

**SISTEM PENCARIAN DAN PERINGKASAN BERITA *ONLINE*
BERBAHASA INDONESIA MENGGUNAKAN METODE *LATENT*
SEMANTIC INDEXING (LSI)**



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer
pada Jurusan Ilmu Komputer/Informatika**

Disusun Oleh:

MUHAMMAD LUKMAN FARIQ

24010313120024

DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2017

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Lukman Fariq

NIM : 24010313120024

Judul : Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia
Menggunakan Metode *Latent Semantic Indexing* (LSI)

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 di dalam daftar pustaka.

Semarang, 20 September 2017



Muhammad Lukman Fariq
24010313120024

HALAMAN PENGESAHAN

Judul : Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia
Menggunakan Metode *Latent Semantic Indexing* (LSI)
Nama : Muhammad Lukman Fariq
NIM : 24010313120024

Telah diujikan pada sidang tugas akhir pada tanggal 7 September 2017 dan dinyatakan lulus pada tanggal 7 September 2017.

Semarang, 20 September 2017

Mengetahui,
a.n. Ketua Departemen Ilmu Komputer/ Informatika
Sekretaris,



Dr. Eng. Adi Wibowo, S.Si, M.Kom
NIP.19820309 200604 1002

Panitia Penguji Tugas Akhir
Ketua,

A handwritten signature in black ink, appearing to read 'R. Kusumaningrum', written over a set of three horizontal lines.

Dr. Retno Kusumaningrum, S.Si, M.kom
NIP. 19810420 200501 2001

HALAMAN PENGESAHAN

Judul : Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia
Menggunakan Metode *Latent Semantic Indexing* (LSI)

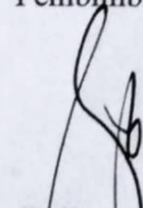
Nama : Muhammad Lukman Fariq

NIM : 24010313120024

Telah diujikan pada sidang tugas akhir pada tanggal 7 September 2017.

Semarang, 20 September 2017

Pembimbing



Sutikno, S.T., M.Cs

NIP. 19790524 200912 1003

ABSTRAK

Jumlah berita *online* yang tersebar di internet semakin bertambah banyak, selain itu banyaknya situs berita yang terdaftar pada *www* (*world wide web*) memungkinkan hasil pencarian berita menjadi sangat luas. Mesin pencari berita berbasis *information retrieval* menjadi solusi untuk mempermudah masyarakat dalam mencari informasi berita tertentu. *Information retrieval* menghasilkan beberapa berita yang diurutkan berdasarkan tingkat relevansi terhadap *query*. Relevan atau tidaknya berita akan diketahui secara jelas setelah pengguna membaca keseluruhan isi berita. Oleh karena itu mesin pencari berita disertai ringkasan berita membantu memudahkan pengguna untuk mencari informasi lebih cepat tanpa membaca keseluruhan isi berita. Peringkasan berita menggunakan peringkasan ekstraktif *Cross Method* dengan memilih kalimat-kalimat penting sebagai representasi ringkasan berita. Penelitian ini menerapkan metode *Latent Semantic Indexing* yang mampu mencari hubungan semantik tiap kata untuk mencari nilai kemiripan antar kalimat maupun *query* dengan dokumen. Data penelitian yang dipakai berupa korpus yang berisi 100 berita dari situs Detik, Kompas dan Tribunnews. Sistem memberikan *output* berupa 10 urutan teratas berita yang dicari disertai dengan ringkasannya. Berdasarkan hasil evaluasi pada peringkasan dan pencarian, nilai akurasi dipengaruhi oleh perubahan parameter nilai *k-rank*. Peringkasan berita yang diuji dengan ringkasan pakar menghasilkan rata-rata nilai *precision* 0.41, nilai *recall* 0.64 dan nilai *F-score* 0.49 pada nilai *k-rank* = 3. Nilai *k-rank* optimal untuk pencarian berita adalah 2 yang menghasilkan nilai *Mean Average Precision* (MAP) sebesar 0.73. Kombinasi penggunaan korpus ringkasan dan parameter *k-rank* untuk pencarian berita menghasilkan nilai *Mean Average Precision* (MAP) sebesar 0.40.

Kata kunci: *information retrieval*, *Cross Method*, *Latent Semantic Indexing*, *k-rank*

ABSTRACT

The number of online news spread over the internet was increasing, in addition to the number of news sites listed on the www (world wide web) allows the search results become very wide news. News-based search engine retrieval into a solution to facilitate the public in searching for certain news information. The information retrieval generates some news that was sorted by the relevance level to the query. Relevant or not the news will be known clearly after the user read the entire contents of the news. Therefore news search engine are accompanied by a summary of the news helps to make it easier for users to find information faster without reading the entire contents of the news. Summarizing news using extractive summarization *Cross Method* by selecting important sentences as a summary representation of news. This research applies Latent Semantic Indexing method which was able to find the semantic relationship of each word to find the value of similarity between sentence and query with document. The research data used was a corpus containing 100 news from Detik, Kompas and Tribunnews website. The system provides output of the top 10 searched newsletters accompanied by a summary. Based on the evaluation results on the summary and search, the accuracy value is influenced by the change of k-rank value parameter. The summary of the news tested with the expert summary resulted an average precision value of 0.41, a recall value of 0.64 and a F-score of 0.49 in k-rank value = 3. The optimal k-rank value for news search is 2 which resulted Mean Mean Precision (MAP) Of 0.73. The combined use of summary cores and k-rank parameters for news search resulted a Mean Mean Precision (MAP) value of 0.40.

Keywords: *information retrieval, Cross Method, Latent Semantic Indexing, k-rank*

KATA PENGANTAR

Puji syukur bagi Allah SWT atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul “Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia Menggunakan Metode *Latent Semantic Indexing* (LSI)”. 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. 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. Sutikno, S.T., M.Cs, 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. Orang tua, keluarga, teman dekat, dan sahabat yang telah mendukung, membantu, dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir ini.
5. 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, 20 September 2017

Penulis

DAFTAR ISI

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	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Tujuan dan Manfaat.....	4
1.4. Ruang Lingkup	4
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian di Bidang <i>Summarization</i> dan <i>Information Retrieval</i>	6
2.2. Peringkasan Teks Otomatis	7
2.2.1. Pengertian Peringkasan Teks Otomatis	7
2.2.2. Teknik Ringkasan	7
2.3. Sistem Temu-Kembali Informasi	8
2.3.1. Pengertian Sistem Temu Balik Informasi.....	8
2.3.2. Komponen Sistem Temu Balik Informasi	8
2.4. <i>Text Preprocessing</i>	10
2.4.1. <i>Sentence Splitting</i>	10
2.4.2. <i>Case Folding</i>	11
2.4.3. <i>Tokenizing</i>	11
2.4.4. <i>Stoplist / Stop Word Removal / filtering</i>	11
2.4.5. <i>Stemming</i>	11
2.5. <i>Latent Semantic Indexing (LSI)</i>	17

2.6. <i>Singular Value Decomposition (SVD)</i>	18
2.7. <i>Sentence Selection</i>	23
2.8. <i>Query Vector Mapping</i>	23
2.9. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	24
2.10. <i>Cosine Similarity</i>	24
2.11. <i>Evaluasi Summarization dan Information Retrieval</i>	25
2.11.1. <i>Precision dan Recall</i>	25
2.11.2. <i>Average Precision (AP)</i>	27
2.11.3. <i>Mean Average Precision (MAP)</i>	27
2.12. <i>Model Pengembangan Perangkat Lunak</i>	27
2.13. <i>Unified Modeling Language (UML)</i>	29
2.13.1. <i>Use Case Diagram</i>	29
2.13.2. <i>Class Diagram</i>	30
2.13.3. <i>Sequence Diagram</i>	31
2.13.4. <i>Entity, Boundary, dan Control Objects</i>	31
BAB III METODOLOGI PENELITIAN	32
3.1. <i>Summarization</i>	33
3.1.1. <i>Preprocessing (Praproses)</i>	33
3.1.1.1. <i>Sentence Splitting</i>	34
3.1.1.2. <i>Tokenization</i>	35
3.1.1.3. <i>Stopword Removal</i>	37
3.1.1.4. <i>Stemming</i>	38
3.1.2. <i>Pembobotan kata TF-IDF</i>	43
3.1.3. <i>Pembentukan Matriks Term-By-Document</i>	45
3.1.4. <i>Perhitungan SVD Matriks A</i>	46
3.1.5. <i>Pembentukan ringkasan</i>	48
3.2. <i>Information Retrieval</i>	50
3.2.1. <i>Pembobotan kata TF-IDF</i>	50
3.2.2. <i>Pembentukan Matriks Term-By-Document</i>	51
3.2.3. <i>Perhitungan SVD Matriks A</i>	52
3.2.4. <i>Preprocessing Query</i>	53
3.2.5. <i>Query Vector Mapping</i>	54
3.2.6. <i>Perhitungan Nilai Similarity dan Ranking Dokumen</i>	55

3.3.	Analisis dan Perancangan	56
3.3.1.	Deskripsi Umum	56
3.3.2.	Analisis Sistem	57
3.3.2.1.	Karakteristik Pengguna	57
3.3.2.2.	Kebutuhan Fungsional	58
3.3.2.3.	Kebutuhan Non-Fungsional	58
3.3.2.4.	<i>Use Case</i>	59
3.3.2.5.	<i>Class Analysis</i>	63
3.3.2.6.	<i>Sequence Diagram</i>	65
3.3.3.	Desain Sistem	69
3.3.3.1.	Pemodelan <i>Class Diagram</i>	69
3.3.3.2.	Perancangan Data	70
3.3.3.3.	Desain Antarmuka	73
BAB IV HASIL DAN PEMBAHASAN		80
4.1	Hasil Pengembangan Sistem	80
4.1.1	Lingkungan Implementasi Sistem	80
4.1.2	Implementasi Data	81
4.1.3	Implementasi <i>Class</i>	82
4.1.4	Implementasi Antarmuka	83
4.2	Skenario Pengujian Sistem	88
4.2.1.	Pengujian Fungsional Sistem	89
4.2.1.1.	Spesifikasi Perangkat	89
4.2.1.2.	Rencana Pengujian Fungsional Sistem	89
4.2.2.	Pengujian Kinerja Sistem	90
4.2.2.1.	Data Eksperimen	90
4.2.2.2.	Eksperimen	91
4.3	Hasil dan Analisa Sistem	93
4.3.1	Pengujian Fungsional Sistem	93
4.3.2	Skenario Eksperimen 1	94
4.3.3	Skenario Eksperimen 2	95
4.3.4	Skenario Eksperimen 3	96
BAB V PENUTUP		98
5.1.	Kesimpulan	98

5.2. Saran	98
DAFTAR PUSTAKA.....	99
LAMPIRAN-LAMPIRAN	103
Lampiran 1. Daftar <i>Stoword List</i> Bahasa Indonesia	104
Lampiran 2. <i>Source Code</i> Fungsi	108
Lampiran 3. Data Kuisisioner Ringkasan Referensi Pakar	116
Lampiran 4. Deskripsi dan Hasil Pengujian Fungsional Sistem	135
Lampiran 5. Perhitungan Nilai <i>Precision</i> , <i>Recall</i> dan <i>F-score</i> Peringkasan	139
Lampiran 6. Perhitungan Nilai MAP Eksperimen 2.....	143
Lampiran 7. Perhitungan Nilai MAP dan Waktu Eksekusi Eksperimen 3.....	176
Lampiran 8. Data Berita Hasil <i>Crawling</i> Situs Berita <i>Online</i>	210

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Information Retrieval (Poernomo & Gunawan, 2015).....	10
Gambar 2.2 Hasil Dekomposisi Matriks USV (Berry, et al., 1995).....	19
Gambar 2.3 Model <i>Waterfall</i> (Pressman, 2001).....	28
Gambar 2.4 <i>Class Diagram</i> (Arlow & Neustadt, 2002)	30
Gambar 2.5 <i>Sequence Diagram</i> (Arlow & Neustadt, 2002).....	31
Gambar 3.1 Gambaran Umum Penelitian	32
Gambar 3.2 <i>Flowchart Preprocessing</i> Dokumen.....	34
Gambar 3.3 <i>Flowchart Tokenization</i> (Bashri, 2017).....	36
Gambar 3.4 Proses <i>Tokenization</i>	36
Gambar 3.5 <i>Flowchart Stopword Removal</i> (Bashri, 2017)	37
Gambar 3.6 Diagram Alur Proses <i>Stemming</i> (Bashri, 2017).....	38
Gambar 3.7 Diagram Alur Subproses <i>Stemmer Sastrawi</i> (Bashri, 2017)	38
Gambar 3.8 Diagram alur subproses <i>stemPluralWord</i> (Bashri, 2017).....	39
Gambar 3.9 Diagram alur subproses <i>stemSingularWord</i>	40
Gambar 3.10 <i>Flowchart</i> Pembobotan Kata TF-IDF.....	43
Gambar 3.11 <i>Flowchart</i> Proses Perhitungan SVD (Irwanto, 2017).....	46
Gambar 3.12 <i>Flowchart</i> Proses Pembentukan Ringkasan	48
Gambar 3.13 <i>Flowchart Preprocessing Query</i>	53
Gambar 3.14 <i>Flowchart Query Vector Mapping</i>	54
Gambar 3.15 Arsitektur Sistem	57
Gambar 3.16 <i>Use Case Diagram</i> SPBP BOBI	60
Gambar 3.17 <i>Sequence Diagram</i> Melakukan <i>Login</i>	66
Gambar 3.18 <i>Sequence Diagram</i> Crawling Berita.....	66
Gambar 3.19 <i>Sequence Diagram</i> Mengelola Berita	67
Gambar 3.20 <i>Sequence Diagram</i> Meringkas Berita	68
Gambar 3.21 <i>Sequence Diagram</i> Mencari Berita	69
Gambar 3.22 <i>Class Diagram</i> SPBP BOBI.....	70
Gambar 3.23 <i>Persistent Class</i>	71
Gambar 3.24 Desain Antarmuka Halaman <i>Login</i>	74
Gambar 3.25 Desain Antarmuka Melakukan <i>Crawling</i> Berita	74
Gambar 3.26 Desain Antarmuka Halaman Mengelola Berita.....	75
Gambar 3.27 Desain Antarmuka Mengelola Berita (Detail).....	75
Gambar 3.28 Desain Antarmuka Mengelola Berita (<i>Delete</i>)	76
Gambar 3.29 Desain Antarmuka Peringkasan Berita.....	76
Gambar 3.30 Desain Antarmuka Halaman Proses Peringkasan Berita	77
Gambar 3.31 Desain Antarmuka Halaman Pencarian Berita (<i>Indexing</i>).....	77
Gambar 3.32 Desain Antarmuka Halaman Pencarian Berita	78
Gambar 3.33 Desain Antarmuka Halaman Utama Pencarian Berita (Pengguna Umum) ...	79
Gambar 3.34 Desain Antarmuka Halaman Hasil Pencarian Berita (Pengguna Umum)	79

Gambar 4.1 Antarmuka Halaman <i>Login</i>	83
Gambar 4.2 Antarmuka Halaman <i>Crawling</i> Berita.....	84
Gambar 4.3 Antarmuka Halaman Mengelola Berita.....	85
Gambar 4.4 Antarmuka Halaman Mengelola Berita (Detail).....	85
Gambar 4.5 Antarmuka Halaman Mengelola Berita (<i>Delete</i>).....	85
Gambar 4.6 Antarmuka Peringkasa Berita.....	86
Gambar 4.7 Antarmuka Halaman Proses Peringkasan Berita.....	86
Gambar 4.8 Antarmuka Halaman Pencarian Berita (<i>Indexing</i>).....	87
Gambar 4.9 Antarmuka Halaman Pencarian Berita (Pengelola Berita).....	87
Gambar 4.10 Antarmuka Halaman Utama Pencarian Berita (Pengguna Umum).....	88
Gambar 4.11 Antarmuka Halaman Hasil Pencarian Berita (Pengguna Umum).....	88
Gambar 4.12 Grafik Jumlah Data Berita Hasil <i>Crawling</i> Situs Berita <i>Online</i>	91
Gambar 4.13 Skenario Eksperimen.....	92
Gambar 4.14 Grafik Nilai <i>Precision</i> , <i>Recall</i> dan <i>F-score</i>	94
Gambar 4.16 Grafik Eksperimen 2 Nilai MAP.....	95
Gambar 4.18 Grafik Nilai MAP Korpus Berita dan Ringkasan.....	96
Gambar 4.19 Grafik Waktu Eksekusi Korpus Berita dan Ringkasan.....	97

DAFTAR TABEL

Tabel 2.1 Perkembangan Penelitian <i>Summarization</i>	6
Tabel 2.2 Perkembangan Penelitian <i>Information Retrieval</i>	6
Tabel 2.3 Kombinasi awalan-akhiran yang dilarang	13
Tabel 2.4 Aturan Pemenggalan Awalan Stemmer Nazief dan Adriani	15
Tabel 2.5 Modifikasi Dan Penambahan Aturan Pemenggalan Awalan Oleh Algoritma <i>Stemming Confix Stripping</i>	16
Tabel 2.6 Daftar Aturan <i>Rule Precedence</i>	16
Tabel 2.7 Modifikasi aturan pemenggalan awalan dan penambahan aturan pemenggalan sisipan oleh algoritma stemming modified <i>Enhanced Confix Stripping</i>	17
Tabel 2.8 Penambahan dan modifikasi aturan pemenggalan awalan <i>stemmer</i> Sastrawi.....	17
Tabel 2.9 Perhiungan <i>Recall</i> dan <i>Precision</i>	25
Tabel 2.10 Komponen <i>Use Case Diagram</i>	29
Tabel 3.1 Hasil <i>Sentence Splitting</i>	35
Tabel 3.2 Hasil <i>Stemming</i> dari Contoh 3.1	42
Tabel 3.3 Pembobotan TF-IDF dari Tabel 3.2	44
Tabel 3.4 Matriks <i>Term-By-Document</i> dari Tabel 3.3.....	45
Tabel 3.5 Matriks U dari Tabel 3.4	47
Tabel 3.6 Matriks S dari Tabel 3.4	47
Tabel 3.7 Matriks V^T dari Tabel 3.4.....	48
Tabel 3.8 Hasil Perubahan Nilai Matriks V^T	49
Tabel 3.9 Hasil Peringkasan	50
Tabel 3.10 Contoh Proses Pembobotan TF-IDF	51
Tabel 3.11 Matriks <i>Term-By-Document</i> dari Tabel 3.10.....	51
Tabel 3.12 Matriks U dari Tabel 3.11	52
Tabel 3.13 Matriks S dari Tabel 3.11	52
Tabel 3.14 Matriks V^T dari Tabel 3.11	53
Tabel 3.15 Hasil Perkalian Query Vector dengan Matriks U dan S^{-1}	54
Tabel 3.16 Vektor Dokumen dari Tabel 3.14.....	55
Tabel 3.17 Karakteristik Pengguna	57
Tabel 3.18 Kebutuhan Fungsional.....	58
Tabel 3.19 Kebutuhan Non Fungsional.....	58
Tabel 3.20 <i>Use Case</i> SPBP BOBI.....	59
Tabel 3.21 <i>Use Case</i> Melakukan Login	60
Tabel 3.22 <i>Use Case</i> Crawling Berita	61
Tabel 3.23 <i>Use Case</i> Mengelola Berita	61
Tabel 3.24 <i>Use Case</i> Meringkas Berita	62
Tabel 3.25 <i>Use Case</i> Mencari Berita.....	62
Tabel 3.26 Identifikasi <i>Class Analysis</i>	63
Tabel 3.27 <i>Entity Objects</i>	64
Tabel 3.28 <i>Boundary Objects</i>	64

Tabel 3.29 <i>Control Objects</i>	65
Tabel 3.30 Desain Tabel <i>Login</i>	72
Tabel 3.31 Desain Tabel Situs.....	72
Tabel 3.32 Desain Tabel Berita	72
Tabel 3.33 Desain Tabel Ringkasan	73
Tabel 4.1 Implementasi <i>Class</i>	82
Tabel 4.2 Rencana Pengujian Fungsional Sistem.....	89

BAB I

PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup tugas akhir mengenai Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia Menggunakan Metode *Latent Semantic Indexing* (LSI).

1.1. Latar Belakang

Informasi menjadi kebutuhan pokok masyarakat saat ini karena dengan informasi masyarakat dapat mengetahui segala peristiwa yang terjadi di dunia. Informasi kini bisa didapatkan melalui internet yang bisa diakses dimanapun dan kapanpun. Berita *online* merupakan salah satu media yang menyajikan informasi terkini dengan berbagai topik berita. Media cetak besar yang ada di Indonesia sekarang mulai memanfaatkan teknologi informasi dengan membuat portal berita *online*. Menurut *Alexa Rank*, situs berita yang menempati peringkat 10 besar situs di Indonesia yang paling banyak pengunjungnya diantaranya Detik.com, Tribunnews.com dan Kompas.com (Alexa, 2016). *Alexa Rank* adalah proses penilaian mengenai kualitas suatu situs *web* atau *blog* yang dilakukan oleh situs alexa.com berdasarkan tolak ukur tertentu seperti jumlah trafik pengunjung maupun kualitas kontennya, lalu diaplikasikan menjadi pemeringkatan atau susunan peringkat/ ranking berupa *widget Alexa Rank* (Widayanti & Dwi, 2015).

Jumlah berita *online* yang tersebar di internet semakin bertambah banyak. selain itu banyaknya situs berita yang terdaftar pada *www (world wide web)* memungkinkan hasil pencarian berita menjadi sangat luas. Situs-situs tersebut belum tentu memberikan sumber berita yang jelas, sehingga pengguna harus memilah berita dari beberapa situs yang terpercaya. Untuk itu, diperlukan suatu sistem pencarian berita dari situs berita nasional yang sudah terpercaya dan populer di masyarakat. Sistem pencarian tersebut menghasilkan beberapa berita yang diurutkan berdasarkan tingkat relevansi terhadap *query*. Relevan atau tidaknya berita akan diketahui secara jelas setelah pengguna membaca satu persatu isi berita yang disajikan sehingga membutuhkan waktu yang banyak. Oleh karena itu dibutuhkan peringkasan berita untuk mempersingkat waktu membaca dan mempermudah pengguna memilah berita yang diinginkan tanpa harus

membaca keseluruhan isi berita. Sistem yang dibutuhkan adalah Sistem Temu-Balik Informasi (*Information retrieval / IR*)

IR adalah proses yang berhubungan dengan representasi, penyimpanan, pencarian, dan pemanggilan informasi yang relevan dengan kebutuhan informasi yang diinginkan pengguna (Ingwersen, 1992). IR memiliki tiga model untuk pencarian teks tidak terstruktur, yaitu model *boolean*, *vector* dan *probabilistic*. Kelebihan dari model *vector* adalah lebih sederhana dan mudah direpresentasikan karena menggunakan *index term* untuk mengindeks dan *me-retrieve* dokumen. Model *vector* terdiri atas *Generalized vector*, *Latent Semantic Indexing* dan *Neural network* (Baeza-Yates & Riberio-Neto, 1999).

Penelitian tentang IR sudah banyak dilakukan seperti penelitian yang dilakukan Fatkhul Amin (2012) untuk pencarian dokumen teks menggunakan metode VSM menghasilkan akurasi rata-rata *recall* dan *precision* masing-masing 0,19 dan 0,54. Kelemahan pada metode VSM adalah menganggap bahwa setiap *term* pada dokumen bersifat independen, yaitu metode ini tidak melihat hubungan makna dengan *term* lain (Wibowo, et al., 2012). Hal ini memungkinkan pencarian hanya berdasarkan kemiripan kata saja tanpa memperdulikan hubungan semantik tiap kata. Padahal banyak kata memiliki kesamaan arti (sinonim) dan kata yang memiliki arti lebih dari satu (polisemi). Masalah lain pada VSM adalah besarnya ruang vektor atau besarnya dimensi pada matrik *term-document* sehingga mengakibatkan penurunan kinerja (Supriyanto & Affandy, 2011). Untuk mengatasi masalah tersebut dibutuhkan metode yang mampu menangani kemiripan kata secara semantik dan menggunakan dimensi matrik yang lebih kecil.

Metode *Latent Semantic Analysis* (LSA) adalah salah satu metode pemrosesan teks berbasis semantik yang bisa digunakan dalam IR dan peringkasan teks. LSA merupakan teknik matematika/statistika untuk mengekstraksi dan menyimpulkan hubungan kontekstual arti kata yang diaplikasikan pada bagian teks yang dibutuhkan (Aji, et al., 2011). Metode LSA dalam konteks aplikasi untuk pencarian informasi (*Information Retrieval*) kadang – kadang juga disebut sebagai *Latent Semantic Indexing* (LSI) (Zulhanif & Anindya, 2015). LSI adalah metode *indexing* pada *information retrieval* yang menggunakan teknik *singular value decomposition* (SVD) untuk mengidentifikasi makna semantik kata-kata berdasarkan pola dan hubungan antara istilah dan konsep-konsep yang terkandung dalam koleksi teks (Wardhana, et

al., 2015). Metode LSI memberikan solusi untuk masalah kata-kata sinonim dan polisemi yang sering terjadi dalam sistem temu balik informasi (Muhammad, et al., 2011). SVD juga digunakan untuk mengatasi masalah besarnya dimensi matrik pada VSM, yaitu dengan mengurai matrik *term-document* menjadi matrik yang berdimensi lebih kecil. SVD ini mempunyai kapasitas reduksi *noise* yang membantu untuk meningkatkan akurasi (Jamrahi, et al., 2014). Oleh karena itu LSI sangat cocok dipakai pada mesin pencari karena meningkatkan relevansi dokumen dibanding menggunakan metode VSM. Selain pencarian teks, metode tersebut juga digunakan untuk membuat ringkasan dokumen berdasarkan keterkaitan semantik antar kata untuk menentukan kalimat penting setiap paragraf. Penelitian yang sudah pernah dilakukan Muhammad (2011) menggunakan algoritma LSI dalam proses IR dapat memberikan hasil pencarian yang relevan dengan nilai *recall* 96,67 % dan *precision* 55,48 % pada batas ambang 0.6. Sedangkan penelitian untuk peringkasan dokumen berita dengan metode CMLSA oleh Winata dan Rainarli (2016) menghasilkan nilai *recall* 66,7 %, *precision* 72,25 % dan *F-Measure* 69,6 %.

Oleh karena itu, pada penelitian tugas akhir ini akan dikembangkan sistem pencarian dan peringkasan berita *online* berbahasa Indonesia menggunakan metode *Latent Semantic Indexing* (LSI). Berita-berita yang disajikan dalam sistem ini bersumber dari situs berita *online* populer di Indonesia. Sistem ini diharapkan mampu menjadi mesin pencari dengan hasil penelusuran berita yang memiliki relevansi tinggi serta melakukan peringkasan berita secara otomatis untuk memudahkan pengguna dalam pencarian informasi.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah disampaikan pada latar belakang, rumusan masalah dalam penelitian ini yaitu bagaimana membuat Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia Menggunakan Metode *Latent Semantic Indexing* (LSI).

1.3. Tujuan dan Manfaat

Penelitian Tugas Akhir ini bertujuan untuk menghasilkan Sistem Pencarian dan Peringkasan Berita *Online* Berbahasa Indonesia Menggunakan Metode *Latent Semantic Indexing* (LSI). Adapun tujuan khusus dari penelitian ini adalah untuk membandingkan sistem pencarian dengan peringkasan dan tanpa peringkasan.

Diharapkan sistem ini bisa bermanfaat membantu pengguna dalam mencari berita secara cepat dan akurat.

1.4. Ruang Lingkup

Ruang lingkup yang akan dibahas dalam penelitian ini adalah:

1. Menggunakan data inputan berita dari situs berita *online* Indonesia.
2. Sistem menghasilkan tipe ringkasan ekstraktif.
3. Sistem dapat menampilkan urutan berita sesuai dengan *query* masukan.
4. Sistem dibangun berbasis *web* dengan bahasa pemrograman PHP dan DBMS MySQL.

1.5. Sistematika Penulisan

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

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup dan sistematika penulisan dalam pembuatan tugas akhir mengenai pembuatan sistem pencarian dan peringkasan berita *online* berbahasa indonesia menggunakan metode *Latent Semantic Indexing* (LSI).

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan tinjauan pustaka yang berhubungan dengan topik tugas akhir. Tinjauan pustaka yang digunakan dalam penyusunan tugas akhir ini meliputi berita, peringkasan teks otomatis, sistem temu-kembali

informasi, *text mining*, *Latent Semantic Indexing* (LSI) dan model pengembangan perangkat lunak.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai langkah-langkah yang dilakukan pada penelitian Tugas Akhir. Langkah-langkah tersebut diawali dengan gambaran umum penelitian kemudian diikuti dengan proses dari peringkasan dan sistem temu balik informasi berita. Pada bab ini juga menjelaskan tentang Analisis dan Desain Sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas mengenai hasil pengembangan sistem berdasarkan analisa dan desain yang dijelaskan pada bab sebelumnya. Selain itu, bab ini juga membahas tentang skenario pengujian baik pengujian fungsional maupun pengujian kinerja beserta dengan hasilnya.

BAB V PENUTUP

Bab ini membahas mengenai kesimpulan dari uraian yang telah dijabarkan pada bab-bab sebelumnya dan hasil dari eksperimen yang telah dilakukan serta saran untuk pengembalian penelitian lebih lanjut.