



# **Web Usage Mining dengan Google Analytics: Studi Kasus Situs Achmatim.Net**

**Disusun Oleh:**  
Achmad Solichin [0806444751]

**Tugas ini dibuat untuk matakuliah  
TEKNOLOGI SISTEM BASIS DATA  
MAGISTER TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS INDONESIA**

**JAKARTA  
SEMESTER GASAL 2008/2009**

**DAFTAR ISI**

|    |                                                          |    |
|----|----------------------------------------------------------|----|
| A. | Data mining.....                                         | 3  |
| B. | Web Mining .....                                         | 4  |
| C. | Google Analytics .....                                   | 8  |
| D. | Analisa Situs Achmatim.Net dengan Google Analytics ..... | 10 |

## Web Usage Mining dengan Google Analytics: Studi Kasus Situs Achmatim.Net

### A. Data mining

*Data mining*, bersama dengan terminologi data *warehouse*, OLTP, OLTP dan BI (*business intelligence*) merupakan topik yang akhir-akhir ini sering dibicarakan oleh kebanyakan orang, terutama di kaum intelektual dan akademisi. *Data mining* sendiri mulai dikenalkan sejak tahun 2000-an. Bahkan sebuah majalah teknologi online ZDNET *News* pada edisi bulan Februari 2001, memprediksikan bahwa *data mining* akan menjadi “*one of the most revolutionary developments of the next decade*”, salah satu perkembangan paling revolusioner untuk dekade berikutnya. Dan pada kenyataannya, prediksi tersebut terbukti saat ini.

Menurut David Hand, Heikki Mannila dan Padhraic Smyth dalam bukunya *Principles of Data mining* (MIT Press, Cambridge, MA, 2001), *data mining* merupakan proses analisa dari sekumpulan (terkadang sangat besar) data pengamatan untuk menemukan adanya hubungan-hubungan yang tidak terduga sebelumnya dan untuk merangkum data yang menjadi bentuk yang mudah dimengerti dan berguna bagi pemilik data. Dari pengertian tersebut dapat ditarik kesimpulan bahwa konsep *data mining* berhubungan dengan data dalam jumlah yang sangat besar. Tujuan dari *data mining* adalah berusaha mencari manfaat dari sekumpulan data tersebut.

Dilihat dari disiplin ilmu yang digunakan, *data mining* merupakan ilmu multi disiplin. *Data mining* menyangkut berbagai disiplin ilmu seperti *database*, kecerdasan buatan (*artificial intelligence*), *information science* (ilmu informasi), *high performance computing*, visualisasi, *machine learning*, statistik, *neural networks* (jaringan syaraf tiruan), pemodelan matematika, *information retrieval* dan *information extraction* serta pengenalan pola. Saat ini *data mining* juga berkembang menjadi berbagai konsep ilmu lain, seperti *web mining*.

## B. Web Mining

Seperti sudah kita pahami bersama bahwa dunia web merupakan dunia yang sangat luas. Internet adalah kumpulan data yang paling banyak di dunia ini dan secara eksponensial data tersebut terus bertambah selama internet masih digunakan. Luasnya jangkauan data yang tersedia di internet tersebut tentunya sangat potensial untuk digali, misalnya dimanfaatkan dalam meningkatkan penjualan (*web marketing*).

Menurut *wikipedia*, *web mining* merupakan suatu aplikasi bagian dari *data mining* yang berusaha menggali pola-pola yang tersedia di dalam web itu sendiri. Jadi antara *data mining* dan *web mining* hanya berbeda dalam hal target data yang dianalisa. *Data mining* umumnya menganalisa data yang berasal dari OLTP (*Online Transactional Process*) dan data transaksi lainnya. Sedangkan *web mining* target analisisnya adalah data dari web, seperti data akses pengunjung, struktur halaman web, format halaman web dan sebagainya.

Berdasarkan target analisisnya, *web mining* dibagi menjadi 3 (tiga) bagian, yaitu:

### 1. *Web structure mining*

*Web structure mining* merupakan proses yang menggunakan teori *graph* untuk menganalisa simpul (*node*) dan keterhubungan struktur dari situs. Menurut tipe dari struktur web, *web structure mining* terbagi menjadi 2 (dua). Jenis pertama adalah mengekstrak dari pola *hyperlink* di web. Sebuah *hyperlink* atau lebih dikenal sebagai *link* merupakan suatu komponen dari web yang memungkinkan suatu halaman terhubung dengan halaman yang lainnya.

Jenis kedua dari *web structure mining* adalah mining terhadap struktur dokumen. Yang dimaksud sebagai struktur dokumen adalah menganalisa struktur dari bahasa yang digunakan dalam web, yaitu bahasa HTML (*Hyper Text Markup Language*), atau XML (*eXtensibel Markup Language*) di dalam halaman.

### 2. *Web content mining*

*Web content mining* adalah proses untuk mendapatkan informasi yang berguna dari isi (*content*) di web. Isi (*content*) dapat berupa *text*, *image*, *audio* dan *video*. *Web content mining* terkadang disebut sebagai *web text mining*, karena teks merupakan bagian dari web yang paling banyak tersedia. Teknologi yang umumnya digunakan dalam *web content mining* adalah NLP (*Natural Language Processing*), dan IR (*Informational Retrieval*).

Secara umum *web content mining* akan berusaha mengubah kumpulan data di web yang begitu besar menjadi pengetahuan (*knowledge*) yang berguna bagi banyak orang.

### **3. *Web usage mining*.**

Menurut Srivastava, *web usage mining* merupakan teknik *data mining* yang berusaha mengungkap pola penggunaan dari halaman web, dalam rangka coba untuk memahami dan meningkatkan pelayanan kebutuhan dari aplikasi berbasis web. Jadi *web usage mining* sedikit berbeda dengan kedua jenis sebelumnya. Pada jenis *structure* dan *content mining*, yang dianalisa atau digali adalah data didalam web itu sendiri, namun pada *web usage mining* yang dianalisa adalah pengguna atau pengunjung dari halaman web. Sehingga karena yang coba dianalisa adalah tingkah laku dari pengunjung (pengguna) dari web maka hasil dari *web usage mining* banyak digunakan dalam e-marketing dan e-commerce. Hasil analisa dapat digunakan untuk meningkatkan layanan dari aplikasi web.

Hasil *web usage mining* antara lain informasi mengenai segmentasi pengunjung dari situs (aplikasi web). Segmentasi dapat dilihat berdasarkan lokasi (negara, kota atau wilayah), waktu akses (pagi, siang, sore atau malam), penggunaan browser dan sebagainya. Dalam situs ecommerce misalnya dapat digunakan untuk melihat pola pengunjung dalam pembelian produk seperti produk apa saja yang paling banyak dibeli (diakses), pengunjung dari mana saja yang banyak melakukan pembelian, dan sebagainya.

Perbandingan dari ketiga jenis *web mining* di atas dapat dilihat dalam tabel berikut ini

Tabel 1. Perbedaan Jenis-jenis Web Mining.

| Topic                  | Web Mining                                                                                                |                                                             |                               |                                                                                |
|------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-------------------------------|--------------------------------------------------------------------------------|
|                        | Web structure mining                                                                                      |                                                             | Web content mining            | Web usage mining                                                               |
|                        | IR View                                                                                                   | DB View                                                     |                               |                                                                                |
| View of Data           | Unstructured, Semistructured                                                                              | Semistructured, Web as DB                                   | Links structure               | Interactivity                                                                  |
| Main Data              | Text documnts<br>Hypertext docs                                                                           | Hypertext docs                                              | Links structure               | -Server logs,<br>-Browser logs                                                 |
| Representation         | -Bag of words, n-grams<br>-Terms, phrase<br>-Concept or ontology<br>-Relational                           | OEM,<br>Relational                                          | Graph                         | -Relational table,<br>-Graph                                                   |
| Method                 | -TFIDF and variants,<br>-Machine Learning,<br>-Statistic (NLP)                                            | Propietary algorithms,<br>ILP,<br>Association rules         | Propietary algorithms         | -Machine Learning,<br>-Statistical,<br>-(Modified) association rules           |
| Application Categories | Categorization,<br>Clustering,<br>Finding extraction rules,<br>Finding pattern in text,<br>User modelling | Finding frequent sub-structure,<br>Website schema discovery | Categorization,<br>Clustering | -Site construction, addaption and management<br>-Marketing,<br>-User modelling |

Berdasarkan tabel di atas, sumber data utama dari *web usage mining* adalah *server logs* dan *browser logs*. *Server logs* merupakan informasi yang dicatat di dalam server web setiap kali pengunjung mengakses suatu halaman web. Dari *log server*, didapat informasi akses web oleh pengunjung yang terdiri dari informasi antara lain:

- Informasi nama domain dari aplikasi situs yang diakses, bisa juga berupa alamat IP
- Waktu akses situs. Umumnya dalam format
- HTTP Request Field yang berisi jenis akses, halaman yang diakses dan jenis browser yang digunakan.
- Status akses berisi informasi status akses, misalnya 404 jika akses halaman tidak ditemukan.

- Ukuran (*byte*) dari halaman yang diakses.

Sumber kedua yang digunakan dalam *web usage mining* adalah *log browser*. *Log browser* dapat berupa *cookies*. *Cookies* berupa teks kecil yang tersimpan di dalam browser *client*. Informasi yang disimpan didalamnya antara lain informasi browser, informasi durasi (lama) pengunjung berada di suatu halaman. *Cookies* juga terkadang digunakan untuk menyimpan informasi sementara misalnya password user, produk yang dibeli (dalam situs ecommerce), dan sebagainya. Selain *cookies*, *browser log* dapat berupa *page tag*, hanya saja *page tag* biasanya ditanam secara sengaja di halaman web dan umumnya berupa *script javascript*. *Page tag* akan mengirimkan data pengunjung ke suatu sumber dimana selanjutnya data yang dikirimkan dapat di-*mining*.

Kedua sumber data tersebut memiliki kelebihan dan kekurangan, terlihat dalam tabel 2 di bawah ini. Secara umum, teknik *server log analysis* digunakan jika memiliki akses penuh terhadap situs dan *server* web yang digunakan. Karena data tersimpan di dalam file, maka *data log* relatif mudah diorganisasikan. Kekurangannya adalah jika terdapat kesalahan dalam pengaturan waktu di server, maka secara otomatis data yang disajikan di log server pun menjadi tidak valid. Sementara itu, teknik penggunaan page tags banyak dipilih jika akses terhadap server web terbatas. Kelebihan dari teknik ini adalah kemudahan dalam penerapan dan keakuratan data yang disajikan. Selain itu, saat ini juga banyak pihak ketiga yang menyediakan fasilitas *web analytic* yang menggunakan teknik ini, sehingga pemilik situs tidak perlu repot-repot dalam menanganinya. Salah satu contohnya adalah Google Analytics.

Tabel 2. Keuntungan dan Kekurangan Teknik Web Usage Mining

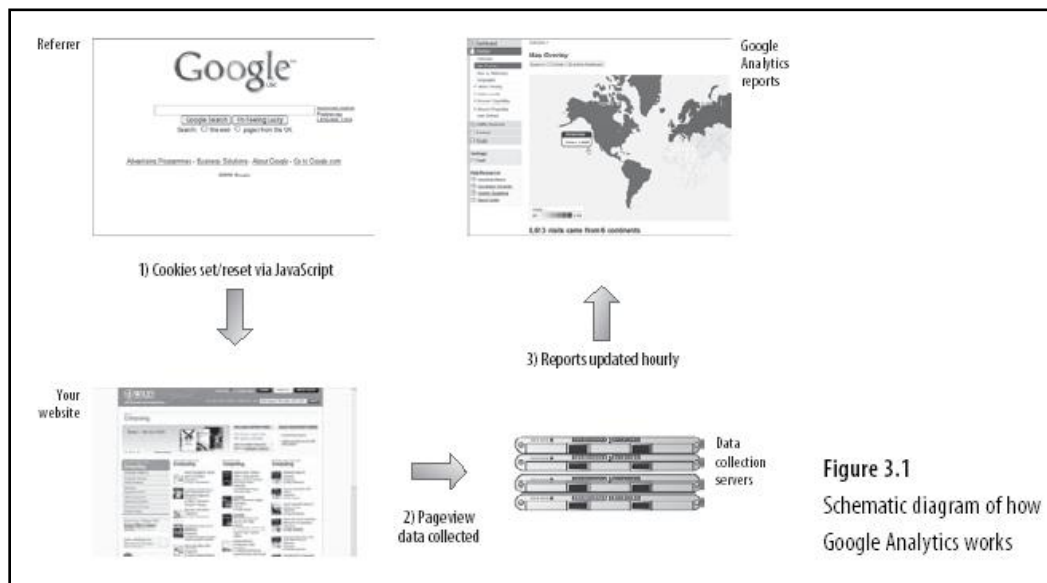
| Methodology               | Advantages                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Disadvantages                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Page tags                 | <ul style="list-style-type: none"> <li>• Breaks through proxy and caching servers—provides more accurate session tracking</li> <li>• Tracks client-side events—e.g., JavaScript, Flash, Web 2.0</li> <li>• Captures client-side e-commerce data—server-side access can be problematic</li> <li>• Collects and processes visitor data in nearly real time</li> <li>• Allows program updates to be performed for you by the vendor</li> <li>• Allows data storage and archiving to be performed for you by the vendor</li> </ul> | <ul style="list-style-type: none"> <li>• Setup errors lead to data loss—if you make a mistake with your tags, data is lost and you cannot go back and reanalyze</li> <li>• Firewall can mangle or restrict tags</li> <li>• Cannot track bandwidth or completed downloads—tags are set when the page or file is requested, <i>not</i> when the download is complete</li> <li>• Cannot track search engine spiders—robots ignore page tags</li> </ul> |
| Logfile analysis software | <ul style="list-style-type: none"> <li>• Historical data can be reprocessed easily</li> <li>• No firewall issues to worry about</li> <li>• Can track bandwidth and completed downloads—and can differentiate between completed and partial downloads</li> <li>• Tracks search engine spiders and robots by default</li> <li>• Tracks mobile visitors by default</li> </ul>                                                                                                                                                     | <ul style="list-style-type: none"> <li>• Proxy and caching inaccuracies—if a page is cached, no record is logged on your web server</li> <li>• No event tracking—e.g., no JavaScript, Flash, Web 2.0 tracking</li> <li>• Requires program updates to be performed by your own team</li> <li>• Requires data storage and archiving to be performed by your own team</li> <li>• Robots multiply visit counts</li> </ul>                               |

### C. Google Analytics

Google Analytics merupakan layanan gratis yang disediakan oleh raksasa mesin pencari Google. Google analytic menyajikan informasi sehubungan dengan pengunjung dari suatu website. Google Analytic merupakan salah satu aplikasi yang menyajikan informasi hasil web usage mining yang menggunakan teknik *page tags*.



Cara kerja dan penggunaan Google Analytics sangatlah mudah. Cukup dengan menyisipkan kode Javascript yang telah disediakan setelah anda menjadi anggota pengguna Google Analytics maka semua statistik halaman web yang telah disisipkan kode tersebut akan diproses oleh Google. Layanan ini memberi kemudahan dan keringanan kerja bagi webmaster atau pemilik situs.



Gambar 1. Cara Kerja Google Analytics

Keringanan kerja yang utama adalah tidak perlunya seorang pemilik situs atau webmaster memasang aplikasi *webservice log-analyzer* (tidak *real-time*), ataupun yang bersifat *real-time* terintegrasi dengan aplikasi situs, yang tentunya menambah kerja proses webservice.

Kemudahan yang lainnya adalah webmaster tidak perlu mengolah dan memilah log webservice karena semua akan dilakukan Google Analytics dengan berbagai parameter penilaian kinerja sebuah web di internet. Plus presentasi hasil pengolahan Google Analytics tampil secara elegan.

Berikut ini beberapa fasilitas yang ditawarkan oleh Google Analytics:

1. Mendukung berbagai bahasa dan tampilan (lebih dari 25 bahasa, termasuk Indonesia).
2. Cukup handal, dilihat dari hasil analisa yang didapat.

3. Dapat digunakan untuk situs skala kecil maupun besar.
4. Dapat diintegrasikan dengan layanan Google lainnya seperti Google Adwords
5. Menyajikan bentuk laporan yang beragam dan dapat dilakukan perbandingan antara beberapa kriteria.
6. Kemampuan untuk menentukan dan mengatur goal dari situs untuk selanjutnya dianalisa apakah goal tersebut tercapai atau tidak.
7. Tampilan laporan dan halaman Google Analytics dapat diatur sesuai keinginan dan kebutuhan.
8. Laporan dapat diekspor ke dalam berbagai format.

Google Analytics dapat diakses di alamat <http://www.google.com/analytics>.

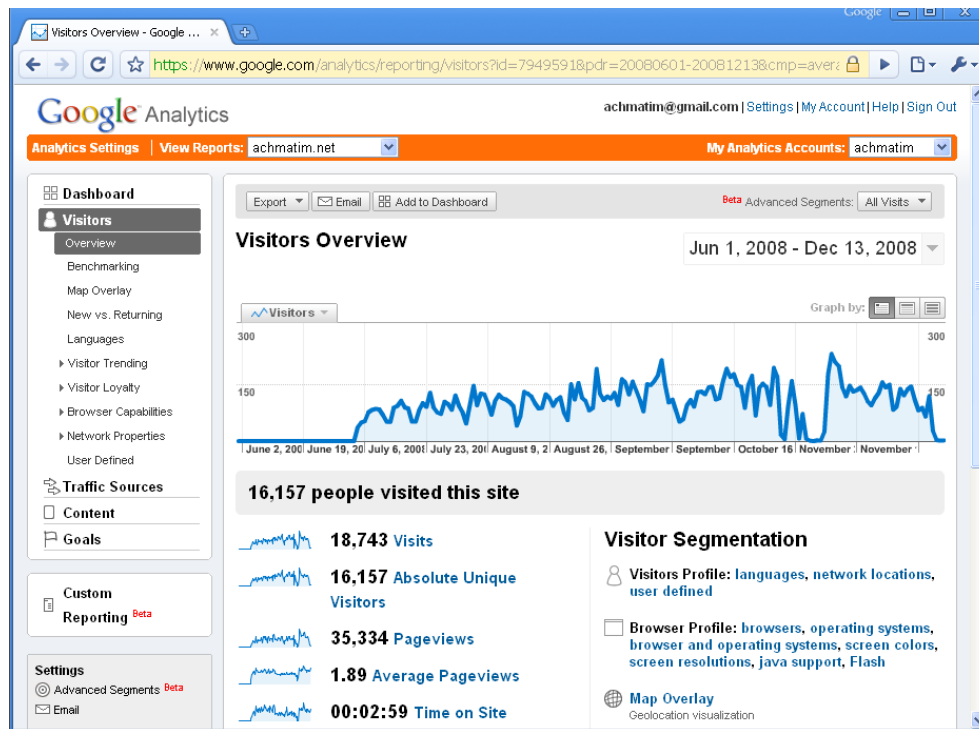
#### **D. Analisa Situs Achmatim.Net dengan Google Analytics**

Achmatim.Net merupakan situs pribadi yang dibangun menggunakan blogging software gratis terkemuka, Wordpress. Di dalam situs ini disajikan mengenai berbagai artikel berhubungan dengan bahasa pemrograman, *web development*, *database* dan materi-materi mengajar yang dimiliki oleh pemiliknya yang kebetulan seorang pengajar. Situs ini sudah online sejak pertengahan tahun 2005.

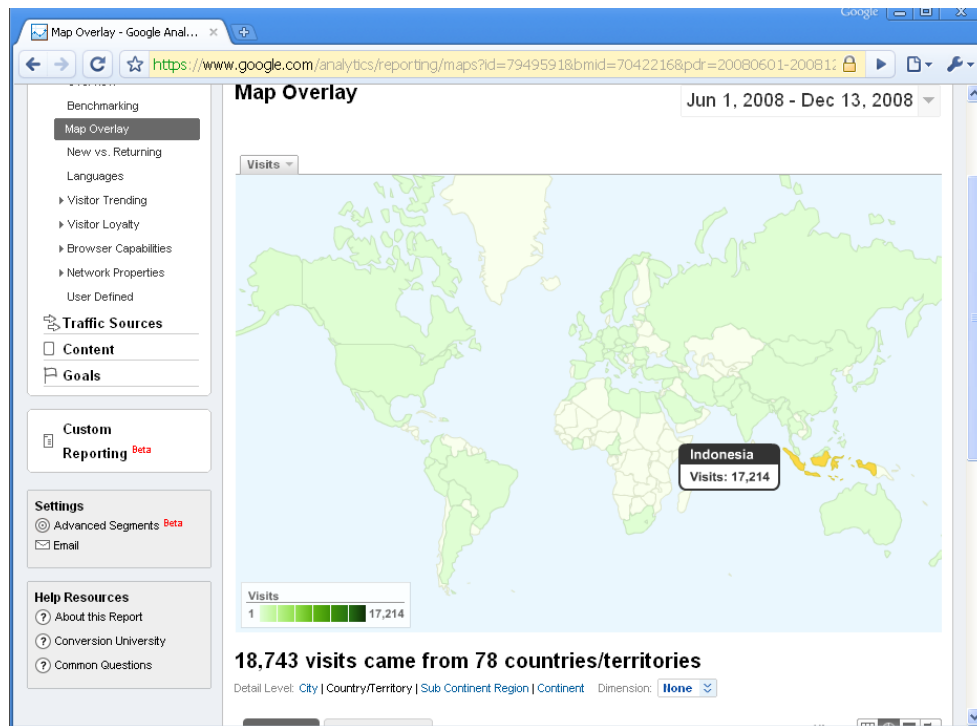
Beberapa informasi pengunjung situs Achmatim.Net dapat diperoleh dengan bantuan Google Analytics. Beberapa informasi tersebut diantaranya:

1. Informasi Jumlah Pengunjung per Periode Waktu (Gambar 2)
2. Informasi Segmentasi Pengunjung Berdasarkan Negara maupun Kota (Gambar 3 dan Gambar 4)
3. Informasi Segmentasi Pengunjung Berdasarkan Tipenya (Pengunjung baru atau pengunjung lama). (Gambar 5)
4. Informasi Presentase Pengunjung berdasarkan Waktu Berkunjung (Jam). (Gambar 6)
5. Informasi Lama Pengunjung Bertahan di Situs (Gambar 7)
6. Informasi Browser yang digunakan oleh Pengunjung (Gambar 8)
7. Informasi darimana Pengunjung Sampai Ke Situs (Gambar 9)

8. Informasi Prosentasi Halaman yang diakses Pengunjung (Gambar 10)
9. Dll.



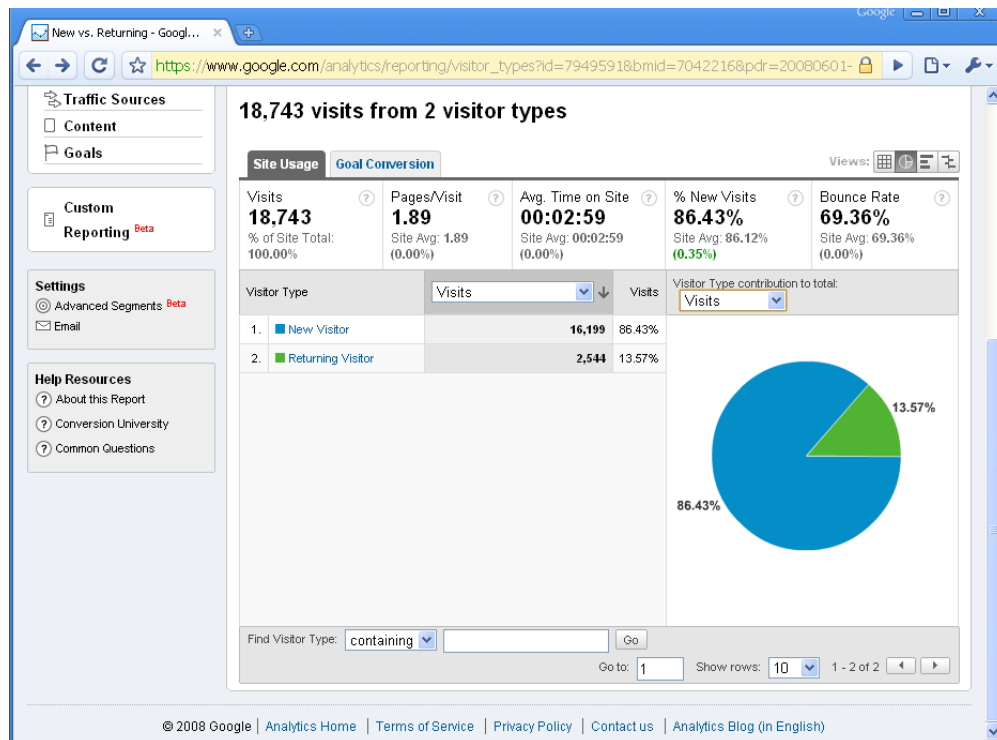
Gambar 2. Informasi Jumlah Pengunjung



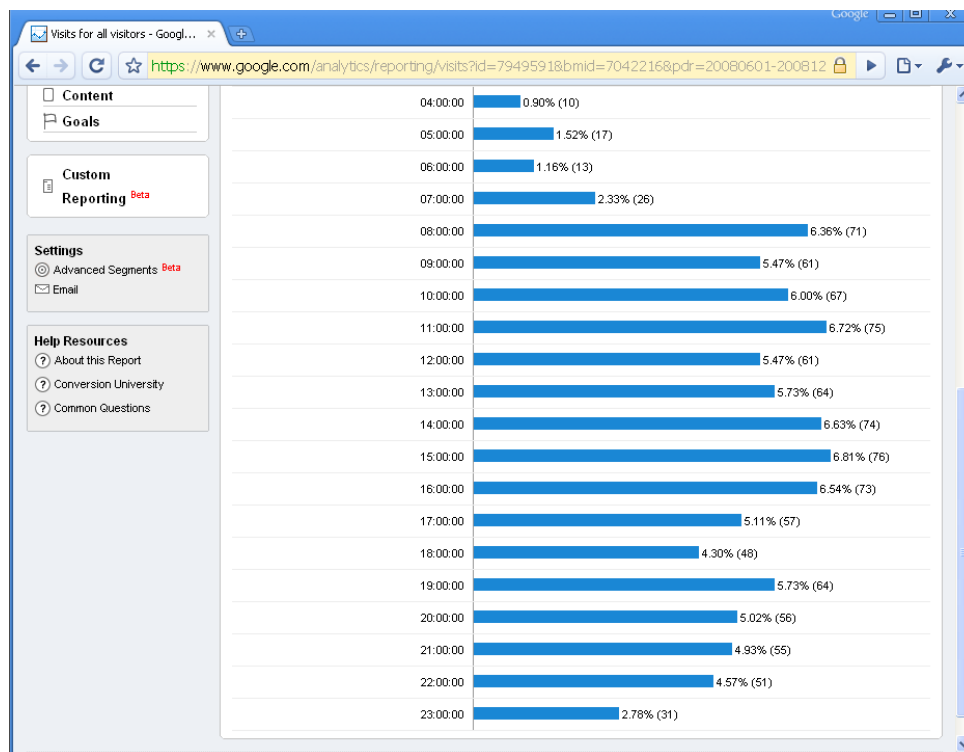
Gambar 3. Informasi Segmentasi Pengunjung berdasarkan Negara



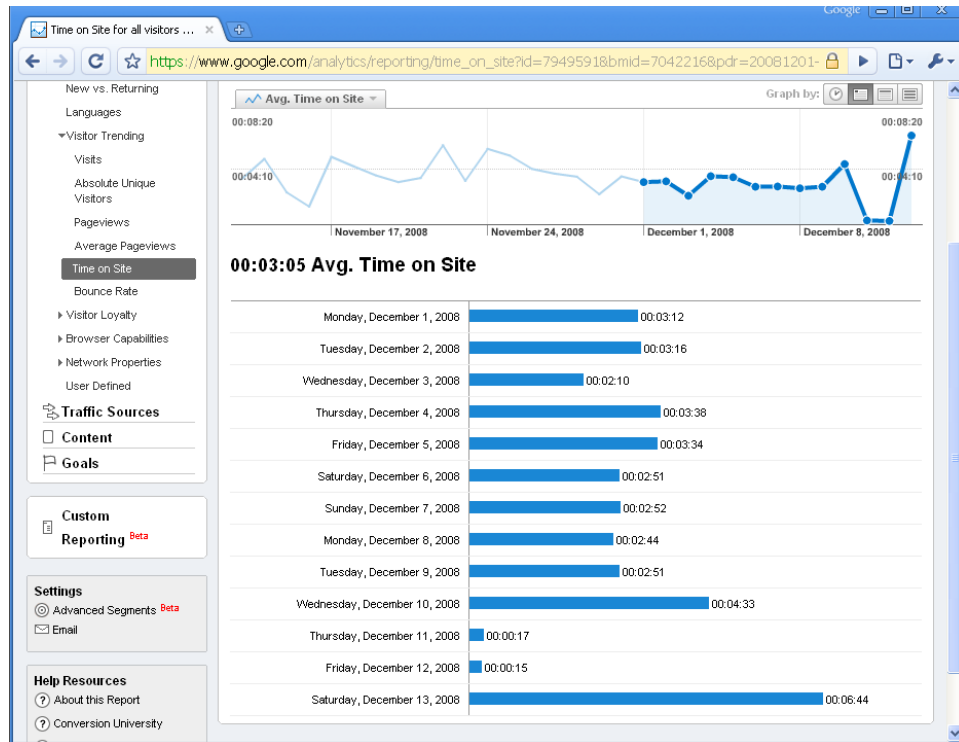
Gambar 4. Informasi Segmentasi Pengunjung Berdasarkan Kota



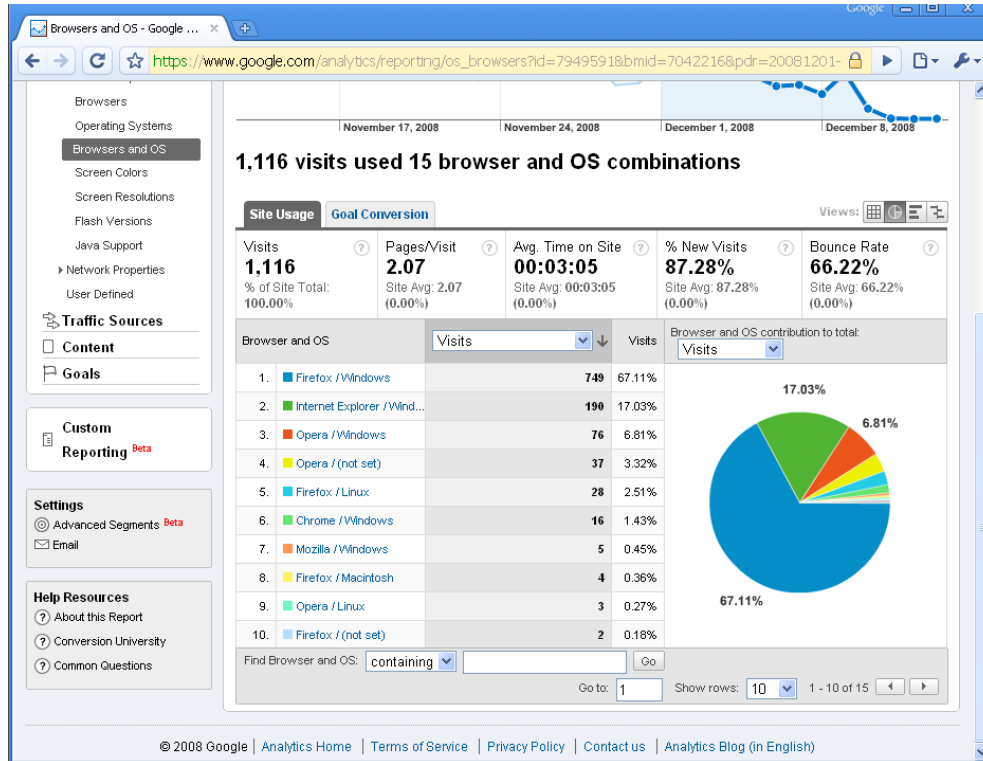
Gambar 5. Informasi Segmentasi Pengunjung berdasarkan Tipe



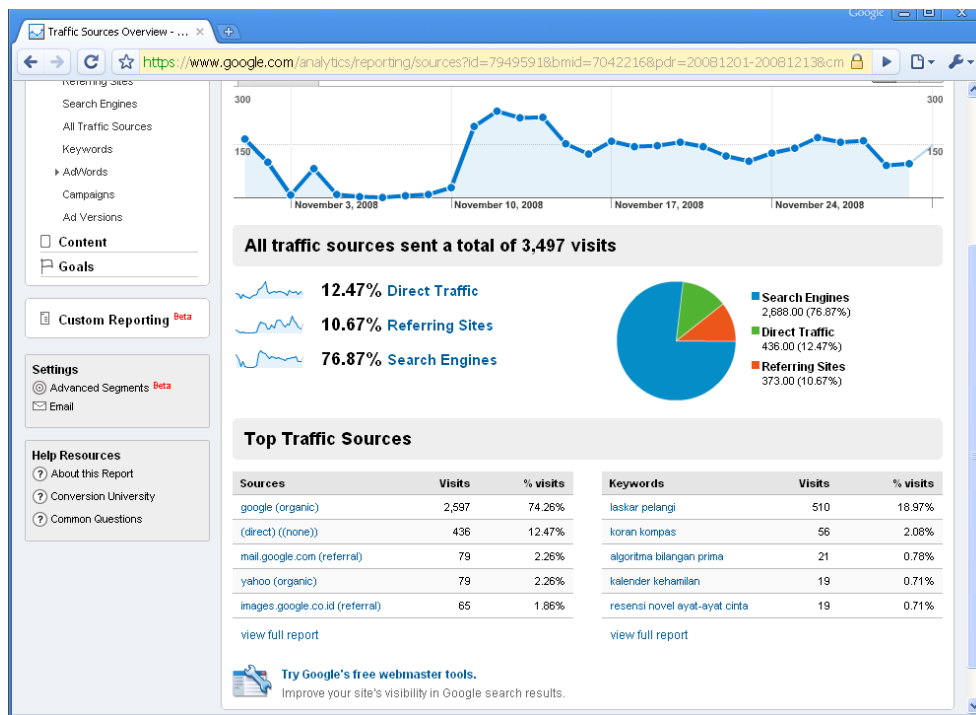
Gambar 6. Informasi Presentase Waktu Kunjung



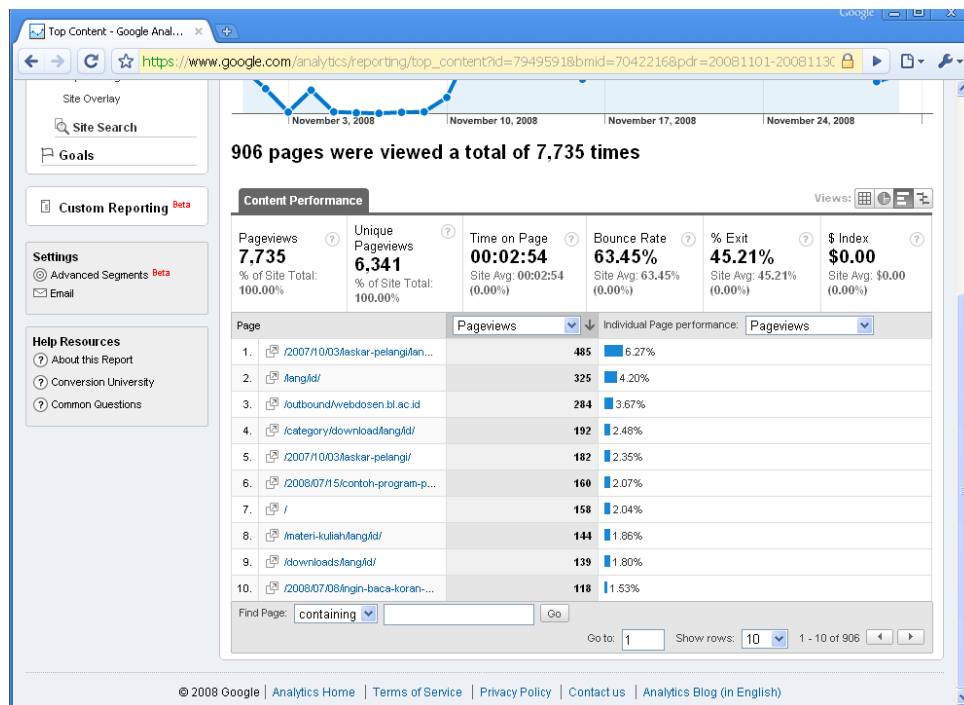
Gambar 7. Informasi Lama Waktu Kunjung di Situs



Gambar 8. Informasi Browser dan OS



Gambar 9. Informasi Asal Situs Refferer



Gambar 10. Informasi Jumlah Kunjungan ke Halaman.

## Kesimpulan

Dari paparan dalam makalah ini, dapat diambil kesimpulan bahwa

1. *Data mining* merupakan bidang ilmu baru yang keberadaannya sangat dibutuhkan dalam berbagai bidang terapan.
2. Web mining merupakan bagian dari *data mining* yang berusaha menggali informasi dari dunia web.
3. Web mining terdiri dari web structure mining, web content mining, dan web usage mining.
4. Google Analytics merupakan salah satu aplikasi web usage mining yang disediakan secara gratis oleh Google.
5. Google Analytics dapat digunakan untuk menganalisa pengunjung suatu situs termasuk performa dari situs tersebut.

## Referensi

1. Achmad Solichin, *Situs Achmatim.Net*, <http://achmatim.net>, 2008
2. Alan K'necht, *Dollars & Sense of Web Analytics*, [http://www.digital-web.com/types/the\\_dollars\\_and\\_sense\\_of\\_it/](http://www.digital-web.com/types/the_dollars_and_sense_of_it/), 2005
3. Brian Clifton, *Advanced Web Metrics with Google Analytics*, Wiley Publishing Inc, 2008
4. Daniel T. Larose, *Data mining Methods And Models*, John Wiley & Sons, Inc, 2006.
5. David Hand, Heikki Mannila dan Padhraic Smyth, *Principles of Data mining*. MIT Press, Cambridge, MA, 2001.
6. Google Inc, *Situs Google Analytics*, <http://www.google.com/analytics>, 2008
7. John E. Simpson, *Analyzing the Web*, <http://www.xml.com/pub/a/2005/07/27/tourist.html>, 2005.
8. Larry Greenfield, *Are Web Analytics Different?*, <http://www.dwinfocenter.org/webdata.html>, 2005
9. Mary E. Tyler and Jerri Ledford, *Google® Analytics*, Wiley Publishing, Inc, 2007.
10. Raymond Kosala, Hendrik Blockeel, *Web Mining Research: A Survey*, Department of Computer Science, Katholieke Universiteit Leuven, 2000



11. Robert Baumgartner dkk, Web Data Extraction for Business Intelligence: the Lixto Approach, Department of Information and Communication, Hochschule der Medien.
12. Wikipedia, Data Mining, [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining), 2008
13. Wikipedia, Google Analytics, [http://en.wikipedia.org/wiki/Google\\_Analytics](http://en.wikipedia.org/wiki/Google_Analytics), 2008
14. Wikipedia, Web Mining, [http://en.wikipedia.org/wiki/Web\\_mining](http://en.wikipedia.org/wiki/Web_mining), 2008
15. Yudho Giri Sucahyo, Data mining : Menggali Informasi yang Hilang.
16. Zdravko Markov and Daniel T. Larose, Data mining The Web : Uncovering Patterns in Web Content, Structure, and Usage, John Wiley & Sons, Inc, 2007.