

**PROSIDING KOMMIT 2012
(KOMPUTER DAN SISTEM INTELIJEN)
Volume 7 – 2012**

**TEKNOLOGI INFORMASI DAN KOMUNIKASI
(TIK) UNTUK KETAHANAN NASIONAL**

ISSN: 2302-3740

PENERBIT

Lembaga Penelitian Universitas Gunadarma

Alamat Editor:

Lembaga Penelitian Universitas Gunadarma
Jl. Margonda Raya 100 Pondok Cina
Depok, 16424
Telp. +62-21-78881112 ext. 455
Fax. +62-21-7872829
e-Mail: kommit@gunadarma.ac.id
Laman: <http://penelitian.gunadarma.ac.id/kommit>

Prosiding KOMMIT, Volume 7 - 2012

Editor:

Tety Elida, Moh. Okki Hardian, Wahyu Rahardjo, Fitriainingsih, Tri Wahyu Retno Ningsih

Disain sampul: Wira Catur

Penerbit: Lembaga Penelitian Universitas Gunadarma

Hak cipta © 2012 oleh Universitas Gunadarma. Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi prosiding ini dalam bentuk apapun, baik secara eletronis maupun mekanis, termasuk memfotocopy, merekam atau dengan sistem penyimpanan lainnya tanpa izin tertulis dari penerbit.

ISSN: 2302-3740

DEWAN REDAKSI

Penanggung Jawab:

Dr. Ir. Hotniar Siringoringo, MSc.

Ketua Dewan Editor:

Dr. Ir. Tety Elida Siregar, MM.

Editor Pelaksana:

Moh. Okki Hardian, ST., MT.

Wahyu Rahardjo, SPsi., MSi.

Fitrianingsih, SKom., MMSi.

Tri Wahyu Retno Ningsih, SSas., MM.

Reviewer:

Prof. Dr. I Wayan Simri Wicaksana, S.Si, M.Eng.

Prof. Dr.rer.nat. Achmad Benny Mutiara, SSi, SKom.

Prof. Dr. Busono Soerowirdjo

Prof. Dr. Sarifuddin Madenda

Prof. Dr. dr. Johan Harlan

Prof. Dr. Ir. Eriyatno MSAE.

Dr. Tb. Maulana Kusuma, SKom., MEngSc.

Dr.-Ing. Adang Suhendra, SSi,SKom,MSc.

Prof. Dr. Ir. Kudang Boro Seminar, MSc.

Drs. Agus Harjoko MSc., PhD.

Dr. Ir. Joko Lianto Buliali

PENERBIT

Lembaga Penelitian Universitas Gunadarma

Jl. Margonda Raya 100 Pondok Cina

Depok, 16424

Telp. +62-21-78881112 ext. 455

Fax. +62-21-7872829

e-Mail: kommit@gunadarma.ac.id

Laman: <http://penelitian.gunadarma.ac.id/kommit>

PANITIA PELAKSANA SEMINAR

Penasehat:

Prof. Dr. E.S. Margianti, S.E., MM.
Prof. Suryadi Harmanto, SSi., M.MS.I.
Agus Sumin, S.Si., MM.

Penanggung Jawab:

Prof. Dr. Yuhara Sukra, MSc.
Prof. Dr. Didin Mukodim, MM.

Ketua Pelaksana:

Dr. Ir. Hotniar Siringoringo, MSc.

Wakil Ketua Pelaksana:

Dr. Bertalya

Sekretariat:

Ida Ayu Ari Angreni, ST., MMT.
Dr. Jacobus Belida Blikololong
MS. Harlina, S.Kom., MM.

Sarana Prasarana:

Drs. Hardjanto Sutedjo, MM.
Rino Rinaldo, SE., MM
Riyanto, ST.

KATA PENGANTAR

Pertukaran informasi merupakan kebutuhan masyarakat modern, sehingga Teknologi Informasi dan Komunikasi (TIK) menjadi hal yang sangat penting. Secara kasat mata, setiap orang dapat menyaksikan perkembangan TIK yang sangat pesat. Perkembangan TIK sampai saat ini masih didominasi oleh negara-negara maju. Kondisi ini harus direposisi.

Indonesia memiliki sumber daya manusia yang handal dan banyak, di antaranya berada di perguruan tinggi. Sumber daya manusia ini terkesan bekerja masih sendiri-sendiri. Penelitian di lingkungan perguruan tinggi maupun litbang sering disalahartikan sebagai pemuas akademis, sementara di kalangan industri lebih tertarik pada penyelesaian ekonomis jangka pendek. Permasalahan ini dapat diatasi dengan memulai kolaborasi antara dunia pendidikan, litbang, industri dan pemerintah.

KOMMIT merupakan seminar nasional di bidang komputer dan teknik yang mendukung pengembangan teknologi komputer maupun aplikasi komputer dalam berbagai bidang. Seminar ini bertujuan menyediakan wadah bagi peneliti, akademisi dan praktisi untuk saling bertukar informasi, berdiskusi dan berkolaborasi sehingga dapat menghasilkan produk siap pakai di dalam bidang sistem informasi.

Topik yang menjadi pembahasan pada KOMMIT ke 7 ini adalah: sistem informasi manajemen, sistem informasi geografis, sistem informasi medis, *enterprise resource planning*, *information retrieval*, matematika aplikasi, sistem keamanan, aplikasi multimedia, pengolahan sinyal dan citra, *computer vision*, *open source & open content*, *e-government*, *e-business*, *e-education*, data semantik, *information system interoperability*, *distributed*, *parallel*, *grid*, *P2Pp*, *mobile information management*, *mobile technology*, *green computing*, telekomunikasi dan jaringan komputer, sistem kontrol, instrumentasi dan diagnosis, mekanika dan elektronika, energi terbarukan, *cognitive science*, *soft computing*, *perceptual science*, bioinformatika dan geoinformatika, *collaborative network*, dan *electron devices*.

Artikel yang disajikan pada seminar ini setelah melalui proses *peer review*, berjumlah seratus satu, yang berasal dari 15 Perguruan Tinggi di Indonesia. Beberapa artikel yang terpilih akan di publikasikan pada Jurnal Ilmiah yang diterbitkan oleh Universitas Gunadarma.

Semoga seminar ini dapat memberikan masukan bagi pengembangan teknologi informasi dan komunikasi di negara kita. Kami ucapkan terima kasih kepada para reviewer yang telah bersedia melakukan review, juga kepada pembicara tamu dan nara sumber yang telah berkontribusi pada acara ini, serta kepada semua pihak yang telah membantu proses produksi prosiding ini.

Ketua Pelaksana
Dr. Ir. Hotniar Siringoringo, MSc.

DAFTAR ISI

DEWAN REDAKSI.....	iii
PANITIA PELAKSANA SEMINAR	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vii
DAFTAR ARTIKEL:	
1. <i>Sistem Informasi Manajemen Penanggulangan Kemiskinan (Studi Kasus Kabupaten Ogan Komering Ilir Provinsi Sumatera Selatan)</i> Ahmad Haidar Mirza.....	1
2. <i>Optimasi Pencarian dengan Knowledge Graph</i> Abidin Ali, Dina Rifdalita, Juliana Putri Lestari, Lintang Yuniar Banowosari	11
3. <i>Analisis Teknik Reduksi Data dan Minimalisasi Ukuran File APK pada Mobile Application Pengenalan Budaya Indonesia Berbasis Android Serta Pengembangannya</i> Adhika Novandya, Debyo Saptono	18
4. <i>Aplikasi Manajemen File Berbasis Web untuk Monitoring Status Kegiatan</i> Akhmad Fauzi, Tri Sulistyorini.....	27
5. <i>Penerapan Metode Dijkstra dalam Pencarian Jalur Terpendek pada Perusahaan Distribusi Film</i> Albert Kurnia, Friska Angelina, Windy Dwiparaswati	36
6. <i>Penyembunyian Informasi (Steganography) Audio Menggunakan Metode LSB (Least Significant Bit) Menggunakan Matlab</i> Ari Santoso, Irfan, Nazori AZ.....	42
7. <i>Standardisasi Sistem Informasi Kesehatan Berjenjang Open E-Health Gunadarma Information System, Mewujudkan Layanan Kesehatan Prima</i> Aries Muslim, AB Mutiara, Teddy Oswari, Riyandari Auror, Irdiah Amsawati	51
8. <i>Pengembangan Web sebagai Upaya Penunjang Optimalisasi Produk Asuransi</i> Armaini Akhirson.....	59
9. <i>Protokol Autentikasi Berbasis One Time Password untuk Banyak Entitas</i> Avinanta Tarigan, D.L. Crispina Pardede	67
10. <i>Peningkatan Keamanan Kartu Kredit Menggunakan Sistem Verifikasi Sidik Jari di Indonesia</i> Bima Shakti Ramadhan Utomo, Denny Satria, Lulu Mawaddah Wisudawati.....	72
11. <i>Rancangan Aplikasi Pencarian Barang Pada Metro Pacific Place dengan Menggunakan Macromedia Dreamweaver 8</i> Triyanto, Bramantyo Sukarno, Miftah Andriansyah.....	78

12.	<i>Sistem Pengambilan Keputusan Bela Negara Non-Fisik untuk Daerah Depok dengan Metode AHP (Analytic Hierarchy Process)</i> Damai Subimawanto, Surya Thiono Wijaya, Yusuf Triyuswoyo, I Wayan Simri Wicaksana, Detty Purnamasari.....	85
13.	<i>Penerapan Teknologi Informasi dan Komunikasi (TIK) pada UMKM dengan Menggunakan Technology Acceptance Model (TAM) (Studi Kasus di Depok dan Qingdao)</i> Deboner Hillery, Dharma Tintri, Pandam R Wulandari.....	94
14.	<i>Faktor Kunci Sukses dalam Pelaksanaan Sistem Enterprise Resource Planning</i> Delvita Dita Putri Anggrayni, Dewi Agushinta R.	101
15.	<i>Model Penentuan Posisi Siaga Lift sebagai Pemanfaatan Penghematan Energi pada Sistem Kerja Lift</i> Denmas Muhammad Ridwan, Donny Ejje Baskoro, Faisal Yafi, Lily Wulandari.....	110
16.	<i>Pemanfaatan Jaringan Akses Telepon sebagai Jaringan Broadband Layanan Internet dengan Teknologi Asymmetric Subscriber Line</i> Djasiodi Djasri.....	116
17.	<i>Evaluasi Website JobsDBTM Mobile dengan Metode Usability Heuristic</i> Esty Purnamasari, Helen Wijayanti, Yosfik Alqadri, Dewi Agushinta Rahayu, Fani Yayuk Supomo	123
18.	<i>Perancangan dan Implementasi Sistem Informasi Peralatan dengan Penerapan Konsep Three Tier (Studi Kasus: Gardu Induk Prabumulih UPT Palembang)</i> Evi Yulianingsih, Marlindawati	131
19.	<i>Faktor-Faktor yang Mempengaruhi Minat Nasabah Menggunakan Internet Banking dengan Menggunakan Anjungan Tunai Mandiri (Studi Kasus pada Bank BCA, BRI dan Bank Syariah Mandiri)</i> Faramita Dwitama, Mohammad Abdul Mukhyi	139
20.	<i>Enkripsi Informasi untuk Pengamanan Pesan Singkat pada Telepon Seluler Berbasis Java MIDP</i> Farid Thalib, Melba Mauludina Novalestari	148
21.	<i>Desain Database e-Supermuseum Batik Indonesia</i> Fikri Budiman, Slamet Sudaryanto Nurhendratno	157
22.	<i>Analisis Perbandingan Kinerja Search Engine Menggunakan Penelusuran Precision dan Recall untuk Informasi Ilmiah Bidang Ilmu Kedokteran</i> Sukei, Fitriainingsih.....	164
23.	<i>Membandingkan Web Pengunduhan Perangkat Lunak</i> Fuji Ihsani, Istiana Idha Aulia, Melisa Chatrine Kamu, Anacostia Kowanda, Trini Saptariani.....	172
24.	<i>Analisis dan Verifikasi Formal Protokol Non-Repudiasi Zhang-Shi dengan Logika SVO-CP</i> Hanum Putri Permatasari, Avinanta Tarigan, D. Lucia Crispina Pardede	178
25.	<i>Implementasi Kebijakan E-Government pada Pemerintah Kota Palembang</i> Hardiyansyah.....	185

26.	<i>Aplikasi Pengingat Jadwal Imunisasi Berbasis Android</i> Hauliza Rindhayanti, Lintang Yuniar Banowosari	193
27.	<i>Model Berbasis Ekstraksi untuk Analisis Gaya Berjalan</i> Hustinawaty, Miftahul Jannah, Rd. Fazlur Rahman.....	201
28.	<i>Metoda Penumbuhan Kreativitas Berbasis Web: Studi Pengembangan Produk Kerajinan Tenun Ikat dalam Upaya Melestarikan dan Meningkatkan Nilai Tambah</i> Iman Murtono Soenhadji, Priyo Purwanto, Ida Astuti, Faisal Reza.....	209
29.	<i>Simulasi dan Optimasi Antrian Pelayanan Agen JNE Buaran</i> Isram Rasal, Hardimen Wahyudi, Nadia Rahmah Al Mukarromah, Yuhilza Nahum	218
30.	<i>Aplikasi Data Mining dengan Teknik Decision Tree untuk Mengklasifikasikan Data Pasien Rawat Inap</i> Julius Santony, Sumijan	226
31.	<i>Integrasi Sumber Data Heterogen Menggunakan Ontologi, Studi Kasus: Data Kependudukan Indonesia</i> Kemal Ade Sekarwati, I Wayan Simri Wicaksana.....	235
32.	<i>Pengenalan Ucapan untuk Belajar Bahasa Menggunakan Perangkat Mobile</i> Kezia Velda Roberta, Raden Supriyanto.....	241
33.	<i>Sistem Pakar Pendeteksi Prediksi Kemungkinan Penyakit Stroke</i> Linda Atika.....	247
34.	<i>Analisis Sektor Unggulan dalam Perekonomian DKI Jakarta</i> Lita Praditha, Mohammad Abdul Mukhyi	254
35.	<i>Kapabilitas Proses Konstruksi Perangkat Lunak pada Perusahaan Pengembang Perangkat Lunak di Bali Menggunakan Kerangka Kerja ISO/IEC 15504</i> Luh Gede Surya Kartika, Kridanto Surendro	262
36.	<i>Sistem New Media pada Aplikasi Internet Radio Berbasis Android</i> Lulu Mawaddah Wisudawati, Avinanta Tarigan.....	269
37.	<i>Kajian Awal Hibridisasi Toyota Soluna dengan Konfigurasi Parallel HEV</i> Mohamad Yamin, Agung Dwi Sapto	276
38.	<i>Pemodelan dan Analisis Rem Cakram dan Rem Tromol dengan Software CATIA V5</i> Mohamad Yamin, Darmawan Sebayang.....	283
39.	<i>Deteksi Sonority Peak untuk Penderita Speech Delay Menggunakan Speech Filing System</i> Muhammad Subali, Tri Wahyu Retno Ningsih, M. Kholiq	289
40.	<i>Penerapan Periklanan di Internet dan Pemasaran Melalui E-Mail untuk Meningkatkan Pemasaran Produk UMKM di Wilayah Depok</i> Mujiyana, Lana Sularto, M. Abdul Mukhyi.....	296
41.	<i>Monitoring Sistem Pengendalian Suhu dan Saluran Irigasi Hydroponik pada Greenhouse Berbasis Web</i> Nia Maharani Raharja, Iswanto.....	303

42.	<i>Disain Rangkaian Detektor Mini Doppler</i> Nur Sultan Salahuddin, Paulus Jambormias, Erma Triawati.....	311
43.	<i>Prototipe Sistem Pemrosesan Limbah Medis</i> Nur Sultan Salahuddin, Adi Hermansyah, RR Sri Poenomo Sari	317
44.	<i>Audit TIK pada Sistem Penerbitan Surat Perjalanan Republik Indonesia (SPRI) di Kantor Imigrasi Bogor</i> Nurul Adhayanti, Karmilasari	323
45.	<i>Aplikasi Pencarian Lokasi Sekolah Menggunakan Telepon Selular Berbasis Android</i> Nuryuliani, Selvi Isni Hadisaputri, Miftah Andriansyah.....	331
46.	<i>Faktor Penentu Efektifitas IT Governance: Studi Kasus pada Perusahaan di DKI Jakarta</i> Pandam Rukmi Wulandari, Samuel David Lee, Renny Nur'ainy.....	340
47.	<i>Aplikasi Mobile Panduan Diet Berdasarkan Golongan Darah Berbasis Android</i> Parno, Swesti Mahardini.....	345
48.	<i>Studi Terhadap Konstruksi Model Pengklasifikasi Regresi Logistik</i> Retno Maharesi.....	352
49.	<i>Karakteristik dan Model Matematika Aliran Lumpur pada Pipa Spiral</i> Ridwan.....	360
50.	<i>Implementasi Mikrokontroler untuk Deteksi Drop Tegangan pada Instalasi Sederhana</i> Rif'an Tsaqif As Sadad, Iswanto.....	368
51.	<i>Analisis Pendeteksian Nodul Citra Sinar-X Paru</i> Rodiah, Sarifuddin Madenda, Dewi Agushinta Rahayu.....	377
52.	<i>Composite Range List Partitioning pada Very Large Database</i> Rosni Gonydjaja, Yuli Karyanti	384
53.	<i>Analisis Perbandingan Waktu untuk Layanan Email dan SMS pada Jaringan Interkoneksi untuk Kajian Efektivitas Dukungan Media Komunikasi Dosen-Mahasiswa</i> S N M P Simamora, Karina Datty Putri, Robbi Hendriyanto.....	389
54.	<i>Desain Prototipe Aplikasi Sistem Keamanan pada Rumah Berbasis Pengenalan Wajah dengan Algoritma Jaringan Saraf Tiruan dan Fitur Fft</i> Shinta Puspasari, Hendra.....	398
55.	<i>Analisis Implementasi Algoritma Propagasi Balik pada Aplikasi Identifikasi Wajah Secara Waktu Nyata</i> Shinta Puspasari, Alfian Sucipta.....	405
56.	<i>Sistem Pemantau Ruangan dengan Penangkapan Gambar Otomatis Menggunakan Sensor Infra Merah Pasif</i> Singgih Jatmiko, R. Supriyanto, R.N. Nasution	412

57.	<i>Sistem Pengenalan Ekspresi Wajah Berdasarkan Citra Wajah Menggunakan Metode Eigenface dan Nearest Feature Line</i> Sulistyo Puspitodjati, Tyas Arie Wirana	418
58.	<i>Ekstraksi Data pada Halaman Web Database Mining Akademik Menggunakan Simple Tree Matching (STM)</i> Sumijan, Julius Santony	426
59.	<i>Perancangan dan Implementasi Software Penyelesaian Persamaan Non Linier dengan Metode Fixed Point Iteration</i> Vivi Sahfitri.....	447
60.	<i>Perhitungan Panjang Janin pada Citra Ultrasonografi untuk Memprediksi Usia Kehamilan</i> Wahyu Supriyatin, Bertalya	456
61.	<i>Model Translator Notasi Algoritmik ke Bahasa C</i> Wijanarto, Achmad Wahid Kurniawan	464
62.	<i>Simulasi Dinamika Molekular Sistem Molekul Argon dan Graphene dengan Menggunakan Perangkat Lunak DL_Poly</i> Ahmad Rifqi Muchtar, Wisnu Hendradjit, Agus Samsi.....	473
63.	<i>Pengidentifikasian Otomatis Bentuk Kista Ovarium Menggunakan Deteksi Circle dan Deteksi Tepi Laplacian dan Prewitt.</i> Yenniwati Rafsyam, Jonifan	482
64.	<i>Pengaruh Karakteristik, Sikap dan Pelatihan terhadap Penggunaan Teknologi Informasi dan Kinerja Pegawai untuk Penerapan Pemerintah Elektronik di Pedesaan</i> Yuventus Tyas Catur Pramudi, Karis Widyatmoko	489
65.	<i>Perancangan Sistem Informasi Alur Kerja (Work Flow) Dokumen Pengajuan Proposal Skripsi</i> Zulfandi, Sarip Hidayatullah, Wahyudianto	500
66.	<i>Aplikasi Pengenalan Budaya dari 33 Provinsi di Indonesia Berbasis Android</i> Adhika Novandya, Ajeng Kartika, Ari Wibowo, Yudhi Libriadiany	508
67.	<i>Sistem Informasi Geografis Bengkel Resmi Mercedes-Benz dan BMW di Kota Jakarta Menggunakan Quantum GIS</i> Agustini Dwi Setia Rahayu, Ana Rizki, Ria Awalliya.....	514
68.	<i>Studi Kasus Konflik PT.XXX dengan Pelanggan Kereta Kelas Ekonomi Berdasar Ilmu Teori Organsisasi Umum</i> Albert Kurnia Himawan, Juliana Putri Lestari, Aris Budi Setiawan.....	517
69.	<i>Aplikasi Pengenalan Dasar-Dasar Bahasa Inggris untuk Anak Usia Dini Menggunakan Adobe Flash CS 3 Professional</i> Alfa Marlin, Siti Andini, Sri Wahyuni	519
70.	<i>Eksplorasi Celah Keamanan Piranti Lunak Web Server Vertrigoserv pada Sistem Operasi Windows Melalui Jaringan Lokal</i> Andrias Suryo Widodo, Maria Magdalena Merry, Stefanus Dwi Putra Medisa	524

71.	<i>Sistem Pengambilan Keputusan Kelayakan Sekolah Mendapatkan Status RSBI Studi Kasus SMA RSBI Di DKI Jakarta</i> Ardhani Reswai Yudistari, Odheta, Tryono Taqwa	529
72.	<i>Penerapan Algoritma Kruskal dan Pengimplementasiannya dalam Kasus Pendistribusian Majalah "UG News" Antar Universitas Gunadarma</i> Ardisa Pramudhita, Mahisa Aji Kusuma, Nur Fisabilillah	535
73.	<i>Implementasi Algoritma Dijkstra untuk Menentukan Rute Terpendek Antar Museum di Yogyakarta Berbasis Web</i> Ardo Rama, Citra Ika Wibawati, Rizka Fajriah	538
74.	<i>Pembuatan Aplikasi Permainan Labirin 2D untuk Handphone</i> Aries Afriliansyah	542
75.	<i>Konfigurasi Trixbox Server Untuk VoIP pada Jaringan Peer to Peer</i> Arif Liberto Jacob, Muhammad Muhijar, Ferry Wisnuargo	547
76.	<i>Sistem Penunjang Keputusan Memilih Kriteria Lagu Pop Indonesia yang Baik</i> Ario Halik, Virgiawan Ananda Pratama.....	550
77.	<i>Evaluasi Algoritma Prim dan Kruskal Terhadap Pemasangan Kabel Telepon di DKI Jakarta</i> Atikah Luthfiyyah, Voni, Wahyu Pratama	553
78.	<i>Aplikasi Pemetaan Pusat Perbelanjaan Kota Bekasi Menggunakan Android</i> Awal Arifianto, Muhammad Yunus, Andrika Siman, Agung Rahmat Dwiardi, Deny Nugroho	556
79.	<i>Penerapan Algoritma Greedy pada Studi Kasus Pencarian Rumah Sakit Terdekat di Jakarta Selatan</i> Bagus Fitroh Alamsyah, Maulana Malik Ibrahim, Prakasita Wigati.....	559
80.	<i>Implementasi Algoritma Dijkstra Guna Optimasi Jalur Pendistribusian Produk Seluler</i> Banu Adi Witono, Dhita Angreny, Randy Aprianggi	561
81.	<i>Face Recognition Menggunakan Metode Linear Discriminant Analysis (LDA)</i> Bayu Adi Yudha Prasetya.....	563
82.	<i>Pembuatan Game Arasen untuk Latihan Soal Tes Potensi Akademik Menggunakan RPG Studio</i> Daisy Patria, Hayu Wasna Sari, Riyandari Asrita	570
83.	<i>Pemodelan Spasial Tingkat Kerawanan Kecelakaan Lalu Lintas di Kota Depok</i> Eriza Siti Mulyani, Muhammad Arsah Novel Simatupang	576
84.	<i>Sistem Log Monitoring Jaringan (LAN) Menggunakan Bahasa Pemrograman Pascal</i> Fendy Christian, Stefanus Goutama, Afrilia Nita Anjani.....	582
85.	<i>Website Surat Pembaca Sebagai Media Komunikasi dalam Penyampaian Aspirasi Masyarakat</i> Hamisati Muftia, Nabiurrahmah.....	584

86.	<i>Aplikasi Pendidikan Bagi Anak di Bawah Umur 7 Tahun</i> Helmi, Muhammad Subentra, Randy Aditiya Yusuf	586
87.	<i>Sistem Pencarian Fasilitas Umum Terdekat Menggunakan Augmented Reality dengan Minimum Spanning Tree</i> Hifshan Riesvicky, Prita Dessica, Tatang Fanji Permana	592
88.	<i>Aplikasi Multimedia Audio Video Player dengan Menggunakan Visual Basic .Net 2008</i> Inggrit Parnandes, Rias Astria, Meilisa Ndaru Hermiyanti.....	595
89.	<i>Aplikasi Energy Usage Calculator untuk Menghitung Penggunaan dan Biaya Energi Listrik Berbasis Python Versi 3.2.3</i> M Haidar Hanif, Herio Susanto.....	599
90.	<i>Implementasi Algoritma Kruskal untuk Optimasi Pengangkutan Sampah</i> Meilidyningtyas Cantika Ryadiani, Nurul Ardianingsih, Robby Matheus.....	602
91.	<i>Pemilihan Aplikasi Permainan untuk Perkembangan Motorik dan Simbolik Anak Usia 1 - 7 Tahun</i> Michael Satrio Prakoso, Detty Purnamasari.....	605
92.	<i>Sistem Informasi Geografis SMA di Bogor</i> Muhamad Ramadani Silatama, Narendra Paskarona, Ary Wahyudi.....	608
93.	<i>Pembuatan Website World Watch Shop Menggunakan Magento Commerce</i> Rahma Eka Putri, Septiana Dewi Saputri, Sheila Rizka	614
94.	<i>Pembuatan Aplikasi Pemetaan Tempat Usaha di Sekitar Kampus Depok Gunadarma Menggunakan Android 2.1</i> Rangga Adhitya Pradiptha, Titik Rahayu Mariani, Winda Utari	616
95.	<i>Aplikasi Penjualan Makanan Khas Garut pada Toko Aneka Sari dengan Menggunakan Visual Basic .Net</i> Rangga Septian Putra, Rion Saputra, Ryan Oktario.....	619
96.	<i>Pengembangan E-Government pada Layanan Informasi Publik Pemerintahan Daerah Sulawesi Barat Menuju Good Governance</i> Rizka Fajriah, Windy Dwiparaswati, Aris Budi Setyawan	625
97.	<i>Perlunya Penerapan Teknologi Web Semantik pada Situs Pencarian Lowongan Pekerjaan di DKI Jakarta</i> Robby Matheus Gultom, Tatang Fanji Permana, Aris Budi Setyawan	628
98.	<i>Program Aplikasi Enkripsi dan Dekripsi SMS pada Ponsel Berbasis Android dengan Algoritma DES</i> Rudy Hendrayanto, A. Ramadona Nilawati	631
99.	<i>Penentuan Keputusan untuk Membantu Program Genre Bagi Pasangan Muda</i> Sandi Agung Harseno, Moh. Ropiyudin, Dessy Wulandari.....	634
100.	<i>Pembuatan Aplikasi Pembelajaran Bahasa Jerman Berbasis Mobile Android</i> Satrio Wibisono, Lisda.....	638
101.	<i>Aplikasi Foodcourt Menggunakan Microsoft Visual Studio 2008</i> Tri Hardiyanti, Shelly Gustika Septiani	644

EKSTRAKSI DATA PADA HALAMAN WEB DATABASE MINING AKADEMIK MENGGUNAKAN SIMPLE TREE MATCHING (STM)

*Sumijan*¹
*Julius Santony*²

*Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Mahasiswa S3 TI-Universitas Gunadarma
Komplek Palm Griya Indah No. 1 Marapalam Padang, Sumatera Barat,
Jl. Raya Lubuk Begalung, Padang, Sumatera Barat, Universitas Putra Indonesia “YPTK”*

¹ *soe@upiypk.org*

² *juliussantony@yahoo.co.id*

Abstrak

Web database mining akademik digunakan untuk menemukan informasi yang relevan yang di inginkan oleh pihak akademik mulai dari calon mahasiswa s.d alumni maupun stakeholder web database mining akademik juga dapat menciptakan pengetahuan dari informasi akademik dan non-akademik yang ada, sehingga informasi dalam jumlah data yang banyak dan besar di situs Sisfokampus (Sistem Manajemen Akademik Kampus) tetapi mudah untuk mencarinya dan mengaksesnya. Informasi tersebut berupa semi-structured dengan kode HTML, yang berisi campuran informasi seperti main content (isi utama), profile kampus, navigation panel, Unstructured data (teks bebas), More structured data (database akademik). Dalam penelitian ini melakukan proses ekstraksi dengan menggunakan metoda Simple Tree Matching (STM). Proses ekstraksi Web database mining akademik dengan menggunakan teknik STM untuk menemukan dan mengekstrak informasi dari database akademik dan seluruh layanan pada web tersebut. Dilakukan pengujian dengan parameter precision, recall serta f-measure untuk mengevaluasi kecocokan informasi yang dibutuhkan, Sehingga didapatkan performansi kinerja dari sisfokampus sesuai dengan relevansi informasi yang dibutuhkan. Penggunaan metoda tersebut mempercepat pencarian data dan layanan web terkait informasi akademik dan non akademik. Tantangan dari web database akademik mining adalah jumlah data atau volume data yang pertambahannya semakin lama semakin banyak untuk mempermudah pengaksesan dengan pengembalian data baik online maupun offline dari source teks dari web data mining. Penelitian ini telah memberikan solusi bahwa dengan ekstraksi web data mining terintegrasi dengan metoda STM memperlancar proses pencarian dan pengaksesan informasi akademik baik yang tidak terstruktur (informasi dari portal) bersifat portal, dan terstruktur (Informasi dari aplikasi).

Kata Kunci : *Ekstraksi, Simple Tree Matching, Database Akademik, Web Database Mining, Recall, Precision, f-measure.*

PENDAHULUAN

Saat ini dunia telah memasuki era baru, dimana informasi yang menjadi pemain utama dalam kehidupan manusia. Siapa pun berusaha mendapatkan informasi dan kemudian mengolahnya untuk kemudian digunakan untuk meningkatkan kemampuannya. Informasi pada saat ini

sangat mudah diperoleh terutama dengan perkembangan teknologi Internet yang menyediakan informasi seakan tanpa batas untuk diakses baik menggunakan makro browser (web base) maupun mikro browser (Mobile base).

Hampir semua aspek kehidupan manusia saat ini mengalami perubahan yang mengarah kepada suatu komunitas

berbasis Internet. Dengan perubahan ini batasan ruang dan waktu semakin kabur. Seseorang dapat melakukan aktivitas keseharian maupun aktivitas kerjanya dari rumah, kantor, atau bahkan dari Warung Internet tanpa adanya kekangan waktu.

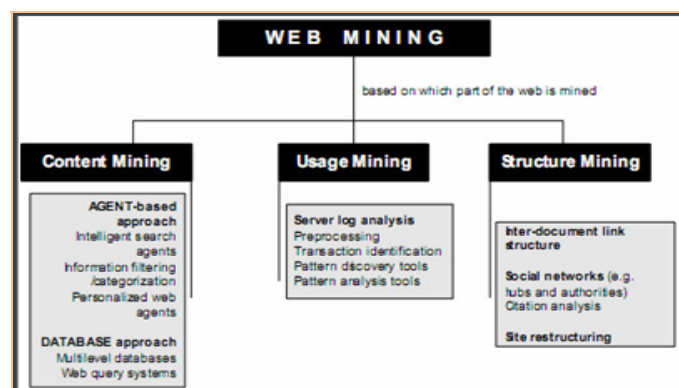
Bidang perhatian yang berkembang dalam teknologi Internet terus berkembang setiap hari, mulai dari bidang hiburan, bidang pariwisata, bidang teknologi, bidang kedokteran, bahkan sampai pendidikan yang dikenal dengan Sistem informasi manajemen akademik. Selanjutnya disingkat SisfoKampus adalah kemampuan untuk mengelola akademik. Sisfo Kampus menjadi pilihan bagi perguruan tinggi yang ingin menerapkan Sistem Informasi Manajemen (SIM) Akademik terpadu sebagai solusi untuk meningkatkan efisiensi dan efektifitas dalam pelayanan kepada stakeholder (Calon mahasiswa, mahasiswa, dosen, manajemen fakultas/universitas, BAUK, karyawan, alumni, dan pengguna lulusan).^{[14][15]}

Web Mining

Perkembangan dan terdistribusinya informasi di Internet merupakan suatu pengetahuan yang sangat bernilai. Dengan semakin pesatnya jumlah informasi yang tersedia, maka dibutuhkan suatu *tool* bagi pengguna untuk mendapatkan informasi tertentu secara efektif dan efisien. Mes-

kipun mesin pencari atau *search engine* dapat menelusuri dan mencari informasi yang diinginkan, tetapi kadang kala dibutuhkan *tool* yang dapat mengambil isi (*content*) dari sebuah halaman Web yang diperoleh. Untuk itu penelitian dan pengembangan teknologi untuk menggali informasi ini terus berkembang. Web mining adalah suatu bidang penelitian yang sangat berkembang saat ini, meskipun saat ini belum diperoleh kesepakatan mengenai definisi Web mining itu sendiri.

Web mining adalah sebuah integrasi beberapa teknologi, seperti data mining, statistik, informatika, dan sebagainya (Jicheng, 1999). Web mining adalah suatu aktifitas untuk mengidentifikasi pola p yang terkandung dalam sebuah koleksi dokumen C , yang dapat dituliskan sebagai pemetaan $\xi : C \rightarrow p$. Definisi tersebut di atas menunjukkan bahwa Web mining mempunyai kemiripan dengan data mining. Tetapi Web mining memiliki karakteristik khusus, antara lain adalah sumber yang digunakan adalah dokumen web. Selain itu pola yang dapat diperoleh dari Web mining adalah isi dari dokumen web atau struktur dari Web, sebab sebuah dokumen Web berisi informasi dan *hyperlink*. Selain itu Web mining mengolah data yang bersifat semi terstruktur sedangkan data mining mengolah data yang lebih terstruktur seperti basis data.



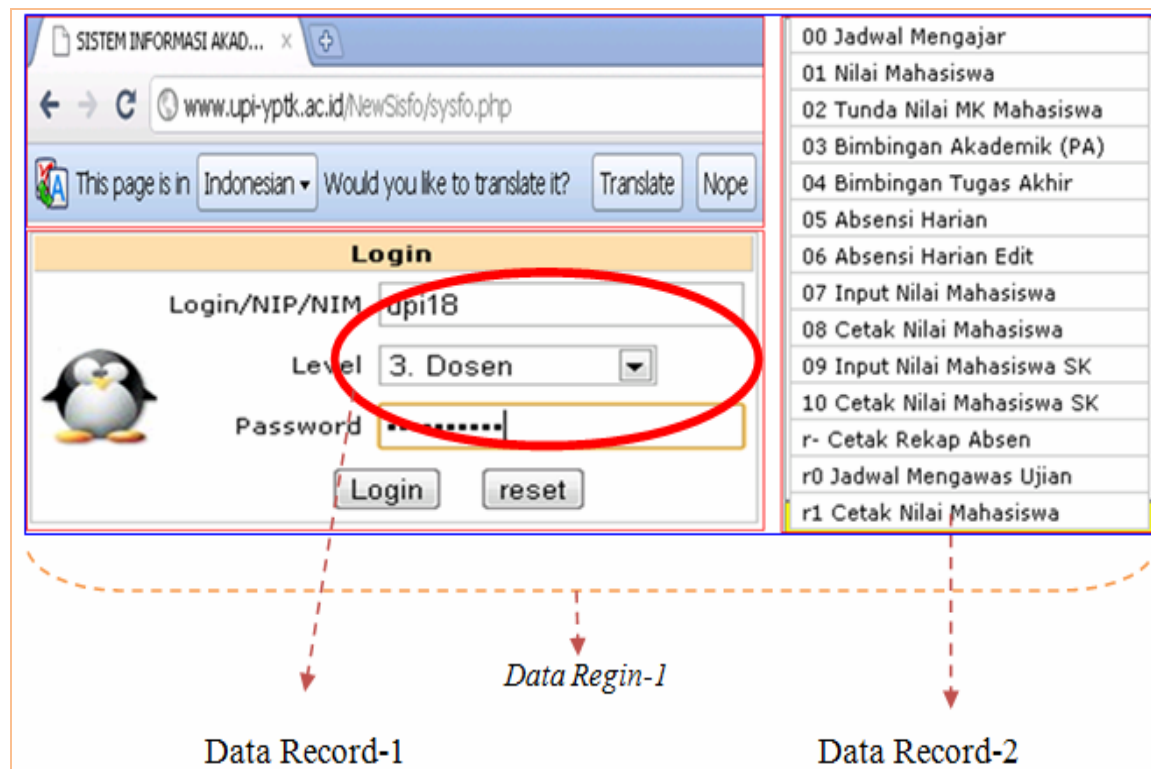
Gambar 1. Taksonomi Web Mining

Sumber: Jicheng (1999)

Pencarian informasi Web adalah sebuah proses untuk memperoleh himpunan bagian dokumen S yang sesuai dengan *query* q tertentu dari kumpulan dokumen C , atau dapat dituliskan sebagai berikut: $\xi : (C, q) \rightarrow S$. Pencarian informasi Web mempunyai tujuan yang berbeda dengan Web mining. Web mining adalah proses lebih lanjut dari pencarian informasi Web. Meskipun demikian keduanya saling melengkapi. Cara pandang ini mencoba untuk memodelkan data pada web dan mengintegrasikannya agar dapat digunakan sebaik mungkin.

1. Cara pandang information retrieval: cara pandang ini membantu atau memperbaiki kualitas informasi yang ditemukan dalam web atau dengan kata lain menyaring informasi didasarkan pada keinginan pemakai. Pada Gambar

2 dapat dilihat pengertian secara umum sebuah *data region* dan sebuah *data record*. Sebuah *data region* adalah daerah yang sangat relevan dari halaman web, seperti daerah pada situs web yang berisi sebuah daftar menu dosen membentuk daerah data (Sumijan 2007). Sebuah *data record* adalah sekumpulan data yang bersama-sama merepresentasikan entitas bermakna yang berdiri sendiri, seperti daftar menu dosen dalam *data region* pada situs web (Sumijan 2007). Algoritma MDR termasuk teknik *unsupervised learning*, yaitu sistem diberikan hanya satu halaman web dengan banyak *data record*, kemudian sistem mengekstrak data secara otomatis.



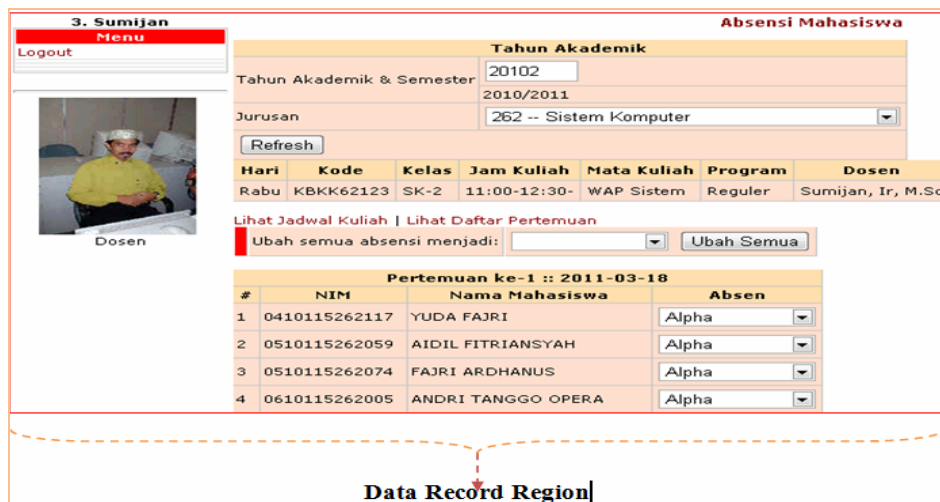
Gambar 2. Halaman Web yang Menjelaskan Data Record dan Region
Sumber : Sumijan (2007)

Menurut paper rujukan (Bing Liu, Robert Grossman, Yanhong Zhai, 2003) berasumsi bahwa *data record* pada halaman web biasanya terdapat pada tag HTML dalam bentuk yang berhubungan dengan tabel dan form, misalnya tag *table*, *form*, *tr*, *td* dan lain sebagainya. Pada penulisan ini, algoritma MDR didasarkan pada dua pengamatan, yaitu:

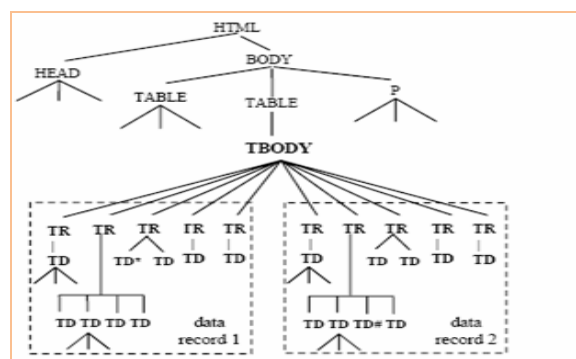
2. *Data region* (atau *data record region*) adalah sekumpulan *data record* berisi deskripsi dari kelompok obyek serupa yang ditampilkan secara khusus pada halaman web dengan region berdekatan dan disusun menggunakan tag HTML yang serupa. Seperti Gambar 3, tiga *data record* ditampilkan pada satu

region yang berdekatan serta disusun menggunakan tag HTML.

3. Struktur bersarang dari tag HTML pada halaman web biasanya membentuk sebuah tag tree dan sekumpulan *data record* serupa dibentuk oleh beberapa node children dari subtree pada node parent yang sama. Contohnya pada Gambar 4, merupakan tag tree untuk halaman web pada Gambar 5. Misalnya setiap notebook (atau sebuah *data record*) pada gambar 4 diekstrak ke dalam 5 node TR dengan bagian tree di bawah node parent TBODY yang sama pada Gambar 5, sehingga terdapat dua *data record* pada dua kotak garis putus-putus.

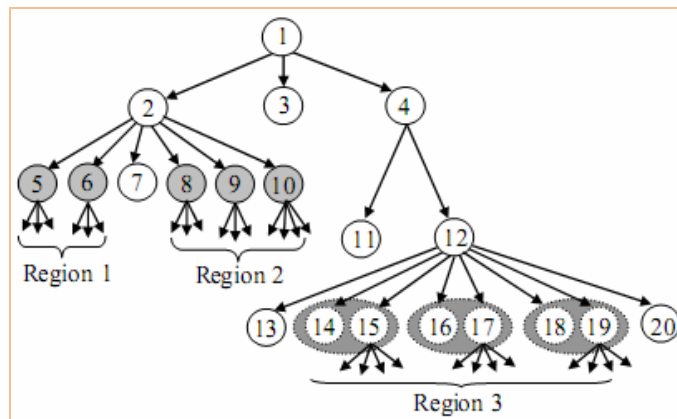


Gambar 3. Halaman Web dengan Dua Data Record
 Sumber: Sumijan (2007)



Gambar 4. Contoh Tag Tree

Sumber : Bing Liu, Robert Grossman, and Yanhong Zhai (2003)



Gambar 5. Ilustrasi dari Node Generalized dan Data Region
Sumber : Bing Liu, Robert Grossman, and Yanhong Zhai (2003)

Pada penulisan ini, untuk melakukan teknik *mining data record* pada sebuah halaman web terdapat tiga langkah, (Abhijit Raorane¹ & R.V.Kulkarni², 2011) yaitu:

1. Membangun *HTML tag tree*.
2. *Mining data region* dalam halaman web dengan menggunakan tag tree dan perbandingan string. *Mining data region* dilakukan terlebih dahulu, karena sangat susah dalam *mining data record* secara langsung. Oleh karena itu, *mining data region* dilakukan untuk mendapatkan *data record* di dalam *data region* tersebut. Contohnya, pada Gambar 3, menemukan satu *data region* di bawah node TBODY.
3. Mengidentifikasi *data record* dari setiap *data region*. Contohnya, pada Gambar 3, langkah ini menemukan *data record* 1 dan *data record* 2 pada *data region* di bawah node TBODY.

Membangun HTML Tag Tree

Pada penulisan ini, hanya menggunakan tag-tag dalam perbandingan string untuk menemukan *data record*. Kebanyakan tag-tag HTML bekerja dalam pasangan. Setiap pasang terdiri dari sebuah tag pembuka (*opening tag*) dan sebuah tag penutup (*closing tag*), masing-

masing diidentifikasi dengan “< >” dan “</ >”. Dalam setiap pasangan tag dapat berhubungan dengan pasangan tag yang lain, sehingga mengakibatkan blok bersarang pada kode HTML. Pembangunan sebuah *tag tree* dengan menggunakan kode HTML secara natural. Pada *tag tree*, setiap pasang dari tag dipertimbangkan menjadi satu node.

Mining Data Region

Langkah ini adalah *me-mining* setiap *data region* pada halaman web yang berisi *data record* serupa, tetapi tidak dapat *me-mining data record* secara langsung, karena susah, pertama kali yang dilakukan adalah *me-mining node generalized* pada halaman web. Sekumpulan *node generalized* yang berdekatan membentuk sebuah *data region*. Dari setiap *data region*, akan mengidentifikasi *data record* yang sesungguhnya.

Node generalized (atau sebuah node kombinasi) dengan panjang r dimana terdiri dari r ($r \geq 1$) node pada *HTML tag tree* dengan dua karakteristik sebagai berikut:

1. Semua node yang mempunyai parent yang sama.
2. Node-node yang berdekatan.

Node generalized menjelaskan bahwa sebuah obyek (atau *data record*) mungkin terisi dalam node-node tag sibling yang jumlahnya lebih daripada satu. *Node generalized* berbeda dengan node tag. Sedangkan node tag adalah setiap node tag pada tag tree pada Gambar-5.

Data region adalah kumpulan dari dua atau lebih *node generalized* dengan mempunyai beberapa karakteristik sebagai berikut :

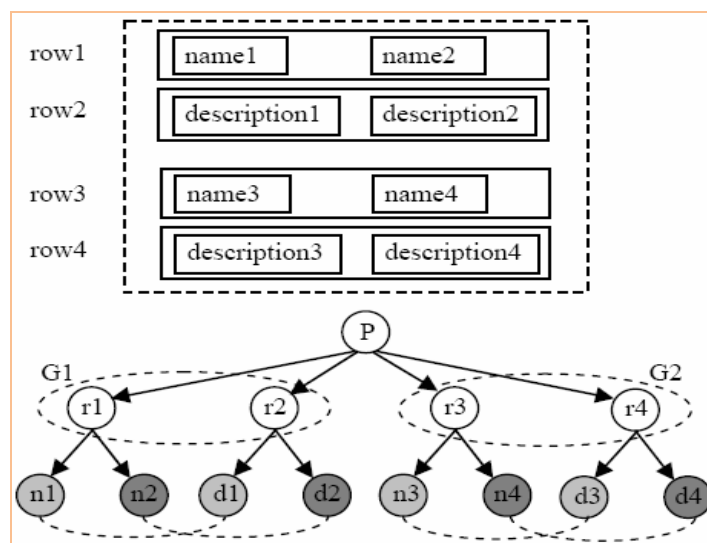
1. Semua *node generalized* mempunyai parent yang sama.
2. Semua *node generalized* mempunyai panjang yang sama.
3. Semua *node generalized* yang berdekatan.
4. Normalisasi edit distance (perbandingan string) antara *node generalized* yang berdekatan lebih kecil daripada batasan yang telah ditentukan.

Pada Gambar 5, menunjukkan bahwa dapat membentuk dua *node generalized*. Pertama terdiri 5 node children dari TR awal untuk TBODY, dan kedua yaitu 5 node children dari TR berikutnya untuk TBODY. Meskipun *node generalized* pada *data region* mempunyai panjang yang sama (mempunyai jumlah

node children yang sama dari satu parent pada tag tree) tetapi node dengan level terbawahnya dapat sangat berbeda.

Perbedaan antara *node generalized* dengan *data region* dijelaskan pada Gambar 5 dengan menggunakan sebuah *tag tree* buatan dan nomer ID yang menggambarkan node tag pada *tag tree*. Daerah gelap adalah *node generalized*. Node 5 dan 6 adalah *node generalized* dengan panjang 1 dan mendefinisikan *data region* dengan label 1 jika kondisi *edit distance* terpenuhi. Node 8, 9, dan 10 adalah *node generalized* dengan panjang 1 dan mendefinisikan *data region* dengan label 2 jika kondisi *edit distance* terpenuhi. Pasangan dari node (14, 15), (16, 17) dan (18, 19) adalah *node generalized* dengan panjang 2 dan mendefinisikan *data region* dengan label 3 jika kondisi *edit distance* terpenuhi.

Pada penulisan ini, mengasumsikan bahwa node-node membentuk sebuah *data region* dari parent yang sama. Contohnya, tidak seperti *data region* dimulai dari node 7 dan akan berakhir pada node 14. Sebuah *node generalized* mungkin bukan mempresentasikan sebuah *data record*, tetapi itu digunakan untuk menemukan *data record*.



Gambar 6. Multiple-record Data Region: masing-masing Node terdiri Lebih dari Satu Non-contiguous Data Record

Sumber: Nitin Jindal and Bing Liu (2007)

Mengidentifikasi Data records

Data yang telah terbentuk dalam region-region (direpresentasikan dengan *node generalized*) sangat bermacam-macam kombinasinya ,berikut 2 kasus utama dalam mengidentifikasi data record

Non – Contiguous Data Records : Case 1

Dalam beberapa halaman web dalam mendeskripsikan sebuah data object (record) tidak dapat dianalisis kedekatannya berdasarkan HTML code-nya (*contiguous segment*) . Amati dalam gambar berikut : untuk kondisi pada Gambar 6 setiap *children nodes* korespondensi - nya pada setiap *tag node* dalam bentuk *generalized node* (dalam kasus *non-cotiguous* data record) dimana *r* merepresentasikan *row* ,*n* merepresentasikan *name* dan *d* merepresentasikan *descriptions*, sedangkan G1 and G2 adalah *generalized nodes* .(n1,d1), (n2,d2), (n3,d3) . [4][5]

Dalam kasus ini ,data regions terdiri dari dua *generalized nodes*, dimana masing-masing *generalized nodes* terdiri dari dua tag nodes (dua rows), yang mengindikasikan bahwa dua *tag node* (*rows*) tersebut diatas tidak ada kesamaan satu sama lain . Tetapi masing-masing *tag node* memiliki jumlah *children node*

yang sama dan *children node* ini memiliki kesamaan satu sama lain.

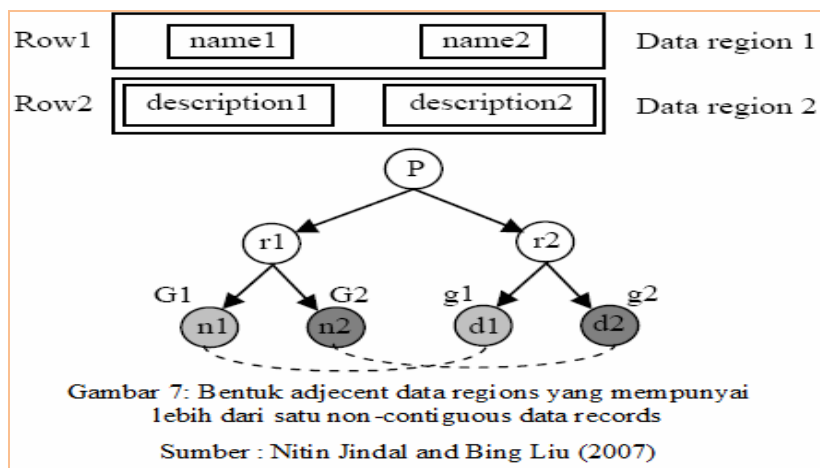
Untuk kasus ini kita dapat menuliskannya menjadi satu *row lists* untuk nama yang diambil dari dua object dalam dua *cells* dan *row lists* selanjutnya. Sehingga dapat ditulis name 1,name 2, description 1,description 2,name 3,name 4, description 3, description 4.

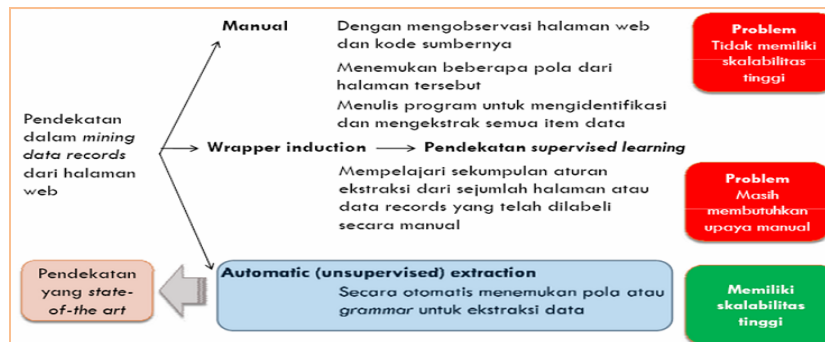
Non-Contiguous Data Records:Case 2

Kasus diatas terdiri dari dua atau lebih *data regions* dari multiple *data records* ,dimana dalam gambar diatas row 1 dan row 2 tidak memiliki kesamaan satu sama lain, dimana bentuk row 1 dari sebuah data regions dan bentuk row 2 berasal dari data region yang berbeda (Gambar 7).

METODE PENELITIAN

Ekstraksi informasi web merupakan masalah mengekstraksi item-item informasi target dari halaman-halaman Web. Dua masalah umum: Ekstraksi informasi dari teks bahasa natural, Ekstraksi data terstruktur dari halaman-halaman web, dengan Program yang mengekstrak data terstruktur dari halaman web disebut *wrapper*.





Gambar 8. Pendekatan dalam Ekstraksi Data

Sumber : Bing Liu, Robert Grossman, and Yamhong Zhai (2003)

Dalam data extraction ini kita akan menerapkan sebuah teknik yang dinamakan dengan *partial tree alignment*, yang kunci pokoknya adalah bagaimana mencocokkan *corresponding data item* atau *field* dari data semua data records. Ada dua langkah penting dalam data extraction:

1. Membuat satu *root tag tree* untuk masing-masing data record :
 Setelah semua data record telah teridentifikasi, *sub-trees* pada masing-masing data records di susun ulang ke dalam *single tree*. Masing-masing data record ada kemungkinan memiliki lebih dari satu *sub-trees* dari sebuah original tag tree pada sebuah halaman, dan masing-masing data record mungkin tidak memiliki kesamaan (Case 1 dan Case 2 pada kasus Pengidentifikasian data record). *Sub-step* ini diperlukan untuk menyusun *single tree* untuk masing-masing data record (sebuah root node buatan yang dapat di tambah setiap saat).
2. *Partial tree alignment: tag trees* dari semua data dalam masing-masing data region di *aligned* menggunakan metode *partial alignment* berdasarkan *tree matching*

Dalam data extraction akan melalui berbagai tahapan yaitu sebagai berikut:

Tree Edit Distance

Tahap *tree edit distance* adalah mengukur kesamaan antara dua *trees* A dan B (*root trees* yang telah terlabel terurut) berdasarkan cost dalam sebuah *minimum set* dari operasi yang diperlukan untuk mentransform A kedalam B. Menurut formula klasik, kumpulan dari operasi yang digunakan untuk menentukan tree edit distance adat tiga tahap: *node removal*, *node insertion*, *node replacement*. Sebuah cost biasanya di assign terlebih dahulu pada masing-masing operasi.

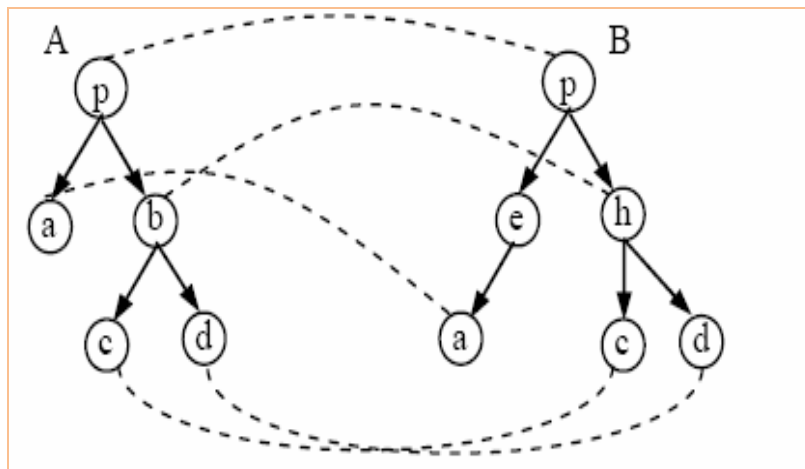
Masalah *tree edit distance* adalah penemuan *minimum-cost mapping* antara dua *tree*, berikut adalah salah satu contoh konsep mapping tersebut diatas:

Misalkan X adalah *tree* dan misalkan $X[i]$ adalah node ke- i dari *tree* X dalam tahapan *preorder*. *Mapping* M antara *tree* A yang berukuran $n1$ dan *tree* B yang berukuran $n2$ adalah kumpulan dari *pairs* (i,j) yang telah terurut dari setiap *tree*, berikut adalah algoritma untuk kondisi $(i1,j1)(i2,j2) \in M$:

- (1) $i1 = i2$ iff $j1 = j2$;
- (2) $A[i1]$ is on the left of $A[i2]$ iff $B[j1]$ is on the left $B[j2]$;
- (3) $A[i1]$ is an ancestor of $A[i2]$ iff $B[j1]$ is an ancestor of $B[j2]$.

Masing-masing node dihilangkan satu kali saat melakukan *mapping* dan diurutkan antara *sibling node* dan *hierarchical relation* antara kedua node yang

telah ada. Gambar 9 adalah gambar yang menunjukkan *mapping*.



Gambar 9. Menunjukkan Mapping Node
 Sumber : Bing Liu, Robert Grossman, and Yamhong Zhai (2003)

Berikut beberapa algoritma yang telah dilakukan untuk mencari *minimum set* dari operasi untuk men-transform satu *tree* ke dalam lainnya, dimana dalam setiap formula memiliki kompleksitas kuadratik ($O(n_1 n_2 h_1 h_2)$), dimana n_1 dan n_2 ukuran sebuah *tree* dan h_1 dan h_2 adalah kedalaman sebuah *tree* hal ini juga ditunjukkan pada kasus *tree* yang tidak teratur.

Simple Tree matching (STM)

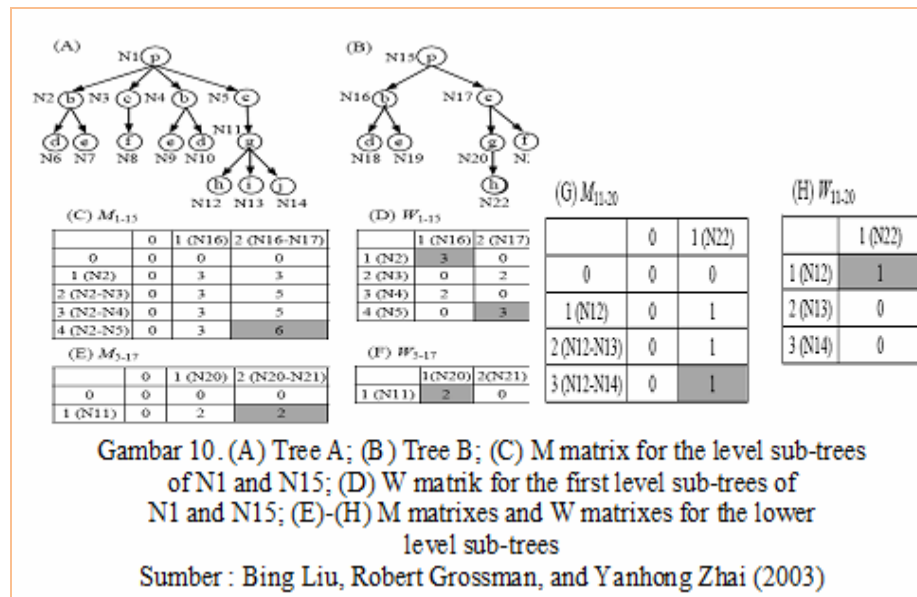
Pada umumnya *mapping* dapat dilakukan antara node a di *tree A* dan node b di *tree B* secara silang, disini juga terdapat *replacement* node b di dalam A node h di dalam B . Dalam kasus ini kita menggunakan *restricted matching algorithm*, yang pertama di usulkan untuk membandingkan dua program computer dalam *software engineering*. Algoritma ini disebut dengan *simple tree matching* (STM). STM ini mengevaluasi similarity dari dua *trees* yang menghasilkan *maximum matching* dalam sebuah dynamic programming dengan complexity $O(n_1.n_2)$, dimana n_1 dan n_2 ukuran dari *trees A* dan B (tidak menggunakan *replacement* dan *no level crossing* masih diijinkan).

Misal A dan B dua buah *tree* dan $i \in A, j \in B$ adalah node di A dan B . A *matching* antara dua *trees* ditujukan untuk *mapping M* dimana setiap pasangan $(i, j) \in M$ dan i dan j adalah *non-root nodes*, $(\text{parents}(i), \text{parent}(j)) \in M$. A *maximum matching* adalah *matching* dengan jumlah pairs maximum.

Misal $A = \langle R_A, A_1, A_2, \dots, A_m \rangle$ dan $B = \langle R_B, B_1, B_2, \dots, B_m \rangle$ dua *trees*, dimana R_A dan R_B adalah root dari A dan B dan A_i, A_j adalah level pertama *sub trees* ke i dan j dari A dan B . ketika R_A dan R_B terdiri dari identical symbol, maximum matching antara A dan B adalah $M_{A,B} + 1$, dimana $M_{A,B}$ adalah *maximum matching* antara $\langle A_1, A_2, \dots, A_m \rangle$ dan $\langle B_1, B_2, \dots, B_m \rangle$. $M_{A,B}$ dapat diperoleh dari *dynamic programming scheme*:

Jika maximum matching antara A_m dan B_m lebih besar dari pada *maximum matching* antara A_m dan $B_i (1 \leq i < m)$, dimana $M_{A,B}$ adalah *maximum matching* antara $\langle A_1, A_2, \dots, A_{m-1} \rangle$ dan $\langle B_1, B_2, \dots, B_{m-1} \rangle$ ditambah *maximum matching* antara A_m dan B_m .

Terkadang, $M_{A,B}$ adalah *maximum matching* antara $\langle A_1, A_2, \dots, A_m \rangle$ dan $\langle B_1, B_2, \dots, B_{n-1} \rangle$ atau antara $\langle A_1, A_2, \dots, A_{m-1} \rangle$ dan $\langle B_1, B_2, \dots, B_n \rangle$



Dalam algoritma Simple tree matching, root dari *A* dan *B* dibandingkan pertama kali (line 1). Jika root terdiri dari *distinct symbol* maka dua *tree* tersebut tidak memiliki kesamaan sama sekali. Jika root terdiri dari *identical symbol* maka algoritma STM secara recursive menemukan *maximum matching* antara *first-level sub-trees* dari *A* dan *B* dan menyimpan hasilnya dalam matriks *W* (line 8). Berdasarkan pada matriks *W*, *dynamic programming maximum matching* antara dua *tree A* dan *B*.

1. **Algorithm:** STM (*A*, *B*)
2. **if** the roots of the two trees *A* and *B* contain distinct symbols
3. **then return** (0);
4. **else** *m* := the number of first-level sub-trees of *A*;
5. *n* := the number of first-level sub-trees of *B*;
6. Initialization: $M[i, 0] := 0$ for $i = 0, \dots, m$;
7. $M[0, j] := 0$ for $j = 0, \dots, n$;
8. **for** $i = 1$ to m **do**
9. **for** $j = 1$ to n **do**
10. $M[i, j] := \max(M[i, j-1], M[i-1, j], M[i-1, j-1] + W[i, j])$;

11. where $W[i, j] = \text{Simple_Tree_Matching}(A_i, B_j)$
12. **endfor**;
13. **endfor**;
14. **return** ($M[m, n] + 1$)
15. **endif**

Tree Matching untuk Regular Part Pada metode STM, sebelum proses tree matching, terlebih dahulu dilakukan proses grouping. Tujuan dari proses grouping adalah untuk mencari tag-tag didalam halaman web database mining akademik yang berada dalam satu sub tree. Metode grouping ini cukup sederhana yaitu dengan mengecek parent dan tree level Tree Matching (TM) yang sibling. Jika TM yang sibling memiliki parent dan level tree yang sama, maka TM tersebut berada dalam subtree yang sama sehingga dimasukkan dalam group yang sama dengan cara memperluas jangkauan group region. Sebaliknya jika tidak memiliki parent dan level tree yang sama maka group yang baru akan diciptakan. Hasil dari proses grouping ini berupa sub tree yang kemungkinan regular atau content parts. Sehingga proses

selanjutnya yang akan dilakukan adalah proses untuk membedakan regular part dari bagian yang lainnya. Penentuan apakah suatu sub tree di dalam Web page adalah regular part dihitung dengan membandingkannya dengan sub tree di Web page lain yang sharing rigid pattern. Fungsi lain yang perlu diperhatikan pada algoritma tree matching adalah fungsi compact Group. Fungsi ini didesign untuk menangani field yang berulang pada regular parts. Pada algoritma compact Group, dilakukan pengecekan apakah node sibling memiliki patten sequence yang sama. Setelah proses tree matching, sistem telah dapat mengidentifikasi baik common parts (dengan sequence matching) dan regular parts. Maka sisa bagian dari halaman web merupakan bagian content parts yang dibutuhkan. Dengan mengembalikan bagian ini akan didapatkan content halaman web akademik yang dibutuhkan.

Pengukuran Performansi

Dalam penelitian ini ada tiga parameter yang digunakan untuk mengetahui tingkat keberhasilan sistem yang dibangun, yaitu precision, recall, dan F Measure menurut rujukan (Prof Joydeep Ghosh dan Prof Dik Lee, 2009). Kebutuhan informasi yang diterjemahkan ke dalam query Relevansi dinilai relatif terhadap informasi tidak perlu query. Misalnya, Informasi yang diutuhkan: Saya mencari informasi apakah IPK mahasiswa jurusan Sistem Informasi dan Teknik Informatika lebih baik dari program studi Sistem Komputer dan Psikologi. Mengevaluasi apakah dokumen yang terkait dengan database akademik memenuhi kebutuhan informasi, apakah ia memiliki IPK yang relevan atau tidak

relevan. Untuk itu IR (Information Retrieval) menguji performansi kinerja ekstraksi web data mining dimana patokan yang digunakan seperti terlihat pada Gambar 10a. dan 10b.

Precision, Recall, dan F-Measure

Precision merupakan perbandingan dari pengembalian dokumen yang relevan. Bisa dikatakan

bahwa precision merupakan pecahan atau fraction dari dokumen yang didapatkan yang relevan dari informasi yang diinginkan sesuai rujukan (Prof Joydeep Ghosh dan Prof Dik Lee, 2009).

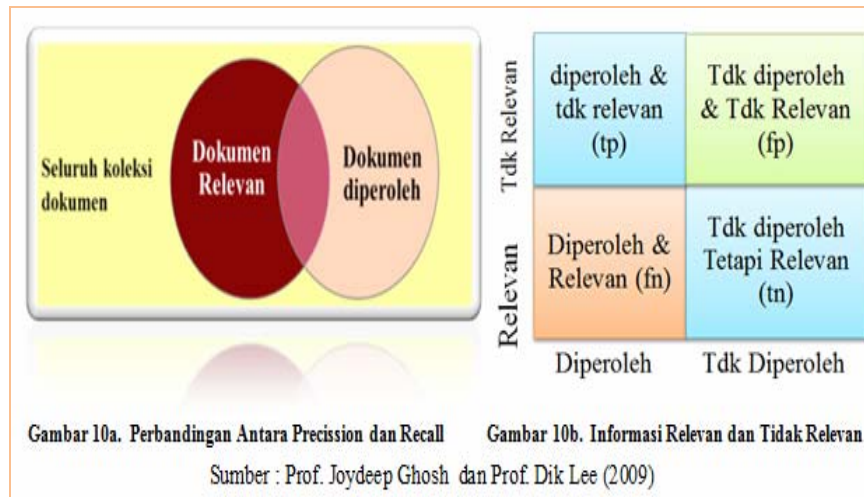
$$Precision(P) = \frac{(tp)}{(tp + fp)} \quad (1)$$

Recall merupakan perbandingan dari dokumen relevan yang sudah dikembalikan. Atau bisa dikatakan bahwa recall merupakan fraction dari dokumen yang relevan berdasarkan query atau permintaan yang sukses dijawab atau dikembalikan. Gambar 2 menyajikan ilustrasi dari Precision dan Recall.

$$Recall(R) = \frac{(tp)}{(tp + fn)} \quad (2)$$

F-Measure didefinisikan sebagai kombinasi dari recall dan precision dengan bobot yang seimbang, rumusan dari F-Measure adalah sebagai berikut :

$$F - Measure(F) = \frac{(tp + tn)}{(tp + fp + fn + tn)} \quad (3)$$



Deskripsi Sistem

Objektivitas yang ingin dicapai dari hasil penelitian ini yakni diharapkan kalangan akademik dapat mengekstrak informasi dari konten web database mining akademik secara otomatis dan mendapatkan nilai akurasi yang dicapai dengan metode STM dan mekanisme kerja sistem ekstraksi secara umum dapat dilihat pada gambar 10c. Sistem dibangun menggunakan PHP dan database MYSQL dengan user interface yang user friendly sehingga hasil dari sistem akan mudah untuk dioperasikan. Data mentah yang masih dalam bentuk halaman web dan masih mengandung common parts dan regular parts akan diletakkan dalam direktori sesuai nama situs www.upi-ytk.ac.id dengan sebelumnya dilakukan proses crawling (crawling yang juga dikenal dengan web spider atau web robot) adalah suatu program atau script otomatis yang menjelajahi WWW dengan menggunakan sebuah metode atau cara yang otomatis. Nama-nama yang jarang digunakan pada sebuah web crawler adalah ants, automatic indexers, bots, worms (Kobayashi & Takeda, 2000). Suatu website dengan menggunakan tools (perkakas) Teleport ataupun secara manual. Setelah dilakukan proses peng-ekstrakan, content yang didapatkan disimpan di database mining akademik untuk dihitung akurasinya dan selanjut-

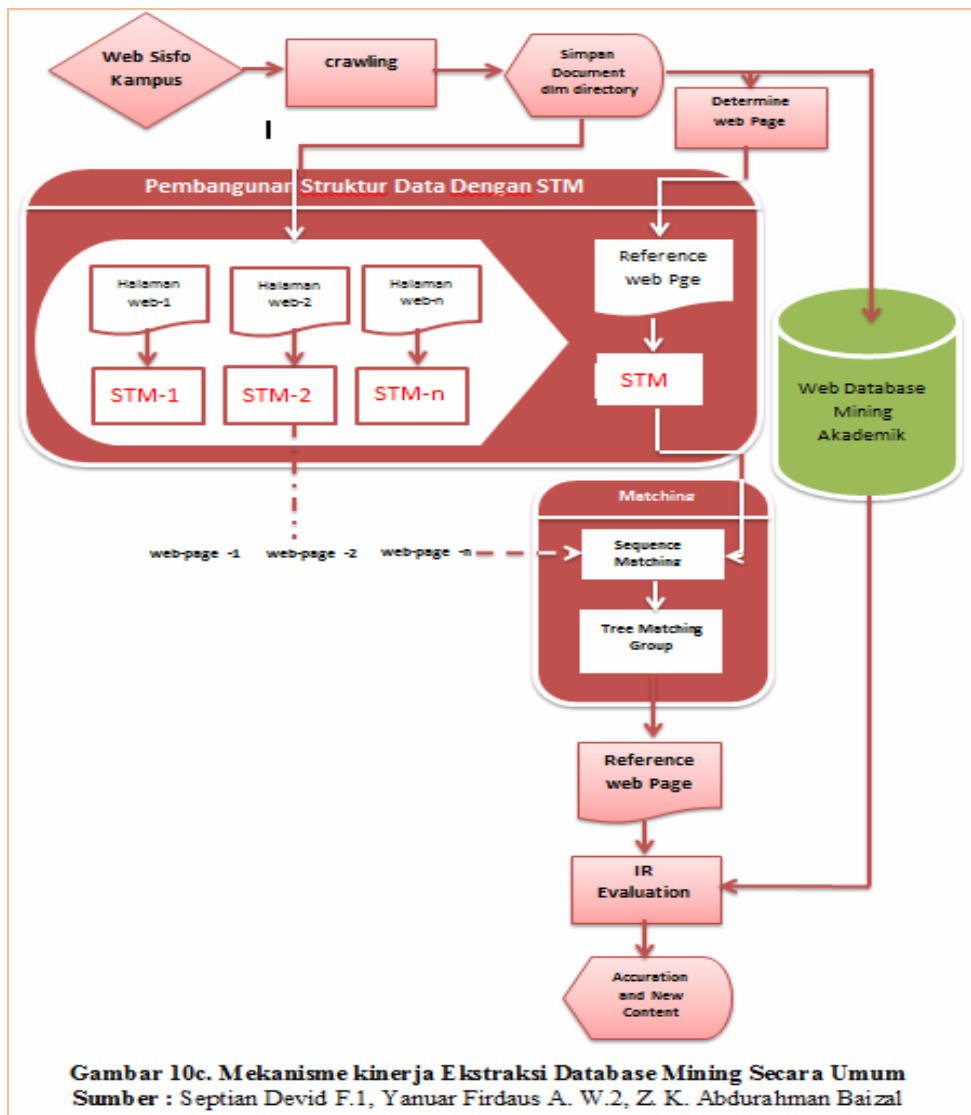
nya disimpan dalam bentuk file txt. Deskripsi sistem ditunjukkan pada sisfo-kampus dapat dijelaskan sbb :

SisfoKampus Universitas/Fakultas/Sekolah Tinggi/Akademik

SisfoKampus meliputi semua proses akademik mahasiswa maupun laporan keuangan mahasiswa secara menyeluruh. Mulai dari Penerimaan Mahasiswa Baru (PMB) sampai dengan alumni. Gambaran lingkup dalam SisfoKampus terlihat pada Gambar 11.

Semua proses *internal* tersebut terintegrasi dengan akunting perguruan tinggi, sehingga dapat dikeluarkan laporan, neraca keuangan, laporan laba rugi dan *cashflow*. Smart SisfoKampus juga mengakomodir semua proses dalam jadwal akademik, secara umum tergambar pada gambar 12. sebagai berikut :

SisfoKampus mampu menjawab dengan tepat mengapa SisfoKampus menjadi pilihan bagi perguruan tinggi yang ingin menerapkan Sistem Informasi Manajemen (SIM) Akademik terpadu sebagai solusi untuk meningkatkan efisiensi dan efektifitas dalam pelayanan kepada *stakeholder* (Calon mahasiswa, mahasiswa, dosen, manajemen fakultas /universitas, BAUK, karyawan, alumni, dan pengguna lulusan).

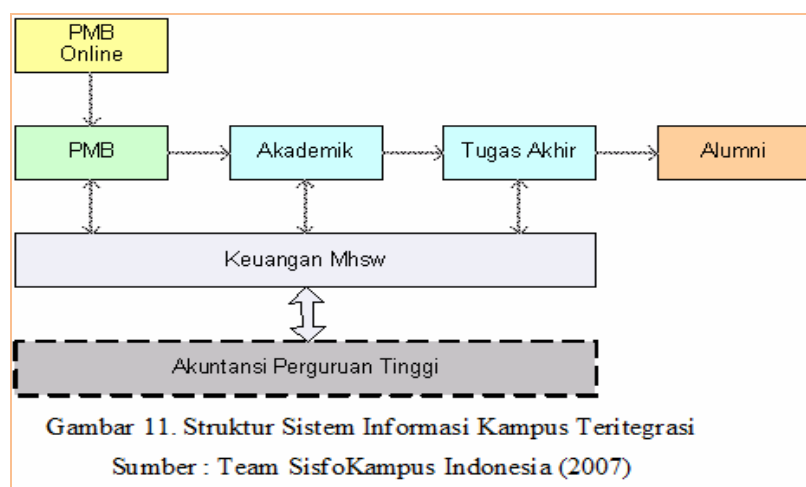


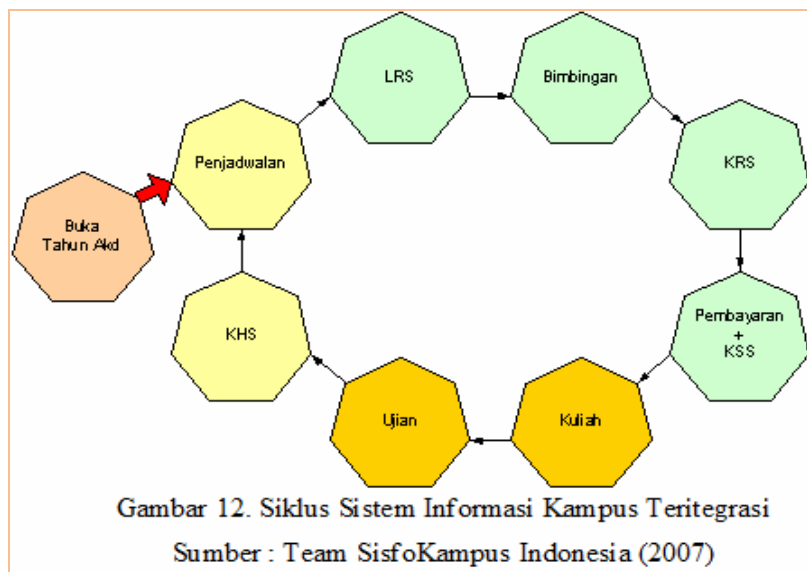
Beberapa keunggulan yang dimiliki oleh Smart SisfoKampus diantaranya adalah sebagai berikut:

1. Sudah terbukti dan teruji (*proven and tested*). Dapat diimplementasi dengan baik di Universitas, Akademi, Sekolah Tinggi, Institut dan Kursus sekalipun.
2. *Open Source*. Kepemilikan *source code* sepenuhnya di tangan *User*, artinya *user* dapat menggunakan dan memodifikasi dengan bebas sesuai azas GNU/GPL (referensi www.gnu.org). Dengan *open source*, *user* mengetahui secara pasti apa yang ada di dalam aplikasi tersebut tanpa harus membeli lisensi.

3. Sistem tumbuh dan berkembang, keunggulan ini sekaligus menjadi *core belief* kami bahwa sistem yang cerdas adalah sistem yang dapat mengikuti perkembangan dari institusi (organisasi), mulai satu fakultas/prodi/kampus sampai menjadi multi fakultas/prodi/kampus. Dengan keunggulan ini, aplikasi dapat dirancang untuk mendukung Lembaga Pelayanan Masyarakat (*business unit*) yang digunakan sebagai wadah pembelajaran lanjutan bagi mahasiswa. Seperti Fakultas Kedokteran yang memiliki program Klinik, atau Fakultas Hukum yang memiliki Lembaga Bantuan Hukum (LBH) kampus.

4. *Integrated.* Semua proses transaksi dari *front desk* (admisi-PMB) sampai *back office* (keuangan) dapat dilakukan secara terintegrasi tanpa dilakukan *double entry data*. Proses integrasi ini membantu perguruan tinggi yang sudah memiliki aplikasi bawaan seperti aplikasi akuntansi, aplikasi PM, aplikadi DIKTI. Selain dengan aplikasi, integrasi juga dapat dilakukan terhadap *tools*(scanner, finger print) dengan menggunakan fasilitas ekspor –impor.
5. *Transparan.* Metode Implementasi Smart SisfoKampus mendorong perguruan tinggi untuk mandiri sekaligus sebagai wadah pembelajaran sehingga dapat menekan *maintenance cost* di kemudian hari. Dengan *source code* yang sepenuhnya dimiliki oleh *user*, dijamin tidak akan ada *hidden cost* yang membebankan *user* seperti yang lazim terjadi dalam implementasi aplikasi.
6. Menjamin terintegrasinya program DIKTI. Smart SisfoKampus mampu mengeluarkan laporan wajib untuk DIKTI baik sekali komprehensif (semua prodi) atau per prodi. Fasilitas ini untuk menghindari terjadi *double entry*, dan laporan dapat diserahkan secara akurat dan tepat waktu.
7. *Features Khusus.* Berbagai feature tambahan tersedia di dalam aplikasi untuk mendukung penerapan di semua level user. Features yang tersedia saat ini diantaranya adalah penyebaran informasi dengan SMS Gateway, akses melalui PDA, auto debet, akses melalui anjungan mandiri, PMB Online, impor data histori akademik, dan sebagainya.
8. *Executive Information System.* Top Manajemen dapat dengan mudah mendapatkan informasi dan laporan tentang seluruh aktifitas (akademik dan keuangan) di dalam perguruan tinggi secara *real time* sehingga memudahkan untuk pengambilan keputusan.
9. *Supporting.* Belajar dari pengalaman dalam implementasi aplikasi bahwa salah satu hambatan implementasi adalah tidak tersedia *manual book* secara komprehensif. Untuk itu, di dalam aplikasi Smart SisfoKampus kami menyediakan *diagram Standar Operation Procedure (SOP)* dan *manual book* di setiap level user. Selain itu, tersedia juga fasilitas *Bugs tracking on line* untuk setiap error yang mungkin terjadi dan langsung dikirim kepada *technical support*.





HASIL DAN PEMBAHASAN

Skenario Pengujian

Secara garis besar pengujian yang dilakukan dibagi dalam dua tahap. Pada tahap pertama (skenario 1) pengujian dilakukan untuk mengetahui pengaruh jumlah dataset berisikan tabel-tabel akademik yang ada pada database akademik terhadap performansi sistem yang sudah dibangun. Pada tahap kedua (skenario 2) pengujian dilakukan untuk mengetahui performansi modul aplikasi dari sisfo-kampus yang sering digunakan dan di akses oleh mahasiswa, dosen, dan pihak akademik. (skenario 3) pengujian dilakukan untuk mengetahui perbandingan performansi antara dataset mentah dan dataset yang telah di validasi sebelumnya dengan metoda tersebut.

Adapun rincian skenario pengujian yang dilakukan terhadap sistem ini adalah sebagai berikut:

- Pengujian terhadap pengaruh banyaknya jumlah dataset yang digunakan terhadap performansi sistem.
- Pengujian terhadap performansi sistem Web News Content Extraction STM

yang telah dibangun pada dataset halaman web sisfokampus (modul-modul yang sering diakses oleh sivitas akademika) yang masih belum memenuhi standar.

- Pengujian dilakukan untuk mengetahui perbandingan performansi antara dataset mentah dan dataset yang telah di validasi sebelumnya dengan metoda tersebut.

Pengujian Terhadap Tabel

Pengujian ini dilakukan untuk mengetahui keterkaitan antara jumlah *dataset* yang digunakan dengan performansi yang dihasilkan oleh sisfokampus. Berdasarkan Tabel 1, 2, 3, dan 4 serta Gambar 13, 14, 15 dan 16 pada tabel dan gambar tersebut dapat diketahui bahwa pada database akademik dan modul sisfokampus, jumlah dari *dataset* yang digunakan tidak terlalu berpengaruh terhadap performansi sistem, hal tersebut terlihat dari grafik yang cenderung stabil walaupun jumlah *dataset* terus bertambah. Selisih nilai minimum dan maksimum F-Measure dari kedelapan situs tersebut juga sangat kecil, yaitu:

Tabel 1.
 Hasil Pengukuran Performans Tabel dengan *Dataset* Mentah

No.	Nama Tabel	Jumlah			
		Records (Byte)	Precision	Recall	F-Measure
1	Mahasiswa	20,265	0.6898	0.0751	0.1192
2	Matakuliah	1,329	0.9052	0.9809	0.9396
3	Jadwal	1,265	0.8764	0.9886	0.9288
4	Kartu Rencana Studi	743,014	0.3264	0.9343	0.476
5	Dosen	556	0.8377	0.9734	0.8947
6	Ruang	67	0.827	0.9838	0.8887
7	Jurusan	18	0.7052	0.7077	0.6492
8	Kartu Hasil Studi	22,139	0.8862	0.9559	0.9149
9	Transaksi	2,259	0.5603	0.2004	0.2255
10	Program	3	0.7198	0.955	0.8178
RATA-RATA		79,092	0,7295	0,7755	0,7295

Error! Not a valid link. Tabel 2.

Hasil Pengukuran Performans Tabel dengan *Dataset* Valid

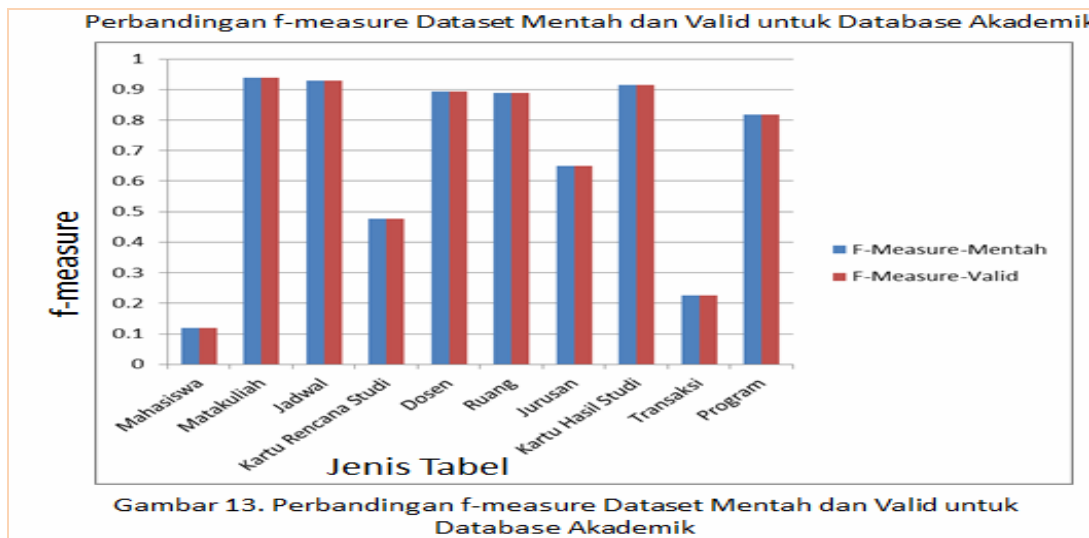
No.	Nama Tabel	Jumlah			
		Records (Byte)	Precision	Recall	F-Measure
1	Mahasiswa	20,265	0.6898	0.0751	0.1192
2	Matakuliah	1,329	0.9052	0.9809	0.9396
3	Jadwal	1,265	0.8764	0.9886	0.9288
4	Kartu Rencana Studi	743,014	0.3264	0.9343	0.476
5	Dosen	556	0.8377	0.9734	0.8947
6	Ruang	67	0.827	0.9838	0.8887
7	Jurusan	18	0.7052	0.7077	0.6492
8	Kartu Hasil Studi	22,139	0.8862	0.9559	0.9149
9	Transaksi	2,259	0.5603	0.2004	0.2255
10	Program	3	0.7198	0.955	0.8178
RATA-RATA		79,092	0,7295	0,7755	0,7295

Dari Tabel 1 dan 2 serta gambar 13 dan 16. dapat diketahui bahwa, hasil yang diperoleh dari dataset valid cenderung sama dengan dataset mentah, tabel-tabel yang mendapatkan nilai tinggi (memiliki kecocokan dengan metode STM) pada dataset mentah, juga mendapatkan nilai tinggi pada dataset valid, begitu pula dengan tabel-tabel yang mendapatkan nilai F-Measure rendah pada dataset mentah juga mendapatkan nilai rendah pada dataset valid, Nilai F-Measure ini bisa sama, karena dalam preprocessing, term-term yang dihilangkan mungkin bukan informasi yang penting dari dataset tersebut.

Hal ini sesuai dengan asumsi penulis bahwa jumlah *dataset* tidak akan mempengaruhi performansi sistem karena masing-masing *dataset* yang diambil adalah *dataset* dari satu situs saja dan *dataset* (halaman *web*) tersebut memiliki *layout* yang sama karena satu sama lain berhubungan secara langsung dari link-link yang mereka miliki, sehingga semakin dekat hubungan halaman *web* yang yang dijadikan *dataset* maka semakin kecil pula pengaruh jumlah *dataset* yang digunakan terhadap performansi sisfokampus atau dengan kata lain berapapun jumlah *dataset* yang digunakan, performansi sisfokampus akan cenderung stabil.

Setelah diperiksa, tidak stabilnya performansi tersebut ternyata terjadi karena walaupun halaman *website* yang dijadikan *dataset* saling berhubungan namun terdapat perbedaan yang cukup besar pada layoutnya yaitu pada bagian *content part*. Sehingga performansi yang ditunjukkan cenderung tidak stabil.

Dari sini dapat diambil kesimpulan bahwa dengan *dataset* yang cocok dan sesuai dengan metode yang digunakan (seperti pada table Mahasiswa, Matakuliah, Jadwal, Kartu Rencana Studi, Dosen, Ruang, Jurusan, Kartu Hasil Studi, Transaksi dan Program), maka penambahan jumlah *dataset* tidak akan mempengaruhi performansi sisfokampus.



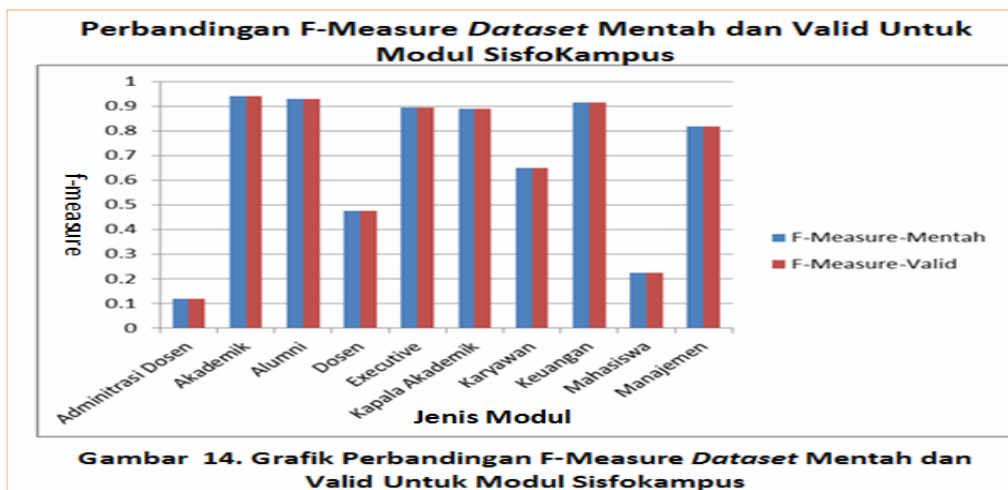
Pengujian Terhadap Modul

Tabel 3.
 Hasil Pengukuran Performans Modul Aplikasi dengan *Dataset* Mentah

ID	GroupModul	Keterangan Modul	Menu	web	cs	Ukuran Modul (Byte)	Precision	Recall	F-Measure
1	Administrasi Dosen	Manajemen Dosen	Y	Y	N	25	0.8377	0.476	0.476
2	Akademik	Manajemen Akademik	Y	Y	N	15	0.3264	0.8947	0.8947
3	Alumni	Manajemen Alumni	Y	Y	N	20	0.8377	0.8887	0.476
4	Dosen	Administrasi Dosen	Y	Y	N	12	0.3264	0.6492	0.8947
5	Executive	Manajemen Info. Executive	Y	Y	N	7	0.7052	0.9149	0.8887
6	Kapala Akademik	Kapala Biro Akademik	Y	Y	N	8	0.8862	0.9149	0.6492
7	Karyawan	Administrasi Karyawan	Y	Y	N	9	0.3264	0.8887	0.9149
8	Keuangan	Manajemen Keu.Fakultas	Y	Y	N	23	0.8377	0.6492	0.9149
9	Mahasiswa	Administrasi Mahasiswa	Y	Y	N	25	0.3264	0.9149	0.476
10	Manajemen	Manajemen Fakultas	Y	Y	N	3	0.8377	0.8887	0.8947
11	Master	Master Seluruh Tabel	Y	Y	N	35	0.8377	0.6492	0.8887
12	PDE Absen	Manajemen Absen	Y	Y	N	13	0.3264	0.476	0.6492
13	PDE Nilai	Manajemen Nilai	Y	Y	N	45	0.8377	0.8947	0.9149
14	Ka. Program Studi	Administrasi Prog. Studi	Y	Y	N	25	0.3264	0.8887	0.6492
15	Administrator Sist.	Administrasi Sistem	Y	Y	N	10	0.8862	0.6492	0.9149
16	SPMB	Manajemen SPMB	Y	Y	N	15	0.8377	0.9149	0.8947
Rata-Rata						18.125	0.64374	0.785	0.774463

Tabel 4.
 Hasil Pengukuran Performans Modul Aplikasi dengan *Dataset Valid*

ID	GroupModul	Keterangan Modul	Menu	web	cs	Ukuran Modul (Byte)	Precision	Recall	F-Measure
1	Adminitrasi Dosen	Manajemen Dosen	Y	Y	N	25	0.8377	0.476	0.476
2	Akademik	Manajemen Akademik	Y	Y	N	15	0.3264	0.8947	0.8947
3	Alumni	Manajemen Alumni	Y	Y	N	20	0.8377	0.8887	0.476
4	Dosen	Administrasi Dosen	Y	Y	N	12	0.3264	0.6492	0.8947
5	Executive	Manajemen Info. Executive	Y	Y	N	7	0.7052	0.9149	0.8887
6	Kapala Akademik	Kapala Biro Akademik	Y	Y	N	8	0.8862	0.9149	0.6492
7	Karyawan	Administrasi Karyawan	Y	Y	N	9	0.3264	0.8887	0.9149
8	Keuangan	Manajemen Keu. Fakultas	Y	Y	N	23	0.8377	0.6492	0.9149
9	Mahasiswa	Administrasi Mahasiswa	Y	Y	N	25	0.3264	0.9149	0.476
10	Manajemen	Manajemen Fakultas	Y	Y	N	3	0.8377	0.8887	0.8947
11	Master	Master Seluruh Tabel	Y	Y	N	35	0.8377	0.6492	0.8887
12	PDE Absen	Manajemen Absen	Y	Y	N	13	0.3264	0.476	0.6492
13	PDE Nilai	Manajemen Nilai	Y	Y	N	45	0.8377	0.8947	0.9149
14	Ka. Program Studi	Administrasi Prog. Studi	Y	Y	N	25	0.3264	0.8887	0.6492
15	Administrator Sist.	Administrasi Sistem	Y	Y	N	10	0.8862	0.6492	0.9149
16	SPMB	Manajemen SPMB	Y	Y	N	15	0.8377	0.9149	0.8947
Rata-Rata						18.125	0.64374	0.785	0.774463



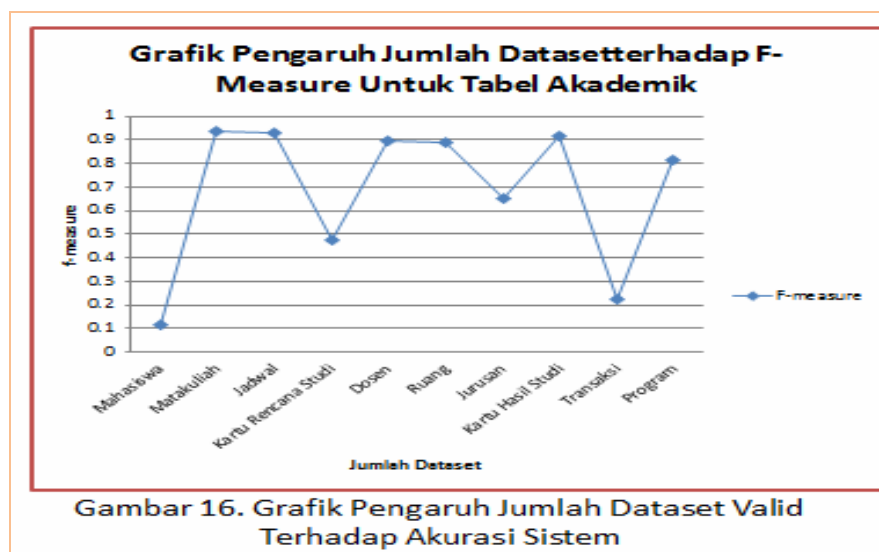
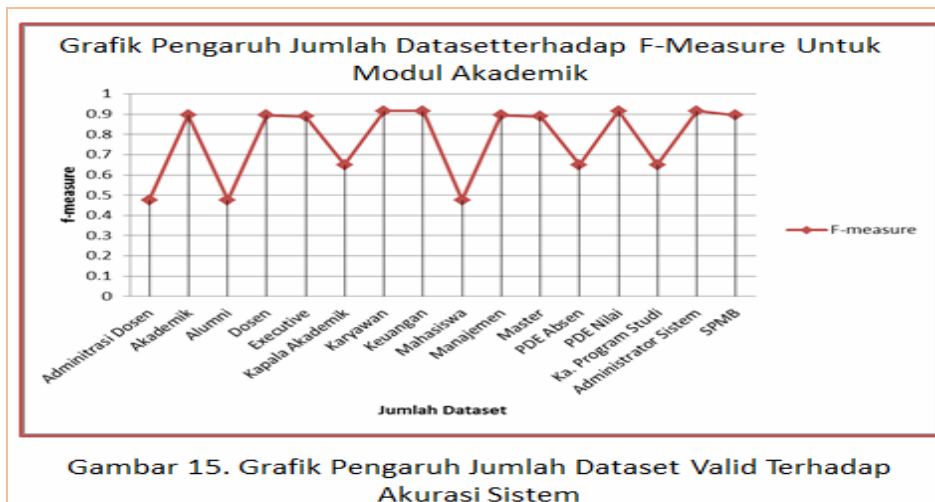
Dari Tabel 3 dan 4 serta gambar 14 dan 15. dapat diketahui bahwa, hasil yang diperoleh dari dataset valid cenderung sama dengan dataset mentah, modul-modul yang mendapatkan nilai tinggi (memiliki kecocokan dengan metode STM) pada dataset mentah, juga mendapatkan nilai tinggi pada dataset valid, begitu pula dengan table-tabel yang mendapatkan nilai F-Measure rendah pada dataset mentah juga mendapatkan nilai rendah pada dataset valid, Nilai

F-Measure ini bisa sama, karena dalam preprocessing, term-term yang dihilangkan mungkin bukan informasi yang penting dari dataset tersebut.

Dari hasil pengujian diatas dapat di simpulkan bahwa web database akademik mining merupakan tempat penyimpanan informasi yang terbesar. Ekstraksi informasi dari web, adalah penting untuk menyediakan alat-alat untuk penemuan yang efisien dan efektif sumber daya dan pengetahuan di web terutama dalam web

database mining akademik, perbedaan informasi dalam web menumbuhkan web mining. Ekstraksi informasi dari web telah dilakukan melalui berbagai penelitian yang menghasilkan algoritma-algoritma (wrapper) yang mampu mengekstrak informasi, yang terstruktur secara sintaksis dan secara otomatis. Informasi yang ditampilkan dalam sebuah halaman web merupakan informasi yang tidak terstruktur atau yang semi terstruktur. Wrapper yang nantinya akan mengekstrak informasi yang tidak terstruktur atau se-

mi terstruktur tersebut. Dengan metoda STM ini dapat mempercepat dalam menemukan dan mengakses informasi di database akademik dan non-akademik karena metoda ini dapat mengurangi waktu komputasi. Dimana tidak perlu dilakukan penyimpanan ulang pola data records actual dalam data ruang khusus yang sudah ditemukan agar dapat digunakan dalam ekstraksi data records dari halaman web dengan encoding template yang sama sehingga tanpa mencari pola lagi.



SIMPULAN DAN SARAN

Berdasarkan analisis dan pengujian terhadap ekstraksi web database mining akademik yang telah dilakukan maka dapat diambil beberapa poin kesimpulan sebagai berikut:

1. Metode STM diterapkan dalam proses pencarian dan pengaksesan terhadap pola otomatis berbasis pencocokan pohon baik menggunakan tabel maupun modul yang tersedia pada Web Database Mining Akademik terutama Sisfokampus.
2. Selanjutnya tabel-tabel dan setiap modul aplikasi yang ada pada sisfokampus dilakukan sistem ekstraksi dengan melakukan pengujian terhadap parameter recall, precision dan f-measure.
3. Hasil yang diperoleh dari pengujian ketiga parameter tersebut didapatkan dataset valid cenderung sama dengan dataset mentah, modul-modul yang mendapatkan nilai tinggi (memiliki kecocokan dengan metode STM) pada dataset mentah, juga mendapatkan nilai tinggi pada dataset valid, begitu pula dengan table-tabel yang mendapatkan nilai F-Measure rendah pada dataset mentah juga mendapatkan nilai rendah pada dataset valid, Nilai F-Measure ini bisa sama, karena dalam preprocessing, term-term yang dihilangkan mungkin bukan informasi yang penting dari dataset tersebut
4. Penggunaan jumlah dataset yang berbeda memiliki pengaruh yang bervariasi terhadap performansi parameter precision, recall, dan F-measure yang ditunjukkan, hal ini tergantung dari tingkat similarity page reference dengan page extracted, Semakin similar maka performansi yang ditunjukkan semakin stabil.
5. Penggunaan dataset mentah maupun valid juga memiliki pengaruh yang bervariasi terhadap performansi precision, recall dan f-measure, tergantun-

ung dari proses validasi yang dilakukan oleh validator.

6. Penggunaan batasan term terhadap suatu group (STM yang berada dalam subtree yang sama/sibling) dapat digunakan sebagai solusi yang sangat efektif untuk meningkatkan performansi sistem aplikasi sisfokampus.

Beberapa saran untuk penelitian lebih lanjut, yaitu:

1. Perlu dilakukan penyimpanan untuk pola data records aktual dalam web data mining khusus yang sudah ditemukan agar dapat digunakan dalam ekstraksi data records dari halaman web dengan encoding template yang sama tanpa harus mencari pola lagi.
2. Perlu diintegrasikan tools (perkakas) sistem ekstraksi informasi berbasis pengolahan bahasa alami untuk meningkatkan nilai precision ekstraksi data records dan penajaran data items, juga agar proses pelabelan kolom dapat dilakukan secara otomatis.
3. Penentuan relevan dan tidak relevan dokumen perlu penekanan dalam penelitian berikutnya, diajurkan menggunakan metoda yang lebih baik untuk pengukuran performansi kinerja aplikasi berbasis web terutama sisfokampus.
4. Sistem ini akan lebih baik jika diimplementasikan secara online (misal sebagai plugin browser).
5. Penentuan page reference akan lebih baik jika secara otomatis dapat mendeteksi tingkat similarity page reference dan page extracted.

DAFTAR PUSTAKA

- Abhijit, R. and Kulkarni2, R.V. 2011 "Data Mining Techniques: A Source For Consumer Behavior Analysis" *International Journal of Database Management Systems (IJDMs)* 3: 3.
- Bing, L., Grossman, R., and Yanhong, Z. 2003 *Mining Data Records in Web*

- Pages* University of Illinois at Chicago
<http://www.cs.uic.edu/~liub/publications/kdd2003-dataRecord.pdf>
didownload pada tanggal 25 Juni 2007.
- Gengxin, M., Junichi, T., Wang-Pin, H., Arsany, S., Louise, E. M. 2007 *Extracting Data Records from the Web Using Tag Path Clustering* Department of Computer Engineering and Information Science.
- Ghosh, J., Lee, Dik. 2009 *Evaluation Lecture 9 Lecture Notes of Information Retrieval* Computer Science and Electrical Engineering University of Maryland
<http://www.cs.umbc.edu/~ian/irF02/lectures/09Evaluation.pdf> didownload pada 2 Juli 2012
- Nitin, J., and Bing, L. 2009 *A Generalized Tree Matching Algorithm Considering Nested Lists for Web Data Extraction* Department of Computer Science, University of Illinois at Chicago
- Pramintya, H. 2008 *Ekstraksi Informasi Pada Halaman Web dengan Memanfaatkan Mining Data Record* Institut Teknologi Telkom Bandung.
- Septian, D. F., Yanuar, F. A. W., Abdurahman, Z. K. B. 2011 *Ekstraksi Informasi Utama Halaman Web Berita Menggunakan Metode Hybrid* didownload pada 26 Juni 2012
- Wahju, A.W. 2001 “RJB Implementasi Agen Untuk Ekstraksi Web Dta” *Jurnal Teknologi Industri* V (3) 163 – 170.