

**ANALISIS DAN PERANCANGAN SEARCH ENGINE DOKUMEN PAPER
BERBASIS WEB**

SKRIPSI



disusun oleh

Aldrik Saddermi

10.11.4055

**JURUSAN TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2014**

**ANALISIS DAN PERANCANGAN SEARCH ENGINE DOKUMEN PAPER
BERBASIS WEB**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S1
pada jurusan Teknik Informatika



disusun oleh

Aldrik Saddermi

10.11.4055

**JURUSAN TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2014**

PERSETUJUAN

SKRIPSI

**ANALISIS DAN PERANCANGAN SEARCH ENGINE DOKUMEN
PAPER BERBASIS WEB**

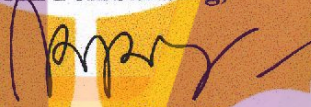
yang dipersiapkan dan disusun oleh

Aldrik Saddermi

10.11.4055

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Maret 2014

Dosen Pembimbing,



Krisnawati, S.Si, MT

NIK. 190302038

PENGESAHAN

SKRIPSI

ANALISIS DAN PERANCANGAN SEARCH ENGINE DOKUMEN

PAPER BERBASIS WEB

yang dipersiapkan dan disusun oleh

Aldrik Saddermi

10.11.4055

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 November 2014

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bambang Sudaryatno, Drs, MM
NIK. 190302029

Krisnawati, S.Si, MT
NIK. 190302038

Yuli Astuti, M.Kom
NIK. 190302146

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 November 2014

KEFUA STMIK AMIKOM YOGYAKARTA

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

PERNYATAAN

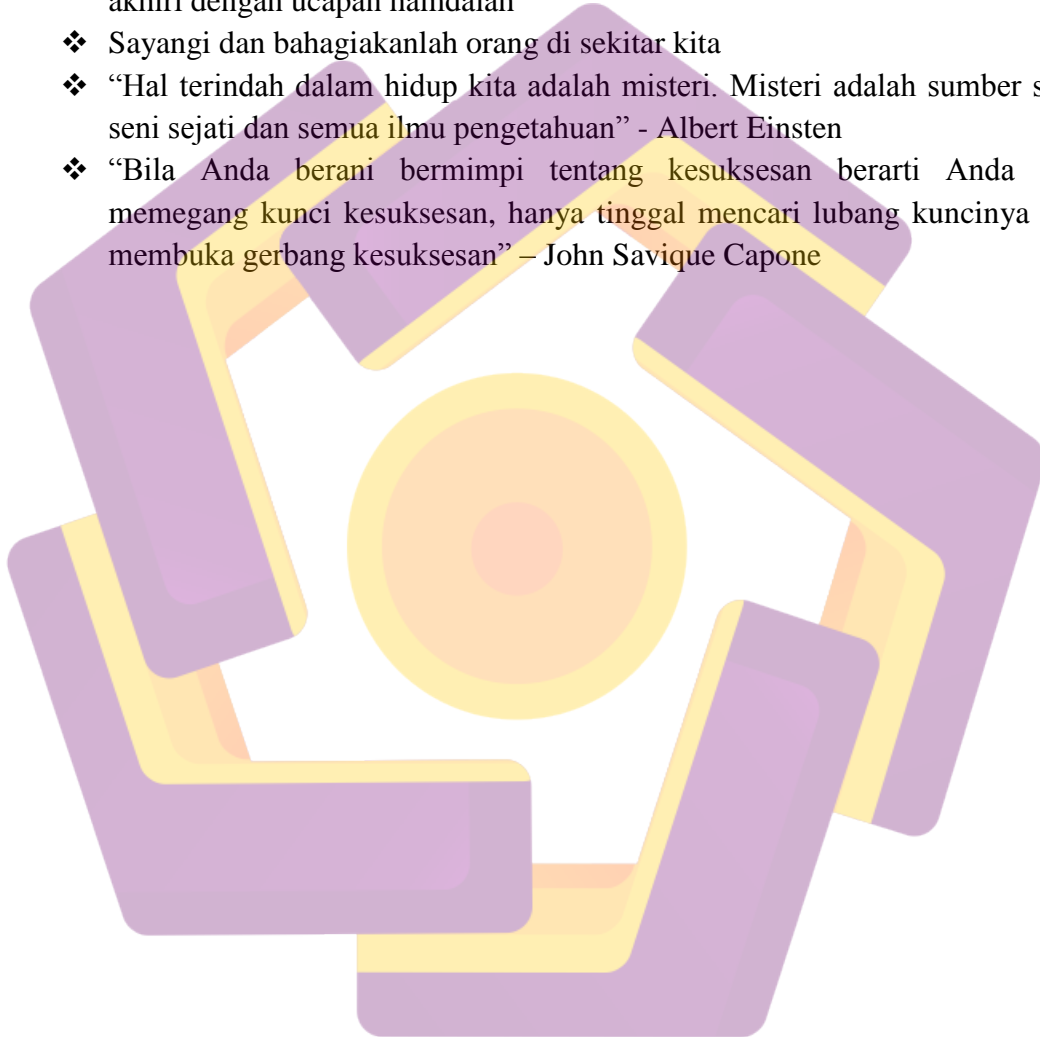
Saya yang bertanda tangan dibawah ini menyatakan bahwa, Skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam Skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain atau kelompok lain untuk memperoleh gelar akademis di suatu Institusi Pendidikan, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan atau diterbitkan oleh orang lain atau kelompok lain, kecuali yang secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Yogyakarta, 26 November 2014

Aldrik Saddermi
10.11.4055

MOTTO

- ❖ Ini dunia, bukan surga. Tidak ada yang sempurna dan kau tak kan bisa sempurna.
- ❖ Kerjakan sesuatu yang baik dengan niat yang tulus dan ucapan basmalah, akhiri dengan ucapan hamdalah
- ❖ Sayangi dan bahagiakanlah orang di sekitar kita
- ❖ “Hal terindah dalam hidup kita adalah misteri. Misteri adalah sumber semua seni sejati dan semua ilmu pengetahuan” - Albert Einstein
- ❖ “Bila Anda berani bermimpi tentang kesuksesan berarti Anda sudah memegang kunci kesuksesan, hanya tinggal mencari lubang kuncinya untuk membuka gerbang kesuksesan” – John Savique Capone



PERSEMBAHAN

Alhamdulillahirobbi'alamin, akhirnya Skripsi ini dapat terselesaikan dengan baik. Hasil karya ini saya persembahkan untuk orang-orang yang saya cintai dan saya sayangi yang selalu memberikan dukungan moral maupun moril kepada saya.

- Kedua orang tuaku yang tercinta yang selalu memberikan dukungan, semangat dan doanya kepadaku. Terimakasih bapak dan ibu, kalian orang tua yang hebat, saya bangga menjadi anak kalian.
- Keluarga, terutama kakak dan adikku yang cantik-cantik. Terimakasih sudah menyemangati adikmu atau mamasmu ini.
- Dosen pembimbing, Krisnawati, S.Si, MT. yang bersedia meluangkan waktu untuk membimbing dan memberikan saran dalam penyusunan skripsi ini.
- Dosen penguji, Bapak Bambang Sudaryatno, Drs, MM dan Ibu Yuli Astuti, M.Kom yang sudah memberikan kritik dan saran pada Skripsi ini.
- Ads-ID Community, terutama Bedit Riyanto yang sudah memberikan banyak hal yang berarti terhadap Skripsi ini dan juga terhadap saya. Terimakasih teman.
- Teman-temanku yang selama ini menemaniku dalam belajar, bergaul dan lain-lain, Zainuri, Hengki, Erni, Tomi Utomo, Nancy dan teman-teman lainnya yang tidak saya sebut baik teman-teman kelas, teman-teman Forum Ads-ID maupun teman-teman main. Terimakasih, kalian takkan terlupakan.
- Semua pihak yang tidak bisa saya sebutkan satu-persatu yang telah banyak membantu saya dalam penyusunan Skripsi ini.

KATA PENGANTAR

Segala puji dan syukur kehadiran Allah SWT, atas rahmat, ridho dan karunia-Nya sehingga skripsi ini dapat terselesaikan. Penyusunan Skripsi ini merupakan persyaratan kelulusan program Sarjana S1 pada jurusan Teknik Informatika di STMIK AMIKOM Yogyakarta.

Dalam penyusunan Skripsi ini, penulis merasa masih menemukan beberapa kesulitan dan hambatan. Disamping itu penulis menyadari bahwa penulisan laporan Skripsi ini masih jauh dari kesempurnaan, maka dari itu penulis mengharapkan saran dan kritik yang membangun dari semua pihak.

Penyelesaian Skripsi ini juga tidak lepas dari bantuan berbagai pihak, untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terimakasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Ketua STMIK AMIKOM Yogyakarta yang selama ini menjadi tempat untuk menuntun ilmu.
2. Ibu Krisnawati, S.Si, MT. selaku dosen pembimbing yang sudah memberikan arahan, masukan dan saran terhadap Skripsi ini.
3. Bapak Bambang Sudaryatno, Drs, MM, Ibu Krisnawati, S.Si, MT dan Ibu Yuli Astuti, M.Kom, selaku dosen penguji yang telah memberikan nilai terbaik untuk saya.
4. Kedua orang tua, kakak, adik, dan teman special yang selalu memberikan dukungan dan doa, sehingga penulis lebih bersemangat dan diberi kemudahan dalam menyelesaikan Skripsi ini.
5. Teman-teman sepermainan dan seperjuangan dari awal sampai akhir, terimakasih atas bantuannya selama ini.
6. Seluruh pihak yang tidak dapat disebutkan satu-persatu yang telah banyak membantu penulis selama ini.

Akhir kata, semoga Allah SWT melimpahkan rahmat dan karunia-Nya kepada pihak-pihak yang telah membantu dalam penyusunan Skripsi ini dan juga semoga Skripsi ini dapat bermanfaat bagi pihak-pihak yang membutuhkan.

Yogyakarta, 26 November 2014

Penulis

Aldrik Saddermi

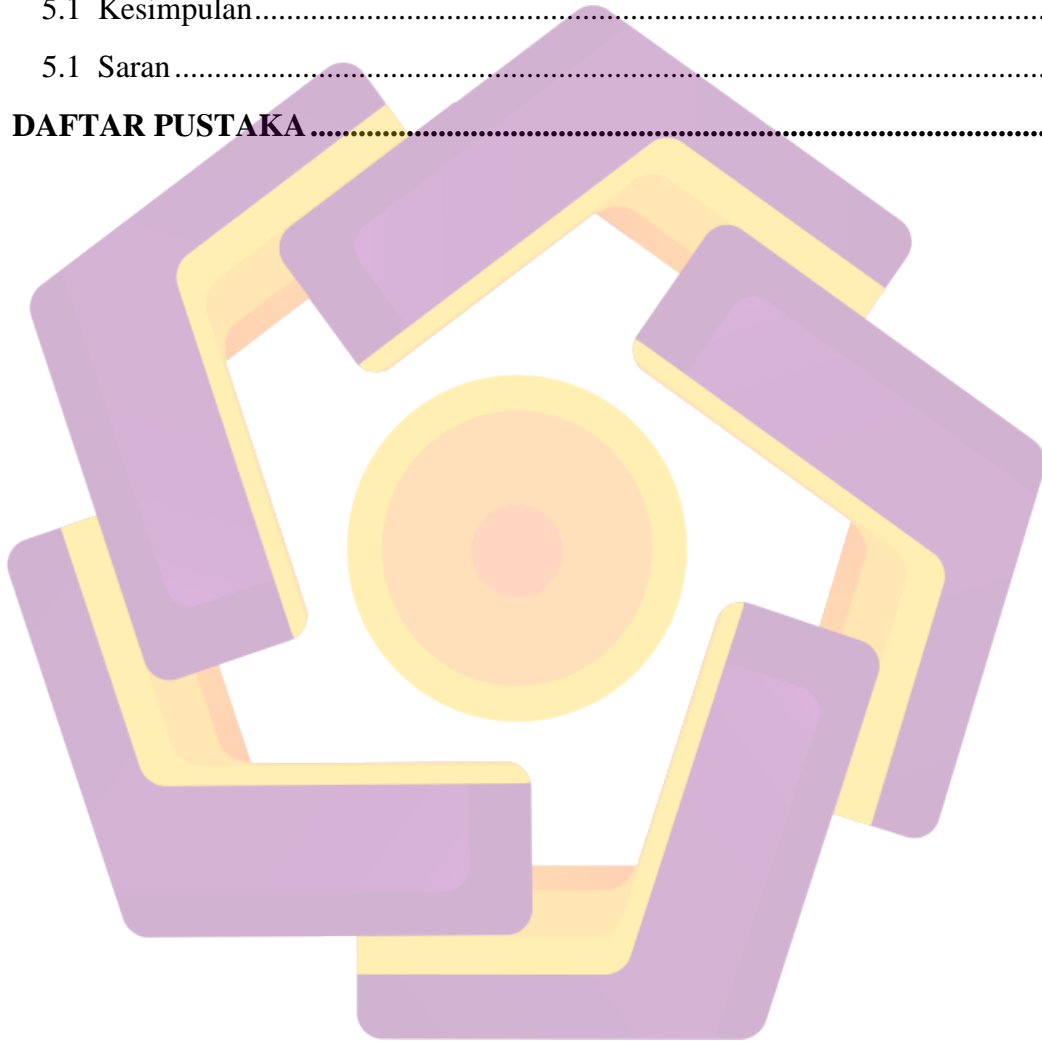
10.11.4055

DAFTAR ISI

COVER	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN.....	v
MOTTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.7 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI	8
2.1 <i>Web</i> dan <i>Situs Web</i>	8
2.1.1 Pengertian <i>World Wide Web</i> (WWW).....	8
2.1.2 Jenis-jenis <i>Web</i>	9
2.1.3 Unsur-unsur atau <i>Situs</i>	10
2.1.4 Istilah-istilah Seputar WWW.....	12

2.2 Aplikasi	14
2.2.1 Perangkat Lunak Aplikasi	14
2.2.2 Aplikasi Berbasiskan <i>Web</i>	15
2.3 Search Engine	18
2.3.1 Pengertian Search Engine	18
2.3.2 Cara Kerja Search Engine	18
2.3.3 Anatomi Robot-The <i>Crawler</i>	20
2.4 Prinsip Umum Search Engine	22
2.4.1 Spider	23
2.4.2 Crawler.....	23
2.4.3 Pengindeks (<i>Indexer</i>).....	27
2.4.4 Basis Data (<i>Database</i>).....	27
2.4.5 Mesin Cari (<i>Result Engine</i>).....	28
2.5 Hypertext Processor (PHP).....	28
BAB III ANALISIS DAN PERANCANGAN	31
3.1 Tinjauan Umum.....	31
3.2 Analisis Sistem.....	33
3.2.1 Analisis <i>PIECES</i>	36
3.2.2 Analisis Kebutuhan Sistem	38
3.2.3 Analisis Kebutuhan Sumber Daya Manusia (<i>SDM</i>)	41
3.3 Perancangan.....	42
3.3.1 Perancangan Proses	43
3.3.2 Perancangan Database.....	56
3.3.3 Perancangan <i>Interface</i>	64
3.3.4 Perancangan Struktur Aplikasi.....	68
BAB IV IMPLEMENTASI DAN PEMBAHASAN	72
4.1 Implementasi	72
4.1.1 Implementasi Basis Data.....	73
4.1.2 Implementasi Koneksi Basis Data	77
4.1.3 Implementasi Tampilan Antar Muka	78
4.1.4 Pengujian Sistem.....	81
4.2 Manual Instalasi	86
4.3 Manual Program	88

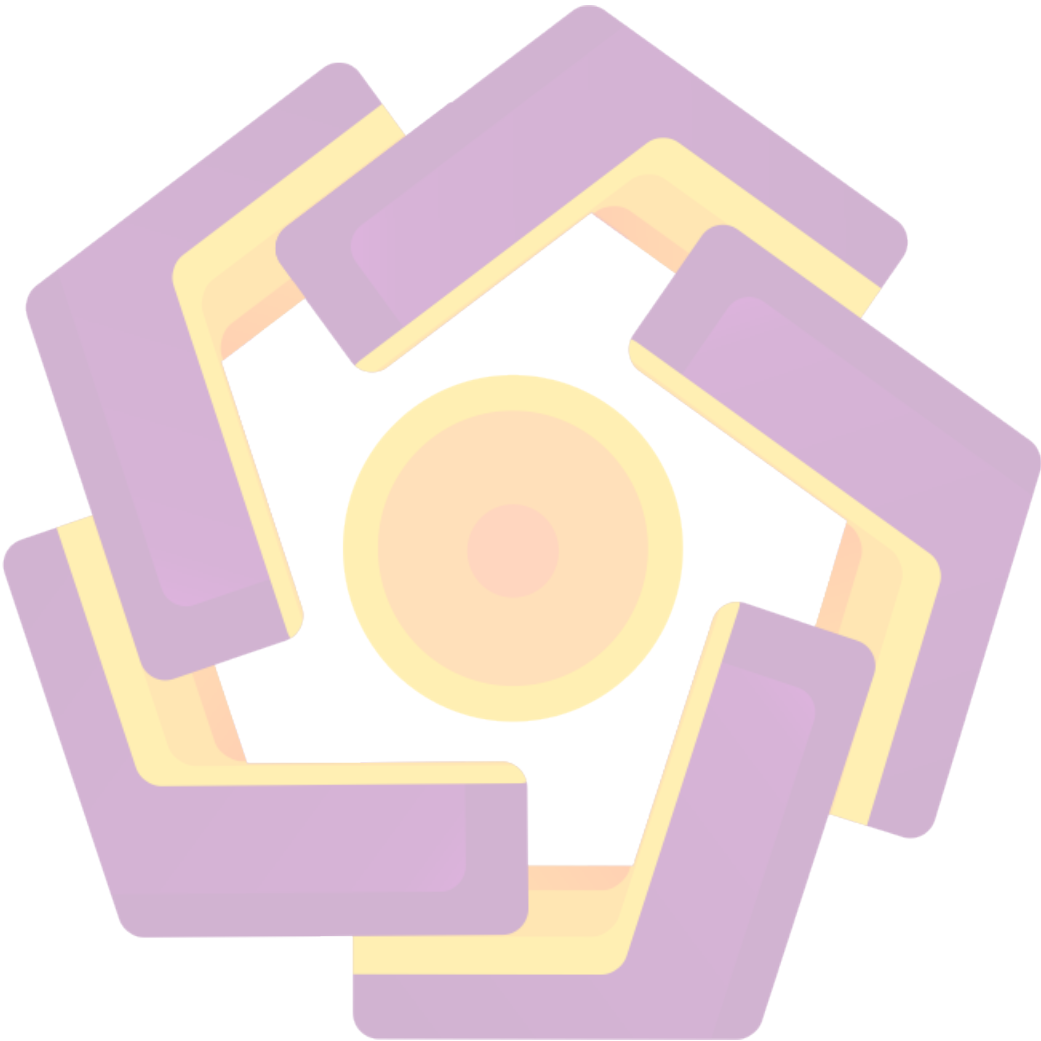
4.4 Pembahasan	94
4.4.1 Pembahasan Listing Program	94
4.4.2 Perbandingan Sistem	95
BAB V PENUTUP	98
5.1 Kesimpulan.....	98
5.1 Saran.....	99
DAFTAR PUSTAKA	100



DAFTAR TABEL

Tabel 3.1 Tabel Domains	57
Tabel 3.2 Tabel Keyword	57
Tabel 3.3 Tabel Links	57
Tabel 3.4 Tabel Link_Keyword0	58
Tabel 3.5 Tabel Link_Keyword1	58
Tabel 3.6 Tabel Link_Keyword2	58
Tabel 3.7 Tabel Link_Keyword3	58
Tabel 3.8 Tabel Link_Keyword4	59
Tabel 3.9 Tabel Link_Keyword5	59
Tabel 3.10 Tabel Link_Keyword6	59
Tabel 3.11 Tabel Link_Keyword7	60
Tabel 3.12 Tabel Link_Keyword8	60
Tabel 3.13 Tabel Link_Keyword9	60
Tabel 3.14 Tabel Link_Keyworda	60
Tabel 3.15 Tabel Link_Keywordb	60
Tabel 3.16 Tabel Link_Keywordc	61
Tabel 3.17 Tabel Link_Keywordd	61
Tabel 3.18 Tabel Link_Keyworde	61
Tabel 3.19 Tabel Link_Keywordf	61
Tabel 3.20 Tabel Pending	62
Tabel 3.21 Tabel Query_Log	62
Tabel 3.22 Tabel Sites	62
Tabel 3.23 Tabel Site_Category	63
Tabel 3.24 Tabel Temp	63
Tabel 3.25 Tabel Categories	63

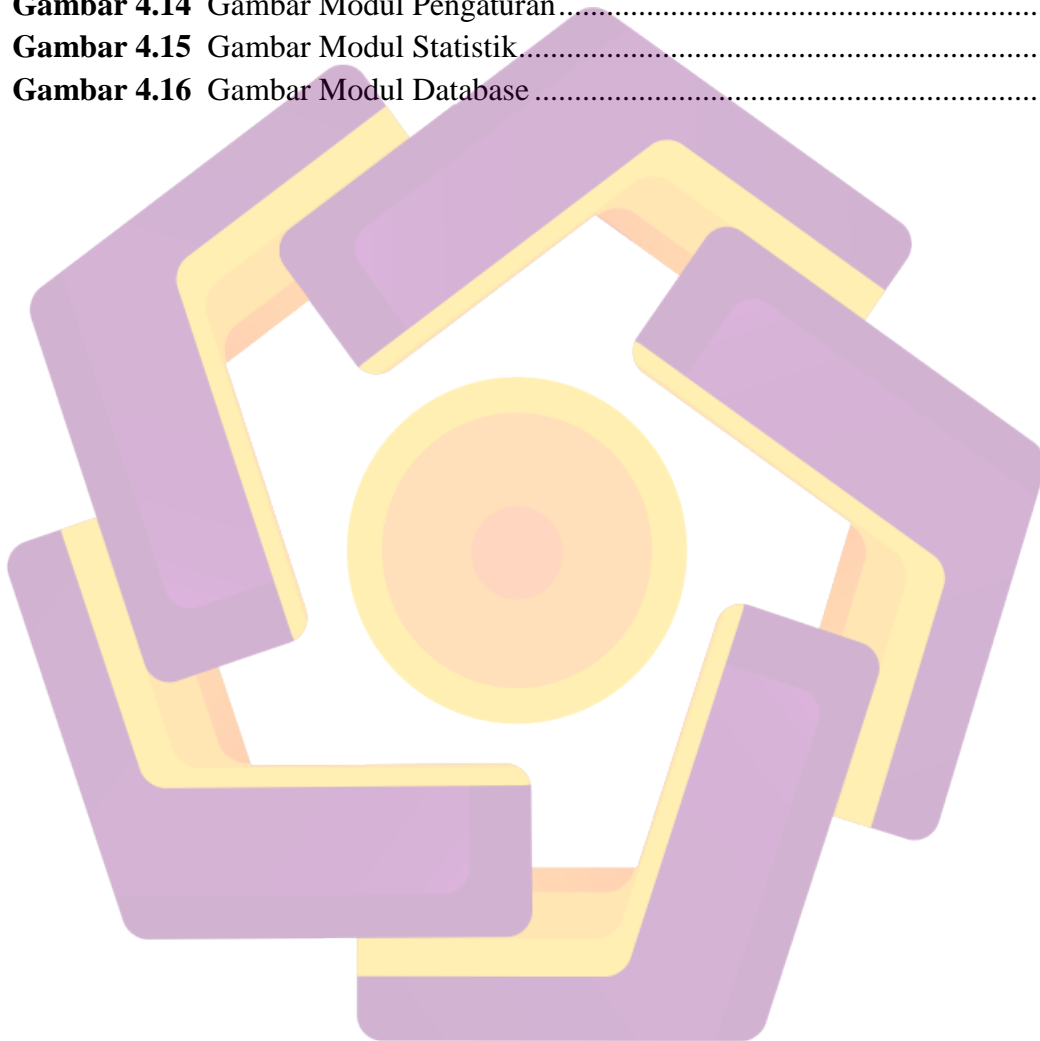
Tabel 4.1 Uji Coba Black Box 81
Tabel 4.2 Pengujian Penggunaan Algoritma Crawler..... 86
Tabel 4.3 Pengujian Perbandingan Sistem..... 96



DAFTAR GAMBAR

Gambar 2.1	Contoh URL Web National Park Service.....	14
Gambar 2.2	Browser dan Server Web (Sumber : Dwiyanto, 2008, 2).....	15
Gambar 2.3	Cara Kerja <i>Search Engine</i>	20
Gambar 2.4	Arsitektur <i>Web Crawler</i>	24
Gambar 3.1	Permodelan system.....	32
Gambar 3.2	Contoh struktur dari <i>web pages</i>	34
Gambar 3.3	Diagram Konteks.....	43
Gambar 3.4	Data Flow Diagram Level 0	44
Gambar 3.5	Data Flow Diagram Level 1 Proses Pengambilan URL.....	45
Gambar 3.6	Data Flow Diagram Level 1 Proses Pengambilan Konten	46
Gambar 3.7	DFD Level 1 Proses Aplikasi Menampilkan Data Pencarian.....	47
Gambar 3.8	Cara Kerja Bot 1	48
Gambar 3.9	Cara Kerja Bot 2	49
Gambar 3.10	Cara Kerja Bot <i>Non Database</i>	50
Gambar 3.11	Gambar Entity Relation Diagram	51
Gambar 3.12	FlowChart dimana aplikasi mengambil data/content	52
Gambar 3.13	FlowChart <i>user</i> dalam menggunakan aplikasi.....	53
Gambar 3.14	FlowChart Data Parsing.....	55
Gambar 3.15	Gambar Relasi Tabel Aplikasi <i>Search Engine</i> Dokumen Paper	56
Gambar 3.16	Antarmuka Halaman Utama	65
Gambar 3.17	Antarmuka Halaman Hasil Pencarian.....	66
Gambar 3.18	Antarmuka Halaman Login Admin	67
Gambar 3.19	Antarmuka Halaman <i>Dashboard</i> Pengaturan Admin.....	67
Gambar 3.20	Gambar Struktur Aplikasi <i>Website</i> untuk <i>user</i>	68
Gambar 3.21	Gambar Struktur Aplikasi <i>Website</i> untuk admin.....	69
Gambar 4.1	Pembuatan Database Baru pada PhpMyAdmin.....	73
Gambar 4.2	Halaman Log in	78
Gambar 4.3	Halaman Admin.....	79
Gambar 4.4	Halaman Home	79
Gambar 4.5	Halaman Hasil Penelusuran.....	80
Gambar 4.6	Black Box Testing pada Navigasi Halaman Pencarian	83
Gambar 4.7	White Box Testing pada Navigasi Halaman Pencarian.....	85
Gambar 4.8	Membuat database baru pada localhost/phpmyadmin.....	87

Gambar 4.9	Cara upload <i>file</i> database pada phpMyAdmin.....	88
Gambar 4.10	Gambar Modul Situs.....	89
Gambar 4.11	Gambar Modul Kategori.....	89
Gambar 4.12	Gambar Modul Index.....	90
Gambar 4.13	Gambar Modul Clean Tables.....	91
Gambar 4.14	Gambar Modul Pengaturan.....	92
Gambar 4.15	Gambar Modul Statistik.....	93
Gambar 4.16	Gambar Modul Database.....	93



INTISARI

Saat ini pertumbuhan informasi secara elektronik di dunia sangatlah cepat. Sejalan dengan itu, informasi dan data tumbuh secara pesat dan sporadis. Perkembangan teknologi yang cukup pesat saat ini memungkinkan untuk mendapatkan aliran data dan informasi dengan mudah dan cepat. Karena data dan informasi tersebut semakin dibutuhkan, maka manusia berusaha untuk mengembangkan teknologi yang mendukung agar pertukaran data semakin cepat dan akurat.

Penulis merancang mesin pencari yang difokuskan untuk mencari dokumen paper dengan menggunakan *Focused Crawler algorithm* sebagai dasar dalam pencarian data kontennya. Sehingga hanya data-data mengenai makalah dan artikel di Indonesia saja yang akan menjadi sumber pencariannya. Selain itu, dilakukan filterisasi pencarian sehingga hasil penelusuran bisa lebih akurat.

Hasil pengujian menunjukan bahwa *focused crawler* didesain hanya untuk mengambil informasi tertentu saja, sehingga mengurangi *traffic* jaringan dan besar data yang dimasukkan kedalam database. Sedangkan *path-ascending crawling* mengambil sebanyak mungkin informasi yang bisa didapat dari website tujuan dengan menggunakan kedalaman *indexing* level 2.

Kata Kunci : Search Engine, Algoritma Crawler, PHP & MySQL

ABSTRACT

Currently, the growth rate of electronic media information was unbelievably fast. Accordingly, the flow of information and data also expanding rapidly and sporadicly. The development of current technology making us to get any new information and data quite fast. Because of the needs of new information and data, humans try to develop a new technology which support information and data transfer not to mention fast and accurate.

The writer designed a search engine that focuses on finding paper documents using Focused Crawler search algorithm as a basis for the data content. Only the data on the papers and articles in Indonesia that will be the source of the search result. In addition, the search result is being filtered first so that the accuracy can be increased.

Testing result show that the focused crawler is designed only to take certain information, thus reducing bandwidth and size data that enters into the database. While the path-ascending crawling take as much information as can be from destination website using indexing depth level 2.

Keywords: *Search Engine, Crawler Algorithm, PHP & MySQL*