



DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
HALAMAN PERSEMBAHAN	iii
HALAMAN MOTTO	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
INTISARI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Metodologi Penelitian	6
1.7 Sistematika Penulisan	7
II TINJAUAN PUSTAKA	9
2.1 <i>Named Entity Recognition</i> dan Ekstraksi Sitasi	9
2.2 Visualisasi Informasi	11
III LANDASAN TEORI	13
3.1 Sistem Temu Balik Informasi	13
3.1.1 Arsitektur dan komponen mesin pencari	13



3.1.2	Proses <i>crawling</i>	14
3.1.3	Proses <i>indexing</i>	16
3.1.4	Pencarian <i>full-text</i>	18
3.1.5	<i>Vector space model</i> (VSM)	19
3.1.6	Pembobotan <i>term</i> dengan skema TF-IDF	19
3.2	Ekstraksi Informasi (<i>Information Extraction</i>)	21
3.2.1	Jenis tugas ekstraksi informasi	22
3.2.2	<i>Finite state automata</i> (FSA)	23
3.2.3	Ekspresi reguler (<i>regular expression</i>)	24
3.3	<i>Named Entity Recognition</i>	25
3.3.1	<i>Named entity recognition</i> berbasis kamus	25
3.3.2	Algoritme <i>spelling correction</i>	26
3.3.3	<i>Edit distance</i>	26
3.4	Visualisasi Informasi	27
3.4.1	Visualisasi graf / network	28
3.5	Visualisasi Sistem Temu Balik Informasi	29
3.5.1	Paradigma dalam visualisasi sistem temu balik informasi	29
3.5.2	Visualisasi hasil pencarian dari mesin pencari	30
3.6	<i>Tools</i> yang Digunakan	30
3.6.1	Scrapy	31
3.6.2	Neo4J	31
3.6.3	Elasticsearch	32
3.6.4	D3.js	33
3.6.5	Flask	34
3.7	Evaluasi Sistem Ekstraksi	34
IV ANALISIS DAN PERANCANGAN		37
4.1	Analisis Permasalahan	37
4.2	Perancangan Umum Sistem	38
4.2.1	Arsitektur kerangka kerja sistem	38
4.2.2	Sumber data	40
4.2.3	Rancangan mesin <i>crawler</i>	42
4.2.4	Rancangan model data graf	47
4.3	Bentuk Sitasi Al-Qur'an di Web	48
4.4	Perancangan <i>Auto Extractor</i> Qur'an Link	49



4.4.1	Pra pemrosesan	50
4.4.2	Proses ekstraksi dengan pencocokan pola menggunakan <i>regular expression</i>	51
4.4.3	Representasi struktur Qur'an Link	53
4.4.4	<i>Spelling correction</i> untuk mencari nomor surat yang eksak	55
4.4.5	Verifikasi sitasi ayat Al-Qur'an	56
4.5	Perancangan Visualisasi	57
4.5.1	Macam visualisasi yang dirancang	57
4.5.2	Rancangan antar muka visualisasi	58
4.6	Perancangan Pengujian	60
4.6.1	Pengujian pencocokan pola (<i>pattern matching</i>)	60
4.6.2	Pengujian <i>spelling correction</i>	61
V	IMPLEMENTASI	62
5.1	Lingkungan Implementasi	62
5.2	Implementasi Akuisisi Data	63
5.2.1	Implementasi <i>crawler</i> untuk artikel Islam di web	63
5.2.2	Implementasi <i>parsing</i> data Al-Qur'an	68
5.2.3	Implementasi <i>crawler</i> untuk tafsir dan asbabun nuzul	72
5.3	Implementasi <i>Auto Extractor</i> Qur'an Link	74
5.3.1	Implementasi pra pemrosesan	74
5.3.2	Implementasi aturan <i>regular expression</i>	75
5.3.3	Implementasi ekstraksi dengan pencocokan pola menggunakan <i>regex</i>	76
5.3.4	Implementasi penanganan ayat ganda	78
5.3.5	Implementasi <i>spelling correction</i>	79
5.3.6	Implementasi verifikasi sitasi ayat Al-Qur'an	81
5.4	Implementasi Visualisasi	82
5.4.1	Implementasi <i>back-end</i> visualisasi	83
5.4.2	Implementasi visualisasi ayat	84
5.4.3	Implementasi visualisasi artikel	85
5.4.4	Implementasi visualisasi hasil pencarian	86
VI	HASIL DAN PEMBAHASAN	87
6.1	Pengujian Sistem dan Hasil	87
6.2	Hasil Pengujian Sistem Ekstraksi	89



6.2.1	Hasil pengujian ekstraksi dengan kombinasi <i>spelling correction</i>	89
6.2.2	Hasil pengujian <i>threshold</i> pada <i>spelling correction</i>	91
6.3	Evaluasi Kesalahan Ekstraksi	94
6.4	Pembahasan Visualisasi Hasil Pencarian	95
6.4.1	Hasil pencarian web	95
6.4.2	Hasil pencarian ayat Al-Qur'an	98
VII KESIMPULAN DAN SARAN		100
7.1	Kesimpulan	100
7.2	Saran	100
DAFTAR PUSTAKA		102