A Framework to Integrate Business Goals in Web Usage Mining

E. Hochsztain[1*], S. Millán[2], B. Pardo[3], J.M. Peña[3], and E. Menasalvas[3+]

[1]Facultad de Ingeniería - Universidad ORT Uruguay
[2]Facultad de Ingeniería. Universidad del Valle. Cali. Colombia
[3]Facultad de Informática UPM. Madrid Spain
esthoc@adinet.com.uy
millan@eisc.univalle.edu.co
bpardo@teldat.es
jmpena@fi.upm.es
emenasalvas@fi.upm.es

Abstract. Web mining is a broad term that has been used to refer to the process of information discovery from Web sources: content, structure, and usage. Information collected by web servers and kept in the server log is the main source of data for analyzing user navigation patterns. Notwithstanding, knowing the most frequent user paths is not enough: it is necessary to integrate web mining with the company site goals in order to make sites more competitive. The concept of **Web Goal Mining** is introduced in this paper to refer to the process information discovery of the relationship between site visitors and sponsor goals.

1 Introduction

Explosive growth in size and usage of the World Wide Web has made it necessary for web site administrators to track and analyze the navigation patterns of web site visitors. The web is a rich source of information about user navigation patterns, content of pages and site structures making it a perfect target for data mining. Web mining is a broad term that has been used to refer to the process of information discovery from Web sources (Web content), discovery of the structure of the Web servers (Web structure) and mining for user browsing and access patterns through logs analysis (Web usage) [2,12]. Information collected by web servers and kept in the server log is the main source of data for analyzing user navigation patterns. Web servers commonly record an entry in a Web log file for every access. Common components of a log file include: IP address, access time, request method, URL of the object or page accessed, data transmission protocol, return code and number of bytes transmitted.

Nevertheless, these data have to be enhanced with information about the business, so that patterns can be extracted to provide the company with information about users and their activities. Adaptive web sites are web sites in which information is custom-

* Research has been partially supported by Programa de Desarrollo Tecnológico (Uruguay).
+ Research has been partially supported by UPM (Spain) under project WEBP-RT.

E. Menasalvas et al. (Eds.): AWIC 2003, LNAI 2663, pp. 28–36, 2003.

ized according to the user. Most of the approaches to build this kind of servers are based on the user's identity and profile.

A lot of approaches [9,11–13,15–20] propose to track and analyze clickstream data to obtain most frequent paths. Most of these approaches calculate user profiles taking into account this information. Many companies have focused more on web site traffic than on their profitability. However, the number of visitors does not necessarily correlate with the number of customers ("those who purchase at the site"). Though marketing has generally focused on the former campaigns should also be focused on customers, those who ultimately determine profits.

Thus, knowing the most frequent user paths is not enough. But, it is necessary to integrate web mining with the company site goals in order to make sites more competitive. One of the reasons web mining may fail is that most of the web mining firms concentrate exclusively on the analysis of clickstream data that contain information about users, their page views, and the timing of their page views. In order to solve this problem, in [7] authors propose an algorithm that considers information on both server logs and business goals to improve traditional web analysis. The focus is on the business and its goals, and this is reflected on the computation of link values, integrating server logs analysis with both the background that comes from the business goals and the available knowledge about the business area. They do not, however, explicitly show how to compute or implement the approach. Taking their proposal as the underlying idea, it is possible to establish that, through web site analysis, both visitors and web site sponsors have to be taken into account.

Other authors [22] present different session reconstruction strategies in terms of either proactive or reactive behavior. They also include a set of different metrics for the evaluation of the reconstruction process. And in [23] a more semantic and context-rich approach based on ontologies is proposed.

It has been shown that web mining must help both users and sponsors of a web site to achieve their goals. Our proposal, unlike traditional web mining approaches, sets the focus on the achievement of site sponsors goals.

A Web site is designed to serve the company needs. Thus, the way the site is used must add value not only for users but also for the company [26]. Former Web Mining approaches focus on site usage: tracking how many visitors there are and how well the site fits their needs. Summing up, the only point of view taking into account is the user's one.

These approaches are based on the idea that if users do not feel comfortable while navigating the company's site, they will not become customers. So, in order to turn users into customers and to avoid current customers to churn, it is necessary to fulfill users' needs.

However, achieving visitors satisfaction is not enough. It must be kept in mind that the site main goal is not just customer satisfaction, but contributing to company success [26]. It is possible that visitors do not use the site in the way it is intended to, nor responding to company needs. For instance, visitors to an electronic book shop may use the site as a catalog to look for books, read some reviews, and then go to a traditional, non web-based book shop to buy the selected items. Probably, that user feels that the site is very useful for him/her, but the goals of the site sponsor (selling books) are not being fulfilled.

Traditionally, clickstream data have been used to measure web sites success. However, using number of hits and page views to evaluate sites success is much like evaluating music quality by its length. Web advertising is another example. Advertis-

ing cost (through banners, for instance) is calculated counting clickthroughs. So, the advertiser pays each time a user clicks on the banner. Consequently, many companies use the same way to measure advertising effectiveness: number of visitors clicking the ad. But customers or sales are not being considered in this measure. In [21], authors state that, if the company goal is selling, the site must attract customers instead of mere visitors. According to this, it is not worth to track user's behavior and sponsor goals separately: it is necessary to analyze how the site is being used and how this usage affects the company's goals as a whole.

Traditional classification of Web Mining into three areas (content, structure and usage) does not consider sites sponsors or companies. Existing approaches do not explicitly consider that web sites are used by sponsor companies as a way to achieve their business goals, and consequently, it is necessary for these companies to measure site effectiveness from that point of view.

Hence, in order to consider this point of view, we introduce a new web mining approach over those already proposed in [2,12] called **Web Goal Mining**. We use the term Web Goal Mining to refer to the process of discovering information about the relationship between site visitors and site goals. The objective of Web Goal Mining is to evaluate the effectiveness of a web site according to the company's goals. In spite of the huge volume of data stored in the web, it is difficult to understand the relationship between user navigation data and site effectiveness in terms of site goals when trying to design "good pages" from the site users' point of view.

Based on navigation pattern discovery, a methodology to improve the success of web sites is proposed in [1, 9]. To improve the success of a site, the authors propose a model that measures and improve the success of the site. The proposed model takes into account the owner's corporate objectives, categorizes users according to their activities in pursuing those goals, and evaluates the site's success. The model can also be used on a regular basis.

In order to evaluate the success of the site, two measures (contact and conversion efficiency) are proposed in [3]. The success of the site is defined by the author in terms of its efficiency for transforming visitors into customers. On the other hand, Eighmey [4] presents a diagnostic method, Website Response Profile, for website planners and designers, that can be used to evaluate visitor reactions to the experience of any website.

The main purpose of this paper is to propose a framework to add value to already existing approaches [24]. Decision-making criteria related to design and content of web sites are needed, so that user behavior matches the objectives and expectations of web site owners. We are proposing therefore a set of criteria (elements) to evaluate web user behavior in terms of web site owners' objectives and to improve web content and design.

The proposed framework takes into account traditional commerce features. In traditional commerce, when a customer arrives at a particular business, her/his acquiring goods/services intention can be easily inferred. Therefore, arriving action has high value. However, if we find that the same customer (or profile) frequently visits the site but never buys anything, the arriving customer action will reduce its value. In the same way, we propose to analyze data of site visitors in order to discover the best ones according to the site business success criteria.

The paper has been organized as follows: Sect. 2 introduces the concept of Web Goal Mining. In Sect. 3 a description of the Web Goal Mining Framework is introduced. Section 4 presents both conclusions and future research lines.

2 Web Goal Mining

In the previous section, the need for considering the business's point of view when performing Web Mining task was presented, and the concept of Web Goal Mining to address this problem has been proposed. Nevertheless, it is necessary to define this concept in a more formal manner.

Web Goal Mining is defined as the process of discovering information formerly unknown and useful for making decisions. These decisions should allow the sponsor company to improve the effectiveness of its web site. Effectiveness is measured in terms of how well the company achieves its (business) goals through its web site. Web Mining has been traditionally divided into three interrelated areas: Web Content Mining, Web Usage Mining and Web Structure Mining [2,12]. In this sense, Web Content Mining has been used to refer to the process of extracting information from web pages content. Web Structure Mining makes reference to the process of information extraction from web topology based on existing links between web pages. Finally, Web Usage Mining has been used in reference to the process of information extraction on how the web is being used. This has been the traditional classification of web mining.

In none of these areas business goals are explicitly considered. This is to say that web sites sponsors use web sites as means for obtaining business's goals. Hence, sponsors need an accurate measure of Web site effectiveness according to the business's goals point of view. The purpose of a web site is to help companies reach the proposed target. Thus, we consider that any of these approaches can be enriched taking into account business's goals

We propose to analyze how well users' navigation fits company aims, how accurate Web site content reflects company's purposes and how much Web site structure contribute to achieve company's goals.

Hence, Web Goal Mining like Web Mining, may be divided into three main areas: Web Usage Goal Mining, Web Content Goal Mining and Web Structure Goal Mining, bringing information regarding the company and its goals together with traditional Web Mining areas.

- Clickstream data (Web logs) + Metadata + (Company + Company Goals) = Web Usage Goal Mining
- Links data + Metadata + (Company + Company Goals) = Web Structure Goal Mining
- Sites content data + Metadata + (Company + Company Goals) = Web Content Goal Mining

Once the Web Goal Mining concept has been defined more accurately, it is necessary to present more comprehensive definitions of what is meant when we refer to goals. We assume definitions for terms regarding objectives like the ones presented in [25]:

- Objective: An objective is a desired end result or condition expressed in measurable terms that can be achieved by the successful performance of one or more business or functional processes.
- Goal: A goal (or target) is the criterion by which you measure the accomplishment of an objective. Every objective must have a quantifiable goal or target.
- Strategy: A strategy is a method or procedure for accomplishing the related objective and achieving the desired goal.
- Performance Measure: A performance measure is an indicator built into a strategy that can measure progress toward satisfying the related strategy.
- Critical Success Factor (CSF): A CSF is a business function or activity that must be performed correctly and completely.
- Key Indicator: A key indicator is a measurement that is readily obtained by observation of a business process or activity, which provides data on how well a process or activity is performed.
- Variance or limits: The degree to which a key indicator can vary and still be within tolerance.

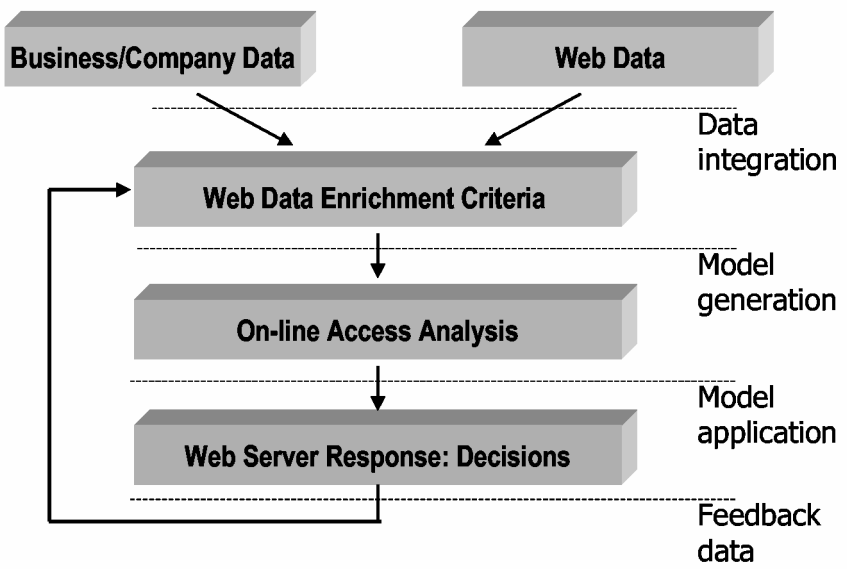
Last, incorporating data regarding the company and its goals to evaluate the web site leads us to consider the following subjects:

- Creating a representation model for company data
- Defining ways to represent company aims
- Analyzing goals at different levels of granularity: company wide goals or department wide ones; short, medium or long term ones and based on users' profiles, for instance.
- Finding the relationship between the web site and company objectives.
- Defining metrics to evaluate whether the site contributes to company goals, and how.

Once Web Goal Mining has been introduced we present an approach than can be categorized as a Web Goal Mining approach overlapping with Web Usage Mining. The approach assumes a webhouse in which information about users, goals, web site, and clickstream data is stored.

3 Web Goal Mining Framework

The proposed framework tries to increase Web Mining effectiveness adding information about business's goals. Web Goal Mining architecture is formed by the modules, depicted by Fig. 1:



- **Business Data**: This module incorporates gathering, preprocessing, modeling and storing data regarding business (goals, different points of view to evaluate those goals and company environment).
- **Web Data**: Module that gathers, preprocesses and stores web data (web logs, user profiles, web environment, navigational patterns, etc.). This module is also present in traditional Web Mining.
- **Web Enrichment Criteria**: This component gathers data on both web and business. Then it generates a common model that will be the base for Web Goal Mining analysis.
- **On-line Access Analysis**: This element applies the former model to current sessions in order to estimate how valuable the session will be for achieving company's goals.
- **Web Server Response**: Decisions and operational strategies to address user navigation towards more profitable web pages or services. These decisions would be mechanisms like special web-page contents, improvement/tuning of web server performance, special offers or advertisements, to name a few.

Web Enrichment Criteria should be reviewed as new decisions are proposed. This feedback data allow web site owner to review how data are integrated and which information is useful in terms of the business's goals. New information and the result of different proposed decisions should be included in a continuous iterative cycle.

The following table represents each module, the inputs it needs and the outputs it must produce.

Module	Input	Output
Business Data Harvesting	Interviews Focus groups Documentation Market Analysis	 Goals Sets Points of View Sets Environment Sets
Web Data Retrieving	Site topology Web Logs Relational Database Data Webhouse	Refined Web Logs
Web Logs Enrichement	Logs Taget, points of view, environments	Link values Page values

4 Conclusions

Under the assumption that "What it is interesting for the web user could not be profitable for the web site company", the Web Goal Mining approach has been defined. The objective of Web Goal Mining, differing from the traditional Web Mining concept, is to evaluate the effectiveness of a web site according to the company's goals.
In this paper we had introduced Web Goal Mining as a new and integrated web mining approach in which the relationship between Web Site and Company's Goals is considered. This contribution defines a new framework for how Business Objectives and Navigation Analysis should be combined. The purpose of this combination is to label web sessions in terms of profit-related criteria, closer to the objectives requested by the web site sponsors/owners. This innovative approach opens new and challenging research lines on the mechanisms and metrics appropriate to merge this information, as well as new decisions and strategies to redirect and seduce potentially profitable clients.

Our contribution identifies the different Web Goal Mining steps and the main targets to be provided for each of the phases this process is divided into. These phases are complex and their procedures require collaboration between both market-analysts and web development experts.

New challenges and future research directions are proposed. These open issues could be developed and addressed by multiple alternatives and the forecoming work on this framework will present them.

References

[1] Spiliopoulou M, Pohle C . Data Mining for Measuring and Improving the Success of Web Sites. Data mining and Knowledge Discovery, (5)1–2, 2001
[2] Han J. and Kamber M. (2001) Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.
[3] Berthon P., Pitt L., Watson R. The World Wide Web as an advertising medium. Journal of Advertising Research 36(1) pp. 43–54,1996

[4] Eighmey J. On the Web: It's What You Say and How You Say It. Greenlee School of Journalism and Communication, Iowa State University http://www.Eighmey.Website.Response.Profile.htm
[5] The CRM Handbook Jill Dyché – Addison Wesley 2001
[6] Hu X., Cercone N. An OLAM Framework for Web Usage Mining and Business Intelligence Reporting – WCCI 2002
[7] Menasalvas E. Hochsztain E. Sessions Value as measure of web site goal achievement. SNPD'2002. Madrid July 2002
[8] Kimball R. Merz, R.The Data Webhouse Toolkit Wiley, 2000
[9] Spiliopoulou M., Pohle C., Faulstich L. (1999) Improving the efectiveness of a web site with web usage mining. In Proceedings WEBKDD99.
[10] Menasalvas E, Millán S, Hochsztain E. (2002) A Granular Approach for Analyzing the Degree of Afability of a Web Site. In International Conference on Rough Sets and Current Trends in Computing RSCTC2002
[11] Mobasher B., Jain N., Han E. and Srivastava J. (1997) Web mining: Pattern discovery from WWW transaction. In Int Conference on Tools with Artificial Intellgence, pages 558–567, New Port.
[12] Srivastava J. , Cooley R., Deshpande M. , and Pang-Ning Tan. (2000) Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations, 1:12–23.
[13] Florescu D., Levy A. , and Mendelzon A. (1998) Database techniques for the World-Wide Web: A survey. SIGMOD Record (ACM Special Interest Group on Management of Data), 27(3):59.
[14] Menasalvas E, Millán S, Hadjimichael M, Hochsztain E (2002) An algorithm to calculate the expected value of an ongoing user session In The 2002 IEEE Data Mining Conference ICDM '02
[15] Shahabi C., Zarkesh A.M. , Adibi J., and Shah V. (1997) Knowledge discovery from user's web-page navigation. In Proceedings of the Seventh International Workshop on Research Issues in Data Engineering, High Performance Database Management for Large-Scale Applications (RIDE'97), pages 20–1, Washington – Brussels – Tokyo, IEEE.
[16] Nasraoiu O., Krisnapuram R., and Joshi A. Mining web access logs using a fuzzy relational clustering algorithm based on a robust estimator.
[17] Mobasher B., Dai H., Luo T., Nakagawa M., and Witshire J.. (2000) Discovery of aggregate usage profiles for web personalization. In Proceedings of the WebKDD Workshop.
[18] Borges J. and Levene M. (2000) A heuristic to capture longer user web navigation patterns. In Proc. Of the First International Conference on Electronic Commerce and Web Technologies, Greenwich, U.K., September.
[19] Pei J., Han J., Mortazavi-Asl B. , and Zhu H. (2000) Mining access patterns eficiently from web logs. In Proceedings Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00).
[20] Wolfang G. and Lars S. Mining web navigation path fragments. In Workshop on Web Mining for E-Commerce – Challenges and Opportunities Working Notes (KDD2000), pages 105–110, Boston, MA, August 2000.
[21] Ansari S. and Kohavi R. and Mason L. and Zheng Z. Integrating E-commerce and Data Mining: Architecture and Challenges. In Workshop on Web Mining for E-Commerce - Challenges and Opportunities Working Notes (KDD2000), pages, Boston, MA, August 2000.
[22] Spiliopoulou M., Mobase B., Berent B. and Nakagawa M. A Framework for the evaluation of session reconstruction heuristics and web Usage Analysis, in I. Journal on Computing pp 1–20, 2002
[23] Berent B. Detail and Context in Web Usage Mining: Coarsing and Visualazing Sequences, In Proc. WEBKDD 2001
[24] Mobasher B. Integrating Web Usage and Content Mining for More Effective Personalization, In Proc. ECWeb 2000

[25] Department of Defense, Framework for Managing Process Improvement, On-line guide: http://www.c3i.osd.mil/bpr/bprcd/policy/fr_toc.html. Dec 2001
[26] Mobasher B., Berent B., Spiliopoulou M.and Wiltshire J. Measuring the Accuracy of Sessionizers for Web Usage Analysis. In Proceedings of the Web Mining Workshop at the First SIAM International Conference on Data Mining, April 2001