

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Persembahan	v
Halaman Motto	vi
PRAKATA	vii
INTISARI	xiv
ABSTRACT	xv
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	2
1.5 Manfaat Penelitian	3
1.6 Keaslian Penelitian	3
1.7 Metodologi Penelitian	3
1.8 Sistematika Penulisan	5
II TINJAUAN PUSTAKA	6
2.1 Penelitian Terkait Web Usage Mining dan Web Content Mining	6
2.2 Penelitian yang Dilakukan	7
III DASAR TEORI	9
3.1 Web Mining	9
3.2 Web Usage Mining	9
3.2.1 Preprocessing	9
3.2.2 Pattern Discovery	11

3.2.3	Pattern Analysis	11
3.3	Web Content Mining	11
3.4	<i>Sequence Clustering</i>	12
3.4.1	<i>Precedence-based sequence</i> matriks	12
3.4.2	Self Organizing Map (SOM)	13
3.5	Latent Dirichlet Allocation (LDA)	14
3.5.1	Proses Generatif	15
3.5.2	Gibbs Sampling	16
3.6	PrefixSpan	17
IV ANALISIS DAN RANCANGAN SISTEM		19
4.1	Analisis Sistem	19
4.1.1	Gambaran Umum Sistem	19
4.1.2	Analisis Kebutuhan Sistem	20
4.2	Rancangan Sistem	22
4.2.1	Web Usage Mining	22
4.2.2	Web Content Mining	31
4.2.3	Kombinasi Web Usage Mining dan Web Content Mining	37
4.3	Perancangan Proses	41
4.3.1	Rancangan Arsitektur Sistem	41
4.3.2	DAD Level Konteks	42
4.3.3	DAD Level 1	43
4.4	Perancangan Antar Muka Sistem	44
4.4.1	Halaman Cek Konten	44
4.4.2	Halaman Pemodelan Topik	44
4.4.3	Halaman Akuisisi Log	45
4.4.4	Halaman Integrasi Log dan Topik	46
4.4.5	Halaman Clustering	47
4.4.6	Halaman Sequential Pattern Mining	48
4.4.7	Halaman Test Profile	49
4.5	Rancangan Pengujian	49
V IMPLEMENTASI		52
5.1	Pembangunan Sistem	52
5.2	Implementasi Scraping	52
5.3	Implementasi Pemodelan Topik	53

5.3.1	Pembuatan Model	54
5.3.2	<i>Assign</i> Topik ke Artikel	55
5.4	Implementasi Pemrosesan Log	56
5.4.1	Pemisahan Bagian Log	56
5.4.2	Pembersihan Data	57
5.4.3	Integrasi Log dan Topik	57
5.4.4	<i>Sessionization</i>	58
5.5	Implementasi Clustering	58
5.5.1	Representasi Input	58
5.5.2	Pelatihan	59
5.5.3	<i>Session Mapping</i>	60
5.6	Implementasi Sequential Pattern Mining	61
5.7	Implementasi Pembuatan Rekomendasi	63
VI HASIL DAN PEMBAHASAN		66
6.1	Kinerja Sistem	66
6.2	Profil Navigasi	68
6.3	Sistem Rekomendasi	71
6.4	Performa Sistem	75
6.4.1	Pengujian Pertama	76
6.4.2	Pengujian Kedua	76
6.4.3	Pengujian Ketiga	77
6.4.4	Pengujian Keempat	77
6.5	Pembahasan	78
VII KESIMPULAN DAN SARAN		81
7.1	Kesimpulan	81
7.2	Saran	81

DAFTAR TABEL

2.1	Daftar Penelitian Terkait dengan Web Usage Mining dan Web Content Mining	8
3.1	Contoh Basis Data Sekuens	18
4.1	Apache Server Log File	23
4.2	Pemisahan server log file	25
4.3	Identifikasi Pengguna	29
4.4	Identifikasi Pengguna 1	29
4.5	Identifikasi Pengguna 2	30
4.6	Identifikasi Pengguna 3	30
4.7	Identifikasi Session Pengguna 1 (Session 1)	30
4.8	Identifikasi Session Pengguna 1 (Session 2)	31
4.9	Penetapan Random Topik	34
4.10	Jumlah Kata Terhadap Topik	34
4.11	Penetapan Random Topik	35
4.12	Input dari Algoritme PrefixSpan Sequence Topik	41
6.1	Pengujian Pemodelan Topik	66
6.2	Pengujian <i>Data Cleaning</i>	67
6.3	Pengujian Sessionization dan Integrasi Log Topik	67
6.4	Pengujian SOM	67
6.5	Pengujian PrefixSpan	68
6.6	Pemisahan server log file	68
6.7	Pemisahan server log file	69
6.8	Profil Navigasi Cluster 3-8	71
6.9	Profil Navigasi Cluster 12-13	71
6.10	Pengujian Pertama	76
6.11	Pengujian Kedua	76
6.12	Pengujian Ketiga	77
6.13	Pengujian Keempat Melihat <i>Precision</i>	78
6.14	Pengujian Keempat Melihat <i>Recall</i>	78
6.15	Pengujian Keempat Melihat <i>F-Measure</i>	78
6.16	Pengujian Total Waktu Komputasi	79

DAFTAR GAMBAR

3.1	Langkah umum dalam Web Usage Mining (Lalani, 2003)	10
3.2	Intuisi dasar dari LDA (Blei, 2012)	16
3.3	Plate Notation LDA (Blei, 2012)	16
4.1	Alur Sistem	20
4.2	Alur Proses Web Usage Mining	22
4.3	Langkah Data Cleaning	27
4.4	Alur Proses Web Content Mining	31
4.5	Area <i>scraping</i> konten <i>gudegnet</i>	32
4.6	Distribusi topik pada konten artikel	36
4.7	Alur Proses Kombinasi Web Content dan Usage Mining	38
4.8	Hasil Clustering SOM	40
4.9	Arsitektur Sistem	42
4.10	DAD Level Konteks	43
4.11	DAD Level 1	45
4.12	Rancangan Antar Muka Halaman Cek Konten	46
4.13	Rancangan Antar Muka Halaman Pemodelan Topik	46
4.14	Rancangan Antar Muka Halaman Akuisisi Log	47
4.15	Rancangan Antar Muka Halaman Integrasi Log dan Topik	47
4.16	Rancangan Antar Muka Halaman Clustering	48
4.17	Rancangan Antar Muka Halaman <i>Sequential Pattern Mining</i>	48
4.18	Rancangan Antar Muka Halaman Test Profile	49
5.1	Implementasi Kode Scraping Konten <i>Gudegnet</i>	53
5.2	Implementasi Kode Pemodelan Topik	55
5.3	Implementasi Kode Pemodelan Topik	55
5.4	Implementasi Pemisahan Bagian Log File	56
5.5	Implementasi Kode Pembersihan Log File	57
5.6	Implementasi Kode Integrasi Log File & Topik	58
5.7	Implementasi Kode <i>Sessionization</i>	59
5.8	Implementasi Kode Representasi Input SOM	60
5.9	Implementasi Pelatihan SOM	61
5.10	Implementasi Session Mapping	62
5.11	Implementasi PrefixSpan	63

5.12 Implementasi Proses Mendapatkan Rekomendasi	64
5.13 Implementasi Mendapatkan Rekomendasi Topik	65
6.1 Tampilan Uji Rekomendasi	72
6.2 Artikel Random	73
6.3 Rekomendasi Artikel	73
6.4 Penjelasan Rekomendasi	73
6.5 Pola Pada Cluster 10-15	74
6.6 Artikel Selanjutnya	74
6.7 Rekomendasi Artikel Kedua	75
6.8 Penjelasan Rekomendasi Kedua	75