World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:1, No:6, 2007

# A Web Designer Agent, Based on Usage Mining Online Behavior of Visitors

Babak Abedin, and Babak Sohrabi

Open Science Index, Economics and Management Engineering Vol:1, No:6, 2007 publications.waset.org/14487/pdf

***Abstract*—**Website plays a significant role in success of an e-business. It is the main start point of any organization and corporation for its customers, so it's important to customize and design it according to the visitors' preferences. Also, websites are a place to introduce services of an organization and highlight new service to the visitors and audiences. In this paper, we will use web usage mining techniques, as a new field of research in data mining and knowledge discovery, in an Iranian government website. Using the results, a framework for web content layour is proposed. An agent is designed to dynamically update and improve web links locations and layout. Then, we will explain how it is used to directly enable top managers of the organization to influence on the arrangement of web contents and also to enhance customization of web site navigation due to online users' behaviors**.**

***Keywords*—**Web usage mining, website design, agent, website customization.

## I. Introduction

WITH the proliferation of the WWW, providing more intelligent Websites has become a major concern in the e-business industry. Recently, this trend has been even more accelerated by the success of Customer Relationship Management (CRM) in terms of product recommendation and self after service, etc. Giving more intelligence to e-commerce sites is popularly recognized as one of the effective strategies that increase customer satisfaction because they react intelligently and can give a personalized response to each customer [1].

The World Wide Web is an immense source of data that can come either from the Web content, represented by the billions of pages publicly available, or from the Web usage, represented by the log information daily collected by all the servers around the world [2].

Data mining is the process of non-trivial extraction of implicit, previously unknown and potentially useful information from data in large databases. Traditionally, data mining has been applied to databases. The wide spread of the World-Wide Web technology has made the large document collection in the World-Wide web a new ground for knowledge discovery[3],[4]. Simply, mining data in this new ground is called web mining. It refers to the use of data mining techniques to automatically retrieve, extract and evaluate (generalize/analyze) information for knowledge discovery from Web documents and services [5]. Web-mining is an increasingly important and very active research field which adapts advanced machine learning techniques for understanding the complex information flow of the World Wide Web [6],[7].

In this paper, we focus on Web usage mining to analyze user behavior while the user interacts with the Web. The general goal of Web Usage Mining is to gather interesting information about users navigation patterns (i.e., to characterize Web users) and improve website personalization and ease of navigation [8]. This information can be exploited later to improve the Web site from the users' viewpoint. The results produced by the mining of Web logs can used for various purposes: (i) to personalize the delivery of Web content; (ii) to improve user navigation through pre-fetching and caching; (iii) to improve Web design; or in e-commerce sites (iv) to improve the customer satisfaction [9].

Usability is one of the major issues in the design and implementation of Web sites. The results produced by Web Usage Mining techniques can provide guidelines for improving the design of Web applications. Adaptive Web sites represents a further step. In this case, the content and the structure of the Web site can be dynamically reorganized according to the data mined from the users' behavior.

In this paper, we will analyze an Iranian firm's website, CAO (Civil Aviation Organization), to extract most frequently visited topics and links. Mining the website data, we also will add critical considerations of CAO and the CEO as the inputs variables to our proposed model. An agent is designed to process these input variables to change the website layout dynamically. Therefore, in addition to cost and time saving in website designing, more user-oriented and CEO-oriented website is obtained. The results are depicted at the end of the paper.

## II. The Problem

Developing ICTs in recent years, most of government organizations in Iran have focused mostly on their websites. In the current environment, the level of quality of web design and web services is a comparison factor between IT

Manuscript received May 19, 2005.

Mr. B. Abedin is with the department of management, University of Tehran – Iran (e-mail: b_abedin@mail.iust.ac.ir ).

Dr. B. Sohrabi is with the department of management, University of Tehran - Iran (phone+98-21-2549950; fax: +98-21-2555850; e-mail: bsohrabi@ut.ac.ir).

Open Science Index, Economics and Management Engineering Vol:1, No:6, 2007 publications.waset.org/14487/pdf

departments of different governmental organizations and firms. CIOs care about website of other organizations , so they spend considerable amount of their budget and time to enhance their own website's appearance and quality of services. In other words, website design and content plays the role of performance measurement criteria of an organization.

During previous months, from July04 to Dec04, the web site of CAO has frequently changed to have better face to users. The main factors in organizing the layout of this website, as the CIO notes, are: *"ease of site navigation and content finding for users, better site appearance, presentation of results of the projects and satisfying top manager's considerations"*. Therefore, a technical committee, including 4 engineers and web designers, was established to decide about the appearance and structure of website.

During this period of time, different homepages were developed. The homepages were dramatically changed and caused a considerable amount of monetary and time cost for the department. On average, during this period, 25% of weekly work time of the participants in the committee was spent, but after frequently changing the arrangement of the website, there was not "the best design" and so they just agreed to the current layout.

In this paper we do not intend to compare these different styles and designs. Here, we are looking for an "acceptable style of web pages", to save cost and time of website development. The "acceptable" level is measured by, first, visitor's satisfaction in site navigation and second, satisfying critical considerations of CAO and the CEO.
Here the main questions are:

- What are the main interesting topics of website for the visitors?
- What are the critical topics for the CEO and also the organization (CAO)?
- What is the acceptable layout of website according to the previous questions?
- How can we manage this layout more economically?

## III. The Dynamic Model

A web site usually contains great amounts of information distributed through hundreds of pages. Without proper guidance, a visitor often wanders aimlessly without visiting important pages, loses interest and leaves the site sooner than expected. This consideration is at the basis of the great interest about web information mining both in the academic and the industrial world [14], as The general goal of Web Usage Mining is to gather interesting information about users navigation patterns (i.e., to characterize Web users) [15], There is currently much debate about what constitutes good web site design. Many detailed usability guidelines have been developed for both general user interfaces and for web page design. However, designers have historically experienced difficulties following design guidelines. Guidelines are often stated at such a high level that it is unclear how to operationalize them [16].

Various personalization schemes and layouts have been suggested in the literature. In this paper, cost efficiency, visitor's satisfaction issues and satisfying the CEO and CIO's considerations are mostly focused. To save time and cost of improving website usability, a dynamic model is proposed. Fig. 1 depicts different parts of the model. In this model, the inputs are, first, data about online behavior of visitors and second, critical topics of the CEO and CIO.

According to fig. 1, the agent receives real-time inputs by its sensors and applies web site improvement criteria and factors, which are identified as rules in the knowledge base. This agent then decides the necessity of new changes and applies them to the layout of web links.

In the next sections, different parts of the model are described in detail.

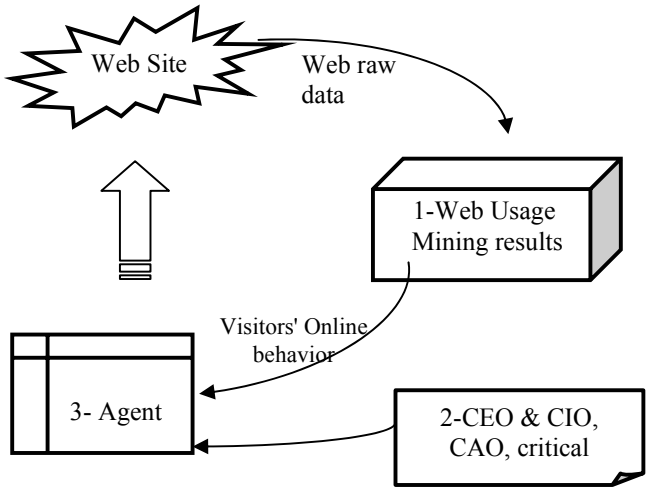



Fig. 1 The dynamic model of website improvement

## IV. Modules 1 and 2: Web Usage Mining and Identifing CAO's Critical Topics

Web users exhibit different types of behavior depending on their information needs and their intended tasks. These tasks are captured implicitly by a collection of actions taken by users during their visits to a site [17].

In this module, online behavior of visitors is mined and visited topics are sorted by click numbers.

In the literature, various web usage mining techniques, including Discovering Association Rules, Discovery of Sequential Patterns, Discovering classification and clustering rules and Path analysis, are introduced [8],[10].

In this case, to obtain the analysis, LiveStat 5.03, a product of Media House Software Inc. is used to discover and sort visited topics and links. LiveStat is log analyzer software which can track online behavior of users in a website and generate various statistical reports. Obtaining the analysis by the LiveStat, the results were not clean. So unnecessary data was eliminated and then the addresses were grouped and sorted. Table I shows addresses, which are visited in Sep04, Oct04, Dec04.

Open Science Index, Economics and Management Engineering Vol:1, No:6, 2007 publications.waset.org/14487/pdf

During last months, the home page of CAO was frequently and dramatically changed and services and head topics were arranged in various styles. But, however the designs of website in this period of time has frequently changed, the most accessed pages are almost alike in different months (circled by dashed oval). Other services and topics in the website are rarely requested and visited.

TABLE I
PERCENTAGE OF ACCESSED PAGES FROM www.cao.ir

| **Dec2004** | | **Oct2004** | | **Sep2004** | |
|---|---|---|---|---|---|
| *Topic* | *%* | *Topic* | *%* | *Topic* | *%* |
| index | 26.7 | index | 28.5 | index | 19.5 |
| Delay rep. | 10.4 | Delay rep. | 8.5 | FIDS | 13.2 |
| FIDS | 11.1 | FIDS | 9.8 | News | 13.4 |
| News | 6.23 | News | 4.7 | Timetable | 11.2 |
| Timetable | 4.2 | Timetable | 2.8 | Statistics | 1.4 |
| Ticket price | 1.7 | Statistics | <1 | Air industry | <1 |
| Statistics | <1 | Air industry | <1 | Regulations | <1 |
| Air industry | <1 | Services | <1 | Newsletter | <1 |
| Services | <1 | Regulations | <1 | About us | <1 |
| Regulations | <1 | About us | <1 | Headlines | <1 |
| e-offices | <1 | Teletex Inf. | <1 | Travel | <1 |
| Weekly Q. | <1 | Phone/SMS | <1 | Links | <1 |

It makes it clear that there are some topics that have very low access level and are not interested for most of users. So, we group them in a "detail list". As a second source, we use "CEO & CIO critical topics". These are the topics that are important for the managers of the organization and *must be highlighted* in the home page of the website. The content of this part of the website may be changed according to manager's implications. At the time of working on this paper, these topics included: *news, announcements and annual civil aviation statistics.*

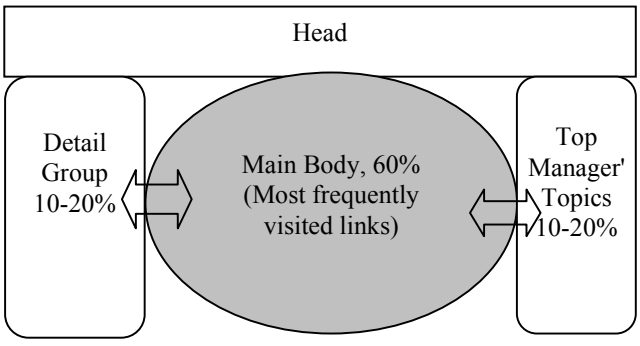



Fig. 2 A framework for website design
And content arrangement

According to Table I, the most frequent visited links were identified. From customer point of view, these links should be more highlighted in the website. We also consider the CEO's and CIO's point of view. There are some topics, according to the previous section, which are critical for CAO and they must be highlighted in the website as well. So, we will have three sections in CAO's website layout. First, the top manager's critical topics, Second, most frequent visited links (main body) and finally less visited links, "detail group".
Fig. 2 depicts the proposed framework for website design and links structure management.

## V. MODULE 3: THE AGENT

In this section we introduce an agent that is responsible to change the priority of site items based on the results of the previously mentioned table. Intelligent agents (or intelligent software agents) are defined as being a software program that can perform specific tasks for a user and possesses a degree of intelligence that permits it to perform parts of its tasks autonomously and to interact with its environment in a useful manner [18]. For Jennings and Wooldridge [19], "an intelligent agent is a computer system that is capable of flexible autonomous action in order to meet its design objectives. By flexible we mean that the system must be responsive […] proactive […], and social […]."

Here, an intelligent entity is needed. This intelligent entity should be able to recognize environmental condition, particularly customer's priority and organizational preferences and is able to change this environment based on mentioned priorities. It results in reduction of human engagement on the system and of course human mistakes. Also it should provide historical information about last changes in site and make it possible to track customer behavior and make future decisions based on analysis of the log file.

Our regarded agent should be able to monitor and perceive the customer regarded priorities and interest and then it should make decision about critical and most frequently visited links (main body) and determine which link should add to them and which link should be omitted and go to the detail group. And finally it should implement the regarded corrective actions. Therefore the agent's architecture is like fig. 3.

As shown in figure there are several modules in regarded the agent that we describe them below:

**The Perceiver Module** of the agent monitors the table of visits' history once a week.

Open Science Index, Economics and Management Engineering Vol:1, No:6, 2007 publications.waset.org/14487/pdf

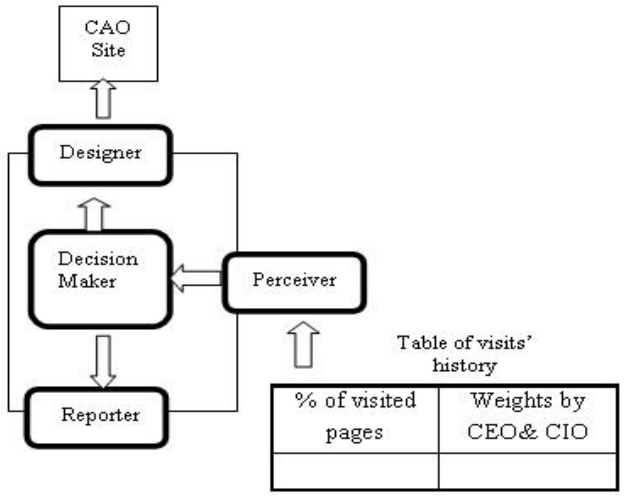



Fig. 3 Web designer agent architecture

This module just monitors the table to find significant changes in main and detail groups and report them to Decision Maker module. The perceiver compares the current table with the previous one and determines if there are any significant differences or not. If so then it sends report to the Decision Maker. The report includes the percentages of all the items.

**The Decision Maker module** of the agent is responsible for decision making about determination of main body and detail items. First it receives the percentage of visited times for each item from perceiver module, and also it receives weights of each item, which is identified by CIO and CEO. Then it ranks the items based on a ranking algorithm. Ranked items should be categorized in two groups. High priority category and low priority category. The first group will be placed in main body and the second one will be in detail group. As said before the main body items are located in more visible space of site. So, this module decides about the more visible spaces of the site and determines which items should be there. It sends a message to the designer module of the agent an ask it to do changes in the base of ranked items. **The Designer Module** of the agent is responsible to implement the decisions made by Decision Maker. It receives the ranked items in two categories and then changes the site in the way that the main body items become more visible. **The Reporter Module** is responsible for capturing the history of changes in site and the history of the customer behavior. Each time the Decision Maker sends a change order to the designer it also sends a report to Reporter. By doing so the reporter makes it easy to capture the changes of site in a period of time. By using provided information it will be easer to analyze customer interests, trends and plan for future implementations.

## VI. Conclusion

Rapid growth of internet has made websites as a huge bunch of raw data. Web mining makes it possible to discover useful knowledge from this web data for organizations. In this paper, the results of web data mining of an Iranian firm was used to propose an efficient and effective web design framework to attract more visitors, make web navigation easier and decrease web site design and development costs. Also, an agent is designed to improve the layout and arrangement of web links. So it will make the organization possible to highlight most visited links and also take critical topics for CEO and CIO into action.

In contrast to traditional way of website development, the advantages of proposed agent are:

- The layout of web contents will be enhanced and improved by the agent in the real-time based on two main factors:
  1. Online behavior of visitors: so it makes it possible to arrange the contents according to the preferences of visitors and make the web navigation easier.
  2. Critical factors for CEO and CIO
- It is possible for CEO and CIO to directly and easily identify their priorities of arranging topics in the website. So the managers can directly influence the arrangement of contents and highlight critical topics in the homepage
- By developing and using the web designer agent, the cost and time of website improvement will be saved.

## Acknowledgment

Authors thank Mr. Emad Farazmand for his kind help in developing the agent's modules.

## References

[1] Kim,Wooju. Song,Yong U. Hong ,June S. "*Web enabled expert systems using hyperlink-based inference (Accepted for publication)*". Expert Systems with Applications. 2004. pp:1-13

[2] Michele Facca, Federico. Luca Lanzi, Pier. "*Mining interesting knowledge from web logs: a survey*". Data & Knowledge Engineering. Accepted for publication. 2004

[3] Hsu, Jeffrey. "*Data mining trends and developments: The Key Data Mining Technologies and Applications for the 21st Century*". Proc. of ISECON 2002

[4] Chakrabarti, Soumen. "*Mining the web discovering knowledge form hypertext data*". San Francisco, CA. Morgan Kaufmann Publishers An imprint of Elsevier Science 2003. pp: 1-13

[5] Arotaritei, Dragos. Mitra, Sushmita. *"Web mining: a survey in the fuzzy framework*". Fuzzy Sets and Systems: vol. 148, 2004. pp: 5–19

[6] Larsen, Jan. Lars Hansen, Kai. Szymkowiak Have, Anna. Christiansen, Torben. Kolenda, Thomas. "*Webmining: learning from the World Wide Web*". Computational Statistics & Data Analysis. 38. 2002. pp: 517–532

[7] Eirinaki, Magdalini. Vazirgiannis, Michalis. "*Web Mining for Web Personalization*". ACM Transactions on Internet Technology: vol. 3, no. 1, 2003. pp: 1–27.

[8] DeYoung, Colin G. Spence, Ian. "*Profiling information technology users: en route to dynamic personalization*". Computers in Human Behavior.. Vol. 20. 2004. pp: 55–65

[9] Cooley, R. Mobasher , B. and Srivastava, J. *Web Mining: Information and Pattern Discovery on the World Wide Web*. Proc of the 9th IEEE Int'l Conf. on Tools with Artificial Intelligence (ICTAI'97), 1997.

[10] Mobasher, B. Jain, N. Han, Eui Hong (Sam). Srivastava, J. "*Web Mining: Pattern Discovery from World Wide Web Transactions*".

Technical Report TR96-050, Department of Computer Science, University of Minnesota, 1996.

[11] Theusinger, Christiane . Huber, Klaus-Peter. *Analyzing the footsteps of your customers*. Sixth ACM SIGKDD Internat. Conf. on Web KDD 2000.

[12] Ho Cho, Yoon. Kyeong Kim, Jae. Hie Kim, Soung. "*A personalized recommender system based on web usage mining and decision tree induction*". Expert Systems with Applications 23 .2002. 329–342

[13] Srivastava, Jaideep. Cooley, Robert. Deshpande, Mukund. Tan, Pang-Ning. "*Web Usage Mining: Discovery and applications of usage patterns from web data*". SIGKDD Explorations. ACM SIGKDD. Vol 1. Issue 2, 2000. pp: 12-23

[14] Albanese. Massimiliano, Picariello. Antonio, Sansone. Carlo, Sansone.Lucio, "*A Web Personalization System based on Web Usage Mining Techniques*". *WWW2004*, New York, USA. May 2004. pp: 288-289

[15] R. Kosala, H. Blockeel, "*Web mining research: a survey*", SIGKDD: SIGKDD explorations: newsletter of the special interest group (SIG) on knowledge discovery & data mining, ACM 2 (1). 2000. 1–15

[16] Y. Ivory. Melody , R. Sinha. Rashmi, A. Hearst. Marti. "*Empirically Validated Web Page Design Metrics*". *SIGCHI'01*, Seattle, WA, USA. April 2001. pp; 53-60

[17] Jin. Xin, Zhou. Yanzan, Mobasher. Bamshad, "*Web Usage Mining Based on Probabilistic Latent Semantic Analysis*", *KDD'04,* Seattle, Washington, USA. Aug- 2004. pp: 197-205

[18] J Decker, K., Sycara, K., Williamson, M., "*Matchmaking and Brokering*", In Proceedings of the second International Conference on Multi-Agent Systems, MIT Press/ AAAI Press. 1996

[19] Saeed M. H. "*Reinforcement learning in Multi-Agent Systems*" McGill university School of computer Science. 2001

**B. Abedin** is M.S. student in Information Technology Management in University of Tehran. He received his B.S. in Industrial Engineering in Iran university of Science and Technology. His research interests are in information systems strategic planning, knowledge management and data mining, intelligent decision support systems and intelligent agents.

**B. Sohrabi** is associate professor of IT management in department of management in University of Tehran. While his undergraduate education is in statistics, he received PhD in management science from university of Leicester at UK. His current research interests are in Vehicle routing problems, Information systems management in developing countries and intelligent agents.

Open Science Index, Economics and Management Engineering Vol:1, No:6, 2007 publications.waset.org/14487/pdf