

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

Skripsi



oleh

LUCIA HYANING PERGIWATI PUTRI KINASIH

71110022

**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2016

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

LUCIA HYANING PERGIWATI PUTRI KINASIH
71110022

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2016

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi 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, 4 Januari 2016



LUCIA HYANING PERGIWATI PUTRI KINASIH
71110022

HALAMAN PERSETUJUAN

Judul Skripsi : EVALUASI WORDNET BAHASA PADA
INFORMATION RETRIEVAL SYSTEM
Nama Mahasiswa : LUCIA HYANING PERGIWATI PUTRI KINASIH
N I M : 71110022
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2015/2016

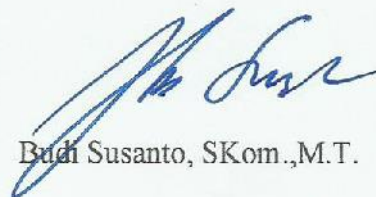
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 4 Januari 2016

Dosen Pembimbing I



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

Dosen Pembimbing II



Budi Susanto, SKom.,M.T.

HALAMAN PENGESAHAN

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

Oleh: LUCIA HYANING PERGIWATI PUTRI KINASIH / 71110022

Dipertahankan di depan Dewan Penguji 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 14 Desember 2015


Yogyakarta, 4 Januari 2016
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Budi Susanto, S.Kom., M.T.
3. Antonius Rachmat C., S.Kom., M.Cs.
4. Yuan Lukito, S.Kom., M.Cs.



 Dekan


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

Ketua Program Studi


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa yang telah menyertai dengan kasih sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Evaluasi Wordnet Bahasa pada Information Retrieval System” dengan baik.

Penulis menyadari bahwa penelitian ini tidak sempurna serta penulis menerima dengan terbuka saran dan masukan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Keluarga terkasih yang senantiasa memberi doa, dukungan, dan semangat kepada penulis dalam proses pengerjaan tugas akhir.
2. Gloria Virginia S.Kom., MAI, Ph.D. dan Budi Susanto S.Kom., M.T. yang telah mendukung, membimbing, dan memberi berbagai masukan bagi penulis terkait dengan pengerjaan tugas akhir.
3. Budi Nugroho yang juga memberikan semangat dan dukungan bagi penulis dalam pengerjaan tugas akhir ini.
4. Teman – teman seperjuangan, Amadea Kristina, Vincentius Billy, Leonardus Dani, dan Michael Himawan, yang sejak awal menjalani proses dan suka duka bersama.
5. Pihak – pihak yang tidak dapat disebutkan satu per satu yang telah membantu dan memberi dukungan baik fisik dan psikis dalam pengerjaan tugas akhir ini.

Akhir kata penulis memohon maaf apabila ada kesalahan dalam penyusunan laporan tugas akhir ini.

Yogyakarta, 27 November 2015

Penulis

INTISARI

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

Semantic similarity adalah sebuah masalah pada hubungan semantik. Hubungan semantik merupakan pendekatan untuk mengetahui bagaimana hubungan dua konsep dalam penggunaan dan relasinya. *Latent Semantic Indexing* (LSI) merupakan sebuah metode pendekatan *dimensional reduction* yang menghasilkan dokumen – dokumen yang relevan, sedangkan Wordnet menyediakan kumpulan data yang menyimpan informasi relasi semantik antar *synset* (*synonym set*).

Pada penelitian ini, penulis mencoba untuk meneliti tingkat performa pencarian dengan menggunakan ekspansi Wordnet Bahasa dibandingkan dengan pencarian dengan menggunakan metode LSI. Setiap jenis data akan dihitung jarak kemiripannya dengan kumpulan *query* yang diberikan menggunakan *cosine similarity*. Setiap hasil pencarian akan dievaluasi menggunakan nilai *recall*, *precision*, dan *f-measure* sebagai satuan ukur evaluasinya.

Pencarian dengan menggunakan LSI mampu mengambil 99 – 100% data relevan, tetapi tidak dapat menghasilkan performa sistem yang baik terlihat dari rata – rata nilai *f-measure* sebesar 0.1174. Hal ini disebabkan oleh rendahnya nilai *precision* karena banyaknya data yang tidak relevan yang juga ikut terambil saat pencarian. Ekspansi Wordnet Bahasa pada data Kompas mampu mengambil rata – rata 96% data relevan dan pada data ICL mampu mengambil rata – rata 49% data relevan. Performa sistem saat proses *retrieval* dengan melakukan ekspansi Wordnet Bahasa untuk kedua jenis data lebih tinggi dibandingkan dengan proses *retrieval* menggunakan LSI dengan rata – rata nilai *f-measure* sebesar 0.2732 untuk data Kompas dan 0.1320 untuk data ICL.

Kata kunci: *cosine similarity*, *f-measure*, *latent semantic indexing*, *precision*, *recall*, Wordnet Bahasa

DAFTAR ISI

HALAMAN JUDUL	i
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. Latar Belakang Masalah	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Metodologi Penelitian	3
1.5.1. Pengumpulan Data	3
1.5.2. Perancangan dan Pembangunan Sistem	3
1.5.3. Pengujian dan Analisis	3
1.6. Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	7
2.2.1. <i>Information Retrieval</i>	7
2.2.2. <i>Latent Semantic Indexing</i>	10
2.2.3. Wordnet dan Wordnet Bahasa	12
2.2.4. Corpus	13

BAB 3 ANALISIS DAN PERANCANGAN SISTEM	16
3.1. Kebutuhan Fungsional dan Non Fungsional Sistem	16
3.1.1. Kebutuhan Fungsional Sistem	16
3.1.2. Kebutuhan Non Fungsional Sistem	17
3.2. <i>Use Case Diagram</i>	17
3.3. Arsitektur Sistem	20
3.4. Diagram Alir	22
3.4.1. <i>Indexing</i> Dokumen	23
3.4.2. Wordnet Bahasa <i>Expansion</i>	24
3.5. Perancangan Basis Data	25
3.6. Perancangan Antar Muka	26
3.7. Perancangan Pengujian Sistem	27
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	28
4.1. Implementasi Antar Muka	28
4.1.1. Halaman Utama	28
4.1.2. Halaman Hasil Evaluasi	29
4.2. Implementasi Kode	30
4.2.1. <i>Document Reading</i> dan <i>Preprocessing</i>	30
4.2.2. <i>Indexing</i> Dokumen	33
4.2.3. <i>Scoring</i>	33
4.2.4. Penghitungan <i>Cosine Similarity</i>	36
4.2.5. Wordnet <i>Expansion</i>	36
4.3. Analisis dan Evaluasi Sistem	37
4.3.1. Evaluasi Manual	37
4.3.2. Evaluasi Sistem	40
BAB 5 KESIMPULAN DAN SARAN	51
5.1. Kesimpulan	51
5.2. Saran	52

DAFTAR PUSTAKA	53
LAMPIRAN	

©UKDW

DAFTAR TABEL

Tabel 3.1. Keterangan <i>Use Case</i> : Pembuatan Indeks Dokumen	18
Tabel 3.2. Keterangan <i>Use Case</i> : Implementasi LSI	18
Tabel 3.3. Keterangan <i>Use Case</i> : Implementasi Wordnet Bahasa	19
Tabel 3.4. Keterangan <i>Use Case</i> : Evaluasi Wordnet Bahasa	19
Tabel 3.5. Tabel Wordnet	25
Tabel 4.1. Tabel perhitungan TFIDF	38
Tabel 4.2. Tabel Vektor Dokumen Baru	38
Tabel 4.3. Tabel Vektor <i>Query</i> Baru	39
Tabel 4.4. Tabel Nilai <i>Cosine Similarity</i>	39
Tabel 4.5. Tabel Nilai <i>Cosine Similarity</i> dengan Implementasi Wordnet	39
Tabel 4.6. Tabel Perbandingan Nilai <i>Recall</i> , <i>Precision</i> , dan <i>F-measure</i>	40

DAFTAR GAMBAR

Gambar 2.1. Proses Tokenisasi	8
Gambar 2.2. Proses <i>Filtering</i>	8
Gambar 2.3. Proses <i>Stemming</i>	9
Gambar 2.4. Skema Reduksi Matriks SVD	11
Gambar 2.5. Tampilan Isi Wordnet Bahasa	13
Gambar 2.6. Tampilan Isi <i>File</i> Informasi Untuk Pengguna	15
Gambar 2.7. Tampilan Isi <i>File Relevance Judgments</i>	15
Gambar 3.1. <i>Use Case</i> Sistem Evaluasi Wordnet Bahasa	17
Gambar 3.2. Arsitektur Sistem Evaluasi Wordnet Bahasa	20
Gambar 3.3. Diagram Alir Sistem Evaluasi Wordnet Bahasa	22
Gambar 3.4. Diagram Alir <i>Indexing</i> Dokumen	23
Gambar 3.5. Diagram Alir <i>Preprocessing</i> Dokumen	24
Gambar 3.6. Diagram Alir Wordnet Bahasa <i>Expansion</i>	25
Gambar 3.7. Rancangan <i>Form</i> Utama Sistem Evaluasi Wordnet Bahasa	26
Gambar 3.8. Rancangan <i>Form</i> Hasil Sistem Evaluasi Wordnet Bahasa	27
Gambar 4.1. Halaman Utama Sistem Evaluasi Wordnet Bahasa	28
Gambar 4.2. Halaman Hasil Sistem Evaluasi Wordnet Bahasa	29
Gambar 4.3. <i>Pseudocode</i> Inisialisasi <i>Stopwords</i>	30
Gambar 4.4. <i>Pseudocode</i> Proses Membaca File TREC	31
Gambar 4.5. <i>Pseudocode</i> Proses Menghilangkan <i>Stopwords</i>	32
Gambar 4.6. <i>Pseudocode</i> Proses <i>Indexing File</i> TREC	32
Gambar 4.7. <i>Pseudocode</i> Proses Perhitungan TFIDF	33
Gambar 4.8. <i>Pseudocode</i> Proses Perhitungan Vektor Dokumen	34
Gambar 4.9. <i>Pseudocode</i> Proses Perhitungan Vektor <i>Query</i>	34
Gambar 4.10. <i>Pseudocode</i> Proses Perhitungan <i>Cosine Similarity</i>	35
Gambar 4.11. <i>Pseudocode</i> Proses Pencarian Wordnet_id	36
Gambar 4.12. <i>Pseudocode</i> Proses Pencarian <i>Lemma</i>	36
Gambar 4.13. Hasil Evaluasi Data Kompas 200	40

Gambar 4.14. Hasil Evaluasi Data Kompas 433	42
Gambar 4.15. Hasil Evaluasi Data ICL 100.....	42
Gambar 4.16. Hasil Evaluasi Data ICL 200.....	43
Gambar 4.17. Hasil Evaluasi Data ICL 300.....	44
Gambar 4.18. Grafik Perbandingan Nilai <i>Recall</i>	45
Gambar 4.19. Grafik Perbandingan Nilai <i>F-measure</i>	47

©UKDW

INTISARI

EVALUASI WORDNET BAHASA PADA INFORMATION RETRIEVAL SYSTEM

Semantic similarity adalah sebuah masalah pada hubungan semantik. Hubungan semantik merupakan pendekatan untuk mengetahui bagaimana hubungan dua konsep dalam penggunaan dan relasinya. *Latent Semantic Indexing* (LSI) merupakan sebuah metode pendekatan *dimensional reduction* yang menghasilkan dokumen – dokumen yang relevan, sedangkan Wordnet menyediakan kumpulan data yang menyimpan informasi relasi semantik antar *synset* (*synonym set*).

Pada penelitian ini, penulis mencoba untuk meneliti tingkat performa pencarian dengan menggunakan ekspansi Wordnet Bahasa dibandingkan dengan pencarian dengan menggunakan metode LSI. Setiap jenis data akan dihitung jarak kemiripannya dengan kumpulan *query* yang diberikan menggunakan *cosine similarity*. Setiap hasil pencarian akan dievaluasi menggunakan nilai *recall*, *precision*, dan *f-measure* sebagai satuan ukur evaluasinya.

Pencarian dengan menggunakan LSI mampu mengambil 99 – 100% data relevan, tetapi tidak dapat menghasilkan performa sistem yang baik terlihat dari rata – rata nilai *f-measure* sebesar 0.1174. Hal ini disebabkan oleh rendahnya nilai *precision* karena banyaknya data yang tidak relevan yang juga ikut terambil saat pencarian. Ekspansi Wordnet Bahasa pada data Kompas mampu mengambil rata – rata 96% data relevan dan pada data ICL mampu mengambil rata – rata 49% data relevan. Performa sistem saat proses *retrieval* dengan melakukan ekspansi Wordnet Bahasa untuk kedua jenis data lebih tinggi dibandingkan dengan proses *retrieval* menggunakan LSI dengan rata – rata nilai *f-measure* sebesar 0.2732 untuk data Kompas dan 0.1320 untuk data ICL.

Kata kunci: *cosine similarity*, *f-measure*, *latent semantic indexing*, *precision*, *recall*, Wordnet Bahasa

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Information Retrieval menyediakan sebuah *task* yang membahas mengenai *semantic similarity*. *Semantic similarity* adalah sebuah masalah pada hubungan semantik. Hubungan semantik merupakan pendekatan untuk mengetahui bagaimana hubungan dua konsep dalam penggunaan dan relasinya. Pendekatan pengukuran kesamaan semantik dapat dilakukan dengan berbagai pendekatan, seperti dengan menggunakan *Latent Semantic Indexing* atau Wordnet.

Latent Semantic Indexing atau yang biasa disingkat dengan LSI merupakan sebuah metode pendekatan *dimensional reduction* yang menghasilkan dokumen – dokumen yang relevan dalam pembentukan konsep similaritasnya. Kelebihannya ialah dapat menemukan relevansi antar data dan menemukan masalah semantik dalam *query* sehingga dokumen – dokumen yang tidak mengandung *term query* pun dapat dimunculkan karena memiliki hubungan keterkaitan makna dengan *query* yang diberikan.

Wordnet merupakan kumpulan data yang menyimpan informasi relasi semantik antar *synset* (sinonim set). *Synset* sendiri digunakan untuk mengelompokkan kata yang memiliki kesamaan makna. Oleh sebab itu diharapkan bahwa penggunaan Wordnet dapat menghindari makna ganda. Biasanya Wordnet dibuat dari kombinasi kamus dan juga Tesaurus. Saat ini perkembangan Wordnet di dunia semakin luas. Pada tahun 2011, tim dari NTU berhasil membuat Wordnet Bahasa yang menggunakan gabungan dari Bahasa Melayu dan Bahasa Indonesia. Tugas akhir ini akan mencoba mengevaluasi performa Wordnet Bahasa jika digunakan dalam dokumen berbahasa Indonesia dibandingkan dengan performa *latent semantic indexing* dalam pembentukan konsep similaritas.

1.2. Rumusan Masalah

Dari latar belakang di atas masalah yang dapat dirumuskan adalah sebagai berikut:

- Seberapa baik tingkat performa tingkat performa *Information Retrieval System* dengan menggunakan ekspansi terhadap Wordnet Bahasa dibandingkan dengan *Information Retrieval System* berbasis *Latent Semantic Indexing* dalam mengembalikan dokumen?

1.3. Batasan Masalah

Penelitian akan dibatasi oleh beberapa hal berikut:

- Data yang digunakan adalah TREC Kompas (Asian, Williams, & Tahaghoghi, 2004) dan TREC ICL (Virginia & Nguyen, 2015).
- *Output* sistem berupa file.eval yang berupa file hasil evaluasi sistem.
- *Preprocessing* dan *indexing* akan dilakukan dengan menggunakan aplikasi *Lucene version 4.1.0*.
- *Stopwords* yang akan digunakan adalah *stopwords* Bahasa Indonesia.
- Fokus penelitian pada pembentukan konsep similaritas.
- Pembentukan konsep similaritas menggunakan metode LSI dan ekspansi dengan Wordnet Bahasa.
- Perhitungan *latent semantic indexing* dilakukan dengan menggunakan *library jama version 1.0.2*.
- Evaluasi akan dilakukan dengan bantuan *library trec_eval version 8.1*.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi Wordnet Bahasa yang dibangun secara kolaboratif oleh Francis Bond, Nuril Hirfana Mohamed Noor, Suerya Sapuan, Muhammad Zulhelmy Mohammad Rosman, Hammam Riza, Tan

Enya Kong, dan Ruli Manurang dalam *Information Retrieval task*. Evaluasi akan dilakukan dengan cara membandingkan tingkat performa hasil pencarian dengan menggunakan metode *latent semantic indexing* dengan hasil pencarian dengan menggunakan ekspansi Wordnet Bahasa.

1.5. Metodologi Penelitian

1.5.1. Pengumpulan Data

Pengumpulan data dimulai dengan studi pustaka yang dilakukan dengan mencari dan mempelajari sumber – sumber pustaka yang berkaitan dengan teori tentang *text mining, information retrieval task, metode Latent Semantic Indexing, query expansion, Wordnet Bahasa, dan teknik evaluasi Precision Recall*.

1.5.2. Perancangan dan pembangunan sistem

Sistem akan dibangun dengan berbasis *desktop*. Sebagai *tools* akan digunakan *tools Java* untuk membantu dalam pembuatan *User Interface*. *Preprocessing* akan menggunakan *apache Lucene (tokenisasi, filtering, indexing)*.

1.5.3. Pengujian dan Analisis

Sistem akan dievaluasi menggunakan *Precision, Recall, dan F-Measure* dengan melihat performa dari metode *Latent Semantic Indexing* dan Wordnet Bahasa.

1.6. Sistematika Penulisan

Laporan tugas akhir ini akan terdiri dari 5 (lima) bab, yaitu:

Bab 1 Pendahuluan, berisi tentang gambaran umum mengenai penelitian. Bab ini terdiri dari latar belakang masalah dari penelitian, rumusan masalah, batasan – batasan masalah, hipotesis, tujuan penelitian, metodologi penelitian, serta sistematika penulisan dari penelitian ini.

Bab 2 Tinjauan Pustaka Dan Landasan Teori, berisi bahasan penelitian dan berbagai referensi mengenai penelitian.

Bab 3 Analisis Dan Perancangan Sistem, berisi rancangan pembangunan sistem, arsitektur sistem, algoritma yang digunakan dalam pembuatan sistem, alur kerja sistem, serta pemenuhan kebutuhan akan hardware maupun software untuk mendukung penelitian yang akan dilakukan oleh penulis.

Bab 4 Implementasi Sistem dan Analisis Sistem, berisi uraian detail implementasi sistem dan analisis dari hasil implementasi sistem.

Bab 5 Kesimpulan Dan Saran, berisi kesimpulan dari hasil penelitian serta saran – saran untuk penelitian yang sejenis.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

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

- Metode *latent semantic indexing* mampu mengambil 99 – 100% data relevan, tetapi tidak dapat menghasilkan performa sistem yang baik terlihat dari rata – rata nilai *f-measure* sebesar 0.1174. Hal ini disebabkan oleh rendahnya nilai *precision* karena banyaknya data yang tidak relevan yang juga ikut terambil saat pencarian.
- Ekspansi Wordnet Bahasa pada data Kompas mampu mengambil rata – rata 96% data relevan dan pada data ICL mampu mengambil rata – rata 49% data relevan. Performa sistem saat proses *retrieval* dengan melakukan ekspansi Wordnet Bahasa untuk kedua jenis data lebih tinggi dibandingkan dengan proses *retrieval* menggunakan LSI dengan rata – rata nilai *f-measure* sebesar 0.2732 untuk data Kompas dan 0.1320 untuk data ICL.
- Penggunaan Bahasa Indonesia formal mampu meningkatkan perolehan data relevan. Hal ini dapat dilihat dari rata – rata nilai *recall* pada penggunaan Kompas-corpora sebesar 0.955. Sedangkan rata – rata nilai *recall* untuk penggunaan ICL-corpora hanya sebesar 0.5533.
- Pada data Kompas, implementasi Wordnet Bahasa tidak dapat meningkatkan performa sistem lebih baik dibandingkan dengan LSI maupun tanpa melakukan ekspansi. Hal ini dapat dilihat dari performa sistem paling tinggi ditunjukkan pada evaluasi tanpa melakukan ekspansi dengan rata – rata nilai *f-measure* sebesar 0.3118.
- Pada data ICL, implementasi Wordnet Bahasa pada kumpulan dokumen mampu meningkatkan performa sistem lebih baik dibandingkan dengan LSI dan tanpa melakukan ekspansi. Hal ini dapat dilihat dari performa sistem

paling tinggi ditunjukkan pada evaluasi dengan melakukan ekspansi Wordnet Bahasa terhadap dokumen maupun dokumen dan query dengan rata – rata nilai *f-measure* sebesar 0.1356.

- Implementasi Wordnet Bahasa pada kumpulan *query* tidak memberikan hasil signifikan pada tingkat performa sistem. Hal ini ditunjukkan dengan perolehan nilai *f-measure* yang sama antara evaluasi tanpa melakukan ekspansi serta evaluasi dengan melakukan ekspansi Wordnet Bahasa terhadap *query*. Begitu pula dengan perolehan nilai *f-measure* antara evaluasi dengan melakukan ekspansi Wordnet Bahasa terhadap dokumen maupun terhadap dokumen dan *query*.

5.2. Saran

Saran yang diajukan oleh penulis untuk perbaikan dan pengembangan sistem yang akan datang.

- Dapat digunakan dokumen yang banyak mengandung Bahasa Indonesia formal sehingga proses pencarian maupun ekspansi dapat dilakukan dengan lebih optimal.
- Dapat digunakan Wordnet khusus Bahasa Indonesia yang lebih lengkap dan mengandung lebih banyak keterkaitan makna agar ekspansi dapat dilakukan lebih optimal.
- Dapat dicari cara untuk mengurangi data tidak relevan ikut terambil saat implementasi *Latent Semantic Indexing*. Misalnya, komputasi LSI dapat mempertimbangkan konteks linguistik, seperti hanya kata-kata dengan konteks sinonim yang diambil, sehingga data yang terambil saat *retrieval* juga didasarkan pada konteks linguistik.
- Penelitian ini dapat digunakan sebagai dasar untuk penelitian lanjutan maupun penelitian serupa dengan konteks atau metode penelitian yang berbeda.

DAFTAR PUSTAKA

- Asian, J., Williams, H., & Tahaghoghi, S. (2004). A Testbed for Indonesian Text Retrieval. *Australasian Document Computing Symposium* (pp. 55-58). Melbourne: University of Melbourne, Department of Computer Science.
- Bond, F., Noor, M., Sapuan, S., & et all. (2014, December 26). *Wordnet Bahasa*. Retrieved from Wordnet Bahasa: <http://wn-msa.sourceforge.net/>
- Büttcher, S., Clarke, C., & Cormack, G. (2010). *Implementing and Evaluating Search Engine*. Cambridge: MIT Press.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Inforation Retrieval*. New York: Cambridge University Press.
- Noviani, L., Ardiyanti, A., & Prima, A. (2008). *Pengklasifikasian Dokumen Berita Berbahasa Indonesia Menggunakan Latent Semantic Indexing (LSI) dan Support Vector Machine (SVM)*. Bandung: Institut Teknologi Telkom.
- Sari, Y., & Puspaningrum, E. (2013). Pencarian Semantik Dokumen Berita Menggunakan Essential Dimension of Latent Semantic Indexing dengan Memakai Reduksi Fitur Document Frequency dan Information Gain Thresholding. *Seminar Nasional Teknologi Informasi dan Multimedia* (pp. 27-32). Yogyakarta: STMIK Amikom.
- Virginia, G., & Nguyen, H. (2015). A Semantic Text Retrieval for Indonesian Using Tolerance Rough Sets Models. *Transaction on Rough Sets XIX*, 138-224.
- Weiss, S., Indurkha, N., Zhang, T., & Damerau, F. (2005). *Text Mining - Predictive Methods for Analyzing Unstructured Information*. New York: Springer.
- Wicaksana, I. (2005). *Membandingkan Pendekatan Latent Semantic terhadap Wordnet untuk Semantic Similarity*. Jakarta: Universitas Gunadarma.