

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Persembahan	v
Halaman Motto	vi
PRAKATA	vii
INTISARI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Keaslian Tesis	6
II TINJAUAN PUSTAKA	7
2.1 Pendekatan Manual	7
2.2 Pendekatan Semiotomatis	7
2.3 Pendekatan Otomatis	11
III LANDASAN TEORI	14
3.1 Teknologi Web	14
3.1.1 <i>HyperText Markup Language</i> (HTML)	14
3.1.2 <i>Document Object Model</i> (DOM)	15
3.1.3 XPath	18
3.1.4 Informasi Visual	21

3.2	<i>Cosine Similarity dan Presentation Style Similarity</i>	23
3.2.1	<i>Cosine Similarity</i>	23
3.2.2	<i>Presentation Style Similarity</i>	24
3.3	<i>Automatic Data Record Mining (AutoRM)</i>	25
3.3.1	Penggalian Semua Himpunan <i>Similar C-Records</i> dari Pohon DOM	27
3.3.2	Penggalian <i>Data-Records</i> dari <i>Similar C-Records</i>	35
3.4	<i>Directed Acyclic Graph - Multi-Tree Matching (DAG-MTM)</i>	35
IV	ANALISIS DAN RANCANGAN SISTEM	41
4.1	Analisis Sistem	41
4.2	Rancangan Sistem	44
4.2.1	Ekstraksi Otomatis	45
4.2.2	Induksi XPath Wrapper dengan Algoritme BUCAPP	49
4.2.3	Ekstraksi Menggunakan XPath Wrapper	59
4.3	Rancangan Antarmuka	59
4.3.1	Ekstraksi Otomatis	59
4.3.2	Pemilihan <i>Data Records</i> dan Pelabelan <i>Data Fields</i>	62
4.3.3	Induksi XPath Wrapper dengan Algoritme BUCAPP	63
4.3.4	Ekstraksi Menggunakan XPath Wrapper	64
4.4	Rancangan Pengujian	66
4.4.1	Penyusunan <i>Dataset</i>	66
4.4.2	Metode Pengujian	69
V	IMPLEMENTASI	73
5.1	Pembangunan Sistem	73
5.2	Implementasi Ekstraksi Otomatis	73
5.2.1	Implementasi Algoritme AutoRM	74
5.2.2	Implementasi Algoritme DAG-MTM	83
5.3	Implementasi Induksi XPath Wrapper dengan Algoritme BUCAPP	86
5.3.1	Implementasi Induksi <i>Data Record</i> XPath	88
5.3.2	Implementasi Induksi <i>Data Item</i> XPath	90
5.4	Implementasi Ekstraksi Menggunakan XPath Wrapper	92
VI	HASIL DAN PEMBAHASAN	94
6.1	Kasus yang Tidak Ditangani oleh Algoritme BUCAPP	94

6.2	Perbandingan Efektivitas	95
6.2.1	<i>Data Records Precision</i>	95
6.2.2	<i>Data Records Recall</i>	96
6.2.3	<i>Data Items Precision</i>	98
6.2.4	<i>Data Items Recall</i>	99
6.2.5	<i>Alignment Accuracy</i>	100
6.2.6	Rangkuman	102
6.3	Perbandingan Efisiensi	103
6.4	Waktu Pelatihan	105
VII KESIMPULAN DAN SARAN		106
7.1	Kesimpulan	106
7.2	Saran	106
A FUNGSI-FUNGSI YANG DIGUNAKAN PADA ALGORITME INDUKSI DATA RECORD XPATH		111
B FUNGSI-FUNGSI YANG DIGUNAKAN PADA ALGORITME INDUKSI DATA ITEM XPATH		113