

A METRICS BASED APPROACH TO ANALYSE WEB USAGE PATTERN

Sukhjinder Singh

Department of IT
Adesh Institute of Engineering &
Technology, Add-Sadiq Road,
Faridkot, Punjab (India).

Amit Makkar

Associate Professor of CSE
Adesh Institute of Engineering &
Technology, Add-Sadiq Road,
Faridkot, Punjab (India).

Navneet Singh

Associate Professor of IT,
Adesh Institute of Engineering &
Technology, Add-Sadiq Road,
Faridkot, Punjab (India).

Abstract

Web metrics are established goals and standards for measuring website performance. Web Analytics can be used to analyze and statistically process user and customer behavior. Web Analytics especially refers to the use of data collected from a Website to determine which aspects of the Website work towards the business objectives. This paper provides a web metrics based approach that can be used to analyze and improve web usage pattern. We define a set of 15 web metrics that can play an important role in understanding web visitors behavior and provide suggestion how these metrics can help in making a website more popular. We describe the approach by considering a case study of the website gndu.ac.in for the data collected over a period of five months.

Keywords: Web Metrics, website, web analytics, web usage pattern, web visitor.

1. Introduction

Lord Kelvin said “If you cannot measure it, you can not improve it”. This fact also applies to analysis of web usage pattern. The more efficient a site is the better the design is since it will take up less of the users’ time. If the website design is more efficient there will be less steps for the users to go through to accomplish tasks and thus less things to remember, making it easier for them to reestablish proficiency even after a long period of not having visited the web. Requirement for heavy traffic to a website is that the website has to be simple and efficient.

In this paper we provide a metrics based approach that can be used to analyze and improve web usage pattern. We first describe what is a web analytics process followed by metrics concern with web. Then we provide a metrics based approach for web usage analysis. We describe the approach by considering a case study of the website gndu.ac.in for the data collected over a period of five months.

2. Web Analytics Process

Web analytics is the process of analyzing the behavior of visitors to a Web site. The use of Web analytics is said to enable a business to attract more visitors, retain or attract new customers for goods or services, or to increase the dollar volume each customer spends.

Web Analytics is the study of the behavior of Website visitors. In a commercial context, Web Analytics especially refers to the use of data collected from a Website to determine which aspects

of the Website work towards the business objectives; for example, which landing pages encourage people to make a purchase.

Web analytics is a tool for measuring website traffic and 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. It helps one to estimate how the traffic to the website changed after the launch of a new advertising campaign. Web analytics provides data on the number of visitors, page views etc to gauge the popularity of the sites which will help to do the market research [1].

2.1 Web Analytics Process

The objective of Web Analytics is to understand and improve the experience of online customers, while increasing revenues for online businesses. Among other techniques, this can be done by studying the ways customers navigate a website. According to the Web Analytics Association, the official definition of Web Analytics is “the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage.” Web Analytics is not a technology to produce reports; it is a process that proposes a virtuous cycle for website optimization.

As illustrated in Figure 1 below, the Web analytics process is composed of the following phases [2]:

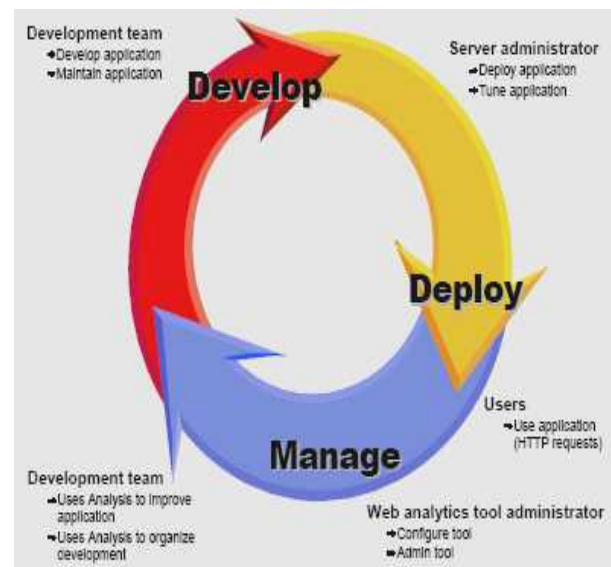


Figure 1: Web analytics process lifecycle [2]

I. Application users use the Web application; Information about this usage is gathered using data gathering methods.

II. Web usage data is analyzed by the Web analytics tool to produce reports; the tool is configured and administered by the Web analytics tool administrator.

III. Developers analyze these reports to define Web application updates.

Updated Web application is tested and deployed.

Analysis process

The analysis and report generation process has following four steps (Figure 2):

- I. Gather Web usage data.
- II. Parse data, and eventually store the data to a database and retrieve previously parsed data from the database.
- III. Analyze data, and eventually store the results.
- IV. Generate and distribute the reports.

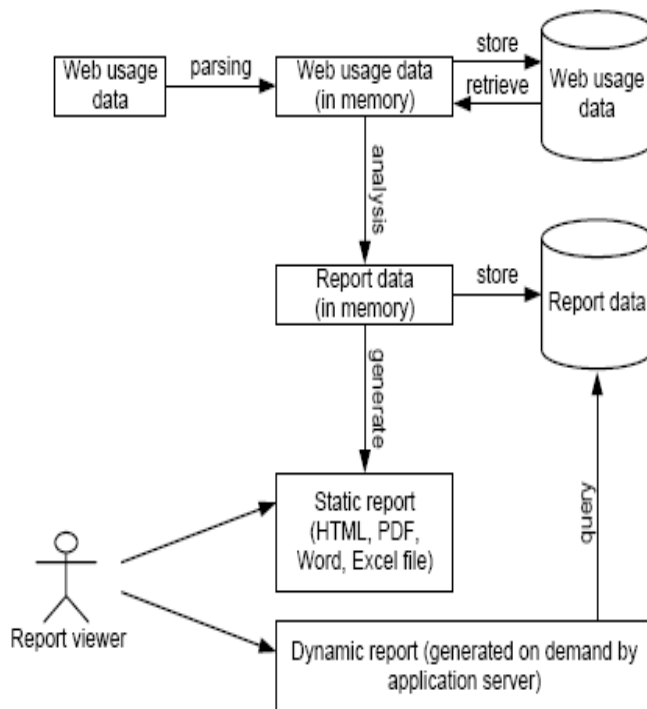


Figure 2: Analysis process [2]

2.2 Different Web analytics Tools

AWStats and Webalizer are the two most common statistics packages available through shared hosting services. Both are fairly basic, offering information about visits over time, most-visited pages, referrers, search strings, and some data about visitor's browsers and locations. Though Webalizer is a bit more popular, AWStats's reports are generally considered somewhat easier to understand [3].

Google Analytics is a class by itself. It offers substantially more functionality than the basic tools above. Unlike tools such as

Webalizer or AWStats, Google Analytics is needed to be installed on site, which involves pasting a chunk of HTML (provided by Google) into every page. This obviously requires a bit of HTML know-how, but if it is known what is to be done, shouldn't require too much effort. Depending on the size of site and how it's set up, installing the Google Analytics code might take anywhere from a couple of minutes to a few hours.

3. A Metrics based approach to analyze & improve web usage pattern

In this section we provide a metrics base approach that can be used to analyze and improve web usage pattern. We describe the approach by considering a case study of the website gndu.ac.in (Guru Nanak Dev University, Amritsar) for the data collected over a period of five months from 17 march 2011 to November 2010 and provide suggestions so that these web metrics can be effectively used to analyze web usage pattern and therefore can help in improving it. The data is generated by Webalizer

1. Number of hits per month

The number of hits per month received by a website is used in analyzing the usage pattern of the website during various months in a year. From figure 3 it is clearly observed that the month of February 2011 received maximum number of hits. The month received total of 747813 hits. Rise in the number of visitors (possibly students) is experienced as more visitors are interested in session end announcements

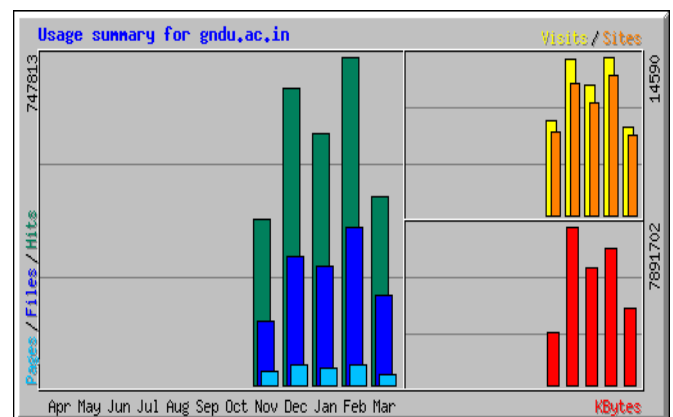


Figure 3: Monthly Usage summary for gndu.ac.in

visitors are accessing the website to know about the date sheets of the upcoming examinations, courses being offered by the university. Visitors are also interested in new vacancies, new tenders for upcoming session. The number of hits in the month of March 2011 is expected to be more than February 2011, but data is available up to 17th March so the number of hits is less in March as compare to February.

Figure 3 also shows that the month of November 2010 received minimum number of hits. The number of hits received in this month is 376690. This is expected as in the month of November least announcements are made by the university. The usage pattern shows that there is increase in the number of hits from the month of November 2010 to February 2011 and is also expected in the coming months. By finding out the months that

received low number of hits, appropriate action should be taken to increase web traffic.

2. Number of pages viewed per month

Pages are those URLs that would be considered the actual page being requested, and not all of the individual items that make it up (such as graphics and audio clips). From figure 3 it is clear that the highest number of pages are viewed in the month of February 2011 and the number is 45642. Since month of February 2011 received maximum number of hits, it is expected that this month should have maximum number of pages viewed. As most visitors are visiting website in this month, so the number of pages are more viewed.

On the other hand month of November received least number of pages viewed and the number is 30928. Since this month has minimum number of hits, so it is expected that the number of pages viewed would be least as less visitors are interested in viewing the pages of the website.

Reasons should be traced to find out the months in which minimum number of pages were viewed so the appropriate action can be taken to increase it. There can be decrease in the number of web pages to be viewed by the visitor by making use of flash/AJAX. The increase in online video, means fewer page views are counted, even though the same amount of content is being looked at.

3. Number of files accessed per month

Files represent the total number of hits (requests) that actually resulted in something being sent back to the user. An observation from figure 3 shows that the maximum number of files is accessed in the month of February 2011 and the number is 358222. The reasons behind this are that the number of hits is maximum in the month of February 2011. Visitors are downloading various files such as date sheets, syllabus, tenders etc. as session 2010–2011 moves towards its end.

On the other hand month of November 2010 received minimum number of files accessed as the number of hits is the minimum in the month of November 2010. More files accessed is a clear indication of popularity of the website and it may be increased after making a analysis of web pattern usage.

4. Number of visits per month

In Webalizer several pages viewed by one person (or search engine) within a half an hour are counted as one visit. Number of visits per month given total number of visits in a specific month. From figure 3 it is clear that the number of visits are maximum in the month of February 2011 and the number is 14590 as the number of hits are maximum in this month. The least number of visits 8049 are in the month of March 2011. This metrics may also help in increasing traffic to a web site if analysis of web usage pattern is made and actions are taken to focus on months having minimum number of visits per month.

5. Number of sites per month

Site is the number of unique IP addresses/hostnames that made requests to the server. Figure 3 shows that the maximum number of sites are in the month of February 2011. This number gives an idea that the most common visitors to the website are in the

month of February 2011. Month of march received minimum number of sites and the number is 7301. This metrics may also help to increase traffic after finding the maximum and minimum number of sites from the data provided.

6. Total Number of KBytes downloaded per month

This metrics is used to show the amount of data that was transferred between the server and the remote machine, based on the data found in the server log. From figure 3 it is observed that maximum number of downloading is done in the month of December 2010. The downloading total is 7891702. Although the month of December does not received maximum number of hits, but still its has maximum downloading. So, the observation confirms that the number of hits and total downloading are not related.

On the other hand month of November 2010 received minimum number of total down loading as in this month less announcements and uploading is made by the university. The count for this month is 2605278. So, files must be up to date in the months that have maximum downloading.

7. Total Unique Sites

Total Unique Sites is each request made to a server coming from a unique “site,” referenced by a name or, ultimately, an IP address. As shown in table 1 the count for total unique sites for the month of March 2011 is 7522. It is important to understand that this does NOT equate to the number of unique “real people” visiting site, which is impossible to

Monthly Statistics for March 2011	
Total Hits	449833
Total Files	213204
Total Pages	26077
Total Visits	8345
Total Kbytes	3990531
Total Unique Sites	7522
Total Unique URLs	811
Total Unique Referrers	587
Total Unique User Agents	2029

Table 1: Monthly Statistics for March 2011

determine via the http protocol and log files, but it is close. This metrics indicates the popularity of the website in the market and can help in increasing the web visitors.

8. Total URLs

This metrics can be used to ascertain the most popular referring URLs to the site (i.e. pages which provide a direct link to the web site. These pages can include search engines). Further, the data indicates the total number of referring or inbound links for the summary period.

Top 10 of 811 Total URLs By KBytes					
#	Hits	KBytes	URL		
1	2668	0.59 %	694445	17.40 %	/image/gndu.swf
2	3048	0.68 %	339464	8.51%	/download/gprospectus.pdf
3	9999	2.22 %	277820	6.96%	/image/stm31.js
4	1811	0.40 %	218298	5.47%	/oncampus/telephone.pdf
5	1306	0.29 %	75223	1.89%	/admission/distance_education.pdf
6	914	0.20 %	70050	1.76%	/admission/affiliatedcolleges.pdf
7	3291	0.73 %	56972	1.43%	/Index.asp
8	9422	2.09 %	54627	1.37%	/include_gndu/menublue.js
9	1335	0.30 %	53530	1.34%	/courses.pdf
10	363	0.08 %	52908	1.33%	/download/Prospectus-2008+N.T..pdf

Table 2: Top 10 of 811 Total URLs By Kbytes

Top 30 of 587 Total Referrers			
#	Hits	Referrer	
1	70262	15.42% http://www.gndu.ac.in/	
2	49973	11.11% http://gndu.ac.in/datesheetsii.asp	
3	33454	7.44% http://department.gndu.ac.in/department.asp	
4	33120	7.34% http://gndu.ac.in/download/appdsheets.asp	
5	24377	5.42% http://department.gndu.ac.in/download/others.asp	
6	23374	5.20% - (Direct Request)	
7	14433	3.21% http://department.gndu.ac.in/admson.asp	
8	12345	2.74% http://gndu.ac.in/	
9	11780	2.42% http://department.gndu.ac.in/download/forms.asp	
10	10017	2.33% http://department.gndu.ac.in/about/gndu.asp	

Table 3: Top 10 of 587 Total Referrers

Table 3 list out that the mostly used referrers are <http://www.gndu.ac.in/> and <http://gndu.ac.in/datesheetsii.asp> which are websites own URLs, so it does not provide much information about outside visitors to the website. The least referred URL is <http://department.gndu.ac.in/mes/registramessage.asp>.

10. Total User Agents

User Agents are a fancy name for browsers. Netscape, Opera, Konqueror, etc. are all User Agents, and each reports itself in a unique way to server. This metrics can be used to ascertain the most popular user agents accessing the site, the operating system/platform on which the software is running (and sometimes the language).

Data collected shows that the highest hits are for the user agent : Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV 1) With a count of 77303 hits and the least one is

Top 10 of 140 Total Entry Pages					
#	Hits	Visits	URL		
1	3291	0.73%	2097	25.29%	/Index.asp
2	2085	0.44%	1070	12.90%	/department.asp
3	1266	0.28%	749	9.03%	/download/others.asp
4	1154	0.24%	607	7.32%	/admson.asp
5	616	0.14%	438	5.28%	/BED09/BED09.asp
6	629	0.14%	268	3.23%	/download/forms.asp
7	573	0.13%	244	2.94%	/about/gndu.asp
8	471	0.10%	208	2.51%	/about/amritsar.asp
9	317	0.07%	206	2.48%	/cdc/cdc.asp
10	358	0.08%	194	2.34%	/DOWNLOAD/Tenders.asp

Table 4: Top 10 of 140 Total Entry Pages

Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.1;+Trent/4.0;+SLCC2;+.NET+CLR+2.0.50727;+.NET+CLR+3.5.30729;+.NET+CLR+3.0.30729;+Media+Center+PC+6.0) With a count of 3492.

11. Number of Entry pages

Entry pages are those pages that were the first requested in a visit. This metrics provides information regarding most and least popular entry page of a website. From the data gathered we have to list the most popular entry pages. So that instead of analyzing and improving only the home page, this page may also be improved because the reality is that many people will arrive deep into website through search engines.

Table 4 list out /Index.asp and /department.asp to the most popular entry pages with a count of 3291 and 2085 hits. The least popular entry page is /DOWNLOAD/Tenders.asp with a count of 358 hits.

12. Number of Exit pages

Exit pages are those pages that were the last requested in a visit. The analysis of usage pattern is to be done to find the most exiting page in the website. Table 5 list out /Index.asp to be the most exiting page in the website with a count of 333291 hits. On the other hand /about/amritsar.asp is the least exiting page with a count of 471 hits. It reveals that this pages on website is driving

people away. Improvement over this page is required so that it does not be marked as the most exiting page in the website.

Top 10 of 156 Total Exit Pages					
#	Hits	Visits		URL	
1	3291	949	0.73%	11.84%	/Index.asp
2	2078	927	0.44%	11.57%	/datesheetsii.asp
3	1266	698	0.28%	8.71%	/download/others.asp
4	1154	512	0.24%	4.38%	/admson.asp
5	2085	353	0.44%	4.41%	/department.asp
6	616	344	0.14%	4.28%	/BED09/BED09.asp
7	629	295	0.14%	3.68%	/download/forms.asp
8	573	237	0.13%	2.94%	/about/gndu.asp
9	358	201	0.08%	2.51%	/DOWNLOAD/Tenders.asp
10	471	189	0.10%	2.34%	/about/amnitsar.asp

Table 5: Top 10 of 156 Total Exit Pages

13. Number of search strings

Search Strings are obtained from examining the referrer string and looking for known patterns from various search engines. According to usage pattern for the given data we have to find out most popular search string, so that it can be used to inform keyword strategies for the web site. Table 6 shows that the most popular search sting for the website is gndu with a count of 32 hits whereas search string gndu.ac.co is the least one with a count of 3 hits.

14. Number of Countries

Countries are determined based on the top level domain of the requesting site. This metrics is used in finding out the place on the earth from where the website is

Top 20 of 139 Total Search Strings			
#	Hits	Search String	
1	32	11.90%	Gndu
2	19	7.04%	gndu amnitsar
3	19	7.04%	guru nanak dev university
4	13	4.83%	gndu.ac.in
5	12	4.44%	www.gndu.ac.in
6	5	1.84%	guru nanak university
7	4	1.49%	guru nanak dev university amnitsar
8	4	1.49%	guru nanak dev university courses
9	4	1.49%	site of guru nanak dev university
10	4	1.49%	www.gndu.ac

Table 6: Top 10 of 139 Total Search Strings

accessed. Table 7 is used in finding out accessed. The topmost country is India from where the maximum access to the website is made. The number of hits for India is 77341. The other

Top 10 of 51 Total Countries							
#	Hits	Files	Kbytes		Country		
1	340071	75.40%	161379	75.48%	2926322	73.33%	Unresolved/Unknown
2	77341	17.18%	35263	14.54%	672253	14.82%	India
3	16719	3.72%	9075	4.24%	196862	4.92%	Network
4	8219	1.83%	4185	1.94%	124395	3.12%	US Commercial
5	1991	0.44%	777	0.34%	15794	0.40%	Australia
6	768	0.17%	192	0.09%	5429	0.14%	New Zealand (Aotearoa)
7	654	0.15%	366	0.17%	8192	0.21%	Canada
8	480	0.11%	220	0.10%	5007	0.13%	US Educational
9	371	0.08%	202	0.09%	4333	0.11%	Pakistan
10	334	0.07%	120	0.04%	6394	0.16%	Non-Profit Organization

Table 7: Top 10 of 51 Total Countries

countries having top access to the website are US Commercial, Australia, New Zealand (Aotearoa) and Canada with 8219, 1991,768 and 654 hits respectively. Steps should be taken to target these countries to increase traffic to the website by creating landing pages/offers or content with this geographic focus.

15. Daily access

This metrics is very important from web usage pattern view as it provides information that when in the day access to the website was high. The analysis of the data collected of the usage pattern shows a very high usage pattern at the beginning of the daily service i.e. at 8-9 AM period which is understandable as most of the users start using internet immediately on arriving at their respective offices. However, the usage of the site decreases with the passage of time during the entire day especially in the month of Feb 2011. However there is an appreciable volume of access of the site before the peak hours (8-9 AM) in the months of Dec 2010 and Jan 2011. There is an unusual high accessibility during 5-8 hours in the months of Dec 2010 and Feb 2011 period which could be attributed to unusual website access which is not explainable.

4. Conclusion and Future Work

The importance of web metrics has dramatically increased due to extremely fast growth in Internet technology. The web metrics is directly related to the purpose of the good website design. Website with poor traffic can easily destroy the purpose of using web metrics. In this paper with the case study of website gndu.ac.in we proposed a set of 15 web metrics that can be used to increase the popularity of the website by increasing its traffic.

Although we have tried to find out the web metrics that are concern with web usage pattern, but still there is a need to find out the more and new ones. Other metrics that can play role are Key Performance Indicator, Search engine optimization, IP address owner etc.

5. References

- [1] Let us analyse your site, Seolytics, Australia's Bespoke Digital Marketing Services Company. Available at <http://www.seolytics.com.au/home/detailed-analysis-for-your-site.html>
- [2] Jean-Pierre Norguet, How to use Web Analytics

- for Improving Web Applications, IBM Corp. 2004, Available at <http://www.redbooks.ibm.com/redpapers/pdfs/redp3820.pdf>
- [3] Laura S. Quinn, A Few Good Web Analytics Tools: Measure your Web site's success, April 19 2007, Available at http://www.techsoup.org/learningcenter/inter_net/page6760.cfm
- [4] www.tnwebtech.com/tennessee/search_engine_marketing.htm
- [5] A survey of Web metrics, Devanshu Dhyani, Wee Keong, Sourav S., Technological University, Singapore, Singapore, ACM Computing Surveys (CSUR) archive Volume 34, Issue 4 (December 2002), Pages: 469 – 503, 2002, ISSN:0360-030
- [6] Visual Analysis of Website Browsing Patterns, Stephen G. Eick Visintuit, Lecture Notes in Computer Science, Springer Berlin/Heidelberg 0302-9743 (Print) 1611-3349 (Online), Volume 2539/2002, ISBN: 978-3-540-00247-5.
- [7] Implementing Web Analytics, Overcoming the Top 5 Challenges , White Paper from Edgewater Technology Summer 2008.
- [8] An Empirical Study of How People Perceive Online Behavioral Advertising, Aleecia M. McDonald¹ and Lorrie Faith Cranor², Carnegie Mellon University November 10, 2009.
- [9] Collaborative Web Recommendation Systems – A Survey Approach, A. Kumar and Dr. P. Thambidurai , Global Journal of Computer Science and Technology Vol. 9 Issue 5 (ver 2.0) , January 2010 Page 30.
- [10] Web Analytics: Empowering Customer Centricity , Waisberg D., Kaushik, a., 2009 semj.org volume2 issue 1.
- [11] Website optimization with web metrics: a case study, Birgit Weischedel, Eelko K. R. E. Huizingh [24], ACM International Conference Proceeding Series; Vol. 156 archive Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet Business – to –business e – commerce track Pages: 463 – 470, 2006, ISBN:1-59593-392-1
- [12] An Empirical Study of How People Perceive Online Behavioral Advertising, Aleecia M. McDonald¹ and Lorrie Faith Cranor², Carnegie Mellon University November 10, 2009.
- [13] Evaluating how users interact with NHS Direct Online, Dr Shirley Large, and Kate Arnold, NHS Direct New Media, Strawberry Fields, Berrywood Business Village, Tollbar Way, Hedge End, Hampshire, SO302UN. <http://www.nhsdirect.nhs.uk>
- [14] Measurement and Analysis of Online Social Networks, Alan Mislove Massimiliano Marcon Krishna P. Gummadi Peter Druschel Bobby Bhattacharjee Max Planck Institute for Software Systems, Rice University.
- [15] Classifying web metrics using the web quality model, Coral Calero, Julia'n Ruiz and Mario Piattini, ALARCOS Research Group, Computer Science Department, University of Castilla-La Mancha, Ciudad Real, Spain
- [16] Web Analytics Key Metrics and KPIs, Web Analytics Association 2300 M Street, Suite 800 Washington DC 20037 standards@webanalyticsassociation.org 2005 Web Analytics Association Authors: Guy Creese & Jason Burby