengukur efektifitas algoritma serta pendekatan yang digunakan.

1.6 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah:

- Metode Pengumpulan Data

Data dikumpulkan melalui sumber sekunder, melalui program *crawler* untuk mendapatkan dokumen di internet. Dengan adanya program *crawler* tersebut, setiap 30 menit *crawler* akan dijalankan untuk mendapatkan *update* berita, lalu program *crawler* akan melakukan proses

parsing kemudian dilakukan proses *index* dan dicocokkan dengan *database* apakah *index* tersebut telah ada atau belum. Jika belum ada maka akan dimasukkan ke dalam *database*. Selain itu, akan dilakukan studi pustaka untuk mendapatkan berbagai informasi yang diperlukan dalam penelitian.

- Metode Peng-*index*-an XML

Sebuah dokumen XML disusun oleh sekumpulan elemen yang memungkinkan struktur bertingkat. Dokumen-dokumen XML biasanya direpresentasikan sebagai akar, tersusun, dan pohon berlabel dimana tiap *node* berkoresponden dengan sebuah elemen dan tiap *edge* merepresentasikan hubungan *parent-child* (Popovici, Ménier, & Marteau, 2005). Oleh sebab itu proses *indexing* dapat dilakukan dengan melakukan *indexing* berdasarkan elemen-elemen yang ada dalam sebuah *item*.

- Metode Evaluasi dengan *presisi* dan *recall*

Evaluasi yang dilakukan melibatkan orang dari luar pembangunan program untuk mengetahui hasil *output* yang dicari. Pada metode evaluasi ini program akan dipunggah (*upload*) ke internet, sehingga orang lain dapat menguji program aplikasi tersebut.

- Konsultasi dan diskusi

Konsultasi dan diskusi bersama dosen dan kakak kelas yang berpengalaman dalam *information retrieval*.

1.7 Sistematika Penulisan

Sistematika penulisan dalam laporan ini dibagi menjadi 3 bagian utama, yaitu bagian awal, bagian utama dan bagian akhir. Pada bagian awal, laporan memuat beberapa poin awal, yaitu halaman sampul depan, sampul dalam, pernyataan keaslian skripsi, halaman persetujuan, halaman pengesahan, ucapan terima kasih, abstrak, daftar isi, daftar tabel, dan daftar gambar.

Pada bagian utama, laporan memuat 5 (lima) bab, yaitu bab pendahuluan, bab tinjauan pustaka, bab analisis dan perancangansistem, bab implementasi dan analisis sistem dan bab kesimpulan dan saran.

Bab pendahuluan berisi latar belakang masalah yang akan diteliti dan rencana penelitian yang akan dilakukan. Bab pendahuluan mempunyai 6(enam) poin penjabaran, yaitu latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode/pendekatan dan sistematika penulisan.

Bab landasan teori terdiri dari dua sub bab utama, yaitu tinjauan pustaka dan landasan teori. Tinjauan pustaka menjelaskan berbagai teori yang didapatkan dari berbagai sumber pustaka yang digunakan untuk penyusunan Tugas Akhir. Landasan teori berisi uraian dari konsep-konsep atau teori-teori yang dipakai sebagai dasar pembuatan tugas akhir ini.

Bab analisis dan perancangan sistem mencakup analisa teori – teori yang digunakan dan bagaimana menerjemahkannya ke dalam suatu sistem yang hendak dibuat. Pada dasarnya bab ini memuat rancangan pembuatan program, prosedur-prosedur yang ada di dalamnya, materi yang dipakai dalam riset, variable yang dipakai dan data yang dikumpulkan, serta algoritma dan flowchart sistem.

Bab implementasi dan analisis sistem berisi penjelasan tentang bagaimana rancangan pada Bab 3 diimplementasikan dalam suatu bahasa pemrograman. Hasil riset disajikan dalam bentuk gambar (*capture*) dan ditempatkan sedekat-dekatnya dengan pembahasan atau analisisnya, sehingga uraian dapat lebih mudah dimengerti.

Bab Kesimpulan dan Saran berisi pernyataan singkat dan tepat yang dijabarkan dari hasil analisis kegiatan riset/implementasi dalam penyusunan tugas akhir. Saran dalam laporan ini memuat teknik-teknik pengembangan yang belum dilakukan dalam riset namun dirasa akan memperbaiki kinerja sistem jika langkah tersebut dilaksanakan pada riset yang mendatang.

Pada bab akhir, laporan memuat daftar pustaka dan lampiran.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi sistem, dapat disimpulkan bahwa sistem pencarian berita berbasis XML dengan metode *Vector Space Model* dengan algoritma SIMNOMERGE dapat menghasilkan dokumen-dokumen yang relevan dengan yang diinginkan oleh *user*. Dan menghasilkan rata-rata presisi sebesar 77,79509% dan nilai *recall* sebesar 95%. Kesimpulan lain yang didapat adalah inputan *query* yang menggunakan frasa (“”) dan *metadata* akan meningkatkan presisi. Semakin detil *query user* akan semakin tepat hasil pencarian yang dilakukan oleh sistem.

5.2 Saran

Saran untuk pengembangan dan perbaikan sistem adalah pengembangan sistem dengan menggunakan *boolean operator* untuk pencarian dokumen. Saran lainnya adalah pengembangan sistem dengan algoritma yang berbeda untuk membandingkan hasil temu kembali. Serta pengembangan sistem untuk memperbaharui proses indeks.

DAFTAR PUSTAKA

- Adam, G., Bouras, C., Pouloupoulos, V. (2009). 'Utilizing the RSS feeds for Crawling the Web'. ICIW '09, International Conference on Internet and Web Applications and Services. hal 211-216.
- Blanken, Henk M., Grabs, T., Schek, H.-J., Schenkel, R., & Weikum, G. (2003). *Intelligent Search on XML Data, Applications, Languages, Models, Implementations and Benchmarks*. New York: Springer.
- Carmel, D., Efraty, N., Landau, G. M., Maarek, Y. S., & Mass, Y. (2002). 'An Extention of the Vector Space Model for Querying XML Documents via XML Fragments'. SIGIR Forum. Vol 36(2).
- Egnor, D., & Lord, R. *Structured Information Retrieval Using XML*. Diakses tanggal 11 Maret 2011 dari <https://www.research.ibm.com/haifa/sigir00-xml/final-papers/Egnor/>
- Feldman, R. & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing an Unstructured Data*. New York: Cambridge University Press.
- Fuhr, N., Großjohann, K., & Kriewel, S. (2003). 'A Query Language and User Interface for XML Information Retrieval'. In Henk Blanken et al. (Eds.), *Intelligent Search on XML Data* (pp. 59-75). Berlin: Springer.
- Fuhr, N., Großjohann, K. (2001). 'XIRQL: A Query Language for Information Retrieval in XML Documents'. SIGIR 2001. Hal 172-180.
- Grossman, D. A., Frieder, O. (2004). *Information Retrieval: Algorithm and Heuristics*. Netherlands: Springer.
- Han, J. & Kamber, M., 2001, *Data Mining: Concepts and Techniques*. San Diego: Academic Press.
- Keizer, P. (2006). 'Indexing Methods for XML Documents'. 4th Twente Student Conference on IT , Enschede 30 January, 2006.

- Kotsakis, E. (2002). 'Structured Information Retrieval in XML Documents'. ACM Symposium on Applied Computing.
- Manning, C. D., Raghavan, P., Schütze, H. (2008). *An Introduction to Information Retrieval*. England: Cambridge University Press.
- Moens, M.-F. (2006). *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Netherlands: Springer.
- Mustaqim, T. (2008). *Analisis dan Implementasi Penerapan Algoritma Genetika pada Information Retrieval*. Diakses tanggal 11 Maret 2011 dari http://www.ittelkom.ac.id/library/index.php?view=article&catid=20%3Ainformatika&id=540%3Ainformation-retrieval&option=com_content&Itemid=15
- Nelson, P. (2010). *The Magic and Wonder of Query Parsing*. Diakses tanggal 11 Maret 2011 dari <http://www.searchtechnologies.com/search-query-parsing.html>
- Olston, C., Najork, M. (2010). 'Web Crawling'. Foundations and Trends in Information Retrieval. Vol 4(3). Hal 175-246.
- Popovici, E., Ménier, G., & Marteau, P.-F. (2005). 'SIRIUS: A Lightweight XML Indexing and Approximate Search System at INEX 2005'. Lecture Notes in Computer Science. 3977, hal 321-335.
- Salton, G. (1975). *Dynamic Information and Library Processing*. Prentice Hall.

