

DAFTAR ISI

DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI.....	14
3.1 <i>SPSE</i>	14
3.2 Web Crawler	16
3.3 Distributed Web Crawler	17
3.4 Web Service	20
3.5 Teori Antrian.....	21
1) <i>PS (Polling System)</i>	21
2) <i>FCFS (First Come First Serve)</i>	21
3.6 Simple Moving Average (SMA).....	22
3.7 GUID/UUID.....	22
3.8 Jarak Jaringan / Network Distance.....	23
BAB IV RANCANGAN PENELITIAN	24
4.1 Langkah Penelitian.....	24
4.1.1 Tinjauan pustaka	24
4.1.2 Analisis dan identifikasi kebutuhan	25
4.1.3 Desain sistem.....	25
4.1.4 Pengembangan sistem	25
4.1.5 Implementasi	25
4.1.6 Pengujian dan penyusunan laporan	25
4.2 Kebutuhan Sistem	25
4.2.1 Perangkat lunak	25
4.2.2 Perangkat keras	26
4.3 Rancangan Sistem	27
4.3.1 <i>Crawler Agent</i>	27
4.3.2 Server Koordinator	28
4.3.3 Komunikasi <i>Crawler Agent</i> dengan Server Koordinator	28
4.3.4 Rancangan Perhitungan Jarak Jaringan	29
4.3.5 Rancangan Metode Pembagian Tugas	30

4.4	Indikator dan Perhitungan	34
4.5	Pengujian Sistem	35
BAB V IMPLEMENTASI		37
5.1	Penyiapan Perangkat Keras	37
5.2	Penyiapan perangkat lunak pendukung	39
5.2.1	Perangkat lunak pendukung pada Server Koordinator	39
5.2.2	Perangkat lunak pendukung pada <i>host Crawler Agent</i>	39
5.3	Sistem DWC (Distributed Web Crawler) pada Server Koordinator	41
5.3.1	User Interface	41
5.3.2	Message Handler	41
5.3.3	Scheduler	42
5.3.4	DB Handler	43
5.4	<i>Crawler Agent</i>	43
5.4.1	Downloader	44
5.4.2	Parser	44
5.4.3	Komunikator	44
5.5	Komunikasi Data	44
5.5.1	Komunikasi <i>Crawler Agent</i> dengan Server Koordinator	45
5.5.2	Komunikasi Internal Server Koordinator	48
5.6	Penyimpanan Data	49
5.6.1	Tabel log	49
5.6.2	Tabel task	49
5.6.3	Tabel task_queue	50
5.7	Implementasi Pembagian Tugas	51
5.8	Perhitungan Kinerja masing-masing Metode Pembagian Tugas	54
BAB VI HASIL DAN PEMBAHASAN		57
6.1	Hasil dan Analisis Akumulasi Jumlah Halaman per Satuan Waktu	57
6.2	Hasil dan Analisis Akumulasi Jumlah Data per Satuan Waktu	60
6.3	Hasil dan Analisis Akumulasi Jumlah Timeout per Satuan Waktu	64
6.4	Hasil dan Analisis Kondisi Jumlah Halaman	66
BAB VII KESIMPULAN DAN SARAN		71
7.1	Kesimpulan	71
7.2	Saran	71
DAFTAR PUSTAKA		73
LAMPIRAN		75

DAFTAR GAMBAR

Gambar 2.1 Distributed Web Crawling (Tomala dkk., 2013).....	7
Gambar 3.1 SPSE pada Universitas Gadjah Mada	14
Gambar 3.2 Proses dalam <i>crawler</i> (Sajja dan Akerkar, 2012).....	17
Gambar 3.3 Ilustrasi struktur halaman dalam website (Ceri dkk., 2013)	18
Gambar 3.4 Urutan proses <i>crawling</i> dengan metode <i>Breadth-first</i>	19
Gambar 3.5 Urutan proses <i>crawling</i> dengan metode <i>Depth-first</i>	19
Gambar 4.1 Langkah Penelitian	24
Gambar 4.2 Rancangan sistem	27
Gambar 4.3 Komunikasi Crawler dengan Server Koordinator	28
Gambar 4.4 Metode perhitungan jarak jaringan	29
Gambar 4.5 <i>Pseude Code</i> untuk metode PS	30
Gambar 4.6 Rancangan Penjadwalan PS	31
Gambar 4.7 <i>Pseude Code</i> untuk metode FCFS	31
Gambar 4.8 Rancangan Penjadwalan FCFS	32
Gambar 5.1 Script launcher.sh	40
Gambar 5.2 Implementasi User Interface	41
Gambar 5.3 Layanan berbasis SOAP Server Koordinator	42
Gambar 5.4 Iterasi mencari nilai minimal jarak jaringan	42
Gambar 5.5 Pemanfaatan ekstensi MySQLi	43
Gambar 5.6 Instalasi <i>Crawler Agent</i> pada <i>host</i>	43
Gambar 5.7 Format task	45
Gambar 5.8 Format result	45
Gambar 5.9 <i>Request</i> layanan registerPeer	46
Gambar 5.10 <i>Response</i> layanan registerPeer	46
Gambar 5.11 <i>Request</i> layanan getTask	47
Gambar 5.12 <i>Response</i> layanan getTask.....	47
Gambar 5.13 <i>Request</i> layanan saveTask	48
Gambar 5.14 <i>Response</i> layanan saveTask	48
Gambar 5.15 Menyimpan <i>network distance</i> pada <i>shared memory</i>	48
Gambar 5.16 Mengakses <i>network distance</i> pada <i>shared memory</i>	49
Gambar 5.17 Implementasi tabel <i>log</i>	49
Gambar 5.18 Implementasi tabel <i>task</i>	50
Gambar 5.19 Implementasi tabel <i>task_queue</i>	51
Gambar 5.20 Implementasi metode PS	52
Gambar 5.21 Implementasi metode FCFS	52
Gambar 5.22 Implementasi metode JJT.....	53
Gambar 5.23 Implementasi perhitungan jarak jaringan	54
Gambar 5.24 Implementasi pemilihan jarak <i>Crawler Agent</i> ke server SPSE terkecil.....	54

Gambar 5.25 Implementasi request untuk <i>crawling</i>	55
Gambar 5.26 Implementasi request untuk <i>crawling</i>	55
Gambar 6.1 Grafik akumulasi jumlah halaman per satuan waktu	60
Gambar 6.2 Grafik akumulasi jumlah data per satuan waktu	63
Gambar 6.3 Grafik konfirmasi akumulasi jumlah halaman dan jumlah data per satuan waktu	63
Gambar 6.4 Grafik akumulasi jumlah timeout per satuan waktu.....	66
Gambar 6.5 Grafik informasi jumlah halaman per satuan waktu	69

DAFTAR TABEL

Tabel 2.1 Perbandingan dengan penelitian terdahulu	9
Tabel 3. 1 Struktur pengalamatan informasi pada server SPSE.....	15
Tabel 3. 2 Perbandingan urutan crawling metode Breadth-first dan Depth-first..	19
Tabel 4. 1 Inisialisasi jarak jaringan	33
Tabel 4. 2 Tabel jarak jaringan	34
Tabel 5.1 Perangkat keras host yang digunakan	37
Tabel 5.2 Spesifikasi perangkat keras yang digunakan	38
Tabel 6.1 Data Server SPSE yang digunakan untuk penelitian	57
Tabel 6.2 Ringkasan rata-rata akumulasi jumlah halaman per satuan waktu	58
Tabel 6.3 Ringkasan rata-rata akumulasi jumlah data per satuan waktu	60
Tabel 6.4 Ringkasan rata-rata jumlah <i>timeout</i> per satuan waktu	64
Tabel 6.5 Ringkasan rata-rata halaman ter- <i>crawling</i> , <i>timeout</i> dan sisa halaman belum ter- <i>crawling</i>	67