

**PENGEMBANGAN SISTEM TEMU KEMBALI BERBASIS  
WEB MENGGUNAKAN ALGORITMA TOLERANCE VALUE  
GENERATOR**

Skripsi



Diajukan oleh:

**REXY DWY BRUNDY**

**71150004**

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2019

**PENGEMBANGAN SISTEM TEMU KEMBALI BERBASIS  
WEB MENGGUNAKAN ALGORITMA TOLERANCE VALUE  
GENERATOR**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika

Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:

**REXY DWY BRUNDY**

**71150004**

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA

2019

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS  
SECARA ONLINE UNIVERSITAS KRISTEN DUTA WACANA**

Saya yang bertanda tangan di bawah ini:

NIM : 71150004

Nama : Rexy Dwy Brundy

Prodi / Fakultas : Informatika / Teknologi Informasi

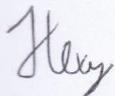
Judul Tugas Akhir : Pengembangan Sistem Temu Kembali Berbasis  
Web Menggunakan Algoritma Tolerance Value  
Generator

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty-free Right*)** serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Noneksklusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk database, merawat, dan memublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 26 Juni 2019

Yang menyatakan,



Rexy Dwy Brundy

71150004

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **PENGEMBANGAN SISTEM TEMU KEMBALI BERBASIS WEB MENGUNAKAN ALGORITMA TOLERANCE VALUE GENERATOR**

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, 16 Juni 2019



REXY DWY BRUNDY  
71150004

## HALAMAN PERSETUJUAN

Judul Skripsi : PENGEMBANGAN SISTEM TEMU KEMBALI  
BERBASIS WEB MENGGUNAKAN ALGORITMA  
TOLERANCE VALUE GENERATOR

Nama Mahasiswa : REXY DWY BRUNDY

N I M : 71150004

Matakuliah : Skripsi (Tugas Akhir)

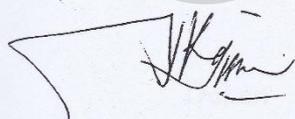
Kode : TIW276

Semester : Genap

Tahun Akademik : 2018/2019

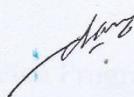
Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 23 Mei 2019

Dosen Pembimbing I



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

Dosen Pembimbing II



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

## HALAMAN PENGESAHAN

### PENGEMBANGAN SISTEM TEMU KEMBALI BERBASIS WEB MENGUNAKAN ALGORITMA TOLERANCE VALUE GENERATOR

Oleh: REXY DWY BRUNDY / 71150004

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

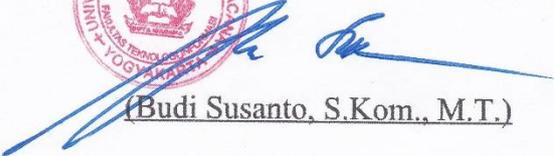
Yogyakarta, 16 Juni 2019  
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Danny Sebastian, S.Kom., M.M., M.T.
3. Lucia Dwi Krisnawati, Dr. Phil.
4. Hendro Setiadi, M.Eng



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 Temu Kembali Berbasis Web Menggunakan Algoritma Tolerance Value Generator”. 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. Teddy Brundy dan Purnamawati selaku orang tua Penulis yang selalu mendukung dan membantu dalam berbagai bentuk agar Penulis dapat menyelesaikan tugas akhir dan studi
2. Bapak Budi Susanto, S. Kom., M.T. selaku dekan Fakultas Teknologi Informasi.
3. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku Ketua Program Studi sekaligus dosen pembimbing Penulis yang tidak lelah dalam memberikan bimbingan dan semangat pada proses pengerjaan tugas akhir Penulis.
4. Bapak Danny Sebastian, S.Kom., M.M., M.T. selaku dosen pembimbing kedua yang selalu memotivasi Penulis dalam mengerjakan tugas akhir dan memberikan solusi permasalahan.
5. Teman-teman kuliah Penulis yang memberi semangat dan menjadi tempat diskusi bagi Penulis
6. Teman-teman ACC yang menjadi tempat bertukar ide bagi 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, 26 Juni 2019

Peneliti

## INTISARI

### Pengembangan Sistem Temu Kembali Berbasis Web Menggunakan Algoritma Tolerance Value Generator

Sistem temu kembali adalah sistem untuk memperoleh informasi dari kumpulan sumber informasi yang relevan dengan keinginan pengguna. Sistem yang dikembangkan menggunakan mesin pencari solr dilengkapi dengan tesaurus. Tesaurus adalah kumpulan kata yang dikelompokkan menjadi satu kelompok berdasarkan hubungan antar kata. Tesaurus yang digunakan pada penelitian ini berbasis pada *Tolerance Rough Set Model* yang dihasilkan melalui algoritma *Tolerance Value Generator*.

Algoritma *Tolerance Value Generator* adalah algoritma yang menghasilkan nilai *theta* yang digunakan sebagai syarat untuk membentuk tesaurus. Pada penelitian ini, nilai *theta* yang dihasilkan bernilai 20 berbeda dengan nilai *theta* yang dihasilkan oleh penelitian lain yang menggunakan metode dan objek yang sama. Pada penelitian ini, nilai *Mean Average Precision* untuk 20 topik utama yang telah ditentukan menghasilkan nilai 0,736095 dan nilai *Mean Average Precision* untuk 30 topik yang diberikan responden melalui API yang telah diimplementasikan menghasilkan nilai 0,621370386. Dapat disimpulkan bahwa pengimplementasian API tesaurus berbasis TRSM dapat mengalami penurunan apabila kata yang digunakan sebagai *query* merupakan kata yang tidak termasuk tesaurus sehingga pengeluaran dokumen tidak relevan. Kekurangan pada sistem ini adalah lama waktu *runtime* untuk pembentukan tesaurus yang besar disebabkan bahasa pemrograman yang digunakan untuk membangun sistem tidak mendukung GPU dan nilai *theta* yang dihasilkan belum optimal disebabkan modifikasi algoritma yang dilakukan peneliti untuk mengurangi waktu *runtime*.

Kata kunci: Solr, Tolerance Value Generator, Information Retrieval, API

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE UNIVERSITAS KRISTEN DUTA WACANA.....	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
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.....	3
1.5. Manfaat Penelitian.....	3
1.6. Metodologi Penelitian.....	3
1.6.1. Studi Literatur.....	3
1.6.2. <i>System Development</i> .....	4
1.6.3. <i>Data Entry</i> .....	4

1.6.4.	<i>Evaluation</i> .....	4
1.7.	Sistematika Penulisan .....	4
BAB 2	TINJAUAN PUSTAKA .....	6
2.1.	Tinjauan Pustaka .....	6
2.2.	Landasan Teori .....	8
2.2.1.	<i>Information Retrieval</i> .....	8
2.2.2.	<i>Preprocessing Text</i> .....	8
2.2.3.	<i>Rough Set Theory</i> .....	9
2.2.4.	<i>Tolerance Rough Set Models</i> .....	9
2.2.5.	<i>Corpus</i> .....	13
2.2.6.	<i>Solr</i> .....	13
2.2.7.	<i>Singular Value Decomposition</i> .....	14
2.2.8.	<i>Application Programming Interface</i> .....	14
2.2.9.	Metode Evaluasi <i>Precision, Recall, dan F-measure</i> .....	15
2.2.10.	Metode Evaluasi <i>Mean Average Precision</i> .....	16
BAB 3	PERANCANGAN SISTEM .....	17
3.1.	Analisis Kebutuhan Sistem .....	17
3.1.1.	Kebutuhan Perangkat Keras .....	17
3.1.2.	Kebutuhan Perangkat Lunak .....	17
3.1.3.	Kebutuhan Teknologi dan <i>Library</i> .....	17
3.2.	Blok Diagram Sistem .....	18
3.3.	Perancangan Desain Antarmuka Sistem .....	20
3.4.	Perancangan Evaluasi Sistem .....	21
BAB 4	IMPLEMENTASI DAN ANALISIS SISTEM .....	23
4.1.	Implementasi Sistem .....	23

4.1.1.	Tampilan Antarmuka .....	23
4.1.2.	Implementasi Sistem .....	25
4.1.3.	Fase Pemrosesan Teks.....	25
4.1.4.	Fase Pembuatan Tesaurus .....	26
4.1.5.	Fase <i>Tolerance Rough Set Model</i> .....	26
4.1.6.	Fase SVD .....	27
4.1.7.	Fase <i>Tolerance Value Generator</i> .....	27
4.1.8.	Fase Implementasi API .....	27
4.2.	Analisis Sistem .....	28
4.2.1.	Pemodelan <i>Tolerance Rough Set Model</i> dengan NodeJS .....	28
4.2.2.	<i>Mean Average Precision</i> .....	29
4.2.3.	<i>Precision, Recall, dan F-Measure</i> .....	30
4.2.4.	Evaluasi API .....	32
BAB 5	KESIMPULAN DAN SARAN .....	35
5.1.	Kesimpulan.....	35
5.2.	Saran.....	36
DAFTAR PUSTAKA	.....	37
LAMPIRAN	.....	38

## DAFTAR TABEL

TABEL 4.1 TABEL TOPIK RESPONDEN YANG TIDAK SESUAI DENGAN TOPIK PADA KOMPAS_433 .....	33
TABEL 4.2 TABEL NILAI <i>MEAN AVERAGE PRECISION</i> UNTUK SETIAP TESAURUS .....	33

©UKDWN

## DAFTAR GAMBAR

GAMBAR 2.1 HASIL <i>TOLERANCE VALUE</i> : .....	6
GAMBAR 2.2. TIGA TUGAS UTAMA <i>INFORMATION RETRIEVAL</i> : .....	8
GAMBAR 2.3. IDE DASAR <i>ROUGH SET THEORY</i> .....	9
GAMBAR 2.4. ALGORITMA <i>TOLERANCE VALUE GENERATOR</i> .....	10
GAMBAR 2.5. PROSES PEMBENTUKAN MATRIKS <i>TOLERANCE</i> YANG BERAWAL DARI MATRIKS TFIDF .....	11
GAMBAR 3.1. ARSITEKTUR SISTEM TEMU KEMBALI YANG DIKEMBANGKAN.....	18
GAMBAR 3.2. ALUR IMPLEMENTASI ALGORITMA <i>TOLERANCE VALUE</i> <i>GENERATOR</i> .....	19
GAMBAR 3.3. ALUR IMPLEMENTASI API.....	20
GAMBAR 3.4. RANCANGAN DESAIN ANTARMUKA <i>WEB</i> .....	20
GAMBAR 3.5. RANCANGAN DESAIN ANTARMUKA <i>MOBILE</i> .....	21
GAMBAR 4.1. TAMPILAN ANTARMUKA SOLR <i>SERVER</i> UNTUK <i>QUERY</i> DOKUMEN .....	23
GAMBAR 4.2. TAMPILAN ANTARMUKA HALAMAN PENCARIAN <i>WEB</i>	24
GAMBAR 4.3. TAMPILAN ANTARMUKA HALAMAN PENCARIAN <i>MOBILE</i> .....	24
GAMBAR 4.4. FORMAT JSON YANG DIKEMBALIKAN OLEH SISTEM ..	28
GAMBAR 4.5 KURVA <i>MEAN AVERAGE PRECISION</i> .....	29
GAMBAR 4.6 HASIL <i>F-MEASURE</i> TOPIK 1 SAMPAI TOPIK 5.....	31
GAMBAR 4.7 GRAFIK <i>AVERAGE PRECISION</i> UNTUK TESAUROS <i>TOLERANCE VALUE GENERATOR</i> , SKOS <i>STEM</i> , SKOS <i>NOSTEM</i> .....	32

## INTISARI

### Pengembangan Sistem Temu Kembali Berbasis Web Menggunakan Algoritma Tolerance Value Generator

Sistem temu kembali adalah sistem untuk memperoleh informasi dari kumpulan sumber informasi yang relevan dengan keinginan pengguna. Sistem yang dikembangkan menggunakan mesin pencari solr dilengkapi dengan tesaurus. Tesaurus adalah kumpulan kata yang dikelompokkan menjadi satu kelompok berdasarkan hubungan antar kata. Tesaurus yang digunakan pada penelitian ini berbasis pada *Tolerance Rough Set Model* yang dihasilkan melalui algoritma *Tolerance Value Generator*.

Algoritma *Tolerance Value Generator* adalah algoritma yang menghasilkan nilai *theta* yang digunakan sebagai syarat untuk membentuk tesaurus. Pada penelitian ini, nilai *theta* yang dihasilkan bernilai 20 berbeda dengan nilai *theta* yang dihasilkan oleh penelitian lain yang menggunakan metode dan objek yang sama. Pada penelitian ini, nilai *Mean Average Precision* untuk 20 topik utama yang telah ditentukan menghasilkan nilai 0,736095 dan nilai *Mean Average Precision* untuk 30 topik yang diberikan responden melalui API yang telah diimplementasikan menghasilkan nilai 0,621370386. Dapat disimpulkan bahwa pengimplementasian API tesaurus berbasis TRSM dapat mengalami penurunan apabila kata yang digunakan sebagai *query* merupakan kata yang tidak termasuk tesaurus sehingga pengeluaran dokumen tidak relevan. Kekurangan pada sistem ini adalah lama waktu *runtime* untuk pembentukan tesaurus yang besar disebabkan bahasa pemrograman yang digunakan untuk membangun sistem tidak mendukung GPU dan nilai *theta* yang dihasilkan belum optimal disebabkan modifikasi algoritma yang dilakukan peneliti untuk mengurangi waktu *runtime*.

Kata kunci: Solr, Tolerance Value Generator, Information Retrieval, API

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

*Information Retrieval* atau temu kembali adalah aktifitas untuk memperoleh informasi dari kumpulan sumber informasi yang relevan dengan keinginan pengguna. Sistem temu kembali dimulai dengan pengguna memasukkan *query* yang akan digunakan untuk mencari informasi pada kumpulan dokumen sumber yang relevan. Sistem temu kembali dapat dibedakan dari cara mencari informasi pada sumber yang disediakan baik secara keseluruhan maupun berdasarkan indeks (Belkin, 1993). Sistem temu kembali dapat mengembalikan informasi dalam berbagai bentuk seperti gambar, suara, teks, atau metadata. Dalam perkembangannya, sistem temu kembali dapat dikatakan sangat lambat di Indonesia dikarenakan pemrosesan teks masih dipandang sebelah mata. Kinerja sistem temu kembali dapat dikembangkan dengan memperkaya korpus yang digunakan dengan tesaurus.

Tesaurus adalah kumpulan kata yang dikelompokkan menjadi satu kelompok berdasarkan hubungan antar kata. Dalam sistem temu kembali, tesaurus sering digunakan sebagai *model domain* untuk melihat kesamaan atau relasi antar informasi dari sumber dan *query* yang dimasukkan pengguna (Hariyanto, Anjasmara, & Budiansyah, 2016). Wordnet Bahasa merupakan salah satu tesaurus dan mencakup gabungan Bahasa Indonesia dan Bahasa Melayu. Pada tahun 2016<sup>1</sup>, penelitian yang dilakukan oleh Kinasih menunjukkan bahwa penggunaan tesaurus terhadap sumber dokumen ICL menghasilkan performa yang lebih baik dibandingkan dengan tidak menggunakan ekspansi Wordnet yang dapat dilihat dari rata-rata *f-measure* baik pada dokumen atau *query*. Tesaurus dapat dibuat dengan cara salah satunya adalah algoritma *Tolerance Value Generator* yang mengenerate tesaurus secara otomatis dari korpus selanjutnya tesaurus akan digunakan dalam sistem temu kembali yang berbasis *Tolerance Rough Set Model*.

---

<sup>1</sup> Kinasih, 2016. "Evaluasi Wordnet Bahasa pada Information Retrieval System"

*Tolerance Rough Set Model* adalah model representasi dokumen berdasarkan perkiraan jarak antar vektor sehingga dokumen dapat ditentukan berdasarkan kumpulan kata yang memiliki toleransi dalam tingkat tertentu. Kelebihan dari tesaurus yang berbasis TRSM adalah sistem temu kembali akan lebih dinamis melihat dari topik tertentu dibandingkan dengan tesaurus yang dibangun pada umumnya akan bersifat statis (Virginia & Nguyen, 2015).

Melalui penelitian ini, peneliti ingin menguji kinerja sistem temu kembali yang menggunakan tesaurus berbasis TRSM menggunakan algoritma *Tolerance Value Generator* yang dibangun pada sebuah API berbasis *web* dan diuji menggunakan *mean average precision* dan *f-measure*.

## 1.2. Perumusan Masalah

Berdasarkan latar belakang yang dijabarkan, masalah yang dapat dirumuskan adalah:

1. Seberapa baik kinerja sistem temu kembali yang menggunakan tesaurus yang berbasis TRSM yang digenerasikan secara otomatis oleh algoritma *Tolerance Value Generator* yang diuji berdasarkan *mean average precision* dan *f-measure*?
2. Seberapa baik kinerja API yang diimplementasikan untuk menghasilkan informasi yang berasal dari dokumen Kompas\_433 yang diperkaya dengan tesaurus berbasis TRSM?

## 1.3. Batasan Masalah

Batasan masalah dalam penelitian ini dijabarkan sebagai berikut:

1. Korpus yang digunakan sebagai tesaurus hanya berasal dari Kompas\_433
2. Preprocessing teks mulai dari tokenisasi, *stemming*, hingga penghilangan *stop word* akan menggunakan library Solr
3. Pada penelitian akan ada antarmuka yang hanya digunakan untuk proses pengujian implementasi API dan bukan menjadi fokus pada penelitian ini

#### **1.4. Tujuan Penelitian**

Penelitian ini bertujuan untuk membuat API sistem temu kembali yang dibangun menggunakan tesaurus berbasis TRSM yang dihasilkan secara otomatis menggunakan algoritma *Tolerance Value Generator*. Sistem diuji berdasarkan efektivitas dan performa sistem dengan alat ukur *precision*, *recall*, *f-measure*, dan *mean average precision*.

#### **1.5. Manfaat Penelitian**

Manfaat penelitian ini bagi peneliti adalah sebagai berikut:

1. Peneliti dapat mengembangkan sistem Information Retrieval Bahasa Indonesia dengan akurat.
2. Sistem yang dikembangkan dapat digunakan pada sistem lain dengan hanya mengakses API sistem

Manfaat penelitian ini bagi pengguna adalah sebagai berikut:

3. Pengguna dapat menggunakan sistem temu kembali untuk diimplementasikan pada aplikasi sendiri dengan mengakses API

#### **1.6. Metodologi Penelitian**

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

##### **1.6.1. Studi Literatur**

Tahap ini adalah tahap pembelajaran dan mengumpulkan teori, buku, artikel-artikel dan penelitian serupa yang berkaitan dengan pembangunan sistem temu kembali yang berbasis *Tolerance Rough Set Model*, pembuatan API, beserta algoritma *Tolerance Value Generator* yang membantu peneliti dalam penggambaran masalah dan penyelesaian penelitian.

### 1.6.2. *System Development*

Tahap ini adalah tahap pembangunan dan pengembangan aplikasi yang akan dilakukan oleh peneliti. Tahap ini meliputi tahap persiapan, pengkodean, pengujian hasil sistem, perbaikan kesalahan pengkodean. Persiapan yang dilakukan dalam penelitian ini adalah pembentukan *tf-idf*, TRSM, pembuatan API, beserta *preprocessing* korpus yang digunakan.

### 1.6.3. *Data Entry*

Tahap ini adalah tahap pemasukan data berupa korpus Kompas\_433.

### 1.6.4. *Evaluation*

Tahap ini adalah tahap dimana peneliti melakukan pengujian terhadap API yang dibangun. Evaluasi sistem dilihat dari tingkat precision, recall, *mean average precision*, dan *F-measure* yang dihasilkan dari memasukkan *query* dari pengguna dan output yang dikeluarkan sistem.

## 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, permasalahan 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.

©UKDW

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan peneliti pada Subbab 4.2, dapat ditarik kesimpulan bahwa penerapan algoritma *Tolerance Value Generator* untuk penentuan nilai *theta* yang digunakan untuk pembuatan tesaurus dapat memberikan hasil yang cukup baik dengan nilai MAP bernilai 0,736095. Nilai rata-rata *f-measure* untuk 10 dokumen teratas pada topik utama bernilai 0,325879046. Dalam pengimplementasian API dengan topik yang diberikan responden, responden lebih memilih tesaurus berbasis SKOS *NOSTEM* dengan nilai MAP bernilai 0,627890028. Adanya penurunan nilai MAP dari 0,736095 menjadi 0,621370386 bisa disebabkan oleh nilai *theta* yang belum optimal.

Hal yang menjadi kekurangan pada penelitian ini adalah sebagai berikut:

1. Jumlah dokumen yang relevan untuk setiap topik berbeda sehingga nilai *precision* yang dikeluarkan apabila telah melebihi jumlah dokumen yang relevan maka nilai *precision* secara otomatis akan menurun.
2. Meskipun nilai *recall* yang dihasilkan untuk setiap topik utama mencapai nilai maksimal yaitu 1 tetapi nilai *precision* untuk beberapa topik menghasilkan nilai yang rendah. Hal ini menunjukkan bahwa sistem menghasilkan dokumen-dokumen yang tidak relevan dengan topik yang bersangkutan.
3. *Runtime* yang lama dalam proses penghasilan nilai *theta*. Hal ini disebabkan oleh bahasa pemrograman yang digunakan yaitu *Javascript* tidak mendukung pengalokasian GPU untuk proses algoritma *Tolerance Value Generator*.
4. Nilai *theta* yang belum optimal dikarenakan adanya modifikasi algoritma untuk mengurangi waktu *runtime*.
5. Tidak ada pembatasan topik dari responden menyebabkan nilai *precision* yang dihasilkan dapat bernilai 0. Hal ini akan menyebabkan turunnya

nilai *mean average precision* sehingga informasi yang dikeluarkan tidak akurat.

Kekurangan dan masalah pada penelitian ini diharapkan dapat digunakan untuk dijadikan penelitian selanjutnya.

## 5.2. Saran

Sistem yang telah dibuat masih dapat dikembangkan dari banyak sisi yang dapat membuat sistem menjadi lebih baik. Berdasarkan hal tersebut dan beberapa kekurangan yang terdapat pada penelitian ini, penulis memberikan beberapa saran yang dapat digunakan untuk mengembangkan sistem ini. Saran tersebut dapat dijabarkan sebagai berikut:

1. Bahasa pemrograman yang mendukung pengalokasian GPU seperti Python dapat digunakan untuk mengganti *Javascript* (NodeJS) dalam proses pembuatan tesaurus yang menggunakan algoritma *Tolerance Value Generator*. Hal ini diharapkan dapat mengurangi waktu pemrosesan tesaurus yang lama.
2. Pembatasan topik dari responden agar topik yang diberikan tetap dalam ruang lingkup korpus yang digunakan. Hal ini diharapkan dapat mengurangi nilai *precision* yang bernilai 0 yang menurunkan akurasi sistem.
3. Sistem yang dikembangkan menerima *query* yang diberikan pengguna sebagai korpus yang akan dihitung dan diberi bobot sehingga sistem dapat melakukan *unsupervised learning* berdasarkan *query* dari pengguna. Selain itu ada fitur yang dapat membantu pengguna seperti contohnya perbaikan *query* secara otomatis.

## DAFTAR PUSTAKA

- Agastya, I. M. (2018). Pengaruh Stemmer Bahasa Indonesia terhadap Performa Analisis Sentimen Terjemahan Ulasan Film. 7.
- Asian, J. (2007). Effective Techniques for Indonesian Text Retrieval.
- Belkin, N. J. (1993). Interaction with Texts: Information Retrieval as Information-Seeking Behaviour. *Information Retrieval*, 55-66.
- Croft, W., Metzler, D., & Strohman, T. (2010). *Search Engines*. Boston: AddisonWesley Professional.
- Dozan, D. J. (2019). Pengembangan Sistem Information Retrieval untuk Bahasa Indonesia Berbasis Web Menggunakan Vector Space Model. *Unpublished*, -.
- Grainger, T., & Potter, T. (2014). *Solr in Action*. New York: Manning Publications Co.
- Hariyanto, Anjasmara, G., & Budiansyah, N. (2016). Temu Kembali Dokumen Berbahasa Indonesia Dengan Memperkaya.
- Jacobson, D., Brail, G., & Woods, D. (2012). *APIs: A Strategy Guide*. United States of America: O'Reilly Media.
- Kinasih, L. H. (2016). *Evaluasi Wordnet Bahasa pada Information Retrieval System*. Yogyakarta: SINTA UKDW.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Meng, X.-J., Chen, Q.-C., & Wang, X.-L. (2009). A Tolerance Rough Set Based Semantic Clustering Method for Web Search Results. 11-12.
- Noviani, L., Suryani, A. A., & Kurniati, A. P. (2008). Pengklasifikasian Dokumen Berita Berbahasa Indonesia Menggunakan Latent Semantic Indexing (LSI) dan Support Vector Machine (SVM). 37-38.
- Vega, V. B. (2001). Information Retrieval for the Indonesian Language.
- Virginia, G., & Nguyen, H. S. (2015). A Semantic Text Retrieval for Indonesian Using Tolerance Rough Sets Models. 89-90.
- Wu, Y., Ding, Y., Wang, X., & Xu, J. (2010). On-line Hot Topic Recommendation Using Tolerance Rough Set Based Topic Clustering. *Journal of Computers*, 7-8.