

**SISTEM TEMU BALIK ARTIKEL ILMIAH MENGGUNAKAN
VECTOR SPACE MODEL DAN DI KELOMPOKKAN DENGAN
*K-MEANS CLUSTERING***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika**

**Disusun oleh:
DZULFIKAR FAUZI
24010313140065**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2018

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Judul : Sistem Temu Kembali Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*

Nama : Dzulfikar Fauzi

NIM : 24010313140065

Dengan ini saya menyatakan bahwa dalam Tugas Akhir/ Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 22 Mei 2018



Dzulfikar Fauzi
NIM 24010313140065

HALAMAN PENGESAHAN

Judul : Sistem Temu Kembali Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*

Nama : Dzulfikar Fauzi

NIM : 24010313140065

Telah diujikan pada sidang tugas akhir pada tanggal 7 Mei 2018 dan dinyatakan lulus pada tanggal 7 Mei 2018.

Semarang, 22 Mei 2018

Mengetahui,

Ketua Departemen Ilmu Komputer/Informatika



Dr. Retno Kusumaningrum, S.Si, M.Kom

NIP. 19810420 200501 2001

Ketua Penguji Tugas Akhir

Helmie Arif Wibawa, S.Si., M.Cs

NIP. 19780516 200312 1001

HALAMAN PENGESAHAN

Judul : Sistem Temu Kembali Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*

Nama : Dzulfikar Fauzi

NIM : 24010313140065

Telah diujikan pada sidang tugas akhir pada tanggal 7 Mei 2018.

Semarang, 22 Mei 2018

Dosen Pembimbing,



Sukmawati Nur Endah, S.Si., M.Kom
NIP. 19780502 200501 2002

ABSTRAK

Sistem Temu Balik Informasi (*Information Retrieval*) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunanya secara otomatis dari suatu koleksi informasi. Dengan banyaknya jumlah dokumen yang ter-*retrieve*, akan lebih mudah jika hasil dokumen ter-*retrieve* itu akan dikelompokkan secara otomatis sesuai dengan kemiripannya. Pada penelitian ini digunakan metode *Vector Space Model* untuk penerapan pencarian dokumen (*information retrieval*) dan *K-Means Clustering* untuk pengelompokkan dokumen. Hasil dari kinerja pencarian dokumen artikel ilmiah pada hasil implementasi *Vector Space Model* mempunyai nilai MAP sebesar 0.955034122, yang mengindikasikan sistem mampu me-*retrieve* dokumen yang relevan terhadap *user* sebesar 95.50%. Kemudian hasil dari kinerja pengelompokkan dokumen artikel ilmiah pada hasil implementasi metode *K-Means Clustering* adalah pada jumlah pengelompokkan 2 *cluster* memperoleh kualitas sebesar 81.09%, dan untuk jumlah pengelompokkan 4 *cluster* memperoleh kualitas sebesar 48.01%.

Kata Kunci : Sistem Temu Balik Informasi, *Vector Space Model*, *K-Means Clustering*, Artikel Ilmiah

ABSTRACT

Information Retrieval (Information Retrieval) is a system used to find information relevant to the needs of its users automatically from a collection of information. With the large number of retrieved documents, it would be easier if the retrieved documents would be automatically grouped according to their similarity. In this research, Vector Space Model is used to apply document retrieval and K-Means Clustering for grouping documents. The results of the scientific paper article search performance on the Vector Space Model implementation results have a MAP value of 0.955034122, indicating a system capable of retrieving relevant documents to the user of 95.50%. Then the result of the performance of grouping of scientific article documents on the result of K-Means Clustering method implementation is on the number of clustering 2 clusters get the quality of 81.09%, and for the number of clustering 4 clusters get the quality of 48.01%.

Kata Kunci : Information Retrieval System, *Vector Space Model*, *K-Means Clustering*, Scientific Articles

KATA PENGANTAR

Segala puji syukur bagi Allah SWT atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul “Analisis Sentimen Menggunakan *Latent Dirichlet Allocation* dan Visualisasi *Topic Polarity Wordcloud*”. Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan tugas akhir ini tentulah telah banyak mendapat bantuan dan dukungan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Ketua Departemen Ilmu Komputer/ Informatika FSM Universitas Diponegoro Semarang.
2. Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/ Informatika FSM Universitas Diponegoro Semarang.
3. Sukmawati Nur Endah, S.Si, M.Kom, selaku dosen Pembimbing yang telah meluangkan waktu dan berkenan memberikan bimbingan, arahan, masukan, serta motivasi yang sangat berharga dan fokus akan tujuan bagi penulis.
4. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu kelancaran penelitian ini, semoga Tuhan yang memberikan balasan yang lebih baik.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Semarang, 22 Mei 2018



Dzafikar Fauzi

DAFTAR ISI

	Hal
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup	3
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1. Penelitian Terkait (<i>State of the art</i>)	5
2.2. Penelusuran Literatur Artikel Ilmiah	7
2.3. Sistem Temu Balik Informasi (<i>Information Retrieval</i>)	9
2.4. <i>Preprocessing</i>	10
2.4.1. Tokenisasi	11
2.4.2. Penghapusan <i>Stopword</i>	11
2.4.3. <i>Stemming</i>	12

2.4.4. <i>Indexing Text</i>	19
2.5. <i>Vector Space Model</i>	19
2.6. <i>K-Means Clustering</i>	21
2.7. Penilaian Relevansi	25
2.8. Penilaian <i>Kappa</i>	27
2.9. <i>Purity</i> untuk <i>Clustering</i>	28
2.10. Model Pengembangan Perangkat Lunak	28
BAB III METODOLOGI PENELITIAN	31
3.1. Tahapan Proses Penelitian	31
3.1.1. Pengumpulan Data Artikel	32
3.1.2. <i>Preprocessing</i>	32
3.1.2.1. Tokenisasi	33
3.1.2.2. Penghapusan <i>Stopword</i>	34
3.1.2.3. <i>Stemming</i>	35
3.1.2.4. Pembobotan Kata	36
3.1.2.5. <i>Indexing Text</i>	36
3.1.3. Proses Pencarian	37
3.1.4. Proses Clustering	38
3.1.5. Evaluasi	39
3.2. Contoh Hasil Proses Penelitian	40
3.3. Analisa dan Perancangan Sistem	51
3.3.1. Analisis Sistem	51
3.3.1.1. Deskripsi Sistem	51
3.3.1.2. Daftar Aktor	52
3.3.1.3. Kebutuhan Fungsional Sistem	52
3.3.1.4. Kebutuhan Non-Fungsional Sistem	53
3.3.2. Perancangan Sistem	53

3.3.2.1. Pemodelan <i>Use Case</i> Sistem	53
3.3.2.1.1. Skenario melakukan pencarian artikel ilmiah	54
3.3.2.1.2. Skenario melihat hasil pengelompokkan artikel ilmiah	56
3.3.2.1.3. Skenario melihat hasil tahapan proses dari tiap metode	57
3.3.2.1.4. Skenario melakukan <i>login</i>	58
3.3.2.1.5. Skenario melakukan <i>logout</i>	59
3.3.2.1.6. Skenario melakukan kelola artikel ilmiah	60
3.3.2.1.7. Skenario melakukan <i>update</i> hasil <i>preprocessing</i>	64
3.3.2.1.8. Skenario melakukan <i>update</i> jumlah <i>cluster</i>	64
3.3.2.2. Realisasi <i>Use Case</i> Tahap Analisis	65
3.3.2.2.1. Melakukan pencarian artikel ilmiah	65
3.3.2.2.2. Melihat hasil pengelompokkan artikel ilmiah	66
3.3.2.2.3. Melihat hasil tahapan proses tiap metode	66
3.3.2.2.4. Melakukan login	67
3.3.2.2.5. Melakukan logout	68
3.3.2.2.6. Melakukan kelola artikel ilmiah	68
3.3.2.2.7. Melakukan <i>update</i> hasil <i>preprocessing</i>	71
3.3.2.2.8. Melakukan <i>update</i> jumlah <i>cluster</i>	72
3.3.2.3. Realisasi <i>Use Case</i> Tahap Perancangan	73
3.3.2.3.1. Perancangan <i>Class Diagram</i>	73
3.3.2.3.1. Perancangan <i>Database</i>	75
BAB IV HASIL DAN ANALISA HASIL PENELITIAN	77
4.1. Implementasi	77
4.1.1. Implementasi <i>Class Diagram</i>	77
4.1.2. Implementasi Antarmuka	78
4.1.2.1. Implementasi Antarmuka <i>Home</i>	78
4.1.2.2. Implementasi Antarmuka <i>Rank Results</i>	79

4.1.2.3. Implementasi Antarmuka <i>Group Results</i>	79
4.1.2.4. Implementasi Antarmuka <i>Process Details</i>	79
4.1.2.5. Implementasi Antarmuka <i>Login</i>	80
4.1.2.6. Implementasi Antarmuka <i>Logout</i>	81
4.1.2.7. Implementasi Antarmuka <i>Manage Articles</i>	81
4.1.2.8. Implementasi Antarmuka <i>Add Article</i>	82
4.1.2.9. Implementasi Antarmuka <i>View Article</i>	83
4.1.2.10. Implementasi Antarmuka <i>Edit Article</i>	83
4.1.2.11. Implementasi Antarmuka <i>Delete Article</i>	84
4.1.2.12. Implementasi Antarmuka <i>Delete All Article</i>	85
4.1.2.13. Implementasi Antarmuka <i>Update Data</i>	85
4.1.2.14. Implementasi Antarmuka <i>Setting</i>	85
4.2. Pengujian	86
4.2.1. Spesifikasi Perangkat	86
4.2.2. Data Penelitian	87
4.2.3. Pengujian Fungsional Sistem	87
4.2.3.1. Rencana Pengujian.....	87
4.2.3.2. Hasil Pengujian	88
4.2.4. Pengujian Kinerja Sistem	88
4.2.4.1. Pengujian Pencarian Dokumen.....	88
4.2.4.2. Pengujian Pengelompokkan Dokumen	91
BAB V KESIMPULAN	94
5.1. Kesimpulan	94
5.2. Saran	94
DAFTAR PUSTAKA	96
LAMPIRAN-LAMPIRAN	100

DAFTAR GAMBAR

	Hal
Gambar 2.1. Tampilan hasil pencarian pada <i>Google Scholar</i>	9
Gambar 2.2. Kerangka dari Sistem Temu Balik Informasi Sederhana	10
Gambar 2.3. Metodologi <i>indexing text</i>	11
Gambar 2.4. Diagram Alur Algoritma <i>K-Means Clustering</i>	24
Gambar 2.5. Proses Pengembangan dengan Pendekatan <i>waterfall</i>	29
Gambar 3.1. Diagram blok proses sistem temu balik artikel ilmiah	31
Gambar 3.2. Tampilan jurnal online UNDIP JMASIF	32
Gambar 3.3. Diagram alur <i>preprocessing</i>	33
Gambar 3.4. Diagram alur proses tokenisasi	34
Gambar 3.5. Diagram alur proses penghapusan <i>stopword</i>	35
Gambar 3.6. Diagram alur proses <i>stemming</i>	36
Gambar 3.7. Diagram alur proses <i>indexing text</i>	37
Gambar 3.8. Diagram alur proses <i>vector space model</i>	38
Gambar 3.9. Contoh proses tokenisasi	41
Gambar 3.10. Contoh proses penghapusan <i>stopword</i>	41
Gambar 3.11. Contoh proses <i>stemming</i>	42
Gambar 3.12. Contoh proses <i>indexing text</i>	44
Gambar 3.13. Diagram <i>use case</i>	54
Gambar 3.14. Sketsa antarmuka <i>use case</i> melakukan pencarian artikel ilmiah (1)	55
Gambar 3.15. Sketsa antarmuka <i>use case</i> melakukan pencarian artikel ilmiah (2)	56
Gambar 3.16. Sketsa antarmuka alternatif 1 <i>use case</i> melakukan pencarian artikel ilmiah	56
Gambar 3.17. Sketsa antarmuka <i>use case</i> menampilkan tahapan proses dari tiap metode	58
Gambar 3.18. Sketsa antarmuka <i>use case</i> melakukan <i>login</i>	59
Gambar 3.19. Sketsa antarmuka <i>use case</i> melakukan <i>logout</i>	60

Gambar 3.20. Sketsa antarmuka <i>use case</i> melakukan kelola artikel ilmiah	62
Gambar 3.21. Sketsa antarmuka alternatif 1 <i>usecase</i> menambah artikel ilmiah	62
Gambar 3.22. Sketsa antarmuka alternatif 2 <i>usecase</i> melihat artikel ilmiah	62
Gambar 3.23. Sketsa antarmuka alternatif 3 <i>usecase</i> mengedit artikel ilmiah	63
Gambar 3.24. Sketsa antarmuka alternatif 4 <i>usecase</i> menghapus artikel ilmiah	63
Gambar 3.25. Sketsa antarmuka alternatif 5 <i>usecase</i> menghapus semua artikel ilmiah ...	63
Gambar 3.26. Sketsa antarmuka <i>usecase</i> melakukan <i>update</i> jumlah <i>cluster</i>	65
Gambar 3.27. <i>Analysis class model</i> melakukan pencarian artikel ilmiah	65
Gambar 3.28. <i>Sequence diagram</i> melakukan pencarian artikel ilmiah	66
Gambar 3.29. <i>Analysis class model</i> melihat hasil pengelompokkan artikel ilmiah	66
Gambar 3.30. <i>Sequence diagram</i> melihat hasil pengelompokkan artikel ilmiah	66
Gambar 3.31. <i>Analysis class model</i> melihat hasil tahapan proses tiap metode	67
Gambar 3.32. <i>Sequence diagram</i> melihat hasil tahapan proses tiap metode	67
Gambar 3.33. <i>Analysis class model</i> melakukan <i>login</i>	67
Gambar 3.34. <i>Sequence diagram</i> melakukan <i>login</i>	68
Gambar 3.35. <i>Analysis class model</i> melakukan <i>logout</i>	68
Gambar 3.36. <i>Sequence diagram</i> melakukan <i>logout</i>	68
Gambar 3.37. <i>Analysis class model</i> alternatif 1 menambah artikel ilmiah	69
Gambar 3.38. <i>Analysis class model</i> alternatif 2 melihat artikel ilmiah	69
Gambar 3.39. <i>Analysis class model</i> alternatif 3 mengedit artikel ilmiah	69
Gambar 3.40. <i>Analysis class model</i> alternatif 4 menghapus artikel ilmiah	69
Gambar 3.41. <i>Analysis class model</i> alternatif 5 menghapus semua artikel ilmiah	70
Gambar 3.42. <i>Sequence diagram</i> alternatif 1 menambah artikel ilmiah	70
Gambar 3.43. <i>Sequence diagram</i> alternatif 2 melihat artikel ilmiah	70
Gambar 3.44. <i>Sequence diagram</i> alternatif 3 mengedit artikel ilmiah	70
Gambar 3.45. <i>Sequence diagram</i> alternatif 4 menghapus artikel ilmiah	71
Gambar 3.46. <i>Sequence diagram</i> alternatif 5 menghapus semua artikel ilmiah	71

Gambar 3.47. <i>Analysis class model</i> melakukan <i>update</i> hasil <i>preprocessing</i>	71
Gambar 3.48. <i>Sequence diagram</i> melakukan <i>update</i> hasil <i>preprocessing</i>	72
Gambar 3.49. <i>Analysis class model</i> melakukan <i>update</i> jumlah <i>cluster</i>	72
Gambar 3.50. <i>Sequence diagram</i> melakukan <i>update</i> jumlah <i>cluster</i>	72
Gambar 3.51. <i>Class Diagram SipLah</i>	73
Gambar 4.1. Implementasi Antarmuka <i>Home</i>	78
Gambar 4.2. Implementasi Antarmuka <i>Rank Results</i>	79
Gambar 4.3. Implementasi Antarmuka <i>Process Details</i>	80
Gambar 4.4. Implementasi Antarmuka <i>Login</i>	80
Gambar 4.5. Implementasi Antarmuka <i>Logout</i>	81
Gambar 4.6. Implementasi Antarmuka <i>Manage Articles</i>	82
Gambar 4.7. Implementasi Antarmuka <i>Add Articles</i>	82
Gambar 4.8. Implementasi Antarmuka <i>View Article</i>	83
Gambar 4.9. Implementasi Antarmuka <i>Edit Article</i>	84
Gambar 4.10. Implementasi Antarmuka <i>Delete Article</i>	84
Gambar 4.11. Implementasi Antarmuka <i>Delete All Article</i>	85
Gambar 4.12. Implementasi Antarmuka <i>Setting</i>	86
Gambar 4.13. Grafik <i>average precision</i> oleh <i>user</i> terhadap <i>query</i>	90

DAFTAR TABEL

	Hal
Tabel 2.1. Penelitian terkait sistem temu kembali informasi teks berbahasa indonesia dan usulan penelitian tugas akhir sistem temu balik artikel ilmiah	5
Tabel 2.2. Aturan pemenggalan awalan stemmer Nazief & Adriani	15
Tabel 2.3. Modifikasi dan tambahan aturan pada Tabel 2.2 oleh algoritma <i>CS Stemmer</i> .	17
Tabel 2.4. Modifikasi dan aturan untuk Tabel 2.2 oleh algoritma <i>ECS stemmer</i>	19
Tabel 2.5. Perbedaan karakteristik umum metode <i>clustering</i>	22
Tabel 2.6. Tingkat relevansi, defenisi dan interpretasi yang digunakan dalam penilaian relevansi	26
Tabel 3.1. Contoh hasil nilai TF-IDF	43
Tabel 3.2. Contoh hasil nilai pembobotan kata tiap dokumen dan <i>query</i>	43
Tabel 3.3. Contoh hasil nilai <i>cosine similarity</i> tiap dokumen	45
Tabel 3.4. Contoh hasil proses <i>k-means clustering</i> iterasi ke-1	46
Tabel 3.5. Contoh hasil proses <i>k-means clustering</i> iterasi ke-2	47
Tabel 3.6. Contoh hasil proses <i>k-means clustering</i> iterasi ke-3/ terakhir	47
Tabel 3.7. Contoh hasil pengelompokkah artikel ilmiah	48
Tabel 3.8. Contoh hasil perhitungan <i>precision</i> dan <i>recall</i>	48
Tabel 3.9. Contoh data pengelompokkan dokumen	50
Tabel 3.10. Contoh data irisan dokumen untuk 2 <i>cluster</i>	51
Tabel 3.11. Daftar Aktor	52
Tabel 3.12. Kebutuhan fungsional	52
Tabel 3.13. Kebutuhan non-fungsional	53
Tabel 3.14. Diagram skenario melakukan pencarian artikel ilmiah	55
Tabel 3.15. Diagram skenario melihat hasil pengelompokkan artikel ilmiah	57
Tabel 3.16. Diagram skenario menampilkan tahapan proses dari tiap metode	57

Tabel 3.17. Diagram skenario melakukan <i>login</i>	58
Tabel 3.18. Diagram skenario melakukan <i>logout</i>	59
Tabel 3.19. Diagram skenario melakukan kelola artikel ilmiah	60
Tabel 3.20. Diagram skenario melakukan <i>update</i> hasil <i>preprocessing</i>	64
Tabel 3.21. Diagram skenario melakukan <i>update</i> jumlah <i>cluster</i>	64
Tabel 3.22. Deskripsi <i>Class Diagram</i>	74
Tabel 3.23. Tabel <i>Articles</i>	75
Tabel 3.24. Tabel <i>Preprocessing</i>	76
Tabel 3.25. Tabel Admin	76
Tabel 4.1. Implementasi <i>class diagram</i>	77
Tabel 4.2. Rencana Pengujian	87
Tabel 4.3. Hasil <i>average precision</i> dan <i>measure average precision</i>	89
Tabel 4.4. Hasil <i>ground truth</i> untuk 2 <i>cluster</i>	91
Tabel 4.5. Hasil <i>ground truth</i> untuk 4 <i>cluster</i>	92
Tabel 4.6. Hasil perhitungan <i>purity</i> untuk 2 <i>cluster</i>	92
Tabel 4.7. Hasil perhitungan <i>purity</i> untuk 4 <i>cluster</i>	92
Tabel L.2.1. Daftar <i>stopword library sastrawi</i>	101
Tabel L.4.1. Daftar dokumen artikel ilmiah	102
Tabel L.4.2. Daftar hasil pengujian sistem	108
Tabel L.4.3. Daftar ranking dokumen artikel ilmiah yang ter- <i>retrieve</i> terhadap <i>query</i> .	111
Tabel L.4.4. Daftar hasil pengujian metode <i>vector space model</i>	115
Tabel L.4.5. Daftar pengelompokkan dokumen oleh kedua penilai untuk 2 <i>cluster</i>	120
Tabel L.4.6. Daftar pengelompokkan dokumen oleh kedua penilai untuk 4 <i>cluster</i>	124
Tabel L.4.7. Daftar hasil perhitungan nilai <i>kappa</i> tiap <i>query</i> untuk 2 <i>cluster</i>	128
Tabel L.4.8. Daftar hasil perhitungan nilai <i>kappa</i> tiap <i>query</i> untuk 4 <i>cluster</i>	130
Tabel L.4.9. Daftar hasil penentuan <i>ground truth</i> tiap <i>query</i> untuk 2 <i>cluster</i>	134
Tabel L.4.10. Daftar hasil penentuan <i>ground truth</i> tiap <i>query</i> untuk 4 <i>cluster</i>	136

DAFTAR LAMPIRAN

	Hal
Lampiran 1. Daftar <i>stopword</i>	101
Lampiran 2. Daftar dokumen artikel ilmiah	102
Lampiran 3. Daftar hasil pengujian sistem	108
Lampiran 4. Daftar hasil pengujian pencarian dokumen	111
Lampiran 5. Daftar hasil pengujian pengelompokkan dokumen	115

BAB I

PENDAHULUAN

Bab pendahuluan ini berisi latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan pelaksanaan Tugas Akhir mengenai Sistem Temu Balik Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*.

1.1. Latar Belakang

Pencarian informasi erat kaitannya dengan kebutuhan akan informasi. Seseorang yang membutuhkan informasi memerlukan waktu untuk berpikir apa yang dibutuhkan, mengingat apa yang dibutuhkan, selanjutnya memutuskan apa yang dibutuhkan (Ilmi, 2014). Proses dan penggunaan alat yang tepat akan menghasilkan informasi yang tepat pula (Umam, 2015). Alat tersebut salah satunya dapat diterapkan menggunakan sistem temu balik informasi. Contohnya adalah penggunaan mesin pencari seperti *Google*. Sistem Temu Balik Informasi (*Information Retrieval*) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunanya secara otomatis dari suatu koleksi informasi. Proses tersebut membutuhkan sebuah model *information retrieval* (Nasution, 2016).

Model *information retrieval* adalah model yang digunakan untuk melakukan pencocokan antara *term-term* dari *query* dengan *term-term* dalam *document collection*. Terdapat 3 jenis model dalam *information retrieval*, yaitu *Set-theoretic models*, *Algebraic model*, dan *Probabilistic model*. Salah satu contoh *algebraic model* adalah *Vector Space Model* (Robinson, 2014). *Vector space model* (VSM) adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan *query*, serta mewakili setiap dokumen dalam sebuah koleksi sebagai sebuah titik dalam ruang (vektor dalam ruang vektor) (Wibowo, 2012).

Polettini (2014) dikutip dalam Karyono & Utomo (2012) menyatakan bahwa keberhasilan dari model VSM ini ditentukan oleh skema pembobotan terhadap suatu *term* baik untuk cakupan lokal maupun global, dan faktor normalisasi. Kelebihan dari metode VSM ini yaitu memberikan bobot kata dengan berbagai jenis bobot yang ditambahkan, dan bobot kata dapat memperbaiki kualitas dari suatu set jawaban. VSM

juga dapat mengukur *similarity* (kemiripan) antara apapun, seperti dokumen dengan *query*, dokumen dengan dokumen, *query* dengan *query*. Adapun pertimbangan pemilihan metode VSM, menurut Amin (2012) cara kerja model ini efisien dan mudah dalam representasinya.

Untuk mesin pencarian yang ada, hasil dokumen yang *ter-retrieve* masih berupa urutan dokumen ter-rangking sesuai pencarian. Selain pencarian dokumen, ada juga penggunaan untuk pengelompokkan dokumen. Penelitian ini mengusulkan adanya mesin pencari yang disertai pengelompokkan dokumen. Latar belakang mengapa dibutuhkannya pengelompokkan pada dokumen, yaitu untuk mempermudah dalam mengelompokkan dokumen artikel ilmiah secara otomatis yang sesuai dengan kemiripannya. Ada beberapa macam algoritma dalam metode pengelompokkan. Pada dasarnya algoritma (*set of rules*) pengelompokkan objek ke dalam klasternya diklasifikasikan menjadi dua metode hierarki dan metode non-hierarki (Rivani, 2010). Salah satu metode non-hierarki yang dapat digunakan untuk mengkluster dokumen adalah *k-means clustering*.

Pemilihan algoritma *k-means clustering* ini disebabkan karena pertimbangan dari segi kualitas, waktu penyelesaian dan jumlah data yang diproses, seperti pada penelitian yang dilakukan oleh Zahrotun (2015) yang berkesimpulan bahwa algoritma *k-means clustering* memiliki hasil lebih baik dari pada metode klasterisasi yang lain seperti model hierarki. Begitu juga menurut Rivani (2010) dalam kajian penelitiannya menyebutkan bahwa pengklasteran menggunakan *k-means clustering* lebih cepat daripada metode hierarki dan lebih menguntungkan untuk jumlah objek atau kasus yang besar.

Selain kebutuhan akan pengimplementasian sistem, penelitian ini juga perlu dievaluasi kinerjanya dalam sistem temu balik informasi yang menggunakan *vector space model* dan *k-means clustering* sebagai implementasinya, terutama pada penentuan jumlah *cluster* yang sesuai representasinya dalam pengelompokkan dokumen. Objek dari penelitian ini adalah penelusuran dokumen artikel ilmiah, dengan tujuan untuk mempermudah akademisi maupun peneliti dalam mencari referensi sebagai bahan rujukan dalam kegiatan penelitian.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, dapat disusun rumusan masalah pada tugas akhir penelitian ini yaitu mengimplementasikan Sistem Temu Balik Artikel Ilmiah dengan menerapkan *Vector Space Model* dalam proses perangkan dokumen artikel ilmiah yang ter-*retrieve* dan *K-Means Clustering* untuk mengelompokkan dokumen artikel ilmiah yang ter-*retrieve* tersebut. Selanjutnya sistem tersebut dievaluasi untuk menentukan jumlah cluster yang representasinya dalam pengelompokkan dokumen.

1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai pada penelitian tugas akhir ini adalah untuk me-*retrieve* artikel-artikel ilmiah dan dapat digolongkan sesuai dengan korelevanan antar dokumen.

Adapun manfaat yang diharapkan dari penelitian Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan Sistem Temu Balik Informasi menggunakan *Vector Space Model* dan *K-Means Clustering*.
2. Mengevaluasi kinerja Sistem Temu Balik Informasi menggunakan *Vector Space Model* dan *K-Means Clustering*.
3. Membantu civitas akademika dalam mendapatkan artikel ilmiah di lingkungan departemen Informatika.

1.4. Ruang Lingkup

Ruang lingkup dari batasan topik pada penelitian Tugas Akhir ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah artikel ilmiah mahasiswa Informatika sebanyak 60 (enam puluh) artikel.
2. Dokumen artikel ilmiah yang diproses menggunakan Bahasa Indonesia.
3. *Input* yang diproses berisi teks dari dokumen artikel ilmiah.
4. Hasil *output* berupa berkas berekstensi pdf dari dokumen artikel ilmiah yang ter-*retrieve*.
5. Aplikasi perangkat lunak yang dibangun berbasiskan *Web*.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu

BAB I PENDAHULUAN

Bab pendahuluan ini berisi latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup pelaksanaan Tugas Akhir mengenai Sistem Temu Balik Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang teori-teori yang digunakan dalam pengerjaan Tugas Akhir Sistem Temu Balik Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan proses penelitian, contoh hasil pemrosesan, analisis dan perancangan sistem yang dilakukan pada pembuatan Sistem Temu Balik Artikel Ilmiah menggunakan *Vector Space Model* dan di kelompokkan dengan *K-Means Clustering*.

BAB IV HASIL DAN ANALISA HASIL PENELITIAN

Bab ini menjelaskan tentang implementasi dan pengujian pada Sistem Temu Balik Artikel Ilmiah menggunakan Metode *Vector Space Model* dan *K-Means Clustering*.

BAB V PENUTUP

Bab ini berisi mengenai kesimpulan dari kegiatan penelitian ini, serta saran yang diajukan untuk pengembangan kegiatan penelitian lebih lanjut.