

LAPORAN PENELITIAN

PEMETAAN AKSES HALAMAN SITUS WEB BERBASIS *LOG-ACCESS* (STUDI KASUS WEB USAGE MINING DI SITUS perpus.akakom.ac.id)



Oleh:

Dison Librado, S.E., M.Kom.

961087/0526096701

Wagito, S.T., M.T.



961080/0522126901

Dilaksanakan Atas Bantuan Biaya Penelitian dari Puslit dan PPM
Semester Genap Tahun Akademik 2017/2018

Sekolah Tinggi Manajemen Informatika dan Komputer AKAKOM
Yogyakarta
2018

HALAMAN PENGESAHAN

Judul Penelitian : Pemetaan Akses Halaman Situs Web Berbasis *Log-access*
(Studi Kasus *Web Usage Mining* di Situs *perpus.akakom.ac.id*)

- 1 Ketua Tim Pengusul
 - a. Nama : Dison Librado
 - b. NPP / NIDN : 961087 / 0526096701
 - c. Jabatan/golongan : Lektor / III C
 - d. Jurusan/Fakultas : Teknik Informatika
 - e. Bidang keahlian : Sistem Informasi
- 2 Anggota Tim Pengusul
 - a. Jumlah anggota : 1
 - b. Nama anggota I : Wagito, S.T.,M.T.
NPP / NIDN : 961080 / 0522126901
Bidang keahlian : Jaringan Komputer
- 3 Jumlah mahasiswa yang terlibat : -
- 4 Jangka Waktu Pelaksanaan : 1 Semester
- 5 Total Biaya yang diusulkan
 - a. STMIK AKAKOM : 
 - b. Sumber Dana Lain : 

Yogyakarta, 8 Februari 2018

Mengetahui / Menyetujui
Ketua Program Studi

Ketua Peneliti,

Ir. M. Guntara, M.T.
891019 / 0509066101

Dison Librado, S.E.,M.Kom
961087 / 0526096701

Menyetujui
Pja. Kepala PUSLIT dan PPM

Edy Prayitno, S.Kom.M.Eng.
151185

DAFTAR ISI

Halaman Judul.....	i
Halaman Pengesahan.....	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
ABSTRAK	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
BAB II TINJAUAN PUSTAKA	5
BAB III LANDASAN TEORI.....	8
3.1 <i>Web Mining</i>	8
3.1.1 <i>Web Content Mining</i>	9
3.1.2 <i>Web Structure Mining</i>	10
3.1.3 <i>Web Usage Mining</i>	10
3.2 Server Web Nginx	13
3.2.1 Fitur Dasar Nginx.....	14
3.2.2 Log Nginx	15
3.3 Situs perpus.akakom.ac.id	18
BAB IV METODE PENELITIAN.....	21
4.1 Tahapan Penelitian.....	21
4.2 Lokasi Penelitian.....	21
4.3 Model yang Digunakan	21
4.3.1 Tahap Preprocessing	22
4.3.2 Tahap Pattern Discovery	23

4.3.3 Tahap <i>Pattern Analysis</i>	24
4.4 Rancangan Penelitian	25
4.5 Teknik Pengumpulan Data dan Analisis Data	25
BAB V HASIL PENELITIAN DAN PEMBAHASAN.....	26
5.1 Hasil Penelitian	26
5.1.1 Konfigurasi server web Nginx	26
5.1.2 Pengumpulan data log akses.....	27
5.1.3 Mengolah data log akses	27
5.1.4 Pengolahan Data MySQL	31
5.2 Pembahasan	35
BAB VI PENUTUP	43
6.1 Kesimpulan	43
6.2 Saran	44
DAFTAR PUSTAKA	45

DAFTAR GAMBAR

Gambar 3.1 Sintaksis Log Format	16
Gambar 3.2 Log Format	17
Gambar 3.3 Sintaksis Access Log.....	17
Gambar 5.1 Format Log.....	26
Gambar 5.2 Log Acces Nginx	28
Gambar 5.3 Sekrip Parsing Log.....	30
Gambar 5.4 Perintah SQL Membuat Tabel Log.....	32
Gambar 5.5 Halaman yang Paling Sering Dikunjungi.....	33
Gambar 5.6 Asal Alamat IP yang Paling Sering mengunjungi situs perpus.....	34
Gambar 5.7 Mencari Sistem Operasi atau Browser yang Dipakai Pengguna.....	34

DAFTAR TABEL

Tabel 5.1 Struktur Tabel Log	29
Tabel 5.2 Daftar URL peringkat 30 terbanyak yang diakses	35
Tabel 5.3 Daftar Menu/Navigasi/Link yang dikunjungi	37
Tabel 5.4 Daftar Halaman yang Diakses	39
Tabel 5.5. Daftar alamat IP peringkat 30 terbanyak yang sering berkunjung.....	40
Tabel 5.6 Daftar Area dan Jenis Akses	41
Tabel 5.7 Daftar Sistem Operasi atau Browser yang digunakan	41

ABSTRAK

Perpustakaan merupakan unit pelayanan teknis yang menyediakan informasi tentang ilmu pengetahuan. Salah satu jenis perpustakaan adalah perpustakaan perguruan tinggi. Perpustakaan STMIK AKAKOM sebagai unit yang berada di kampus baru menerapkan teknologi informasi yang berbasis internet secara menyeluruh mulai tahun 2016 dengan memanfaatkan aplikasi yang dikembangkan oleh bagian Sistem Informasi sebagai *front-end* yang menghubungkan dengan aplikasi perpustakaan lainnya. Berbagai menu yang disediakan dalam situs resmi perpustakaan ini adalah Home, Kontak, Tautan, Layanan, Profil, Katalog Online, dan Digital Library. Seberapa efektifkah penyediaan teknologi informasi ini dapat melayani penggunaannya? Efektivitas ini akan dapat diukur dengan banyaknya pengguna yang mengakses halaman web ini

Penelitian terdiri dari beberapa tahap yaitu studi pustaka dari sumber referensi yang relevan dengan topik yang akan diteliti, mengonfigurasi server web Nginx, mengumpulkan data *log access* selama waktu tertentu, kemudian mengolah data *log access* tersebut dan membuat pembahasan hasil pengolahan data, serta membuat kesimpulan. Penelitian menyajikan hasil pengumpulan dan pengamatan data dari kunjungan para pengguna web dengan tujuan menyajikan informasi dari jumlah pengunjung berdasarkan kunjungan dan keragaman pengunjung pada situs web perpustakaan STMIK AKAKOM untuk melihat pola kunjungan ke web dan mengidentifikasi halaman apa yang sering dikunjungi oleh pengunjung situs dengan memanfaatkan *log* dari penggunaan situs, sehingga dapat diketahui pola dan minat pengguna di situs perpus.akakom.ac.id.

Hasil penelitian menunjukkan bahwa halaman situs yang paling sering dikunjungi berdasarkan urutan jumlah aksesnya adalah 'Home', 'Berita', 'Digital Library', 'Koleksi', 'Tautan', 'Layanan', 'Katalog Online', dan 'Kontak'. Sistem Operasi atau Browser yang paling banyak digunakan untuk mengakses halaman situs adalah Windows, Android, Linux, dan iPhone. Berdasarkan pengamatan IP, paling banyak IP berasal dari luar Akakom

Kata kunci: *log-access*, situs web, *web usage mining*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perpustakaan adalah unit yang berfungsi sebagai penyedia informasi bagi yang membutuhkan dalam bentuk buku, majalah, maupun dalam bentuk digital. Perpustakaan di perguruan tinggi berfungsi sebagai unit yang memberikan pelayanan untuk mendukung pelaksanaan Tri Dharma Perguruan Tinggi. Perpustakaan yang terorganisasi secara baik dan sistematis secara langsung atau pun tidak langsung dapat memberikan kemudahan bagi proses belajar mengajar di perguruan tinggi tempat perpustakaan tersebut berada. Perkembangan teknologi informasi yang demikian cepat dapat mempengaruhi berbagai bidang kehidupan dan pekerjaan. Hal ini menyebabkan berubahnya sistem pada instansi atau perusahaan yang juga mengubah cara kerja mereka. Teknologi informasi banyak digunakan untuk pengelolaan pekerjaan karena efektivitas dan efisiensinya yang sudah terbukti mampu mempercepat kerja. Kecepatan kerja pada akhirnya akan meningkatkan keuntungan baik finansial maupun non-finansial. Penggunaan teknologi informasi dalam kehidupan sehari-hari dapat mempermudah pertukaran informasi dan data antarwilayah sehingga penyebaran pengetahuan menjadi begitu cepat. Kemajuan paling terlihat adalah pada penggunaan teknologi informasi dalam proses pengolahan data menjadi informasi yang menjadi lebih cepat dan otomatis. Menurut Sulistyo-Basuki dalam Angelia (2015) Teknologi Informasi merupakan teknologi yang berbasis komputer yang digunakan untuk bekerja secara efisien dengan

memanfaatkan media internet. Perkembangan teknologi sudah memberikan banyak dampak keseluruh lembaga atau instansi, salah satunya adalah perpustakaan. Perpustakaan merupakan unit pelayanan teknis yang menyediakan informasi tentang ilmu pengetahuan. Salah satu jenis perpustakaan adalah perpustakaan perguruan tinggi. Buku Pedoman Perpustakaan Perguruan Tinggi (1994:3) menyatakan bahwa Perpustakaan Perguruan Tinggi merupakan perpustakaan yang didirikan di lingkungan universitas, fakultas maupun jurusan. Dalam perkembangan teknologi informasi di dalam perpustakaan perguruan tinggi, perpustakaan melakukan kegiatannya menggunakan sistem terkomputerisasi guna untuk memudahkan pustakawan di dalam bekerja untuk mengolah, menyimpan di dalam suatu database yang di miliki oleh perpustakaan itu sendiri. Pemanfaatan Teknologi Informasi dalam perpustakaan memiliki peran untuk mengotomatiskan kinerja dari layanan perpustakaan. Dengan penerapan teknologi informasi, pustakawan dan pengguna dapat semakin cepat dalam mengakses berbagai layanan perpustakaan seperti pencarian katalog, pendaftaran anggota, peminjaman dan pengembalian, dan rekapitulasi serta pembuatan laporan.

Perpustakaan STMIK AKAKOM sebagai unit yang berada di kampus baru menerapkan teknologi informasi yang berbasis internet secara menyeluruh mulai tahun 2016 dengan memanfaatkan aplikasi yang dikembangkan oleh bagian Sistem Informasi sebagai *front-end* yang menghubungkan dengan aplikasi perpustakaan lainnya. Berbagai menu yang disediakan dalam situs resmi perpustakaan ini adalah Home, Kontak, Tautan, Layanan, Profil, Katalog Online, dan Digital Library. Di

dalam setiap menu ini masih terdapat beberapa sub menu yang berhubungan dengan menu utamanya. Selain dari menu-menu utama yang berada pada baris navigasi ini juga terdapat menu lain yang terletak di panel bagian kanan yaitu Koleksi Terbaru dan di bagian bawah terdapat panel Berita, Pengumuman, dan Tautan Terkait. Seberapa efektifkah penyediaan teknologi informasi ini dapat melayani penggunanya? Efektivitas ini akan dapat diukur dengan banyaknya pengguna yang mengakses halaman web ini.

1.2 Rumusan Masalah

Dari uraian dalam Latar Belakang selanjutnya dapat dirumuskan masalah dalam penelitian ini, yaitu berapa jumlah pengunjung dalam kurun waktu kira-kira dua bulan semester ganjil tahun 2017 dan halaman apa yang sering dikunjungi oleh pengunjung situs. Studi ini didasarkan log dari kunjungan ke situs web Perpustakaan STMIK AKAKOM. Dari situ akan dapat dilihat bagaimana pola dan minat pengguna di situs perpus.akakom.ac.id tersebut.

1.3 Batasan Masalah

Jumlah pengakses situs perpus.akakom.ac.id sangat banyak terhitung sejak tahun 2016. Untuk mengurangi jumlah pengamatan data yang demikian banyak maka penelitian ini membatasi cakupan datanya pada semester genap tahun akademik 2016/2017 dengan rincian pengamatan yang tidak sampai pada pengidentifikasian lokasi dari mana pengguna mengakses.

1.4 Tujuan

Penelitian ini akan menyajikan hasil pengumpulan data dan pengamatannya terhadap pengunjung web dengan tujuan

- 1) menyajikan informasi dari jumlah pengunjung berdasarkan keragaman tujuan kunjungan pada situs web Perpustakaan STMIK AKAKOM
- 2) untuk melihat pola kunjungan ke web dan mengidentifikasi halaman apa yang sering dikunjungi oleh pengunjung situs dengan memanfaatkan *log* dari penggunaan situs sehingga dapat diketahui pola dan minat pengguna di situs perpus.akakom.ac.id.

1.5 Manfaat

Dengan melakukan studi dan pengamatan pada data statistik kunjungan web ini diharapkan dapat digunakan sebagai bahan evaluasi terhadap pengembangan situs web Perpustakaan STMIK AKAKOM dan peningkatan pelayanan kepada para pengunjung.

BAB II

TINJAUAN PUSTAKA

Dalam penelitiannya, Angelia (2015) Penelitian ini akan menganalisis tentang Pemanfaatan Senayan Library Information Management System (SLIMS) Di Perpustakaan Pusat Universitas Warmadewa dalam mengelola management yang dilaksanakan oleh Perpustakaan Pusat Universitas Warmadewa. Berdasarkan penelitiannya, sebelum mengaplikasikan SLIMS Perpustakaan Pusat Universitas Warmadewa menggunakan aplikasi e-lib untuk membantu kegiatan yang dilakukan oleh perpustakaan. E-lib merupakan sebuah sistem informasi yang terdiri dari perangkat keras (*hardware*) dan perangkat lunak (*software*), pengelolaan, pelayanan serta penyediaan (akses) informasinya dilakukan dengan menggunakan perangkat elektronik yang berupa komputer. Seiring berjalan waktu Perpustakaan Pusat Universitas Warmadewa mengubah sistemnya menjadi Senayan Library Information Management System (SLIMS). Disimpulkan bahwa Perpustakaan Pusat Universitas Warmadewa sudah memiliki kesiapan yang cukup matang untuk mengaplikasikan SLIMS dan penerapan SLIMS di Perpustakaan Pusat Universitas Warmadewa mampu memberikan manfaat untuk pengelola perpustakaan.

Penelitian tentang web usage mining sudah dilakukan dengan konsentrasi dan objek yang berbeda-beda. Suratno (2011) dalam tulisannya menyampaikan materi tentang penggalian data di situs www.faperta.unja.ac.id dengan cara mengembangkan sistem dengan membuat program metode *browser sniffing* (metode tanpa men-

submitform) yang digunakan untuk menangkap alamat log yang dihasilkan oleh user secara otomatis. Pengembangan program tersebut menggunakan php dan mysql yang disisipkan di file dalam situs tersebut. Data kunjungan ke situs tersebut yang tersimpan dalam database mysql, selanjutnya diproses secara manual menggunakan program bantu Microsoft Excel dan Matlab 2008a. Dari log file yang telah digali dapat diketahui jumlah pengunjung situs yang ditampilkan melalui web traffic serta pattern discovery-nya untuk digunakan sebagai rekomendasi bagi perancang web untuk mengatur penempatan item atau modul yang ada di situs.

Berkaitan dengan perkembangan internet yang telah memicu peningkatan aktivitas belanja online, Hermanto () membuat penelitian yang menerapkan *web usage mining* untuk membantu perusahaan dalam meningkatkan penjualan. Metode *web usage mining* yang akan diimplementasikan ke dalam sistem *ecommerce* itu dapat memberikan satu solusi baru untuk menganalisis transaksi yang terjadi di toko online maupun pangsa pasar suatu produk berdasarkan pembelian konsumen, dan kunjungan ke web. Simpulan yang dibuat dari *web usage mining* itu menjadi data yang dapat digali untuk menunjang analisis penjualan agar dapat menjadi alat bantu bagi para pemilik perusahaan sebagai fitur analisis perkembangan alur distribusi perdagangan di perusahaannya.

Dalam penelitian yang lain, Indrayani (2011) melakukan penelitian tentang pengelolaan sistem informasi akademik perguruan tinggi berbasis teknologi informasi dan komunikasi (TIK). Efektivitas aplikasi TIK dalam proses manajemen

kelembagaan sering terhambat oleh banyak faktor non teknis yang tidak dipersiapkan lembaga. Mulai dari penyiapan orang, budaya, mekanisme organisasi, bahkan teknis pemeliharannya. Tak selamanya SIA yang berbasis TIK bisa meningkatkan kinerja pengelolaan administrasi akademik, manakala lembaga hanya menganggap bahwa implementasi TIK untuk SIA hanya sekedar menyiapkan perangkat keras TIK. Selanjutnya dengan memperhatikan hal-hal tersebut diatas, penelitian ini mencoba untuk memperoleh data empirik yang cukup lengkap dan dapat dipercaya untuk menggambarkan tentang keadaan faktor-faktor yang terkait dengan pengelolaan Sistem Informasi Akademik berbasis TIK terhadap kinerja perguruan tinggi dan dampaknya terhadap prestasi akademik mahasiswa. Data yang telah diperoleh juga dapat digunakan untuk mengembangkan alternatif model Sistem Informasi Manajemen Akademik yang dapat memberi dukungan pada setiap proses pelayanan akademik maupun pengambilan keputusan baik di lingkungan internal maupun yang terkait dengan *stakeholders*.

BAB III

LANDASAN TEORI

3.1 *Web Mining*

Web merupakan kumpulan data dan informasi yang sangat berpotensi untuk dilakukan penambangan (*mining*) agar menghasilkan pengetahuan (*knowledge*) yang dapat berguna bagi masyarakat maupun pihak-pihak tertentu. Data dan informasi yang tersimpan di dalam web memiliki karakteristik yang berbeda dengan data yang tersimpan dalam penyimpanan konvensional. Web memiliki ukuran yang terlalu besar sehingga tidak terlalu efektif jika diterapkan dengan menggunakan *data warehouse* dan *data mining* biasa. Ukuran data dalam web mencapai ribuan *terabyte* dan akan terus berkembang. Saat ini begitu banyak perusahaan dan organisasi yang memublikasikan berbagai informasi perusahaan di sebuah website. Dengan ukuran yang begitu besar, akan sangat sulit jika harus membangun sebuah *data warehouse* yang akan menyimpan data dan informasi tersebut (Suratno, 2011).

Web mining merupakan ekstraksi pola-pola penting dan bermanfaat namun tersimpan secara implisit pada kumpulan data yang relatif besar pada layanan world wide web. *Web mining* terdiri atas tiga bagian yaitu: *web content mining*, *web structure mining*, dan *web usage mining* (Tyagi et al., 2010). Teknik *Web Mining* pertama kali dikenalkan oleh Etzioni Oren pada tahun 1996. *Web mining* dapat didefinisikan sebagai suatu usaha untuk mengimplementasikan teknik *data mining* untuk menggali dan kemudian mengkaji atau mengekstraksi informasi yang berguna

dari data yang tersimpan di web secara otomatis. Dalam implementasinya *web mining* hampir mirip akar terminologinya dengan *data mining* tetapi pada dasarnya ada sedikit perbedaan. Perbedaan *web mining* yaitu memiliki hubungan sifat data yang tidak terstruktur dan sumber data yang tidak tersimpan dalam suatu *data warehouse* melainkan tersebar di berbagai sumber.

3.1.1 *Web Content Mining*

Web Content Mining adalah suatu proses otomatis untuk menemukan informasi yang berguna dari dokumen atau data. Pada prinsipnya teknik ini mengekstraksi kata kunci yang terkandung pada dokumen. Teknik *Web Content Mining* berfokus pada penggalian informasi dari isi (*content*) yang ada pada sebuah web, seperti teks, gambar, audio, video, metadata dan *hyperlink*. Ada dua strategi yang umum digunakan: pertama langsung melakukan mining terhadap data; kedua melakukan pencarian serta mengimprove hasil pencarian seperti layaknya search engine (Srivastava et al, 2000). *Web Content Mining* juga dapat disebut sebagai *Web Text Mining* yang dalam implementasi biasanya penggalian data terletak pada daerah teks. Teknik ini juga lebih banyak berhubungan dengan disiplin *information retrieval (IR)*, tetapi dengan inovasi yang terus menerus dan seiring dengan perkembangan dunia teknologi informasi, secara dramatis akan meningkatkan akurasi dalam analisis dan mampu menghemat biaya (Hermanto dan Sartono, _).

3.1.2 *Web Structure Mining*

Web Structure Mining adalah teknik yang digunakan untuk menemukan struktur link dari *hyperlink* sehingga dapat diketahui keterkaitan antara suatu halaman web dengan halaman web lainnya, kemudian akan digunakan untuk membangun rangkuman website dan halaman web. Jadi kategori kedua dari Web Mining ini berfokus pada penggalian informasi dari isi (*content*) yang disajikan dalam web.

3.1.3 *Web Usage Mining*

Kategori ketiga dari *Web Mining* adalah *Web Usage Mining*. Model ini merupakan teknik yang selalu mengedepankan penggalian pada pola (*pattern*) dari pengguna (*user*) dalam mengakses halaman-halaman web yang ditandai melalui informasi dari log, click stream, cookies, dan query. Devi and Sreevani dalam Suratno (2011) menyatakan, manfaat *web usage mining* adalah untuk melakukan perubahan halaman berdasarkan profil pengguna, menentukan ketertarikan pelanggan terhadap produk tertentu, dan menentukan target market yang sesuai.

Dalam implementasi *Web Usage Mining* terdapat tiga (3) tahap/fase yang harus dilakukan untuk mendapatkan pustaka–pustaka maupun sumber informasi yaitu:

i. Tahap *Preprocessing*

Tahap *Preprocessing* ini adalah tahapan awal untuk memulai proses keseluruhan dari *Web Usage Mining*. Tahap ini sangat penting dikerjakan untuk melakukan standarisasi data dan menghilangkan bagian–bagian data yang tidak diperlukan

dalam proses mining. Tahap *Preprocessing* ini masih terbagi lagi menjadi beberapa tahapan, yaitu:

- i. *Data Cleaning*, yaitu tahapan yang berfungsi untuk membersihkan *file log* yang tidak relevan dari data dengan proses *mining*, seperti data multimedia dan *script* CSS maupun Javascript.
- ii. *User Identification*, yaitu tahapan yang berfungsi sebagai proses pengidentifikasian *user*. Dalam realitanya, beberapa *user* dimungkinkan untuk menggunakan komputer (*host*) yang sama.
- iii. *Session Identification*, yaitu tahapan yang dimulai setelah *user* teridentifikasi. Halaman yang diakses juga harus dibagi ke dalam sesi tertentu agar dapat tercipta sesi tunggal untuk setiap *user* yang mengakses.
- iv. *Path Completion*, yaitu tahapan untuk melengkapi *path* yang mungkin belum lengkap karena tersimpan pada *file log*.
- v. *Transaction Identification*, yaitu tahapan untuk mengidentifikasi sejumlah sesi tertentu agar dapat menunjukkan proses transaksi yang dilakukan oleh *user*.

ii. Tahap *Pattern Discovery*

Tahap kedua dari *web usage mining* ini berupa pencarian pola akses yang dilakukan oleh *user* (pengguna) aplikasi. Tahap ini adalah tahap yang paling penting dan sangat menentukan keluaran dari proses *web usage mining*. Tahap ini juga memiliki beberapa algoritma dan teknik yaitu

- i. *Statistical Analysis* (analisis statistik), yaitu teknik yang paling banyak digunakan dalam mendapatkan informasi atau pengetahuan dari pola akses user. Analisis statistik dapat disajikan dalam berbagai bentuk analisis dengan beragam variabel yang menjadi parameter analisis. Contoh analisis yang dapat dihasilkan adalah pola akses *user* yang dilihat dari waktu akses untuk setiap harinya.
- ii. *Association Rules*, yaitu teknik yang dapat diterapkan pada *Web Usage Mining*, dengan contoh keluaran yang dihasilkan berupa pola akses terhadap halaman-halaman web. Dari pola tersebut kemudian dapat diketahui halaman mana saja yang selalu diakses secara bersamaan oleh user. Hal tersebut dapat digunakan sebagai dasar untuk perancangan atau penyusunan kembali halaman web agar menjadi lebih efektif.
- iii. *Clustering*, yaitu teknik yang berupa proses pengelompokan sekumpulan objek fisik maupun abstrak ke dalam kelas tertentu berdasarkan kesamaannya. Dalam kaitannya dengan *web usage mining*, teknik clustering sering digunakan untuk menentukan segmentasi pasar pengunjung suatu situs ecommerce berdasarkan kesamaan pola akses maupun demografinya.
- iv. *Classification*, yaitu teknik yang berupa proses pengelompokan berdasarkan kelas yang sudah didefinisikan sebelumnya. *Classification* terbagi menjadi dua, yaitu proses membangun model sesuai dengan kelas yang sudah ditentukan dan proses menerapkan model untuk mengklasifikasi sekumpulan data.

- v. *Sequential Pattern*, yaitu teknik yang digunakan untuk menganalisis pola urutan akses halaman web oleh *user*. Tren urutan pola akses user dapat digunakan untuk memprediksikan tren di masa yang akan datang atau untuk mengatur penempatan menu.
- vi. *Dependency Modeling*, yaitu teknik yang berusaha mencari ketergantungan antara satu variabel dengan variabel yang lain dalam web. Hal ini berguna untuk memprediksikan pola di masa mendatang.

iii. Tahap *Pattern Analysis*

Tahap ke tiga ini merupakan fase terakhir dalam *web usage mining*. Pada fase ini dilakukan proses visualisasi hasil analisis pola yang telah dikerjakan pada langkah sebelumnya. Penyajian data menjadi hal yang penting dalam langkah ini, tergantung pada kebutuhan user dan bisnis. Dari hasil visualisasi tersebut, dapat dibuat suatu keputusan (*action*) misalnya keputusan untuk mengubah tampilan website, melakukan optimasi navigasi website, meningkatkan kemampuan website dengan melakukan *caching* halaman-halaman tertentu yang sering dikunjungi.

3.2 Server Web Nginx

Nginx [dibaca: engine x] adalah HTTP dan reverse proxy server, server proxy email, dan server proxy TCP / UDP generik, yang aslinya ditulis oleh Igor Sysoev. Untuk waktu yang lama, telah berjalan di banyak situs Rusia yang banyak dimuat

termasuk Yandex, Mail.Ru, VK, dan Rambler. Menurut Netcraft, nginx melayani atau proxied 30,46% situs tersibuk di bulan Januari 2018. Berikut adalah beberapa kisah suksesnya: Dropbox, Netflix, Wordpress.com, FastMail.FM.(Nginx Foundation, 2018)

3.2.1 Fitur Dasar Nginx

Nginx memiliki fitur dasar sebagai berikut

1. Melayani file statis dan indeks, autoindexing; buka file descriptor cache;
2. Accelerated reverse proxying dengan caching; load balancing dan fault tolerance;
3. Dukungan yang dipercepat dengan caching server FastCGI, uwsgi, SCGI, dan memcached; load balancing dan fault tolerance;
4. Arsitektur modular. Filter meliputi gzipping, rentang byte, respon chunked, XSLT, SSI, dan filter transformasi gambar. Beberapa inklusi SSI dalam satu halaman dapat diproses secara paralel jika ditangani oleh server proxy atau FastCGI / uwsgi / SCGI;
5. Dukungan SSL dan TLS SNI;
6. Dukungan untuk HTTP / 2 dengan prioritas berbasis bobot dan ketergantungan.

Server web Nginx sudah mendukung beberapa sistem operasi. Di antara sistem operasi yang sudah diuji menggunakan Nginx adalah sebagai berikut.

1. FreeBSD 3 — 11 / i386; FreeBSD 5 — 11 / amd64;
2. Linux 2.2 — 4 / i386; Linux 2.6 — 4 / amd64; Linux 3 — 4 / armv6l, armv7l, aarch64, ppc64le;
3. Solaris 9 / i386, sun4u; Solaris 10 / i386, amd64, sun4v;
4. AIX 7.1 / powerpc;
5. HP-UX 11.31 / ia64;
6. macOS / ppc, i386;
7. Windows XP, Windows Server 2003.

Selain fitur dasar yang sudah disebutkan pada paragraf sebelumnya. Server nginx juga mempunyai fitur dasar yang berkaitan dengan akses database dan akses e-mail. Fitur-fitur ini berkaitan dengan kemampuan server Nginx sebagai reverse proxy, sehingga dapat digunakan sebagai pelindung pada jaringan private.

3.2.2 Log Nginx

Server web Nginx mendukung pencatatan log yang dapat diatur formatnya. Pengaturan pencatatan log pada server web Nginx menggunakan tiga pengarah yaitu pengarah `log_format`, `access_log` dan `open_log_file_cache` (dokumentasi Nginx). `Log_format` digunakan untuk mengatur format pencatatan log yang dihasilkan, sedangkan `access_log` digunakan untuk menentukan nama file tempat menyimpan pencatatan log.

Sintaksis cara penggunaan pengarah `log_format` beserta nilai default dapat dituliskan sebagai berikut.

```
Syntax:
log_format name [escape=default|json] string ...;

Default:
log_format combined "...";
```

Gambar 3.1 Sintaksis Log Format

Parameter `escape` memungkinkan pengaturan menggunakan format json atau pemisahan karakter dalam variabel. Secara default digunakan `escape default`. Format log dapat mengandung variable umum dan variabel yang hanya ada pada waktu log ditulis. Variabel yang bisa ada pada format log antara lain sebagai berikut.

1. `$bytes_sent`: Jumlah byte yang dikirim ke client
2. `$connection`: connection serial number
3. `$connection_requests`: jumlah request sesaat yang dibuat melalui suatu connection
4. `$msec`: waktu dalam second dengan resolusi millisecond pada waktu penulisan log yang memerlukan waktu lama.
5. `$pipe`: "p" jika request dipipakan, "." jika request tidak dipipakan
6. `$request_length`: panjang request (termasuk baris request, header, dan body request)

7. `$request_time`: waktu pemrosesan request dalam second dengan resolusi millisecond; Rentang waktu sejak time diluncurkan dengan bytes pertama yang dibaca dari client and penulisan log write setelah bytes terakhir yang dikirim client
8. `$status`: response status
9. `$time_iso8601`: local time dalam bentuk standar format ISO 8601
10. `$time_local`: local time dalam format log umum

Header line yang dikirim ke client punya prefix “sent_http_”, misalnya `$sent_http_content_range`. Pada konfigurasi selalu menyertakan predefined format “combined” misalnya sebagai berikut.

```
access_log logs/access.log combined;
ed '$remote_addr - $remote_user
[$time_local] '"$request" $status $body_bytes_sent
'"$http_referer"'"$http_user_agent"';
```

Gambar 3.2 Log Format

Pengarah berikutnya dalam kaitannya dengan pencatatan log adalah `access_log`. Sintaksis cara penggunaan pengarah `access_log` adalah sebagai berikut.

```
Syntax:
access_log path [format [buffer=size] [gzip[=level]] [flush=time]
[if=condition]];
access_log off;
Default:
access_log logs/access.log combined;
```

Gambar 3.3 Sintaksis Access Log

Access_log digunakan untuk pengaturan path, format, dan konfigurasi buffer untuk penulisan log. Beberapa log dapat ditentukan pada level yang sama. Logging ke syslog dapat dikonfigurasi menggunakan pengarah prefiks "syslog:" pada parameter pertama. Nilai khusus pengarah access_log adalah off, cancels atau all berada pada level sesaat. Jika format tidak ditentukan, maka nilai format predefined "combined" yang digunakan. Nginx Foundation, 2018. Jika salah satu nilai parameter buffer atau gzip digunakan, penulisan log akan ditampung dalam buffer. Ukuran buffer harus lebih besar dibanding nilai terkecil penulisan ke disk file. Untuk FreeBSD, ukuran tersebut unlimited.

3.3 Situs perpus.akakom.ac.id

Perpustakaan STMIK AKAKOM, merupakan salah satu unit yang berada di kampus, baru menerapkan teknologi informasi yang berbasis internet secara menyeluruh terintegrasi mulai tahun 2016 dengan memanfaatkan aplikasi yang dikembangkan oleh bagian Sistem Informasi sebagai *front-end*. Aplikasi *front-end* ini diberi nama domain perpus.akakom.ac.id yang berfungsi untuk menampilkan halaman utama dengan menu-menu navigasinya dan berfungsi untuk menjembatani pengunjung dengan aplikasi perpustakaan lainnya. Menu utama yang terdapat pada baris navigasi disediakan dalam situs resmi perpustakaan ini adalah Home, Kontak, Tautan, Layanan, Profil, Katalog Online, dan Digital Library. Di dalam setiap menu

ini terdapat beberapa sub menu yang menghubungkan pengunjung dengan halaman-halaman selanjutnya dari situs ini.

Menu ‘Home’ adalah tautan yang menampilkan halaman utama dari situs perpustakaan ini. Jadi jika pengunjung sedang berada pada satu halaman dalam situs tersebut dan ingin menuju kembali ke halaman utama dapat menggunakan menu ini. Menu ‘Kontak’ adalah tautan yang menampilkan informasi identitas perpustakaan baik berupa tulisan alamat lengkap maupun tampilan gambar peta lokasi. Menu selanjutnya adalah ‘Tautan’ yang akan menyediakan submenu di bawahnya seperti OPAC, Digital Library, Jurnal Dikti, Proquest, Ebsco Host, Jiko AKAKOM, Directory of Open Access Journals, Gale dan Cengage, E-resources Perpustakaan Nasional, Akses Aplikasi Buku Digital, dan Indonesia Onesearch. Menu berikutnya adalah ‘Layanan’ yang akan menyediakan submenu berupa layanan untuk CD, Pustaka Rujukan/Referensi, Buku (Sirkulasi), Majalah dan Koran, Penyerahan Proyek Akhir/Skripsi, Bebas Perpustakaan, Penelusuran Informasi, dan Pesan Buku Online. Selanjutnya terdapat menu ‘Profil’ yang berisi submenu Visi Misi, Sejarah, Tata Tertib, dan Staf Perpustakaan. Dua (2) menu berikutnya adalah ‘Katalog Online’ dan ‘Digital Library’. Menu ‘Katalog Online’ ini adalah merupakan tautan yang menghubungkan pengunjung menuju aplikasi OPAC (*Open Public Access Catalogue*) atau aplikasi yang lebih dikenal dengan SLiMS, sedangkan menu ‘Digital Library’ akan menghubungkan pengunjung menuju aplikasi *Digital Library* atau eprints. Dalam setiap tampilan dari menu yang ada selalu akan menyediakan tautan

navigasi menu Home, Koleksi, Berita, dan Pengumuman untuk memudahkan pengunjung. Menu ini dipandang merupakan tautan yang paling sering diakses.

Selain dari menu utama pada halaman utama ini terdapat juga menu lain yang terletak di panel bagian kanan yaitu Koleksi Terbaru dan di bagian bawah terdapat panel Berita, Pengumuman, dan Tautan Terkait. Menu ‘Koleksi’ berguna untuk menghubungkan pengunjung dengan halaman informasi daftar koleksi perpustakaan yang terbaru. Menu ‘Berita’ berguna untuk menampilkan panel dan halaman informasi tentang aktivitas dan kegiatan yang dilaksanakan oleh perpustakaan atau perihal yang terkait dengan pihak luar perpustakaan. Menu ‘Pengumuman’ berguna untuk menampilkan informasi tentang aktivitas atau kegiatan yang berhubungan langsung dengan keberadaan para pustakawan. Menu ‘Tautan Terkait’ berfungsi untuk menyediakan link-link menuju laman-laman yang erat kaitannya dengan kebutuhan pustaka, baik dalam lingkup lokal, nasional, maupun internasional.

BAB IV METODE PENELITIAN

4.1 Tahapan Penelitian

Pengerjaan penelitian ini akan terdiri dari beberapa tahap yaitu

1. Studi pustaka bersumber dari referensi yang relevan dengan topik yang akan diteliti
2. Konfigurasi server web Nginx
3. Mengumpulkan data *log access* selama waktu tertentu
4. Mengolah data *log access*
5. Membuat pembahasan hasil pengolahan data
6. Membuat kesimpulan

4.2 Lokasi Penelitian

Penelitian ini menggunakan objek situs perpus.akakom.ac.id yang merupakan situs resmi dari perpustakaan STMIK AKAKOM. Pengamatan dilakukan selama satu semester yaitu semester ganjil tahun akademik 2017/2018.

4.3 Model yang digunakan

Model yang digunakan dalam penelitian ini adalah *Web Usage Mining*. Model ini merupakan teknik ini selalu mengedepankan penggalian pada pola (*pattern*) dari *user* dalam mengakses halaman-halaman web. Dalam implementasi *Web Usage*

Mining terdapat tiga (3) fase yang harus dilakukan untuk mendapatkan pustaka–pustaka maupun sumber informasi yaitu:

4.3.1 Tahap *Preprocessing*

Tahap *Preprocessing* ini adalah tahapan awal untuk memulai proses keseluruhan dari *Web Usage Mining*. Tahap ini sangat penting dikerjakan untuk melakukan standarisasi data dan menghilangkan bagian–bagian data yang tidak diperlukan dalam proses *mining*. Tahap *Preprocessing* ini terbagi atas beberapa tahapan, yaitu:

- i. *Data Cleaning*, yaitu tahapan yang berfungsi untuk membersihkan *file log* yang tidak relevan dari data dengan proses *mining*, seperti data multimedia dan *script* CSS maupun Javascript.
- ii. *User Identification*, yaitu tahapan yang berfungsi sebagai proses pengidentifikasian *user*. Dalam realitanya, beberapa *user* dimungkinkan untuk menggunakan komputer (*host*) yang sama.
- iii. *Session Identification*, yaitu tahapan yang dimulai setelah *user* teridentifikasi. Halaman yang diakses juga harus dibagi ke dalam sesi tertentu agar dapat tercipta sesi tunggal untuk setiap *user* yang mengakses.
- iv. *Path Completion*, yaitu tahapan untuk melengkapi *path* yang mungkin belum lengkap karena tersimpan pada *file log*.
- v. *Transaction Identification*, yaitu tahapan untuk mengidentifikasi sejumlah sesi tertentu agar dapat menunjukkan proses transaksi yang dilakukan oleh *user*.

4.3.2 Tahap *Pattern Discovery*

Tahap kedua dari *web usage mining* ini berupa pencarian pola akses yang dilakukan oleh *user* (pengguna) aplikasi. Tahap ini adalah tahap yang paling penting dan sangat menentukan keluaran dari proses *web usage mining*. Tahap ini juga memiliki beberapa algoritma dan teknik yaitu

- i. *Statistical Analysis* (analisis statistik), yaitu teknik yang paling banyak digunakan dalam mendapatkan informasi atau pengetahuan dari pola akses user. Analisis statistik dapat disajikan dalam berbagai bentuk analisis dengan beragam variabel yang menjadi parameter analisis. Contoh analisis yang dapat dihasilkan adalah pola akses *user* yang dilihat dari waktu akses untuk setiap harinya.
- ii. *Association Rules*, yaitu teknik yang dapat diterapkan pada *Web Usage Mining*, dengan contoh keluaran yang dihasilkan berupa pola akses terhadap halaman-halaman web. Dari pola tersebut kemudian dapat diketahui halaman mana saja yang selalu diakses secara bersamaan oleh user. Hal tersebut dapat digunakan sebagai dasar untuk perancangan atau penyusunan kembali halaman web agar menjadi lebih efektif.
- iii. *Clustering*, yaitu teknik yang berupa proses pengelompokan sekumpulan objek fisik maupun abstrak ke dalam kelas tertentu berdasarkan kesamaannya. Dalam kaitannya dengan *web usage mining*, teknik clustering sering digunakan untuk menentukan segmentasi pasar

pengunjung suatu situs ecommerce berdasarkan kesamaan pola akses maupun demografinya.

- iv. *Classification*, yaitu teknik yang berupa proses pengelompokan berdasarkan kelas yang sudah didefinisikan sebelumnya. *Classification* terbagi menjadi dua, yaitu proses membangun model sesuai dengan kelas yang sudah ditentukan dan proses menerapkan model untuk mengklasifikasi sekumpulan data.
- v. *Sequential Pattern*, yaitu teknik yang digunakan untuk menganalisis pola urutan akses halaman web oleh *user*. Tren urutan pola akses user dapat digunakan untuk memprediksikan tren di masa yang akan datang atau untuk mengatur penempatan menu.
- vi. *Dependency Modeling*, yaitu teknik yang berusaha mencari ketergantungan antara satu variabel dengan variabel yang lain dalam web. Hal ini berguna untuk memprediksikan pola di masa mendatang.

4.3.3 Tahap *Pattern Analysis*

Tahap ke tiga ini merupakan fase terakhir dalam *web usage mining*. Pada fase ini dilakukan proses visualisasi hasil analisis pola yang telah dikerjakan pada langkah sebelumnya. Penyajian data menjadi hal yang penting dalam langkah ini, tergantung pada kebutuhan user dan bisnis. Dari hasil visualisasi tersebut, dapat dibuat suatu keputusan (*action*) misalnya keputusan untuk mengubah tampilan website,

melakukan optimasi navigasi website, meningkatkan kemampuan website dengan melakukan *caching* halaman-halaman tertentu yang sering dikunjungi.

4.4 Rancangan Penelitian

Penelitian ini dipersiapkan dengan rancangan berikut.

1. menentukan objek penelitian
2. mengkonfigurasi server web
3. mengumpulkan *log access* selama kurun waktu tertentu
4. konversi data ke dalam format MySQL
5. mengolah data akses
6. membuat kesimpulan

4.5 Teknik Pengumpulan Data dan Analisis Data

Pengerjaan penelitian ini melibatkan kegiatan-kegiatan dengan rincian

1. mengaktifkan mode log konfigurasi web server Nginx
2. mengamati log akses server web selama waktu tertentu
3. mengkonversi log akses server web dalam database MySQL
4. mengelompokkan data berdasar halaman situs
5. menghitung jumlah akses untuk tiap halaman situs
6. memetakan distribusi akses halaman situs

BAB V

HASIL PENELITIAN DAN PEMBAHASAN

5.1 Hasil Penelitian

5.1.1 Konfigurasi server web Nginx

Konfigurasi server web Nginx secara umum dituliskan dalam file `/etc/nginx/nginx.conf`. Dalam distro-distro Linux tertentu, letak file konfigurasi ditempatkan pada direktori yang bervariasi. Namun demikian, isi file konfigurasi dapat dikatakan sama. Isi konfigurasi sangat banyak mulai dari nama server, alamat IP kerja, port kerja, letak dokumen HTML, pengaturan log baik log error maupun log akses, *virtual host*, pengaturan izin akses direktori dan sebagainya. Pengaturan konfigurasi yang berkaitan dengan penelitian ini adalah pengaturan log server khusus berkaitan dengan log akses pada situs.

Supaya server web Nginx dapat menghasilkan log akses, maka pada file konfigurasi perlu dituliskan pengarah yang mengatur pencatatan log. Pengaturan terutama berupa format log serta file pencatatan log. Pengaturan yang dilakukan berbentuk sebagai berikut.

```
log_format main '$remote_addr - $remote_user [$time_local]
"$request"
'$status $body_bytes_sent "$http_referer"
'"$http_user_agent"$http_x_forwarded_for"';

access_log /var/log/nginx/access.log main;
```

Gambar 5.1 Format Log

Konfigurasi tersebut, terutama mengatur bentuk catatan log yang dihasilkan dan file tempat menyimpan catatan log tersebut. Format catatan log diatur pada pengarah `log_format`. Letak file tempat menyimpan catatan log diatur pada pengarah `access_log`. Pada pengaturan tersebut, catatan log disimpan dalam file teks `access.log`.

5.1.2 Pengumpulan data log akses

Data log akses dikumpulkan selama waktu tertentu yang dalam penelitian ini dilakukan dalam kurun waktu lebih dari dua bulan. Selama kurun waktu tersebut web Akakom akan diakses secara acak oleh pengguna baik dari lingkungan kampus maupun dari luar kampus. Dalam penelitian tidak diperhatikan dari mana dan oleh siapa secara detail web perpustakaan diakses. Beberapa hal yang diperhatikan antara lain asal alamat IP (apakah lokal atau luar AKAKOM), halaman mana yang diakses, serta diakses menggunakan sistem apa.

5.1.3 Mengolah data log akses

Log akses Nginx mengandung beberapa informasi seperti alamat IP user yang mengakses, waktu akses, halaman yang diakses, *browser* untuk akses, dan sebagainya. Informasi yang diperlukan dalam penelitian ini terutama adalah halaman akses, *browser*, dan sistem operasi yang dipakai pengguna. Sebagai informasi tambahan dapat juga dilibatkan alamat IP dan waktu akses. Informasi tersebut dapat diperoleh pada kolom satu untuk IP, kolom empat untuk waktu, kolom enam untuk halaman akses dan kolom sebelas untuk *browser* yang dipakai pengguna. Kolom

sebelas juga dapat menunjukkan sistem operasi yang dipakai pengguna. Kolom-kolom tersebut berdasarkan pemisah spasi dengan prioritas tanda double quote ("). Contoh isi file access.log adalah sebagai berikut.

```
5.45.207.50 - - [17/Nov/2017:04:03:36 +0700] "GET /site/index.php?r=berita
%2Fview&id=27&slug=pustakawan-stmik-akakom-yogyakarta-menjadi-pemakalah-
dialog-ilmiah-perpustakaan-ke-5 HTTP/1.1" 200 18316 "-"Mozilla/5.0
(compatible; YandexBot/3.0; +http://yandex.com/bots)"
40.77.167.149 - - [17/Nov/2017:05:18:26 +0700] "GET / HTTP/1.1" 302 5
-"-"Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)"
66.249.77.28 - - [17/Nov/2017:05:19:02 +0700] "GET /robots.txt HTTP/1.1" 404
162 "-"Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"
66.249.77.27 - - [17/Nov/2017:05:19:04 +0700] "GET /site/index.php?
r=koleksi-terbaru%2Fview&id=20820 HTTP/1.1" 200 16788 "-"Mozilla/5.0
(Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML,
like Gecko) Chrome/41.0.2272.96 Mobile Safari/537.36 (compatible; Googlebot/
2.1; +http://www.google.com/bot.html)"
34.228.199.214 - - [17/Nov/2017:05:29:21 +0700] "GET / HTTP/1.1" 302 5
-"-"Mozilla/5.0 (compatible; SurdotlyBot/1.0; +http://sur.ly/bot.html)"
34.228.199.214 - - [17/Nov/2017:05:29:22 +0700] "GET ./site HTTP/1.1" 301
178 "-"Mozilla/5.0 (compatible; SurdotlyBot/1.0; +http://sur.ly/bot.html)"
34.228.199.214 - - [17/Nov/2017:05:29:24 +0700] "GET /site/ HTTP/1.1" 200
33340 "-"Mozilla/5.0 (compatible; SurdotlyBot/1.0;
+http://sur.ly/bot.html)"
34.228.199.214 - - [17/Nov/2017:05:29:26 +0700] "GET
/site/img/logo_akakom.png HTTP/1.1" 304 0 "-"Mozilla/5.0 (X11; U; Linux
x86_64; en-US) AppleWebKit/532.9 (KHTML, like Gecko) Chrome/5.0.307.7
Safari/532.9"
5.45.207.50 - - [17/Nov/2017:05:38:21 +0700] "GET /site/img/Book-icon.png
HTTP/1.1" 304 0 "-"Mozilla/5.0 (compatible; YandexImages/3.0;
+http://yandex.com/bots)"
46.229.164.100 - - [17/Nov/2017:05:38:47 +0700] "GET /robots.txt HTTP/1.1"
404 162 "-"Mozilla/5.0 (compatible; SemrushBot/1.2~bl;
+http://www.semrush.com/bot.html)"
```

Gambar 5.2 Log Access Nginx

Untuk mempermudah pengolahan, data access.log perlu diubah atau dimasukkan dalam sistem database MySQL. Untuk keperluan tersebut, terutama perlu disiapkan tabel untuk menampung data access.log. Tabel yang diperlukan punya struktur sebagai berikut.

Tabel 5.1 Struktur Tabel Log

Column	Type	Comment
id	int(10) <i>Auto Increment</i>	
ip	varchar(20)	
1	varchar(2)	
2	varchar(2)	
date	varchar(40)	
method	varchar(5)	
site	varchar(200)	
version	varchar(10)	
7	varchar(3)	
8	varchar(7)	
9	varchar(200)	
browser	varchar(200)	

Untuk mengubah atau memasukkan data access.log dalam sistem database MySQL sesuai format data pada Tabel 4.1 digunakan skrip PHP yang berbentuk sebagai berikut.

```
<?php
$con=mysqli_connect("localhost","root","");
mysqli_select_db($con,"web_perpus");

for($i=1;$i<=23;$i++)
{
if($i<10)
$file="perpus.access.log-2018010".$i;
else
$file="perpus.access.log-201801".$i;

$f=fopen($file, "r") or die("Unable to open file!");
while(!feof($f))
{
$row=fgets($f);
$str1="";
$pos1=0;
$sql="";
while($pos2=strpos($row,$str1,$pos1))
{
$col=substr($row,$pos1,$pos2-$pos1);
$col=str_replace(";","",$col);
if(strpos("_".$col,"GET")==1) $col=str_replace("",";",
$col);
$sql=$sql.";".$col;
}
```

```

if($str1=="|" || $str1=="\"") $pos2=$pos2+1;
$pos1=$pos2+1;

if($pos2>strlen($row)-2) continue;
if($row[$pos2+1]=="[")
{
$str1="]";
$pos1++;
}
else if($row[$pos2+1]=="\"")
{
$str1="\"";
$pos1++;
}
else
{
$str1="";
}

$sql=str_replace(";", "','', $sql);
if($sql<>'')
{
$sql="insert into log values ('".$sql."')";
$result=mysqli_query($con,$sql);
}

fclose($f);
}
mysqli_close($con);
?>

```

Gambar 5.3 Sekrip Parsing Log

Bagian pertama sekrip adalah koneksi menuju server database. Untuk koneksi ke server database diperlukan nama host, username dan password. Pada bagian ini digunakan nama host localhost, username xxxxx dan password yyyy. Username dan password yang dituliskan pada makalah ini sengaja disamarkan.

Bagian kedua sekrip adalah untuk membuka file catatan log server web Nginx. Bentuk nama file catatan log server web Nginx adalah perpus.access.log-20180102. Bagian paling depan adalah nama penanda bahwa file tersebut sebagai file log akses situs perpustakaan, sedangkan bagian akhir adalah tanggal backup file log. Tanggal backup sesuai dengan periodisasi backup file log dari sistem pewaktu server

web. Untuk memudahkan penyusunan sekrip, pengolahan file dilakukan untuk tiap bulan, sehingga dilakukan pembukaan file untuk setiap bulan pembukaan secara berulang dari awal tanggal sampai akhir tanggal.

Bagian ketiga adalah proses parsing untuk memisahkan bagian bagian file log. Parsing dilakukan untuk tiap baris isi file log. File log memisahkan antar bagian kolom menggunakan spasi, namun pemisahan ini dikombinasikan dengan penyatu string ("...") dan kurung siku ([...]). Bagian yang dipisahkan dengan spasi adalah alamat IP asal, Bagian yang dibatasi dengan kurung siku adalah waktu akses. Bagian yang dibatasi dengan penanda string adalah halaman yang diakses dan peralatan yang dipakai pengguna untuk akses. Bagian parsing mengelompokkan bagian-bagian ini supaya menjadi format SQL yang dapat diterapkan dalam database MySQL.

Hasil keluaran sekrip ini berupa suatu tabel log yang tersimpan dalam sistem database MySQL. Selanjutnya pengolahan data dilakukan melalui sistem database MySQL.

5.1.4 Pengolahan Data MySQL

Pengolahan data pada sistem database MySQL terutama dilakukan menggunakan perangkat lunak bantu pengolahan database adminer. Adminer adalah program pengolah database yang disusun oleh Jakub Vrána. Tool adminer mendukung banyak sistem database yaitu MySQL, MariaDB, PostgreSQL, SQLite, MS SQL, Oracle, Firebird, SimpleDB, Elasticsearch dan MongoDB.(Vrána, J., 2017)

Langkah awal pengolahan data melalui MySQL adalah membuat database. Database disiapkan bersama username serta password yang digunakan. Nama database yang dipakai adalah web_perpusdb. Selanjutnya disiapkan tabel log untuk menampung hasil pengubahan catatan file akses log menggunakan sekrip yang sudah diuraikan pada paragraf sebelumnya. Struktur seperti yang ditunjukkan pada Table 4.1 dibuat menggunakan perintah SQL pada sistem database MySQL sebagai berikut.

```
CREATE TABLE `log` (  
  `id` int(10) NOT NULL AUTO_INCREMENT,  
  `ip` varchar(20) NOT NULL,  
  `1` varchar(2) NOT NULL,  
  `2` varchar(2) NOT NULL,  
  `date` varchar(40) NOT NULL,  
  `method` varchar(5) NOT NULL,  
  `site` varchar(200) NOT NULL,  
  `version` varchar(10) NOT NULL,  
  `7` varchar(3) NOT NULL,  
  `8` varchar(7) NOT NULL,  
  `9` varchar(200) NOT NULL,  
  `browser` varchar(200) NOT NULL,  
  PRIMARY KEY (`id`)  
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

Gambar 5.4 Perintah SQL Membuat Tabel Log

Setelah data berhasil masuk pada database MySQL, pengolahan selanjutnya dilakukan menggunakan perintah SQL. Perintah SQL dapat dijalankan melalui Adminer pada kolom SQL command. Hasil tabel log dimodifikasi untuk mempermudah pengolahan dengan pengubahan nama field sesuai dengan yang diperlukan dalam penelitian ini. Nama tabel yang digunakan hasil modifikasi adalah log_modi.

Informasi yang diinginkan adalah jumlah akses pada suatu halaman web selama kurun waktu tertentu. Dengan demikian harus dihitung cacah akses setiap halaman yang sama. Hasilnya harus berupa halaman akses versus jumlah akses. Untuk mendapatkan informasi tersebut, dapat digunakan perintah SQL sebagai berikut.

```
select site, count(*) as count from log_modi group by site  
order by count desc limit 105
```

Gambar 5.5 Halaman Yang Paling Sering Dikunjungi

Maksud perintah SQL tersebut adalah menampilkan field site dan cacah site sama sebagai jumlah. Hasilnya dikelompokkan oleh site. Selanjutnya hasil diurutkan berdasarkan hasil jumlah akses secara urut turun (descending). Hasilnya difilter lagi supaya hanya ditampilkan 105 urutan jumlah akses terbesar. Hasilnya dapat diekspor dalam file csv yang dapat diolah lebih mudah, karena dibatasi sampai 105 data terbesar.

Pada penelitian juga dicoba untuk melempokkan asal alamat IP yang paling banyak menggunakan situs perpustakaan. Untuk mendapatkan cacah alamat Ip paling sering menggunakan situs perpustakaan digunakan perintah SQL sebagai berikut.


```
select ip, count(*) as count from log_modi group by ip order  
by count desc
```

Gambar 5.6 Asal Alamat IP yang Paling Sering mengunjungi situs perpustakaan

Dari hasil pengolahan ini bisa dikelompokkan lagi untuk menentukan apakah alamat IP berasal (atau lebih tepat menggunakan) dari jaringan lokal Akakom atau menggunakan jaringan luar Akakom. Alamat IP yang berasal dari jaringan Akakom adalah 172.17.0.0/16, 172.18.0.0/16 dan 10.0.0.0/8. Selain alamat IP tersebut dianggap berasal dari jaringan luar Akakom.

Pengolahan selanjutnya adalah berdasarkan peralatan yang dipakai oleh pengguna untuk mengakses situs perpustakaan. Untuk memperoleh data ini digunakan perintah SQL sebagai berikut.

```
SELECT * FROM `log_modi` WHERE `browser` LIKE '%Windows%'
```

Gambar 5.7 Mencari Sistem Operasi atau Browser yang Dipakai Pengguna

Contoh SQL tersebut untuk memperoleh jumlah pengguna yang menggunakan sistem operasi Windows. Untuk memperoleh pengguna yang menggunakan sistem operasi lain, misal Linux, parameter filter diganti dengan x86. Pencarian data menggunakan filter-filter: windows, Linux, Android, Go-http-client, Googlebot-Image, MJ12bot, Yandex, bingbot, AhrefsBot, Macintosh, Baiduspider dan iPhone. Untuk memastikan masing-masing filter berasal dari peralatan apa, perlu dilakukan pencermatan lebih

lanjut. Hal demikian disebabkan variasi jumlah peralatan yang dipakai pengguna sangat banyak. Dalam penelitian ini belum dilakukan pencermatan terhadap variabel tersebut.

5.2 Pembahasan

Berdasarkan hasil pengolahan data MySQL tersebut dilakukan analisis sebagai berikut. Dalam rentang waktu 17/Nov/2017:04:03:36 +0700 sampai dengan 23/Jan/2018:02:19:24 +0700, secara keseluruhan tercatat sebanyak 93.286 kali akses yang menuju situs. Dengan dibatasi pada peringkat 30 terbanyak URL yang diakses dapat ditampilkan dalam Tabel 5.2.berikut.

Tabel 5.2. Daftar URL peringkat 30 terbanyak yang diakses

No.	site	jumlah pengakses
1	/site	29619
2	/site/	29547
3	/	1299
4	/site/index.php?r=category%2Fview&id=10	182
5	/site/index.php?r=koleksi-terbaru%2Findex	106
6	/site/index.php?r=category%2Fview&id=9	79
7	/site/index.php?r=site%2Findex	76
8	/wp-login.php	76
9	/site/index.php	73
10	/site/index.php?r=post%2Fview&id=10	73
11	/site/_file/berita/berita_28_2866_dXBkYXRl.png	71
12	/site/_file/berita/berita_29_6997_Y3JlYXRl.png	67
13	/pustakawan/	63
14	/site/index.php?r=berita%2Findex	55
15	/site/_file/berita/berita_23_2330_Y3JlYXRl.png	55
16	/site/index.php?r=site%2Fcontact	50
17	/site/index.php?r=berita %2Fview&id=29&slug=perpustakaan-stmik-akakom-resmi- launching-aplikasi-buku-digital	50
18	/site/index.php?r=post%2Fview&id=3	44
19	/site/_file/berita/berita_27_1508_Y3JlYXRl.png	43
20	/site/_file/berita/berita_26_1115_Y3JlYXRl.png	40
21	/site/index.php?r=site%2Findex&page=2&per-page=3	40
22	/pustakawan	39
23	/pustakawan/index.php	37
24	/site/_file/berita/berita_19_5716_dXBkYXRl.png	36
25	/site/_file/berita/berita_22_9726_Y3JlYXRl.png	35
26	/site/index.php?r=koleksi-terbaru %2Findex&page=2&per-page=10	33
27	/site/_file/berita/berita_20_7704_Y3JlYXRl.png	33
28	/.well-known/apple-app-site-association	29
29	/apple-app-site-association	28
30	/site/index.php?r=berita %2Fview&id=28&slug=perpustakaan-stmik-akakom- segera-launching-aplikasi-buku-digital	27

Berdasarkan URL *address* yang tercatat tersebut dapat diidentifikasi halaman-halaman yang dituju dalam situs. Pemeriksaan halaman yang dituju dikerjakan dengan menuliskan setiap URL pada URL *address* pada browser dan membukanya. Informasi yang diperoleh kemudian disusun berdasarkan urutan terbanyak akses halaman yang dituju. Dari informasi itu dipilih 30 halaman yang terbanyak seperti ditampilkan pada Tabel 5.3 berikut.

Tabel 5.3 Daftar Menu/Navigasi/Link yang dikunjungi

No.	Menu	jumlah	detail halaman dituju
1	Home	29619	
2	Home	29547	
3	Home	1299	
4	digital library	182	
5	Koleksi	106	koleksi terbaru
6	katalog online	79	
7	Home	76	
8	link tidak ditemukan	76	
9	Home	73	
10	Layanan	73	Penyerahan Proyek Akhir / Skripsi
11	Berita	71	koleksi digital
12	Berita	67	sosialisasi buku digital
13	Tautan	63	login pustakawan
14	Berita	55	halaman utama berita
15	Berita	55	koleksi foto perpustakaan
16	Berita	50	acara launching buku digital
17	Kontak	50	
18	Layanan	44	buku (sirkulasi)
19	Berita	43	koleksi foto kegiatan pendidikan pemakai
20	Berita	40	koleksi foto kegiatan pendampingan unggah mandiri
21	Home	40	
22	Tautan	39	login pustakawan
23	Tautan	37	login pustakawan
24	Berita	36	seminar nasional
25	Berita	35	koleksi foto
26	Berita	33	koleksi foto
27	Koleksi	33	koleksi terbaru
28	link tidak ditemukan	29	
29	link tidak ditemukan	28	
30	Berita	27	perpustakaan akakom segera launching buku digital

Dari Tabel 5.3 tersebut dapat dideskripsikan bahwa halaman yang paling banyak dikunjungi adalah akses dari menu 'Home' dengan total jumlah pengunjung selama pengamatan sebesar 60.654 kali (dengan URL /site, /site/, atau /). Kondisi ini dapat diterima karena sebagian pengunjung yang menuju ke situs perpustakaan akakom.ac.id kebanyakan akan melalui menu 'Home' ini. Halaman yang sering dikunjungi berikutnya adalah dari link 'Berita' dengan jumlah pengakses sebanyak 512 kali pada detail halaman koleksi digital (71 kali), sosialisasi buku digital (67 kali), halaman utama berita (55 kali), koleksi foto perpustakaan (55 kali), acara *launching* buku digital (50 kali), koleksi foto kegiatan pendidikan pemakai (43 kali), koleksi foto kegiatan pendampingan unggah mandiri (40 kali), koleksi foto (68 kali), dan berita tentang perpustakaan akakom yang akan segera me-*launching* buku digital (27 kali). Selanjutnya terdapat menu 'Digital Library' yang diakses sebanyak 182 kali dalam periode yang telah ditentukan. Menu ini berhubungan dengan tampilan halaman Koleksi Terbaru. Menu selanjutnya yang banyak dikunjungi adalah 'Koleksi' dengan jumlah pengunjung sebesar 139 kali. Jumlah yang sama (139 kali) juga ditujukan untuk menu 'Tautan'. Menu ini berhubungan dengan permintaan halaman untuk Login sebagai pustakawan. 139 kali dalam kurun waktu tersebut menunjukkan bahwa pustakawan cukup sering berkunjung ke situs yang merupakan media tempat lingkup pekerjaannya berlangsung. Selanjutnya terdapat 133 kali akses pada URL yang ternyata tidak ditemukan hasilnya (404 Not Found) yaitu dari URL /.well-known/apple-app-site-association, /apple-app-site-association, dan /wp-login.php. Berikutnya terdapat menu 'Layanan' yang telah diakses sebanyak 117 kali.

Menu ‘Layanan’ ini berhubungan dengan Penyerahan Proyek Akhir / Skripsi dan Sirkulasi Buku. Menu selanjutnya adalah ‘Kalatog Online’ yang dikunjungi sebanyak 79 kali dan menu ‘Kontak’ yang diakses sebanyak 50 kali. Daftar urutan terbanyak halaman tersebut ditampilkan dalam Tabel 5.4 berikut.

Tabel 5.4 Daftar Halaman yang Diakses

Halaman	jumlah
Home	60654
Berita	512
digital library	182
Koleksi	139
Tautan	139
link tidak ditemukan	133
Layanan	117
katalog online	79
Kontak	50

Bersumber dari data IP paling sering berkunjung, dapat diperhatikan bahwa dalam kurun waktu yang telah ditentukan, pengunjung yang mengakses situs perpustakaan.akom.ac.id berasal dari 1.301 alamat *internet protocol* (IP address), dengan total jumlah akses sebanyak 93.286 kali. Dari data tersebut jika dibatasi hanya pada peringkat 30 besar asal alamat IP yang sering mengunjungi situs dapat disajikan dalam tabel seperti pada Tabel 5.5 berikut.

Tabel 5.5. Daftar alamat IP peringkat 30 terbanyak yang sering berkunjung

No.	IP	jumlah akses
1	88.198.23.179	3585
2	46.4.58.130	2916
3	144.76.38.73	2800
4	144.76.8.66	2678
5	5.9.60.241	2671
6	5.9.61.111	2100
7	192.151.157.210	2029
8	136.243.68.226	1800
9	5.9.144.234	1798
10	144.76.30.241	1795
11	69.197.177.50	1772
12	5.9.147.133	1768
13	46.4.32.75	1761
14	5.9.89.170	1756
15	158.69.252.176	1740
16	46.4.91.109	1672
17	173.208.157.186	1643
18	69.30.213.202	1543
19	91.209.51.22	1474
20	79.211.69.33	1349
21	62.210.157.129	1323
22	69.30.198.178	1277
23	62.210.157.95	1227
24	193.164.132.237	1147
25	173.208.128.34	1139
26	88.198.66.230	1079
27	144.76.12.66	1048
28	46.4.0.246	1015
29	5.9.156.74	1013
30	158.69.254.103	999

Berdasarkan Tabel 5.5 dapat disebutkan bahwa asal pengunjung yang terbanyak, sejumlah 3.585 kali ke situs `perpus.akakom.ac.id` adalah yang berasal dari IP *address* 88.198.23.179. Jika mengacu pada data IP *address* di lokasi penelitian, maka seluruh IP *address* tersebut berasal dari lokasi di luar kampus STMIK AKAKOM. Jika ingin diketahui asal IP pengunjung dari lokal STMIK AKAKOM, maka data tersebut diproses kembali sehingga menghasilkan informasi asal pengakses berdasarkan area dan jenisnya di lokal STMIK AKAKOM yaitu.

Tabel 5.6 Daftar Area dan Jenis Akses

No.	Alamat IP	Area akses	Jumlah
1	172.17.	gedung timur	465
2	172.18.	laboratorium	40
3	10.10.	hotspot akakom	501

Deskripsinya adalah, terdapat sejumlah 465 akses dari pengunjung lokal akakom via LAN dengan alamat IP 172.17 yang berarti akses tersebut berasal dari area gedung timur. Berikutnya terdapat sejumlah 40 akses pengunjung via LAN dengan alamat IP 172.18 yang berasal dari laboratorium, dan 501 akses situs dari alamat IP 10.10, yaitu akses lokal akakom yang berasal dari jalur *hotspot*. Dengan demikian terdapat total jumlah pengakses sebanyak 1.006 kali, atau sekitar 1% dari 93.286 akses yang berasal dari alamat IP lokal akakom.

Analisis selanjutnya dilakukan dengan perintah SQL

```
SELECT * FROM `log_modi` WHERE `browser` LIKE '%Windows%'
```

yang hasilnya adalah sebagai berikut.

Tabel 5.7 Daftar Sistem Operasi atau Browser yang digunakan

Sistem operasi pengguna	jumlah
windows	2.411
Linux	194
Android	1.634
Go-http-client	2
Googlebot-Image	26
MJ12bot	85.680
Yandex	432
bingbot	270
AhrefsBot	244
Macintosh	231
Baiduspider	48
iPhone	70

Hasil pengolahan tabel tersebut menunjukkan Sistem Operasi atau *browser* apa yang paling banyak digunakan oleh pengguna untuk mengakses situs. Tercatat sebanyak 2.411 kali akses yang berasal dari pengunjung dengan sistem operasi Windows, 1.634 kali akses dengan sistem operasi Android, 194 kali akses yang berasal dari peranti dengan sistem operasi Linux, dan sebanyak 70 kali akses berasal dari peranti bersistem operasi iPhone. Log akses yang tercatat dengan Go-http-client, Googlebot-Image, MJ12bot, YandexBot, bingbot, AhrefsBot, dan Baiduspider adalah lebih dikenal sebagai Robot/spider/*webcrawler*/mesin pencari yang bekerja secara otomatis mengindex halaman situs. Whois Lookup — Domain Names Search, Registration and Availability

BAB VI PENUTUP

6.1 Kesimpulan

Setelah analisis dan pembahasan dilakukan dalam bab sebelumnya, maka dapat dibuat simpulan untuk penelitian ini sebagai penyelesaiannya. Beberapa simpulan yang dibuat adalah

1. Dalam rentang waktu 17/Nov/2017:04:03:36 +0700 sampai dengan 23/Jan/2018:02:19:24 +0700, secara keseluruhan tercatat sebanyak 93.286 kali akses yang menuju situs.
2. Halaman situs yang paling sering dikunjungi berdasarkan urutan jumlah aksesnya adalah 'Home', 'Berita', '*Digital Library*', 'Koleksi', 'Tautan', 'Layanan', 'Katalog Online', dan 'Kontak'
3. Berdasarkan IP address pengunjung dari 30 peringkat terbanyak, asal kunjungan ke situs adalah dari luar kampus,
4. Berdasarkan IP address pengunjung yang berasal dari lokal STMIK AKAKOM, terdapat sejumlah 465 akses dari area gedung Timur, 40 akses dari laboratorium, dan 501 akses dari jalur *hotspot*, dan
5. Sistem Operasi atau Browser yang paling banyak digunakan untuk mengakses halaman situs adalah Windows, Android, Linux, dan iPhone

6.2 Saran

Dengan simpulan yang telah dibuat maka dapat diberikan saran sebagai bahan pertimbangan untuk pengembangan penelitian maupun manajemen dari Perpustakaan beserta bagian yang terkait.

1. Untuk penelitian selanjutnya mengenai analisis *log access* ini dapat dilakukan pengidentifikasian lokasi/asal dari mana Alamat IP pengakses berada,
2. Disarankan juga agar analisis tersebut dapat mengidentifikasikan provider apa pemilik dari Alamat IP yang berkunjung itu, dan
3. Lebih lanjut dapat dilakukan pencermatan lebih detail tentang browser yang dipakai pengguna.

DAFTAR PUSTAKA

- Achour, M., dkk, 2017, *PHP Manual documentation*, <http://php.net/manual/en/>
- Angelia, Putu Laras, 2015, *Pemanfaatan Senayan Library Information Management System (SLIMS) Di Perpustakaan Pusat Universitas Warmadewa*, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Udayana, Jimbaran.
- Departement Pendidikan Nasional RI, 2004, *Buku Pedoman Perpustakaan Perguruan Tinggi*, Departemen Pendidikan Nasional RI, Jakarta.
- Dokumentasi Nginx, 2018, *Module ngx_http_log_module*, https://nginx.org/en/docs/http/ngx_http_log_module.html
- Hermanto, A., Sartono, Sandra E., __, *Pemanfaatan Web Usage Mining pada e-commerce untuk Peningkatan Penjualan pada Retailer Online*, Universitas 17 Agustus Surabaya 1945, Surabaya.
- Indrayani, E., 2011, *Pengelolaan Sistem Informasi Akademik Perguruan Tinggi Berbasis Teknologi Informasi dan Komunikasi (TIK)*, Jurnal Penelitian Pendidikan, vol. 12 no.1, Universitas Pendidikan Indonesia, Bandung.
- Nginx Foundation, 2018, *nginx*, <https://nginx.org/en/>
- Suratno, T., Prahasto, T., Rochim, A.F., 2011, *Web Usage Mining, Pattern Discovery dan Log File*, Jurnal Sistem Informasi Bisnis 02, Universitas Diponegoro, Semarang.
- Vrána, J., 2017, *Adminer, Database management in a single PHP file*, <https://www.adminer.org/en/>