

**APLIKASI PERINGKAS TEKS BAHASA INDONESIA OTOMATIS  
MENGUNAKAN *VECTOR SPACE MODEL* BERBASIS WEB**



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Komputer  
Pada Jurusan Ilmu Komputer/ Informatika**

**Disusun oleh:**

**HARIS DWI ARFIANTO**

**24010311130038**

**JURUSAN ILMU KOMPUTER/ INFORMATIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**2015**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Haris Dwi Arfianto

NIM : 24010311130038

Judul : Aplikasi Peringkat Teks Bahasa Indonesia Otomatis Menggunakan *Vector Space Model* Berbasis Web

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.



## HALAMAN PENGESAHAN

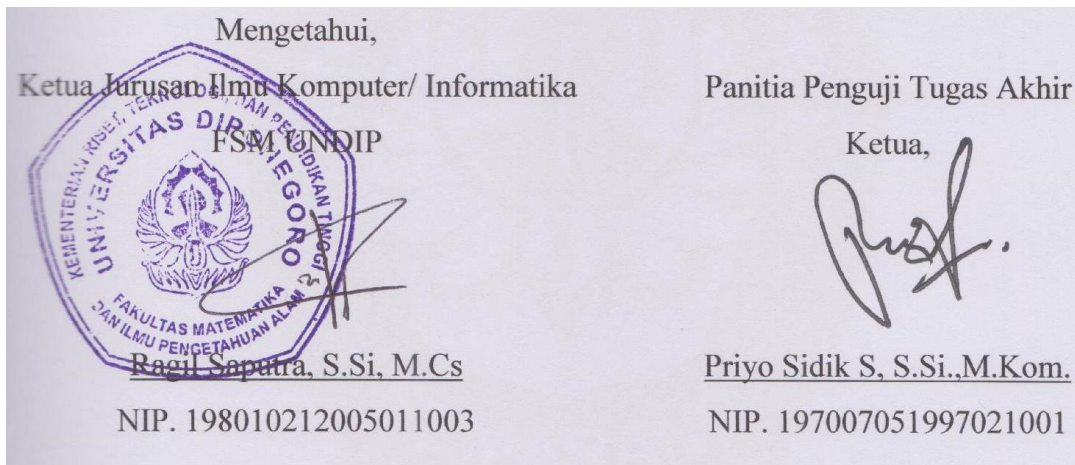
Judul : Aplikasi Peringkat Teks Bahasa Indonesia Otomatis Menggunakan *Vector Space Model* Berbasis Web

Nama : Haris Dwi Arfianto

NIM : 24010311130038

Telah diujikan pada sidang tugas akhir tanggal 6 Oktober 2015 dan dinyatakan lulus pada tanggal 4 Desember 2015

Semarang, 10 Desember 2015



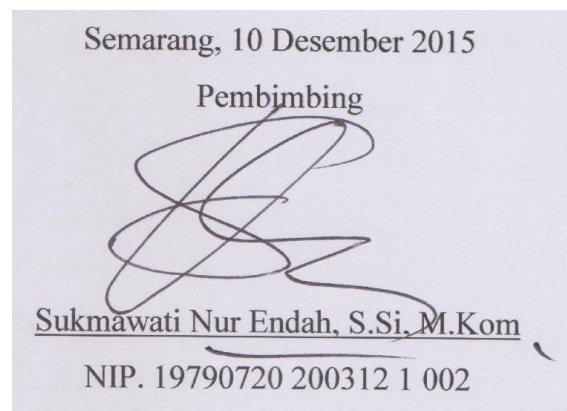
## HALAMAN PENGESAHAN

Judul : Aplikasi Peringkat Teks Bahasa Indonesia Otomatis Menggunakan *Vector Space Model* Berbasis Web

Nama : Haris Dwi Arfianto

NIM : 24010311130038

Telah diujikan pada sidang tugas akhir tanggal 06 Oktober 2015.



## ABSTRAK

Ringkasan menyajikan informasi dari suatu teks secara ringkas sehingga dapat mengurangi usaha dalam memahami banyaknya teks. Data teks yang banyak mengakibatkan proses meringkas secara manual memakan waktu yang lama. Pembuatan ringkasan ini dapat dipermudah dengan adanya peringkasan teks secara otomatis bahasa Indonesia. Peringkasan teks otomatis dapat dilakukan dengan menentukan nilai kemiripan antar kalimat. Nilai kemiripan ini dapat ditentukan berdasarkan sudut vektor kalimat pada *vector space model*. Beberapa kalimat dengan nilai kemiripan tertinggi dipilih sebagai representasi ringkasan teks. Pengujian dengan menggunakan data dari 25 teks terdiri dari teks berjenis argumentasi, deskripsi, eksposisi, narasi, persuasi telah menghasilkan rata-rata nilai *precision*, *recall* dan *F-Score* masing-masing sebesar 0.55, 0.49, 0.51 dari narasumber pakar serta 0.55, 0.48, dan 0.50 dari narasumber pengguna umum. Hasil ringkasan memiliki tingkat *informativeness* yang sama dengan hasil ringkasan ideal. Hasil *precision* tertinggi merupakan teks yang berjenis argumentasi dengan rata-rata *precision* sebesar 0.52 dari narasumber pakar dan 0.46 dari pengguna umum.

**Kata kunci** : peringkasan teks otomatis, *vector space model*

## **ABSTRACT**

Summary provides information from a text briefly to reduce the amount of effort in understanding the text. The amount of data in the text resulting longer summarization process. This process can be easily done with automatic text summarization based on Indonesian language. Automatic text summarization can be done by determining the value of similarity between sentences. This similarity value can be determined based on the vector angle sentence on the vector space model. Some sentences with the highest similarity value chosen as a representation of the text summaries. Based on testing that used data from 25 text consisted of argument, description, exposition, narration, and persuasion text that resulted in an average value of precision, recall and F-Score each 0.55, 0.49, 0.51 from expert interviewees and 0.55, 0.48, and 0.50 from common user interviewees. The summary had the same informativeness level with the expert interviewees summaries. Precision with the highest result obtained from argument text with an average precision of 0.52 from expert interviewees and 0.46 from common user interviewees.

**Keywords** : automatic text summarization, vector space model

## KATA PENGANTAR

Segala puji syukur bagi Tuhan Yang Maha Esa atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan penulisan laporan tugas akhir yang berjudul “Aplikasi Peringkat Teks Bahasa Indonesia Otomatis Menggunakan *Vector Space Model* Berbasis Web”. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Jurusan Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

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

1. Ragil Saputra, S.Si, M.Cs selaku Ketua Jurusan Ilmu Komputer/Informatika
2. Helmie Arif Wibawa, S.Si, M.Cs selaku Koordinator Tugas Akhir
3. Sukmawati Nur Endah, S.Si, M.Kom selaku dosen pembimbing
4. Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir, yang tidak dapat penulis sebutkan satu persatu.

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 dan penulis pada umumnya.

Semarang, November 2015

Penulis,

Haris Dwi Arfianto

24010311130038

## DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN .....	iii
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I     PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan dan Manfaat .....	2
1.4. Ruang Lingkup.....	2
1.5. Sistematika Penulisan.....	3
BAB II    TINJAUAN PUSTAKA .....	4
2.1. Ringkasan .....	4
2.2. <i>Information Retrieval</i> .....	4
2.3. <i>Arsitektur Information Retrieval System</i> .....	5
2.4. <i>Vector Space Model</i> .....	9
2.5. <i>Model Unified Process</i> .....	11
2.6. UML.....	14
2.6.1. <i>Things</i> .....	14
2.6.2. <i>Relationship</i> .....	15
2.6.3. <i>Diagram</i> .....	18
2.7. Evaluasi .....	20
2.9.1. <i>Compression Ratio</i> .....	20



	2.9.2. <i>Informativeness</i> .....	20
	2.9.3. <i>Precision and Recall</i> .....	21
BAB III	FASE <i>INCEPTION</i> DAN FASE <i>ELABORATION</i> .....	23
	3.1. Fase <i>Inception</i> .....	23
	3.2.1. Deskripsi Aplikasi .....	23
	3.2.2. Kebutuhan Fungsional .....	24
	3.2.3. Kebutuhan Non-Fungsional .....	24
	3.2.4. <i>Use Case Modeling</i> .....	24
	3.2.4.1. Daftar Aktor .....	24
	3.2.4.2. Daftar <i>Use Case</i> .....	25
	3.2.4.3. Diagram <i>Use Case</i> .....	25
	3.2.4.4. Detail <i>Use Case</i> .....	25
	3.2. Fase <i>Elaboration</i> .....	27
	3.3.1. Arsitektur Aplikasi .....	27
	3.3.2. <i>Design Model</i> .....	27
	3.3.2.1. <i>Class Diagram</i> .....	27
	3.3.2.2. <i>Sequence Diagram</i> .....	28
	3.3.3. <i>Data Model</i> .....	30
	3.3.4. Menyusun Rancangan Proses .....	30
	3.3.4.1. <i>Input</i> .....	31
	3.3.4.2. Praproses .....	31
	3.3.4.3. Menghitung Bobot .....	33
	3.3.4.4. Menghitung Nilai <i>Similarity</i> .....	33
	3.3.4.5. <i>Output Ringkasan</i> .....	33
	3.3.4. Menyusun Prototipe Antarmuka .....	37
	3.3.5. Menyusun Rencana Pengujian .....	40
	3.3.5.1. Rencana Pengujian Fungsional Sistem .....	41

	3.3.5.2. Rencana Pengujian Validitas .....	41
BAB IV	FASE <i>CONSTRUCTION</i> DAN FASE <i>TRANSITION</i> .....	43
	4.1. Fase Construction .....	43
	4.1.1. Lingkungan Implementasi Sistem .....	43
	4.1.2. Implementasi Data .....	43
	4.1.3. Implementasi Class .....	44
	4.1.4. Implementasi Antarmuka .....	44
	4.2. Fase Transaction .....	49
	4.2.1. Lingkungan Pengujian .....	49
	4.2.2. Pengujian Fungsi .....	49
	4.2.3. Pengujian Validitas .....	50
	4.2.4. Analisa Pengujian .....	51
BAB V	PENUTUP .....	54
	5.1 Kesimpulan .....	54
	5.2 Saran .....	54
	DAFTAR PUSTAKA .....	55
	LAMPIRAN-LAMPIRAN .....	57

## DAFTAR GAMBAR

Gambar 2.1.	Proses <i>Information Retrieval</i> .....	5
Gambar 2.2.	Proses Pengindekan Kata .....	6
Gambar 2.3.	<i>Cosine</i> dari $\theta$ Didapatkan dari $d_j$ dan $q$ .....	9
Gambar 2.4.	Alur Kerja <i>Unified Process</i> .....	12
Gambar 2.5.	Siklus Hidup <i>Unified Process</i> .....	12
Gambar 2.6.	Hubungan Fase dan Alur Kerja dalam <i>Unified Process</i> .....	13
Gambar 2.7.	<i>Dependency</i> Antara Class ‘ <i>Filmclip</i> ’ dan ‘ <i>Channel</i> ’ .....	15
Gambar 2.8.	Contoh Penggunaan <i>Name</i> Asosiasi Antara Class ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’ .....	16
Gambar 2.9.	Contoh Penggunaan <i>Role</i> dari Asosiasi Antara Class ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’ .....	16
Gambar 2.10.	Contoh Penggunaan <i>Multiplicity</i> dari Asosiasi Antara Class ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’ .....	17
Gambar 2.11.	Contoh Penggunaan <i>Aggregation</i> Antara Class ‘ <i>Company</i> ’ dan ‘ <i>Department</i> ’ .....	17
Gambar 2.12.	<i>Generalization</i> : Class ‘ <i>Rectangle</i> ’, ‘ <i>Circle</i> ’, ‘ <i>Polygon</i> ’ Spesialisasi dari Class ‘ <i>Shape</i> ’ .....	17
Gambar 2.13.	Contoh <i>Class Diagram</i> Pemesanan Barang .....	18
Gambar 2.14.	Simbol <i>Use Case</i> .....	19
Gambar 2.15.	Simbol <i>Actor</i> .....	19
Gambar 2.16.	Contoh <i>Sequence Diagram</i> untuk Proses Pemesanan Barang .....	20
Gambar 3.1.	Tahapan Alur Proses Aplikasi Peringkat Teks .....	23
Gambar 3.2.	Diagram <i>Use Case</i> Aplikasi .....	25
Gambar 3.3.	Arsitektur Aplikasi Peringkat Teks Otomatis .....	27
Gambar 3.4.	<i>Class Diagram</i> Aplikasi Peringkat Teks .....	28
Gambar 3.5.	<i>Sequence Diagram</i> Input Teks .....	29
Gambar 3.6.	<i>Sequence Diagram</i> Peringkat .....	29
Gambar 3.7.	Desain Antarmuka Halaman Beranda Aplikasi Peringkat Teks Otomatis .....	38
Gambar 3.8.	Desain antarmuka menu <i>input</i> dengan <i>text field</i> .....	38
Gambar 3.9.	Desain antarmuka menu <i>input</i> dengan <i>upload file</i> .....	39

Gambar 3.10. Desain antarmuka halaman hasil ringkasan .....	39
Gambar 3.11. Desain antarmuka detail praproses .....	40
Gambar 3.12. Desain antarmuka detail menghitung bobot .....	40
Gambar 3.13. Desain antarmuka detail menghiung nilai <i>similarity</i> .....	40
Gambar 4.1. Struktur Tabel Tb_katadasar pada MySQL .....	43
Gambar 4.2. Struktur Tabel Tb_stoplist pada MySQL.....	44
Gambar 4.3. Antarmuka Halaman Beranda.....	45
Gambar 4.4. Antarmuka Halaman <i>Input Text Field</i> .....	46
Gambar 4.5. Antarmuka Halaman <i>Upload File</i> .....	46
Gambar 4.6. Antarmuka Halaman Hasil Ringkasan.....	47
Gambar 4.7. Antarmuka Halaman Detail Praproses.....	47
Gambar 4.8. Antarmuka Halaman Detail Menghitung Bobot .....	48
Gambar 4.9. Antarmuka Halaman Detail Menghitung Nilai <i>Similarity</i> .....	48
Gambar 4.10 Grafik Precision, Recall dan F-Score dari Narasumber Pakar Berdasarkan Jenis teks .....	52
Gambar 4.11 Grafik Precision, Recall dan F-Score dari Narasumber Pengguna Umum Berdasarkan Jenis teks .....	53

## DAFTAR TABEL

Tabel 2.1. Kombinasi Awalan Akhiran yang tidak Diizinkan.....	8
Tabel 2.2. Nilai Bobot Kata dari Contoh <i>Vector Space Model</i> .....	10
Tabel 2.3 Jenis <i>Relationship</i> Pada <i>Use Case Diagram</i> .....	19
Tabel 2.4. Perhitungan <i>Recall</i> dan <i>Precision</i> .....	21
Tabel 3.1. Daftar Aktor Aplikasi .....	24
Tabel 3.2. Daftar <i>Use Case</i> Aplikasi .....	25
Tabel 3.3. Detail <i>Use Case Input</i> Teks .....	26
Tabel 3.4. Detail <i>Use Case</i> Meringkas Teks .....	26
Tabel 3.5. Detail <i>Use Case</i> Menampilkan Hasil.....	26
Tabel 3.6. Struktur Tabel <i>tb_stoplist</i> .....	30
Tabel 3.7. Struktur Tabel <i>tb_katadasar</i> .....	30
Tabel 3.8. Pembagian Dokumen.....	34
Tabel 3.9. Kata-Kata Dasar dari Tahap Praproses .....	35
Tabel 3.10. Nilai Bobot Kata dari Setiap Dokumen .....	35
Tabel 3.11. Nilai <i>Similarity</i> dari Setiap Kombinasi Kalimat.....	37
Tabel 3.12. Rencana Pengujian Fungsi Aplikasi .....	41
Tabel 4.1. Implementasi <i>class</i> .....	44
Tabel 4.2. Hasil dan Evaluasi Pengujian Fungsi Aplikasi Peringkasan Teks Bahasa Indonesia.....	49
Tabel 4.3 Data Statistik Grup .....	50
Tabel 4.4 Data Hasil Uji Independen.....	50

## DAFTAR LAMPIRAN

Lampiran 1. Tabel Daftar <i>Stopword</i> .....	57
Lampiran 2. Daftar Kata Dasar .....	59
Lampiran 3. Data Kuesioner Ringkasan Teks .....	62
Lampiran 4. <i>Source Code</i> Fungsi-Fungsi .....	98
Lampiran 5. Tabel Hasil Perhitungan Nilai <i>Compression Ratio</i> .....	102
Lampiran 6. Nilai <i>Precision, Recall</i> , dan <i>F-Score</i> Data Narasumber.....	103

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sebuah teks memiliki banyak informasi yang terkandung di dalamnya. Informasi ini dapat diperoleh dari berbagai macam sumber mulai dari berita, karya ilmiah dan buku-buku. Informasi yang terkandung di dalam tulisan tersebut tidak sepenuhnya penting. Penyajian teks terdiri dari kalimat utama yang merupakan inti dari sebuah tulisan dan kalimat penjelas yang merupakan pelengkap dari tulisan tersebut. Informasi akan lebih mudah diterima apabila langsung diperoleh dari kalimat-kalimat utama dalam sebuah bentuk ringkasan.

Ringkasan yang ditulis dengan baik dapat mengurangi pekerjaan dalam memahami teks yang banyak. Namun, dengan banyaknya data teks yang ada, pembuatan ringkasan secara manual dapat memakan waktu yang lama. Pembuatan ringkasan ini dapat dipermudah dengan menggunakan aplikasi yang dapat membuat ringkasan secara otomatis berbasis web, sehingga dapat digunakan oleh siapa saja, kapan saja dan dimana saja. Proses peringkasan dilakukan dengan memanfaatkan sistem temu kembali informasi. Teks yang terdiri dari beberapa paragraf akan diproses untuk didapatkan hasil ringkasan secara otomatis.

Aplikasi peringkasan teks sudah mulai berkembang pada tahun 1958 (Luhn, 1958). Peringkasan teks bahasa Inggris telah dikembangkan menggunakan beberapa metode antara lain *sentence scoring*, *cluster based*, *LSA*, *fuzzy logic*, *vector space model* dan lain-lain (Gupta & Lehal, 2010). Peringkasan teks bahasa Indonesia dikembangkan dengan metode diantaranya *sentence scoring*, *cluster based*, *LSA*.

Peringkasan teks bahasa Indonesia belum ditemukan dengan menggunakan *vector space model*, sedangkan peringkasan teks bahasa Inggris yang menggunakan *vector space model* telah terbukti menghasilkan tingkat keakuratan 57,86% (Kageback et al., 2014). Nilai tersebut tidak menunjukkan nilai keakuratan yang tinggi, namun bahasa Indonesia dan bahasa Inggris memiliki struktur yang berbeda. Perbedaan ini yang mendasari dilakukan penelitian untuk peringkasan teks menggunakan *vector space model* pada teks bahasa Indonesia.

*Vector space model* merupakan proses pencocokan *query* (kata kunci untuk mencari, berupa kata/kalimat) dengan teks yang ada. Salah satu kelebihanannya adalah mudah beradaptasi dengan metode pembobotan (Kowalski G & Maybury M, 2002). Kelebihan tersebut mengakibatkan proses mencari kesamaan antar kalimat bisa lebih mudah dilakukan dengan metode pembobotan yang digunakan. Proses peringkasan teks otomatis dilakukan dengan membandingkan setiap kalimat yang ada di dalam teks. Beberapa kalimat yang memiliki nilai kemiripan tertinggi diambil untuk dijadikan sebagai ringkasan teks tersebut.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan dapat dibuat rumusan masalah yaitu bagaimana membangun sebuah aplikasi yang mampu membuat ringkasan secara otomatis menggunakan *vector space model* dengan pembobotan *Term Frequency/Inverse document Frequency* (TF/IDF) serta mengimplementasikan ke aplikasi berbasis *web*.

## 1.3. Tujuan dan Manfaat

Tujuan dari penelitian tugas akhir ini adalah sebagai berikut :

1. Menghasilkan aplikasi yang dapat meringkas teks bahasa Indonesia secara otomatis dengan menggunakan *vector space model*.
2. Mengetahui tingkat *informativeness* hasil ringkasan dan mengukur ketepatan aplikasi dalam mengambil kalimat yang tepat sebagai hasil ringkasan.
3. Mengetahui jenis teks yang memiliki ketepatan terbaik terhadap hasil ringkasan aplikasi.

Adapun manfaat dari penelitian tugas akhir ini adalah aplikasi yang dikembangkan dapat digunakan untuk meringkas teks dan mendapatkan inti dari informasi dalam teks tersebut dengan waktu yang lebih efisien.

## 1.4. Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut :

1. Aplikasi hanya membuat ringkasan secara ekstraktif.
2. Teks yang diringkaskan merupakan teks bahasa Indonesia.
3. Masukan teks berupa *file* dengan ekstensi *.txt* maupun *input* langsung ke dalam aplikasi melalui *input* teks.
4. Masukan teks hanya berupa teks/dokumen tunggal.



5. Setiap paragraf dipisahkan oleh pergantian baris dan tidak mengandung *tag* HTML atau gambar.

## 1.5. Sistematika Penulisan

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

### BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup dan sistematika dalam pembuatan tugas akhir mengenai aplikasi peringkasan teks bahasa Indonesia otomatis menggunakan *vector space model* berbasis web.

### 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 pengertian ringkasan, *Information Retrieval*, *Arsitektur Information Retrieval System*, *Vector Space Model*, model *Unified Process*, *Unified Modeling Language (UML)*, PHP, MySQL, dan *Precision Recall*.

### BAB III FASE *INCEPTION* DAN FASE *ELABORATION*

Bab ini menyajikan mengenai pembahasan tahapan dari model pengembangan perangkat lunak menggunakan model *Unified Process*. Bab ini disajikan dua fase awal yaitu *Inception* dan *Elaboration*.

### BAB IV FASE *CONSTRUCTION* DAN FASE *TRANSITION*

Bab ini menyajikan mengenai pembahasan tahapan dari model pengembangan perangkat lunak menggunakan model *Unified Process*. Bab ini disajikan fase *Construction* dan *Transition*.

### BAB VI PENUTUP

Bab ini merupakan kesimpulan dari bab-bab yang dibahas sebelumnya dan saran untuk pengembangan penelitian tugas akhir lebih lanjut.