

A Seminar on Web Analytics

Mahesh Kumar Bhandari
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science and Technology
The ICFAI University
Dehradun

26 February, 2016

Session Objectives

- 1 Introduction
- 2 Background
- 3 Data Collection and Analysis
- 4 Future Trends
- 5 Conclusion

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

- Web analytics applications help companies measure the results of traditional print or broadcast advertising campaigns.

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

- Web analytics applications help companies measure the results of traditional print or broadcast advertising campaigns.

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

- Web analytics applications help companies measure the results of traditional print or broadcast advertising campaigns.
- It helps one to estimate how traffic to a website changes after the launch of a new advertising campaign.

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

- Web analytics applications help companies measure the results of traditional print or broadcast advertising campaigns.
- It helps one to estimate how traffic to a website changes after the launch of a new advertising campaign.

Example:

Google Analytics offers a set of instruments for reviewing and evaluation online campaigns.

What is Web Analytics?

Web Analytics Association - Definition

The measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage.

- Web analytics applications help companies measure the results of traditional print or broadcast advertising campaigns.
- It helps one to estimate how traffic to a website changes after the launch of a new advertising campaign.

Example:

Google Analytics offers a set of instruments for reviewing and evaluation online campaigns.

- Web analytics provides information about the number of visitors to a website and the number of page views.

Categories of Web Analytics

- There are two categories of web analytics:

Categories of Web Analytics

- There are two categories of web analytics:

On-site web analytics

- Data collection on the current site
- Used to effectively measure many aspects of direct user-website interactions
- Example: **number of visits, time on site, click path, etc.**

Categories of Web Analytics

- There are two categories of web analytics:

On-site web analytics

- Data collection on the current site
 - Used to effectively measure many aspects of direct user-website interactions
 - Example: **number of visits, time on site, click path, etc.**
- **Google Analytics** is the most widely used on-site web analytics service, although new tools like **heat maps** and **session replay** are emerging that provide additional layers of information.

Categories of Web Analytics

- There are two categories of web analytics:

On-site web analytics

- Data collection on the current site
 - Used to effectively measure many aspects of direct user-website interactions
 - Example: **number of visits, time on site, click path, etc.**
- **Google Analytics** is the most widely used on-site web analytics service, although new tools like **heat maps** and **session replay** are emerging that provide additional layers of information.

Off-site web analytics

- Usually offered by third party companies(Twitalyzer or Sweetspot)
- It includes data from other sources which measures website's potential audience(opportunity), share of voice(visibility), and buzz(comments)
- Example: **surveys, market report, competitor comparison, public information, etc.**

Basic Steps of the Web Analytics Process

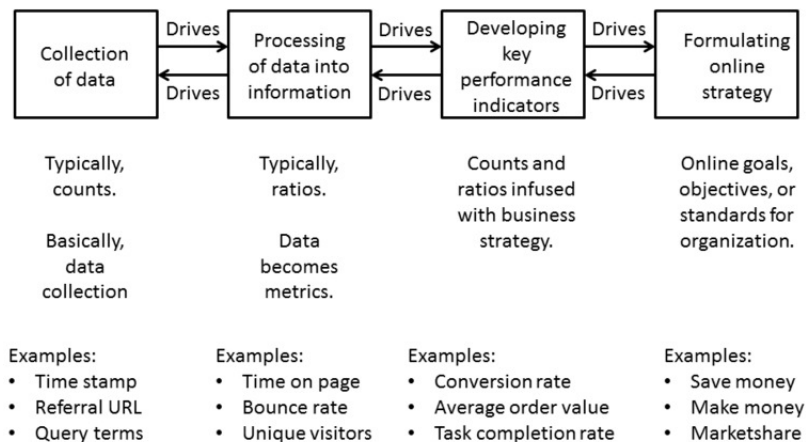


Figure 1: Basic Steps in Web Analytics

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.
- **WebTrends** was one the pioneers of web log analysis and created the first commercial website analytics software

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.
- **WebTrends** was one the pioneers of web log analysis and created the first commercial website analytics software
- **Dr. Stephen Turner** created Analog, the first free log file analysis software in 1995.

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.
- **WebTrends** was one the pioneers of web log analysis and created the first commercial website analytics software
- **Dr. Stephen Turner** created Analog, the first free log file analysis software in 1995.
- **WebSideStory** offered hit counter as a service for websites that would display a banner.

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.
- **WebTrends** was one the pioneers of web log analysis and created the first commercial website analytics software
- **Dr. Stephen Turner** created Analog, the first free log file analysis software in 1995.
- **WebSideStory** offered hit counter as a service for websites that would display a banner.
- **Web server logs** have some limits in types of data collected, as they could not provide information about visitors screen size, user interactions with page elements, mouse events, etc.

Background

- **Log files** have been used to keep track of web requests since WWW emerged and Mosaic was released in 1993.
- **WebTrends** was one the pioneers of web log analysis and created the first commercial website analytics software
- **Dr. Stephen Turner** created Analog, the first free log file analysis software in 1995.
- **WebSideStory** offered hit counter as a service for websites that would display a banner.
- **Web server logs** have some limits in types of data collected, as they could not provide information about visitors screen size, user interactions with page elements, mouse events, etc.
- **Page Tagging** is able to overcome the limitations and gets more popular recently.

Major Web Analytics Usages are:

Major Web Analytics Usages are:

- 1 Improving website/application design and user experience

Major Web Analytics Usages are:

- 1 Improving website/application design and user experience
- 2 Optimizing e-Commerce and e-CRM on customer orientation, acquisition, and retention

Major Web Analytics Usages are:

- ➊ Improving website/application design and user experience
- ➋ Optimizing e-Commerce and e-CRM on customer orientation, acquisition, and retention
- ➌ Tracking and measuring success of actions and programs such as commercial campaigns

Major Web Analytics Usages are:

- ➊ Improving website/application design and user experience
- ➋ Optimizing e-Commerce and e-CRM on customer orientation, acquisition, and retention
- ➌ Tracking and measuring success of actions and programs such as commercial campaigns
- ➍ Identifying problems and improving performance of web applications

Data and Sources

Data and Sources

- Dimensional model is used to study the data

Data and Sources

- Dimensional model is used to study the data
- There are two major types of data under dimensional model:

Data and Sources

- Dimensional model is used to study the data
- There are two major types of data under dimensional model:
 - ① Facts or Measurement data - are the quantities you want to measure. Visits, page views, hits, bounce rate and other items that can be quantified numerically.
 - ② Dimensional data - are those things you want to track. They're referrers, pages, country of origin, product category and other items whose attributes are often non-numerical.

Data and Sources

- Dimensional model is used to study the data
- There are two major types of data under dimensional model:
 - ① Facts or Measurement data - are the quantities you want to measure. Visits, page views, hits, bounce rate and other items that can be quantified numerically.
 - ② Dimensional data - are those things you want to track. They're referrers, pages, country of origin, product category and other items whose attributes are often non-numerical.
- Both measurement data and dimensional data come from a number of sources, which can be categorized into the following:

Data and Sources

- Dimensional model is used to study the data
- There are two major types of data under dimensional model:
 - ① Facts or Measurement data - are the quantities you want to measure. Visits, page views, hits, bounce rate and other items that can be quantified numerically.
 - ② Dimensional data - are those things you want to track. They're referrers, pages, country of origin, product category and other items whose attributes are often non-numerical.
- Both measurement data and dimensional data come from a number of sources, which can be categorized into the following:
 - ① Direct HTTP request data.
 - ② Application level data sent with HTTP requests.
 - ③ Network level and server generated data associated with HTTP requests.
 - ④ External data

Direct HTTP request data

- Data directly comes from the HTTP request messages

Direct HTTP request data

- Data directly comes from the HTTP request messages
- An HTTP request is a message sent by a web client to web server to request a resource.

Direct HTTP request data

- Data directly comes from the HTTP request messages
- An HTTP request is a message sent by a web client to web server to request a resource.
- Example: Request Command to get “www.spsu.edu/itdegrees”.

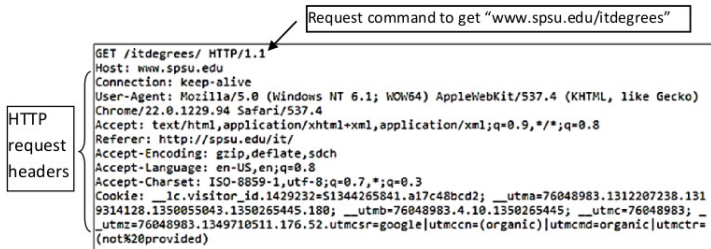


Figure 2: HTTP request header sample

Application level data sent with HTTP requests

- Data is generated and processed by application level programs such as JavaScript, PHP, and ASP.Net.

Application level data sent with HTTP requests

- Data is generated and processed by application level programs such as JavaScript, PHP, and ASP.Net.

Some common examples are:

- Sessions
- Referral
- User action
- Client/browser side

Application level data sent with HTTP requests

- Data is generated and processed by application level programs such as JavaScript, PHP, and ASP.Net.

Some common examples are:

- Sessions
 - Referral
 - User action
 - Client/browser side
-
- Application level data is usually embedded in HTTP requests. There are three common places to hold this information:

Application level data sent with HTTP requests

- Data is generated and processed by application level programs such as JavaScript, PHP, and ASP.Net.

Some common examples are:

- Sessions
 - Referral
 - User action
 - Client/browser side
-
- Application level data is usually embedded in HTTP requests. There are three common places to hold this information:
 - 1 Appended to a request URL as URL parameters.
 - 2 Sent as the HTTP cookie header.
 - 3 Included in the HTTP request body when an HTTP "Post" method is used(common for form submission)

Application level data sent with HTTP requests contd..

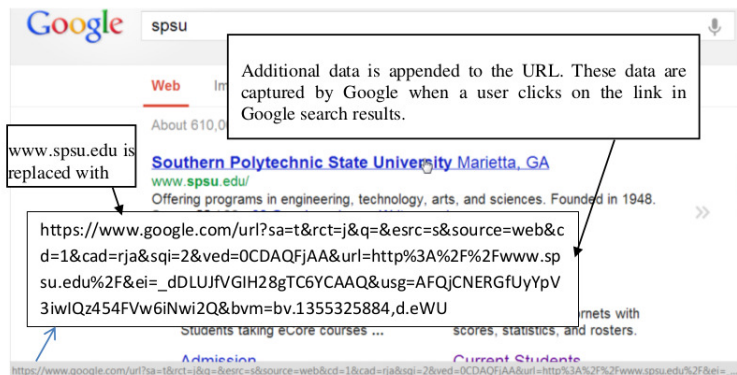


Figure 3: Google uses a transmission URL when redirecting a link to an external target[1]

Network level and server generated data

- Network level data

Network level and server generated data

- Network level data
 - It is not part of an HTTP request, but it is required for successful request transmission.
 - Example: IP address of a requester.
 - IP address and port number are required in order to return a response.
 - This information is sent at the TCP/IP level and is logged by a web server.

Network level and server generated data

- Network level data
 - It is not part of an HTTP request, but it is required for successful request transmission.
 - Example: IP address of a requester.
 - IP address and port number are required in order to return a response.
 - This information is sent at the TCP/IP level and is logged by a web server.
- Server generated data
 - Usually used for internal reference and is recorded in server log files.
 - Example: file size, processing time, server IP, request events other than HTTP request, etc.

- External data can be combined with on-site data to help interpret web usage.

- External data can be combined with on-site data to help interpret web usage.
- Example 1: IP addresses are usually associated with Geographic regions and ISP
- Third party databases or services provides such mapping, e.g., GeoIP, Geolite, IPInfoDB, GeoBytes, and hostip.info

- External data can be combined with on-site data to help interpret web usage.
- Example 1: IP addresses are usually associated with Geographic regions and ISP
- Third party databases or services provides such mapping, e.g., GeoIP, Geolite, IPInfoDB, GeoBytes, and hostip.info
- Example 2: User information collected and stored during a separate process(eg. registration)

Web Analytics Major Data and Source Summary

Table 1: Web Analytics Major Data and Source Summary

Data	Type	Source
Page view	Measurement	HTTP request
Client profile/User-Agent (browser, OS)	Dimension	Application, HTTP request
User action (keyboard and mouse)	Measurement	Application
Geo location	Dimension	External
Visit or session	Measurement, dimension	Application
Referrer (preceding webpage)	Dimension	HTTP request (“referrer” header)
Referral (channel identification)	Dimension	Application
Client profile (screen size, color depth)	Dimension	Application
IP address	Dimension	Network protocol
User profile	Dimension	External
Revenue or profit	Measurement	External

Collection and Tracking Methods - Web Server Logging

- A traditional method of usage data collection

Collection and Tracking Methods - Web Server Logging

- A traditional method of usage data collection
- A log file is generated by web server to record server activities and HTTP headers in textual format.

Collection and Tracking Methods - Web Server Logging

- A traditional method of usage data collection
- A log file is generated by web server to record server activities and HTTP headers in textual format.
- There are various formats of log files. Most commonly logged data in NCSA Common Log Format are:
 - Server IP
 - date/time
 - HTTP request command
 - response status
 - response size

Collection and Tracking Methods - Web Server Logging

- A traditional method of usage data collection
- A log file is generated by web server to record server activities and HTTP headers in textual format.
- There are various formats of log files. Most commonly logged data in NCSA Common Log Format are:
 - Server IP
 - date/time
 - HTTP request command
 - response status
 - response size
- Log analysis software can be used to extract and analyze log file. Some popular tools are:
 - Analog
 - Deep Log Analyzer
 - Webalizer
 - AWStats

Collection and Tracking Methods - Web Server Logging contd..

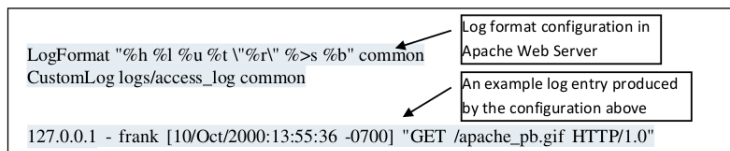


Figure 4: Common Log Format Example in Apache Web Server 2.2

Collection and Tracking Methods - Page Tagging

- Most recent method uses client side programs such as embedded scripts, browser add-ons and plug-ins.

Collection and Tracking Methods - Page Tagging

- Most recent method uses client side programs such as embedded scripts, browser add-ons and plug-ins.
- Example: A piece of JavaScript code is included in a page tracks user activity and stores information in a cookie.

Collection and Tracking Methods - Page Tagging

- Most recent method uses client side programs such as embedded scripts, browser add-ons and plug-ins.
- Example: A piece of JavaScript code is included in a page tracks user activity and stores information in a cookie.
- The information is sent to a processing server(may be a separate server) using web beacons or web service.

Collection and Tracking Methods - Page Tagging

- Most recent method uses client side programs such as embedded scripts, browser add-ons and plug-ins.
- Example: A piece of JavaScript code is included in a page tracks user activity and stores information in a cookie.
- The information is sent to a processing server(may be a separate server) using web beacons or web service.
- This method is commonly used by third party service providers such as Google Analytics and Open Web Analytics.

Common Analysis and Reports

- Meaningful and measurable metrics must be defined in order to analyze web traffic and relate it to business goals.

Common Analysis and Reports

- Meaningful and measurable metrics must be defined in order to analyze web traffic and relate it to business goals.
- The most common traditional metrics used in web analytics are:
 - **Visit count:** page view, visit, unique visitor
 - **Visit duration:** time on page, time on site.
 - **bounce rate**

Common Analysis and Reports

- Meaningful and measurable metrics must be defined in order to analyze web traffic and relate it to business goals.
- The most common traditional metrics used in web analytics are:
 - **Visit count:** page view, visit, unique visitor
 - **Visit duration:** time on page, time on site.
 - **bounce rate**
- The most basic analysis is the dimensional analysis involving measures and dimensions and it is the fundamental piece of other analyses and reports. Most Common types of analyses include:
 - Trend analysis
 - Distribution analysis
 - User activity or behavior analysis
 - Engagement analysis
 - Clickstream analysis
 - Visitor interest/attention analysis
 - Conversion analysis
 - Performance analysis

- Web 2.0 has brought many changes to the Web analytics industry

Future Scope

- Web 2.0 has brought many changes to the Web analytics industry
- AJAX changed how users interact with websites, and the future analytics will be more focused on event data rather than just based on HTTP requests.

Future Scope

- Web 2.0 has brought many changes to the Web analytics industry
- AJAX changed how users interact with websites, and the future analytics will be more focused on event data rather than just based on HTTP requests.
- Mobile web has also become a major trend in the last few years.

- Web 2.0 has brought many changes to the Web analytics industry
- AJAX changed how users interact with websites, and the future analytics will be more focused on event data rather than just based on HTTP requests.
- Mobile web has also become a major trend in the last few years.
- Web analytics traditionally was used for e-commerce sites, but recently expanded into other areas such as:
 - Learning analytics - Blackboard
 - Video analytics - YouTube
 - Search analytics - Google
 - Social media analytics - Facebook and LinkedIn

Conclusion

- Web analytics is a field of web traffic data collection and analysis.
- It has gained wide adoption and become one of the important tools to help web application management and business analysis.
- With the continuing expansion of data sources diversity of client systems, Web/digital analytics will play an even more important role in the future.