

STUDY AND IMPLEMENTATION OF ADAPTIVE WEB INTERFACES

Yadwinder Singh¹ (M. Tech Student), Himanshu Aggarwal² (Professor)
Department of Computer Engineering
Punjabi University
Patiala, Punjab, India.

Abstract: *The World Wide Web (WWW) tends to become the main medium supporting access to information and services to an expanding group of users. The increasing complexity of existing web applications and the volume of data offered often result in user disorientation. Adaptive interaction is proposed by many researchers as an effective means for tackling this problem. The main goal of this research paper is to present the study and implementation of adaptive interface which adapts itself according to user browsing behavior. Adaptive interface is the requirement of today's Web in which the interface of the site changes according to the previous results by mining the log files.*

Keywords: *Web usage mining, Web server log, Data mining, Adaptive Interface, and Pattern Analysis.*

I. INTRODUCTION

Data are pervasive. Machines record our program schedules, preferences, and events, achievements, buying and selling. Even our comings and goings are recorded today. As the volume of data increases, it becomes a more difficult task to comprehend it. Information technology has made it possible to manage a huge volume of data electronically and to be able to search for potentially very useful knowledge hidden inside this deep ocean of data. Data Mining, the methodology for the extraction of knowledge from data seems the only solution to this ever growing problem. Web servers store information of each page requested by web visitors in a file called the web access log. Web Usage Mining addresses the problem of extracting behavioral patterns from one or more web access logs [1]. Adaptive web interface is the technique in which the log files are mined to have the useful information and by using that information the interface of the site changes automatically. The entire process can be divided into three steps.

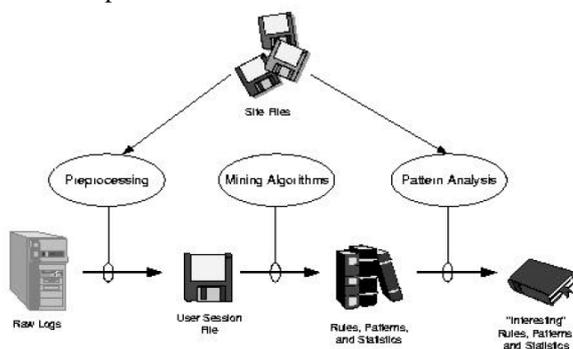


Fig. 1: Mining process

As can be seen in figure 1, the first step, preprocessing, is the task of accurately identifying pages accessed by web visitors. This is a very difficult task because of page caching and accesses by web crawlers. The second step, pattern discovery, involves applications of data mining algorithms to the preprocessed data to discover patterns. The last step, pattern analysis, involves analysis of patterns discovered to judge their interestingness. After having the desired knowledge it is used for interfaces to be adaptive.

II. ADAPTIVE INTERFACE

Adaptive interfaces are the interfaces that changes according to the user requirements. Suppose there is a site for online shopping consisting of hardware items like processors, speakers, hard disks. And if the purchase of one particular item is in large extent, then this item must be on the top of the website, is the main idea behind the adaptive web interfaces. By mining the log files and extracting the useful patterns from the log files, results into the information, that is further used in making the interfaces adaptive to the users and serves the users to meet their requirements in a better way.

III. RELATED WORK

It is quite common for Web sites to allow users to customize the site for themselves. Common manual customizations include lists of favorite links, stock quotes of interest, and local weather reports. Slightly automated customizations include records of previous interactions with the site and references to pages that have changed since the previous visit. Some sites also allow users to describe interests and will present information — news articles, for example — relevant to those interests.

A. The Index Page Synthesis Problem

Index page synthesis problem [6]: given a web site and visitor access log, create new index pages containing collections of links to related but currently unlinked pages. A web site is restricted to a collection of HTML documents residing at a single server – dynamically generated pages or multiple servers are not yet handled. An access log is a document containing one entry for each request answered by the web server. Each request lists at least the origin of the request, the URL requested, and the time of the request. Related but unlinked pages are pages that share a common topic but are not currently linked at the site; two pages are considered linked if there exists a link from one to the other

or if there exists a page that links to both of them. Rather than attempting to understand the content of every page at a site and to figure out which are related, approach is based on analysis of each visit [6]. Define a visit to be an ordered sequence of pages accessed by a single visitor in a single session. Visit coherence assumption is made: the pages a user visits during one interaction with the site tend to be conceptually related [6].

B. Page Gather Algorithm

The Page Gather algorithm [7] uses cluster mining to find collections of related pages at a web site. In essence, Page Gather takes a web server access log as input and maps it into a form ready for clustering; it then applies cluster mining to the data and produces candidate index-page contents as output. The algorithm comprises of five steps:

- a. Process access log into visits.
- b. Compute the co-occurrence frequencies between pages and create a similarity matrix.
- c. Create the graph corresponding to the matrix and find maximal cliques in graph.
- d. Rank the cluster found.
- e. For each cluster, create a web page consisting of links to the documents in the cluster.
- f. Present the clusters to the Webmaster for evaluation.

Drawback of Page Gather algorithm is that it relies on the human Web masters to determine the appropriateness of the generated index pages in final check. This will create a bottleneck for the workflow, especially for sites that have many web pages to be indexed [7].

C. Avanti Project

The Avanti Project [2] focuses on dynamic customization based on users' needs and tastes. As with the Web Watcher, Avanti relies partly on users providing information about them when they enter the site. Based on what it knows about the user, Avanti attempts to predict both the user's eventual goal and his/her likely next step. Avanti will prominently present links leading directly to pages it thinks a user will want to see. Additionally, Avanti will highlight links that accord with the user's interests.

D. Strudel Approach

It attempts to separate the information available at a web site from its graphical presentation. Instead of manipulating web sites at the level of pages and links, web sites may be specified using Strudel's view-definition language [3]. With all of the site's content so encoded, its presentation may be easily adapted.

E. Collaborative Approach

Rating is done by the users according to their interests and then customization is done and is known as collaborative filtering. In this type of customization, rating of objects is done by users then a pattern is found that which users have the same type of interest and then an object purchased by one user is recommended to the another user having the same type of interest. A simple form of collaborative filtering is

used by, for example, Amazon.com; the Web page for a particular book may have links to other books commonly purchased by people who bought this one. Firefly uses a more individualized form of collaborative filtering in which members may rate hundreds of CDs or movies, building up a very detailed personal profile; Firefly then compares this profile with those of other members to make new recommendations.

F. Client side customization

It is a technique in which a user has his/her own associated agent who learns about his/her interests and customizes his/her web experience accordingly. The AiA project [4, 5] explores the customization of web page information by adding a "presentation agent" who can direct the user's attention to topics of interest.

IV. METHODOLOGY AND IMPLEMENTATION

For implementation I have constructed a website consisting of the hardware items for sale, consisting of wide variety of products.

Flowchart for proposed work

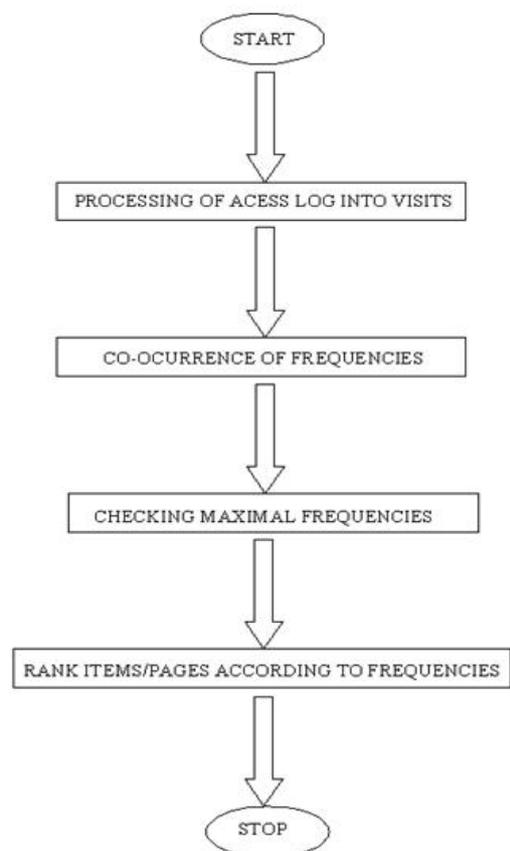


Fig. 2: Flowchart of proposed work

The whole process of adaptive interface is described by the above defined figure 1, in which the first task is to process access log into visits, then calculating the frequencies and then maximum frequencies are checked and then the ranking of the items is done according to the frequency. And the final

webpage is resulted to the user with the changes that have taken by extracting the knowledge from the log collected.

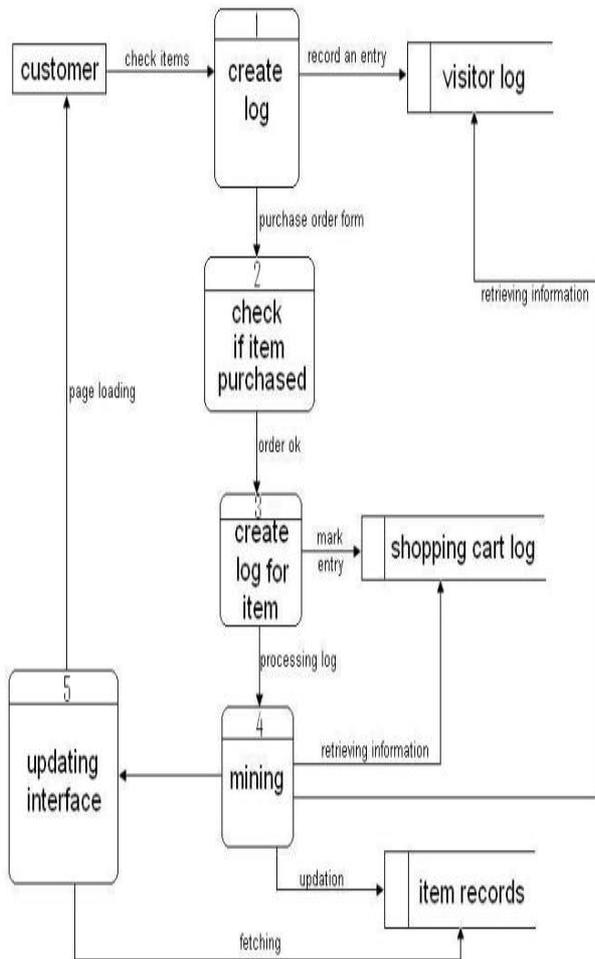


Fig. 3: DFD of proposed work

Above described figure 2 describes the data flow in the whole project. Let us explain the data flow diagram, in which a visitor's IP address is tracked while he checks into the website, and then further it is checked that the item is purchased or not, if the item is purchased then further an entry to the shopping cart file is maintained, and after that a procedure for mining is called in that mines the visitor's log and items purchased log from which the frequency for various items purchased is traced and is used to update the item's database and finally rank updation of the items is done which further updates the front end displayed to the user.

To design the mechanism following steps are taken

- A. Log creation
- B. Mining the log
- C. Frequency updation
- D. Rank updation
- E. Front end updation

A. Log creation – The very first step of the implementation is log creation, in which the IP address of the visitor is tracked and is stored in a log file after that it is checked that whether an item is purchased or not if yes then an entry for that item

is marked in the log file.

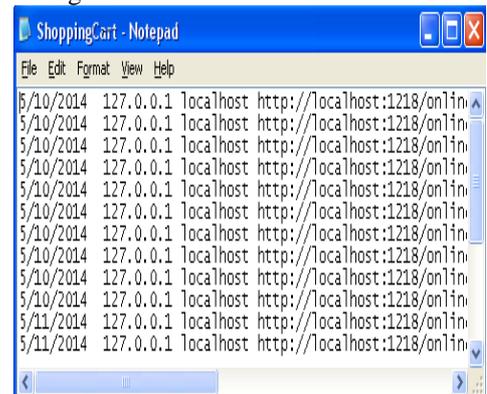


Fig. 4: Log file

B. Mining the log- Mining the log is an important step of the implementation of adaptive web interface, in which the useful information is acquired then is used for implementation.

C. Frequency updation- After mining the log, it comes into knowledge that which item is accessed how many times and which one is not accessed at all and from that the frequency of the access of the items is tracked and stored in a file for further reference.

D. Rank updation – Rank is the parameter that is attached to each and every item in a project, it describes the position where the item will appear in the front end and will be displayed to user. Rank is updated according to the frequency that the item that has maximum frequency attached with it will be assigned the 1st rank, and similarly the item with the minimum frequency will be assigned the minimum rank and will be present at last of the website page according to its rank. And similarly the ranks for the another items have been considered.

E. Front end updation- This is the last step in the implementation in which the updated front end will result to the user, this front end will get updated when ranks for items will be updated.

V. CONCLUSION

Adaptive web interfaces are investigated in this paper. Throughout the access log is taken as an input to the whole project in which the log files are mined and required knowledge is gathered and later on that results are used for updating the ranks of the items and which will further change the front end of the site. Using the whole mechanism, it is shown that adaptive web interfaces can be implemented and can serve the user in a better way.

VI. FUTURE SCOPE

Adaptive web interfaces can serve the users in the better and much more attractive way. These can also reduce the loads from the servers by keeping only required data that must be displayed to the users. Transactions between the numbers of pages will also reduce by the drastic amount as the required thing will be present in the starting and this thing will also help the users to explore the things in a better way.

VII. ACKNOWLEDGEMENT

The author is grateful to Dr. Himanshu Aggarwal, Professor, Department of Computer Engineering, Punjabi University, Patiala for his support and guidance.

REFERENCES

- [1] R. Khare and A. Rifkin, XML: a door to automated Web applications, *IEEE Internet Computing* 1(4) (1997) 78–87.
- [2] A. Fox, S. Gribble, Y. Chawathe, and E. Brewer. Adapting to Network and Client Variation Using Infrastructural Proxies: Lessons and Perspectives. *IEEE Personal Communications*, 5(4):10-19, 1998
- [3] M. Fernandez, D. Florescu, J. Kang, A Levy and D. Suci, System demonstration — Strudel: a Web-site management system, in: *ACM SIGMOD Conf. on Management of Data*,1997
- [4] E. Andre´, W. Graf, J. Mu¨ller, H.-J. Profitlich, T. Rist and W. Wahlster, AiA: adaptive communication assistant for effective infobahn access, Document, DFKI, Saarbru¨cken, 1996.
- [5] T. Rist, E. Andre´ and J. Mu¨ller, Adding animated presentation agents to the interface, in: *Proc. 1997 Int. Conf. on Intelligent User Interfaces*, Orlando, FL, 1997, pp. 79–86.
- [6] Mike Perkowitz, Oren Etzioni, Department of computer science and Engineering, Box 352350, University of Washington, Seattle, WA,98195{map,etzioni}@cs.washington.edu,(206)616-845Fax(206)543-969, Adaptive web sites: Automatically Synthesizing Web pages
- [7] Qiang Yang, Charles X. Ling and Jianfeng Gao, Department of computer science Hong Kong University of science and technology Clearwater bay, Kowloon Hong Kong, Department of Computer Science, University of western Ontario, Ontario N6A 5B7, Canada, Microsoft Research Asia, China, Mining Web logs for Actionable Knowledge.