

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan dari pembahasan-pembahasan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan dari Tesis Pembangunan Perangkat Lunak Peringkasan Dokumen dari Banyak Sumber Berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF, sebagai berikut:

1. Penerapan Metode TF-IDF dapat digunakan untuk mengelompokkan kata di dalam sebuah dokumen/artikel sehingga dapat dihasilkan sebuah ringkasan yang sesuai dengan dokumen yang diringkaskan. Penggunaan Metode TF-IDF ini membantu proses peringkasan dokumen menjadi lebih cepat dan mudah.
2. Perangkat Lunak Peringkasan Dokumen dari Banyak Sumber Berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF telah berhasil dibangun. Sistem ini dibuat berbasis *website* dengan menggunakan bahasa pemrograman PHP sehingga proses komputasi yang dijalankan dapat berjalan lebih cepat.

5.2 Saran

Beberapa saran dan masukan yang dapat disampaikan penulis terhadap pembangunan perangkat lunak peringkasan dokumen dari banyak sumber berbasis

web menggunakan *sentence scoring* dengan metode TF-IDF ini dimasa yang akan datang adalah :

1. Kalimat yang diambil dari masing-masing dokumen sebaiknya dicek kemiripannya agar tidak terjadi duplikasi kalimat.
2. Pengambilan kalimat yang akan digunakan sebagai ringkasan akhir sebaiknya diambil dari masing-masing dokumen sumber sehingga inti dari masing-masing dokumen terlihat pada ringkasan akhir.
3. Penambahan metode untuk menggabungkan kalimat agar kesinambungan antar kalimat dapat menjadi lebih baik lagi.

DAFTAR PUSTAKA

- Aizawa, A. 2003. An information-theoretic perspective of TF-IDF measures. *Inf. Process. Manage.* 39, 1, 45–65.
- Berger, A et al (2000). Bridging the Lexical Chasm: Statistical Approaches to Answer Finding. In *Proc. Int. Conf. Research and Development in Information Retrieval*, 192-199.
- Carbonell J, Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. SIGIR*, Melbourne, Australia, Aug. 24-28, 1998, pp.335-336.
- Feldman, R. and Dagan, I. 1995. Knowledge discovery in textual databases (KDT). In *proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, Montreal, Canada, August 20-21, AAAI Press, 112-117.
- Gupta, Virendra Kumar, Tranveer J. Siddiqui, 2012, *Multi-Document Summarization Using Sentence Clustering*, IEEE, India.
- Hearst, M. A. 1997. Text data mining: Issues, techniques, and the relationships to information access. Presentation notes for UW/MS workshop on data mining, July 1997.
- ISC, "ISC Internet Domain Survey, <http://ftp.isc.org/www/survey/reports/current/>, diakses tanggal 28 Agustus 2013.

- Jing H. 2000. Sentence reduction for automatic text summarization. In: Proceedings of 6th conference on applied natural language processing (ANCL'00), pp 310–315.
- Knight K, Marcu D. 2002. Summarization beyond sentence extraction: a probabilistic approach to sentence compression. *Artif Intell* 139(1):91–107.
- Long, Chong, Huang, Min-Lie, Zhu, Xiao-Yan and Li, Ming. 2010. A New Approach for Multi-Document Update Summarization. *JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY* 25(4): 739-749 July 2010. DOI 10.1007/s11390-010-1057-8.
- Mani I . 2001. Automatic summarization. John Benjamin's Publishing Company, Amsterdam.
- Otterbacher JC, Radev DR, Luo A. 2002. Revisions that improve cohesion in multi-document summaries: a preliminary study. In: Proceedings of the ACL02 workshop on automatic summarization, pp 27 – 36.
- Radev, D.R. and Fan, W. 2000, “Automatic summarization of search engine hit lists”. Proceedings of the ACL-2000 workshop on recent advances in natural language processing and information retrieval, Hong Kong, 2000, pp. 99-109.
- Radev, D. R., Hovy, E., and McKeown, K. 2002. “Introduction to the special issue on summarization. *Computational Linguistics*”, 28(4):399-408.

- Radev, D. R., Jing, H., & Budzikowska, M. (2000). Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In ANLP/NAACL Workshop on Summarization Seattle, WA.
- Radev D R, Jing H, Stys M, Tam D. Centroid-based summarization of multiple documents. *Information Processing and Management*, 2004, 40(6): 919-938.
- Ramos, Juan. 2000. Using TF-IDF to Determine Word Relevance in Document Queries. Department of Computer Science, Rutgers University, 23515 BPO Way, Piscataway, NJ, 08855.
- Salton, G., Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24(5):513–523.
- Salton, G., Yang, C.S. 1973. On the specification of term values in automatic indexing. *J. Documentation* 29(4):351–372.
- Simoudis, E. 1996. Reality check for data mining. *IEEE Expert*, 11(5).
- Tan, Ah-Hwee. Text mining: The state of the art and the challenges. In *proceedings, PAKDD Workshop on Knowledge discovery from Advanced Databases (KDAD'99)*, pp. 71-76, Beijing, April 1999.
- Turner, J. and E. Charniak, “Supervised and unsupervised learning for sentence compression,” in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ser. ACL '05. Stroudsburg, PA, USA:

Association for Computational Linguistics, 2005, pp. 290– 297. [Online].

Available: <http://dx.doi.org/10.3115/1219840.1219876>.

Wu, H. C., Luk, R. W. P., Wong, K. F., and Kwok, K. L. 2008. Interpreting TF-IDF term weights as making relevance decisions. *ACM Trans. Inform. Syst.*

26, 3, Article 13 (June 2008), 37 pages.

<http://doi.acm.org/10.1145/1361684.136168>.

Zajic D., B. J. Dorr, J. Lin, and R. Schwartz, “Multi-candidate reduction: Sentence compression as a tool for document summarization tasks,” *Inf. Process. Manage.*, vol. 43, pp. 1549–1570, November 2007. [Online]. Available:

<http://portal.acm.org/citation.cfm?id=1284916.1285161>.

SKPL

SPEKIFIKASI KEBUTUHAN PERANGKAT LUNAK


**Pembangunan Perangkat Lunak Peringkat Dokumen
dari Banyak Sumber berbasis Web menggunakan
Sentence Scoring dengan Metode TF-IDF**

Dipersiapkan oleh:

Fabianus Hendy Evan / 125301915

Program Studi Magister Teknik Informatika

Universitas Atma Jaya Yogyakarta

	Program Studi Magister Teknik Informatika	Nomor Dokumen		Halaman
		<i>SKPL-DOSUM</i>		<i>1/22</i>

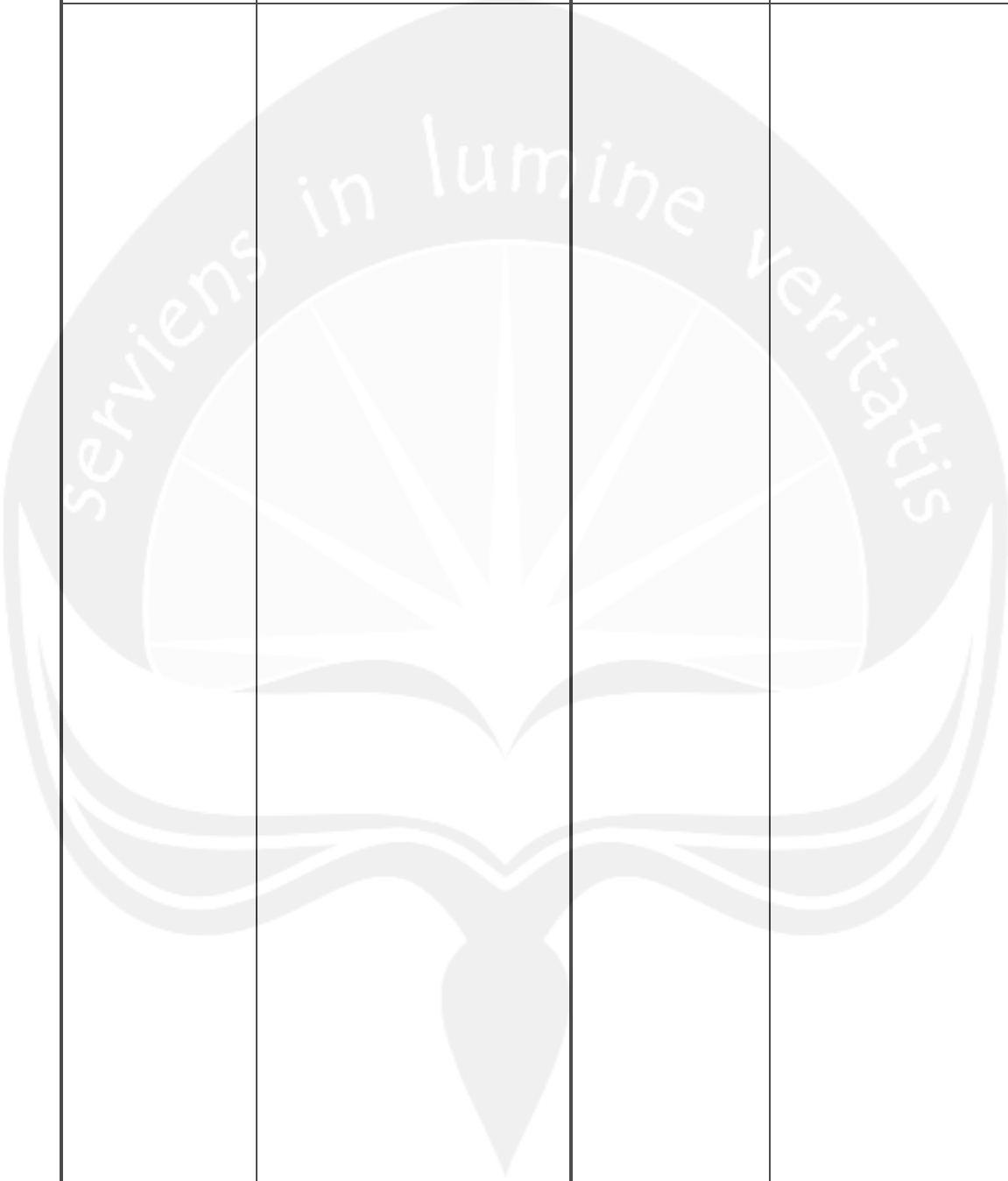
DAFTAR PERUBAHAN

Revisi	Deskripsi
A	
B	
C	
D	
E	
F	

INDEX TGL	-	A	B	C	D	E	F	G
Ditulis oleh	FHE							
Diperiksa oleh								
Disetujui oleh								

Daftar Halaman Perubahan

Halaman	Revisi	Halaman	Revisi



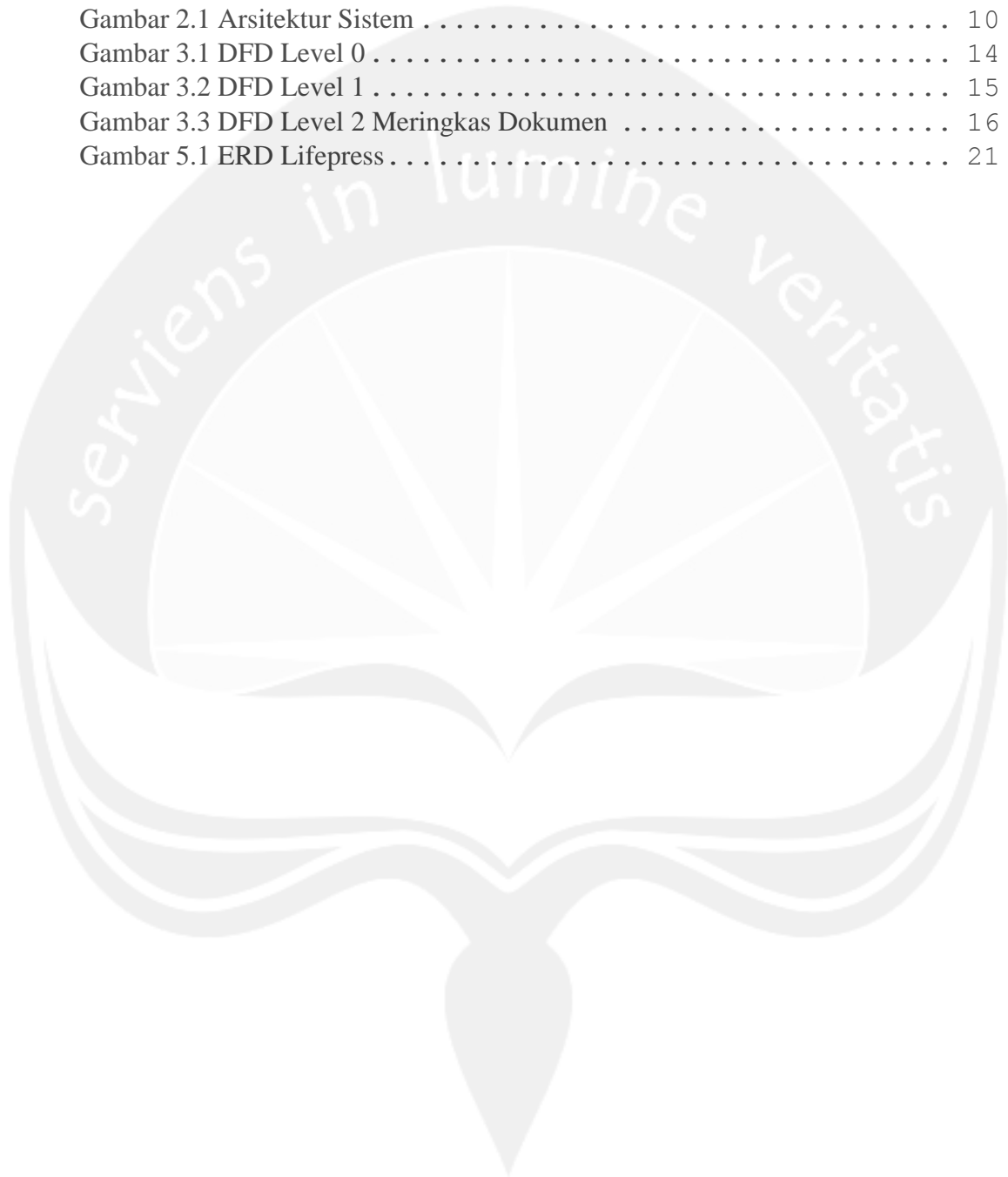
Daftar Isi

1	PENDAHULUAN	7
1.1	Tujuan	7
1.2	Lingkup Masalah	7
1.3	Definisi, Akronim dan Singkatan	7
1.4	Referensi	8
1.5	Deskripsi umum (Overview)	8
2	DESKRIPSI KEBUTUHAN	9
2.1	Perspektif produk	9
2.2	Fungsi Produk	11
2.3	Karakteristik Pengguna	11
2.4	Batasan-batasan	12
2.5	Asumsi dan Ketergantungan	12
3	KEBUTUHAN KHUSUS	12
3.1	Kebutuhan antarmuka eksternal	12
3.1.1	Antarmuka pemakai	12
3.1.2	Antarmuka perangkat keras	12
3.1.3	Antarmuka perangkat lunak	13
3.2	Kebutuhan fungsionalitas	14
3.2.1	Data Flow Diagram	14
3.2.1.1	DFD Level 0 DOSUM	14
3.2.1.1.1	Entitas Data	14
3.2.1.1.2	Proses	14
3.2.1.1.3	Topologi	14
3.2.1.2	DFD Level 1 DOSUM	15
3.2.1.2.1	Entitas data masukan	15
3.2.1.2.2	Proses	15
3.2.1.2.3	Topologi	15
3.2.1.3	DFD Level 2 Meringkas Dokumen	15
3.2.1.3.1	Entitas Data Masukan	15
3.2.1.3.2	Proses	16
3.2.1.3.3	Topologi	16
3.2.2	Deskripsi Proses	17
3.2.2.1	Proses Meringkas Dokumen	17
3.2.2.1.1	Entitas data masukan	17
3.2.2.1.2	Algoritma atau formula dari proses	17
3.2.2.1.3	Entitas data terlibat	17
3.2.2.2	Proses Membagi Kalimat	17
3.2.2.2.1	Entitas data masukan	17
3.2.2.2.2	Algoritma atau formula dari proses	17
3.2.2.2.3	Entitas data terlibat	17

3.2.2.3	Proses Mencari Dokumen.....	17
3.2.2.3.1	Entitas data masukan	17
3.2.2.3.2	Algoritma atau formula dari proses	18
3.2.2.3.3	Entitas data terlibat	18
3.2.2.4	Proses Membagi Kata	18
3.2.2.4.1	Entitas data masukan	18
3.2.2.4.2	Algoritma atau formula dari proses	18
3.2.2.4.3	Entitas data terlibat	18
3.2.2.5	Proses Pembobotan Kata	18
3.2.2.5.1	Entitas data masukan	18
3.2.2.5.2	Algoritma atau formula dari proses	18
3.2.2.5.3	Entitas data terlibat	18
3.2.2.6	Proses Perhitungan Skor Kalimat	19
3.2.2.6.1	Entitas data masukan	19
3.2.2.6.2	Algoritma atau formula dari proses	19
3.2.2.6.3	Entitas data terlibat	19
3.2.2.7	Proses Pembentukan Ringkasan	19
3.2.2.7.1	Entitas data masukan	19
3.2.2.7.2	Algoritma atau formula dari proses	19
3.2.2.7.3	Entitas data terlibat	19
4	KAMUS DATA.....	19
4.1	Data Ringkasan	19
4.2	Data Dokumen	20
4.3	Data Query	21
5	ENTITY RELATIONSHIP DIAGRAM (ERD).....	21

Daftar Gambar

Gambar 2.1 Arsitektur Sistem	10
Gambar 3.1 DFD Level 0	14
Gambar 3.2 DFD Level 1	15
Gambar 3.3 DFD Level 2 Meringkas Dokumen	16
Gambar 5.1 ERD Lifepress	21



1 Pendahuluan

1.1 Tujuan

Dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL) ini merupakan dokumen spesifikasi kebutuhan perangkat lunak Peringkasan Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF untuk mendefinisikan kebutuhan perangkat lunak yang akan dikembangkan meliputi antarmuka eksternal (antarmuka antara sistem dengan sistem lain perangkat lunak dan perangkat keras, dan pengguna), tempat penyimpanan yang dibutuhkan, serta keakuratan), dan atribut (*feature-feature* tambahan yang dimiliki sistem), serta mendefinisikan fungsi perangkat lunak. SKPL-DOSUM ini juga mendefinisikan batasan perancangan perangkat lunak.

Dokumen ini digunakan oleh pengembang perangkat lunak sebagai acuan teknis pengembangan perangkat lunak pada tahap selanjutnya.

1.2 Lingkup Masalah

Perangkat Lunak Peringkasan Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF dikembangkan dengan tujuan untuk:

1. Mengelompokkan kalimat dari beberapa dokumen menggunakan metode TF-IDF.
2. Membangun perangkat lunak untuk membuat ringkasan yang bersumber dari satu atau lebih dokumen berbasis web.

1.3 Definisi, Akronim dan Singkatan

Daftar definisi akronim dan singkatan :

Keyword/Phrase	Definisi
SKPL	Merupakan spesifikasi kebutuhan dari perangkat lunak yang akan dikembangkan.
SKPL-DOSUM-XXX	Kode yang merepresentasikan kebutuhan pada DOSUM (Peringkasan Dokumen dari Banyak Sumber berbasis Web

	menggunakan Sentence Scoring dengan Metode TF-IDF) dimana XXX merupakan nomor fungsi produk.
DOSUM	Perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan Sentence Scoring dengan Metode TF-IDF.
<i>Document Summarization</i>	<i>Document Summarization</i> merupakan istilah umum yang dipakai untuk menunjuk sebuah proses penyaringan informasi penting dari dokumen untuk menghasilkan ringkasan yang singkat dan mudah dipahami.

1.4 Referensi

Referensi yang digunakan pada perangkat lunak tersebut adalah:

1. Hendy Evan, Fabianus, *Spesifikasi Kebutuhan Perangkat Lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan Sentence Scoring dengan Metode TF-IDF*. Universitas Atmajaya Yogyakarta, 2014.

1.5 Deskripsi umum (Overview)

Secara umum dokumen SKPL ini terbagi atas 2 bagian utama. Bagian utama berisi penjelasan mengenai dokumen SKPL tersebut yang mencakup tujuan pembuatan SKPL, ruang lingkup masalah dalam pengembangan perangkat lunak tersebut, definisi, referensi dan deskripsi umum tentang dokumen SKPL ini.

Bagian kedua berisi penjelasan umum tentang perangkat lunak Perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF yang akan dikembangkan, mencakup perspektif produk yang akan dikembangkan, fungsi produk perangkat lunak, karakteristik pengguna, batasan dalam penggunaan perangkat lunak dan asumsi yang dipakai dalam pengembangan perangkat lunak Perangkat lunak Peringkat

Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF tersebut.

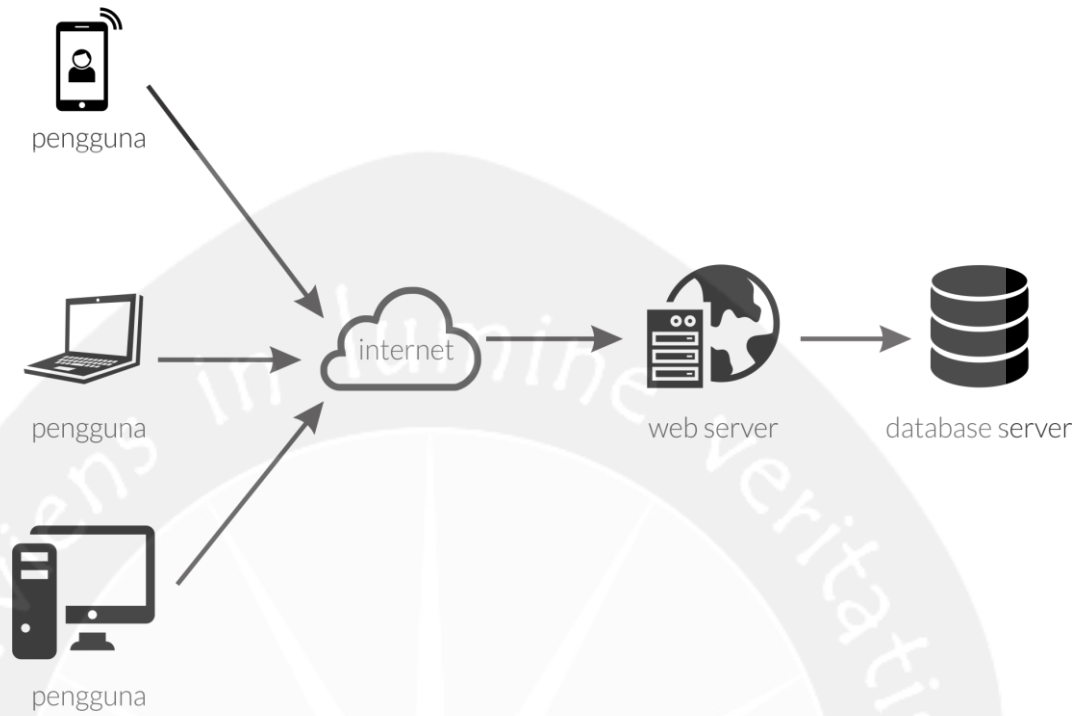
2 Deskripsi Kebutuhan

2.1 Perspektif produk

Perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF merupakan perangkat lunak yang dikembangkan untuk membantu pengguna dalam membuat sebuah ringkasan yang dari banyak dokumen sehingga pengguna dapat memahami isi dokumen dengan lebih mudah.

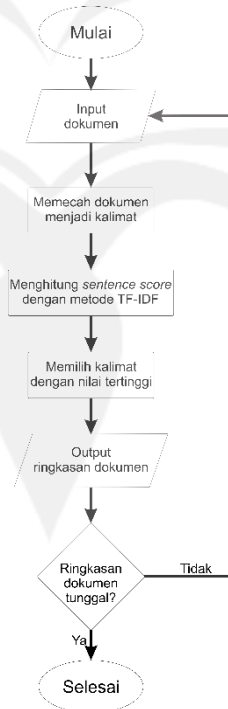
Perangkat lunak ini berjalan pada lingkungan multiplatform berbasis web dan dibuat dengan menggunakan bahasa pemrograman PHP. Sedangkan untuk lingkungan pemrogramannya menggunakan Sublime Text, *framework* yang digunakan untuk membuat aplikasi adalah CodeIgniter.

Pengguna akan berinteraksi dengan sistem melalui antarmuka GUI (*Graphical User Interface*) yang ditampilkan pada Web browser dari masing-masing pengguna. Pengguna dapat mengakses data yang ada di server secara *online* dengan memanggil *web service* pada website yang tersedia di *web server*. Pengguna juga dapat melakukan pencarian data di dalam *database server* yang selanjutnya dikirimkan ke pengguna yang melakukan *request* melalui *web server*.



Gambar 2.1 Arsitektur Sistem

Flow chart perangkat lunak DOSUM ini ditunjukkan pada gambar 2.2 di bawah ini:



Gambar 2.2 Flow Chart DOSUM

Flow chart dimulai ketika terdapat inputan berupa dokumen dari pengguna dengan jumlah minimal 1. Setelah itu masing-masing dokumen yang diinputkan akan dipecah menjadi ke dalam bentuk per kalimat, hal ini dimaksudkan untuk mempermudah dalam pemberian skor dan juga dalam pemilihan kalimat untuk menjadi ringkasan. Kemudian kalimat-kalimat tadi akan dipecah lagi menjadi per kata untuk kemudian diberi bobot menggunakan metode TF-IDF dan skor dari masing-masing kalimat didapat dari penjumlahan bobot kata yang terdapat di dalam masing-masing kalimat. Kalimat tersebut akan dipilih beberapa dengan skor tertinggi untuk menjadi ringkasan. Jika ringkasan yang dihasilkan masih berupa ringkasan dari masing-masing dokumen, maka ringkasan tersebut akan digabungkan menjadi satu dan di proses lagi hingga membentuk ringkasan akhir.

2.2 Fungsi Produk

Fungsi produk perangkat lunak Perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF adalah sebagai berikut:

1. Fungsi *Summarize* (SKPL-DOSUM-001)

Merupakan fungsi yang digunakan oleh pengguna untuk menampilkan hasil ringkasan dari banyak sumber dokumen.

2. Fungsi Cari (SKPL-DOSUM-002)

Merupakan fungsi yang digunakan oleh pengguna untuk mencari artikel dari berbagai media online yang tersimpan di database untuk dijadikan ringkasan.

2.3 Karakteristik Pengguna

Karakteristik dari pengguna perangkat lunak Perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF adalah sebagai berikut :

1. Memahami pengoperasian Personal Komputer.
2. Mengerti tentang internet.
3. Memahami pengoperasian *web browser*.

2.4 Batasan-batasan

Batasan-batasan dalam pengembangan perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF tersebut adalah:

1. Kebijakan Umum

Berpedoman pada tujuan dari pengembangan perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF.

2. Keterbatasan perangkat keras

Dapat diketahui kemudian setelah sistem ini berjalan (sesuai dengan kebutuhan).

2.5 Asumsi dan Ketergantungan

Sistem ini dapat dijalankan di berbagai perangkat keras seperti PC, Netbook, Notebook dan *mobile device* yang mempunyai *web browser*.

3 Kebutuhan Khusus

3.1 Kebutuhan antarmuka eksternal

Kebutuhan antar muka eksternal pada perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF meliputi kebutuhan antarmuka pemakai, antarmuka perangkat keras, antarmuka perangkat lunak, antarmuka komunikasi.

3.1.1 Antarmuka pemakai

Pengguna berinteraksi dengan antarmuka yang ditampilkan dalam bentuk form-form di dalam halaman web.

3.1.2 Antarmuka perangkat keras

Antarmuka perangkat keras yang digunakan dalam perangkat lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan *Sentence Scoring* dengan Metode TF-IDF adalah:

1. Perangkat *mobile device* yang mempunyai *web browser*.
2. Perangkat komputer yang mempunyai *web browser*.

3. Jaringan internet

3.1.3 Antarmuka perangkat lunak

Perangkat lunak yang dibutuhkan untuk mengoperasikan perangkat lunak DOSUM adalah sebagai berikut:

1. Nama : MySQL
Sumber : Sun Microsystem

Sebagai database yang dibutuhkan dalam mengoperasikan perangkat lunak DOSUM. Database management system atau dikenal sebagai basis data yang berfungsi untuk menyimpan data dari sistem. Pembangunan perangkat lunak ini menggunakan database management system dengan nama MySQL dengan versi ke 5.

2. Nama : Apache
Sumber : Apache Software Foundation

Sebagai web server. Web server merupakan sebuah perangkat lunak server yang berfungsi menerima permintaan HTTP atau HTTPS dari klien yang dikenal dengan web browser dan mengirimkan kembali hasilnya dalam bentuk halaman-halaman web yang umumnya berbentuk dokumen HTML dengan menggunakan Apache. Apache adalah salah satu web server yang paling populer, yang dapat menjalankan script dari PHP.

3. Nama : PHP
Sumber : The PHP Group

Sebagai bahasa pemrograman yang digunakan untuk membangun perangkat lunak DOSUM.

4. Nama : Firefox, Chrome, Safari, Opera, Internet Explorer

Sebagai web browser yang digunakan untuk mengakses perangkat lunak DOSUM.

3.2 Kebutuhan fungsionalitas

3.2.1 Data Flow Diagram

3.2.1.1 DFD Level 0 DOSUM

3.2.1.1.1 Entitas Data

Entitas eksternal yang terlibat dalam pengembangan perangkat lunak DOSUM tersebut dinyatakan dalam tabel sebagai berikut:

NAMA	KODE
Pengguna aplikasi DOSUM	Pengguna

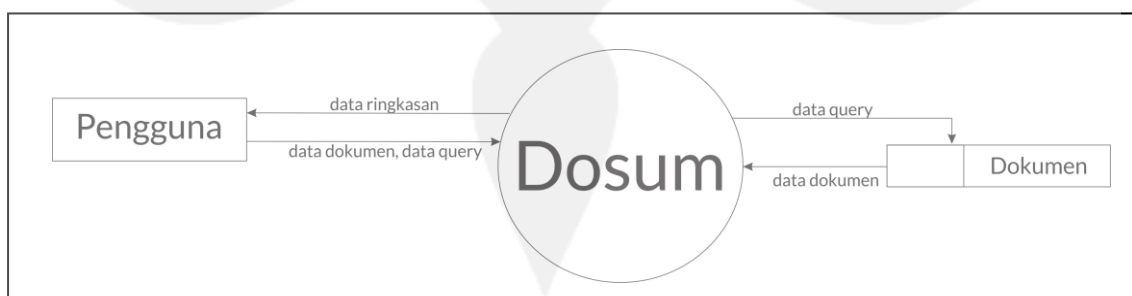
Entitas yang didefinisikan dalam tabel tersebut merupakan entitas yang terlibat dalam proses – proses yang terjadi dalam perangkat lunak DOSUM tersebut.

3.2.1.1.2 Proses

Proses yang terjadi dalam perangkat lunak DOSUM tersebut adalah menerima input dari pengguna yang selanjutnya diproses menjadi informasi yang di kehendaki.

3.2.1.1.3 Topologi

Topologi dari proses perangkat DOSUM dapat dilihat pada Gambar 3.1.



Gambar 3.1 DFD Level 0

3.2.1.2 DFD Level 1 DOSUM

3.2.1.2.1 Entitas data masukan

Entitas data eksternal sesuai dengan entitas data pada DFD level 0.

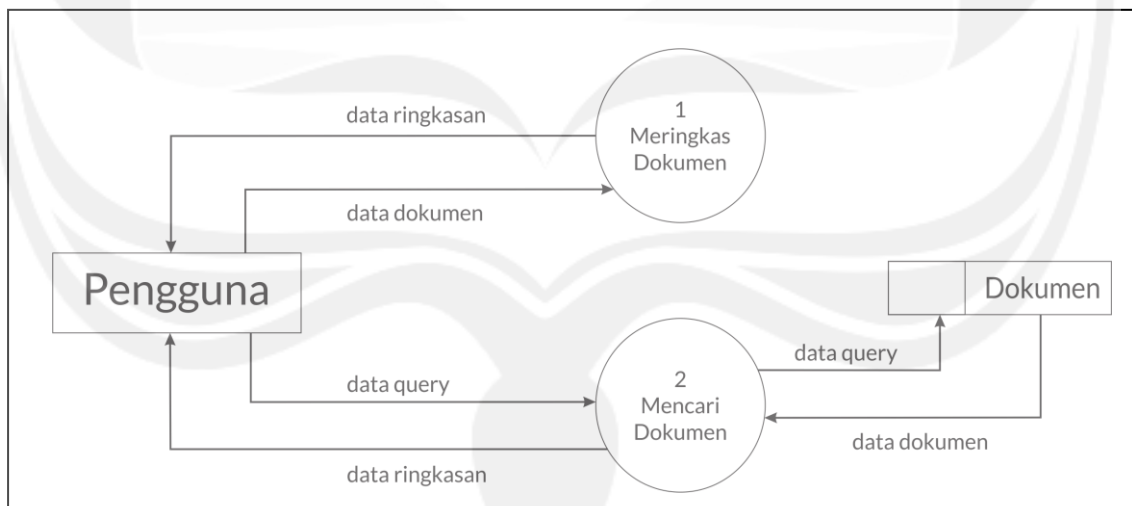
3.2.1.2.2 Proses

Proses yang terjadi dalam DFD Level 1 mencakup 2 proses, antara lain:

1. Meringkas Dokumen, adalah proses untuk meringkas dokumen yang di inputkan pengguna ke dalam sistem.
2. Mencari Dokumen, adalah proses untuk menampilkan dokumen sesuai dengan *query* pencarian yang diinputkan pengguna ke dalam sistem.

3.2.1.2.3 Topologi

Topologi dari proses perangkat lunak DOSUM dapat dilihat pada gambar 3.2.



Gambar 3.2 DFD Level 1

3.2.1.3 DFD Level 2 Meringkas Dokumen

3.2.1.3.1 Entitas Data Masukan

Entitas data eksternal sesuai dengan entitas data pada DFD level 1.

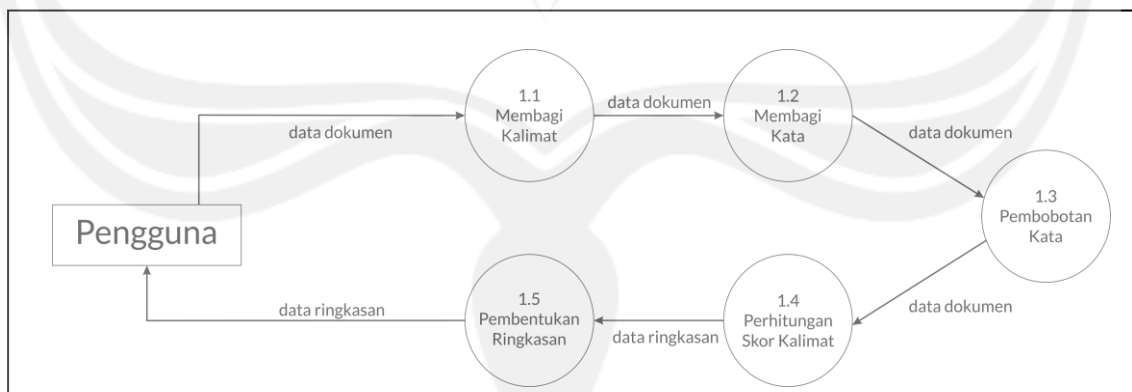
3.2.1.3.2 Proses

Proses yang terjadi dalam DFD Level 2 ini mencakup 5 proses, antara lain:

1. Membagi Kalimat, adalah proses untuk memecah dokumen inputan menjadi kalimat-kalimat sendiri.
2. Membagi Kata, adalah proses untuk membagi kalimat yang sudah dipecah dari dokumen ke dalam bagian kata.
3. Pembobotan Kata, adalah proses untuk memberikan bobot/nilai untuk suatu kata di dalam sebuah dokumen.
4. Perhitungan Skor Kalimat, adalah proses untuk menghitung skor dari masing-masing kalimat dengan cara menjumlahkan bobot/nilai dari kata yang dimiliki masing-masing kalimat.
5. Pembentukan Ringkasan, adalah proses pemilihan beberapa kalimat dengan skor tertinggi untuk kemudian dijadikan sebuah ringkasan.

3.2.1.3.3 Topologi

Topologi dari proses DFD level 2 Meringkas Dokumen perangkat lunak DOSUM dapat dilihat pada gambar 3.3.



Gambar 3.3 DFD Level 2 Meringkas Dokumen

3.2.2 Deskripsi Proses

3.2.2.1 Proses Meringkas Dokumen

3.2.2.1.1 Entitas data masukan

Entitas data masukan dalam proses Meringkas Dokumen yaitu dokumen yang diinputkan oleh pengguna ke dalam sistem.

3.2.2.1.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa dokumen yang berjumlah 1 atau lebih yang akan diproses dan ditampilkan menjadi sebuah ringkasan.

3.2.2.1.3 Entitas data terlibat

Entitas data yang terlibat adalah data ringkasan dan data dokumen.

3.2.2.2 Proses Membagi Kalimat

3.2.2.2.1 Entitas data masukan

Entitas data masukan dalam proses Membagi Kalimat yaitu dokumen yang diinputkan pengguna ke dalam sistem.

3.2.2.2.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa dokumen dari pengguna dan akan dibagi menjadi kalimat-kalimat yang berdiri sendiri.

3.2.2.2.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen.

3.2.2.3 Proses Mencari Dokumen

3.2.2.3.1 Entitas data masukan

Entitas data masukan dalam proses Mencari Dokumen yaitu *query* pencarian yang diinputkan pengguna ke dalam sistem.

3.2.2.3.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa *query* pencarian dan akan ditampilkan hasil dari pencarian dokumen dari basis data.

3.2.2.3.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen dan data *query*.

3.2.2.4 Proses Membagi Kata

3.2.2.4.1 Entitas data masukan

Entitas data masukan dalam proses Membagi Kata yaitu dokumen yang diinputkan pengguna ke dalam sistem.

3.2.2.4.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa dokumen dari pengguna yang sudah dibagi ke dalam bentuk kalimat dan akan dibagi menjadi kata yang berdiri sendiri.

3.2.2.4.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen.

3.2.2.5 Proses Pembobotan Kata

3.2.2.5.1 Entitas data masukan

Entitas data masukan dalam proses Pembobotan Kata yaitu dokumen yang diinputkan pengguna ke dalam sistem.

3.2.2.5.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa dokumen dari pengguna yang sudah dibagi ke dalam bentuk kata. Kata tersebut kemudian akan diberi nilai/bobot menggunakan metode TF-IDF.

3.2.2.5.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen.

3.2.2.6 Proses Perhitungan Skor Kalimat

3.2.2.6.1 Entitas data masukan

Entitas data masukan dalam proses Perhitungan Skor Kalimat yaitu dokumen yang diinputkan pengguna ke dalam sistem.

3.2.2.6.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa dokumen dari pengguna yang sudah dibagi ke dalam bentuk kata. Kata yang sudah diberi bobot/nilai tersebut selanjutnya akan dijumlahkan sesuai dengan susunan kata di dalam sebuah kalimat untuk mendapatkan skor dari masing-masing kalimat.

3.2.2.6.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen.

3.2.2.7 Proses Pembentukan Ringkasan

3.2.2.7.1 Entitas data masukan

Entitas data masukan dalam proses Mencari Dokumen yaitu dokumen yang diinputkan pengguna ke dalam sistem.

3.2.2.7.2 Algoritma atau formula dari proses

Proses tersebut akan menerima masukan berupa kalimat-kalimat dari dokumen yang diinputkan pengguna yang sudah memiliki skor masing-masing. Kalimat tersebut akan dipilih beberapa yang mempunyai skor tertinggi untuk kemudian dijadikan ringkasan.

3.2.2.7.3 Entitas data terlibat

Entitas data yang terlibat adalah data dokumen dan data ringkasan.

4 Kamus Data

4.1 Data Ringkasan

Data Element : sum_all

Definition : *ringkasan akhir dokumen*

Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z|] *Legal Character*
 Range : tidak ada

Data Element : sum_sentence_score
 Definition : *skor kalimat dari ringkasan akhir*
 Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z|] *Legal Character*
 Range : tidak ada

Data Element : sum
 Definition : *ringkasan masing-masing dokumen*
 Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z|] *Legal Character*
 Range : tidak ada

Data Element : sentence_score
 Definition : *skor kalimat dari ringkasan masing-masing dokumen*
 Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z|] *Legal Character*
 Range : tidak ada

4.2 Data Dokumen

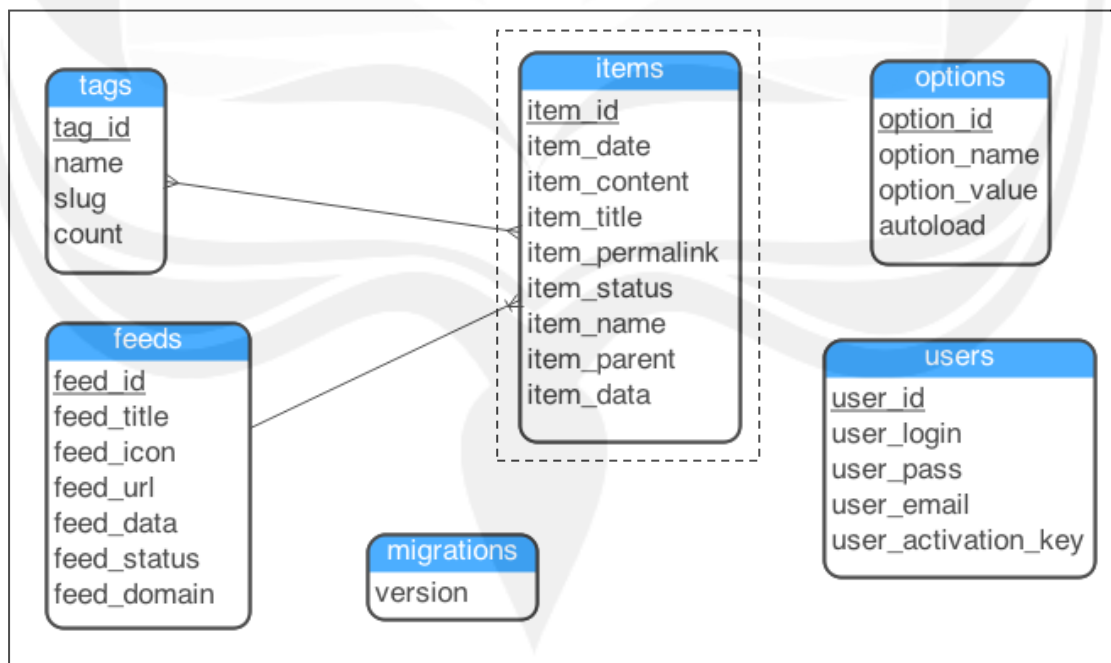
Data Element : doc
 Definition : *dokumen inputan dari pengguna*
 Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z|] *Legal Character*
 Range : tidak ada

4.3 Data Query

Data Element : query
 Definition : *query dari pengguna untuk parameter pencarian dokumen*
 Alias : tidak ada
 Format : LC + LD
 LC : [0|1|...|9] *Legal Digits*
 Range : 005000000 to 995999999
 LC : [A-Z|a-z| |] *Legal Character*
 Range : tidak ada

5 Entity Relationship Diagram (ERD)

ERD (*Entity Relationship Diagram*) di bawah ini merupakan ERD yang menggambarkan hubungan dari entitas yang ada pada perangkat lunak Lifepress. Lifepress merupakan perangkat lunak yang digunakan untuk mengumpulkan artikel-artikel dari media online menggunakan RSS. ERD dari Lifepress ditunjukkan pada gambar 5.1 berikut ini:



Gambar 5.1 ERD Lifepress

Perangkat lunak DOSUM mengakses 1 table dari tabel yang dimiliki oleh Lifepress. Tabel yang digunakan perangkat lunak DOSUM adalah tabel items,

dimana tabel tersebut digunakan untuk menyimpan artikel-artikel ke dalam basis data yang didapat dari media online oleh Lifepress.



DPPL

DESKRIPSI PERANCANGAN PERANGKAT LUNAK


**Pembangunan Perangkat Lunak Peringkat Dokumen
dari Banyak Sumber berbasis Web menggunakan
Sentence Scoring dengan Metode TF-IDF**

Dipersiapkan oleh:

Fabianus Hendy Evan / 125301915

Program Studi Magister Teknik Informatika

Universitas Atma Jaya Yogyakarta

	Program Studi Magister Teknik Informatika	Nomor Dokumen		Halaman
		<i>DPPL-DOSUM</i>		<i>1/14</i>

Program Studi Magister Teknik Informatika	DPPL-DOSUM	1 / 14
Dokumen ini dan informasi yang dimilikinya adalah milik Program Studi Magister Teknik Informatika-UAJY dan bersifat rahasia. Dilarang untuk me-reproduksi dokumen ini tanpa diketahui oleh Program Studi Magister Teknik Informatika		

DAFTAR PERUBAHAN

Revisi	Deskripsi
A	
B	
C	
D	
E	
F	

INDEX TGL	-	A	B	C	D	E	F	G
Ditulis oleh	FHE							
Diperiksa oleh								
Disetujui oleh								

Daftar Halaman Perubahan

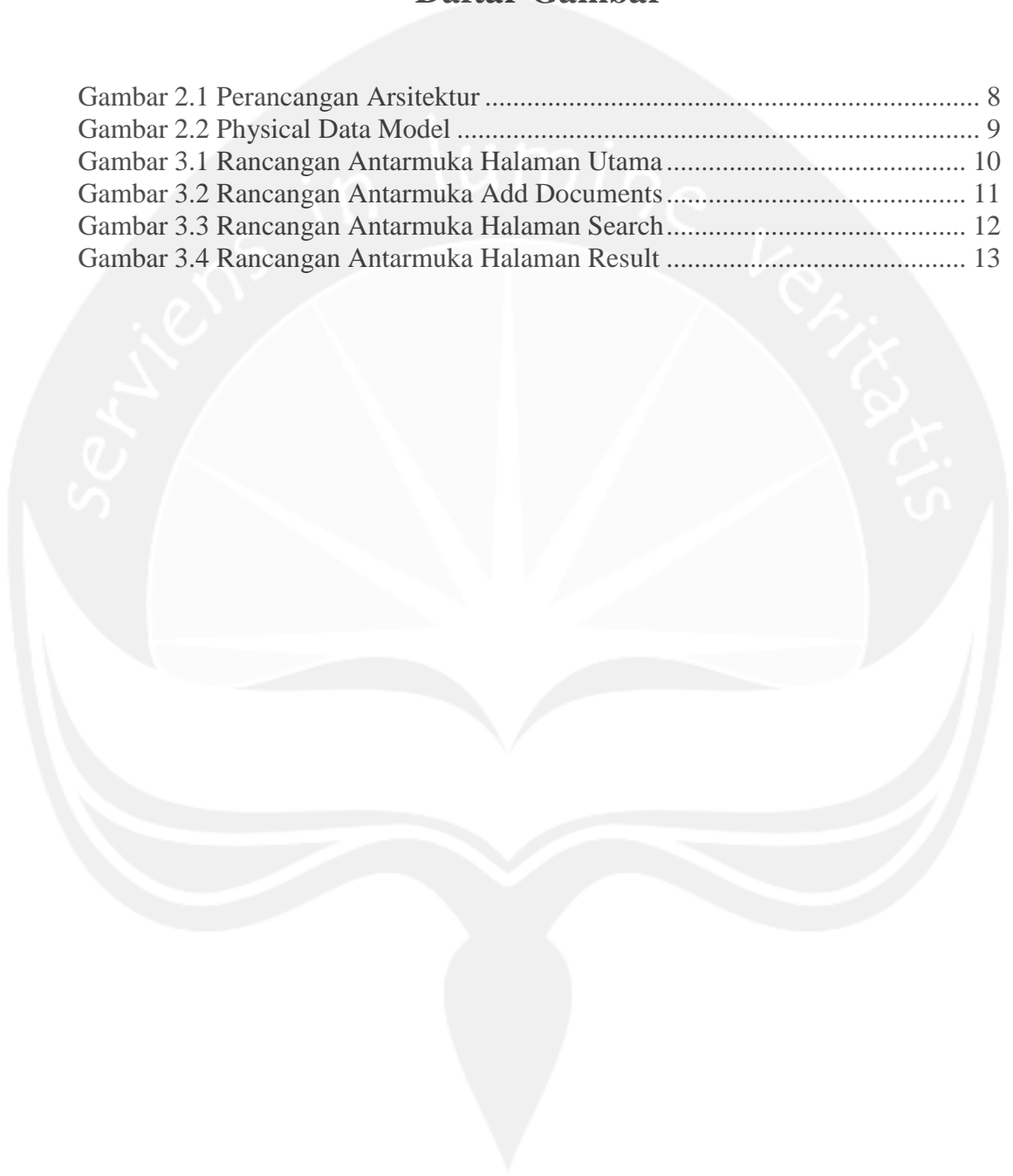
Halaman	Revisi	Halaman	Revisi

Daftar Isi

1 PENDAHULUAN.....	6
1.1 Tujuan	6
1.2 Ruang Lingkup.....	6
1.3 Definisi dan Akronim.....	6
1.4 Referensi	7
2 DESKRIPSI DEKOMPOSISI	8
2.1 Dekomposisi Modul.....	8
2.1.1 Rancangan Arsitektur	8
2.2 Dekomposisi Data	8
2.2.1 Dekomposisi Entitas items	8
2.3 Physical Data Model	9
3 PERANCANGAN ANTARMUKA SISTEM	10
3.1 Halaman Utama.....	10
3.2 Halaman Add Documents	11
3.3 Halaman Search	12
3.4 Halaman Result	13

Daftar Gambar

Gambar 2.1 Perancangan Arsitektur	8
Gambar 2.2 Physical Data Model	9
Gambar 3.1 Rancangan Antarmuka Halaman Utama	10
Gambar 3.2 Rancangan Antarmuka Add Documents	11
Gambar 3.3 Rancangan Antarmuka Halaman Search	12
Gambar 3.4 Rancangan Antarmuka Halaman Result	13



1 Pendahuluan

1.1 Tujuan

Dokumen Deskripsi Perancangan Perangkat Lunak (DPPL) bertujuan untuk mendefinisikan perancangan perangkat lunak yang akan dikembangkan. Dokumen DPPL tersebut digunakan oleh pengembang perangkat lunak sebagai acuan untuk implementasi pada tahap selanjutnya.

1.2 Ruang Lingkup

Perangkat Lunak Peringkat Dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF dikembangkan dengan tujuan untuk:

1. Mengelompokkan kalimat dari beberapa dokumen menggunakan metode TF-IDF.
2. Membangun perangkat lunak untuk membuat ringkasan yang bersumber dari satu atau lebih dokumen berbasis web.

Perangkat lunak ini berjalan pada semua lingkungan *platform* yang memiliki web browser.

1.3 Definisi dan Akronim

Daftar definisi akronim dan singkatan:

Keyword/Phrase	Definisi
DPPL	Deskripsi Perancangan Perangkat Lunak disebut juga <i>Software Design Description</i> (SDD) merupakan deskripsi dari perancangan produk/perangkat lunak yang akan dikembangkan.
DOSUM	Perangkat Lunak Peringkat Dokumen dari Banyak Sumber berbasis Web menggunakan <i>Sentence Scoring</i> dengan Metode TF-IDF.

<p><i>Document Summarization</i></p>	<p><i>Document Summarization</i> merupakan istilah umum yang dipakai untuk menunjuk sebuah proses penyaringan informasi penting dari dokumen untuk menghasilkan ringkasan yang singkat dan mudah dipahami.</p>
--------------------------------------	--

1.4 Referensi

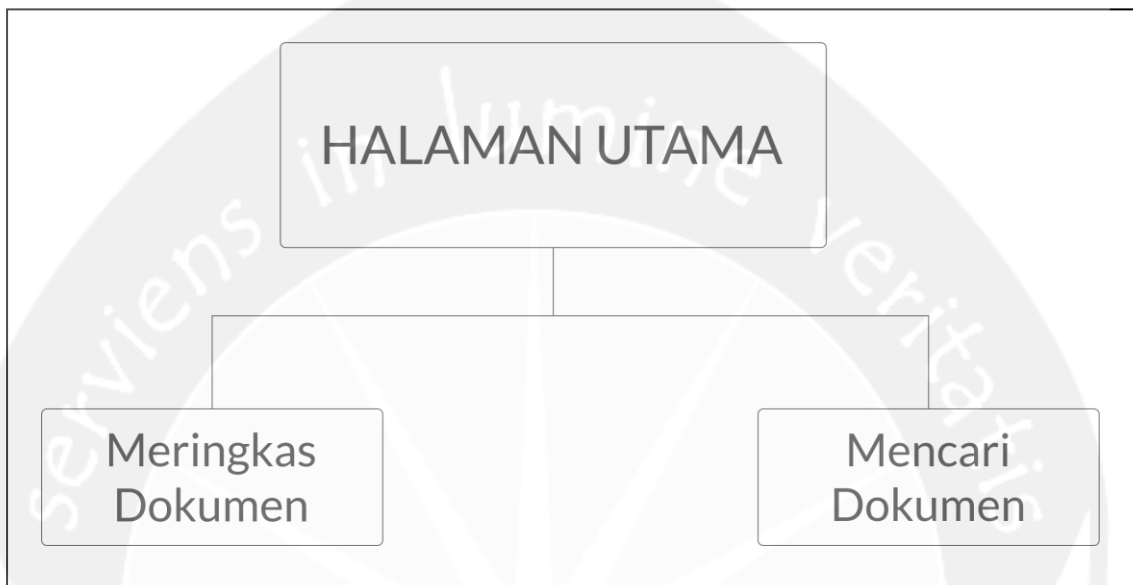
Referensi yang digunakan pada perangkat lunak tersebut adalah:

1. Hendy Evan, Fabianus, *Spesifikasi Kebutuhan Perangkat Lunak Peringkas Dokumen dari Banyak Sumber berbasis Web menggunakan Sentence Scoring dengan Metode TF-IDF*, Universitas Atma Jaya Yogyakarta, 2012.

2 Deskripsi Dekomposisi

2.1 Dekomposisi Modul

2.1.1 Rancangan Arsitektur



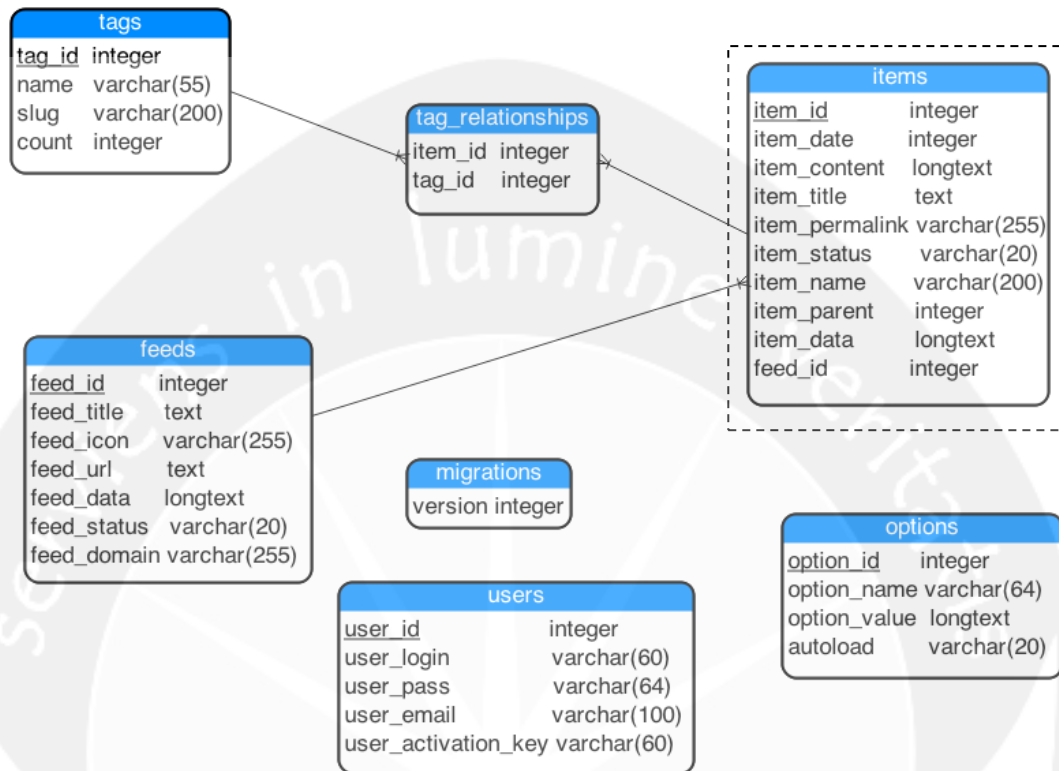
Gambar 2.1 Perancangan Arsitektur

2.2 Dekomposisi Data

2.2.1 Dekomposisi Entitas items

Nama	Tipe	Panjang	Keterangan
item_id	int	11	ID item, Primary Key, nilai digenerate secara otomatis
item_date	int		Tanggal item di simpan
item_content	longtext		Isi dari item
item_title	text		Judul dari item
item_permalink	varchar	255	Alamat url item
item_status	varchar	20	Status item publish/tidak
item_name	varchar	200	Nama item pada url
item_parent	int		Parent dari item
item_data	longtext		

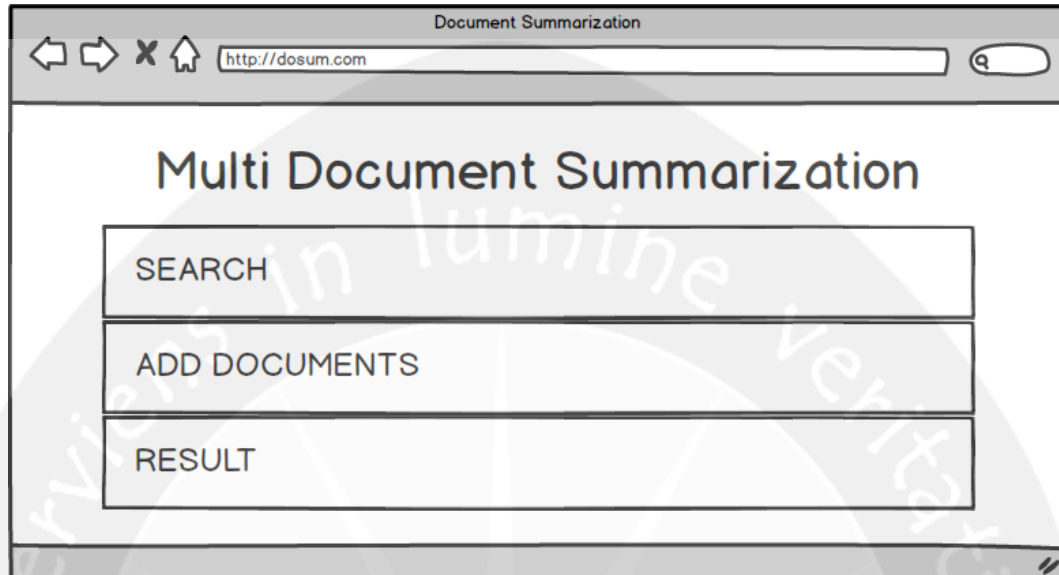
2.3 Physical Data Model



Gambar 2.2 Physical Data Model

3 Perancangan Antarmuka Sistem

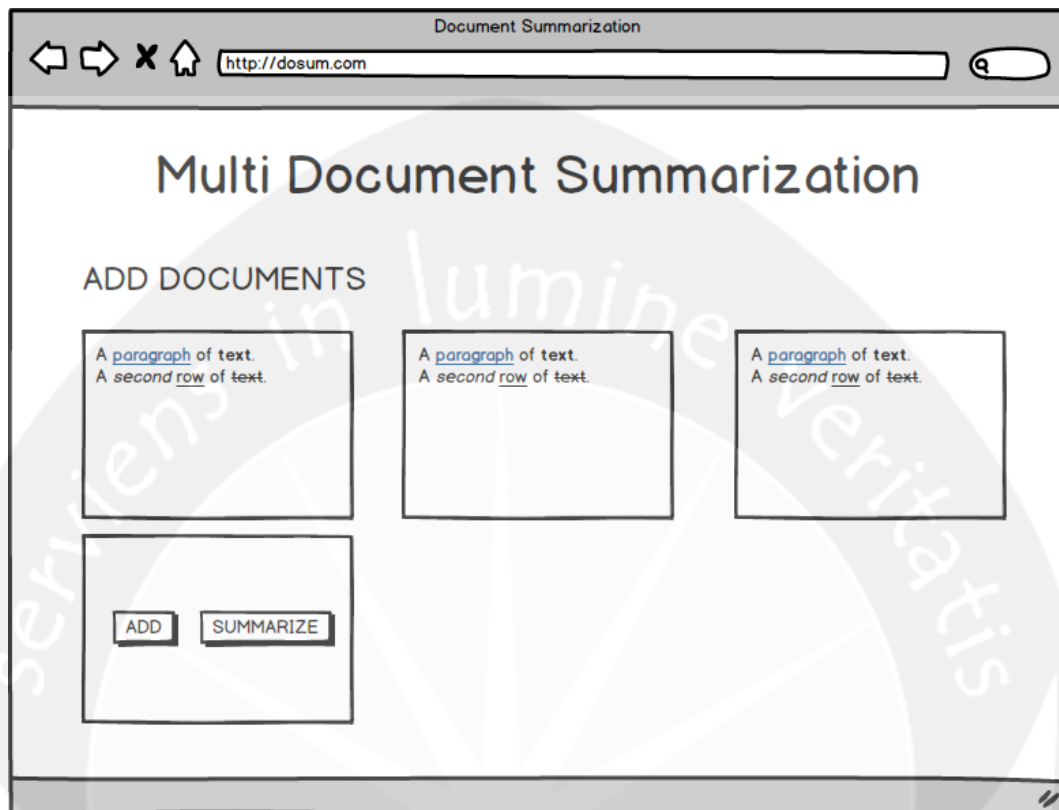
3.1 Halaman Utama



Gambar 3.1 Rancangan Antarmuka Halaman Utama

Gambar 3.1 merupakan halaman utama dari perangkat lunak DOSUM. Pada halaman ini terdapat 3 menu yaitu Search, Add documents dan Result.

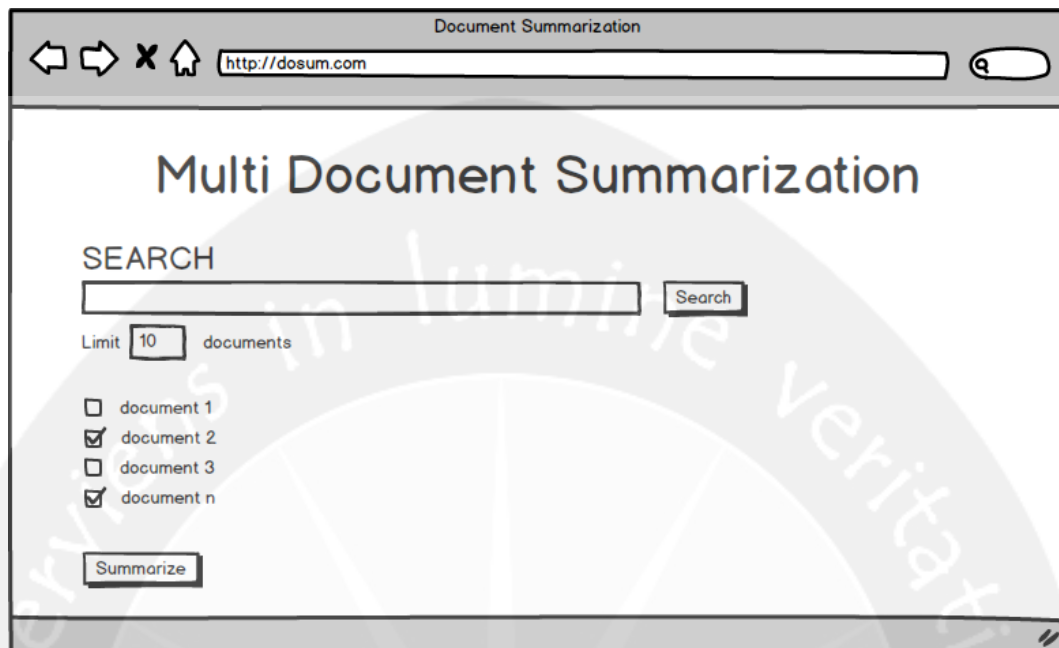
3.2 Halaman Add Documents



Gambar 3.2 Rancangan Antarmuka Add Documents

Gambar 3.2 merupakan halaman menu Add Documents dari perangkat lunak DOSUM yang digunakan untuk menginputkan dari pengguna pada *text area*. Terdapat banyak teks area untuk memungkinkan pengguna menginputkan banyak dokumen. Pengguna juga dapat menambah teks area dengan menekan tombol ADD. Tombol SUMMARIZE digunakan untuk membuat ringkasan dari seluruh dokumen yang diinputkan oleh pengguna.

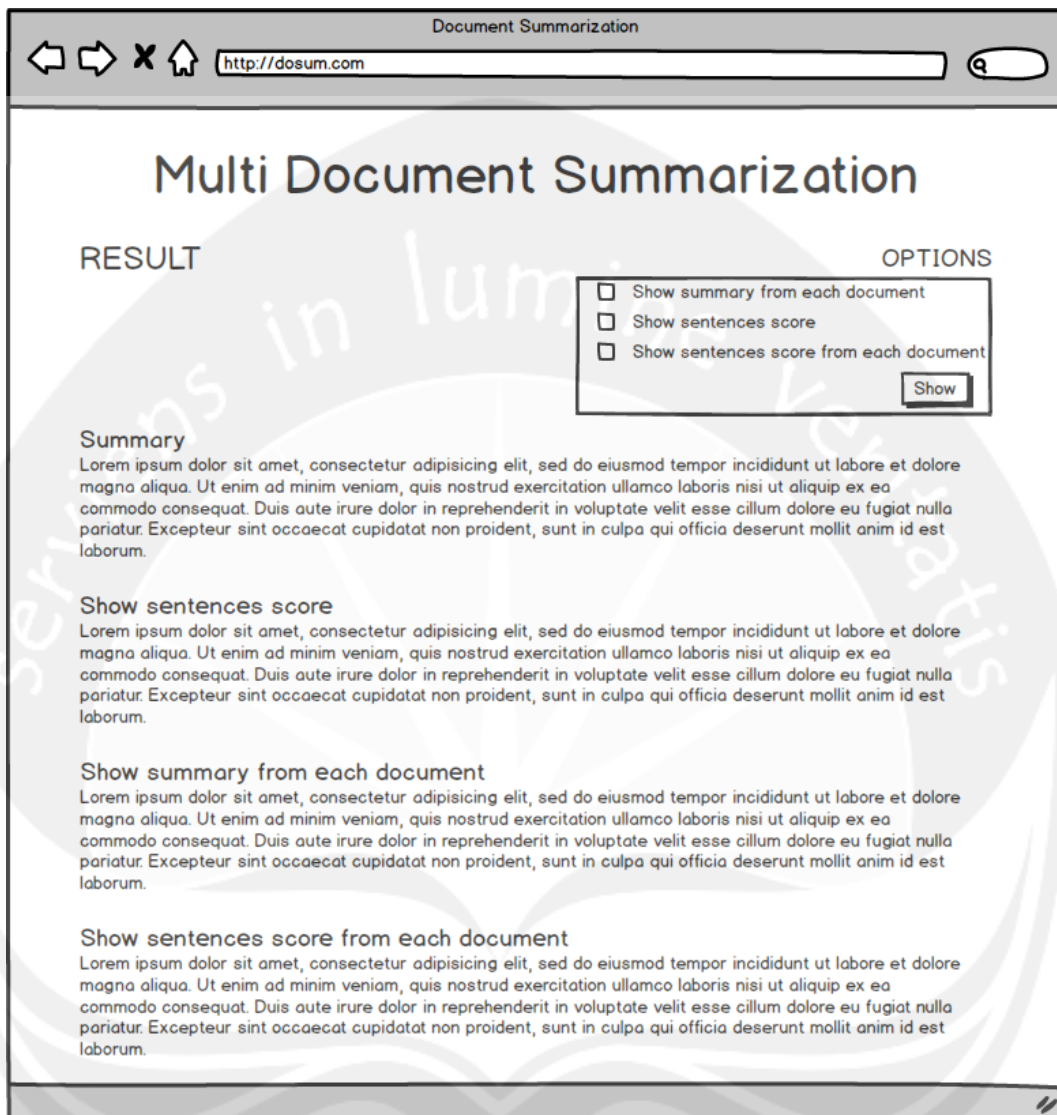
3.3 Halaman Search



Gambar 3.3 Rancangan Antarmuka Halaman Search

Gambar 3.3 merupakan halaman menu Search dari perangkat lunak DOSUM yang digunakan untuk melakukan pencarian dokumen dari database. Pengguna cukup memasukkan kata kunci text input yang tersedia dan menekan tombol search. Limit digunakan untuk memberikan batas dokumen yang akan dicari di dalam database. Pengguna kemudian memilih dokumen mana saja yang akan diringkas berdasar judul dengan mencentang *checkbox* di samping judul dokumen. Kemudian pengguna harus menekan tombol Summarize untuk meringkas dokumen yang telah dipilih tersebut.

3.4 Halaman Result



Gambar 3.4 Rancangan Antarmuka Halaman Result

Gambar 3.4 merupakan halaman menu Result dari perangkat lunak DOSUM yang digunakan untuk menampilkan hasil ringkasan dari dokumen inputan maupun dari dokumen hasil pencarian. Hasil ringkasan akan ditampilkan dalam beberapa bagian yaitu ringkasan akhir (Summary), skor masing-masing kalimat dari ringkasan akhir (Show sentences score), ringkasan masing-masing dokumen (Show summary from each document) dan juga skor masing-masing kalimat dari ringkasan masing-masing dokumen (Show sentences score from each document). Pada antarmuka ini juga terdapat Option, yaitu pilihan untuk

menampilkan atau menyembunyikan bagian Show sentences score, Show summary from each document dan Show sentences score from each document.

