Google Analytics: website traffic analysis in a Brazilian journalistic company

Beatriz Cristina Munhoz (beatrizmunhoz@gmail.com)
Production Engineering Dept. of Polytechnic Engineering School, University of São Paulo (USP), Brazil.

Renato de Oliveira Moraes (remo@usp.br)
Production Engineering Dept. of Polytechnic Engineering School, University of São Paulo (USP), Brazil.

Abstract
With increased use of the Internet as a tool for obtaining information, news organizations have raced to launch their websites and, more recently, to develop apps for smartphones and tablets. This article uses a case study to examine how a Brazilian journalism company analyzes the data that it collects from its homepage and how that information influences its decisions on issues such as the revenues earned from selling advertising and partnerships; these data are also important for setting targets in different areas.

Keywords: Business Intelligence, Google Analytics as a tool, Web Analytics

Introduction
In the past, due to the dearth of available media and the difficulty of finding information, printed newspapers were the main source of information. Today, newspapers are divided into sections, including news (concerning current affairs), advertising (advertisements for various products and services), services (weather, exchange rates) and classified ads (real estate, cars, jobs), and generate their revenues through subscriptions, sales, advertising sales and classified ads.

According to the National Newspaper Association (ANJ), the emergence of media, such as cable TV and the Internet, has changed how citizens choose to access information. Brazilian newspapers have replaced their managements to improve efficiency and now compete in a changed market. Although the overall newspaper circulation continues to grow, the largest Brazilian newspapers have seen their circulations dwindle (Table 1).

One hypothesis for this reduction in circulation is the emergence of other means of obtaining news and services that were previously provided, nearly exclusively, by newspapers but are now available from other sources, many of which are free.
Thus, several newspaper companies have begun to introduce products and services with the help of information technology that was initially guided by Internet portals but has more recently been guided by mobile applications for tablets and smart phones.

Table 1: Circulation of Brazilian Newspapers

<table>
<thead>
<tr>
<th>Ranking</th>
<th>State</th>
<th>2002</th>
<th>2012</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>Total</td>
<td>-</td>
<td>6,972,000</td>
<td>8,802,000</td>
</tr>
<tr>
<td>1</td>
<td>Folha de São Paulo</td>
<td>SP</td>
<td>346,333</td>
<td>297,650</td>
</tr>
<tr>
<td>2</td>
<td>Super Notícia</td>
<td>MG</td>
<td>Not Available</td>
<td>296,799</td>
</tr>
<tr>
<td>3</td>
<td>O Globo</td>
<td>RJ</td>
<td>266,185</td>
<td>277,876</td>
</tr>
<tr>
<td>4</td>
<td>O Estado de São Paulo</td>
<td>SP</td>
<td>268,433</td>
<td>235,217</td>
</tr>
<tr>
<td>5</td>
<td>Extra</td>
<td>RJ</td>
<td>286,655</td>
<td>209,556</td>
</tr>
<tr>
<td>6</td>
<td>Zero Hora</td>
<td>RS</td>
<td>169,669</td>
<td>184,674</td>
</tr>
<tr>
<td>7</td>
<td>Diário Gaúcho</td>
<td>RS</td>
<td>125,430</td>
<td>166,221</td>
</tr>
<tr>
<td>8</td>
<td>Daqui</td>
<td>GO</td>
<td>Not Available</td>
<td>159,022</td>
</tr>
<tr>
<td>9</td>
<td>Correio do Povo</td>
<td>RS</td>
<td>183,077</td>
<td>149,562</td>
</tr>
<tr>
<td>10</td>
<td>Meia Hora</td>
<td>RJ</td>
<td>Not Available</td>
<td>118,257</td>
</tr>
</tbody>
</table>

The new media landscape has made it imperative that companies focus on new strategies and on developing new products and services to respond quickly to the changed environment in which they find themselves.

Thus, this study evaluates the importance of measuring access to an organization's website and analyzes how a specific newspaper company is evolving digitally.

This article is organized as follows. First, the main concepts of Web Traffic Information and Business Intelligence are discussed. Second, a case study is presented of an organization seeking to determine how to apply measurement and analysis tools and how those tools will benefit the company. Third, limitations are identified based on the number of Web searches. Nonetheless, these limitations can be used to inform future research.

Web Traffic Information and Business Intelligence

With today's heightened access to and availability of information, many organizations have become interested in Web Traffic Information and Business Intelligence. Fleisher and Bensoussan (2003) note that Business Intelligence (BI) is gaining in popularity due to its ability to collect, store and process information for decision making.

Halliman apud Fleisher and Bensoussan (2003) note that BI is a process linked mainly to technology stocks and that it consists of locating data in a variety of databases within an organization and extracting these data. Another definition of BI is given by Kohavi et al. (2004), who stress that BI is a set of techniques that can be used in various processes: collection, validation, processing, understanding, building models, data mining, knowledge sharing, reporting and visualization.

By examining the Internet scenario, where access has become popular in the last decade though its importance has long been recognized, it becomes evident that many systems properties, such as how people interact and the extent to which culture affects those
interactions, are not fully understood (GONÇALVES and RAMASCO, 2008). The same authors believe that a better understanding of how individuals use websites is economically important. Chang and Tsai (2009) argue that it is necessary to analyze the data collected from websites because that data can be used in conjunction with a company’s marketing strategy to generate site visitors and convert those visits into sales. Chang and Tsai further argue that traffic analysis can yield valuable information for website administrators who want to customize the information that is hosted on the Web server to reach a wider audience.

Hasan, Probets and Morris (2009) believe that ease of use is the most important feature that websites can offer and that analyzing web traffic information, which involves collecting, measuring, monitoring, analyzing and reporting usage data, allows companies to understand the visitor experience that they are offering. Web analytics can help websites achieve business goals or improve customer satisfaction and loyalty.

According to Gonçalves and Ramasco (2008), many studies describe the structure of a Web statistical perspective, but the biggest challenge lies in identifying how users navigate websites. Meiss et. al. (2008) argue that the most comprehensive source of data is real web traffic data. This information tells analysts how users navigate the Internet. These users are people who visit websites via links, people who visit websites by typing the URL directly, or even people who access pages from search engines such as Google or Yahoo. The authors also mention the possibility of measuring the number of hits per hour.

This information is accessed through website servers. For a user to access a webpage, he must make a request on the concerned website (URL). The computer that the user is using is then identified via its IP address, which is a single record, by the website’s server. The server identifies the IP address, and the request is fulfilled by delivering the requested data. The server is capable of identifying the URL (requested page), the user and the time of the request. Additional information is also recorded and stored (GONÇALVES and RAMASCO, 2008).

Phippen, Sheppard and Furnell (2004) believe that the basic metrics used in website analysis, such as the number of visitors or page views, can create inaccuracies. The authors argue for improvements in basic metrics to improve definitions and specific goals for reducing or eliminating the misinterpretation of information. The authors argue that consumers are the key to a website’s success and that understanding how users and consumers use websites is essential to developing focused metrics.

Web Analytics is a tool that collects, measures, analyzes and reports Internet data for the purpose of understanding and optimizing web usage. The data are collected automatically from the system with high accuracy. Google Analytics, a web analytics tool, offers statistical information in a visual way, allowing one to understand the interactions between a website and its visitors. Other examples of web analytics tools are VisiStat, StatCounter and ClickTracks (FANG, 2007).

**Methodology**

Although there are several tools for analyzing information flow on the Web, the difficulties associated with analyzing such information are debated. This article examines how a newspaper company monitors its website.
The study was developed through a literature review with the goal of advancing theories that encompass theory concerning Web traffic and BI. A single qualitative case study with a low level of involvement of the researcher was later applied to this study (MICHAEL, 2009).

The following criteria were considered in selecting the studied company: (i) the company had to have a dedicated Business Intelligence and Metrics department and (ii) the company had to be a major newspaper in Brazil.

The purpose of the case study was to determine whether the following proposition, "the newspapers have control of the visited pages, execute regular access analysis for decision making and use this information to increase its revenues," is valid.

**Case Study**

The company selected for examination in this study is a high-circulation national Brazilian newspaper that has a department dedicated to Internet projects, including the designing of apps for smartphones and tablets.

The study data were collected through a semi-structured interview with a company project manager who was engaged in the development and implementation of digital projects, and with a company metrics expert who was responsible for analyzing data access to the company’s website.

The respondents reported that follow-up is performed on the website using Google Analytics, a free BI tool that measures the number of hits that a website receives. This tool was implemented by the company in 2007 and has since gained greater importance in its operations.

According to the project manager, BI analysis is fundamental to understanding the behavior of users toward information across the website, such as greater access to news, origin of access and the time (day or time) at which the user visited the website. This information allows the company to share specific news articles in a timely manner so that they reach a large number of readers who can use the information. As a result, the company can direct news content and advertising.

Additionally, access information is relevant for the advertising sales team because it allows advertisers, through the information contained in the reports, to choose the appropriate times for advertising. Furthermore, the company can use the information as a price adjustment tool because it can measure the number of hits the website receives.

The journalism company’s website data were analyzed for this study by Google Analytics between July 1, 2013, and July 31, 2013.

**Observations**

Figure 1 demonstrates that 22% of the accesses were made by new users (new visitors) and that 78% were made by returning visitors during the study period (returning visitors). The average number of page visits was 6 per user, and each user averaged 14 minutes on the website.
Figure 1: Visits to homepage. Source: Google Analytics

Based on this information, it is interesting to note the number of people accessing the newspaper’s website from countries where Portuguese is not spoken. This leads us to believe that these particular users are Brazilians living abroad. This observation is shown in Figure 2.

Figure 2: Visits to the homepage by country. Source: Google Analytics

Figure 3 shows the daily visits to the company’s homepage. It is interesting to note that vouchers represent the weekend and that the greatest number of visits occurred on Mondays. The per-hour analysis presented in Figure 4 shows that the visits began to increase at 6:00 a.m., peaked at 9:00 a.m. and fluctuated until 5 p.m., at which point the number of visits peaked a second time and subsequently declined.

The project manager commented on other important information related to the browsers used in computers and the operating systems used in smartphones and tablets. This information is presented in figures 5 and 6. The information is relevant to the programming area of the
website, which requires various architectures to enable the correct loading of pages so that visitors can read website content on different platforms.

It is interesting to note that Google's Chrome browser is the top choice of users, accounting for 38.4% of the company's website hits. When comparing the study period to the same period in 2012, an increase of 3.1 percentage points in Chrome users is evident. This increase is equivalent to an increase of 8.7%. The reverse is observed for Microsoft Internet Explorer, whose users decreased by 20.8%, equivalent to 7.0 percentage points. Apple's Safari browser usage increased by 5.4 percentage points, equivalent to an increase of 54.0%. Android browser usage was not examined in 2012.
Based on the information of the types of browsers used to access the website, the development of Internet Explorer 6 was discarded by the company due to its low volume of hits and the high cost of investing in its programming.

![Pie chart showing browser usage]

**Figure 6: Visits by Operational Systems (Smartphones and Tablets). Source: Google Analytics**

It is interesting to note in figure 6 that 62.4% of the website hits were made using Apple's iOS for smartphones and tablets, followed by Android's smartphone and tablet platform with 28.6%. When comparing the same period in 2012, it is observed that Apple’s iOS usage decreased by 17.9%, the equivalent of 13.6 percentage points, in 2013, and that Android’s platform usage increased by 50.7%, the equivalent of 9.6 percentage points.

Figure 7 shows how users access the website. Access methods include the various search engines used to find the website, referral traffic references from other websites and direct traffic based on typing the website’s URL. A representative number of hits originated from referrals on social media networks.

The use of social media has increased significantly in user access. By analyzing the same period in 2012, it is observed that only 14.8 percentage points of users were directed to the company’s website through links. In 2013, that figure jumps to 25 percentage points, an increase of 70%. Social media is responsible for those numbers. The number of website hits obtained through social media in July 2012 increased by 2.5% and 7.0% in July 2013. Figure 8 shows that Facebook accounted for 84.6% of the hits, followed by Twitter, with 12.2%.
Analysis Results

The company examined in this study has access control for its users on its website. A series of metrics has been generated by Google Analytics, which allow for decision making by the editors responsible for various sections of the newspaper, such as economics, politics and sports. These editors, apart from selecting the themes that will be highlighted on the website, analyze website hits. If any piece of news garners a significant number of hits, it can be featured on the homepage. Similarly, if a specific piece does not garner the expected number of views, it can be replaced or its content can be reviewed to attract more hits. Such an analysis can be performed in real time.

As reported by the respondents, the number of website visits is critical to the company’s advertising sales and partnerships for the development of applications. The sale value of advertisements is based on the number of hits, which has a cost per thousand hits.

The company has used the information obtained from Google Analytics strategically, where the prioritization of developing mobile applications is based on the operating system that incurs the
largest volume of requests. In this case, it was iOS. Other relevant decisions, such as investment in development, are made according to performance variation. Official access targets for each editorial section are used. To define access goals for each of these sections, market development forecasts have been made, considering the variations that are likely to occur in connection to major events in Brazil in 2014, such as the World Cup and the elections. Beyond official targets, access analysis promotes internal competition that benefits the organization as a whole. Identifying product access and outlining strategies to increase the number of website visits were also mentioned by the respondents. A separate study found that a product experienced higher access during the week via computer browsers, whereas the same product experienced a drop in views on weekends. It was determined that the product was accessed over the weekend via smartphones. Using this information, it was possible to change the product’s strategy by incorporating Facebook and Twitter usage. This created greater access of the product during the desired period.

Final Remarks
The capture and analysis of Web traffic has recently garnered attention from academia. However, the research findings are not easily applied in organizations. Some specific tools, such as Google Analytics, allow for decisions using relevant information. Importantly, social media has become relevant in the direction that news websites take because the proportion of hits originating from social media has grown significantly in the last year.

The main objective of this study was verifying the importance of measuring access to a newspaper company’s website and determining how that information is critical to the company’s business.

This study possesses limitations due to its use of a single case study that interviewed only two newspaper company employees. Future research should examine how the analysis of access to a company’s homepage impacts its advertising and partnership revenues. Additionally, future research should examine access analysis at other companies in the news industry to observe the types of decisions these companies make based on the results they obtain from analytics tools.

Acknowledgement
The authors thank FAPESP (Foundation for Research Support of the State of São Paulo) for their support in the publication of this work.

References
Associação nacional de Jornais (ANJ) – Imprensa Brasileira – dois séculos de história
Associação Nacional de Jornais (http://www.anj.org.br/)


