

**PEMBANGUNAN TESAURUS BAHASA INDONESIA
MENGUNAKAN SIMPLE KNOWLEDGE ORGANIZATION
SYSTEM**

Skripsi



oleh

MICHAEL HIMAWAN HANDOKO
71110053

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA 2015

PEMBANGUNAN TESAURUS BAHASA INDONESIA MENGUNAKAN SIMPLE KNOWLEDGE ORGANIZATION 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

MICHAEL HIMAWAN HANDOKO
71110053

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA 2015

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PEMBANGUNAN TESAURUS BAHASA INDONESIA MENGGUNAKAN SIMPLE KNOWLEDGE ORGANIZATION 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, 28 September 2015



MICHAEL HIMAWAN HANDOKO
71110053

HALAMAN PERSETUJUAN

Judul Skripsi : PEMBANGUNAN TESAURUS BAHASA
INDONESIA MENGGUNAKAN SIMPLE
KNOWLEDGE ORGANIZATION SYSTEM

Nama Mahasiswa : MICHAEL HIMAWAN HANDOKO

N I M : 71110053

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2015/2016

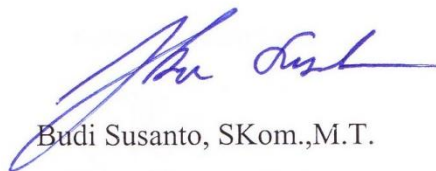
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 7 September 2015

Dosen Pembimbing I



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

Dosen Pembimbing II



Budi Susanto, SKom.,M.T.

HALAMAN PENGESAHAN

PEMBANGUNAN TESAURUS BAHASA INDONESIA MENGGUNAKAN SIMPLE KNOWLEDGE ORGANIZATION SYSTEM

Oleh: MICHAEL HIMAWAN HANDOKO / 71110053

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 18 September 2015

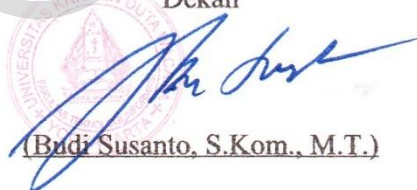
Yogyakarta, 28 September 2015
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Budi Susanto, S.Kom., M.T.
3. Willy Sudiarto Raharjo, S.Kom., M.Cs.
4. Junius Karel, M.T.

DUTA WACANA

Dekan


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

Ketua Program Studi


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji syukur kepada Tuhan yang Maha Esa atas berkat dan rahmat-Nya, skripsi yang berjudul “Pembangunan Tesaurs Bahasa Indonesia menggunakan *Simple Knowledge Organization System*” dapat terselesaikan. Penulis menyadari banyak masukan, saran, dan campur tangan dari berbagai pihak. Untuk itu penulis hendak berterima kasih kepada:

1. Gloria Virginia, S.Kom., MAI, Ph.D selaku dosen pembimbing pertama yang banyak membimbing penulis dalam menyelesaikan tugas akhir ini.
2. Budi Susanto, S. Kom., M. T. selaku dosen pembimbing kedua yang telah banyak membimbing penulis dalam menyelesaikan tugas akhir ini,
3. Keluarga terkasih yang selalu memberikan dukungan motivasi kepada penulis,
4. Rekan-rekan yang membantu dan memberikan masukan selama masa pengerjaan tugas akhir ini.
5. Untuk semua pihak yang tidak dapat penulis sebutkan satu-persatu di sini.

Yogyakarta, September 2015

Penulis

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa atas berkat dan rahmat-Nya penulis dapat menyelesaikan program dan laporan skripsi tentang Pembangunan Tesaurs Bahasa Indonesia menggunakan *Simple Knowledge Organization System*.

Penulisan skripsi ini diajukan untuk memenuhi salah satu syarat kelulusan dalam jenjang perkuliahan Strata I Universitas Kristen Duta Wacana. Dalam proses pengerjaan skripsi ini pastilah terdapat hambatan dan kesulitan. Untuk menghadapi hambatan dan kesulitan tersebut penulis mendapat banyak bantuan, masukan, dan saran dari berbagai pihak yang membantu terselesaikannya penulisan skripsi ini.

Tulisan ini pun tentunya tidak lepas dari kekurangan, baik dari aspek kualitas maupun aspek kuantitas dari materi penelitian yang disajikan. Oleh karena itu penulis membutuhkan kritik dan saran yang bersifat membangun untuk perbaikan tulisan ini.

Semoga hasil dari tulisan ini dapat bermanfaat untuk bidang penelitian dalam bidang teks mining dan semantik web, dan dapat digunakan sebagai referensi penelitian selanjutnya.

Yogyakarta, September 2015

Penulis

ABSTRAKSI

PEMBANGUNAN TESAURUS BAHASA INDONESIA MENGGUNAKAN *SIMPLE KNOWLEDGE ORGANIZATION SYSTEM*

Penelitian ini muncul karena masih belum adanya tesaurus bahasa Indonesia yang dibangun dengan *Simple Knowledge Organization System*. Padahal dalam bahasa Inggris sudah tersedia wordnet yang dibangun dengan berbasiskan database atau RDF, tetapi untuk bahasa Indonesia sendiri baru tersedia wordnet bahasa (campuran bahasa Indonesia dan Melayu). Karena belum adanya kamus yang tersimpan dalam bentuk SKOS maka penulis mengadakan penelitian ini, sehingga kamus yang dihasilkan dapat digunakan untuk penelitian lain di bidang teks atau semantik web.

Proses pembangunan SKOS dari tesaurus bahasa Indonesia ini, dimulai dari proses digitalisasi kamus tercetak. Setelah itu dilakukanlah penyesuaian pada hasil digitalisasi tersebut agar sistem dapat membaca dan mengubah kamus dalam bentuk SKOS. Setelah berbentuk SKOS, dilakukan pengecekan apakah data asli kamus yang diubah menjadi bentuk SKOS sama atau tidak, apabila tidak sama, maka dilakukan pengecekan ulang agar SKOS yang dibuat sama dengan data asli.

SKOS yang telah dibangun dievaluasi dengan melihat aspek viii ontology dan aspek aplikatif (sistem *information retrieval*). Dari aspek ontology ternyata ditemukan 2 aspek yang tidak terpenuhi yaitu aspek *consistency* dan *completeness*, sedangkan 1 aspek, yaitu aspek *conciseness* berhasil dipenuhi. Dalam pengujian pada aspek aplikatif, didapati bahwa pencarian dengan bantuan kamus menghasilkan nilai yang kurang baik dibandingkan pencarian tanpa bantuan kamus.

Kata Kunci: tesaurus bahasa Indonesia, SKOS, ontology, *information retrieval*

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	vii
ABSTRAKSI	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	3
1.5 Metodologi Penelitian	3
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori.....	9
2.2.1 Tesaurus	9
2.2.2 Semantik Web	10
2.2.3 <i>Resource Description Framework</i>	11
2.2.4 <i>Simple Knowledge Organization System</i>	12
2.2.5 <i>Corpus</i>	13
2.2.6 Ontologi	16
2.2.7 SPARQL	17
2.2.8 <i>Query Expansion</i>	18

2.2.9 <i>Interpolated Precision-Recall</i>	19
2.2.10 <i>F-Measure</i>	20
BAB III ANALISIS DAN PERANCANGAN SISTEM	21
3.1 Analisis Kebutuhan	21
3.1.1 Analisis Data	21
3.1.2 Spesifikasi Perangkat Keras	24
3.1.3 Spesifikasi Sistem	24
3.2 Pengolahan Tesaurs	25
3.3 Perancangan Sistem	32
3.3.1 Use Case Sistem	32
3.3.2 Arsitektur Sistem	34
3.3.3 Diagram Alir (Flowchart) Sistem	35
3.3.4 Proses Pencarian Sinonim kata	40
3.3.5 Perancangan Basis Pengetahuan	41
3.3.6 Perancangan Antarmuka Sistem	41
3.3.7 Perancangan Pengujian Sistem	45
BAB IV IMPLEMENTASI DAN ANALISIS.....	48
4.1 Implementasi Sistem	48
4.1.1 Antarmuka	48
4.1.2 Pembuatan Tesaurs Terformat	53
4.1.3 Pseudocode Sistem	55
4.2 Implementasi Basis Pengetahuan	61
4.3 Pemakaian Kamus.....	63
4.4 Analisis Sistem	63
4.3.1 Analisis ontology	63
4.3.2 Analisis performa	67
BAB V KESIMPULAN DAN SARAN.....	72
5.1 Kesimpulan	72
5.2 Saran	73
DAFTAR PUSTAKA	74

DAFTAR GAMBAR

Gambar 2.1. Contoh sebuah kata pada tesaurus yang memiliki beberapa label	10
Gambar 2.2. Contoh RDF	11
Gambar 2.3. Contoh SKOS dalam mengidentifikasi kata.....	12
Gambar 2.4. contoh dokumen dengan format sesuai kesepakatan TREC pada corpus Kompas.....	14
Gambar 2.5. contoh topik dengan format sesuai kesepakatan TREC pada corpus Kompas	15
Gambar 2.6. contoh relevance judgement pada corpus Kompas	15
Gambar 2.7. kategorisasi ontologi menurut Lasila dan McGuiness (2001).....	16
Gambar 2.8. contoh SPARQL query	18
Gambar 2.9. hasil kembalian SPARQL query pada gambar 2.6.....	18
Gambar 2.10. grafik precision-recall dengan interpolasi	19
Gambar 3.1. Hasil scan Tesaurus dengan scanner	21
Gambar 3.2. Contoh dokumen dalam corpus Kompas	22
Gambar 3.3. Contoh topik dalam corpus Kompas	23
Gambar 3.4. Contoh Relevance Judgement pada corpus Kompas untuk dokumen news10002-html.....	23
Gambar 3.5. dokumen news10015-html yang memiliki relevansi pada beberapa topik	24
Gambar 3.6. hasil pembacaan kamus untuk kata abad.....	27
Gambar 3.7. hasil pembacaan kata abad yang telah diubah.....	26
Gambar 3.8. hasil pembacaan untuk kata aba	26
Gambar 3.9. kata aba yang telah dirubah.....	26
Gambar 3.10. hasil pembacaan untuk kata abah dan abstrak.....	26
Gambar 3.11. hasil penyesuaian kata abah dan abstrak	27
Gambar 3.12. hasil pembacaan untuk kata abstrak	27
Gambar 3.13. hasil penambahan nomor pada kata abstrak.....	27
Gambar 3.14. hasil pembacaan kata afeksi	27

Gambar 3.15. hasil penambahan atribut pada kata afeksi	28
Gambar 3.16. hasil pembacaan kata agak	28
Gambar 3.17. hasil pengubahan kata agak	28
Gambar 3.18. hasil pembacaan kata agen	28
Gambar 3.19. hasil pengubahan kata agen	28
Gambar 3.20. hasil pembacaan kata adam	29
Gambar 3.21. hasil pengubahan kata adam	29
Gambar 3.22. hasil pembacaan kata agen	29
Gambar 3.23. hasil perubahan kata kantor cabang	29
Gambar 3.24. hasil pembacaan kata acu	29
Gambar 3.25. hasil bentukan kata acu dan mengacu	30
Gambar 3.26. hasil pembacaan kata pengabdian	30
Gambar 3.27. hasil penyesuaian kata pengabdian	30
Gambar 3.28. hasil pembacaan kata acap	30
Gambar 3.29. hasil penambahan tambahan kata pada akhiran kali	31
Gambar 3.30. hasil pembacaan kata berada	31
Gambar 3.31. hasil penambahan kata pada akhiran buku tangan	31
Gambar 3.32. hasil pembacaan kata caplak	31
Gambar 3.33. hasil perubahan kata dengan tanda garis miring	31
Gambar 3.34. Use case diagram	32
Gambar 3.35. arsitektur sistem pembuatan tesaurus dalam bentuk SKOS	34
Gambar 3.36. flowchart pembuatan file SKOS	35
Gambar 3.37. flowchart pembuatan file SKOS (lanjutan)	36
Gambar 3.38. Flowchart pengambilan data dari Sesame Server	37
Gambar 3.39. Flowchart sistem untuk mengindex dokumen	38
Gambar 3.40. Flowchart sistem untuk melakukan information retrieval	39
Gambar 3.41. Flowchart sistem untuk menggambar grafik precision-recall	40
Gambar 3.42. query yang diberikan untuk melakukan pencarian persamaan kata	41
Gambar 3.43. grafik representasi pengetahuan tesaurus (T-Box)	41
Gambar 3.44. Mock-up form pembentukan file SKOS	42
Gambar 3.45. Mock-up form untuk melakukan query	42

Gambar 3.46. Form untuk melakukan indexing dan pencarian dokumen dengan bantuan lucene.....	43
Gambar 3.47. Form yang menampilkan hasil pencarian berupa daftar dokumen 4dan scoringnya	43
Gambar 3.48. Form yang digunakan untuk memasukkan data hasil evaluasi TREC Eval	44
Gambar 3.49. Form yang menunjukkan perbandingan dari dua jenis pencarian...	44
Gambar 3.50. contoh evaluasi completeness pada kata bukat tesaurus bahasa Indonesia	46
Gambar 3.51. contoh evaluasi completeness pada kata cemar tesaurus bahasa indonesia	46
Gambar 3.52. contoh evaluasi conciseness pada tesaurus bahasa indonesia	47
Gambar 3.53. tampilan evaluasi sistem pencarian	47
Gambar 4.1. form pembuatan taksonomi	48
Gambar 4.2. form pencarian persamaan kata	49
Gambar 4.3. tampilan pencarian kata yang tidak mendapatkan persamaannya....	49
Gambar 4.4. form indexing dan information retrieval	50
Gambar 4.5. form hasil pencarian	51
Gambar 4.6. form perbandingan hasil pencarian	51
Gambar 4.7. form hasil evaluasi pencarian	52
Gambar 4.8. form evaluasi konversi kamus tercetak ke bentuk taksonomi.....	53
Gambar 4.9. hasil scan halaman kamus	53
Gambar 4.10. hasil pembacaan software Abbyy Fine Reader	54
Gambar 4.11. hasil formatting tesaurus terformat secara manual.....	54
Gambar 4.12. daftar kata yang tidak merujuk ke persamaan kata yang tidak ada.	55
Gambar 4.13. pseudocode untuk membuat file taksonomi berdasarkan file tesaurus terformat.....	55
Gambar 4.14. contoh perubahan kata ayan pada tesaurus terformat ke taksonomi	56
Gambar 4.15. contoh perubahan kata ayan pada tesaurus terformat ke taksonomi (lanjutan)	57

Gambar 4.16. pseudocode untuk mengambil daftar persamaan kata pada Sesame Server	57
Gambar 4.17. SPARQL query untuk mengambil daftar persamaan kata pada server	58
Gambar 4.18. pseudocode to store DOC in lucene index	59
Gambar 4.19. pseudocode pencarian pada Lucene	59
Gambar 4.20. perintah yang digunakan untuk mendapatkan hasil evaluasi TREC Eval	60
Gambar 4.21. pseudocode membuat tampilan evaluasi TREC Eval	61
Gambar 4.22. struktur class basis pengetahuan yang dibangun.....	61
Gambar 4.23. struktur class basis pengetahuan yang dibangun (lanjutan)	62
Gambar 4.24. individual yang dibuat berdasarkan class pada basis pengetahuan	62
Gambar 4.25. owl:ObjectProperty yang menghubungkan kata dengan atribut	62
Gambar 4.26. implementasi relasi antara kata dengan atribut dalam ontologi	63
Gambar 4.27. form evaluasi tesaurus	64
Gambar 4.28. kata aba yang memiliki persamaan abi	64
Gambar 4.29. kata abi yang tidak memiliki persamaan aba	64
Gambar 4.30. beberapa daftar kata yang tidak saling ada.....	65
Gambar 4.31. contoh konsep yang disengaja sama.....	66
Gambar 4.32. query untuk menghitung jumlah concept dalam Sesame Server	66
Gambar 4.33. hasil query Sesame Server	66
Gambar 4.34. hasil query Sesame Server	66
Gambar 4.35. gambar evaluasi performa oleh TREC Eval pada dataset 433 dokumen.....	67
Gambar 4.36. gambar evaluasi performa oleh TREC Eval pada dataset 1000 dokumen.....	67
Gambar 4.37. contoh pencarian dengan dokumen, query, dan daftar persamaan kata	68
Gambar 4.38. contoh hasil pencarian dengan ekspansi query	68
Gambar 4.39. contoh hasil pencarian dengan ekspansi dokumen.....	68
Gambar 4.40. grafik interpolasi precision-recall pada dataset 433 dokumen	70

Gambar 4.41. grafik interpolasi precision-recall pada dataset 1000 dokumen70

©UKDWN

DAFTAR TABEL

Tabel 3.1. Tabel <i>use case</i> membuat file SKOS.....	32
Tabel 3.2. Tabel <i>use case</i> untuk mencari persamaan kata	33
Tabel 3.3. Tabel <i>use case</i> untuk proses information retrieval.....	33
Tabel 3.4. Tabel <i>use case</i> membuat grafik precision-recall.....	34

©UKPDW

ABSTRAKSI

PEMBANGUNAN TESAURUS BAHASA INDONESIA MENGGUNAKAN *SIMPLE KNOWLEDGE ORGANIZATION SYSTEM*

Penelitian ini muncul karena masih belum adanya tesaurus bahasa Indonesia yang dibangun dengan *Simple Knowledge Organization System*. Padahal dalam bahasa Inggris sudah tersedia wordnet yang dibangun dengan berbasiskan database atau RDF, tetapi untuk bahasa Indonesia sendiri baru tersedia wordnet bahasa (campuran bahasa Indonesia dan Melayu). Karena belum adanya kamus yang tersimpan dalam bentuk SKOS maka penulis mengadakan penelitian ini, sehingga kamus yang dihasilkan dapat digunakan untuk penelitian lain di bidang teks atau semantik web.

Proses pembangunan SKOS dari tesaurus bahasa Indonesia ini, dimulai dari proses digitalisasi kamus tercetak. Setelah itu dilakukanlah penyesuaian pada hasil digitalisasi tersebut agar sistem dapat membaca dan mengubah kamus dalam bentuk SKOS. Setelah berbentuk SKOS, dilakukan pengecekan apakah data asli kamus yang diubah menjadi bentuk SKOS sama atau tidak, apabila tidak sama, maka dilakukan pengecekan ulang agar SKOS yang dibuat sama dengan data asli.

SKOS yang telah dibangun dievaluasi dengan melihat aspek viii ontology dan aspek aplikatif (sistem *information retrieval*). Dari aspek ontology ternyata ditemukan 2 aspek yang tidak terpenuhi yaitu aspek *consistency* dan *completeness*, sedangkan 1 aspek, yaitu aspek *conciseness* berhasil dipenuhi. Dalam pengujian pada aspek aplikatif, didapati bahwa pencarian dengan bantuan kamus menghasilkan nilai yang kurang baik dibandingkan pencarian tanpa bantuan kamus.

Kata Kunci: tesaurus bahasa Indonesia, SKOS, ontology, *information retrieval*

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Tesaurus merupakan buku referensi berupa daftar kata dengan sinonimnya, dengan adanya daftar kata dan sinonim ini tesaurus menjadi buku yang dapat membantu proses *information retrieval*. Tesaurus membantu memberikan pengayaan dalam pencarian informasi karena tesaurus menyimpan daftar persamaan kata. Daftar persamaan kata ini dapat digunakan untuk memperkaya dokumen atau *query* pencarian. Apabila dalam *query* diminta kata X dan ternyata tidak ada, dengan adanya tesaurus maka kata X tadi dapat dicari padanannya dan didapatkan hasil pencarian berupa padanan kata X.

Melihat kegunaan tadi pembangunan tesaurus digital tidak bisa dilakukan dengan sembarangan. Ada banyak tesaurus yang telah dicoba dibangun, salah satu tesaurus yang cukup sering digunakan adalah Wordnet. (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) Wordnet adalah sistem referensi leksikal online, yang desainnya terinspirasi oleh teori *psycholinguistic* dari *memory* leksikal manusia. Pembangunan tesaurus yang dilakukan oleh Wordnet memiliki beberapa model, misalnya data tesaurus disimpan dalam bentuk database, selain itu data tesaurus juga ada yang disimpan dalam bentuk Resource Description Framework. Agar tesaurus digital yang dibangun dapat dibaca langsung oleh mesin, maka diperlukan sebuah representasi pengetahuan yang mampu menyimpan satu kata beserta relasinya (padanan atau lawan kata) yang memungkinkan dilakukan pembacaan langsung. *Simple Knowledge Organization System* menjadi representasi pengetahuan yang dipilih penulis karena *Simple Knowledge Organization* merupakan sebuah model data yang umum untuk merepresentasikan sistem organisasi pengetahuan seperti tesaurus, skema klasifikasi, taksonomi, dan lain-lain.

Penelitian ini akan membangun sistem yang mampu menghasilkan file tesaurus dalam representasi pengetahuan *Simple Knowledge Organization System*

dan menguji tesaurus yang telah dihasilkan dengan melakukan pencarian pada sekumpulan data yang disiapkan. Dengan menguji tesaurus pada sistem pencarian diharapkan tesaurus ini dapat diaplikasikan langsung dalam sistem pencarian yang akan dibuat setelah penelitian ini.

1.2 Rumusan Masalah

Berdasar latar belakang masalah di atas, maka permasalahan yang akan diteliti, antara lain:

- Bagaimana membangun kamus dengan menggunakan *Simple Knowledge Organization System*?
- Bagaimana evaluasi dari kamus yang digunakan berdasarkan kriteria ontologi?
- Bagaimana evaluasi hasil pencarian antara data yang menggunakan kamus dan data normal (tidak menggunakan kamus)?

1.3 Batasan Masalah

Ruang lingkup permasalahan dalam penelitian ini dibatasi oleh hal-hal berikut:

1. Taksonomi kamus yang disimpan tidak melibatkan cara baca sebuah kata.
2. Kamus yang dirubah dalam bentuk SKOS, hanya memetakan kata dalam bentuk *Concept* yang sifatnya umum, sehingga semua kata yang memiliki relasi baik itu, bisa dimasukkan dalam *Concept*.
3. Pencarian persamaan kata dalam kamus hanya dilakukan 1 level.
4. Pencarian sinonim kata hanya memperhatikan kata yang dicari dan padanannya tanpa melihat makna sebuah kata dalam kalimat.
5. Konten dokumen uji yang digunakan berasal dari TREC Kompas.
6. Pencarian dokumen dan evaluasi pencarian akan dilakukan oleh Lucene dan *TREC Eval*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah membangun tesaurus bahasa Indonesia yang *ready to use* dengan menggunakan representasi pengetahuan *Simple Knowledge Organization System*. Tesaurus yang telah dibangun akan dilihat performanya dengan membandingkan hasil pada *information retrieval* di 2 dataset pengujian, selain itu tesaurus yang dibangun akan dievaluasi menurut 3 kriteria ontologi.

1.5 Metode Penelitian

1.5.1 Tahap Studi Pustaka

Untuk dapat melakukan penelitian ini diperlukan sebuah studi pustaka yang bertujuan untuk mengumpulkan dan mendapatkan keterangan dan pemahaman yang tepat mengenai apa itu *Simple Knowledge Organization System*, *Query Ekspansion*, dan evaluasi *information retrieval* menggunakan *TREC Eval*. Studi pustaka ini dilakukan dengan membaca artikel jurnal dan buku yang memberikan pemahaman mengenai teori dan implementasi ketiga bagian ini.

1.5.2 Tahap Pembangunan Tesaurus

A. Pengambilan data tesaurus

Tesaurus yang berbentuk buku dilakukan scanning perhalamannya lalu dibaca dengan menggunakan software Optical Character Recognition (OCR)¹. Hasil dari OCR tadi akan diperiksa terlebih dahulu kesesuaiannya dengan buku teks asli. Setelah itu maka dilakukan *formatting* pada teks hasil OCR, misalnya pemisah antar kata diberikan jarak(enter) secara manual agar dapat dibaca oleh sistem.

B. Pembangunan SKOS

Sebelum memberikan tag-tag SKOS, dilakukan terlebih dahulu analisa tesaurus, pada tahap analisa ini thesaurus dilihat dan diidentifikasi data item

¹ OCR merupakan teknologi yang dapat mengubah gambar yang terdapat teks di dalamnya menjadi file teks(.txt, .doc, .pdf) yang dapat diubah-ubah (Optical Character Recognition (OCR) – How it works , 2012)

apa saja yang menjadi karakter dari sebuah tesaurus bahasa indonesia. Setelah menganalisa tesaurus tersebut, tahap yang kedua adalah mapping data item. Pada tahap ini data item yang pada tahap pertama telah diidentifikasi dibuat *property/classnya* dalam notasi SKOS. Setelah semua data item memiliki notasi SKOS, tahap yang ketiga adalah pembangunan program konversi. Program ini akan membaca thesaurus yang telah dihasilkan dari OCR dan diberikan tag-tag SKOS sesuai identifikasi dan mapping yang telah dilakukan. (Assem, Malaise, Miles, & Schreiber, 2004)

1.5.3. Tahap Pembangunan Sistem

A. Perancangan awal

TREC dokumen kompas yang akan membantu pengujian kamus yang telah dibangun dilakukan preprocessing terlebih dahulu. Preprocessing ini dilakukan untuk membuat 2 jenis dokumen uji, yaitu dokumen yang tidak diperkaya dengan kamus dan dokumen yang diperkaya dengan kamus. Kemudian kedua jenis dokumen tadi diindex dengan bantuan Lucene.

Dokumen-dokumen yang telah siap (telah terindex di Lucene) diberikan dua jenis query, yaitu query langsung dan query yang telah diekspansi. Lucene akan memberikan urutan dokumen yang dianggap memenuhi query yang diberikan.

B. Query Ekspansi

Semua query yang akan dimasukkan dalam Lucene dilakukan ekspansi. Ekspansi tersebut dilakukan dengan bantuan kamus yang telah dibuat sebelumnya. Melalui bantuan *SPARQL* setiap term(kata) akan dicari padanannya pada tesaurus yang telah dimodelkan dengan *SKOS*. Setelah itu dilakukanlah query pada sistem sehingga hasil yang didapat dari query dapat digunakan untuk melakukan evaluasi dengan menggunakan *TREC Eval*.

1.5.4. Metode Evaluasi

Evaluasi sistem dibuat dengan menggunakan bantuan *TREC Eval*. *TREC Eval* akan mencocokkan hasil pencarian yang diberikan Lucene dan *relevance judgement* yang telah disediakan oleh corpus Kompas. Dengan dicocokkan dengan *relevance judgement*, *TREC Eval* akan memberikan beberapa nilai yang sering digunakan untuk evaluasi pencarian, seperti *interpolated precision-recall*, *F1-measure*, *GMAP*, dan lain-lain. Nilai interpolasi tertinggi pada titik tertentu menandakan bahwa pencarian tersebut memiliki hasil yang paling baik. Selain itu kamus yang dibuat juga akan dilakukan evaluasi dengan menggunakan 3 unsur ontologi, yaitu *consistency*, *completeness* dan *conciseness*.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Pada bab ini membahas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisi dua sub bab, yaitu tinjauan pustaka dan landasan teori. Pada tinjauan pustaka diuraikan tentang penelitian-penelitian yang telah dilakukan terkait dengan topik penelitian ini. Pada landasan teori dijelaskan mengenai teori-teori yang mendukung penelitian.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini dijelaskan tentang data-data yang akan digunakan dalam penelitian ini, apa saja variabel yang akan dipakai, dan perancangan sesuai dengan masalah yang telah diutarakan

BAB IV IMPLEMENTASI AND ANALISIS SISTEM

Pada bab ini dijelaskan mengenai hasil implementasi dari perancangan pada bab sebelumnya, sistem yang telah dibuat dianalisis dan dievaluasi hasilnya sesuai rumusan masalah yang ada.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan-kesimpulan yang didapat dari hasil penelitian dan berisi saran-saran yang sesuai dengan permasalahan yang diteliti.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil implementasi dan analisis yang telah dibahas pada bab sebelumnya, maka penulis menyimpulkan beberapa hal terkait penelitian ini, kesimpulan tersebut antara lain:

1. Penelitian ini menghasilkan sebuah tesaurus bahasa Indonesia yang *ready to use* dalam representasi pengetahuan *Simple Knowledge Organization Sistem*. *Ready to use* yang dimaksud adalah tesaurus yang dapat digunakan untuk mencari daftar persamaan kata, atau file SKOS yang dihasilkan dapat digunakan sebagai bagian terpisah.
2. Kamus yang dibuat dalam format *Simple Knowledge Organization Sistem* dipetakan dengan memberikan *skos:Concept* sebagai pengenalan sebuah kata. Untuk memberikan atribut sebuah kata diberikan *wordattr:atribut*. Sebagai pengenalan dari *Concept* yang dibuat diberikan *skos:prefLabel*. Sedangkan untuk mengagipit daftar persamaan kata diberikan *skos:altLabel*.
3. Dalam penelitian ini kamus yang digunakan sudah memenuhi aspek *conciseness* (ringkas/padat) hal ini ditunjukkan dengan tidak adanya *Concept* yang berulang pada sistem (Sesame Server). Namun kamus yang digunakan masih memiliki kekurangan pada aspek *completeness* dan *consistency*. Terbukti masih ada 78108 kata yang hilang dari kamus tersebut.
4. Pada kedua dataset (433 dokumen dan 1000 dokumen) nilai *F-Measure* untuk data normal lebih tinggi dibandingkan data yang menggunakan kamus. Kamus yang digunakan untuk ekspansi menyebabkan besarnya dokumen kembalian dari pencarian. Hal ini menyebabkan nilai *F-Measure*

pada data yang menggunakan kamus menjadi lebih kecil daripada data normal.

5.2. Saran

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

1. Ketidaklengkapan pada kamus yang masih ada dapat dibuat sebuah sistem yang mampu melengkapi kamus tersebut. Dengan kamus yang lebih lengkap pencarian yang dilakukan dapat lebih luas.
2. Pada saat dilakukan ekspansi data, dapat dilakukan penambahan pengenalan penugasan (verba, nomina, arkais, cakapan) pada sebuah kata. Sehingga saat melakukan ekspansi data, dapat dipilih persamaan yang penugasannya sama.

Daftar Pustaka

- Asian, J., Williams, H. E., & Tahaghoghi, S. (2004). A Testbed for Indonesian Text Retrieval.
- Assem, M. v., Malaise, V., Miles, A., & Schreiber, G. (2004). A Method to Convert Thesauri to SKOS. *Proc. of the 3rd Int'l Semantic Web Conf. (ISWC'04)* , 17–31.
- Broughton, V. (2006). *Essential Thesaurus Construction*. London: Facet Publishing.
- DuCharme, B. (2013). *Learning SPARQL 2nd Edition*. Sebastopol, California: O'Reilly Media, Inc.
- Endarmono, E. (2006). *Tesaurus Bahasa Indonesia*. Jakarta: Gramedia Pustaka Utama.
- Faro, S., Biasiotti, M. A., & Francesconi, E. (2011). Thesaurus Mapping for Promoting Semantic Interoperability of European Public Services. *eChallenges e-2011 Conference Proceedings* , 1.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering*. London: Springer-Verlag.
- Haslhofer, B., Martins, F., & Magalhães, J. (2013). Using SKOS vocabularies for improving Web Search. *International World Wide Web Conference*, (pp. 1253-1258). Rio de Janeiro.
- Imran, H., & Sharan, A. (2009). Thesaurus and Query Expansion. *International Journal of Computer science & Information Technology (IJCSIT)* , 89-97.
- Lasila, O., & McGuinness, D. (2001). *The Role of Frame-Based Representation on the Semantic Web*. Stanford, California: Knowledge Systems Laboratory, Stanford University.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York, USA: Cambridge University Press.
- McGuinness, D. L., & Noy, N. F. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. California: Stanford Knowledge Systems Laboratory Technical Report.
- Mikhalenko, P. (2005, June 22). *Introducing SKOS*. Retrieved November 1, 2014, from O'Reilly Media, Inc.: <http://www.xml.com/lpt/a/1593>

Miles, A., & Bechhofer, S. (2011, 08 06). *w3.org*. Retrieved 07 3, 2015, from <http://www.w3.org/2009/08/skos-reference/skos.html>

National Institute of Standards and Technology. (2015, Feb 23). *Text REtrieval Conference*. Retrieved Ag 3, 2015, from Text REtrieval Conference: <http://trec.nist.gov/pubs/call2015.html>

Palmer, S. B. (2001, Sept). *The Semantic Web: An Introduction*. Retrieved Sept 24, 2015, from <http://infomesh.net/2001/swintro/>

Pastor-Sanchez, J.-A., Mendez, F. J., & Rodríguez-Muñoz, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) comparedwith proposed alternatives. *Information Research*, vol 14 .

©UKPDW