

Integrated Web 2.0 Style HTTP Referer Analytics

Stephanie Ng

Ateneo de Manila University
Unit 200, 319-H Katipunan Avenue
Loyola Heights, Quezon City
+63-2-9273925

stephanietanng@gmail.com

Ann-Gretchelle Santos

Ateneo de Manila University
58 Unit K Scout Madrinian Street,
South Triangle, Quezon City
+63-2-9267875

gretch.santos@gmail.com

William Emmanuel S. Yu

Novare Technologies – MDI Holdings Inc.
Ateneo de Manila University, Loyola
Heights, Quezon City
+63-2-4266001

wyu@ateneo.edu

ABSTRACT

Little to no attention has been given to referer profiling which is highly essential in the analysis of monitoring Web site visits. Web sites involved in e-business are specifically meticulous about where their visitors come from. This referer information may provide company Web sites with the adequate knowledge of visitor interests and what could keep them coming back. This thesis intended to pay heed to the importance of referer analysis. It thus created a Web analytics system using Web 2.0 technology with the aid of semantic Web services. Some of the finest Web services available have been harnessed not only for data mine precision but also for the expansion of the existing tools. The developers present an implementation of the work with the use of the page tagging method in the approach of monitoring visitor activity in the company Web site. Together with a scheduled task, this direct method allows for real-time data processing and the elimination of disk space and bandwidth compromise on the company's side. Visual reports through the Web analytics interface are also provided for company viewing. The usefulness of the work is realized through the company's need for a sophisticated but extensible tool that will deeper acquaint them of their visitors.

Keywords

Web 2.0, semantic Web service, referer, Web analytics.

1. INTRODUCTION

1.1 Background of the Study

With the relentless advancement of Internet technology, the World Wide Web (WWW) we know today has taken a big leap from just a mere network of communication for a small group of engineers to a universal network catering to billions of users worldwide. It is an avenue for information interchange where everyone can add to and benefit from the insurmountable data it holds. Geographical boundaries have also been bridged. Electronic business, in particular, has laid great advantage to this. Most companies, regardless of enterprise size, have established online stores catering to an international audience. eBay and Amazon have also been instrumental in bringing commodities closer to the consumer.

Even after the burst of the dot-com bubble, companies are aware that the WWW is still an indispensable venue for marketing their products and services. Hence, the concern for Web presence still prevails. Companies develop Web sites to bridge the gap between clientele and suppliers. In order to better address the latter's needs,

they harness available Web tools that analyze their users. However, current tools and technology provide limited user information. Examples of such are simple counters and user logging which no longer fulfill the needs of the companies.

Companies are becoming more aggressive in gathering timely information not only about their visitors but also about their referers. The referer or the referring page is the URL of the previous Web page from which a link to the company's Web site was followed. In other words, the referer specifies the page from which the visitor accessed the current page. It answers the question "Where do people come from?" and "What do they like visiting?"

Referers are important because they signify the community in which a particular Web site belongs to. As the old adage goes, "I can know you by knowing your friends." Referers are friendly sites. They link and draw visitors to the company's Web site. Not only are referers important aspects in getting more visitors, they are also the essential key of getting to know them and what topics interest them most. This information can help companies devise the optimal online strategy in designing their Web sites.

However, today's Web statistics focus on user profiling with little to no emphasis on referer profiling. Referer profiling primarily deals with categorizing the referer Web sites. Its variables include title, content, search engine ranking, Web site keywords, and other pertinent information. The referer area is often overlooked but as mentioned, it can be the difference between succeeding and failing online. The developers aim to supply this necessity in the Web market.

1.1.1 Benefits

1.1.1.1 Significance to Society

This thesis has noteworthy applicability in the domain of E-business. Companies benefit from receiving valuable insight and knowledge on the whereabouts and activities of their site visitors. This Web analytics is able to determine the pages both frequently and infrequently visited by the company's prospect customers. Tracking visitors' behavior helps the management devise ways in increasing usability to ensure that visitors stay engaged and return to their site.

The Web analyzer not only gathers useful information about the visitors but also increases the companies' revenue and conversion. By understanding their visitor's Web usage pattern, the companies are able to predict an accurate return of investment (ROI). By monitoring the traffic of customers, the companies may translate this information into direct sales [1].

Companies are given feedback on how well their Web site is doing rather than just having it stale online. They are not only guided to making excellent Web design decisions but also enabled to “view real data and start using numbers to make educated marketing decisions [2].” The bottom line is “understanding your customer is tantamount to making any money in business [2].”

This thesis gives emphasis to referer profiling because it is a sound and beneficial metric in addressing the common question of who visits the site. Knowing the referers present linking opportunities for companies. A mutually beneficial arrangement such as an affiliate partnership may be created with the top referers. Both Web sites benefit from the increase in amount of visitors. For companies that sell advertisements on their Web sites, this allows them to cater their advertisements to the target market. If the top referers are Web sites of luxury automobiles, then it would be wise to place advertisements intended for cars enthusiasts. Moreover, the Web analytics guides the companies in making strategic decisions on banner placement.

Through referer profiling, reciprocal linking opportunities may be explored. Linking has always been fundamental to the Web. Furthermore, in the past few years, the value of links has significantly increased. This phenomenon is attributed to the major search engines, specifically Google's PageRank. “Links have become the currency of the Web. With this economic value they also have power, affecting accessibility and knowledge on the Web [3].” Links now correspond directly to the value, more commonly known as rank, of a particular site. The more Web sites that link to a particular site A, the higher the rank of site A becomes. Not only that, if the linking Web site's rank is exceptionally high, then it directly increases the rank of site A.

1.1.1.2 Significance to Computer Science

The Web analytics system contributes a novel method of utilizing semantic Web services in developing the analytics engine. The use of the semantic Web services is a smarter form of analysis which goes beyond simple Web counters. Moreover, these semantic Web services enable the exploration of a new and in depth strategy of referer profiling. This is an edge because heavy data processing is done elsewhere rather than having it done on the server side. This method not only puts existing services into good use, but it also helps the developers build and improve on these services.

1.1.2 Scope and Limitation

The analytics tool uses JavaScript for the page tag, as will be discussed later in the paper. This poses grave problems in cases when visitors have disabled JavaScript from their browsers. This thesis also encompasses the use of existing semantic Web services but only a small subset of the available resources is utilized.

The project employs semantic Web services such as Google, Yahoo!, and MSN search APIs. It also employs software tools such as phpWhoIS and MaxMind GeoIP (IP Address Location Technology). However, utilizing semantic Web services brings about some limitations. The failure or malfunction of the Web services has a serious repercussion on the functionality of the Web analytics tool. In addition, the three search engines mentioned above limit the frequency by which any computer may

access their Web services. If the limit is reached, then the whole Web analytics is impaired.

It is also important to emphasize that the primary objective of the project is to develop a system. Whatever patterns derived from the result is not the focus.

1.2 Project Descriptions

The primary objective of this thesis is to create an integrated Web analytics using Web 2.0 technology. Multiple best-of-the-breed semantic Web services, namely Google SOAP Search API, Google Maps API, Yahoo! Search Web Services and MSN Search Web Service are utilized to gather timely and accurate information. Several software tools such as phpWhoIS and MaxMind GeoIP are also used. The Web application keeps track of visitor statistics, such as the number of unique visitors, most popular entry pages, and top referers to a site. The distinctive quality of this tool is its concentration on referer profiling and real-time generation of results.

Through the generated results of the analytics, companies have a concrete measurement of the behavior of visits to their Web site. They are able to study their visitors - who they are, where they are from, and how to meet their needs, just to name a few. Armed with these valuable information, they are able to increase the usability of their Web site. One specific benefit for the companies is that they are informed of the landing pages linked by referers. This, in turn, aids them in making strategic decisions on banner placement. Companies largely benefit from boost in traffic, sales, and revenues.

1.3 Research Questions

1. How can companies monitor the visitor activity of their Web site in near real-time?
2. How can companies analyze their visitors without compromising domain disk space and bandwidth?
3. How can referers be profiled with the use of integrated semantic Web services?

2. METHODOLOGY

2.1 Architectural Design

Semantic Web services such as Google, Yahoo!, and MSN APIs are utilized in the creation of the analytics. The Web analytics requests necessary data from these services and processes the output. Google, Yahoo!, and MSN Search APIs analyze keywords and return referer page ranks which are used to profile the Web site's referers. Also utilized are software tools such as MaxMind GeoIP - which yields the user and referer's location - and phpWhoIS - which gives detailed information about the referer's domain.

The collection of the Web analytics data makes use of the page tagging method rather than the traditional logfile analysis. Caching presents a problem in logfile analysis. The use of page tagging, then, ensures that information retrieval is both accurate and timely.

2.2 Description of Components

Apache

Apache is an open source Web server software that is used to handle Web requests and serve up Web resources. It runs mostly on Unix-based operating systems as well as on Windows platforms. It is the counterpart of Microsoft IIS. Apache is used to run the developers' PHP pages.

PHP

PHP Hypertext Preprocessor is a Web programming language that resides on the server. Developers pass data via the Common Gateway Interface (CGI) from HTML forms for dynamic content processing which includes database interaction. PHP scripts are embedded in HTML files. The integrated analytics project is a Web application therefore PHP is used for easier Web interface implementation. Specifically, the developers choose the version 5 because most, if not all, of the Web services are compatible with the latest PHP version.

MySQL

MySQL is a widely used relational database management system (RDBMS) that utilizes Structured Query Language (SQL). It is the most popular open source database because of its consistent fast performance, high reliability and ease of use [4]. The developers of the Web analytics employ this database for the storage of valuable client, user, and referer information.

JavaScript

JavaScript is a scripting language interpreted on the client's Web browser. The script is embedded into a Web page which is not run on the server side, but on the client side.

2.2.1 Page Tag

The project's method for acquiring the log of a particular Web page is page tagging. A snippet of JavaScript code, known to 'tag' the visitors, has to be embedded in the client's Web page.

2.2.2 Tracker

An Apache Log Format workalike is utilized to track the hits of a Web page. The page tag inserted into the client's own Web page invokes this Javascript code residing on the server. This code, called the Tracker, gathers the necessary user and referer information returned [5]. Examples of such are entry URL, referer URL, screen resolution, color depth, and visitor IP address.

2.2.3 Logger

The Logger is a short PHP code whose main function is to store all the information collected by the Tracker into the Job Queue.

2.2.4 Job Queue

The Job Queue is simply a database where basic information gathered by the Tracker is stored. These data are accumulated for later processing.

2.2.5 Scheduled Task

To achieve timely information, data gathering and data processing are executed independently. Thus, data filed in the Job Queue are scheduled for processing at a five-minute interval.

2.2.6 Analytics Engine

The Analytics Engine is responsible for pooling collected data from the Job Queue. It consolidates these data to retrieve even more information such as title, metatags, and category of the referer. The detailed records are then entered into the main database. This is where the referer profiling takes place.

2.2.7 Database

A database of only five tables is used in the implementation. The tables include: client, logs, visitor, referer, and keyword information. The *client* table contains the company's account information in order for the system to be able to determine which site to keep track of. The *logs* table contains the date of access to the site made by a visitor and the entry URL (landing page) he visited. The *visitor* table holds the IP address and the browser information of the user. The *referer* table contains the URL, base url, IP address, title, and category of the referer site. On occasions when the client's Web site is directly typed and accessed onto the URL address, the referer data is empty. The *keyword* table contains the keywords of each referer obtained through the referer profiling process. The following is an entity relationship diagram (ERD) of the developers' database.

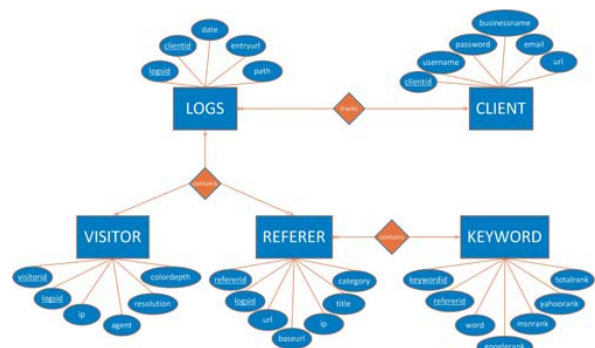


Figure 2.1. Entity Relationship Diagram.

2.2.8 Analytics Interface Engine

The Analytics Interface Engine organizes records from the database into useful statistical information that are instantaneously displayed on the Analytics Web Interface. Statistics are obtained through SQL Database Manipulation Language (DML) or queries made to the database. This solution is performed in PHP and is responsible for determining information such as the top entry pages, referers and other pertinent visitor information from the information system or database.

2.2.9 Analytics Web Interface

This Web interface component is the result of the seamless combination of multiple semantic Web services and software tools. Clients are able to register their Web site in order to avail of the integrated Web analytics. One client is restricted to enrolling only one URL. They may log on to the Web application to view the analyzed data of their Web site's user behaviors and activities. With the provided at-a-glance reports, the clients are able to take appropriate actions for the improvement of their site performance.

2.2.10 Semantic Web Services

The developers use the following semantic Web services in the retrieval of the categories and other pertinent information of client's referers. Google, Yahoo!, and MSN searches provide different sets of data. Integrating them generates more reliable results. The developers aim to ensure that the information retrieved is accurate. If one search engine fails to generate a result set, the two others may compensate.

2.2.10.1 Google SOAP Search API

The Google API is a free experimental Web tool that developers can use to find and manage information on the Web [cite{google}]. It is a service in which Google offers their resources to developers and researchers for the latter's own application development. The software applications created by the developers remotely connect to the service through an XML-based system, SOAP.

2.2.10.2 Google Maps API

The Google Maps JavaScript API lets developers embed Google Maps in their own Web pages. These maps are customizable to the needs of the users. Markers and lines can be drawn on the map to specify certain locations.

2.2.10.3 Yahoo! Search Web Services

Like Google API, Yahoo! Search Web Services allows access to Yahoo content and services. However, unlike Google, Yahoo! uses REST instead of SOAP. Yahoo! claims that it has a lower barrier to entry, easier to use, and entirely sufficient for their services in contrast to SOAP. Yahoo! Web Service is also language independent which gives developers freedom and ease in integrating the service into their application [7].

Under the umbrella of Yahoo! Search Web Services, three data sources are particularly utilized. First is the Web Search service that allows developers to search the Internet for Web pages. Second is the Term Extraction service which provides a list of significant words or phrases extracted from a larger content. It sifts out all the irrelevant words and returns the keywords for the content passed. It uses the very algorithm Yahoo! Search engine uses to rank pages. Third is the Site Explorer service which provides access to information on individual sites. The InLink data service is a subset of this and gives Web site owners a good grasp of who is linking to their site. It retrieves the information about the pages linking to a particular Web page.

2.2.10.4 MSN Search Web Service

MSN Search Web Service, like Google, also uses SOAP. It also allows users access to some of MSN's services. It also allocates a separate search query quota to each IP address allowing third-party users to use developed applications at a more extensive level [8].

2.2.11 Software Tools

The developers also employ several open source software tools to provide clients more details regarding their Web sites. Given the IP address, MaxMind GeoIP determines geographical location and phpWhoIS retrieves the domain owner information.

2.2.11.1 MaxMind GeoIP (IP Address Location Technology)

Geographic location of a website or website visitor is identified real-time by tracking the user's Internet Protocol address. GeoIP determines the country, region, city, postal code, and area code of the visitor as well as provide information such as longitude/latitude, connection speed, ISP, company name, domain name, and whether the IP address is an anonymous proxy or a satellite provider [9].

A new "Know Your Customer" law has been implemented by regulatory entities in both US and Europe. As a result, banks, software vendors and other online enterprises are subject to compliance, thus providing a more secure environment. Geolocation is used as an investigatory tool by the security teams to track Internet routes of online assailant and prevent future assault from the same location. Hosts of live video streaming such as internet movie vendors and online broadcasters are able to monitor their viewers on licensing regulation conformance. This technology is very functional in several industries, including e-retail, banking, media, online gaming and law enforcement, for preventing online fraud, complying with regulations, and managing digital rights. It also provides location-based content such as the user's language, currency, and pricing.

GeoIP obtains its dataset from the user's area or zip code information entered upon the filling out an online form. The data is then run through a series of algorithms that acquires sets of IP addresses in the particular location. The next time the site is visited by the user's neighbor, GeoIP already guesses his/her location.

This is utilized in identifying the referer's location, specifically country, city, longitude and latitude, in this project's attempt to be an integrated analytics.

2.2.11.2 phpWhoIS

phpWhoIS is a software tool that identifies the owner of a domain given its URL. This is used for additional information gathering.

2.3 Implementation Details

Clients first need to embed a JavaScript code, called a page tag, into every page of their Web site that they wish to track. The page tag invokes a tracker, another JavaScript code responsible for gathering basic referer data such as referer URL, entry URL, browser resolution, and color depth. The tracker then passes the referer data to the logger, a PHP code, of which its only task is to store these data into a Job Queue for later processing. Data in the Job Queue does not take long, however, since the records are processed every five minutes. A component relationship diagram is shown in Figure 2.2.

The analytics engine processes the information from the Job Queue. It also acquires more information about the referring Web site. It gets the meta tags of the referer Web page and parses the actual contents to obtain its title and body.

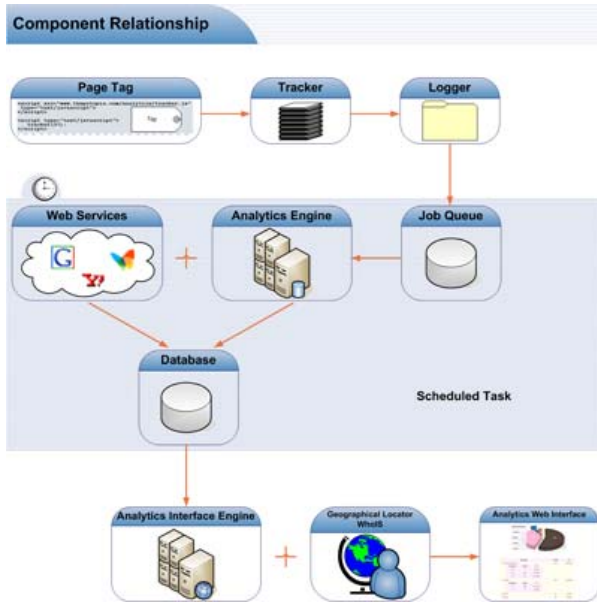


Figure 2.2. Component Relationship.

The engine also executes the first phase of the analysis process which is the referer profiling. Referer profiling simply means categorizing the referer Web sites. However, this is the most complex part of the implementation process. Keywords are the vital determinants of categories. Three different sources are mined to resolve the keywords: meta tags - specifically the keywords and descriptions, the title, and the body itself. The combined text of the sources is passed to the Yahoo! Term Extraction Web Service which returns the list of significant words and phrases in order of importance. The top ten words and phrases are the keywords. Each one is run through the three search engines. The search engines will return the position or rank of the referer Web site based on the keyword used. The site may be any from the first through the 100th search result displayed. Google, Yahoo, and MSN all rank the pages differently. The developers seek to ensure accuracy which is why all three are utilized. The numeric ranks are added to get the total page rank for the particular keyword used. After all the keywords have been processed, the total page ranks are compared. The least, that is, the most popular, keyword will be utilized as the category.

All the mined information are then entered into the database. This real-time process guarantees accuracy of the results. Once the data gathering and profiling is done, the database is populated.

The Analytics Interface Engine executes the second phase of the analysis process. It simply evaluates information previously filed in the database by organizing them for display. It accomplishes this task simply by SQL queries.

The clients can log in to the Analytics Web Interface or the Web application to view their Web site statistics. From this point, Web Services and software tools continue to process and return data. The geographical locator gathers the country, city, longitudinal and latitudinal information from the IP addresses stored. Google Maps API gives a visual perspective to these geographical locations. Special emphasis is given to the referers, their entry pages, and possible banner placement.

Everything is developed using PHP scripts understood from Google, Yahoo!, and MSN APIs and other tutorials found online.

3. RESULTS AND DISCUSSION

3.1 Near Real-Time Monitoring of Visitor Activity

The approach for companies in monitoring the visitor activity of their Web site is through page tagging. The task is as simple as inserting a snippet of JavaScript code in every page of the Web site that the company wishes to submit for Web analysis. The code then invokes another JavaScript code, the tracker with the company's user ID as its request parameter for identification. The company's Web page containing the page tag must be rendered for the page tag to execute and invoke the tracker script to capture data. The tracker gathers basic referer data which is passed to a PHP code, the logger. The logger is responsible for storing the data amassed into a job queue.

Although data in the job queue is stored for later processing, it does not take long since the records are processed every five minutes. Five minutes is near real-time. A shorter time frame will cause too much overhead and overlapping in the database.

The following is a sample of the page tag script (Figure 3.1) that is returned to the user upon registering for an analytics account. Following it is the scheduled task (Figure 3.2), specifically the cron daemon, that configures the job queue into processing every five minutes.

The page tagging that the developers use is specifically designed for near real-time processing. In addition, not all

```
<script src="www.thepctopia.com/analytics/tracker.js"
  type="text/javascript">
</script>
<script type="text/javascript">
  tracker(37);
</script>
```

Figure 3.1. Sample Page Tag Script with Company ID 37

```
*/5 * * * * /usr/bin/php /home/thepngl/public_html/analytics/process.php
```

Figure 3.2. Scheduled Task Script Used to Trigger Job Queue Every Five Minutes

information that may be acquired from Web services are stored, but are rather retrieved when the user views the reports to ensure the timeliness of data.

3.2 No Compromise in Domain Disk Space and Bandwidth

In contrast to a hybrid scheme of collecting Web analytics data through combined logfile analysis and page tagging, dedicated page tagging is employed in the implementation of this thesis. It does not make use of the conventional 1x1 image requested by the page tag thus creating an entry in the client's logfile. Moreover, Web server logs are not designed to mine data but simply to debug Web servers. They are flat files on multiple file systems and even stored in different time zones. It is even

more complicated if Web servers are distributed geographically in different time zones, especially for large sites that have multiple servers - each logging data into separate files and on different file systems - since combined data requires it to be in one time zone (usually GMT). Web server logs also store large amounts of structured data inefficiently because they are typically in ASCII. Redundant and irrelevant information may also be contained in Web server logs. This is evident in every request for images found in a page. Using server logfiles for analyzing visitor activities is therefore inefficient [10]. Logfile analysis also poses inaccuracies due to browser caching. If a visitor revisits a page, the second request is oftentimes retrieved from the browser's cache so that request is received by the Web server only once. This traditional method does not only deposit large amounts of data into the client's logfile and waste disk space but it also leaves the data for later analysis.

Dedicated page tagging allows companies to gain reports of visitor profiles without compromising their domain disk space and bandwidth since there is neither client nor server-side processing involved. Rather, computation and analysis are remotely done on the developer's technology and stored in a dedicated domain thus saving the user from expensive space that can be better used in storing their own data. The job queue, in particular, makes the task possible because background processing is done every five minutes. As a result, the clients outsource the analytics system and by doing so, they free up their own system.

Another advantage for the analytics to be located external to the company's domain is that the reports need not be restricted to the viewing of the Web site developer. The analytics caters its organized reports to the owner, manager or any marketing contact of the company. These reports are statistics derived from SQL queries made to the database.

3.3 Referrer Profiling with the use of Integrated Semantic Web Services

Referrer profiling simply means categorizing the referer Web sites. Keywords are the vital determinants of categories. In order to establish the profiling procedure, the analytics engine obtains the necessary information about the referer. The information consists of the meta tags, the Web site title, and the body itself. These three are integrated in order to produce a collection of texts to be passed to the Yahoo! Term Extraction Web Service. The Term Extraction Web Service returns the list of keywords arranged in order of importance. To limit the scope and to expedite the processing function, only the top ten keywords are considered. Each keyword is run through Google, Yahoo, and MSN search Web services. They return the rank of the referer Web site based on the particular keyword searched. Since the three search engines all have different algorithms and distinct results, all are utilized to ensure accuracy. The sum of the three ranks is the total rank for that keyword. After all keywords have undergone the same procedure, the keyword with the highest rank is deemed as the category.

Using best-of-the-breed semantic Web services is a novel idea which places heavy data processing particularly on Google, Yahoo, and MSN. It is quite advantageous because the developer's server becomes even more dedicated to the Web analytics itself. It is able to give more processing power to gathering information and producing the client Web site reports.

4. CONCLUSION

In this thesis, the developers have discussed the importance and benefits of Web analytics. The analytics built by the developers has a particular emphasis on referers. Analyzing referers is an advantage in knowing a visitor's background of interest, thus gaining knowledge on what types of visitors frequent a Web site. The construction of the analytics has also been possible with Web 2.0 technology. Harnessing an array of semantic Web services has not only allowed a more precise data yield but has also extended the development of the services. This method has also kept heavy data from being processed on the client or on the server's end. Although the software produced uses a variety of semantic Web services, the implementation is not a Web service itself, but rather, a consumer of it. In principle, the accomplishment is part of the Web service eco-system, if termed as such. The task has also been to create a tool - although not any ordinary tool - the referer Web analytics tool.

Companies can monitor the visitor activity of their Web site in near real-time. It has been conferred that data is processed and stored near real-time thus preserving the integrity of the information. The use of the technology, in consequence, will give users access to timely referer profiles. This will thus inform company users of strategic banner placement, for example. Companies can also analyze their visitors without compromising their domain disk space and bandwidth. There is neither client nor server-side processing involved because clients outsource the analytics system and by doing so, they free up their own system. These have been possible through the page tagging method and the scheduled task. Referers can be profiled with the use of integrated semantic Web services. The three search engines employed in the development all perform different algorithms and distinct results, all of which are used to ensure accuracy.

The developers have undoubtedly presented a sophisticated, yet extensible Web referer analytics with the stated features. It is without reservation that a tool such as the current implementation is vastly essential in e-business today.

5. RECOMMENDATIONS

The work has been an initial attempt at referer profiling in Web analytics. Although the technology is sufficient in its right, the developers propose additions to the work to further extend its possibilities. This includes adding wider ranges of Web services and accordingly tracking more information. Security issues must also be resolved such as those involved in the page tag. The user id must be encrypted, for example, and the logger script information must not be disclosed.

6. ACKNOWLEDGMENTS

We thank God who has gone with us in the creation up to the completion of this thesis. We thank Him for giving us wisdom and strength. We also thank the Department of Information Systems and Computer Science of the Ateneo de Manila University and our thesis adviser Mr. William Emmanuel S. Yu for their constant support during the entire process of this thesis.

7. REFERENCES

- [1] THINKMETRICS. Thinkmetrics web analytics - providers of website analysis and statistics. Available from <http://www.thinkmetrics.com>.

- [2] WAA. The web analytics association. Available from <http://www.webanalyticsassociation.org/>.
- [3] WALKER, J. Links and power: The political economy of linking on the web. ACM Hypertext Conference 2002, ACM, ACM Press, pp. 78–79.
- [4] MYSQL. Mysql ab :: The world's most popular open source database. Available from <http://www.mysql.com/>.
- [5] ATTERER, R., WNUK, M., AND SCHMIDT, A. Knowing the users every move user activity tracking for website usability evaluation and implicit interaction. In World Wide Web Conference (Edinburgh, Scotland, May 2006), ACM, ACM Press, pp. 203–212.
- [6] GOOGLE. Google soap search api. Available from <http://www.google.com/apis>.
- [7] YAHOO! Yahoo! developer network - frequently asked questions. Available from <http://developer.yahoo.com/faq/>.
- [8] MSN. Live: Why msn search? Available from <http://msdn.microsoft.com/live/msnsearch/whysearch/default.aspx>.
- [9] MAXMIND. Maxmind - geoip — ip intelligence solution. Available from <http://www.maxmind.com/app/ip-locate.htm>
- [10] IMEDIA CONNECTION. Web analytics 101. Available from <http://www.imediaconnection.com/content/4509.asp>.