



Università degli Studi di Bergamo  
Centro Studi per il Turismo e l'Interpretazione del Territorio (CeSTIT)  
Dipartimento di Scienze dei linguaggi, della comunicazione  
e degli studi culturali  
Facoltà di Lingue e Letterature straniere

# *Analytics, metrics, and databases*

Roberto Peretta

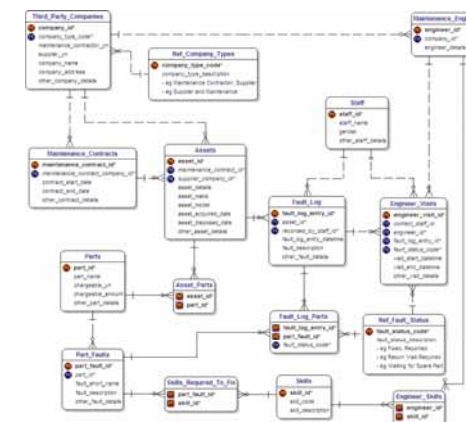
09.01 :: UniBg 44111 :: IT for Tourism Managers  
Thursday, December 17, 2011



# Analytics, metrics, and databases

*What are we talking about?*

1. Analytics
2. Metrics
3. A tool called Alexa
4. Toolbars
5. Databases





## **Analytics, metrics, and databases**

### *Web analytics*

According to Wikipedia, **Web analytics** is the measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing **web usage**.

Web analytics is not just a tool for measuring website **traffic** but can be used as a tool for business research and **market research**. Web analytics applications can also help companies measure the results of traditional print advertising campaigns.

Web analytics provides information about the **number of visitors** to a website and the number of **page views**. It helps gauge traffic and popularity trends which is useful for market research.



## Analytics, metrics, and databases

### *Google Analytics*

Although **Google Analytics (GA)** is obviously only one among many tools collecting data on visits to websites, it is widely recognized as a reliable and a leading source on information.

According to Wikipedia, GA is a free service offered by Google that generates detailed **statistics** about the visitors to a website. The product is aimed at **marketers** as opposed to webmasters and technologists from which the industry of web analytics originally grew. It is the **most widely used** website statistics service, currently in use on around 57% of the 10,000 most popular websites.



## Analytics, metrics, and databases

### *Analytics through page tagging*

Google Analytics is activated by the website's manager by adding a **javascript tracking code** to the html source of every page the traffic of which is to be recorded.

```
<script type="text/javascript">

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-XXXXX-X']);
_gaq.push(['_trackPageview']);

(function() {
  var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
  ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analytics.com/ga.js';
  var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

</script>
```

Google Analytics is therefore a web analytics technology based on **page tagging**.



## **Analytics, metrics, and databases**

### *Analytics through logfile analysis*

Other web analytics technologies are based instead on **logfile analysis**.

Web servers record some of their transactions in a logfile. These logfiles can be read by a program to provide data on the **popularity** of the website.

According to Wikipedia, **two units of measure** were introduced in the mid 1990s to gauge the amount of human activity on web servers.

These units were **page views** and **visits**, or sessions.



## Analytics, metrics, and databases

### *Metrics: page views and visits*

This brings us to the question of **metrics** in web analytics technologies.

Let's consider how the very first units of measure in web analytics technologies were defined in the mid 1990s.

A **page view** was defined as a **request** made to the web server **for a webpage**.

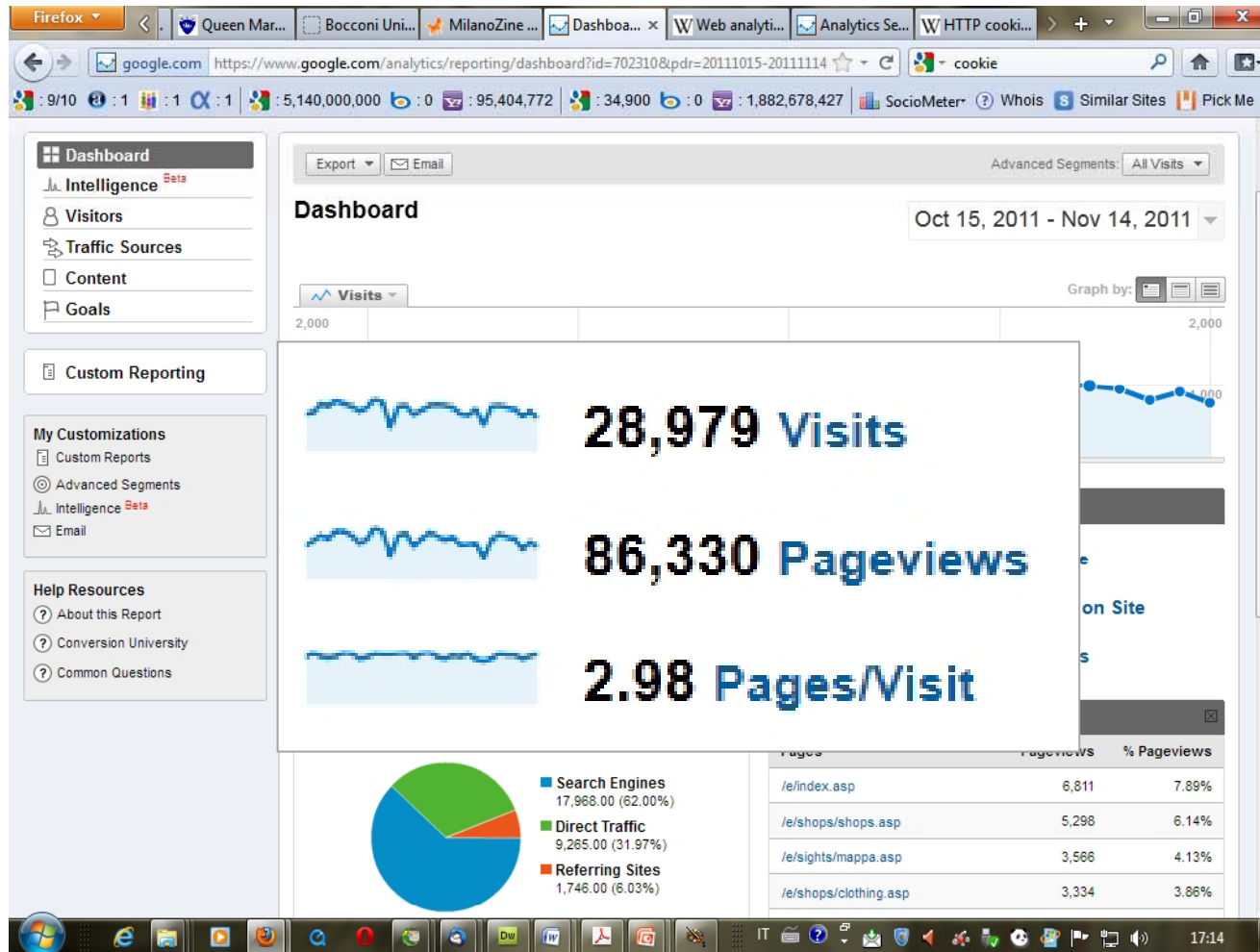
A **visit** was defined as a sequence of requests from a uniquely identified client that **expires after a certain amount of inactivity**, usually 30 minutes.

Now, let's see how page views and visits are displayed by a web analytics application's interface.



# Analytics, metrics, and databases

## *Metrics: a Google Analytics dashboard*





## Analytics, metrics, and databases

### *Metrics: bounce rate and unique visitors*

Independently on whether they are based on page tagging or logfile analysis, today's analytics technologies consider **more metrics** than simply the **page views** and the **visits** units of measure.

Here we list some of them, according to Wikipedia.

**Bounce rate.** The percentage of visits where the visitor enters and exits at the same webpage **without visiting any other pages** on the website.

**Unique visitors.** The uniquely identified client generating requests on the web server (log analysis) or viewing pages (page tagging) **within a defined time period** (i.e. day, week or month).

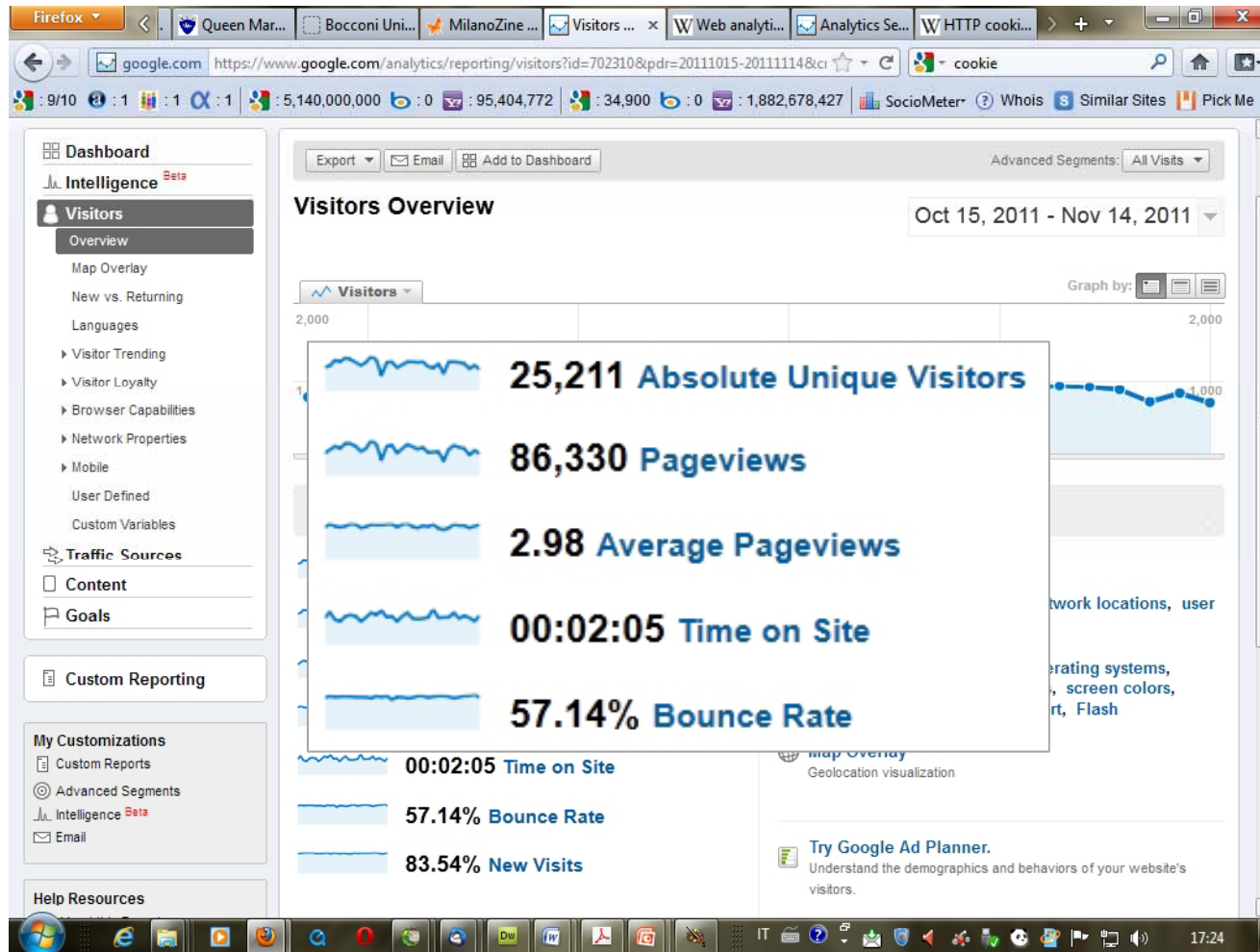
A Unique Visitor counts once within the timescale.

A visitor can make multiple visits.



# Analytics, metrics, and databases

## *Metrics: another Google Analytics dashboard*





## **Analytics, metrics, and databases**

### *Metrics and cookies*

Unique visitors' identification is made to the visitor's computer, not the person – usually via **cookie**.

By the way... what's a cookie?

According to Wikipedia, a **cookie** is used for an origin website to **send state information** to a user's browser and for the browser to **return the state information** to the origin site.

The state information can be used for **authentication**, identification of a user session, user's preferences, **shopping cart contents**, or anything else that can be accomplished through storing text data on the user's computer.



## **Analytics, metrics, and databases**

### *Cookies and tracking*

Cookies are not software: just alphanumeric strings. They **cannot be programmed**, cannot carry viruses, and cannot install malware on the host computer.

However, they can be used by spyware to **track user's browsing activities** — a major privacy concern that prompted European and US law makers to take action.

Cookies can also be **stolen by hackers** to gain access to a victim's web account.

Now, let's see where cookies are stored (and can be deleted...), for instance in a Mozilla Firefox browser.



## Analytics, metrics, and databases

### *Cookies: where are they?*

Search:

The following cookies are stored on your computer:

Site	Cookie Name
1254.virgilio.it	
27esimaora.corriere.it	
27esimaora.corriere.it	OAS_SC1
a2a.solution.weborama.fr	
abmr.net	

Name: <no cookie selected>  
Content: <no cookie selected>  
Host: <no cookie selected>  
Path: <no cookie selected>  
Send For: <no cookie selected>  
Expires: <no cookie selected>



## **Analytics, metrics, and databases**

*Metrics: session duration, time on page*

Now let's leave cookies, and go back to our short list of **more metrics** according to Wikipedia.

**Session Duration.** Average amount of time that visitors spend on the site each time they visit.

This metric can be complicated by the fact that analytics programs cannot measure the length of the final page view.

**Page View Duration / Time on Page.** Average amount of time that visitors spend on each page of the site.

As with Session Duration, this metric is complicated by the fact that analytics programs can not measure the length of the final page view unless they record a page close event, such as `onUnload()`.



## **Analytics, metrics, and databases**

*Metrics: active time, page depth, etc.*

**Active Time / Engagement Time.** Average amount of time that visitors spend actually interacting with content on a web page, based on mouse moves, clicks, hovers and scrolls.

Unlike Session Duration and Page View Duration / Time on Page, this metric can accurately measure the length of engagement in the final page view.

**Page Depth / Page Views per Session.** Page Depth is the average number of page views a visitor consumes before ending their session.

This metric is calculated by dividing total number of page views by total number of sessions and is also called Page Views per Session or PV/Session.



## Analytics, metrics, and databases

### *Metrics: ranking.*

The evaluation **search engines** provide of a webpage, also called the webpage's "position", is measured in terms of **ranking**.

Although ranking is referred to a single webpage, it is widely assumed that **a whole website's** ranking is that website's home page ranking.

According to Wikipedia, a good ranking in the search engines provides a steady supply of interested visitors/customers, while a poor one sinks any idea into oblivion, no matter how good or relevant.

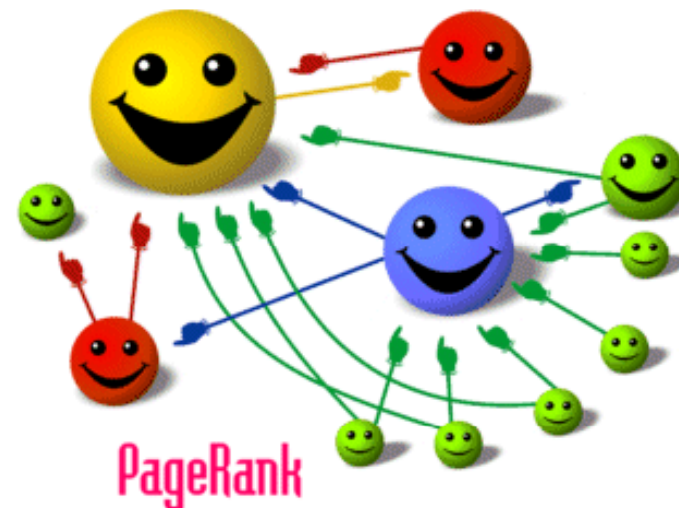
There is a whole industry called **SEO, search engine optimizing**, dedicated to obtain good web rankings. The most simple unit of measure for ranking is Google's **PageRank**.



## Analytics, metrics, and databases

*Metrics: PageRank.*

According to Wikipedia, PageRank is a link analysis **algorithm**, named after Larry Page and used by the Google Internet search engine, that assigns a **numerical weighting** to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of “measuring” its **relative importance** within the set.





## **Analytics, metrics, and databases**

*Metrics: industry and academy.*

The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element  $E$  is referred to as the PageRank of  $E$  and denoted by  $PR(E)$ .

The name “PageRank” is a trademark of **Google**, and the PageRank process has been patented (U.S. Patent 6,285,999).

However, the patent is assigned to Stanford University and not to Google. Google has **exclusive license rights** on the patent from Stanford University.

The university received 1.8 million shares of Google in exchange for use of the patent; the shares were sold in 2005 for \$336 millions.



## **Analytics, metrics, and databases**

### *Ranking tools*

How do you come to know the **PageRank** of a specific webpage, or – somehow – of a specific website?

As we'll see later on, some dedicated **tools** are available on the Web, to retrieve information on how and how much webpages and websites are visited.

Moreover, some **toolbars** do exist, that provide an easy and relatively comprehensive access to this sort of ranking data.

However, if you simply want to know the PageRank of a specific webpage, the most diffused relevant tool is **prchecker**.



# Analytics, metrics, and databases

## *Prchecker*

PR CHECKER .info

Page Rank checker is a free service to check Google™ page rank instantly via online PR checker or by adding a PageRank checking button to your web pages

← PAGERANK CHECKER - /> → CHECK PAGE RANK → NEWS & UPDATES → LINK TO US → CONTACT US → ADD TO BOOKMARKS

+f 1.4k Like

CHEAP HOSTING

Cheap Web Hosting - \$4.95/month

- Host many websites
- Unlimited web space
- Unlimited bandwidth
- PHP, ASP.Net, SSH
- FREE Blogs, Forums

Bonus: a FREE Domain

PHP 6 HOSTING plans

PHP6 Web Hosting - \$6.95/month

Their Best hosting plan:

- Unlimited host space
- Unlimited transfer
- PHP 5, PHP 6, MySQL 5
- Spam/Virus Protection

Bonus FREE Site Builder

UNLIMITED HOSTING


Check PAGE RANK of Web site pages Instantly

In order to check pagerank of a single web site, web page or domain name, please submit the URL of that web site, web page or domain name to the form below and click "Check PR" button.

Check PR

You are about to check PageRank for the URL below:

Anti-bot: please verify you are a human - submit a code shown on the image below:

 Anti-bot code:

If you can't read the code, [click here](#)

Verify Now



## Analytics, metrics, and databases

### *A tool called Alexa*

A widely used web tool providing data on visits and website ranking is **Alexa**.

According to Wikipedia, **Alexa Internet, Inc.** is a California-based subsidiary company of Amazon.com that is known for its toolbar and website. Once installed, the toolbar **collects data on browsing behavior** which is transmitted to the website where it is **stored and analyzed** and is the basis for the company's **web traffic reporting**.

The basic parameters of Alexa are

- ✓ **Traffic Rank**
- ✓ **Reputation**





## Analytics, metrics, and databases

### *Alexa Traffic Rank*

### **Alexa Traffic Rank**

A rough estimate of the website's popularity.

The rank is calculated using a combination of average **daily visitors** to the website and **pageviews** on the website **over the past 3 months**.

The site with the highest combination of visitors and pageviews is ranked #1

Updated daily





## Analytics, metrics, and databases

### *Alexa Reputation*

### **Alexa Reputation** or Sites Linking In.

The number of **links** to the website from sites visited by users in the Alexa traffic panel. Links that were not seen by users in the Alexa traffic panel are not counted.

Multiple links from the same site are only counted once.

Updated monthly





## Analytics, metrics, and databases

### *Toolbars: WebRank*

For those of you who rely on Mozilla Firefox as their browser, a **toolbar** may be recommended:

✓ **WebRank**

The screenshot shows a Mozilla Firefox browser window with the following elements:

- Browser Tabs:** USI - Uni..., Aree | MH D..., ict and sust..., Queen Mar..., Bocconi Uni..., MilanoZine..., Google.
- Address Bar:** www.usi.ch/en
- Toolbar:** Includes search engines (Google, Bing, Yahoo), social media (Facebook, Twitter), and analytics tools (SocioMeter, Whois, Similar Sites, Pick Me).
- Website Content:**
  - Language selector: It | En
  - Search bar: People directory, Search in the site, GO
  - USI open days banner
  - Image of two students studying
  - Navigation menu: FOR YOU, UNIVERSITY, EDUCATION, FACULTIES, RESEARCH, HIGHLIGHTS
  - Faculty list: ACADEMY of ARCHITECTURE, FACULTY of COMMUNICATION SCIENCES, FACULTY of ECONOMICS, FACULTY of INFORMATICS
  - Footer: Accessibility, Site map, Contact us
  - USI prospective students box
  - Header: *International, interdisciplinary, innovative*
  - Text: Welcome to USI Università della Svizzera italiana



## Analytics, metrics, and databases

*Databases: just a mention...*

All data like these – as most data worldwide – are collected, stored and retrieved by **databases**.

According to Wikipedia, a **database** is an **organized collection of data** for one or more purposes, usually in digital form.

The data are typically organized to **model** relevant aspects of reality (for example, the availability of rooms in hotels), in a way that supports **processes** requiring this information (for example, finding a hotel with vacancies).

This definition is very general, and is independent of the technology used.