

ANALISIS WEB SERVER LOG DALAM PENCARIAN POLA PENGUNJUNG WEB DENGAN TEKNIK ASSOCIATION RULES

Muhammad Dahria^{#1}, Muhammad Syahril^{#2}

^{#1,2}Program Studi Sistem Informasi, STMIK Triguna Dharma

Jl. A.H. Nasution No. 73 F-Medan

E-mail: ^{#1}m.dahria@gmail.com

Abstrak

Sebuah Web Server Log akan menyimpan seluruh aktifitas pengunjung web site dalam sebuah server. Banyaknya interaksi yang dilakukan oleh pengunjung web, akan meningkatkan ukuran rekaman data log yang tersimpan dalam sebuah web, akan meningkatkan ukuran rekaman data log yang tersimpan dalam sebuah web server, sehingga pembengkakan volume data tidak dapat dihindari. Sayangnya, kebanyakan web administrator tidak mengetahui bagaimana manfaat data log yang tersimpan tersebut agar menggali informasi yang bermanfaat untuk pengembangan sebuah situs web. Data mining merupakan suatu cara untuk menambang data sehingga informasi yang sebelumnya tidak diketahui dapat diperoleh. Informasi yang dimasukkan difokuskan pada penemuan pola kunjungan web server log dengan menggunakan teknik Association Rules dan pemanfaatan Algoritma Apriori. Analisis web server log yang dilakukan mampu memberikan informasi mengenai pola kunjungan dan menemukan beberapa aturan asosiasi yang berguna untuk peningkatan efektifitas sebuah web site dimasa yang akan datang, serta memberikan layanan informasi dan komunikasi terbaik bagi publik.

Kata Kunci: Web Server Log, Association Rules, Algoritma Apriori

Abstract

A Web Server Log will store all over the web site visitor activity in a server. Number of interactions that are performed by web visitors, will increase the size of the log data records stored in a web, will increase the size of the log data records stored in a web server, thus swelling the volume of data is inevitable. Unfortunately, most web administrators do not know how the benefits of log data stored in order to dig up information that is beneficial to the development of a web site. Data mining is a way to mine the data so that the information previously unknown can be obtained. The information included is focused on the discovery of a pattern of visits the web server logs by using the technique of Association Rules and exploiting a priori Algorithm. Web server log analysis is conducted to provide information on the pattern of visits and found some useful Association rules for the enhancement of the effectiveness of a web site in the future, as well as providing the best information and communication service to the public.

Keywords: Web Server Logs, Association Rules, Apriori Algorithm

PENDAHULUAN

Internet telah dimanfaatkan oleh berbagai perusahaan atau institusi untuk menyediakan sumber daya informasi maupun layanan melalui situs *Web*. Pengunjung suatu situs *web* akan berinteraksi melalui serangkaian permintaan (*request*) yang bertalian dan berhubungan selama melakukan suatu kunjungan tunggal yang sering disebut dengan istilah *session*. Interaksi ini dilakukan untuk mendapatkan informasi maupun layanan yang diinginkan oleh pengunjung situs *web*.

Pengunjung situs *web* dalam sebuah *session* dapat berinteraksi dengan berbagai fungsi yang telah disediakan. Interaksi yang dilakukan oleh pengunjung yang satu dengan pengunjung yang lainnya mungkin berbeda baik dalam urutan rangkaian permintaannya maupun frekuensi kunjungannya. Pengunjung yang sama dalam suatu *session* yang berbeda mungkin melakukan rangkaian permintaan yang berbeda. Tingkah laku atau interaksi pengunjung situs *Web* dalam sebuah *session* dapat pula dimodelkan dengan suatu *graph*.

Demikian halnya dengan layanan informasi yang disediakan dalam situs resmi Pemerintah Kota Medan (<http://www.pemkomedan.go.id>). Seluruh aktifitas yang dilakukan oleh *user* pada sebuah *website* ini akan tersimpan dalam *web server log*. Semakin banyak kunjungan yang dilakukan pada sebuah *web*, semakin besar pula data yang terekam dalam *web server log*. Ukuran data yang tersimpan dalam *web server log* tidak hanya dalam ukuran megabyte, tetapi dapat juga *terabyte* atau bahkan sampai *petabyte*. Karena jumlah data yang ada pada data *web server log* cukup besar, maka diperlukan analisis terhadap data tersebut, sehingga informasi yang tersembunyi dibalik

data *web server log* dapat digali. Analisis yang dilakukan terhadap *web server log* diharapkan akan dapat memberikan informasi untuk peningkatan efektifitas sebuah *website*.

Analisis ini tentunya juga bermanfaat bagi pengelola *web* bilamana pada waktu tertentu diperlukan adanya pengembangan maupun perancangan ulang sebuah layanan informasi seperti halnya *website* milik Pemerintah Kota Medan. Salah satu cara yang dapat dipakai untuk melakukan analisis, lebih dari sekedar analisis statistik, adalah menggunakan *webmining*. *Webmining* menerapkan teknik dalam *data mining*. *Data mining* merupakan suatu cara untuk 'menambang' data sehingga informasi yang sebelumnya tidak diketahui, akan diperoleh. Karena *webmining* menggunakan teknik data mining, maka teknik yang dapat digunakan adalah *association rule*, *sequential pattern*, *clustering*, *decision tree*, *classification* dan masih banyak lagi.

Begitu juga dengan pencarian pola tingkah laku pengunjung situs *Web* yang dapat digunakan untuk mengetahui bagaimana pola kunjungan untuk masing-masing pengunjung dalam sebuah *session*. Informasi pola kunjungan ini dapat digunakan untuk mengetahui fungsi-fungsi mana saja yang sering dilakukan ataupun yang jarang dilakukan oleh pengunjung. Hal ini terkait dengan bagaimana pengaturan distribusi beban pada sumber daya yang ada sehingga layanan yang disediakan dalam situs *Web* dapat optimal.

LANDASN TEORI

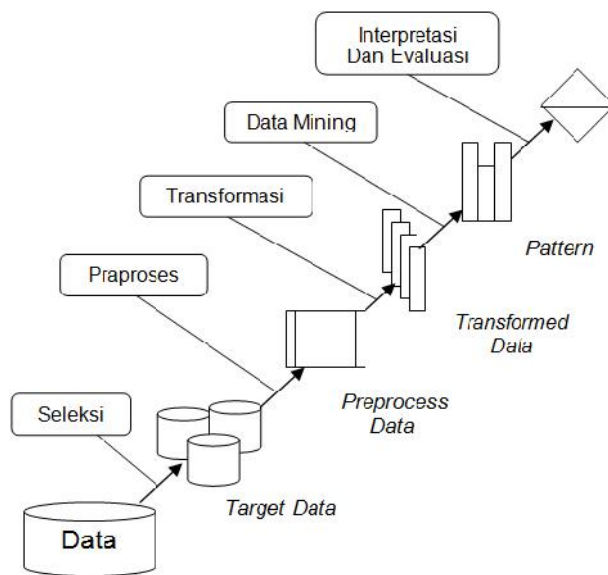
1. Knowledge Discovery in Databases

Knowledge Discovery yang sering disebut *Data Mining* secara sederhana dapat dikatakan sebagai proses mengekstrak atau menggali pengetahuan atau informasi yang

berharga (*interesting knowledge*) dari sejumlah besar data, baik yang disimpan di dalam *database*, *datawarehouse* maupun media penyimpanan informasi lainnya (Han, Kamber, 2001). Dari definisi tersebut didapatkan berbagai istilah yang beredar dan memiliki arti hampir sama dengan *Data Mining* yaitu *knowledge mining from databases*, *knowledge extraction*, *data / pattern analysis*, *data archeology*, *data dredging*, *information harvesting* dan *business intelligence*.

2. Data Mining

Banyak orang menyamakan istilah *Data Mining* dengan *Knowledge Discovery in Databases (KDD)*. Sebenarnya *Data Mining* merupakan salah satu tahap yang terdapat di dalam *knowledge discovery*. Proses *Data Mining* seperti yang digambarkan pada gambar 1



Gambar 1. Proses Data Mining

3. Web Mining

Web Mining merupakan penerapan teknik data mining terhadap *web* dengan tujuan untuk memperoleh pengetahuan dan informasi lebih dari dalam *web*.

Web mining dapat dikategorikan ke dalam tiga ruang lingkup yang berbeda, yaitu *web content mining*, *web structure mining* dan *web usage mining* (Srivastava et al., 2000).

Sedangkan menurut Colley et al. (1997b) dalam jurnalnya yang berjudul "*Web Mining : Information and Pattern Discovery on World Wide Web Browsing*" mendefinisikan *web mining* secara umum sebagai pencarian dan analisis informasi yang berasal dari *World Wide Web*. *Web mining* secara garis besar dibagi menjadi dua macam, yaitu *web content mining*, yang merupakan proses menambang isi dari sumber online seperti *web*, dan *web usage mining*, yang mencari informasi dari penggunaan sumber online seperti *web*.

4. Web Server

Web Server adalah software server yang menjadi tulang belakang dari *World Wide Web (WWW)*. *Web server* menunggu permintaan dari client yang menggunakan browser seperti *netscape navigator*, *Internet Explorer*, *mozilla*, dan program browser lainnya. Jika ada permintaan dari browser, maka *Web Server* akan memproses permintaan itu dan kemudian memberikan hasil prosesnya berupa data yang diinginkan kembali ke browser. Data ini mempunyai format yang standar disebut dengan format *SGML (Standard General Markup Language)*. Data yang berupa format ini kemudian akan ditampilkan oleh browser sesuai dengan kemampuan browser itu.

Web Server, untuk berkomunikasi dengan clientnya (*web browser*) mempunyai protokol sendiri yaitu *HTTP (Hyper Text Transfer Protocol)*. Dengan protokol ini, komunikasi antar *Web Server* dengan (*browser*) dapat saling dimengerti dan lebih mudah. Seperti telah dijelaskan di atas, Standar format data pada *World Wide*

Web adalah SGML. Tapi sudah menjadi hal yang umum bahwa para pengguna internet lebih banyak menggunakan format HTML (*Hyper Text Markup Language*) karena penggunaannya yang lebih sederhana dan mudah dipelajari. Kata *Hyper Text* mempunyai arti bahwa seorang pengguna internet dengan *web* browsernya dapat membuka dan membaca dokumen-dokumen yang ada dalam komputernya atau bahkan komputer yang jauh tempatnya sekalipun. Hal ini memberikan cita rasa dari suatu proses yang tridimensional, artinya pengguna internet dapat membaca dari satu dokumen ke dokumen yang lain hanya dengan mengklik beberapa bagian dari halaman-halaman dokumen (*web*) itu.

Proses yang dimulai dari permintaan *webclient* (browser), diterima *web server*, diproses, dan dikembalikan hasil prosesnya oleh *web server* ke *webclient* lagi dilakukan secara transparan. Setiap orang dapat dengan mudah mengetahui apa yang terjadi pada tiap-tiap proses. Secara garis besarnya *Web Server* hanya memproses semua masukan yang diperolehnya dari *web client*nya.

5. Web Server Log

Seluruh aktifitas yang dilakukan oleh user pada sebuah *website* ini akan tersimpan dengan sendirinya dalam *web server log*. Semakin banyak kunjungan yang dilakukan pada sebuah *web*, semakin besar pula data yang terekam dalam *web server log*. Ukuran data yang tersimpan dalam *web server log* tidak hanya dalam ukuran megabyte, tetapi dapat juga terabyte atau bahkan sampai petabyte. Karena jumlah data yang besar dan pentingnya data *web server log*, maka diperlukan analisis terhadap data tersebut, sehingga informasi yang tersembunyi dibalik data *web server log* dapat digali. Analisis yang

dilakukan terhadap *web server log* diharapkan akan dapat memberikan informasi untuk peningkatan efektifitas sebuah *website*.

6. Association Rules

Association rules merupakan salah satu teknik *data mining* yang berfungsi untuk menemukan asosiasi antar variabel, korelasi atau suatu struktur diantara *item* atau objek-objek didalam *database* transaksi, *database* relasional, maupun pada penyimpanan informasi lainnya. *Association rules Discovery* seringkali digunakan dalam *Market Basket Analysis* dan juga sebagai *Affinity Grouping*. *Association rules* mining digunakan untuk mencari pola-pola yang sering muncul, asosiasi-asosiasi, korelasi atau hubungan sebab akibat diantara himpunan dari *item-item* atau objek-objek dalam *database*.

Fungsi ini paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis. Tipe *association rule* bisa dinyatakan sebagai misal : "70% dari orang-orang yang membeli mie, jus dan saus akan membeli juga roti tawar".

Aturan asosiasi mengcapture *item* atau kejadian dalam data berukuran besar yang berisi data transaksi. Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "basket data." Aturan asosiasi yang didefinisikan pada basket data, digunakan untuk keperluan promosi, desain katalog, segmentasi customer dan target pemasaran.

Secara tradisional, aturan asosiasi digunakan untuk menemukan trend bisnis dengan menganalisa transaksi *customer*. Dan dapat digunakan secara efektif pada bidang *Web Mining* yang diilustrasikan sebagai berikut : pada *Webaccess log*, kita

menemukan bahwa aturan asosiasi : "A and B implies C," memiliki nilai *confidence* 80%, dimana A, B, dan C adalah halaman *Web* yang bisa diakses. Jika seorang user mengunjungi halaman A dan B, maka terdapat 80% kemungkinan dia akan mengunjungi halaman C juga pada session yang sama, sehingga halaman C perlu diberi *direct link* dari A atau B. Informasi ini dapat digunakan untuk membuat link secara dinamik ke halaman C dari halaman A atau B sehingga user dapat melakukan *direct link* ke halaman C. Informasi semacam ini digunakan untuk melakukan link ke halaman produk yang berbeda secara dinamik berdasarkan interaksi *web visitors*.

PEMBAHASAN

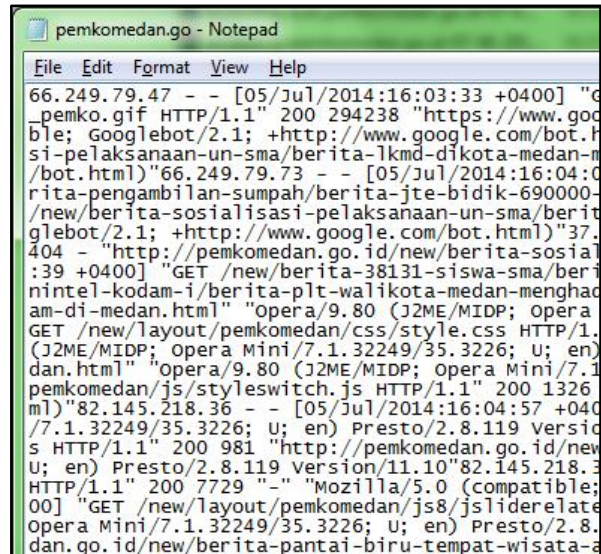
1. Penyedia Data Sumber

Format *web server log* yang digunakan dalam penelitian adalah *ECLF (Extended Common Log Format)*. Sumber data yang diambil sebagai bahan penelitian berasal dari tools *raw access log* yang terdapat dalam *web server log* disajikan dalam format *text* merupakan data mentah, yang siap untuk di olah dalam proses data mining untuk mengeksplorasi informasi yang menarik, maupun *knowledge* yang bermanfaat bagi pihak yang membutuhkannya.

Tabel 1. Nama dan Isi Elemen Data Log

Nama Elemen	Isi Elemen
Host	124.195.18.244
Indent	-
Authuser	-
Time	[05/Jul/2014:00:55:36 +0000]
Request	"GET /images/iconpdf.gif HTTP/1.0"
Status	200
Bytes	253
Refferer	http://www.pemkomedan.go.id/database.php
User Agent	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9) Gecko/2008052906 Firefox

Adapun arti dari setiap isi elemen adalah Elemen-elemen yang telah diterangkan pada Tabel 1 merupakan bagian keterangan dari sebuah pembacaan log yang menjadi target penelitian. Secara lengkap bentuk data mentah yang akan dikupas memiliki format bertipe teks (Log*.txt) seperti yang terlihat pada gambar2 berikut ini.



Gambar 2. Format Data Log

2. Pra Processing (Cleaning) Data

Di dalam *Knowledge Discovery Database(KDD)* diperlukan adanya pemrosesan pembersihan data yang merupakan operasi dasar seperti meliputi penghapusan *noise* yang dilakukan. Sebelum proses *data mining* dapat dilaksanakan perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data seperti kesalahan cetak (tipografi). Pada proses persiapan ini data asli dari web server log disaring sehingga hanya data yang valid saja yang akan diolah. Proses persiapan yang dilakukan adalah :

1. Menghapus seluruh *request* dari data-data dengan jenis file grafik, yaitu file-file request yang berekstensi gif, jpeg, jpg, bmp, ico, swf dan format gambar lainnya. Juga menghapus *request* file *sound*, *video* dan *file* sejenis lainnya.
2. Menghapus seluruh *request* yang *error*. Untuk mengetahui apakah sebuah *request* berstatus *error* atau tidak, dapat diisi dari
3. elemen status dan kata awal dari isi elemen *request*. Kode status menunjukkan respon dari *server* terhadap permintaan user. Jika kode statusnya bernilai 200 berarti bahwa *request* adalah valid, sedangkan kode status lainnya seperti 404 berarti bahwa dokumen yang diminta tidak ada di dalam website. Kata awal dari elemen *request* menunjukkan permintaan *user* kepada *server*. Jika diawali dengan kata *GET* berarti *user* meminta sebuah dokumen untuk ditampilkan. Sedangkan jika kodenya adalah head berarti *user* hanya menginginkan informasi dari dokumen, tetapi tetapi bukan dokumennya itu sendiri. Arti kode status dan kata awal dari *request* merupakan ketentuan dari protokol *HTTP (Hyper Text Transfer Protocol)*

3. Transformasi Data

Data log di atas akan dikonversikan dalam bentuk database yang telah memiliki baris kolom, hasil dari pengelompokan data ini seperti yang diperlihatkan pada tabel 2.

Tabel 2. Data Akses Log Setelah di Konversi

Host	Time	Request	Status	Bytes	Referrer	UserAgent
124.195.18.244	05/Jul/2014:00:55:30	GET /images/dalam1bckagustus.jpg HTTP/1.0	200	3510	http://www.pernikomedan.go.id/layang_profil.php	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9) (Gecko/2008052906 Firefox
124.195.18.244	05/Jul/2014:00:55:30	GET /images/icon_aas.gif HTTP/1.0	200	150	http://www.pernikomedan.go.id/prc/il.pdf	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9) (Gecko/2008052906 Firefox
124.195.18.244	05/Jul/2014:00:55:30	GET /images/gas.gif HTTP/1.0	200	52	http://www.pernikomedan.go.id/layang_profil.php	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9) (Gecko/2008052906 Firefox

Data *log* yang ada dapat melewati proses *enrichment* yaitu “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD seperti data atau informasi eksternal.

Adalah perlu memanfaatkan data sumber untuk keperluan pengayaan sumber data baru yang dapat memberikan informasi tambahan kepada para pihak yang membutuhkan. Untuk itu data sumber yang sudah disiapkan akan dikelola lagi untuk menghasilkan informasi berkenaan dengan :

- 1) Berapa jumlah pengunjung dalam sehari ?
- 2) Bagaimana respon server dalam menangani request dari pengunjung ?
- 3) Berapa banyak halaman yang dikunjungi untuk seorang pengunjung dalam waktu yang sama ?
- 4) Berapa banyak pengunjung yang membuka satu halaman tertentu pada waktu tertentu ?
- 5) Bagaimana pola kunjungan pengunjung pada waktu tertentu ?

Informasi diatas dapat diperoleh dengan melakukan filter dan pengelolaan data yang berasal dari data log sumber. Data Pengayaan Web Log terlampir pada bagian akhir penelitian ini.

4. Penemuan Pola Kunjungan

Karena data yang diambil adalah data untuk pencarian halaman halaman yang dikunjungi oleh alamat *ip/host*, maka data yang diambil dibatasi pada :

- a. Alamat Host
- b. Date-Time
- c. Halaman Tujuan

Tabel 3. Contoh Data yang akan Dianalisis

Host	Date/Time	Halaman Tujuan
24.81.66.130	[05/Jul/2014:00:55:36	http://www.pemkomedan.go.id/produk_hukum.php
24.81.66.130	[05/Jul/2014:00:55:38	http://www.pemkomedan.go.id/news_detail.php
24.81.66.130	[05/Jul/2014:00:55:56	http://www.pemkomedan.go.id/bukutamu.php
24.81.66.130	[05/Jul/2014:00:55:59	http://www.pemkomedan.go.id/info_detail.php

5. Analisa Data dan Inisialisasi

Untuk menggambarkan keberadaan sebagian besar halaman yang akan diteliti, maka dibawah ini akan ditampilkan daftar nama file yang ada di dalam sebuah web site server.

Tabel 4. Contoh Data yang Dianalisis

No.	LEVEL 2 - Hal Permintaan	LEVEL 3 - Halaman Tujuan	INISIAL
...
2	Sejarah Kota Medan	selayang_sejarah.php	A1
	Informasi Umum	selayang_informasi.php	A2
	Kependudukan	selayang_kependudukan.php	A3
	Potensi Daerah	selayang_potensi.php	A4
	Profil Pejabat Daerah	selayang_profil.php	A5
	Lambang Daerah	selayang_lambang.php	A6
	Hari Jadi Kota Medan	selayang_hari.php	A7

Kebutuhan data untuk mencari asosiasi pada suatu kunjungan situs web, dibatasi pada Field Host dan Refferer. Sehingga data yang telah dikumpulkan mempunyai format sebagai berikut

Tabel 5. Hasil Pola Kunjungan

Host	Nama Halaman	Nama File	Inisial File
24.81.66.130	Produk Hukum	produk_hukum.php	E1
24.81.66.130	Berita Pemko	news_detail.php	ANS1
24.81.66.130	Buku Tamu	bukutamu.php	P3
24.81.66.130	Info Pemko	Info_detail.php	AIN1

Proses menghitung Assosiation Rules dengan menggunakan algoritma apriori yang akan dilakuka nanti akan lebih mudah jika proses inisialisasi telah dilakukan untuk semua pola kunjungan setiap alamat host yang melakukan interaksi dengan situs web . Adapun bentuk yang akan digunakan untuk perhitungan algoritma tersebut terhadap setiap host seperti yang digambarkan dalam tabel 6.

Tabel 6. Hasil Pola Kunjungan Tujuan

Host	Inisial Halaman	Pola Kunjungan
24.81.66.130	E1,ANS1,P3,AIN1	Produk Hukum - Berita Pemko - Buku Tamu - Info Pemko

6. Association Rules yang Ditemukan

Berdasarkan data yang telah dianalisis, maka hasil sementara pola kunjungan untuk data satu hari pada tanggal 5 Juli 2014 adalah seperti pada tabel 7.

Tabel 7. Pola Kunjungan Web

No.	ID	Items
1	124.195.18.244	FB7,E1,FA6
2	24.81.66.130	E1,ANS1,P3,AIN1
3	125.162.3.211	AIN1,C4,A5,C3
4	202.134.2.83	ANS1,P3
5	202.55.170.34	A5,ANS1,E1
6	222.124.59.80	AIN1,C4,ANS1,C3,FA6
7	74.6.22.182	FB7,E1,P3
8	61.247.8.66	A5,C4
9	74.208.16.198	ANS1,B2,A5
10	202.59.166.171	C3,C4,P3
11	43.244.175.199	P3,ANS1
12	125.162.40.39	E1,FA6,C3
13	61.247.8.86	C4,B2,P3,A5
14	202.152.175.182	ANS1
15	202.152.170.240	P3,E1,B2

7. Algoritma Apriori

Algoritma Apriori digunakan untuk menghasilkan *candidate itemset* yang menyebabkan tidak seluruh *itemset* diolah pada proses selanjutnya, melainkan hanya yang memenuhi syarat saja yang akan diproses pada tahap berikutnya.

Berdasarkan data yang telah diproses terdapat jumlah item tunggal yang akan diteliti lebih lanjut. Adapun jumlah item

tunggal yang dimaksud adalah seperti yang ditunjukkan oleh Tabel 1.

Tabel 8. Daftar Index dan Jumlah Item

No.	Nama Halaman	Inisial
1	Profil Pejabat Daerah	A5
2	Agenda Kegiatan	AIN1
3	Berita Kota	ANS1
4	Struktur Organisasi Pemerintah Kota	B2
5	Profil Anggota Dewan	C3
6	Daftar Dinas Daerah	C4
7	Produk Hukum	H1
8	Prosedur Pengurusan Akte Kelahiran	FA6
9	Prosedur Pengurusan Izin Mendirikan Bangunan	FB7
10	Komentar	P3
Total Indeks Item		10

Dari hasil pengujian dengan menggunakan tools program, telah ditemukan sebanyak 10 besar item yang sering sekali dikunjungi oleh pengunjung web. Dan jumlah IP pengunjung setelah diurut berdasarkan banyaknya transaksi yang dilakukan berjumlah 17 pengunjung dengan alamat IP Address yang berbeda-beda.

8. Penentuan Frequent Item Set

Adapun pola pencarian *FIS (Frequent Item Set)* dilakukan melalui beberapa pengujian dan penghitungan dengan urutan-urutan tertentu seperti yang diuraikan di bagian di bawah ini.

a. Penentuan Nilai Support Kunjungan

Support dari sebuah pola asosiasi mengacu pada prosentase jumlah transaksi-transaksi data yang relevan untuk pola tertentu. Sintak Umum pencarian support sebagai berikut :

$$Support_{A \rightarrow B} = \frac{\#_tuples_containing_both_A_and_B}{total_#_of_tuples}$$

Untuk permasalahan yang sedang dibahas maka perhitungannya adalah sebagai berikut :

Tabel 9. Database Pola Kunjungan :

No.	ID	Items
1	124.195.18.244	FB7,E1,FA6
2	24.81.66.130	E1,ANS1,P3,AIN1
3	125.162.3.211	AIN1,C4,A5,C3
4	202.134.2.83	ANS1,P3
5	202.55.170.34	A5, ANS1,I1
6	222.124.59.80	AIN1,C4,ANS1,C3,FA6
7	74.6.22.182	FB7,E1,P3
8	61.247.8.66	A5,C4
9	74.208.16.198	ANS1,B2,A5
10	202.59.166.171	C3,C4,P3
11	43.244.175.199	P3,ANS1
12	125.162.40.39	E1,FA6,C3
13	61.247.8.86	C4,B2,P3,A5
14	202.152.175.182	ANS1
15	202.152.170.210	P3,E1,B2
16	74.208.16.19	FA6,E1,C4,A5
17	91.203.96.17	P3,ANS1

Untuk mencari Nilai Support :

$$Support \{A5\} = \frac{\text{Banyaknya Kunjungan Ke } \{A5\}}{\text{Total Jumlah Transaksi Host}}$$

$$Support \{A5\} = \frac{6}{17} * 100 \%$$

$$Support \{A5\} = 35,29\%$$

Hasil perhitungan persentase untuk tiap-tiap item kunjungan adalah seperti yang dipaparkan pada tabel 10.

Tabel 10. Hasil Support Item Kunjungan

No.	Item	Jumlah	Total Item	Support
1	{A5}	6	17	35,29%
2	{AIN1}	3	17	17,64%
3	{ANS1}	8	17	47,05%
4	{B2}	3	17	17,65%
5	{C3}	4	17	23,53%
6	{C4}	6	17	35,29%
7	{E1}	7	17	41,17%
8	{FA6}	4	17	23,53%
9	{FB7}	2	17	11,76%
10	{P3}	8	17	47,05%

b. Penentuan Nilai Minumun Support

Penentuan nilai Minimum Support sebuah kunjungan adalah bergantung pada kebutuhan user. Artinya setiap user bebas untuk menentukan berapa nilai minimum support yang layak untuk dimasukkan sebagai ketentuan pengisian sebuah support. Sebagai pengujian awal penulis akan memasukkan nilai 20% untuk masing-masing *minimum support* dan *minimum confidence*.

c. Data Analisis Association Rule

Untuk mencari hubungan asosiasi dari tiap-tiap item kunjungan. Jika kita akan menguji seorang Host akan mengunjungi halaman BERITA (ANS1), maka ia juga akan mengunjungi halaman KOMENTAR (P3), maka formulasi untuk mencari asosiasi antara keduanya adalah :

- *Support*ANS1 => P3 adalah :

$$= \frac{\text{Banyaknya Kemunculan Pasangan ANS1 dan P3}}{\text{Total Jumlah Transaksi Host}}$$

$$= \frac{4}{17} = 23,53\%$$

Ketentuan berlaku umum :

$$\boxed{\text{Support ANS1} \Rightarrow \text{P3}} = \boxed{\text{Support P3} \Rightarrow \text{ANS1}}$$

$$= \frac{\text{Support Pasangan ANS1} \Rightarrow \text{P3}}{\text{Support ANS1}}$$

$$= \frac{23,53\%}{47,05\%} = 50\%$$

Ketentuan :

$$\boxed{\text{Confidance ANS1} \Rightarrow \text{P3}} \neq \boxed{\text{Confidance P3} \Rightarrow \text{ANS1}}$$

- Sehingga *Confidence* P3 => ANS1 :

$$= \frac{\text{Support Pasangan ANS1} \Rightarrow \text{P3}}{\text{P3}}$$

$$= \frac{23,53\%}{47,05\%} = 50\%$$

Demikian proses perhitungan Support dan Confidence antar item set yang lain.

9. Frequent Item Set yang Ditemukan

Penentuan Frequent Item Set (FIS) bergantung pada nilai *minimum support* dan *minimum confidence* yang dimasukkan sejak awal perhitungan ini, yaitu :

Minimum Support : 20%

Minimum Confidence : 20%

sehingga sebuah item akan dikatakan interest jika nilai Support nya >= Nilai Minimum Support. Hasil penentuan ini seperti yang dihasilkan pada Tabel 11.

Tabel 11. Penentuan Interest dan Uninterest

No.	Item	Jumlah	Total Item	Support	Interest Uninterest	Ket
1	{A5}	6	17	35,29%	Interest	✓
2	{AIN1}	3	17	17,64%	Uninterest	x
3	{ANS1}	8	17	47,05%	Interest	✓
4	{B2}	3	17	17,65%	Uninterest	x
5	{C3}	4	17	23,53%	Interest	✓
6	{C4}	6	17	35,29%	Interest	✓
7	{E1}	7	17	41,17%	Interest	✓
8	{FA6}	4	17	23,53%	Interest	✓
9	{FB7}	2	17	11,76%	Uninterest	x
10	{P3}	8	17	47,05%	Interest	✓

Data yang memenuhi syarat akan dipisahkan untuk segera diproses kembali untuk penentuan pencarian FIS lebih lanjut. Suatu subset dari FIS harus juga Frequent atau Interest.

Tabel 12. Penentuan Data Interest

No.	Item	Jumlah	Total Item	Support	Interest Uninterest
1	{A5}	6	17	35,29%	Interest
2	{ANS1}	8	17	47,05%	Interest
3	{C3}	4	17	23,53%	Interest
4	{C4}	6	17	35,29%	Interest
5	{E1}	7	17	41,17%	Interest
6	{TA6}	4	17	23,53%	Interest
7	{P3}	8	17	47,05%	Interest

Berdasarkan data di atas, maka ditemukan daftar Interest Association Rules hanya sampai pada 2 item set. Data yang dihasilkan pada sampel penelitian yang telah dibahas adalah :

Tabel 13. Tabel Interest Association Rules

2 Item Set Data	Bobot	Hubungan Asosiasi	Ket	Eliminasi Frequent Item Set
{A5,C4}	= 4/17	= 23.529%	Interest	✓
{ANS1,P3}	= 4/17	= 23.529%	Interest	✓

Jika kita telah menguji asosiasi kunjungan halaman BERITA (ANS1), terhadap kunjungan halaman KOMENTAR (P3), maka lebih lanjut kita akan membahas bagaimana hubungan asosiasi antara halaman DINAS DAERAH (C4) dengan halaman PROFIL PEJABAT DAERAH (A5), maka untuk menghitung *Minimum Support* dan *Minimum Confidence* dari dua asosiasi diatas dapat diuraikan dengan paparan di bawah ini :

Rules {C4}=> {A5}:

Diketahui :

$$Support A5 = \frac{6}{17} * 100\% = 35,29\%$$

$$Support C4 = \frac{6}{17} * 100\% = 35,29\%$$

Support {A5}=> {C4} adalah :

$$= \frac{\text{Banyaknya Kemunculan Pasangan A5 dan C4}}{\text{Total Jumlah Transaksi}}$$

$$= \frac{3}{17} = 17,64\%$$

Ketentuan umum :

$$\boxed{\text{Support A5} \Rightarrow \text{C4}} = \boxed{\text{Support C4} \Rightarrow \text{A5}}$$

Confidence {C4}=>{A5} adalah :

$$= \frac{\text{Support Pasangan A5} \Rightarrow \text{C4}}{\text{Support C4}}$$

$$= \frac{17,64\%}{35,29\%} = 50\%$$

Confidence {A5}=> {C4} adalah :

$$= \frac{\text{Support Pasangan A5} \Rightarrow \text{C4}}{\text{Support A5}}$$

$$= \frac{17,64\%}{35,29\%} = 50\%$$

Ketentuan umum :

$$\boxed{\text{Confidence A5} \Rightarrow \text{C4}} \neq \boxed{\text{Confidence C4} \Rightarrow \text{A5}}$$

10. Interesting Rules yang Ditemukan

Berdasarkan perhitungan dan analisis yang telah dilakukan terhadap data sampel yang ada maka ditemukanlah beberapa aturan yang dianggap diterima, yaitu :

A5->C4 [23.53%,66.67%]	<FIS Association Rules>
C4->A5 [23.53%,66.67%]	<FIS Association Rules>
ANS1->P3 [23.53%,50.00%]	<FIS Association Rules>
P3->ANS1 [23.53%,50.00%]	<FIS Association Rules>

11. Hasi Akhir

Dari daftar *Interest Association Rules* yang dihasilkan di atas, maka ditetapkanlah beberapa aturan sebagai berikut :

1. Aturan asosiasi antara halaman **Berita** yang diinisialisasikan dengan simbol **ANS1**, dengan halaman **Komentar** yang diinisialisasikan dengan simbol **P3**, memiliki dukungan data *Support* sebesar 23,53%. Dengan kekuatan hubungan diantara keduanya bernilai 50%. Artinya bahwa setiap kunjungan yang dilakukan oleh pengunjung untuk meminta halaman **Berita**, selalu diikuti dengan permintaan halaman halaman **Komentar** memiliki kekuatan hubungan 50%.
2. Berdasarkan data komposisi tiap menu pada halaman web, kedua halaman **Berita** dan **Komentar** selalu dikunjungi secara bersama dan disajikan pada kedalaman level yang sama pula, sehingga untuk asosiasi ini penyajian menu halaman website Pemerintah Kota Medan khususnya halaman **Berita** dan halaman **Komentar** sudah dapat dikatakan efektif.
3. Berdasarkan data komposisi tiap menu pada halaman web, kedua halaman Profil Pejabat Daerah dan halaman Dinas Daerah selalu dikunjungi secara bersama namun disajikan pada kedalaman level yang berbeda, sehingga untuk asosiasi ini penyajian menu halaman website Pemerintah Kota Medan khususnya

halaman **Profil Pejabat Daerah** dan halaman **Dinas Daerah** belum dapat dikatakan efektif, karena seorang pengunjung harus melewati level menu berbeda untuk mencari halaman yang memiliki aturan asosiasi yang kuat. Sehingga untuk keperluan pengembangan maupun perancangan ulang komposisi menu pada halaman web site Pemerintah Kota Medan dimasa yang akan datang perlu memposisikan kedua halaman tersebut kedalam level yang sama untuk memudahkan pengunjung mengakses informasi yang diinginkan. Adalah perlu juga mempertimbangkan aspek beban akses dan perancangan antar muka pada halaman-halaman web yang masuk dalam *interesting association rules* ini dengan tujuan agar setiap halaman dapat diakses lebih cepat dan dapat disajikan secara lebih menyenangkan.

SIMPULAN

Dengan mengacu kepada analisis *web server log* yang dilakukan dan implementasi system yang ada didapatkan beberapa kesimpulan, antara lain :

1. Data Mining mampu dijadikan sebagai alat bantu untuk proses analisis web server log dalam pencarian pola kunjungan website Pemerintah Kota Medan dengan menerapkan metode yang tepat yaitu metode *Association Rule* yang mampu menghasilkan aturan-aturan asosiasi yang menggambarkan pola kunjungan terhadap website Pemko Medan.
2. Penerapan teknik *Association Rules* menggunakan Algoritma *Apriori* yaitu dengan menentukan nilai *Support* (jumlah dukungan) dan *Confidence* (kekuatan hubungan) terhadap data kunjungan yang ada dalam *web server log*.

3. Bahwa telah ditemukan pola pengunjung halaman situs web pemerintah Kota Medan dengan pemanfaatan *Data Mining* dan teknik *Association Rules* dan berhasil menemukan asosiasi yang mengandung informasi berharga untuk dimanfaatkan dalam melakukan pengembangan atau restrukturisasi halaman web yang lebih baik dimasa yang akan datang.
4. Tidak semua data dari *web server log* merupakan data yang siap untuk digunakan dan sangat sedikit sekali dari data tersebut yang merupakan data yang 'berarti'.
5. Berdasarkan data komposisi tiap menu pada halaman web, kedua halaman **Berita** dan **Komentor** selalu dikunjungi secara bersama dan disajikan pada kedalaman level yang sama pula, sehingga untuk asosiasi ini penyajian menu halaman website Pemerintah Kota Medan khususnya halaman **Berita** dan halaman **Komentor** sudah dapat dikatakan efektif.
6. Berdasarkan data komposisi tiap menu pada halaman web, kedua halaman **Profil Pejabat Daerah** dan halaman **Dinas Daerah** selalu dikunjungi secara bersama namun disajikan pada kedalaman level yang berbeda, sehingga untuk asosiasi ini penyajian menu halaman website Pemerintah Kota Medan khususnya halaman **Profil Pejabat Daerah** dan halaman **Dinas Daerah** belum dapat dikatakan efektif, karena seorang pengunjung harus melewati level menu berbeda untuk mencari halaman yang memiliki aturan asosiasi yang kuat. Sehingga untuk keperluan pengembangan maupun perancangan ulang komposisi menu pada halaman web site Pemerintah Kota Medan dimasa yang akan datang perlu memposisikan kedua halaman tersebut kedalam level yang sama untuk

memudahkan pengunjung mengakses informasi yang diinginkan

DAFTAR PUSTAKA

- Jiawei Han and Micheline Kamber. 2000. *Data Mining: Concept and Techniques*, Intelligent Database System Research Lab School of Computing Science, Canada, Chap. 6.
- Ah-Hwee Tan. 1999. *Text Mining: Promises and Challenges*. Kent Ridge Digital Labs Singapore: Email, ahhwee@krdl.org.sg
- R. Cooley B. Mobahseeer, and J. Srivastava. 1998. *Web Mining : Information and Pattern Discovery on the World Wide Web*, Departemen of Computer Science and Engineering Univ. Of minnesota, 55455, USA
- R. Agrawal, R. Srikant. 2001. *Mining Sequential Patterns*. Proc. IEEE Int. Conf. On Data Engineering (ICDE '01), 1995
- Ramakharisnan Srikant, Yunghui Yang. 2001. *Mining Web Logs to Improve Website Organization*. Hongkong: Journal
- Yongjian Fu, mario Creado, Chunhua Ju. 1970. *Reorganizing Web Sites Based on User Acces Patterns*. USA: Journal, University of Missouri-Rolla Miner Circle.
- Gregoris S. Budhi, Ibnu Gunawan, Ferry Yuwono. 2007. *Algoritma Porter Stemmer For Bahasa Indonesia Untuk Pre-Processing Text Mining Berbasis Metode Market Basket Analysis*: Surabaya: Jurnal UK Petra jurusan Teknik Informatika.
- Arrummaisha Adrifina, Juwita UP, I Wayan SW. 2008. *Pemilihan Artikel Berita Dengan Text Mining*.