Web mining framework for user profile evaluation

[1]M.Nagaraj, [2]D.Suresh
PG Scholar, Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India.
E-mail: nagarajnga@gmail.com
Professor, Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India
E-mail:Sureshdgll@gmail.com

***Abstract*— This is an innovative work for the field of web usage mining. The main feature of the paper, this is a complete framework and findings in mining Web usage patterns from Web log files of a real Web site that has all the difficult aspects of real-life Web usage mining, including developing user profiles and external data describing an ontology of the Web content. Presenting a method for discovering and tracking evolving user profiles. Profiles are also enriched with other domain-specific information facets that give a panoramic view of the discovered mass usage modes. An objective validation plan is also used to assess the quality of the mined profiles, in particular their adaptability in the face of evolving user behavior.**

***Keywords*— Ontology, CRM, Web usage mining**.

1. I. INTRODUCTION

Customer Relationship Management (CRM) can use data from within and outside an organization to allow an understanding of its customers on an individual basis or on a group basis such as by forming customer profiles. An improved understanding of the customer's habits, needs, and interests can allow the business to profit by, for instance, "cross selling" or selling items related to the ones that the customer wants to purchase. Hence, reliable knowledge about the customers' preferences and needs forms the basis for effective CRM. As businesses move online, the competition between businesses to keep the loyalty of their old customers and to attract new customers is even more important, since a competitor's Web site may be only one click away. The fast pace and large amounts of data available in these online settings have recently made it imperative to use automated data mining or knowledge discovery techniques to discover Web user profiles. Using Web usage mining techniques that can automatically extract frequent access patterns from the history of previous user click streams stored in Web log files. These profiles can later be harnessed toward personalizing the Web site to the user or to support targeted marketing. Although there have been considerable advances in Web.

In this paper, present a complete framework and a summary of our experience in mining Web usage patterns with real-world challenges such as evolving access patterns, dynamic pages, and external data describing an ontology of the Web content and how it relates to the business actors (in the case of the studied Web site, the companies, contractors, consultants, etc., in corrosion). The Web site in this study is a portal that provides access to news, events, resources, company information (such as companies or contractors supplying related products and services), and a library of technical and regulatory documentation related to corrosion and surface treatment. The portal also offers a virtual meeting place between companies or organizations seeking information about other companies or organizations. Without loss of generality, in the rest of this paper, we will refer to all the Web site participants (organizations, contractors, consultants, agencies, corporations, centers, agencies, etc.) simply as companies.

II. RELATED WORKS

There was no system presenting a fully integrated approach to mine a real Web site with the challenging characteristics of today's Web sites, such as evolving profiles, dynamic content, and the availability of taxonomy or databases in addition to Web logs. The runtime also accelerates developer productivity. For example, programmers can write applications in their development language of choice, yet take full advantage of the runtime, the class library, and components written in other languages by other developers. The runtime is designed to enhance performance. Although the common language runtime provides many standard runtime services, managed code is never interpreted. A feature called just-in-time (JIT) compiling enables all managed code to run in the native machine language of the system on which it is executing. Finally, the runtime can be hosted by high-performance, server-side applications, such as Microsoft SQL Server and Internet Information Services (IIS).

III. PROBLEM DEFINITION

Here we present a complete framework for mining Web usage patterns with real-world challenges such as evolving access patterns, dynamic pages, and external data describing ontology of the Web content and how it relates to the business actors. The Web site in this study is a portal that provides access to news, events, resources, company information and a library. The Web site in our study is managed by a nonprofit organization that does not sell anything but only provides free information. Here we perform clustering of the user sessions extracted from the Web logs to partition the users into several homogeneous groups with similar activities and then extract user profiles from each cluster as a set of relevant URLs. Data mining techniques have been applied to extract usage patterns from Web log data, this process is known as **Web usage mining.**

IV. METHODS AND MATERIALS

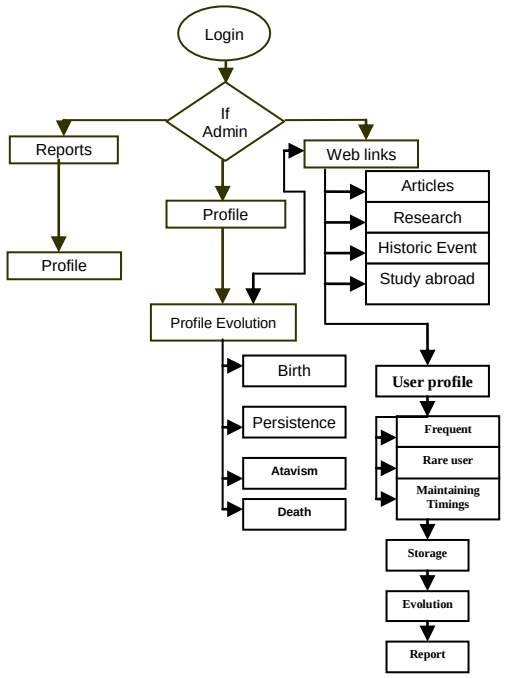

Fig. 1 Proposed Web mining architecture

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system.

Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly.

Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which provides components and the operating procedures of the system.

The proposed system consist the following modules:

1. Sign in
2. Authentication module.
3. Customer Relationship Module.
4. User Interface Module.
5. Profile Evolution Module.
6. Report Module

Sign in: In this module the user signing in to the link authentically. For this we are providing a user id and password for the user.

Authentication module: This module deals with the authentication reports of user or admin. In user authentication module, it stored the verification details of the user. The type of user and their usage, identification details like idno, name, address etc. In admin authentication the verification of the user occurs. The detail of this module helps to verify the user information.

Customer relationship module: This module deals with the customer relationship management (CRM). It can receive data from within and out side the organization to understand its customer on individual basis or on group basis by forming customer profiles. It helps a better understanding of customer's habit, needs and interests can help the business to profit.

User interface module: In this module it's all about the designs that come on the user side. It must be attractive and user friendly. If it is very complicating the user will not use the link. Users are of different types from different levels. So the link should be designed in user's point of view.

Profile evolution module:

This module is all about the user details. Like

a) Type of the user,
b) For what purpose they are using the site,
c) From when they are using the site,
d) Is they are a regular user,
e) The login time and logout time,

All this details helps to categories the user in web usage mining

Report Module: A report is used to vies and print information from the database. The report can ground records into many levels and compute totals and average by checking values from many records at once. Also the report is attractive and distinctive because we have control over the size and appearance of it.

V. RESULTS AND DISCUSSUION

In this paper, analyzed in phase and business proposal is put forth with a very general plan for the project and some cost estimates. During analysis the proposed system is to be carried out and ensure that the system is not a burden to the company. The proposed method carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely

available. Only the customized products had to be purchased. The proposed method checked the technical feasibility and technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, only minimal or null changes are required for implementing this system. The aspect of acceptance of the system by the user is easy. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

VI. CONCLUSION

Presented a framework for mining, tracking, and validating evolving multifaceted user profiles on Web sites that have all the challenging aspects of real-life Web usage mining, include evolving user profiles and access patterns, dynamic Web pages, and external data describing ontology of the Web content. A multifaceted user profile summarizes a group of users with similar access activities and consists of their viewed pages, search engine queries, and inquiring and inquired companies. The choice of the period length for analysis depends on the application or can be set, depending on the cross-period validation results. Even though we did not focus on scalability, the latter can be addressed by following an approach similar to, where Web click streams are considered as an evolving data stream, or by mapping some new sessions to persistent profiles and updating these profiles, hence eliminating most sessions from further analysis and focusing the mining on truly new sessions. Every application has its own merits and demerits. The paper has covered almost all the requirements. Further requirements and improvements can easily be done since the coding is mainly structured or modular in nature. Changing the existing modules or adding new modules can append improvements. Further enhancements can be made to the application, so that the web site functions very attractive and useful manner than the present one.

REFERENCES

[1] R. Cooley, B. Mobasher, and J. Srivastava, “Web Mining: Information and Pattern Discovery on the World Wide Web,” Proc. Ninth IEEE Int’l Conf. Tools with AI (ICTAI ’97), pp. 558-567, 1997.

[2] O. Nasraoui, R. Krishnapuram, and A. Joshi, “Mining Web Access Logs Using a Relational Clustering Algorithm Based on a Robust Estimator,” Proc. Eighth Int’l World Wide Web Conf. (WWW ’99), pp. 40-41, 1999.

[3] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, “Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data,” SIGKDD Explorations, vol. 1, no. 2, pp. 1-12, Jan. 2000.

[4] M. Spiliopoulou and L.C. Faulstich, “WUM: A Web Utilization Miner,” Proc. First Int’l Workshop Web and Databases (WebDB ’98), 1998.

[5] T. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, “From User Access Patterns to Dynamic Hypertext Linking,” Proc. Fifth Int’l World Wide Web Conf. (WWW ’96), 1996.

[6] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/ The MIT Press, 1996.

[7] J. Ross Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.

[8] Michael Berry and Gordon Linoff, "Data Mining Techniques (For Marketing, Sales, and Customer Support), John Wiley & Sons, 1997.

[9] Sholom M. Weiss and Nitin Indurkhya, "Predictive Data Mining: A Practical Guide", Morgan Kaufmann Publishers, 1998.

[10] Alex Freitas and Simon Lavington, "Mining Very Large Databases with Parallel Processing", Kluwer Academic Publishers, 1998.

[11] A. K. Jain and R. C. Dubes, "Algorithms for Clustering Data", Prentice Hall, 1988.

[12] V. Cherkassky and F. Mulier, "Learning From Data", John Wiley & Sons, 1998.

[13] Rakesh Agrawal and Tomasz Imielinski, "Database Mining: A Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, 5(6):914-925, December 1993.

[14] Usama Fayyad, David Haussler, and Paul Stolorz, "Mining Scientific Data", Communications of the ACM, vol. 39, no. 11, pp. 51-57, November 1996.

[15] David J. Hand, "Data Mining: Statistics and more?", The American Statistician, vol. 52, no. 2, pp 112-118, May 1998.

[16] Clark Glymour, David Madigan, Daryl Pregibon, and Padhraic Smyth, "Statistical Inference and Data Mining", Communications of the ACM, vol. 39, no. 11, pp. 35-41, November 1996.

[17] G. Berger and A. Tuzhilin, "Discovering Unexpected Patterns in Temporal Data Using Temporal Logic", in "Temporal Databases: Research and Practice", O. Etzion, S. Jajodia, and S. Sripada (eds.), Springer-Verlag, 1998.

[18] B. Padmanabhan and A. Tuzhilin, "A Belief-Driven Method for Discovering Unexpected Patterns", In Proceedings of the 4rd International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998.

