

COMBINING WEBSITE SEARCH ENGINE OPTIMIZATION WITH ADVANCED WEB LOG ANALYSIS

Dromey, Seamus, National University of Ireland - Cork, O'Rahilly Building, UCC,
Ireland, sdromey@bismail.ucc.ie

Heavin, Ciara, National University of Ireland - Cork, O'Rahilly Building, UCC, Ireland,
c.heavin@ucc.ie

Neville, Karen, National University of Ireland - Cork, O'Rahilly Building, UCC, Ireland,
kneville@afis.ucc.ie

Abstract

This paper provides a clear guideline to the development of an online decision-making tool. The importance of ranking for an organizations virtual presence through search engines is also discussed. The system described illustrates the complexity of the competition between organizations to be highly ranked by leading search engines. The system not only reports the rankings of the owners but compares an organization with its competitors and enables it to decisively formulate an online development strategy in improving its ranking and therefore increasing its audience or critical mass. The system (Googalyser) utilizes Web logs and content analysis to provide decisive information to Web developers in order to improve the cases ranking through for example www.Google.com.

Keywords: Decision-making, Web Log Analysis, Competitive Advantage, data mining, Web ranking and development

1 INTRODUCTION

Organizational change in today's competitive landscape is no longer just an option, but a fundamental strategy for success (Neville and Heavin, 2004). "*Successful firms no longer just add value, they invent it*" (Chattel, 1998; pg. 59). In order for firms to do so they must reverse the traditional value chain thinking by which businesses define themselves in terms of the products and services that they produce (Porter, 1985). However, the creation of value in the digital era requires a reversal of this value chain thinking, as for firms to succeed on-line they need to find new and unique ways of doing business, often challenging accepted ways of trading in the process (Fingar, 1998). This reversal of thinking reinforces the rationale behind organizational centric sites. In this particular value chain model information is seen as a critical element in value creation (Rayport & Sviokla, 1995 pg. 174). The exploitation of the virtual value chain has become a digital based value network for enterprises to reach out through technology to their customers, suppliers, and even competitors (Lawrence *et al.*, 1998). In order to develop a close relationship with these groups, it is essential to have a unified view of the target audience (Chase, 2000) to personalize the Website to their customer needs and to compete (Toffler, 1991). Integration is the key to creating this '*unified view*' the information gathered through customer interaction channels can be shared throughout the organization. Key data can and is derived from Web traffic to determine online customer behaviour or Web-click analysis (Greenstein *et al.*, 2000) and ultimately to pull customers away from competitors to a particular site. To facilitate this value added approach to strengthening an organization's relationship with customers a strong Web presence is vital in communicating the strategy employed by an organization. This paper focuses on the development of a system that utilizes Web logs to aid managerial decision-making in an attempt to increase its online ranking through leading search engines, as approximately eighty-five *percent* of browsers discover Websites through search engines. This research outlines the different strategies that can be employed to secure a higher search engine ranking and therefore a higher customer audience. It also highlights the potential of the system to overcome the virtual barriers in increasing an organizations presence and even to analyse a competitor's strategy towards search engine optimization. The system can, when properly structured and maintained, facilitate any type of organization from an educational to a financial case. Further research would also indicate the justification of the Googalyser in increasing the online presence of any organization.

2 THEORETICAL FOUNDATION

Shanks and Tay (2001) define information quality as '*fitness for use*'; customer information should be viewed as a corporate asset that can be used to deliver competitive advantage and support business initiatives to focus on the customer. It is now seen as critical to use customer information in order to meet the challenges of increased competition and to compete in markets that were once driven by product offerings, but are now dictated by customer needs and preferences (Greenstein *et al.*, 2002). However, customer information used needs to be of a high quality if a system is to work to its full potential and produce the desired results. Maoz (2001) emphasises the critical importance of quality information to the success of an online customer initiative when he states that "*while a centralized customer data repository is a critical foundation.... it is the quality of data that will determine the level of success.*" It is extremely beneficial to the organization in question as high quality customer information is very difficult to obtain, but has a direct impact on cost reduction, revenue generation and competitive advantage (English, 1999). Therefore quality information is vital in order to support the long-term relationships with customers, and maximize the competitive advantage that can be gained from

customer data (Shanks and Darke, 1998). However Crowder *et al.*, (2001) points out that dealing with the data deluge is a dilemma, as it comes fast and furious from online channels. The development of organizational technologies such as data mining tools has alleviated the issue. Data mining technology is capable of discovering patterns, associations, changes, anomalies and significant structures from large amounts of data stored in databases or other information repositories (Hui and Jha, 2000). Intelligent, in-depth analysis requires consistent customer data as a starting point, with all enterprise applications participating in the analytic environment (Magic Software Enterprises, 2000). It is important to note that added value is only achieved when the analysis results are fed back to management and throughout the organization. It is then that enterprise decision makers must leverage this knowledge to make more informed and timely business decisions. Data mining tools are a specific class of computerised information system that support decision-making activities (Ha *et al.*, 1998). “*They condense large amounts of data into a form in which they can be analysed*” (Laudon *et al.*, 2004; p 45). Data mining has evolved to accommodate the growing need for data at the strategic level of a business unit (Laudon *et al.*, 2004). This has evolved into sophisticated systems with Executive Support components to generate sophisticated reports and graphics to summarise key queries (Adam *et al.*, 1998). They serve to “*...isolate and present the relevant information*” (Adam *et al.*, 1998; p 5) needed to support organizational decision-making. The drill down facilities and integral intelligence of the systems are essential in delivering the reporting that matters and identifying cause and effect relationships for an organization. The importance of this type of analysis has increased significantly with the advent of customer driven strategies. The strategic advantages of implementing this technology in an organization is becoming increasingly evident, as there is considerable competitive advantage to be gained by utilizing a data store if implemented and used effectively (Sammon *et al.*, 2003). Today, virtually all organizations in a broad range of industries utilize some form of decision support system (Shin *et al.*, 2002) to discover customer patterns through both on or offline interactions. The technology under consideration in this paper acts as an online data mining tool that could enable organizations to enhance the effectiveness of their Web presence through the analysis of customer movements online.

2.1 Broadening Web Presence

Web technology can be used as a tool for disseminating information about products and services, facilitate communication or to directly integrate with the core business processes and transactions (Nambisan and Ming Wang, 2000). However, it is not always the first to market that will succeed but in many cases those companies who are first to achieve a critical mass (Aubert & Hamel, 2001). In order to understand how businesses can best utilize the Web it is useful to listen to customers and adapt to changing customer needs (Kalakato and Robinson, 1999). Being the best often involves reinventing business processes and raising relationships with customers and suppliers to unprecedented levels of intimacy. The building of such levels of intimacy, as cited by Tapscott (1995, pg.87) “*has allowed enterprises to reach out through technology to customers, suppliers, affinity groups and even competitors*”. The type of presence chosen by firms to adapt to their particular business is determined by their level of commitment to the potential of the Web to create business value. By failing to use the interactive capabilities of the Web companies will never be able to generate return from the benefits of the Web and turn their investment into a profit maximising opportunity (Chattel, 1998). This technology permits information to flow in both directions between the customer and the company creating the feedback loop that integrates the customer into the organization (Greenstein *et al.*, 2000). This integration allows a company to own a particular market and permits customization and dialogue between the two (McKenna, 1991; Chattel, 1998). As noted by Chard (1999 pg. 3) “*...communication is very important to the process of building trust relationships with clients and for decision-making...*” Adopting a Web presence is no longer a new strategy but an accepted norm. With such input and knowledge of

customers, firms are able to produce goods and services to suit the particular needs of different market niches through customization, which is an underlying motivation of firms in establishing and maintaining a loyal audience and business centric sites (McKenna, 1995) as it can be up to six times more expensive to attract a new customer than to retain an existing one (Rosenberg & Czepial, 1984). Therefore, the development of a Web presence or system is vital to any organization and should be approached methodically to align its function with the focus of the business.

2.2 Developing a Web Presence

The Internet Commerce Development Methodology (ICDM), unlike other methodologies, provides the Web development process with a business focus (Standing, 2001). The first step of the method is the Web Management Structure, which is subdivided into three separate tiers. The first tier is the meta-development and management tier, the second tier is concerned with the development of the components of the Website and the third deals with the development and implementation of the system. The second of the components and features of the ICDM is the strategy development phase. This phase requires the assessment of the organization and its environment using a competitive analysis technique known as a SWOT analysis (Strengths, Weaknesses, environmental Opportunities, and Threats) (Thompson and Strickland, 1995). The third feature of the ICDM is user involvement, which is vital at the strategy development and analysis stages in eliciting requirements of the system, identifying barriers to using the Web and evaluating design issues. The fourth component of the ICDM is the meta-development strategy, which outlines the *'amount of regulation or control that is desired, both for content and design'*, of the Website/presence and is decided upon by the Web management team (Standing, 2001, pg. 553). Once the meta-development strategy has been finalized the next stage is the site and component development. The next stage of the ICDM is the analysis phase, which utilizes requirements analysis techniques and a functional requirements framework to obtain both the logical and functional requirements of the system. The next component is the physical architecture framework. Standing (2001) states that *'there are three fundamental types of Web systems: document publishing systems, basic interactive systems, and complex transaction systems'* (pg. 554), which influence the techniques that are used to define the requirements of the system. The next stage is the implementation phase, which relates to the implementation of the system. The final stage of the ICDM is the evolutionary phase, which reflects the continual and evolutionary nature of Web applications, which need to be managed by the Web management team (Standing, 1999 & 2001). In this paper we present the development process of the 'Googalyser' and discuss how this type of application can act as an online data mining tool that has the potential to provide a wide variety of organizations with strategic information that can generate and support their competitive edge in the marketplace through enhancing the customer experience and evaluating different development strategies with recommendations for improvement. The tool mines through Web logs to discover patterns in browser traffic, in addition to analysing Website content in order to optimize keyword usage. The system provides management with information to aid in deciding the most effective design and content of the corporate Website in increasing its search engine ranking and therefore performing a SWOT analysis of different design strategies, including that of its competitors.

3 RESEARCH OBJECTIVE & APPROACH

This research study outlines the development of a decision-making tool for any type of organization utilizing a commercial Website. Essentially, the system or 'Googalyser' mines through volumes of current and historical data to aid managerial decision-making regarding the formulation of an online strategy. The paper illustrates the reversal of the traditional 'value chain'

highlighting that information, not the service or the product, is vital to the success of an organizations strategy to maximize its online presence. The system not only monitors Web traffic it analyses key user usage (for example customer demographics), optimizes search engine rankings and tracks the owners as well as competitor’s online ranking through up-to-date reporting. The study produced a data analysis tool that monitors Web logs and site content to provide a competitive advantage to the owner of the system. The researchers sought to produce a dynamic as well as a generic tool to mine through seemingly valueless data to discover value added knowledge regarding an organization’s Website as well as that of its competitor and evaluate one in comparison to the other. The researchers examined the development of the Googalyser to aid in the design of the different Websites and have also begun to monitor the initial results generated by the tool in two different environments: an academic Website and an investment management site. The researchers briefly refer to these recent test cases in section 6. The authors investigated the degree to which the benefits of the system met the requirements of the current leading search engines, particularly Google.com. Figure 1 illustrates the data-mining tool that has been created to support managerial decision-making as well as internal Website developer’s requirements, as determined through ongoing discussions and Web log analysis.

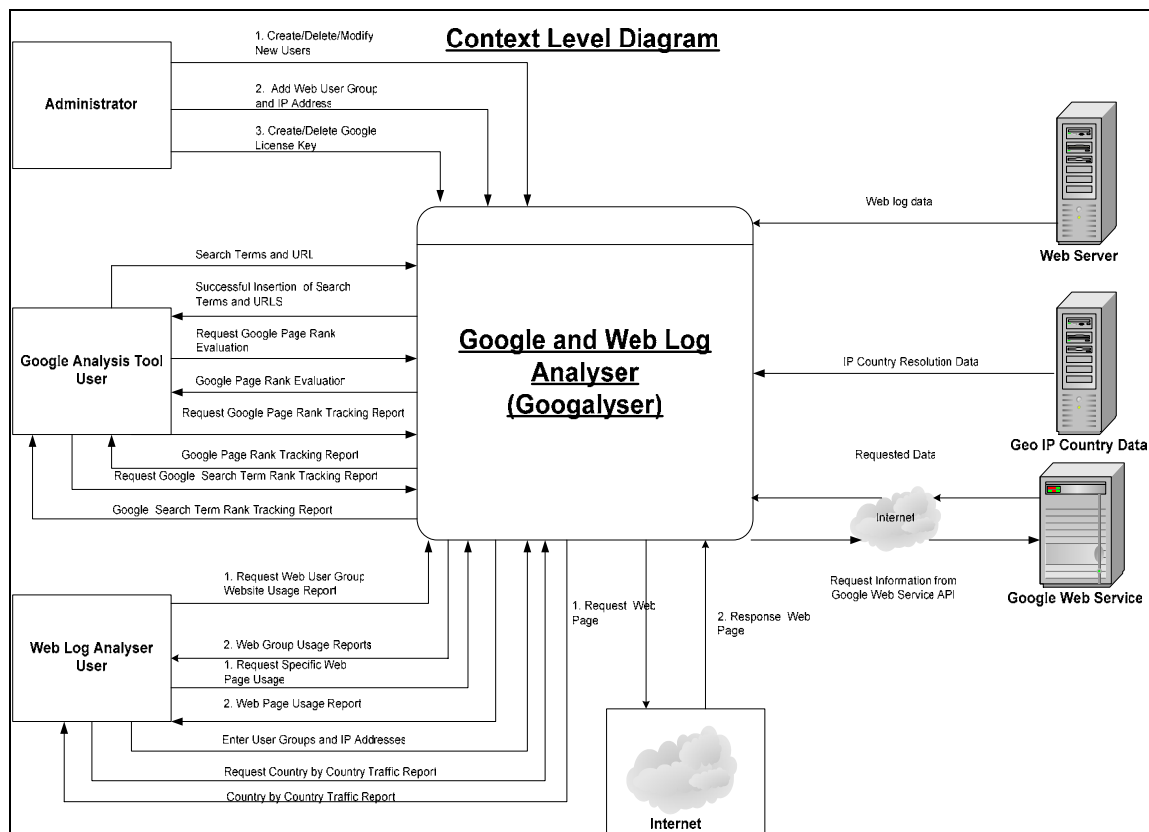


Figure 1 Googalyser Context Diagram

However, both the approach and objective of the study has limitations due primarily to the secrecy of the algorithm used and designed by Google to evaluate or determine the ranking of individual Websites and key search terms. Therefore, the researchers formulated assumptions to guide the development of the tool identified from sourced optimization sites as outlined in Section 4. In addition, further studies must be carried out in the test case environments as this research only outlines initial feedback from the systems implementation.

4 THE ANALYSIS – GOOGLE.COM

Google is the most widely used ‘spidering’ search engine operating on the Internet. It receives over two-hundred and fifty million search engines a day and it is the leading search engine as it runs on a unique combination of advanced hardware and software (www.google.com). The success of the search engine can be attributed to the efficiency of their search algorithm. Not surprisingly Google are less than generous in publicly making their algorithm available, which determines the ranking of key phrases/words and therefore Websites (see Figure 1). Some important factors have been established that can allow Web designers to manipulate their Websites in order to improve their ranking. Firstly, Google examines and analyses Web link structures (see Section 2.2). This is believed to be a critical determinate of the Google Page Rank. It examines the number of pages linked and interprets a link from Page A to Page B as a vote by Page A for Page B. However, it also analyses the pages that cast the vote, votes cast by pages that are themselves significantly weighted more when evaluated and therefore help to make other pages significant. This allows Webmasters a certain amount of scope to create internal links within their sites or from affiliations with other Websites to improve page rankings. It is believed that high quality sites will receive a higher Page Rank, which Google (or any other engine) notes each time it conducts a search. However, page structures are meaningless to users if the pages do not match the search term entered. Google therefore combines Page Rank with sophisticated keyword matching techniques to find pages that are both important and relevant to the users search. The researchers identified this factor as a requirement that could be fulfilled by a data-mining tool. Through the analysis of the content of pages that Google ranks highly for a particular search term and the pages that a Web master wishes to be ranked highly for that search term, it is possible to derive the areas where a Web page’s content can be modified and thus a higher ranking can be achieved for a particular search term. The researchers investigated a number of sources such as search engine optimization sites and as a result compiled a list of factors, which are essential for a Web master to easily compare and analyse pages competing for top position in the Google search engine through an automated method. These factors include keywords in the ‘URL’ (domain name), ‘Title’, ‘Headings’, ‘Body text’ (especially near the beginning), ‘Meta tags-descriptions’ and keywords, in the ‘alt tag’, ‘text links’, ‘external links’ (backward links) and finally in the ‘Common theme’ throughout the site. The only way to access this information for the proposed system without breaking the Google Terms of Service is to use the Google API Web service. The service supports several SOAP (Simple Object Access Protocol) methods; which are described in Google WSDL (Web Service Description Language) files. This provides a formal description of a Web service that is imported into a Web programming development environment. If an organization is not ranked in the top ten search results for a search phrase it is seven times less likely to receive a hit from a Web browser using for example the Google Search engine. So obviously search engine optimization is a fundamental factor for any organization with an online presence to achieve a critical mass. However, one final query needed to be addressed by the researchers prior to systems development, which was the consideration of purchasing a link in the ‘sponsored links’ of search engines, in order to theoretically guarantee a high search ranking. Primarily this approach, excluding the high cost as links are auctioned, is not adopted by the majority of organizations because organic links (links that are not purchased) in search engines tend to receive a much higher ‘click through rate’ than sponsored (paid for) links from Web browsers. Clearly organic search engine optimization is the optimal approach to undertake when attempting to achieve a competitive advantage in this increasingly competitive e-business environment. The researcher’s initial idea grew from the need to develop a system that could contribute to achieving an easy to use search engine optimization tool (combining a data-mining backend with a Web front-end). Initially to identify the requirements of the system the researchers performed a detailed use-case investigation in the early stages of the development. This entailed determining the systems functions regarding business events such as the current

online strategy of both test cases and therefore the requirements of the different cases. From this research, it was possible to outline the actors (anything that interacts with the system to exchange information) and high level uses cases (major behaviourally related sequence of steps, both automated and manual, for the purpose of completing a single business task). This allowed the researchers to construct a context diagram to define the scope and boundaries of the system (See Figure 1).

5 DEVELOPING THE GOOGLE ANALYSER

The aim of Googalyser is to provide an organization with commercial Web presence vital information that will allow them to gain a competitive advantage in their respective markets by aiding the organization in performing search engine optimization. Through the provision of vital information that can be derived from Web traffic to an organization's Website and the source of this traffic the system allows managers and developers to identify both the strengths and weaknesses of the corporate Website and identify the opportunities and threats by monitoring design changes and key user traffic through web log analysis. The Googalyser was constructed to support all of the users within both cases and to aid in the optimization of different Websites. The system is designed or customized for the requirements of any type of organization, Web designers and management by providing an integrated view of external/internal Web browser usage. The utilization of the Web varies as illustrated by the test cases, however, the same factors determine their ranking. The system enables any organization to determine the ranking of their Website and to provide data regarding its success or failure in attracting customers/browsers. Figure 1 illustrates the scope and boundaries for the Googalyser in the context of the requirements identified through the test cases and the assumptions made regarding the Google evaluation methods. To simplify the development process the context diagram illustrated one process. External agents are drawn around the perimeter. Data flows define the interaction of the system with the boundaries and the external data stores as determined through the interviews carried out in the test cases. An external agent is defined as a person, organization unit, other system or other organization that lies outside the scope of the project but interacts with the system. External agents provide the net inputs into the system and receive net outputs from the system. From the researchers analysis the main external agents interacting with the system are as follows: (1) Web Log Analyser User – this external agent will typically be the system hosts organization's webmaster. Their primary aim would be discovering information regarding the organizations web site traffic/usage. (2) Google Analysis Tool User – this external agent is any authorized system user who has a role/interest in the host organizations search engine optimization activities. The system provides in real time intelligent search engine optimization suggestions for any page on the organization web site, subsequently the user can monitor any web pages ranking for a host of key search phrases. In addition it is also possible to track how organizations web pages are ranked against direct competitors web pages for key search phrases. (3) Administrator – the role of the Administrator in the context of the system is to oversee system user management and web user group management. *Note:* A web user group in effect is a grouping of IP addresses that are considered to be representative of the computers used by a particular group of users when accessing the organizations web site. (4) Google Web Service -this is the service that provides the system with the capability to perform an automated search of Google through a software program. (5) Web Server - the system retrieves web log data for analysis from the Web Server on a daily basis. (6) Geo IP Database – the system interacts with this agent to map an IP Address to its country of origin. This information is used to generate reports on the distribution of an organization's web traffic on a country-by-country basis.

5.1 Googalyser Functionality

The following section outlines the key features that the Googalyser offers to the participants. Figure 2 depicts an overview of these characteristics, which were identified as a result of extensively scoping the functional requirements of this data-mining tool. A fundamental enabler of key system functionality is the ability to monitor website traffic. The primary function of the tool is to deliver advanced search engine optimization capabilities, in direct association with detailed web log analysis. The system has been developed to run a backend process on a daily basis to analyze the previous day's Web log file to determine the number of hits received to each Web page on the organization's Website. In addition, the system also automatically detects new pages added to the organization's Website and determines the level of traffic to these pages. The process also determines the country from which each Web page request was made and records this information. To derive value from this functionality a number of reports are generated to aid decision-making: (1) Individual Web Page Traffic Report: this report allows the user to view the number of requests per week to any page in the organization's Website. This report is used in determining which pages in the Website receive the highest levels of Web traffic at different periods in time. (2) Country-by-Country Website Traffic Report: this report allows the system user to view the countries from which the highest numbers of requests are made to your organizations Website between any two dates. This information can have a number of positive impacts for an organization. Examples include: the ability to focus marketing campaigns at countries with particularly high levels of interest in the organizations Website. Alternatively, if an important marketplace is outside the top ten then this information again can be used to target a particular country. Additionally, the added possibility to tailor a Website's content to match that of a "strategically important" country where a low number of hits are currently being received provides an organization with a competitive advantage in targeting key browsers. For example a graphics design company, could attract an American audience to the site by modifying words such as 'colour' to 'color'. The second requirement, as identified by the researchers, is to *analyze key user-group site usage* through the submission of the IP Addresses of Key Web User Groups to the system. The backend daily summary process then monitors the organizations Web log files on a daily basis to determine the level of Web traffic generated to the organizations Website by these key web user groups. This requirement also produces a report (Web User Group Site Usage Report) to allow managers to view the different pages on the organization's Website that a key user group accesses the most frequently. For example, if the Webmaster of the educational case wanted to determine which pages a group of students were accessing most frequently, by submitting this user group computer IP addresses the Webmaster can track the most popular pages visited by this particular user group. This provides valuable information on the usage trend of any user group. The third requirement is the *search engine optimization function* that provides real time page rank evaluation to the owners of the system. This allows a system user to submit a search term and a URL of a Web page on the organization's Website and receive a real time Web page evaluation report that compares this Web page to the highest-ranking pages by Google for the search term. This complex evaluation involves performing link and content analysis of the top ranking web pages for a particular search term and the specified web page on an organization Website. The system then, based on this information, generates intelligent page rank improvement suggestions. The suggestions offer key ways in which the specified Web page's content can be modified in order to improve its search engine ranking. These suggestions can in turn be appended to the evaluation report and distributed to a Web development team in Excel or .pdf format. To aid competitor analysis the researchers also identified the need to *track competitor's page ranking*. This functionality allows the system user to track the Google Search Engine ranking of a particular Web page for all search terms that the user has associated with for example a competitor. Additionally, if the specified Web page is located on an organization's Website then the system will also automatically generate a Web page traffic report for this Web

page so that a correlation can be derived between a change in a Web pages Google Search engine ranking and the traffic received to a Web page. This functionality also intermingles with the search engine optimization functionality, as it allows a user to determine if modifications to a Web page based on system generated suggestions have actually made an impact on a Web pages search engine ranking. The 'search term rank tracking' report offers a reverse of the functionality of the 'Web page rank tracking' report. It allows the user to select a search term they have been tracking in association with a Web Pages URL and instead track the Google search ranking of all the URLs associated with the search term. This report offers a subtle means through which a company (the cases) can track how it is performing in the Google Search Engine against its competitors (other educational courses and investment banks). Although the four broad functions of the system as outlined above are described independently and from an end user perspective and can be used in a mutually exclusive manner, this is in fact not the most efficient manner in which to use the system. The maximum benefits from the system are derived from interconnecting the different functionality of the system (see Figure 2).

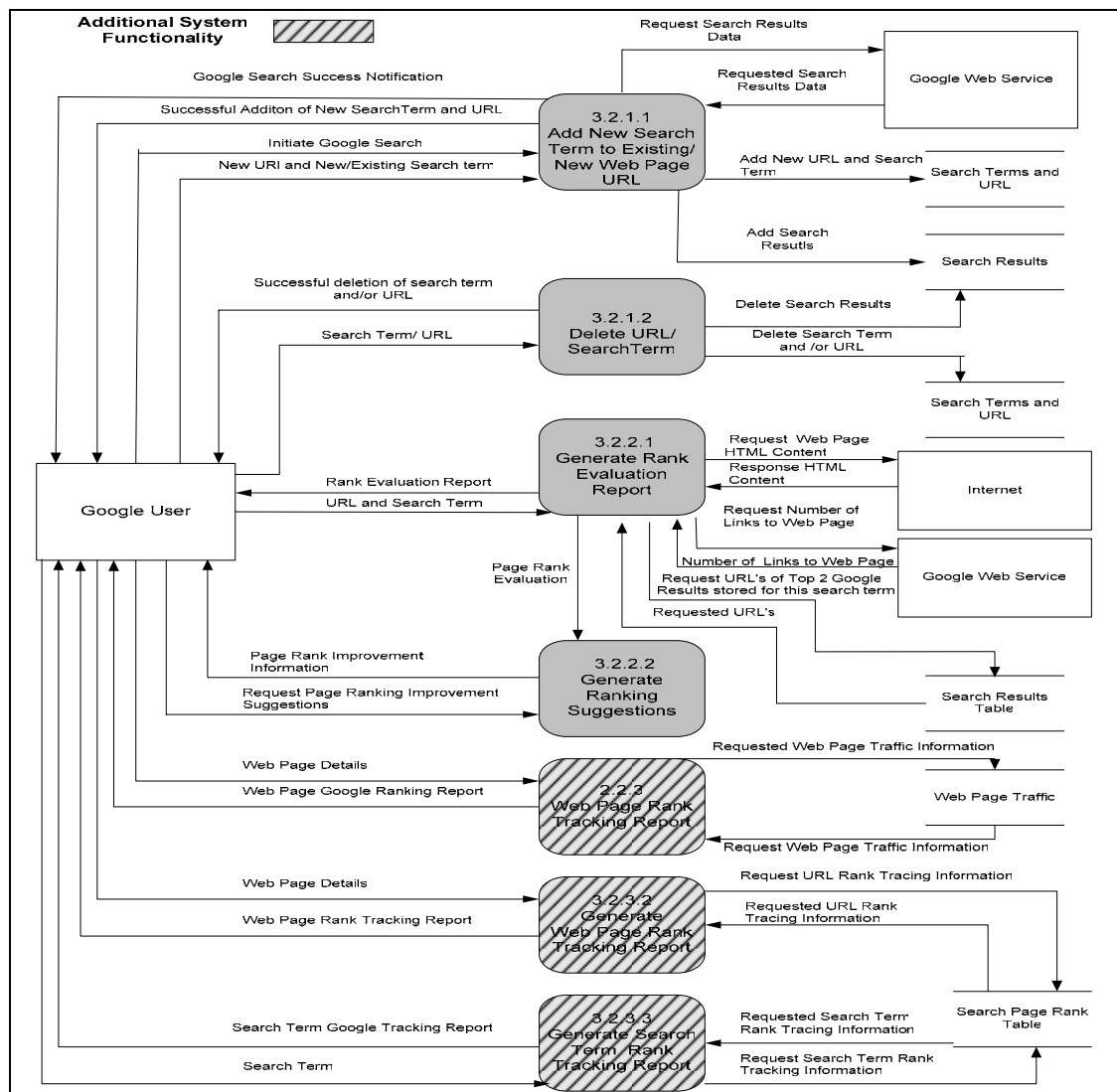


Figure 2. Googalysers Functionality

So, for example, if a user uses the system Website traffic monitoring capabilities to determine which pages are not receiving the required level of Web traffic, the manager could then use the key user group analysis functionality to determine which Web user groups are visiting this Web page. Subsequently, the user could modify the content of the Web page to maximize its benefit to a particular user group. After this the user could use the search engine optimization functionality to help them to optimize the Web page's Google ranking for a range of keywords; finally the user can track the Web page's Google ranking to detect any improvements across all previously identified keywords. Subsequently, the user can loop back to the start of the process again and determine if they have actually achieved a higher level of traffic to the targeted Web page.

6 DEVELOPMENT PROBLEMS AND CHALLENGES

From the researchers early research into the Google API Web service, it became apparent that Google have placed certain restrictions/barriers to the access of the Google API. These include a results limitation through which the Google API returns only *ten* results per query. Effectively when a user enters a specified search term such as "Mutual Funds" for example only the top *ten* search results are returned for the search term. Initially, this posed a significant problem to the system, however the researchers added an automated query which queries the API a number of times for each search term until the required number of search results are retrieved. The search engine also enforces a limitation regarding the number of requests users can make. In order to access the Google API it is necessary that a system user obtain a Google License Key from the search engine (<http://www.google.com/apis/>). Each time a request is sent to the Google API a license key value is also provided as an argument in the request. Each license key is limited to make *one thousand* license key requests to the Google API per day. These restrictions posed serious issues for the researchers and required careful thought in order to effectively overcome the problem. The implication for the system included the number of search results retrieved. Due to the restriction, it simply was not practical to retrieve all search results for a particular search term from the Google API. This restriction has altered the Google daily summary process results as the API will only retrieve the top *two hundred* search results for a search term. Thus, it will only check for a URL associated with a search term by a system user in the top *two hundred* search results. Therefore, if a Web page is outside the top *two hundred* search results its Google Ranking position cannot accurately be tracked. In addition to the above, another added complication was that the API is not always available as Google periodically 'take down' the Web service in order to undertake maintenance work. During this time it will not be possible to perform a Google Search via the API. To deal with this scenario the researchers implemented an error handler that notifies the system user if the Google API is unavailable. In addition, the system also has the ability to notify all of the users when the Google API is available again.

7 FINDINGS

While this research primarily focuses on the development of the Googalyser tool the researchers have received some initial feedback from two of our test case sites that have fully implemented the application. Although this preliminary data is not as extensive as we would like it provides us with a strong indication of some of the benefit these organizations can derive from the application as well as those highlighted through the development stage. University College Cork (UCC) uses a number of Websites from in-house developments to off-the-shelf packages to promote the university. Each department within the university has the option of developing their own site, which is directly linked to the Home page. Therefore the origin and type of Website for the different departments varies from one department to the next. Individual departments, such as Business Information Systems (BIS), have developed customized Websites, which are geared

towards the promotion of internal courses as well as facilitating/supporting student e-learning requirements. In fact, the Web designer within BIS was seeking to refine both the content and structure of the BIS Website in order to improve its relevance to both current and prospective students and industrial partners. Prior to the implementation of the Googalyser the department did not have any data regarding the different user groups visiting the Website. For example it was unaware of the type of content searches carried out by current Industrial partners such as the type of subjects taught to the graduates or research projects undertaken within the department. Utilizing existing data analysis of existing Web logs was identified as a source of value to the designer of the Website, which would determine its usage. The Bank test case, like the majority of banks, has invested heavily in the development of Web enabled technologies. The majority of the services it provides to its stakeholders are available over the Web. In order to maximize the number of potential clients who access the company's Website, the company constantly strives to improve its ranking through major search engines. A bank was selected as a test case for the system as it represents a sector that strives to promote online services and it is in direct contrast to the educational case. It, like its competitors, monitors a number of search terms that are related to its business. Currently both cases, as do other companies, manually enter a search term into the 'Google Search Engine' and then scroll through the search results in order to find the position/ranking of its Website in relation to a key 'search term'. Subsequently, if a Web page is not within at least the top ten results a developer would have to view the HTML code of the corporate Web page and the top ranking pages in Google for a particular search term and then try to identify factors such as: (1) the number of links to each page, (2) the occurrence of keywords in the title and content of the page, (3) the occurrence of key phrases in the title and content of the page and (4) finally, the relevance of text placed between anchor tags on each page. Having performed this operation, it was then necessary to perform an analysis of the collected data in order to determine methods through which the Web page could be modified in order to improve its search ranking. This situation prior to system implementation had a number of drawbacks for the cases. Firstly, the process as it was performed prior to the system implementation was extremely inefficient. It was very time consuming and monotonous work to compile the reports on particular search terms. Furthermore, the labour resource intensity in undertaking this work could certainly be applied to more progressive work elsewhere within the cases. In addition, as the system was paper-based, there was no centralized repository where all the search term information could be tracked. Finally, we are looking forward to reporting detailed case analysis from both of the research sites identified here. However, as a direct result of the implementation both cases have increased Website rankings for chosen key search terms and analyzed the strategies of specific competitors to leverage this knowledge to make more informed and timely business decisions.

8 DISCUSSION AND CONCLUSIONS

The obvious conclusion that can be drawn from this evidence is that through the use of a tool such as the Googalyser, a company can gain enhanced visibility on the web by achieving a high search engine ranking. Clearly, organic search engine optimization is the optimal approach to undertake when attempting to achieve competitive advantage in this increasingly competitive e-business environment. The need for the development of a system that could contribute to achieving this goal drove this research. It has resulted in the development of an easy to use search engine optimization tool. The objective of the Googalyser is to provide strategic information that will allow an organization to gain competitive advantage in their respective market place by aiding users to perform search engine optimization analyses. Data from Web-browser traffic to an organizations Website is analyzed to determine the source of the traffic, the group accessing the site and their path to provide vital information to the decision-makers within the organization. The goal of the results from the patterns discovered in the data is not only to

increase the ranking of the site but to align the focus of the business with the corporate Website. The focus of the business is incorporated into the design, with the aid of the system, through the provision of recommendations to increase the ranking of the keywords, which describe the organization such as 'Mutual Funds' or 'IS Graduates'. The cases presented here are prime examples of the utilization of the system within *any* environment in promoting an online presence that can and will ensure ongoing success. An effective system can provide a university or an organization with a strategic advantage in the virtual market. It provides a unified view of the customers (group specific) online path and interests to enable a company to target specific groups and niches. Therefore, the tool can facilitate the management of Web traffic to generate value adding information, and the expectations of customers in the provision of relevant searchable terms. The tool essentially provides organizations with the ability to deal with the complexity of optimizing an organizations online ranking. The ranking as well as the quality of the online service dictates the retention of an organization's or a university's customer base. This paper focuses on the use of the Googalyser in pulling a wider audience to the Website under analysis as the higher a term or site is ranked the higher the potential 'click through rate'. The research outlines the potential of the system which incorporates a data mining approach to analysing both web log files and web content through a user-friendly GUI (graphical user interface) to provide any organization with a competitive advantage. Previous research into increasing the Web presence of an organization has identified theoretical online strategies and models with rich case examples. However, this study illustrates the design and use of a system that examines Web traffic to discover patterns of customer behaviour and provide management with relevant data pertaining to their individual Websites with suggestions for improving and evaluating the design. Further research would also indicate the justification of this system as an integral part of a methodology, such as the ICDM, to aid developers to consistently re-evaluate their designs. It is the intention of the researchers to measure the level of competitive advantage achieved through the implementation of the system into suitable case environments.

9 REFERENCES

- Adam, Frédéric, & Murphy, Ciaran (1998) Executive Information Systems, The Executive Business Review, Executive Systems Research Centre, Issue 16.
- Aubert, B. & Hamel, G., (2001), Adoption of smart cards in the medical sector: the Canadian experience, Social Science & Medicine, volume 53, (2001), pages 879–894.
- Chard, S. (1999) "Data Management: The Foundational for Web Development in the Retail and Service Sector" http://www.firstmonday.dk/issues/issue4_9/chard/index.html
- Chase, P.R. (2000) "Why CRM implementations fail....and what to do about it" Scribe Software Corporation.
- Chattel, A. (1998) "Creating Value in the Digital Era: Achieving Success through Insight, Imagination and Innovation", McMillan Business Press.
- Crowder, H.P., Dinkelacker, J., Hsu, M. (2001) "Predictive Customer Relationship Management: Gaining Insights about customers in the electronic economy" DM Review Feb 2001
- English, L. (1999) Improving data warehouse and business information quality, Wiley Computer Publishing, New York.
- Fingar, P. (1998) A CEO's Guide to eCommerce Using Intergalactic Object-Oriented Intelligent Agents, July 1998. <http://home1.gte.net/pfingar/eba.htm>
- Ghosh, S. (1998), "Making Business Sense of the Internet", Harvard Business Review, March-April 1998, pg. 126-135

- Greenstein, M. and Vasarhelyi M., (2002), *Electronic Commerce; Security, Risk Management and Control*, 2nd Edition, Published by McGraw Hill 2002.
- Ho Ha, Sung; Chan Park, Sang; (1998). "Application of Data Mining tools to Hotel Datamart on the Intranet for Database Marketing"; *Expert Systems with Applications*; Issue 15; pg 1-31.
- Hui and Jha, 2000 Hui , S.C., Jha, G.(2000). Data mining for customer service support, *Information and Management*, October, 38(1), 1-13.
- Kalakota, R., Robinson, M. (1999) e – Business Roadmap for Success, Addison – Wesley.
- Laudon, Kenneth C.; Laudon Jane P. (2004). "Essentials of Management Information Systems: Managing the Digital Firm"; 6th Edition; Prentice Hall
- Lawrence E., Corbitt B., Tidewell A., Fisher J. and Lawrence J. (1998), "Internet Commerce: Digital Models for Business", Wiley & Sons Publishing.
- Magic Software Enterprises (2000) "The CRM Phenomenon" White Paper www.magic.sw.com
- Maoz, M. (2001) "CRM: What's right for customers is the correct approach", Gartner Group Inc. www.gartner.com
- McKenna, R. (1995), "Real Time Marketing", *Harvard Business Review*; July- Aug.1995, pg. 87-95
- McKenna, R. (1991), "Marketing is Everything", *Harvard Business Review*; Jan-Feb 1991, pg.65-79 Nambisan & Ming Wang, 2000
- Neville, K. and Heavin, C., (2004) e-learning: Academia's Approach To The CRM Challenge, *Proceedings of the 2004 IFIP International Conference on Decision Support Systems (1 - 3 Jul 2004)*.
- Porter, M. (1985), "Competitive Advantage: Creating and Sustaining Performance", Free Press, New York.
- Rayport and Sviokla (1994), "Managing in the Marketspace", *Harvard Business Review*; 1994
- Sammon, David; Adam, Frederic; Carton, Fergal; (2003). "Benefit Realisation Through ERP: The Re-Emergence of Data Warehousing"; *Electronic Journal of Information Systems Evaluation*; Volume 6; Issue 2; p 155-164
- Shanks, G., and Tay, E. (2001) "The role of knowledge management in moving to a customer-focused organization", *The 9th European Conference on Information Systems, Bled, Slovenia, June 27 – 29, 2001*.
- Shanks, G., Darke, P. (1998) "Understanding Data Quality in a data warehouse", *Australian Computer Journal (November)*, (30:4) pp 122 –128
- Shin and Bongsik (2002). "A Case of Data Warehousing Project Management"; *Information & Management*; vol 39; pg 581-592.
- Standing, C. (1999). *Managing and Developing Internet Commerce Systems with ICDM*. In *Proceedings of the Xth Australasian Conference on Information Systems, 1999*, pg. 850-862, 1st-3rd December 1999, Wellington, New Zealand.
- Standing C. (2001). The requirements of methodologies for developing web applications. In *Proceedings of the 9th European Conference on Information Systems (ECIS 2001): "Global Co-operation in the New Millennium"* Vol 1, pp 548-557, 27th-29th June 2001, Bled, Slovenia.
- Thelwell, M. (2000), "Commercial Web Sites: Lost in CyberSpace", *Internet Research: Electronic Networking Applications and Policy*, Vol. 10 No. 2, pg.150-159.

Thompson, A.A., and Strickland, A.J. III. (1995). *Strategic Management: Concepts and Cases*, (8th ed.), Homewood: Richard D. Irwin, Inc.

Toffler, A. (1991) *Powershift: Knowledge, Wealth, and Power at the Edge of the 21st Century*, Bantam Books, Reprint Edition 1991.