Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

Team project ©2017
Dony Pratidana S. Hum | Bima Agus Setyawan S. IIP

BAB V

PENGUJIAN DAN PEMBAHASAN

5.1 Pengujian

5.1.1 Data

Dataset crawling yang digunakan adalah hasil *crawling* pada situs www.umn.ac.id per tanggal 26 Desember 2011. *Dataset crawling* dari tv.umn.ac.id tidak digunakan dalam pengujian implementasi *real system* karena didapat temuan pada tahap pengujian implementasi *coding* ternyata metode pencarian pada situs tv.umn.ac.id berbeda dengan skema penelitian ini. Adapun jumlah artikel dalam tv.umn.ac.id juga terlalu sedikit, yaitu 6 artikel per tanggal 28 Desember 2011.

Dalam proses pengujian implementasi *coding*, dilakukan pengujian atas skrip dalam stemmer.php dan query.php. Digunakan 2 kata dasar, 2 kata ulang, 16 kata berimbuhan awalan, 9 kata berimbuhan akhiran, dan 4 kata berimbuhan awalan akhiran untuk diujicobakan pada form *stemmer* untuk memastikan bahwa penulisan algoritma *stemming* Nazief-Andriani sudah tepat dalam stemmer.php. Digunakan juga 10 *queries* yang masing-masing mengandung operator pencarian AND, OR, *quote pharse*, dan *tilde*, serta *stop words* untuk diujicobakan pada form *generate query* untuk memastikan bahwa penulisan langkah-langkah *generate query* dalam query.php sudah tepat.

Dalam proses pengujian *stemming* digunakan 872 kata yang dikelompokkan ke dalam 190 kelompok semantik. Data ini digunakan untuk mendapat UI dan OI.

Web Search Engine..., Stephani Marcelli Djojo, FTI UMN, 2012

Dalam proses pengujian interpretasi *query* digunakan 37 *queries* menggunakan 3 kelas kata (kata tunggal/jamak, kata gabung, dan kata kerja dengan banyak akhiran), operator pencarian dan *stop words*. Data ini diujicobakan untuk memastikan *query-query* yang di-*generate* interpretasinya tepat.

Dalam proses pengujian relevansi hasil pencarian digunakan 50 *queries* yang rata-rata mengandung 2-3 kata setiap *query*-nya. Data ini diujicobakan pada model yang dikembangkan dan pada *search engine* situs UMN untuk dibandingkan jumlah hasil pencarian yang relevan guna mendapat nilai peningkatan jumlah hasil pencarian relevan yang dilakukan oleh model yang dikembangkan.

Data yang digunakan selama pengujian disajikan dalam Lampiran D Tabel 2 hingga Tabel 8.

5.1.2 Kamus

Pada penelitian ini, kamus kata dasar diambil dari daftar kata dasar pada kamus Bahtera, yang dapat diunduh dari situs bahtera.org. Kamus Bahtera adalah kamus bahasa Indonesia yang menjadi rujukan sesuai Kamus Besar Bahasa Indonesia (KBBI). Kata dasar yang terdapat di Bahtera sebanyak 28.526 kata.

5.1.3 Stop Words

Stopwords yang digunakan dalam penelitian ini terdapat 329 kata, yang dapat diunduh dari fpmipa.upi.edu/staff/yudi/stop_words_list.txt.

UNIVERSITAS MULTIMEDIA NUSANTARA

5.1.4 Variabel Pengujian

Variabel penelitian dibatasi pada dua parameter metode evaluasi Paice, yaitu *understemming index* dan *overstemming index* untuk pengujian *stemming*.

Di sisi output, hasil pencarian diuji untuk memastikan menginterpretasikan *query* secara benar dan menyajikan judul, URL, dan potongan konten yang tepat. Pengujian interpretasi *query* menggunakan *form generate query*.

5.2 Hasil Pengujian

Pada pengujian implementasi *coding* didapat bahwa stemmer.php sudah mengimplementasikan algoritma *stemming* Nazief-Andriani dengan beberapa perbaikan yang ditambahkan (Agusta, 2009) diwakili dengan 35 kata. Skrip dalam query.php sudah mengimplementasikan langkah-langkah yang benar untuk *generate query* diwakili dengan 7 *queries* yang mengandung operator pencarian AND, OR, *quote pharse*, dan *tilde* serta *stop words*. Hasil pengujian implemantasi *coding* dapat dilihat dalam lampiran.

Pada pengujian *stemming*, setiap kelompok semantik memiliki nilai DMT, UMT, dan DNT masing-masing. Begitupun dengan setiap *stemmed group* memiliki nilai WMT-nya masing-masing. Hasil pengujian *stemming* dapat dilihat dalam Lampiran D. Hasil *stemmed word* yang tidak berhasil didapat kata dasarnya diberi tanda huruf tebal (*bold*). Pada bab ini disajikan 8 kelompok semantik hasil pengujian *stemming* beserta perhitungan manual nilai DMT, UMT, DNT, dan WMT.

UNIVERSITAS MULTIMEDIA NUSANTARA

Tabel 5.1 *Semantic group* dan hasil *stemming*

	Full word	Stemmed words	N_g	n_g	DMT	UMT	DNT
g1	sekolah	sekolah	5	4	10	2	2167.5
	bersekolah	**seko**					
	menyekolahkan	sekolah					
	persekolahan	sekolah					
	disekolahkan	sekolah					
g10	proses	proses	5	4	10	2	2167.5
	memproses	proses					
	pemprosesan	proses					
	diprosesi	**prosesi**					
	terprosesnyalah	proses					
g20	ketik	ketik	5	3	10	6	2167.5
	mengetik	**etik**					
	pengetik	**etik**					
	ketikan	ketik					
	diketikkan	ketik					
g22	kerja	kerja	5	4	10	2	2167.5
	kerjaanku	kerja					
	bekerja	kerja					
	mempekerjakan	**mempekerjakan**					
	pekerja	kerja					
g168	lompat	lompat	5	5	10	0	2167.5
	melompat	lompat					
	dilompati	lompat					
	melompati	lompat					
	lompatannyalah	lompat					
g173	prosesi	prosesi	1	1	0	0	435.5
g174	etik	etik	1	1	0	0	435.5
g190	seko	seko	1	1	0	0	435.5

Dalam hasil *stemming semantic group* g1 terlihat bahwa kata [bersekolah] memiliki hasil *stemming* yang salah karena menemukan bentuk dasar [seko]. Hasil *stemming* [bersekolah] akan tepat jika hasilnya [sekolah]. Berikut Gambar 5.1 adalah *print screen* dari form *stemmer* yang memperlihatkan proses pemenggalan imbuhan dari kata [bersekolah].

Gambar 5.1 Proses *stemming* kata [bersekolah]

Berikut adalah perhitungan manual untuk *semantic group* g1.

$DMT_{g1} = 0.5\ N_{g1}\ (N_{g1} - 1)$

$DMT_{g1} = 0.5\ .\ 5\ (\ 5\text{-}1\ )$

$DMT_{g1} = 10$

$$UMT_{g1} = 0.5 \sum_{i \in [1..f_g]} n_{g1\ i}\ (N_{g1} - n_{g1\ i})$$

$UMT_{g1} = 0.5\ (\ n_{g1\ 1}\ (N_{g1} - n_{g1\ 1})\ )$

$UMT_{g1} = 0.5\ (\ 4(5\text{-}4)\ )$

$UMT_{g1} = 2$

$DNT_{g1} = 0.5\ N_{g1}\ (\ W - N_{g1}\ )$

$DNT_{g1} = 0.5\ .\ 5\ (872 - 5)$

$DNT_{g1} = 2167.5$

Dalam hasil *stemming semantic group* g10 terlihat bahwa kata [diprosesi] memiliki hasil *stemming* yang salah karena menemukan bentuk dasar [prosesi]. Hasil *stemming* [diprosesi] akan tepat jika hasilnya [proses]. Berikut Gambar 5.2 adalah *print screen* dari form *stemmer* yang memperlihatkan proses pemenggalan imbuhan dari kata [diprosesi].

Gambar 5.2 Proses *stemming* kata [diprosesi]

Berikut adalah perhitungan manual untuk *semantic group* g10.

$DMT_{g10} = 0.5\ N_{g10}\ (N_{g10} - 1)$

$DMT_{g10} = 0.5\ .\ 5\ (\ 5\text{-}1\ )$

$DMT_{g10} = 10$

$$UMT_{g10} = 0.5 \sum_{i \in [1..f_g]} n_{g10\ i}\ (N_{g10} - n_{g10\ i})$$

$UMT_{g10} = 0.5\ (\ n_{g10\ 1}\ (N_{g10} - n_{g10\ 1})\ )$

$UMT_{g10} = 0.5\ (\ 4(5\text{-}4)\ )$

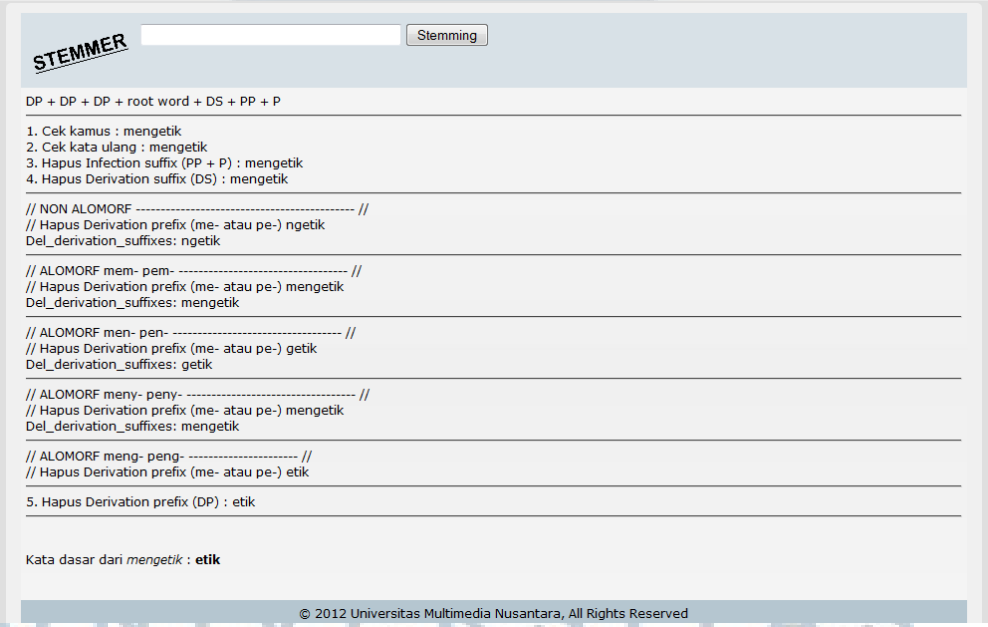
$UMT_{g10} = 2$

$DNT_{g10} = 0.5\ N_{g10}\ (W - N_{g10})$

$DNT_{g10} = 0.5 \cdot 5\ (872 - 5)$

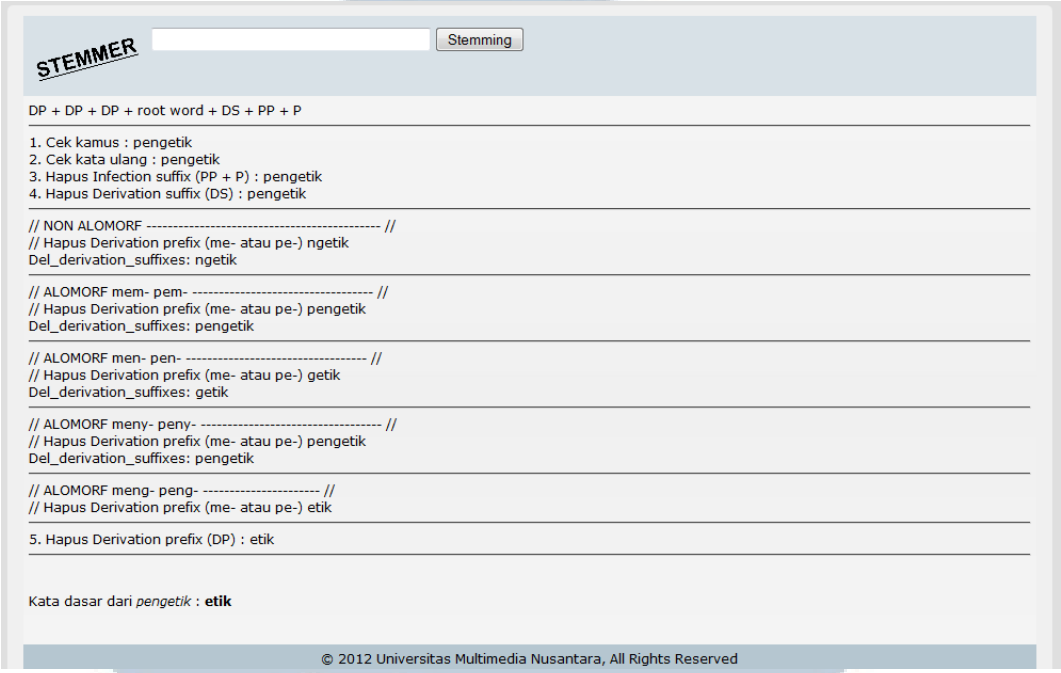
$DNT_{g10} = 2167.5$

Dalam hasil *stemming semantic group* g20 terlihat bahwa kata [mengetik] dan [pengetik] memiliki hasil *stemming* yang salah karena menemukan bentuk dasar [etik]. Hasil *stemming* [mengetik] dan [pengetik] akan tepat jika hasilnya [ketik]. Berikut Gambar 5.3 dan Gambar 5.4 adalah *print screen* dari form *stemmer* yang memperlihatkan proses pemenggalan imbuhan dari kata [mengetik] dan [pengetik].



Gambar 5.3 Proses *stemming* kata [mengetik]

UNIVERSITAS MULTIMEDIA NUSANTARA



Gambar 5.4 Proses *stemming* kata [pengetik]

Berikut adalah perhitungan manual untuk *semantic group* g20.

$DMT_{g20} = 0.5\ N_{g20}\ (N_{g20} - 1)$

$DMT_{g20} = 0.5\ .\ 5\ (\ 5\text{-}1\ )$

$DMT_{g20} = 10$

$$UMT_{g20} = 0.5 \sum_{i \in [1..f_g]} n_{g20\ i}\ (N_{g20} - n_{g20\ i})$$

$UMT_{g20} = 0.5\ (\ n_{g20\ 1}\ (N_{g20} - n_{g20\ 1}) + n_{g20\ 2}\ (N_{g20} - n_{g20\ 2})\ \ )$

$UMT_{g20} = 0.5\ (\ 3(5\text{-}3) + 3(5\text{-}3)\ )$

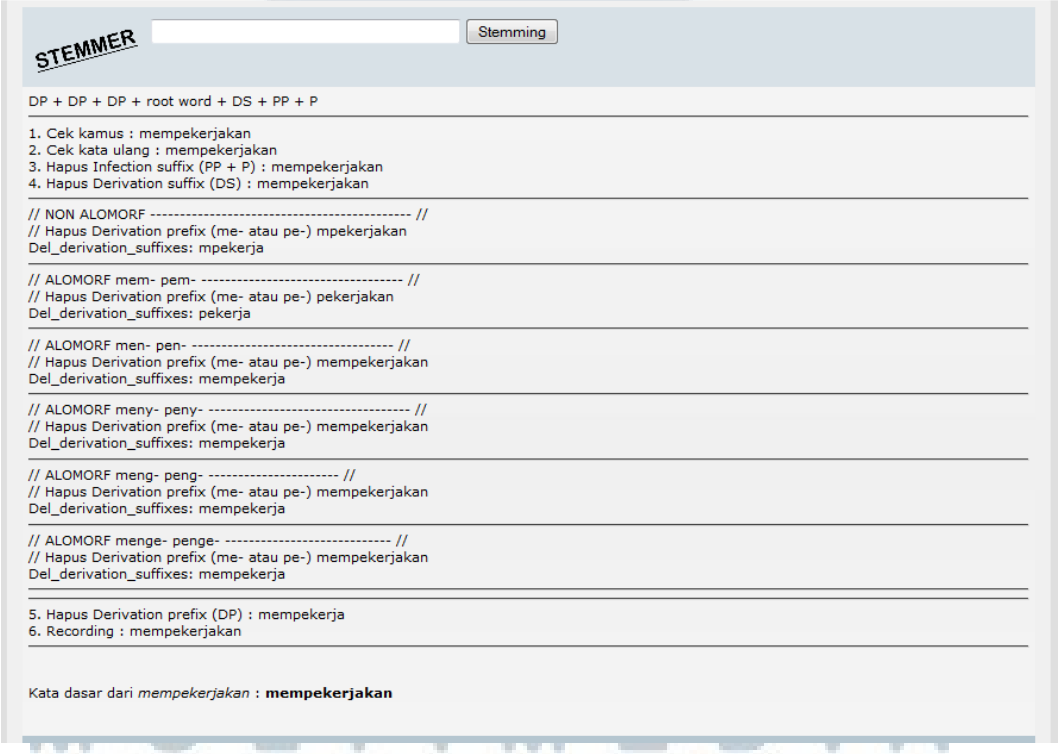
$UMT_{g20} = 6$

$DNT_{g20} = 0.5\ N_{g20}\ (W - N_{g20})$

$DNT_{g20} = 0.5 \cdot 5\ (872 - 5)$

$DNT_{g20} = 2167.5$

Dalam hasil *stemming semantic group* g22 terlihat bahwa kata [mempekerjakan] tidak berhasil mendapat bentuk kata dasarnya. Hasil *stemming* [mempekerjakan] akan tepat jika hasilnya [kerja]. Berikut Gambar 5.5 adalah *print screen* dari form *stemmer* yang memperlihatkan proses pemenggalan imbuhan dari kata [mempekerjakan].



Gambar 5.5 Proses *stemming* kata [mempekerjakan]

Berikut adalah perhitungan manual untuk *semantic group* g22.

$DMT_{g22} = 0.5\ N_{g22}\ (N_{g22} - 1)$

$DMT_{g22} = 0.5\ .\ 5\ (\ 5\text{-}1\ )$

$DMT_{g22} = 10$

$$UMT_{g22} = 0.5 \sum_{i\ \epsilon\ [1..f_g]} n_{g22\ i}\ (N_{g22} - n_{g22\ i})$$

$UMT_{g22} = 0.5\ (\ n_{g22\ 1}\ (N_{g22} - n_{g22\ 1})\ )$

$UMT_{g22} = 0.5\ (\ 4(5\text{-}4)\ )$

$UMT_{g22} = 2$

$DNT_{g22} = 0.5\ N_{g22}\ (\ W - N_{g22}\ )$

$DNT_{g22} = 0.5\ .\ 5\ (872 - 5)$

$DNT_{g22} = 2167.5$

Dalam *semantic group* g168, semua kata berhasil di-*stemming* ke bentuk kata dasar yang tepat. Berikut adalah perhitungan manual untuk *semantic group* g168.

$DMT_{g168} = 0.5\ N_{g168}\ (N_{g168} - 1)$

$DMT_{g168} = 0.5\ .\ 5\ (\ 5\text{-}1\ )$

$DMT_{g168} = 10$

$UMT_{g168} = 0$

$DNT_{g168} = 0.5\ N_{g168}\ (\ W - N_{g168}\ )$

$DNT_{g168} = 0.5\ .\ 5\ (872 - 5)$

$DNT_{g168} = 2167.5$

Berikut adalah perhitungan manual untuk *semantic group* g173.

$DMT_{g173} = 0$

$UMT_{g173} = 0$

$DNT_{g173} = 0.5\ N_{g173}\ (W - N_{g173})$

$DNT_{g173} = 0.5 . 1 (872 - 1)$

$DNT_{g173} = 435.5$

Berikut adalah perhitungan manual untuk *semantic group* g174.

$DMT_{g174} = 0$

$UMT_{g174} = 0$

$DNT_{g174} = 0.5\ N_{g174}\ (W - N_{g174})$

$DNT_{g174} = 0.5 . 1 (872 - 1)$

$DNT_{g174} = 435.5$

Berikut adalah perhitungan manual untuk *semantic group* g190.

$DMT_{g190} = 0$

$UMT_{g190} = 0$

$DNT_{g190} = 0.5\ N_{g190}\ (W - N_{g190})$

$DNT_{g190} = 0.5 . 1 (872 - 1)$

$DNT_{g190} = 435.5$

UNIVERSITAS MULTIMEDIA NUSANTARA

Berikut Tabel 5.2 adalah hasil susun ulang dari *stemmed words* untuk delapan *semantic group*.

Tabel 5.2 *Stem group* sebagai hasil susun ulang dari *stemmed words*

Full word	**Stemmed words**	**Original semantic group**	**N_s**	**n_s**	**WMT**
etik	etik	g174	3	1	2
mengetik	**etik**	**g20**			
pengetik	**etik**	**g20**			
ketik	ketik	g20	3	3	0
ketikan	ketik	g20			
diketikkan	ketik	g20			
kerja	kerja	g22	4	4	0
kerjaanku	kerja	g22			
bekerja	kerja	g22			
pekerja	kerja	g22			
lompat	lompat	g168	5	5	0
melompat	lompat	g168			
dilompati	lompat	g168			
melompati	lompat	g168			
lompatannyalah	lompat	g168			
mempekerjakan	**mempekerjakan**	**g22**	1	0	0
proses	proses	g10	4	4	0
memproses	proses	g10			
pemprosesan	proses	g10			
terprosesnyalah	proses	g10			
diprosesi	**prosesi**	**g10**	2	1	0.5
prosesi	prosesi	g173			
bersekolah	**seko**	**g1**	2	1	0.5
seko	seko	g190			
sekolah	sekolah	g1	4	4	0
menyekolahkan	sekolah	g1			
persekolahan	sekolah	g1			
disekolahkan	sekolah	g1			

UNIVERSITAS MULTIMEDIA NUSANTARA

Berikut adalah perhitungan manual untuk *stem group* [etik].

$$WMT_{s1} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s1\,i} (N_{s1} - n_{s1\,i})$$

$$WMT_{s1} = 0.5 \; (n_{s1\,1} (N_{s1} - n_{s1\,1}) + n_{s1\,2} (N_{s1} - n_{s1\,2}))$$

$$WMT_{s1} = 0.5 \; (1(3-1) + 1(3-1))$$

$$WMT_{s1} = 2$$

Berikut adalah perhitungan manual untuk *stem group* [ketik].

$$WMT_{s2} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s2\,i} (N_{s2} - n_{s2\,i})$$

$$WMT_{s2} = 0$$

Berikut adalah perhitungan manual untuk *stem group* [kerja].

$$WMT_{s3} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s1\,i} (N_{s3} - n_{s3\,i})$$

$$WMT_{s3} = 0$$

Berikut adalah perhitungan manual untuk *stem group* [lompat].

$$WMT_{s4} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s4\,i} (N_{s4} - n_{s4\,i})$$

$$WMT_{s4} = 0$$

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Berikut adalah perhitungan manual untuk *stem group* [mempekerjakan].

$$WMT_{s5} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s5\,i} (N_{s5} - n_{s5\,i})$$

$$WMT_{s5} = 0.5 (n_{s5\,1} (N_{s5} - n_{s5\,1}))$$

$$WMT_{s5} = 0.5 (0(1-0))$$

$$WMT_{s5} = 0$$

Berikut adalah perhitungan manual untuk *stem group* [proses].

$$WMT_{s6} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s6\,i} (N_{s6} - n_{s6\,i})$$

$$WMT_{s6} = 0$$

Berikut adalah perhitungan manual untuk *stem group* [prosesi].

$$WMT_{s7} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s7\,i} (N_{s7} - n_{s7\,i})$$

$$WMT_{s7} = 0.5 (n_{s7\,1} (N_{s7} - n_{s7\,1}))$$

$$WMT_{s7} = 0.5 (1 (2-1))$$

$$WMT_{s7} = 0.5$$

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Berikut adalah perhitungan manual untuk *stem group* [seko].

$$WMT_{s8} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s8\,i} (N_{s8} - n_{s8\,i})$$

$$WMT_{s8} = 0.5 (n_{s8\,1} (N_{s8} - n_{s8\,1}))$$

$$WMT_{s8} = 0.5 (1 (2-1))$$

$$WMT_{s8} = 0.5$$

Berikut adalah perhitungan manual untuk *stem group* [sekolah].

$$WMT_{s9} = 0.5 \sum_{i \in [1 \ldots f_s]} n_{s9\,i} (N_{s9} - n_{s9\,i})$$

$$WMT_{s9} = 0$$

Dari semua nilai DMT, UMT, DNT, dan WMT didapat nilai total GDMT, GUMT, GDNT, dan GWMT, serta UI dan OI. Berikut perhitungan manualnya.

$$GDMT = \sum_{i \in n_g} DMT_{gi}$$

$$GDMT = 1699$$

$$GUMT = \sum_{i \in n_g} UMT_{gi}$$

$$GUMT = 142.5$$

UNIVERSITAS MULTIMEDIA NUSANTARA

UI = GUMT/GDMT

UI = 142.5 / 1699 = 0.08387287 = 8.39 x 10^{-2}

$$GDNT = \sum_{i \in [1 \ldots n_g]} DNT_{gi}$$

GDNT = 378057

$$GWMT = \sum_{i \in n_s} WMT_{si}$$

GWMT = 24

OI = GWMT/GDNT

OI = 24/378057 = 0.00006348 = 6.35 x 10^{-5}

Berikut 5 hasil pengujian interpretasi *query* yang dilakukan. Keseluruhan hasil pengujian interpretasi *query* dimuat dalam lampiran.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 5.3 Hasil pengujian interpretasi *query*

	Query	**Keterangan Query**	**Interpreted Query**	**Ketepatan**
Kata tunggal/jamak	memenangkan lomba-lomba	- Mencari artikel yang mengandung kata/*stem* [memenangkan] dan [lomba-lomba] - Urutan kata dalam artikel kata/*stem* [memenangkan] terlebih dahulu kemudian [lomba-lomba] - Di antara kata/*stem* [memenangkan] dan kata/*stem* [lomba-lomba] boleh sembarang karakter lain	memenangkan%lomba-lomba memenangkan%lomba menang%lomba-lomba menang%lomba	Tepat
	menang per-lombaan-perlombaan	- Mencari artikel yang mengandung kata/*stem* [menang] dan [perlombaan-perlombaan] - Urutan kata dalam artikel kata/*stem* [menang] terle-bih dahulu kemudian [perlombaan-perlombaan] - Di antara kata/*stem* [menang] dan kata/*stem* [per-lombaan-perlombaan] boleh sembarang karakter lain	menang%perlombaan-perlombaan menang%lomba	Tepat
Kata gabung	~menarinarikan	- Operator tilde (~) membuat *query single word* dicari juga bentuk *stem*-nya - Query memiliki bentuk *stem* [tari] - Mencari artikel yang mengandung kata [menarinarikan] atau [tari]	menarinarikan tari	Tepat
	~dijualbelikan	- Operator *tilde* (~) membuat *query single word* dicari juga bentuk stemnya - *Query* memiliki bentuk *stem* [jual-beli] - Mencari artikel yang mengandung kata [dijualbelikan] atau [jual-beli]	dijualbelikan	Kurang tepat, karena bentuk stem [jual-beli] tidak diin-terpretasikan

UNIVERSITAS MULTIMEDIA NUSANTARA

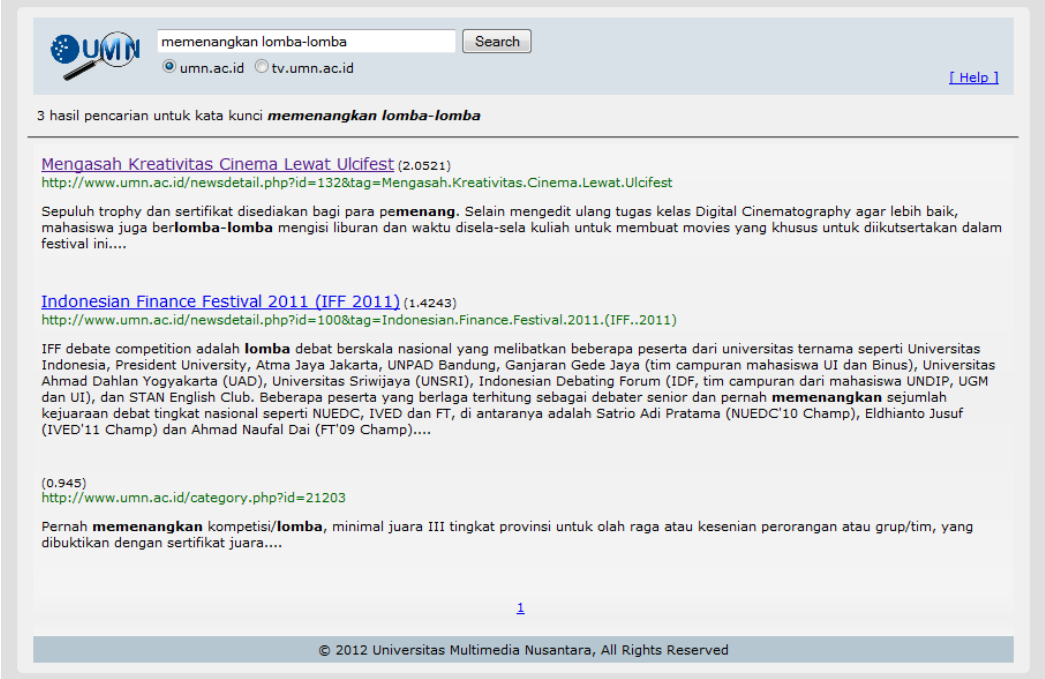
Tabel 5.3 Hasil pengujian interpretasi *query* (lanjutan)

	Query	**Keterangan Query**	**Interpreted Query**	**Ketepatan**
Kata kerja banyak akhiran	ditampilkannyalah gramedia	- Mencari artikel yang mengandung kata/*stem* [ditampilkammyalah] dan [gramedia] - Urutan kata dalam artikel kata/*stem* [ditampilkannyalah] terlebih dahulu kemudian [gramedia] - Di antara setiap kata dalam *query* boleh sembarang karakter lain	ditampilkannyalah%gramedia tampil%gramedia	Tepat

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Berikut Gambar 5.6 dan Gambar 5.7 adalah *print screen* dari *form generate query* dan hasil pencarian dari lima *query* [memenangkan lomba-lomba].

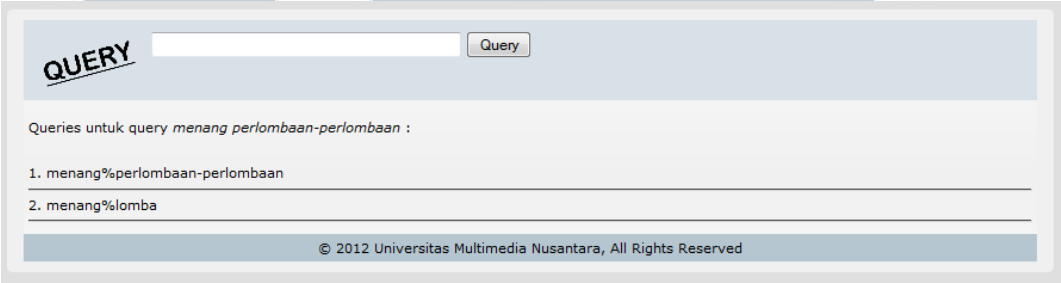
Gambar 5.6 Hasil *generate query* [memenangkan lomba-lomba]



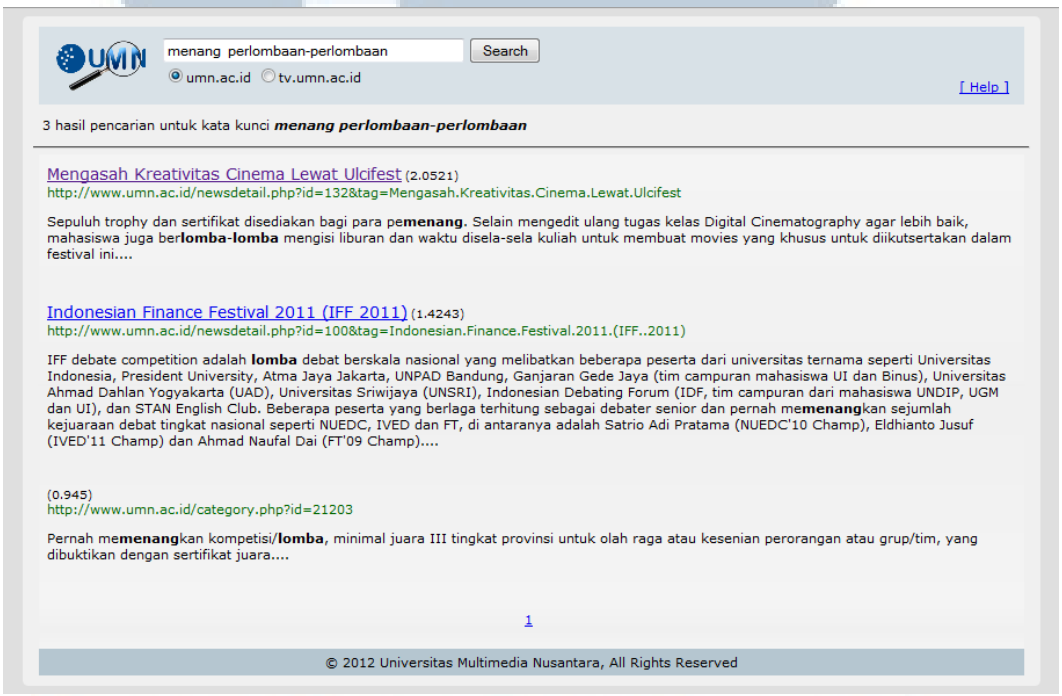
Gambar 5.7 Hasil pencarian *query* [memenangkan lomba-lomba]

UNIVERSITAS MULTIMEDIA NUSANTARA

Berikut Gambar 5.8 dan Gambar 5.9 adalah *print screen* dari *form generate query* dan hasil pencarian dari lima *query* [menang perlombaan-perlombaan].



Gambar 5.8 Hasil *generate query* [menang perlombaan-perlombaan]



Gambar 5.9 Hasil pencarian *query* [menang perlombaan-perlombaan]

UNIVERSITAS MULTIMEDIA NUSANTARA

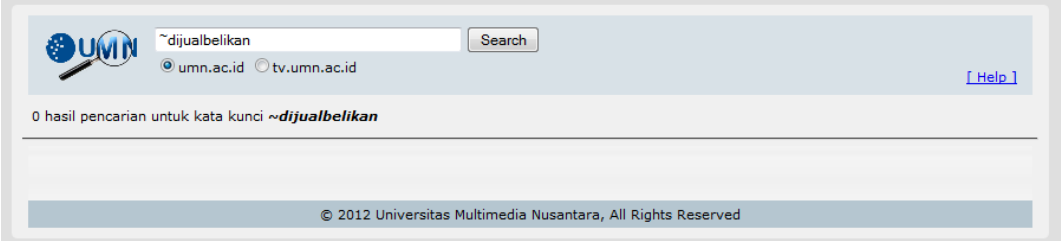
Berikut Gambar 5.10 dan Gambar 5.11 adalah *print screen* dari *form generate query* dan hasil pencarian dari lima *query* [~menarinarikan].

Gambar 5.10 Hasil *generate query* [~menarinarikan]

Gambar 5.11 Hasil pencarian *query* [~menarinarikan]

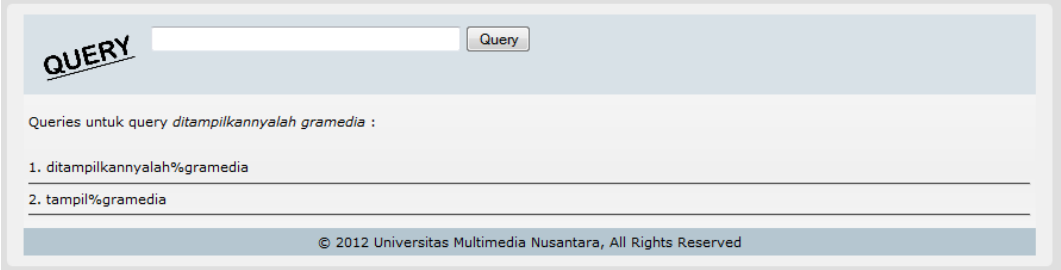
Berikut Gambar 5.12 dan Gambar 5.13 adalah *print screen* dari *form generate query* dan hasil pencarian dari lima *query* [~dijualbelikan].

Gambar 5.12 Hasil *generate query* [~dijualbelikan]



Gambar 5.13 Hasil pencarian *query* [~dijualbelikan]

Berikut Gambar 5.14 dan Gambar 5.15 adalah *print screen* dari *form generate query* dan hasil pencarian dari lima *query* [terpikirkannyakah kreatif].



Gambar 5.14 Hasil *generate query* [terpikirkannyakah kreatif]

Gambar 5.15 Hasil pencarian *query* [terpikirkannyakah kreatif]

Berikut hasil pengujian relevansi hasil pencarian yang dilakukan.

Tabel 5.4 Hasil pengujian relevansi hasil pencarian

No	Query	Jumlah Hasil Pencarian				Pe-ning-katan
		Situs UMN		Model pengem-bangan		
		Match content	Relevan content	Match content	Relevan content	
1	menyukseskan UMN	0	0	2	2	2
2	ninok menjadi rektor	0	0	2	2	2
3	penampilannya teater katak	0	0	1	1	1
4	menyongsong masa depan	0	0	1	1	1
5	UMN pengajarannya berkualitas	0	0	1	1	1
6	pendaftaran online	7	5	3	3	-2
7	kemajuan teknologi informasi	0	0	3	3	3
8	peresmian kampus UMN	0	0	2	1	1
9	lokasi ke UMN	0	0	2	2	2
10	pemprograman android	0	0	1	1	1

Tabel 5.4 Hasil pengujian relevansi hasil pencarian (lanjutan)

No	Query	Jumlah Hasil Pencarian				Pe-ning-katan
		Situs UMN		Model pengem-bangan		
		Match content	Relevan content	Match content	Relevan content	
11	UMN membantu yang terkena musibah	0	0	1	1	1
12	lulusan sudah bekerja	0	0	6	6	6
13	guru-guru belajar	0	0	3	1	1
14	unggulan di bidang ICT	1	1	1	1	0
15	berkuliah di UMN	0	0	8	8	8
16	pembangunan gedung-gedung UMN	0	0	1	1	1
17	kawasan UMN pendidikan	0	0	1	1	1
18	lulusan diwisuda	0	0	1	1	1
19	menjuarai perlombaan	0	0	3	3	3
20	berhasil membawa piala	0	0	1	1	1
21	beasiswa dari kompas grame-dia	0	0	3	3	3
22	menonton teater	0	0	1	1	1
23	pengembangan pembinaan	0	0	1	1	1
24	atlet perenang	0	0	1	1	1
25	permainan karnaval	0	0	1	1	1
26	riset penelitian	0	0	1	1	1
27	kekeluargaan keakraban	0	0	1	1	1
28	menyekolahkan putra-putri ke UMN	0	0	1	1	1
29	ajang ide kreatif	0	0	2	2	2
30	kunjungan duta besar	1	1	1	1	0
31	pembangunan green building	0	0	1	1	1
32	UKM taekwondo mendapat medali	0	0	1	1	1
33	pentingnya pendidikan	1	1	2	2	1
34	kedatangan pelajar SMA	0	0	1	1	1
35	perolehan medali	0	0	3	3	3
36	persyaratan keuangan	0	0	3	3	3
37	ujian memasuki UMN	0	0	1	1	1
38	kejuaraan tingkat nasional	1	1	5	5	4
39	mengundang SD Cihuni	0	0	1	1	1
40	mahasiswa DKV menggambar	0	0	1	1	1

Tabel 5.4 Hasil pengujian relevansi hasil pencarian (lanjutan)

No	Query	Jumlah Hasil Pencarian				Pening-katan
		Situs UMN		Model pengemban-gan		
		Match content	Relevan content	Match content	Relevan content	
41	prosedur pendaftaran online	1	1	2	2	1
42	siap bekerja	1	1	4	4	3
43	dihadiri siswa sekolah	0	0	1	1	1
44	memotivasi untuk berkarya	0	0	1	1	1
45	yohanes surya meresmikan	0	0	2	2	2
46	dormitory UMN	1	1	2	2	1
47	kerja sama dengan kompas gramedia	1	1	8	8	7
48	karakteristik masyarakat	1	1	1	1	0
49	intelektual mahasiswa UMN	0	0	1	1	1
50	mahasiswa menggali poten-sinya	0	0	1	1	1
Total			**14**		**96**	**82**
Persentase peningkatan						**585.71**

$$\text{Persentase peningkatan} = \frac{(\text{relevan content model - relevan content situs UMN})}{\text{relevan content situs UMN}} * 100\%$$

$$= \frac{(96 - 14)}{14} * 100\%$$

$$= \mathbf{585.71\ \%}$$

Dengan kata lain, persentase peningkatan yang sebesar 585.71% menunjukkan model yang dibangun memberi peningkatan jumlah hasil pencarian relevan 5 kali lebih besar dari jumlah hasil pencarian relevan di *search engine* situs UMN.

UNIVERSITAS MULTIMEDIA NUSANTARA

5.3 Pembahasan

Algoritma *stemming* Nazief-Andriani menggunakan acuan kamus untuk memeriksa apakah pemenggalan imbuhan sudah sampai ke bentuk dasar kata. Oleh karena itu, keberhasilan proses *stemming* yang dilakukan sangat bergantung pada kamus kata dasar yang digunakan. Dengan menggunakan kamus Bahtera, dari 190 kelompok semantik, ditemukan semua kata dasar kelompok semantik tersebut. Dengan kata lain, dari percobaan ini tidak ditemukan kendala dari keterbatasan jumlah kata dasar dalam kamus.

Hasil *stemming* dari pengujian tidak semuanya tepat karena keterbatasan algoritma Nazief-Andriani dalam mengenkapsulasi dan menyusun aturan morfologi Bahasa Indonesia.

Keterbatasan algoritma *stemming* Nazief-Andriani ini bisa dilihat dari proses *stemming* kata [mengetik] dan [pengetik] menghasilkan kata dasar [etik]. Kata [etik] dan [ketik] keduanya terdapat dalam kamus Bahasa Indonesia. Namun pemenggalan imbuhan dengan algoritma Nazief-Andriani menemukan kata [etik] terlebih dahulu dan ditemukan juga dalam kamus. Hal ini yang menyebabkan hasil *stemming* dari kata [mengetik] dan [pengetik] adalah [etik]. Masalah ini dapat dilihat melalui Gambar 5.16 dibawah ini:

UNIVERSITAS
MULTIMEDIA
NUSANTARA

```
if($huruf_awal=='a'||$huruf_awal=='i'||$huruf_awal=='u'||
   $huruf_awal=='e'||$huruf_awal=='o'||$huruf_awal=='g'||
   $huruf_awal=='h') {
      if(Cek_kamus($_kata)) {
            $ketemu=1;
            $return['kata'] = $_kata;
            $return['ketemu'] = $ketemu;
            return $return;
       }
       if(Cek_kamus("k".$_kata)) {
            $ketemu=1;
            $return['kata'] = "k".$_kata;
            $return['ketemu'] = $ketemu;
            return $return;
       }
}
```

Gambar 5.16 Potongan *source code* stemmer.php

Dengan melihat alur dari algoritma Nazief-Andriani tersebut bukan berarti mendahulukan pemeriksaan dengan membubuhi huruf ‘k’ terlebih dahulu maka algoritma Nazief-Andriani sudah dapat mengatasi masalah untuk alomorf “meng-(k)”. Hal ini bisa dibuktikan dengan mengujikan kata [mengerakan]. Kata dasar [mengerakan] adalah [era]. Hasil *stemming* dari algoritma Nazief-Andriani tepat menemukan kata dasar [era]. Kata [kera] terdapat dalam kamus Bahasa Indonesia. Jika dilakukan perbaikan dengan mendahulukan pemeriksaan kamus dengan pembubuhan huruf ‘k’ terlebih dahulu, maka hasil *stemming* dari kata [mengerakan] adalah [kera]. Dapat dikatakan hal ini sebagai ketidakteraturan dari Bahasa Indonesia.

Keterbatasan algoritma *stemming* Nazief-Andriani juga dapat dilihat dari proses *stemming* kata [mempekerjakan]. Algoritma Nazief-Andriani tidak berhasil menemukan kata dasar [mempekerjakan] sehingga dilakukan *recording*. Padahal dalam algoritmanya, pemenggalan *derivation prefix* diiterasi sebanyak tiga kali. Kesalahan ini dikarenakan aturan awalan “me-” yang diikuti kata dasar

yang berawalan huruf ‘p’ maka huruf ‘p’ tersebut luluh. Tetapi lain halnya jika awalan “me-” + “pe-”, huruf ‘p’ dari imbuhan “pe-” tidak luluh dalam morfologi Bahasa Indonesia.

Kesalahan lain yang didapat dari hasil *stemming* adalah kata [bersekolah] dan [diprosesi]. Kedua kata tersebut mendapat kata dasar yang salah karena benar mengikuti tahapan algoritma Nazief-Andriani. Dari kata [bersekolah], algoritma ini melakukan pemenggalan partikel (*-lah, -kah, -tah*) paling awal dibandingkan pemenggalan imbuhan lainnya. Oleh algoritma Nazief-Andriani, “-lah” dari kata [bersekolah] dianggap sebagai partikel. Sedangkan kata [diprosesi], algoritma ini melakukan pengecekan ke kamus sebelum pemenggalan akhiran dan ternyata kata [prosesi] terdapat dalam kamus sehingga dianggap sebagai kata dasar.

Dari hasil pengujian *stemming* diperoleh dua nilai parameter metode evaluasi Paice, yaitu *understemming index* sebesar 8.39×10^{-2} dan *overstemming index* sebesar 6.35×10^{-5}. Dengan mengacu pada metode evaluasi Paice, yaitu algoritma *stemming* yang baik memiliki nilai *understemming index* dan *overstemming index* sekecil mungkin, maka dapat dikatakan algoritma Nazief-Andriani berhasil melakukan proses *stemming* yang diwakili dengan 190 kelompok semantik yang terdiri dari 872 kata.

Pengujian interpretasi *query* yang dilakukan pada penelitian ini menunjukkan bahwa interpretasi *query* kurang tepat jika dalam *query* terdapat kata gabung yang bukan berasal dari kata ulang berimbuhan (awalan dan akhiran) yang melalukan proses *stemming*, seperti kata gabung [dijualbelikan]. Kurang

tepatnya interpretasi *query* ini karena algoritma Nazief-Andriani dengan perbaikannya tidak mendukung pemenggalan imbuhan dari kata gabung.

Dalam pengujian yang dilakukan, operator pencarian, AND, OR, *tilde* (~), dan *quote pharse* (“ ”) sudah berhasil menginterpretasikan *query* secara tepat. Begitu juga dengan *stop words* yang terdapat dalam *query* pencarian berhasil dihapus dari *query* pencarian. Namun jika dilihat dari hasil pencarian sepertinya terjadi kesalahan dalam menginterpretasi *query* yang mengandung operator AND baik eksplisit maupun implisit. Dalam hasil pencarian, kata kunci dibuat menjadi huruf tebal (*bold*). Gambar 5.7 adalah *print screen* dari hasil pencarian *query* [memenangkan lomba-lomba]. Secara implisit antara kata [memenangkan] dan [lomba-lomba] terdapat operator AND yang berarti urutan dalam hasil pencarian harus kata/*stem* [memenangkan] terlebih dahulu kemudian kata/*stem* [lomba-lomba]. Namun terlihat di hasil pencarian nomor keduanya kata [lomba] ditemukan lebih dahulu kemudian kata [memenangkan]. Hal ini bukan kesalahan dari interpretasi *query* karena jika artikel tersebut dibaca kelanjutannya, kata [lomba] terdapat setelah kata [memenangkan]. Berikut *print screen* dari artikel tersebut.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

News

Monday, Mar/21/2011

Indonesian Finance Festival 2011 (IFF 2011)

Masih tentang *Debate Competition* yang sedikit demi sedikit menunjukkan eksistensi intelektual dari para mahasiswa UMN. Di bulan Maret 2011, sejumlah mahasiswa UMN berlaga dalam *Indonesian Finance Festival* 2011 (IFF 2011) yang diadakan oleh STAN (Sekolah Tinggi Akuntansi Negara) di Bintaro, Tangerang Selatan.

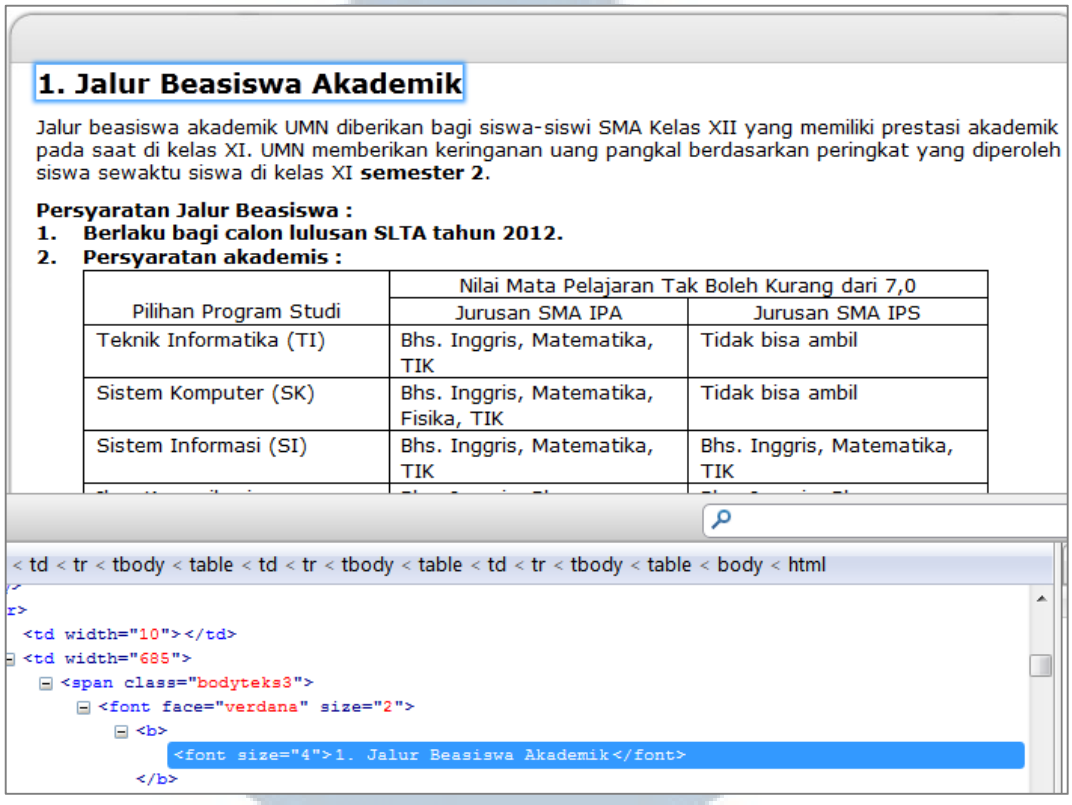
IFF debate competition adalah lomba debat berskala nasional yang melibatkan beberapa peserta dari universitas ternama seperti Universitas Indonesia, President University, Atma Jaya Jakarta, UNPAD Bandung, Ganjaran Gede Jaya (tim campuran mahasiswa UI dan Binus), Universitas Ahmad Dahlan Yogyakarta (UAD), Universitas Sriwijaya (UNSRI), *Indonesian Debating Forum* (IDF, tim campuran dari mahasiswa UNDIP, UGM dan UI), dan STAN English Club. Beberapa peserta yang berlaga terhitung sebagai *debater* senior dan pernah memenangkan sejumlah kejuaraan debat tingkat nasional seperti NUEDC, IVED dan FT, di antaranya adalah Satrio Adi Pratama (NUEDC'10 *Champ*), Eldhianto Jusuf (IVED'11 *Champ*) dan Ahmad Naufal Dai (FT'09 *Champ*). Sementara, para mahasiswa UMN yang berpartisipasi dalam lomba ini masih terhitung sebagai orang baru (*novices / newbies*) di ajang debat mahasiswa nasional. Mereka adalah Helga Liliani Cakra Dewi (Ilkom '10), Jessica Kristie Vancy Longdong (Ilkom '10), dan Joseph Reinaldo (DKV '08).

Gambar 5.17 Artikel Indonesian Finance Festival 2011

Penelitian ini tidak fokus pada *string matching* kata kunci di hasil pencarian. Potongan konten diperoleh dengan cara mengambil per kalimat yang mengandung kata kunci, tidak mempedulikan *term order* dari kata kunci.

Adapun dalam hasil pencarian terdapat beberapa artikel yang tidak berhasil diambil judul artikel atau isi kontennya, seperti hasil pencarian ketiga dari *query* [memenangkan lomba-lomba] yang dapat dilihat pada Gambar 5.7. Hal ini dikarenakan judul artikel tidak memakai tag HTML yang umumnya digunakan oleh judul artikel-artikel lainnya atau karena tidak ada judul atau berisi gambar.

Judul artikel dalam www.umn.ac.id umumnya diawali dengan tag “<span class=\"bodyteks11\"><b>”. Sedangkan artikel pada URL http://www.umn.ac.id/category.php?id=21203 ini menggunakan awalan tag “<b><font size=”4”>”. Berikut Gambar 5.18 adalah hasil *print screen* dari hasil inspeksi elemen dari URL http://www.umn.ac.id/category.php?id=21203.



1. Jalur Beasiswa Akademik

Jalur beasiswa akademik UMN diberikan bagi siswa-siswi SMA Kelas XII yang memiliki prestasi akademik pada saat di kelas XI. UMN memberikan keringanan uang pangkal berdasarkan peringkat yang diperoleh siswa sewaktu siswa di kelas XI **semester 2**.

Persyaratan Jalur Beasiswa :

1. **Berlaku bagi calon lulusan SLTA tahun 2012.**
2. **Persyaratan akademis :**

Pilihan Program Studi	Nilai Mata Pelajaran Tak Boleh Kurang dari 7,0	
	Jurusan SMA IPA	Jurusan SMA IPS
Teknik Informatika (TI)	Bhs. Inggris, Matematika, TIK	Tidak bisa ambil
Sistem Komputer (SK)	Bhs. Inggris, Matematika, Fisika, TIK	Tidak bisa ambil
Sistem Informasi (SI)	Bhs. Inggris, Matematika, TIK	Bhs. Inggris, Matematika, TIK

```
< td < tr < tbody < table < td < tr < tbody < table < td < tr < tbody < table < body < html
r>
  <td width="10"></td>
  <td width="685">
    <span class="bodyteks3">
      <font face="verdana" size="2">
        <b>
          <font size="4">1. Jalur Beasiswa Akademik</font>
        </b>
```

Gambar 5.18 Inspeksi elemen dari URL

http://www.umn.ac.id/category.php?id=21203

Ada juga artikel yang tidak ditemukan adanya judul, seperti yang dapat dilihat di URL http://www.umn.ac.id/category.php?id=21204. Berikut *hasil print screen* artikel dalam URL tersebut.

UNIVERSITAS MULTIMEDIA NUSANTARA

Jalur Prestasi UMN adalah jalur penerimaan mahasiswa baru **tanpa tes**. Setiap mahasiswa yang memenuhi persyaratan diterima langsung dan berhak mendapatkan grade kelulusan : **Grade A**, yakni grade tertinggi dengan uang pangkal terendah.

Persyaratan Jalur Prestasi :
1. Berlaku bagi calon lulusan SLTA tahun 2012.
2. Memenuhi persyaratan akademik sebagai berikut :

Pilihan Program Studi	Nilai Mata Pelajaran Tak Boleh Kurang dari 7,0 atau KKM	
	Jurusan SMA IPA	Jurusan SMA IPS
Teknik Informatika (TI)	Bhs. Inggris, Matematika	Tidak bisa ambil
Sistem Komputer (SK)	Bhs. Inggris, Matematika, Fisika	Tidak bisa ambil
Sistem Informasi (SI)	Bhs. Inggris, Matematika	Bhs. Inggris, Matematika
Ilmu Komunikasi (Ilkom)	Bhs. Inggris, Bhs. Indonesia	Bhs. Inggris, Bhs. Indonesia
Manajemen	Bhs. Inggris, Matematika	Bhs. Inggris, Matematika
Akuntansi	Bhs. Inggris, Matematika	Bhs. Inggris, Matematika
Desain Komunikasi Visual (DKV)	Bhs. Inggris, Matematika	Bhs. Inggris, Matematika

Catatan :
- Jurusan Bahasa bisa mengambil : Prodi Ilmu Komunikasi dan DKV.
- SMK Jurusan TI/Multimedia bisa ambil : Prodi TI, SI dan DKV.
- SMK Jurusan Elektronika bisa ambil : Prodi SK.
- SMK Jurusan Bisnis dan Akuntansi bisa ambil : Prodi Manajemen dan Akuntansi
- SMK Jurusan Sekretaris bisa ambil : Prodi Ilkom.

Prosedur Pendafataran :
1. Mengisi formulir pendaftaran atau daftar online di website : www.umn.ac.id
2. Melampirkan fotocopy raport kelas XI semester 1 dan 2.

Gambar 5.19 Tidak adanya judul artikel dari URL

http://www.umn.ac.id/category.php?id=21204

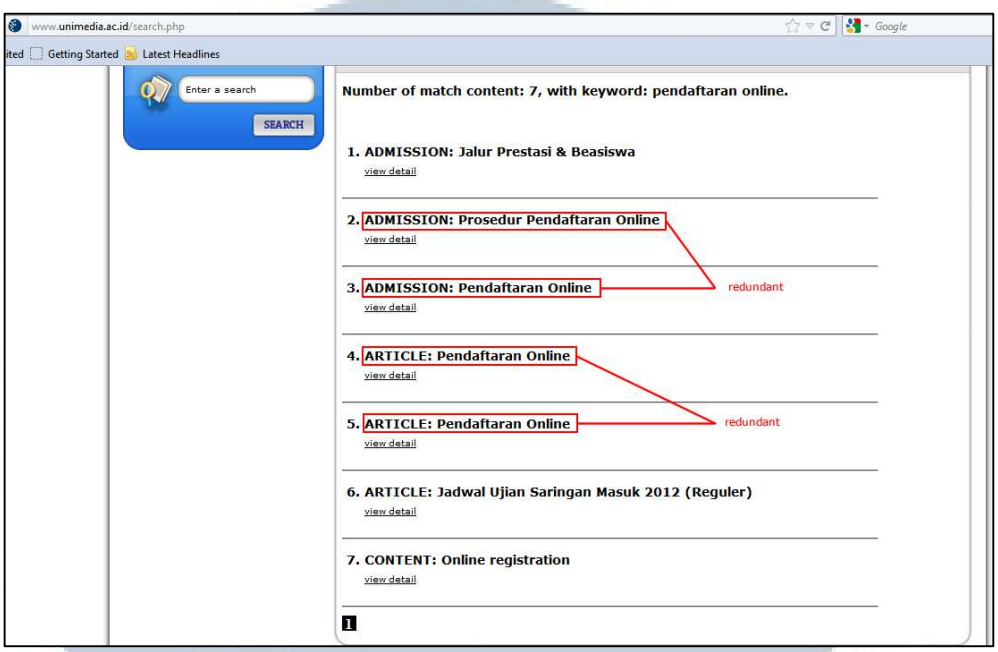
Konsekuensi dari *search engine* yang mengimplementasikan *stemming* dapat dilihat pada contoh *query* [~menarinarikan] di Gambar 5.11. *Query* [~menarinarikan] mendapat 7 hasil pencarian di mana artikel hasil pencariannya tidak semua mengandung kata [tari] seperti yang dimaksud. Pencarian kata [tari] membuat artikel yang sebenarnya mengandung kata [sekretaris] pun ikut terambil. Hal ini tidak dapat dihindari dalam implementasi *stemming* pada *search engine*. Sesuai dengan telaah literatur yang dilakukan sebelumnya, kebiasaan pengguna *search engine* akan menginputkan *query* yang rata-rata terdiri dari dua

sampai tiga kata. Oleh karena itu, *query* yang hanya terdiri dari satu kata tidak dilakukan *stemming* kecuali diberi karakter *tilde* (~). Untuk mendapat hasil pencarian yang akurat dapat dilakukan dengan mempertajam *query*, yaitu dengan menambahkan kata kunci dalam *query* pencarian.

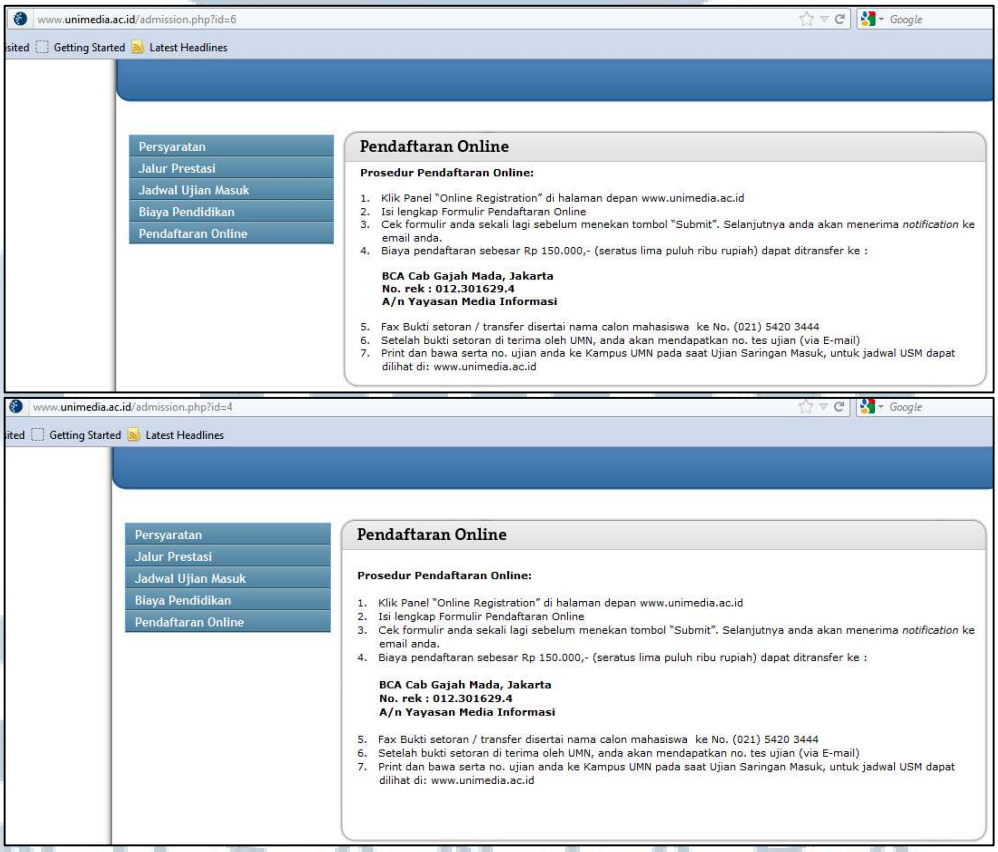
Pengujian relevansi hasil pencarian didapat bahwa model *web search engine* yang dikembangkan memberi peningkatan jumlah hasil relevan 5 kali lebih besar dari hasil pencarian relevan *search engine* situs UMN. Seperti yang tercantum dalam telaah literatur, pengukuran relevan atau tidaknya hasil pencarian dengan *query* dilihat dari ada tidaknya penyimpangan kata kunci (konsekuansi *stemming*) dalam hasil pencarian. Adanya penyimpangan kata kunci dalam hasil pencarian maka hasil pencarian tersebut dianggap tidak relevan.

Jumlah hasil pencarian dari model *search engine* tidak selalu lebih besar dibandingkan jumlah hasil pencarian dari *search engine* situs UMN, bahkan memungkinkan jumlah hasil pencarian dari model lebih sedikit. Penyebabnya beberapa hasil pencarian di situs UMN *redundant*. Hal ini dapat dilihat dari pengujian *query* [pendaftaran online] seperti dalam Gambar 5.20 dan Gambar 5.21.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 5.20 Hasil pencarian *query* [pendaftaran online] dalam situs UMN



Gambar 5.21 *Redundant* antara "Prosesdur pendaftaran online" dan "Pendaftaran online"

Dengan beberapa hasil pencarian yang *redundant* membuat hasil pencarian di situs UMN terkesan mendapat hasil pencarian lebih banyak. Oleh karena itu, dalam kolom relevan *content* dari tabel pengujian relevansi, jumlah hasil pencarian yang relevan dari *query* [pendaftaran online] hanya lima.

Alasan lain yang menyebabkan hasil pencarian di situs UMN lebih banyak adalah permasalahan klasik yang ada pada penelitian sebelumnya, yaitu ada beberapa *link* yang tidak terindex oleh *crawler* karena situs UMN menggunakan JavaScript untuk memasukkan beberapa *link* tertentu. Hal ini terjadi pada *query* [pendaftaran online]. Diperoleh jumlah hasil relevan di situs UMN sebanyak lima sedangkan di model yang dikembangkan hanya ditemukan tiga.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA