

## ABSTRAK

*Information Retrieval System* adalah bidang yang berkaitan dengan struktur, analisis, organisasi, penyimpanan, pencarian, dan pengambilan informasi. Dalam tugas akhir ini dikembangkan sebuah aplikasi untuk mendukung perkuliahan Pengantar Temu Balik Informasi dengan menggunakan contoh koleksi novel yang ditulis oleh Rider Haggard. Metode-metode yang diimplementasikan dalam aplikasi ini meliputi metode pencarian *Vector Space Model*(VSM) dan *Boolean*. Kedua metode tadi diterapkan dalam aplikasi dengan menggunakan Lucene. Hasil evaluasi menunjukkan bahwa urutan penemuan dokumen secara teoretis bersesuaian dengan hasil dari aplikasi. Hal ini menunjukkan bahwa aplikasi telah dapat berperan sebagai perangkat temu balik dokumen. Dengan tugas akhir ini, diharapkan dapat menunjukkan cara kerja *VSM*, *Boolean*, dan *Lucene*, sehingga dapat digunakan untuk pengenalan mata kuliah Sistem Temu Balik Informasi.

Kata Kunci : *Lucene, Boolean, Vector Space Model*, Sistem Temu Balik Informasi

## ABSTRACT

*Information Retrieval System is a field that deals with the structure, analysis, organization, storage, search, and retrieval of information. In this thesis, developed an application to support the course of Introduction to Information Retrieval by using an example, a collection of novels written by Rider Haggard. The methods implemented in this application includes search methods of Vector Space Model (VSM) and Boolean. Both methods are applied by using Lucene. The evaluation shows that the order of documents is theoretically consistent with the results of the application. This indicates that the application has been able to act as information retrieval of the documents. With this thesis, it is expected to show how the VSM, Boolean, and Lucene works, so that it can be used for the Introduction of Information Retrieval System course.*

*Keyword : Lucene, Vector Space Model, Boolean, Information Retrieval System*

## DAFTAR ISI

LEMBAR PENGESAHAN.....	i
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN.....	ii
SURAT PERNYATAAN ORISINALITAS KARYA.....	iii
PRAKATA.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
DAFTAR RUMUS.....	xii
DAFTAR KODE.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	1
1.3 Tujuan.....	2
1.4 Batasan Masalah.....	2
1.5 Sistematika Pembahasan.....	2
BAB II DASAR TEORI.....	3
2.1 Sistem Temu Balik Informasi.....	3
2.1.1 Boolean Retrieval.....	4
2.1.2 Vector Space Model.....	4
2.2 Lucene.....	8
2.4 Sumber Data.....	10
BAB III ANALISIS DAN DESAIN.....	12
3.1 Contoh Kasus.....	12
3.2 Use Case.....	15
3.3 Activity Diagram.....	16
3.3.1 Melakukan Pencarian.....	16
3.3.2 Indeksasi.....	17

BAB IV PENGEMBANGAN PERANGKAT LUNAK.....	19
4.1 Implementasi.....	19
4.1.1 Indeksasi.....	20
4.1.2 Melakukan Pencarian.....	22
4.1.3 Clear.....	29
BAB V PENGUJIAN.....	30
5.1 Pengujian.....	30
5.2 Membuat Indeks.....	31
5.3 Kendala.....	35
BAB VI KESIMPULAN DAN SARAN.....	36
6.1 Kesimpulan.....	36
6.2 Saran.....	36
DAFTAR PUSTAKA.....	37

## DAFTAR GAMBAR

Gambar 3.1 Use Case.....	15
Gambar 3.2 Melakukan Pencarian.....	16
Gambar 3.3 Indeksasi.....	17
Gambar 3.4 Rancangan Antarmuka Aplikasi Temu Balik.....	18
Gambar 4.1 Tampilan Awal.....	20
Gambar 4.2 Pesan Indeks.....	22
Gambar 4.3 Hasil Pencarian VSM.....	27
Gambar 4.4 Hasil Pencarian Boolean.....	28
Gambar 4.5 <i>Highlights</i> dalam isi <i>file</i> .....	29
Gambar 5.1 Testing Boolean.....	31
Gambar 5.2 <i>Folder Sumber Data</i> .....	32
Gambar 5.3 <i>Folder Indeks</i> .....	33
Gambar 5.4 Hasil <i>VSM</i> .....	34
Gambar 5.5 Hasil <i>Boolean</i> .....	34

## DAFTAR TABEL

Tabel 3.1 Perhitungan Koleksi Dokumen.....	13
Tabel 3.2 Perhitungan Kueri.....	14
Tabel 3.3 Cosine Similarity.....	15
Tabel 4.1 Membuat Indeks.....	20
Tabel 4.2 Melakukan Pencarian.....	23
Tabel 5.1 Perbandingan.....	30

## DAFTAR RUMUS

Rumus 2.1 Sudut Vektor.....	5
Rumus 2.2 <i>Dot Product</i> .....	5
Rumus 2.3 Panjang Vektor.....	5
Rumus 2.4 IDF.....	5
Rumus 2.5 TFIDF.....	5
Rumus 2.6 Gabungan.....	6
Rumus 2.7 <i>Cosine Similarity</i> .....	6
Rumus 2.8 <i>Average Precision</i> .....	7
Rumus 2.7 <i>Scoring Lucene</i> .....	9

## DAFTAR KODE

Kode 4.1 Membuat Indeks.....	21
Kode 4.2 Melakukan Pencarian.....	25
Kode 4.3 Highlights.....	26
Kode 4.4 genHTML.....	26
Kode 4.5 MyHighlighter.....	27