

**PENGEMBANGAN SISTEM INFORMATION RETRIEVAL  
UNTUK BAHASA INDONESIA BERBASIS WEB  
MENGUNAKAN VECTOR SPACE MODEL**

Skripsi



Diajukan oleh:

**DANNY JOE DOZAN**

**71130027**

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2019

**PENGEMBANGAN SISTEM INFORMATION RETRIEVAL  
UNTUK BAHASA INDONESIA BERBASIS WEB  
MENGUNAKAN VECTOR SPACE MODEL**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika

Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:

**DANNY JOE DOZAN**

**71130027**

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA

2019

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **PENGEMBANGAN SISTEM INFORMATION RETRIEVAL UNTUK BAHASA INDONESIA BERBASIS WEB MENGGUNAKAN VECTOR SPACE MODEL**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi 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 dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 24 Mei 2019



DANNY JOE DOZAN

71130027

## HALAMAN PERSETUJUAN

Judul Skripsi : PENGEMBANGAN SISTEM INFORMATION  
RETRIEVAL UNTUK BAHASA INDONESIA  
BERBASIS WEB MENGGUNAKAN VECTOR  
SPACE MODEL

Nama Mahasiswa : DANNY JOE DOZAN

N I M : 71130027

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 24 Mei 2019

Dosen Pembimbing I



Danny Sebastian, S.Kom., M.M., M.T.

Dosen Pembimbing II



Gloria Virginia, S.Kom., MAI, Ph.D.

## HALAMAN PENGESAHAN

### PENGEMBANGAN SISTEM INFORMATION RETRIEVAL UNTUK BAHASA INDONESIA BERBASIS WEB MENGGUNAKAN VECTOR SPACE MODEL

Oleh: DANNY JOE DOZAN / 71130027

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 14 Juni 2019

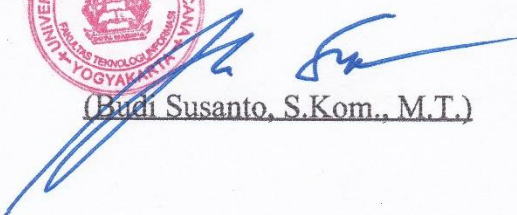
Yogyakarta, 1 Juli 2019  
Mengesahkan,

Dewan Penguji:


1. Danny Sebastian, S.Kom., M.M., M.T.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Budi Susanto, SKom., M.T.
4. Joko Purwadi, M.Kom



Dekan

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

Ketua Program Studi

  
(Gloria Virginia, Ph.D.)



## UCAPAN TERIMA KASIH

Pertama-tama Penulis mengucapkan syukur kepada Tuhan karena atas berkat-Nya, Penulis dapat menyelesaikan tugas akhir dengan judul “Pengembangan Sistem Information Retrieval untuk Bahasa Indonesia Berbasis Web menggunakan Vector Space Model”. Selama proses penulisan tugas akhir, penulis mendapatkan banyak dukungan dan bantuan dari banyak pihak. Oleh sebab itu, penulis ingin mengucapkan banyak terima kasih kepada:

1. David Joe Dozan dan Juliawaty selaku orang tua dari Penulis yang selalu mendoakan dan menyemangati supaya Penulis dapat menyelesaikan tugas akhir dan studinya
2. Bapak Danny Sebastian, S.Kom., M.M., M.T. selaku dosen pembimbing Penulis
3. Ibu Gloria Virginia, S.Kom, MAI, Ph.D selaku Ketua Program Studi dan dosen pembimbing kedua Penulis
4. Raxy Dwy Brundy selaku teman dan partner dalam penelitian ini
5. Teman-teman kuliah Penulis yang memberi semangat kepada Penulis

Akhir kata, Penulis sekali lagi mengucapkan banyak terima kasih kepada pihak yang telah disebutkan dan yang tidak dapat disebutkan satu per satu. Semoga hasil penelitian ini dapat digunakan untuk berbagai pihak yang terkait dan pembaca.

Yogyakarta, 1 Juli 2019

DANNY JOE DOZAN  
71130027

## INTISARI

### Pengembangan Sistem Information Retrieval untuk Bahasa Indonesia Berbasis Web menggunakan Vector Space Model

Sistem temu kembali adalah sistem untuk memperoleh informasi dari kumpulan sumber informasi yang relevan dengan keinginan pengguna. Salah satu pengembang mesin pencarian yang bernama Solr menyediakan fitur untuk memberi *term expansion* berupa *synonym*. Fitur ini digunakan pada penelitian ini untuk memasukkan tesaurus. Tesaurus merupakan hubungan antar kata yang dikelompokkan menjadi satu kelompok. Tesaurus yang digunakan pada penelitian ini menggunakan hasil penelitian dari Himawan (2015) yang berbentuk file RDF.

Pada penelitian ini ditambahkan fitur kata majemuk untuk menggabungkan dua kata menjadi satu supaya pada saat *query matching* tidak terpisah. Untuk melihat tingkat keefektifitas saat *re-trieve* dokumen, penulis membandingkan sistem jika ditambah *stemming* dengan tidak ditambah *stemming*. Sistem yang dibangun pada penelitian ini dibentuk menjadi sebuah *search engine* yang berbasis *javascript*.

Sistem yang dibangun diuji *F-Measure* untuk 20 topik utama yang telah ditentukan dan menghasilkan nilai 0.3728 untuk sistem dengan *stemming* dan 0.3716 untuk sistem tanpa *stemming*. Setelah itu, sistem yang dibangun diuji kepada responden dengan 30 topik yang berbeda dan dihitung *Mean Average Precision*-nya yang bernilai 0.6225 untuk sistem dengan *stemming* dan 0.6279 untuk sistem tanpa *stemming*. Dapat disimpulkan bahwa sistem yang dibangun dengan *ad-hoc evaluation* lebih baik sistem dengan *stemming*, namun jika dicoba langsung kepada responden lebih baik sistem tanpa *stemming*.

Kata kunci: Solr, SKOS, Tesaurus, Search Engine, Information Retrieval

## DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE UNIVERSITAS KRISTEN DUTA WACANA.....	<b>Error! Bookmark not defined.</b>
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1    PENDAHULUAN.....	1
1.1    Latar Belakang Masalah.....	1
1.2    Perumusan Masalah.....	2
1.3    Batasan Masalah.....	2
1.4    Tujuan Penelitian.....	2
1.5    Manfaat Penelitian.....	3
1.6    Metode / Pendekatan.....	3
1.6.1    Tahap Studi Literatur.....	3
1.6.2    Tahap pembangunan basis data kata majemuk.....	3
1.6.3    Tahap pembangunan sistem.....	4
1.6.4    Evaluasi.....	4



1.6.5	Final Report.....	4
1.7	Sistematika Penulisan.....	4
<b>BAB 2</b>	<b>TINJAUAN PUSTAKA .....</b>	<b>6</b>
2.1	Tinjauan Pustaka .....	6
2.2	Landasan Teori .....	6
2.2.1	Information Retrieval .....	6
2.2.2	Preprocessing Text .....	7
2.2.3	<i>Corpus</i> .....	7
2.2.4	Solr .....	8
2.2.5	<i>Resource Description Framework</i> .....	9
2.2.6	<i>Simple Knowledge Organization System</i> .....	9
2.2.7	Metode Evaluasi Precision dan Recall.....	10
2.2.8	F-measure.....	11
<b>BAB 3</b>	<b>PERANCANGAN SISTEM.....</b>	<b>12</b>
3.1	Analisis Kebutuhan .....	12
3.1.1	Analisis Data.....	12
3.1.2	Kebutuhan Perangkat Keras.....	15
3.1.3	Kebutuhan Perangkat Lunak.....	16
3.1.4	Kebutuhan Teknologi dan Library.....	16
3.2	Arsitektur Sistem.....	16
3.3	Perancangan Desain Antarmuka Sistem.....	17
3.4	Perancangan Evaluasi Sistem.....	18
<b>BAB 4</b>	<b>IMPLEMENTASI DAN ANALISIS SISTEM.....</b>	<b>19</b>
4.1	Implementasi Sistem .....	19
4.1.1	Tampilan Antarmuka .....	19

4.1.2	Bagan Implementasi Sistem.....	20
4.1.3	Fase perubahan tesaurus dalam RDF menjadi file teks.....	21
4.1.4	Fase pembentukan basis data kata majemuk.....	21
4.1.5	Fase Implementasi sistem .....	22
4.2	Analisis Sistem .....	23
4.2.1	Validasi Sistem .....	23
4.2.2	<i>Precision, Recall, dan F-Measure</i> .....	25
4.2.3	Evaluasi sistem.....	28
BAB 5	KESIMPULAN DAN SARAN .....	30
5.1	Kesimpulan.....	30
5.2	Saran.....	30
	DAFTAR PUSTAKA .....	31
	LAMPIRAN.....	33

## DAFTAR TABEL

TABEL 3.1. TABEL JUMLAH KATA BERSINONIM PADA SETIAP ABJAD .....	13
---	----

©UKDW

## DAFTAR GAMBAR

GAMBAR 2.1. TIGA TUGAS UTAMA INFORMATION RETRIEVAL:.....	7
GAMBAR 2.2. CONTOH DOKUMEN DALAM KORPUS KOMPAS_433.....	8
GAMBAR 2.3. CONTOH RDF .....	9
GAMBAR 2.4. CONTOH PENYIMPANAN TESAUROS DALAM SKOS MILIK HANDOKO (2015) .....	10
GAMBAR 3.1. TESAUROS DALAM FILE RDF MILIK HANDOKO (2015).....	12
GAMBAR 3.2. KATA-KATA YANG TIDAK MEMILIKI SINONIM .....	13
GAMBAR 3.3. CONTOH DOKUMEN DALAM KORPUS KOMPAS.....	14
GAMBAR 3.4. CONTOH TOPIK DALAM KORPUS KOMPAS.....	14
GAMBAR 3.5. CONTOH RELEVANT JUDGEMENT PADA KORPUS KOMPAS .....	15
GAMBAR 3.6. ARSITEKTUR SISTEM TEMU KEMBALI YANG DIKEMBANGKAN.....	17
GAMBAR 3.7. RANCANGAN DESAIN ANTARMUKA .....	17
GAMBAR 4.1. TAMPILAN ANTARMUKA SOLR SERVER UNTUK QUERY DOKUMEN.....	19
GAMBAR 4.2. TAMPILAN ANTARMUKA HALAMAN PENCARIAN .....	20
GAMBAR 4.3. BAGAN UTAMA ALUR IMPLEMENTASI SISTEM .....	20
GAMBAR 4.4. CONTOH FORMAT SYNONYMS.TXT PADA SOLR.....	21
GAMBAR 4.5. CONTOH HASIL PENGUBAHAN TESAUROS DARI RDF KE TEKS.....	21
GAMBAR 4.6. CONTOH PEMBENTUKAN BASIS DATA KATA MAJEMUK .....	22
GAMBAR 4.7. CONTOH KALIMAT .....	22
GAMBAR 4.8. URL PEMANGGILAN SISTEM .....	23
GAMBAR 4.9. GRAFIK INTERPOLASI PRECISION-RECALL MILIK HANDOKO (2015) PADA KORPUS KOMPAS 433.....	24
GAMBAR 4.10. GRAFIK INTERPOLASI PRECISION-RECALL MILIK PENULIS PADA KORPUS KOMPAS 433 .....	24
GAMBAR 4.11. HASIL F-MEASURE UNTUK TESAUROS BERBASIS SKOS DENGAN STEMMING DARI TOPIK 1 SAMPAI TOPIK 5 .....	26
GAMBAR 4.12. HASIL F-MEASURE UNTUK TESAUROS BERBASIS SKOS TANPA STEMMING DARI TOPIK 1 SAMPAI TOPIK 5 .....	27
GAMBAR 4.13. GRAFIK AVERAGE PRECISION UNTUK TESAUROS BERBASIS SKOS DENGAN STEMMING DAN TESAUROS BERBASIS SKOS TANPA STEMMING .....	28

## INTISARI

### Pengembangan Sistem Information Retrieval untuk Bahasa Indonesia Berbasis Web menggunakan Vector Space Model

Sistem temu kembali adalah sistem untuk memperoleh informasi dari kumpulan sumber informasi yang relevan dengan keinginan pengguna. Salah satu pengembang mesin pencarian yang bernama Solr menyediakan fitur untuk memberi *term expansion* berupa *synonym*. Fitur ini digunakan pada penelitian ini untuk memasukkan tesaurus. Tesaurus merupakan hubungan antar kata yang dikelompokkan menjadi satu kelompok. Tesaurus yang digunakan pada penelitian ini menggunakan hasil penelitian dari Himawan (2015) yang berbentuk file RDF.

Pada penelitian ini ditambahkan fitur kata majemuk untuk menggabungkan dua kata menjadi satu supaya pada saat *query matching* tidak terpisah. Untuk melihat tingkat keefektifitas saat *re-trieve* dokumen, penulis membandingkan sistem jika ditambah *stemming* dengan tidak ditambah *stemming*. Sistem yang dibangun pada penelitian ini dibentuk menjadi sebuah *search engine* yang berbasis *javascript*.

Sistem yang dibangun diuji *F-Measure* untuk 20 topik utama yang telah ditentukan dan menghasilkan nilai 0.3728 untuk sistem dengan *stemming* dan 0.3716 untuk sistem tanpa *stemming*. Setelah itu, sistem yang dibangun diuji kepada responden dengan 30 topik yang berbeda dan dihitung *Mean Average Precision*-nya yang bernilai 0.6225 untuk sistem dengan *stemming* dan 0.6279 untuk sistem tanpa *stemming*. Dapat disimpulkan bahwa sistem yang dibangun dengan *ad-hoc evaluation* lebih baik sistem dengan *stemming*, namun jika dicoba langsung kepada responden lebih baik sistem tanpa *stemming*.

Kata kunci: Solr, SKOS, Tesaurus, Search Engine, Information Retrieval

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Thesaurus* merupakan buku referensi yang berupa daftar kata dari sinonim kata. Tesaurus dapat membantu proses *information retrieval* karena fungsinya untuk memperkaya dokumen atau *query* pencarian. *Query* yang misalkan hanya satu kata jika ditambah tesaurus dapat menjadi beberapa kata. Hal ini dapat membantu saat proses *matching* antar dokumen dengan *query* yang dicari (Hariyanto, Anjasmara, & Budiansyah, 2016).

Salah satu *thesaurus* yang sering digunakan adalah WordNet. WordNet adalah basis data leksikal online yang dirancang untuk digunakan di bawah kendali program (Miller, 1992). WordNet yang menggunakan bahasa campuran bahasa Indonesia dan bahasa Melayu disebut WordNet bahasa. Diperlukan media basis data untuk menyimpan data tesaurus tersebut. Salah satu cara tersebut dengan cara disimpan dalam bentuk *Resource Description Framework* (RDF). RDF merupakan salah satu cara untuk memodelkan data untuk pertukaran data. *Simple Knowledge Organization System* (SKOS) menjadi representasi pengetahuan yang dipilih untuk penelitian ini.

Pada penelitian sebelumnya, Handoko (2015) telah melakukan penelitian tentang SKOS. Hasil dari penelitian tersebut adalah basis data tesaurus berbasis SKOS yang akan digunakan pada penelitian ini.

*Information Retrieval System* adalah aktifitas untuk memperoleh informasi yang dibutuhkan dari kumpulan sumber informasi. Sebagai contoh yang sering kita temui adalah *search engine*, *Question-Answer system*, *extraction system*, dan sebagainya. Untuk penelitian ini, *Information Retrieval System* yang akan dibangun adalah sebuah *search engine* yang menggunakan tesaurus berbasis SKOS.

Penelitian ini akan membangun sebuah *search engine* berbasis web yang menggunakan SKOS untuk basis data tesaurus bahasa Indonesia dan menguji

*precision*, *recall*, dan *F-measure* dari hasil pencarian pada sekumpulan data yang disiapkan.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang dijabarkan, masalah yang dapat dirumuskan adalah seberapa baik kinerja *search engine* berbasis web yang menggunakan *thesaurus* Bahasa Indonesia yang disimpan dalam bentuk RDF jika ditambah fitur kata majemuk dan dilakukan *stemming*?

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dijabarkan sebagai berikut:

1. Korpus yang digunakan sebagai data uji hanya berasal dari Kompas\_433 (Asian, 2007)
2. *Indexing* beserta pencarian dokumen dan evaluasi pencarian akan dilakukan oleh Solr dan *Trec\_Eval*
3. Bahasa pemrograman yang digunakan dalam pembangunan aplikasi adalah Javascript
4. *Stemming* kata akan menggunakan library jsastrawi (Librian & Kuku, 2017)
5. *Stopword* yang digunakan berasal dari penelitian Vega (2001) yang berisikan 556 kata
6. *Thesaurus* yang berbasis SKOS berasal dari penelitian Handoko (2015).

## 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun *search engine* yang *ready to use* dibangun menggunakan *thesaurus* Bahasa Indonesia yang disimpan dalam file RDF yang ditambah fitur kata majemuk. Sistem yang dibangun akan divalidasi dengan



hasil milik Handoko (2015) untuk mengetahui apakah sistem yang dibangun sudah memenuhi standar yang ada. Evaluasi sistem dilihat dari tingkat *Mean Average Precision* yang dihasilkan dari memasukkan *query* dari responden dan output yang dikeluarkan sistem.

## **1.5 Manfaat Penelitian**

Manfaat penelitian ini bagi peneliti adalah sebagai berikut:

1. Peneliti dapat mengembangkan sistem Information Retrieval Bahasa Indonesia yang lebih baik.
2. Sistem yang dikembangkan dapat digunakan dalam sistem lain.
3. User dapat menggunakan sistem Information Retrieval Bahasa Indonesia yang lebih baik.

## **1.6 Metode / Pendekatan**

Dalam penelitian ini, metode / pendekatan yang digunakan dibagi menjadi 5 tahap yaitu:

### **1.6.1 Tahap Studi Literatur**

Untuk dapat melakukan penelitian ini diperlukan sebuah studi pustaka untuk mengumpulkan informasi tentang Simple Knowledge Organization System, query expansion, REST, dan evaluasi information retrieval menggunakan TREC Eval. Studi pustaka ini dilakukan dengan membaca artikel jurnal dan buku yang memberikan pemahaman mengenai teori dan implementasi kelima bagian ini.

### **1.6.2 Tahap pembangunan basis data kata majemuk**

Penulis mengumpulkan kata-kata di database tesaurus yang diperoleh dari Handoko (2015) seperti: ibu kota, naik daun, anak kandung, dan sebagainya.

Kumpulan kata tersebut digunakan untuk mengecek kata-kata pada query dan dokumen corpus yang dapat berupa kata majemuk.

Pada tahap ini, kata kata yang berupa kata majemuk akan digabungkan menggunakan garis bawah. Sebagai contoh, “ibu kota” menjadi “ibu\_kota”, “naik daun” menjadi “naik\_daun”, “anak kandung” menjadi “anak\_kandung”.

### 1.6.3 Tahap pembangunan sistem

Penulis merancang aplikasi search engine yang berbasis web dengan menggunakan tesaurus berbasis SKOS. Pada tahap ini, penulis akan menggunakan *library* eksternal untuk proses stemming. Stemming adalah cara untuk mengambil kata dasar dari kata yang berimbuhan. *Library* tersebut diambil dari repositori Github dengan nama sastrawi (Librian & Kukuh, 2017).

### 1.6.4 Evaluasi

Tahap ini adalah tahap dimana peneliti melakukan pengujian terhadap sistem yang dibangun. Evaluasi sistem dilihat dari tingkat interpolasi *precision-recall* yang dihasilkan dari memasukkan *query* dari *query relevant judgement* dan output yang dikeluarkan sistem untuk validasi dengan hasil *output* milik Handoko (2015). Setelah itu dilakukan evaluasi sistem jika sistem diuji langsung ke pengguna untuk dilihat performanya.

### 1.6.5 Final Report

Laporan berupa laporan penelitian yang akan dibungkus dalam bentuk laporan tugas akhir yaitu laporan skripsi.

## 1.7 Sistematika Penulisan

Dalam mempermudah penjelasan dan pembahasan mengenai penelitian ini, laporan skripsi terbagi menjadi 5 bab yaitu Bab I Pendahuluan Bab II Tinjauan

Pustaka, Bab III Perancangan Sistem, Bab IV Implementasi dan Analisis Sistem dan Bab V Kesimpulan dan Saran.

BAB I, bab ini berisi tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode / pendekatan penelitian dan sistematika penulisan. Bab ini menjelaskan mengenai latar belakang diadakan penelitian ini dan gambaran secara garis besar penelitian serta manfaat bagi semua pihak.

BAB II, bab ini berisi tentang tinjauan pustaka dan landasan teori. Tinjauan pustaka menjelaskan mengenai penelitian sejenis yang menjadi referensi peneliti. Landasan teori berisi kumpulan teori dan definisi yang dikutip melalui berbagai buku dan jurnal.

BAB III, bab ini berisi pembahasan mengenai perancangan sistem, uraian mengenai data yang akan digunakan dan diterapkan pada sistem.

BAB IV, bab ini berisi implementasi dari perancangan sistem yang telah dibuat, uji coba sistem dan analisis sistem dari penelitian yang telah dilakukan secara terpadu.

BAB V, bab ini berisi kesimpulan yang merupakan hasil analisis dari penelitian yang telah dilakukan. Bab ini juga berisi saran yang diajukan untuk penelitian selanjutnya dan diharapkan dapat berguna untuk menghasilkan hasil penelitian yang lebih baik.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil implementasi dan analisis yang telah dibahas pada bab sebelumnya, penulis dapat menyimpulkan terkait penelitian ini, antara lain:

1. Sistem yang dibangun telah divalidasi dengan milik Handoko (2015). Untuk hasil yang tanpa *term expansion*, sistem yang dibangun sudah menghasilkan hasil yang menyerupai dengan milik Handoko (2015). Akan tetapi, hasil yang dibangun dengan *term expansion* menghasilkan hasil yang lebih baik daripada milik Handoko (2015).
2. Sistem yang dibangun telah memenuhi rancangan yang dibangun dengan urutan fitur-fitur yang akan diimplementasi, seperti tokenisasi, pengubahan kata majemuk, *term expansion* menggunakan tesaurus, dan *stemming*. Dari hasil uji coba para responden, sistem yang dibangun dengan tesaurus berbasis SKOS tanpa *stemming* menjadi pilihan para responden dengan *Mean Average Precision* 0.6279.

#### 5.2 Saran

Saran yang diberikan penulis untuk mengembangkan sistem ini adalah sebagai berikut:

1. Penambahan fitur pada saat *query* seperti penggunaan tanda petik (“...”) untuk mengelompokkan kata-kata supaya saat pemrosesan tidak terpisah, tanda tambah (+) untuk pencarian dokumen yang wajib memiliki kata yang dicari, dan tanda kurang (-) untuk pencarian dokumen yang wajib tidak memiliki kata yang dicari.
2. Penambahan fitur akronim untuk menjabar kata-kata akronim supaya *term matching* lebih tinggi.

## DAFTAR PUSTAKA

- Asian, J. (2007). *Effective Techniques for Indonesian Text Retrieval*. Thesis of Doctoral. RMIT University.
- Chinchor, N. (1992). MUC-4 Evaluation Metrics. *Proceedings of the Fourth Message Understanding Conference*, (pp. 22-29). doi:10.3115/1072064.1072067
- Croft, W., Metzler, D., & Strohman, T. (2010). *Search Engines*. Boston: AddisonWesley Professional.
- Handoko, M. H. (2015). *Pembangunan Tesaurus Bahasa Indonesia Menggunakan Simple Knowledge Organization System*. Undergraduate thesis. Universitas Kristen Duta Wacana.
- Hariyanto, Anjasmara, G., & Budiansyah, N. (2016). *Temu Kembali Dokumen Berbahasa Indonesia dengan Memperkaya Query menggunakan Persamaan Kata pada Tesaurus*. Institut Teknologi Sepuluh Nopember. doi:10.13140/RG.2.2.21986.12485
- Kesuma, H. W. (2016). *Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang*. Undergraduate thesis. Universitas Negeri Semarang.
- Lestari, K. (2013). *Query Expansion pada Sistem Temu Kembali Informasi dengan Model Ruang Vektor*. Undergraduate thesis. Universitas Islam Negeri Sultan Syarif Kasim.
- Librian, A., & Kukuh, R. (2017, Mei 4). *Github - sastrawi/sastrawi: High quality stemmer library for Indonesian Language (Bahasa)*. Retrieved from Github: <https://github.com/sastrawi/sastrawi>
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Miles, A., & Bechhofer, S. (2011, Agustus 6). *SKOS Simple Knowledge Organization System Namespace Document - HTML Variant, 18 August 2009 Recommendation Editon*. Retrieved from World Wide Web Consortium (W3C): <https://www.w3.org/2009/08/skos-reference/skos.html>
- Miller, G. A. (1992). WORDNET: A Lexical Database for English. *Commun. ACM*, 38, 39-41.
- Pardede, J., Barmawi, M., & Pramono, W. (2013). Implementasi Metode Generalized Vector Space Model pada Aplikasi Information Retrieval. *Jurnal Informatika*, 4(1), 35.

- Polettini, N. (2004). The Vector Space Model in Information Retrieval - Term Weighting Problem.
- Roberts, D., & Johnson, R. (1996). Evolving Frameworks. *Pattern languages of program design*, 3.
- Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor mater*, 1(5), 1-5.
- Siregar, R., Sinaga, F., & Arianto, R. (2017). Aplikasi Penentuan Dosen Pengui Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Journal of Computer Science and Information Systems*, 171-186.
- Stylos, J., & Myers, B. (2007). Mapping the space of API design decisions. *Visual Languages and Human-Centric Computing* (pp. 50-60). IEEE.
- Vega, V. (2001). *Information Retrieval for the Indonesian Language*. M.S. thesis. National University of Singapore.
- Virginia, G., & Nguyen, H. S. (2015). A Semantic Text Retrieval for Indonesian Using Tolerance Rough Sets Models. *LNCS Transactions on Rough Sets*, 8988, 138-224. doi:10.1007/978-3-662-47815-8\_9
- World Wide Web Consortium. (2014, Februari 25). *RDF - Semantic Web Standards*. Retrieved from Semantic Web Standards: <https://www.w3.org/RDF/>