

A proposal for design personalization of WEB page

Raghad Mohammed Hadi

Iraqi commission for computer and informatics
informatics of postgraduate studies.

Abstract

Web personalization process viewed as an application of data mining requiring support for all the phases of a typical data mining cycle. These phases include data collection and preprocessing, pattern discovery and evaluation, and finally applying the discovered knowledge in real-time to mediate between the user and the Web. So the proposed system specially designed to availability for changing visual setting of sites from user and saving the changes (colors, fonts and positions of some parts of site). This setting different from user to another and the Site is always display automatically all information and links that the user need and likes. This can be done from studding the behavior of user through interval of time (like one week) and the site containing suitable design of database (expandable, can be used from more than one user at the same time and can be connect with it locally or remotely). So the required files are filtered in an efficient manner according to our algorithm, moreover from this algorithm we improve our system's efficiency, frequency and memory.

Keywords:

Search engine – Personalization – Visual effects – Web – Databases – Static Sites – Dynamic Sites – Scripts – HTML (Hyper Text Markup Language) – ASP (Active Server Pages).

1. Introduction

The continuous growth in the size and use of the World Wide Web imposes new methods of design and development of online information services. Most Web structures are large and complicated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. On the other hand, the e-business sector is rapidly evolving and the need for Web marketplaces that anticipate the needs of the customers is more evident than ever. [1]

Therefore, the requirement for predicting user needs in order to improve the usability and user retention of a Web site can be addressed by personalizing it. Web personalization is defined as any action that adapts

the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational behavior and individual interests, in combination with the content and the structure of the Web site. The objective of a Web personalization system is to "provide users with the information they want or need, without expecting from them to ask for it explicitly".[2]

Web data are those that can be collected and used in the context of Web personalization. These data are classified in four categories:

- Content data are presented to the end-user appropriately structured. They can be simple text, images, or structured data, such as information retrieved from databases.
- Structure data represent the way content is organized. They can be either data entities used within a Web page, such as HTML or XML tags, or data entities used to put a Web site together, such as hyperlinks connecting one page to another.
- Usage data represent a Web site's usage, such as a visitor's IP address, time and date of access, complete path (files or directories) accessed, referrers' address, and other attributes that can be included in a Web access log.
- User profile data provide information about the users of a Web site. A user profile contains demographic information (such as name, age, country, marital status, education, interests etc.) for each user of a Web site, as well as information about users' interests and preferences. Such information is acquired through registration forms or questionnaires, or can be inferred by analyzing Web usage logs. [3]

The overall process of usage-based Web personalization consists of five modules, which correspond to each step of the process. These are as follows.

- User profiling stored in Web server logs is processed by applying data mining techniques in order to (a) extract statistical information and discover interesting usage patterns, (b) cluster the users into groups according to their navigational behavior, and (c) discover potential correlations between Web pages and user groups. This process of extracting information concerning the browsing behavior of the users can be regarded as part of the user profiling process. It is therefore evident that the user profiling and Web usage mining modules overlap.

- Content management: This is the process of classifying the content of a Web site in semantic categories in order to make information retrieval and presentation easier for the users. Content management is very important for Web sites whose content is increasing on a daily basis, such as news sites or portals.
- Web site publishing: A publishing mechanism is used in order to present the content stored locally in a Web server and/or some information retrieved from other Web resources in a uniform way to the end-user. Different technologies can be used to publish data on the Web.
- Information acquisition and searching: In many cases information provided by a Web site is not physically stored in the Web site's server. In the case of a Web portal or vortal (vertical portal), users are interested in information from various Web sources. [1]

2Theoretical background:- web personalization

What is Personalization?

Personalization can be defined as the ability for a web user to customize the content and layout of their own portal web page. It is one of the most popular ways of increasing traffic at portal sites today, and helps to ensure return customers. Today, the structure of the web causes the user go from site to site looking news and information. A personalized page brings news and information to the user. It allows to see what the user want, when he want. With a personalized page, the user can check ours stocks, ours weather, and ours sports teams. With information scattered all over the web, the Ability to create his own on-line, real-time interactive newspaper is quite valuable. Even more, the user can now financially benefit from the web. For years, his Investments have been constrained by having to deal with stockbrokers on their schedule. The user can now manage his investments on his schedule. [2]

The web personalization process is illustrated in Figure 1. Using the four aforementioned sources of information as input to pattern discovery techniques, the

System tailors the provided content to the needs of each visitor of the web site. The

Personalization process can result in the dynamic generation of recommendations, the

Creation of index pages, the highlighting of existing hyperlinks, the publishing of targeted advertisements or emails, etc. In this research we

focus on personalization systems that aim at providing personalized recommendations to the web site's visitors. Furthermore, since the personalization algorithms we propose in this work are generic and applicable to any web site, we assume that no explicit knowledge involving the users' profiles, such as ratings or demographic information, is available.

The steps of a Web personalization process include: (a) the collection of Web data, (b) the modeling and categorization of these data (preprocessing phase), (c) the analysis of the collected data, and (d) the determination of the actions that should be performed. The ways that are employed in order to analyze the collected data include content-based filtering, collaborative filtering, rule-based filtering, and Web usage mining. The site is personalized through the highlighting of existing hyperlinks, the dynamic insertion of new hyperlinks that seem to be of interest for the current user, or even the creation of new index pages.

Content-based filtering systems are solely based on individual users' preferences. The system tracks each user's behavior and recommends items to them that are similar to items the user liked in the past.

Collaborative filtering systems invite users to rate objects or divulge their preferences and interests and then return information that is predicted to be of interest to them. This is based on the assumption that users with similar behavior (e.g. users that rate similar objects) have analogous interests.

In rule-based filtering the users are asked to answer a set of questions. These questions are derived from a decision tree, so as the user proceeds to answer them, what he finally receives as a result (e.g. a list of products) is tailored to his needs. Content-based, rule-based, and collaborative filtering may also be used in combination, for deducing more accurate conclusions.[1]

Information retrieval technologies have matured in the last decade and search engines do a good job of indexing content available on the Internet and making it available to users, if the user knows exactly what he is looking for but often, search engines themselves can return more information than the user could possibly process. Also, most widely used search engines use only the content of Web documents and their link structures to assess the relevance of the document to the user's query. Hence, no matter who the user of the search engine is, if the same query is provided as input to the search engine, the results returned will be exactly the same.

The goal of personalization is to provide users with what they want or need without

Requiring them to ask for it explicitly. This does not in any way imply a fully automated process, instead it encompasses scenarios where the user is not able to fully express exactly what they are looking for but in interacting with an intelligent system can lead them to items of interest.[4]

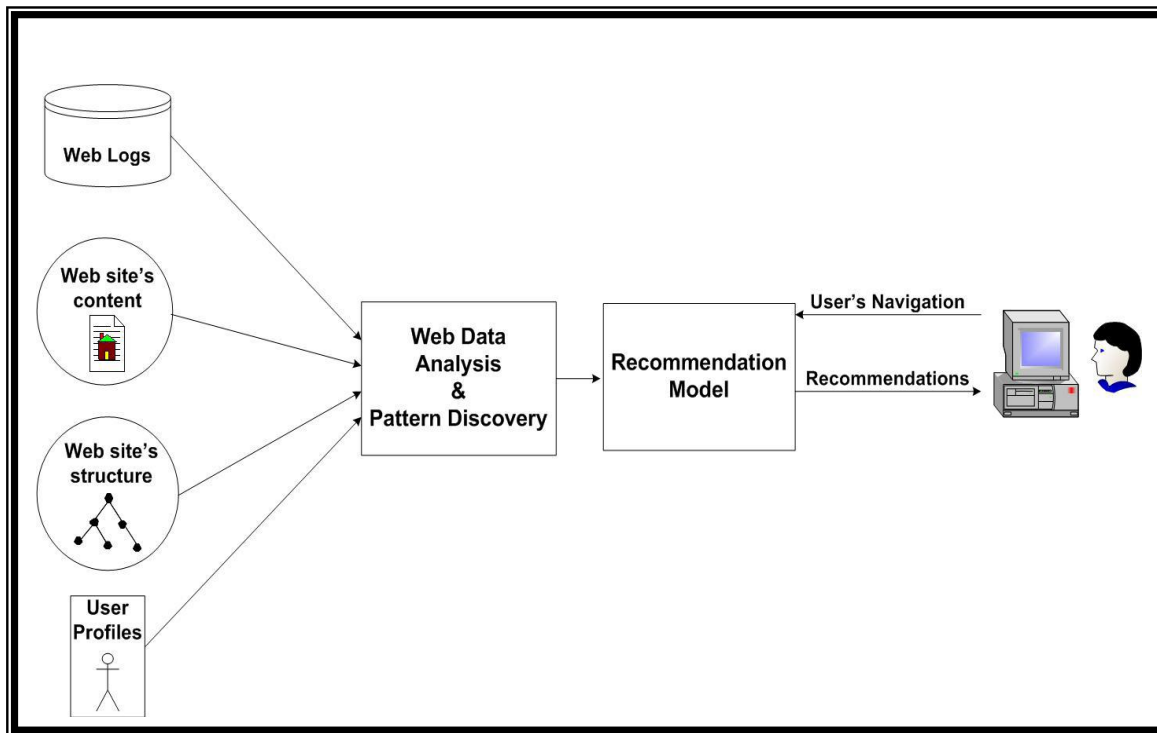


Figure 1. The web personalization

Common Features

Common content features provided by the portals include local and national news, weathers, sports, horoscopes, TV listings, movie listings, ski reports, maps, people finders, yellow pages and favorite links. Some even offer office suitelets, such as an address book and a calendar. Common layout features include selecting the number of columns, ordering of content areas, multiple pages, selecting a greeting and setting colors. [2]

3. Classifications of approaches to personalization

In this section the user discusses various dimensions along which personalization systems can be classified based on the data they utilize, the learning paradigm used, the location of the personalization and the process that the interaction takes with the user.

3.1 Individual Vs Collaborative

The term personalization impresses upon the individuality of users and the need for systems to adapt their interfaces to the needs of the user.

This requires data collected on interactions of users with the system to be modelled in a user-centric fashion. Typically, data is collected by the business with which the user is interacting and hence the business has access to data associated with all its customers.

A personalization system may choose to build an individual model of user likes and dislikes and use this profile to predict/tailor future interactions with that user. This approach commonly requires content descriptions of items to be available and are often referred to as *content-based filtering systems*. The process of building a profile for a user requires the transformation of each article into a bag or words representation, with each token being assigned a weight using some learning method such as *tfidf* [10] or minimum description length [11]. The profile is then used to recommend articles to the user.

3.2 Reactive Vs Proactive

Reactive approaches view personalization as a conversational process that requires explicit interactions with the user either in the form of queries or feedback that is incorporated into the recommendation process, refining the search for the item of interest to the user. Reactive systems can be further classified based on the types of feedback they expect from the user. Common feedback mechanisms used by these systems include value elicitation, rating and preference feedback.

3.3 User Vs Item Information

Personalization systems vary in the information they use to generate recommendations. Typically, the information utilized by these systems includes:

- Item Related Information: This includes content descriptions of the items being recommended and a product/ domain ontology
- User Related Information: This includes past preference ratings and behavior of the user, and user demographics Systems that use item related information generally deal with unstructured data . Once this data has been processed, into relational form such as a bag-of-words representation commonly used for textual data, a user profile is generated. [2]

4Automatic Personalization and Data Mining

The ability of a personalization system to tailor content and recommend items implies that it must be able to infer what a user requires based on previous or current interactions with that user, and possibly other users. The personalization task can therefore be viewed as a prediction problem: the system must attempt to predict the user's level of interest in,

or the utility of, specific content categories, pages, or items, and rank these according to their predicted values. Furthermore, the task of delivering personalized content is often framed in terms of a recommendation task in which the system recommends items with the highest predicted interest values or utilities to an active user. In general, a personalization system can be viewed as a mapping of users and items to a set of "interest values". The view of personalization function as a prediction task comes from the fact that this mapping is not, in general, defined on the whole domain of user-item pairs, and thus requires the system to estimate the interest values for some elements of the domain. Automatic personalization systems, generally, differ in the type of data and the method used to create user profiles, and in the type of algorithmic approaches used to make predictions. [3]

5. General system structure

The web, it is a small word written in English, but really now through the web we can do any thing (learning, chatting, browsing sites, searching for any things, read information and news). Any information always updated everyday so it is up to date. The World Wide Web (also called "WWW") are collections of sites designed to do some jobs running remotely through the special software called "Browser", these sites have special address called Universal Resource Locator "URL" used to call it. Each site is collection of pages designed usually using programming languages and HTML. These sites of two type (Static Sites and Dynamic Sites). Always dynamic site is one that uses databases. In our research we use dynamic web sites to personalize the web site through personalization of visual affect of site and the information that interested from user of this site, this make the site dynamically changed from user to another but it is one copy of site installed in one sever and available for any user. This mechanism can be implemented in any site. This site need database used to save updated setting for each user that use this site. This can be done from studding the behavior of user through interval of time (like one week), we uses search engine in our research as one of important types of sites, and uses Database design and programming, this is very important for learning how to design suitable database for saving any information about user that is differs from one user to another. One of the usually used is Microsoft SQL Server 2000 Enterprise Edition. Our system is design to be applicable to any existing website, Web site content including user profiles, web pages and usage logs. Although the issue of user privacy will arise, since our system only performs actions on the user information locally, we think

it is still within safety concerns. A larger system that will work on multiple websites for user overall browsing personalization is beyond the scope of this research could be investigated in future works. The system can be divided into two major stages:

- **The client side:** which the data we collected. In order to perform user personalization, web data must be collected. web data can be classified Into four major categories:
 - 1 **Content data** represent the content of the website. These are presented to the end-user appropriately structured. They can range from simple text to images to even structured data such as information retrieved from web databases.
 - 2 **Structure data** represent the way content is organized within a web site. They can be either data entities used within a webpage, such as HTML or XML tags, or data entities used to put a website together, such as hyperlinks connecting one page to another.
 - 3 **Usage data** represent general information collected by the website on its users, such as a visitor's IP address, time and date of access, complete path (files or directories) accessed, referrer's address, and other attributes that can be included in a Web access log.
 - 4 **User profile** data provide information on specific users of the website. A user profile contains demographic information (such as name, age, country, marital status, education, interests, etc.) for each user of the website, as well as information on user interests and preferences. Such information is typically acquired either explicitly through registration forms or questionnaires, or can be implicitly inferred by analyzing web usage logs. In the client side we used content data and user profile as data collected.
- **The server side:** in order to personalization a web site, the system should be able to describe each distinct user. And we call this process user profiling. Our system use approaches that create a knowledge base that contains user preferences, characteristics and activities. And we exclude the user activities and put it into the user behavior part since it is dynamically changing and does not fit our offline data preprocessing concept. Although user preference sometimes changes, it is much less likely when compared with user browsing activities. User profiling could simply be a registration step. Users describe their preferences and characteristics through registration forms. If the website doesn't require a registration step, we can internally create an empty one for each user

containing the user IP address in order to distinguish different log-ins. Based on different functionalities.

The architecture of the proposed system can be explained by the diagram in the figure (2). And the Client side shown in flowchart (1), the Server side shown in flowchart (2), and the software requirements in our research are:

- 1-Windows XP with IIS (Internet Information Services) installed or any Windows Server.
- 2- Microsoft SQL Server 2000.
- 3-Microsoft Visual Interdev 6.0 (for programming Java script, ASP active server page and HTML).

And in order to examine the viability of our framework, we incorporate and implement it over a well designed simulation environment.

- 1-Two PCs (database and my site can be in the same PC) or more. Ram 512 MB for (client and database) computers, 1 GB for server computer. CPU 2 GHz for each PC and any size of Hard Disk drive.
- 2-LAN (Local Area Network) Cards with UTP cables for connections and one HUB device. Or Wireless Network Cards.

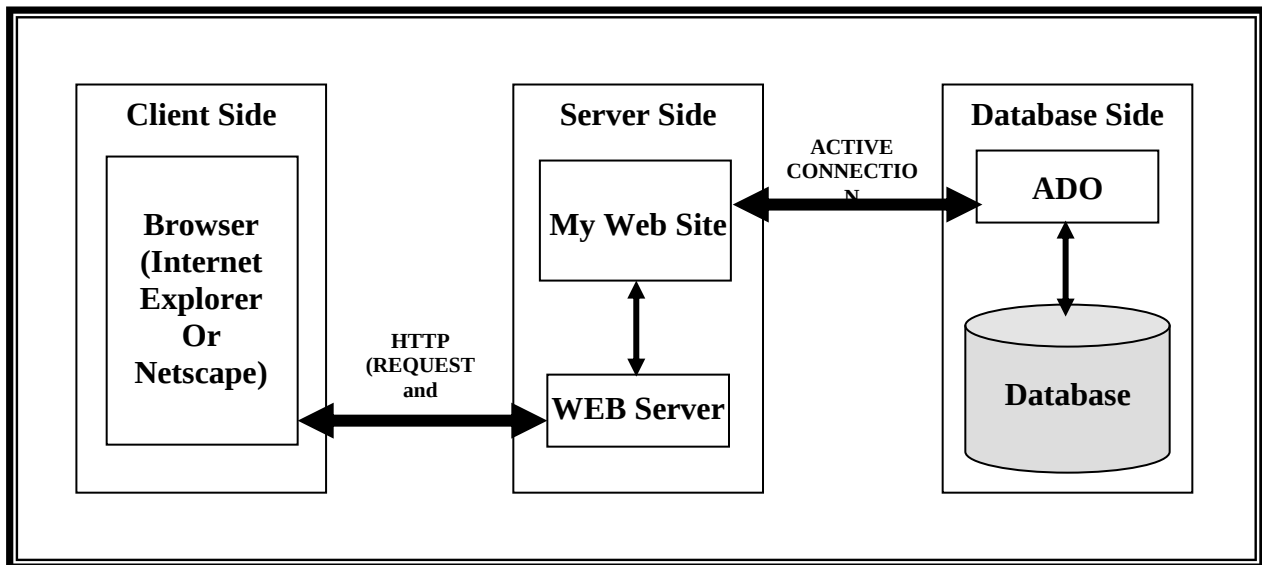
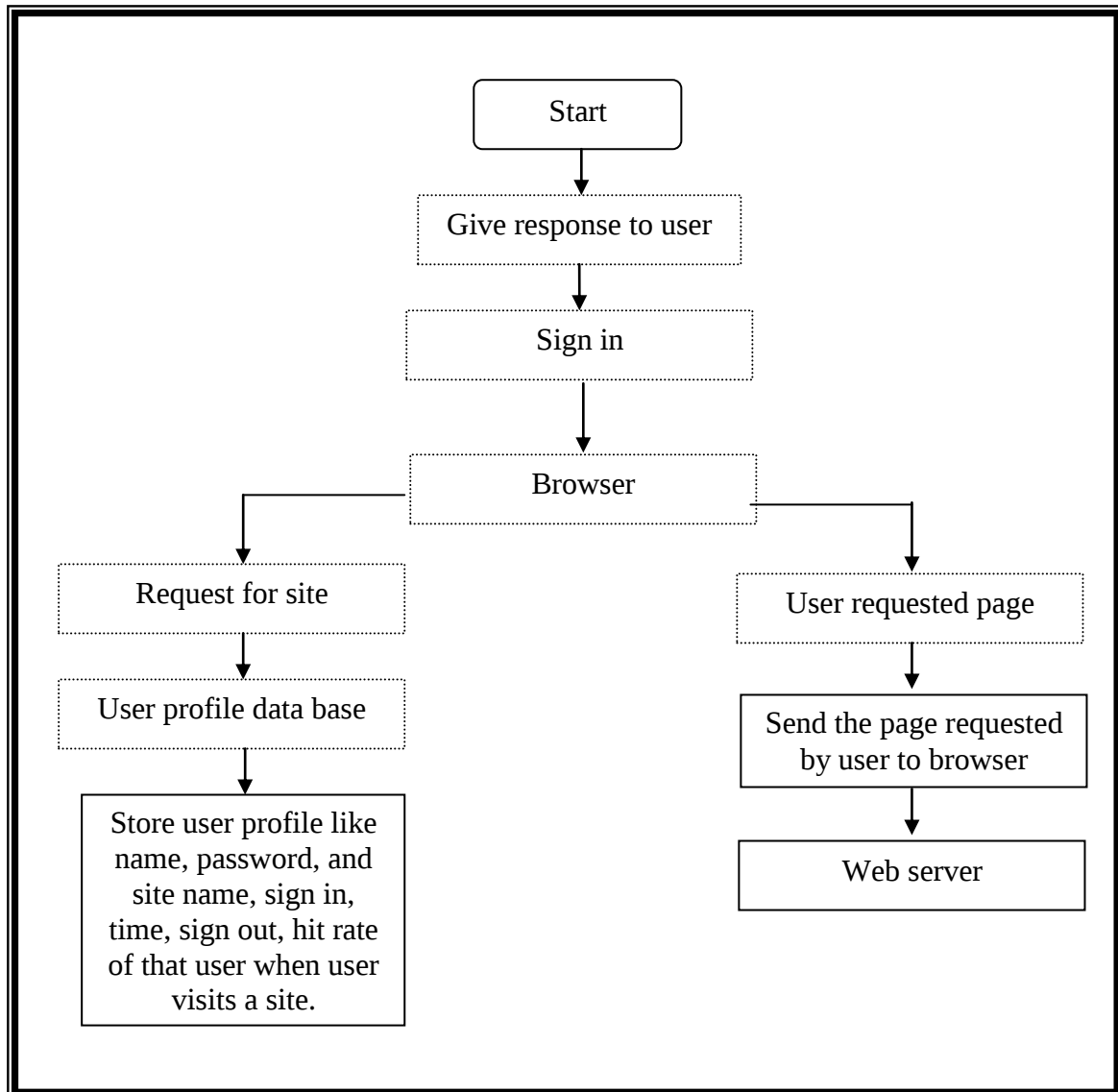
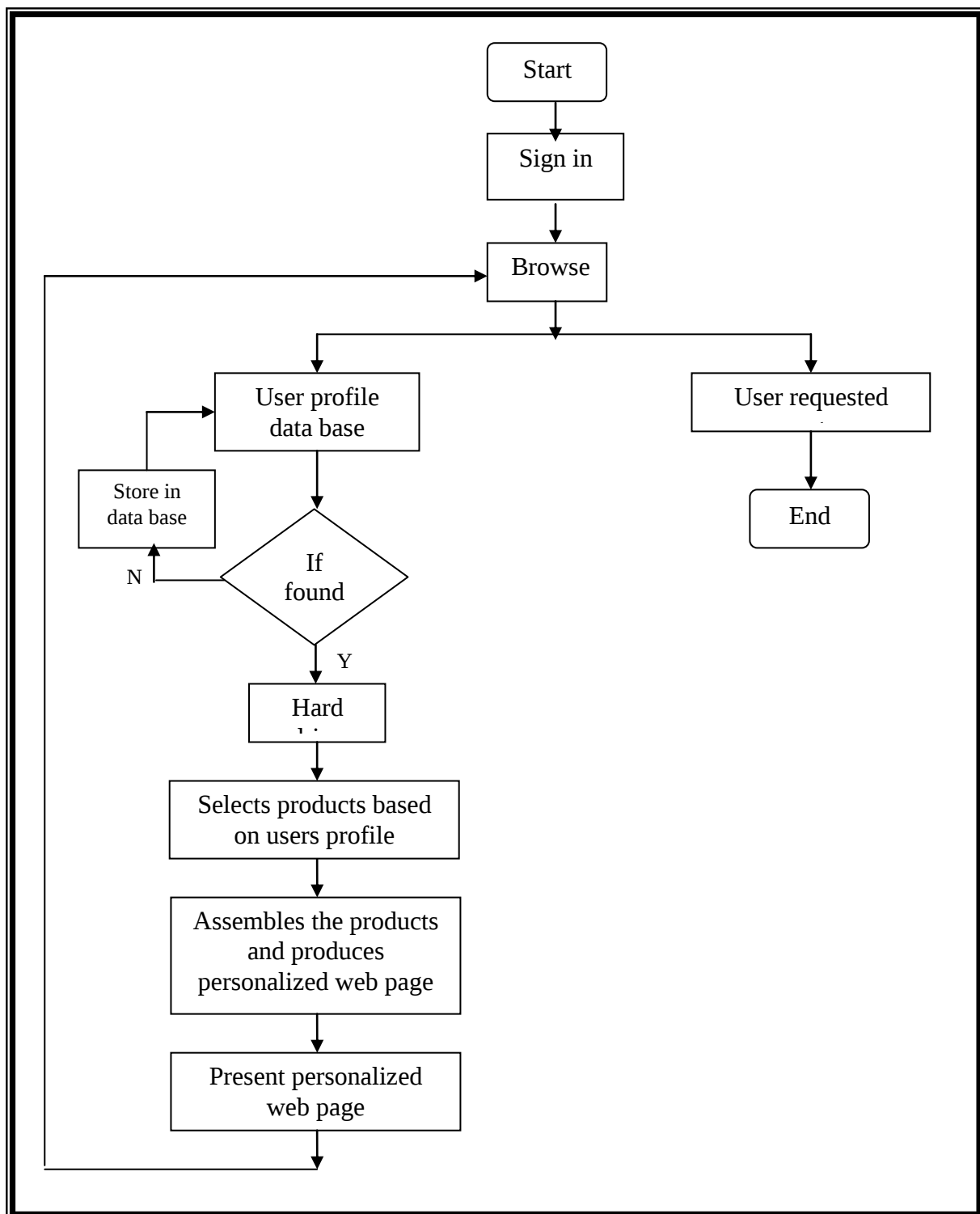


Figure (2): Proposed System Architecture



Flowchart (1): The client side



Flowchart (2): The server

6. Conclusions

1. The ability to provide personalized web content is vital to the success of the web. Personalization designers should strive to make these features simple and efficient. Two things to keep in mind during design are to keep graphics to a minimum, and to design pages for 800X600 display resolution. For most users, the web is still a very slow experience.
2. Web personalization is the process of customizing the content and structure of a Web site to the specific and individual needs of each user, without requiring them to ask for it explicitly. This can be achieved by taking advantage of the user's navigational behavior, as revealed through the processing of Web usage logs, as well as the user's characteristics and interests. Such information can be further analyzed in association with the content of a Web site, resulting in improvement of the system performance, users' retention, and/or site modification.
3. The overall process of Web personalization consists of five modules, namely: user profiling, log analysis and Web usage mining, information acquisition, content management, and Web site publishing.
4. User profiling is the process of gathering information specific to each visitor to a Web site either implicitly, using the information hidden in the Web logs or technologies such as cookies, or explicitly, using registration forms, questionnaires, and the like. Such information can be demographic, personal, or even information concerning the user's navigational behavior.
5. The main component of a Web personalization system is the usage miner. Log analysis and Web usage mining is the procedure where the information stored in the Web server logs is processed by applying statistical and data mining techniques such as clustering, association rules discovery, classification, and sequential pattern discovery, in order to reveal useful patterns that can be further analyzed.

7. References

- [1] Magdalini Eirinaki and Michalis Vazirgiannis "Web mining for web personalization" department of informatics, Athens university of economics and business, 2008.
- [2] Harris Kravatz "Designing web personalization features", HK interface design Inc, 2007.
- [3] Bamshad Mobasher, "Data mining for web personalization", Depaul University, Chicago, Illinois, USA, 2009.
- [4] Sarabjot Singh Anand and Bamshad Mobasher, "Intelligent Techniques for web personalization", Department of Computer Science University of Warwick, 2008.

مقترح لتصميم صفحة ويب متخصصة

رغد محمد هادي *

*الهيئة العراقية للحاسبات والمعلوماتية-معهد المعلوماتية للدراسات العليا.

الخلاصة

عملية التخصيص الويب ينظر إليها على أنها تطبيق لاستخراج البيانات التي تحتاج إلى دعم لجميع مراحل دورة استخراج البيانات النموذجية. هذه المراحل تشمل جمع البيانات وتجهيزها ، واكتشاف نمط والتقييم ، وتطبيق المعرفة وأخيرا اكتشفت في الوقت الحقيقي للتوسط بين المستخدم والشبكة العالمية. لذلك النظام المقترح صمم خصيصا لتوفير التغيير وضع البصرية من المواقع من المستخدم وحفظ التغييرات (الألوان والخطوط والمواقف من بعض أجزاء من الموقع). هذا التغييرات مختلفة من مستخدم لآخر ، والموقع هو دائما تلقائيا يعرض جميع المعلومات والروابط التي تلبي الحاجة المستخدم ويحبها. ويمكن القيام بذلك من الترتيب سلوك المستخدم خلال فترة من الزمن (مثل أسبوع واحد) ويمكن استخدام موقع يحتوي على تصميم مناسب لقاعدة البيانات (قابلة للتوسيع ، من أكثر من مستخدم واحد في نفس الوقت ، ويمكن أن تواصل معها محليا أو عن بعد). هكذا يتم تصفية الملفات المطلوبة بطريقة فعالة وفقا لأنظمتنا ، إضافة على ذلك ان هذه الخوارزمية قد حسنت نظامنا بالكفاءة اعلى ، وتردد الذاكرة.

الكلمات المفتاحية:

محرك البحث -التخصيص- تاثيرات بصرية- ويب- قواعد البيانات- مواقع ثابتة- مواقع الحيوية- مخطوطات- HTML (لغة ترميز النصوص التشعبية)- ADO (احدث وجوه البيانات) .