An Overview of Personalized Recommendation System to Improve Web Navigation

S. Ancy[1], V. Subhashini[2], R. Pooja Karpagam[3], S. Sujeethra[4], L. Jayachitra[5]
Department of Information Technology, Jeppiaar Institute of Technology, India
ancys@jeppiaarinstitute.org, subhashinivelusamy@gmail.com, poojarajan1895@gmail.com, sujeethra94@gmail.com,
jayachitraloganathan@gmail.com

Abstract:-We present a new personalized recommendation system, which means the searches of each user is done according to their interest which is based on ranking or preference method. It also maintains the logs which records the sessions of each user and brings out the exact data required by the user. This is done by fetching the data that is already stored in the database. Web server logs maintains history of page results and consists of a log file which automatically creates and maintains the list of activities performed by the users. For extracting the data according to the user's previous searches, we are using Stemming Algorithm. The Stemming Algorithm is a process where the exact, meaningful words are extracted from the URL. Because of this process the user's search time will be reduced. It also improves the quality of web navigation and overcomes the limitation of existing system. In the proposed system we extract user's behaviour from web server logs in the actual process whereas, in the anticipated system, the user's behaviour is done with the help of cognitive user model and we perform the comparison between the two usage processes. The data produced from this comparison can help the users to discover usability issues and take actions to improve usability. In the anticipated usage the cognitive user model is done that can be used to simulate or predict human behaviour or by performance and task. Finally, the system is executed by using the top-k ranking algorithm. The advantage of this system are accuracy and better processing speed. The user's convenience deals with the ease of navigation which helps the users to interact with their interface.

_____***** _____

I. INTRODUCTION

The process of examining data from different views and abstracting it into useful information and the information that can be used to intensify rate, minimize cost, or together is called Data Mining. The software Data Mining is a tool for analyzing data. It helps users toanalyze data from many different dimensions, differentiate it, and summarize the relationships identified. The process of finding associations or designs among dozens of fields in large relational database is called data mining. Data that the user enters can be facts, numbers, or text that the computer processes. Today, organizations are gathering huge and developing bulk of data in different formats. The operational or transactional data such as, sales, cost, inventory, payroll, and accounting and Non-operational data, such as industry sales, forecast data, and macro-economic data are contained in it. Data about the data itself, such as rational database design or data dictionary definitions is called as Meta-Data. The information is provided by the patterns, association, or relationships among this data. Analysis of selling point of sale transaction data can yield information on which products are selling and when can be said as an example. Information canbe converted into knowledge about ancient patterns and future trends.

Considerable approach in data capture, processing power, data transmission, and storage capabilities are enabling organizations to merge their various databases into data warehouse. The process of collecting data management and recovery is called Data warehousing. Like data mining, Data Warehousing is a relatively new term, though the concept itself has been around for years. Data Warehousing represents an imaginary vision of maintaininga central repository of all organizational data. Centralization of data is needed to expand user gains and analysis. Considerable technological approaches are making this vision a reality for many companies. And, equally fair advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports Data Mining. While large-scale information technology has been evolving separate transaction and logical systems, data mining provides the connection between the two. Data mining software analyses relationships and patterns in stored deal is data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Four types of relationship exists. Data can be stored as classes and is used to locate data in predetermined groups. Clusters means data items are grouped according to logical relationships or consumer preferences. Association's data can be mined to identify associations. Sequential patterns data is mined to anticipate behaviordesign and trends.

Web Navigation is important to the victory of your website user's familiarity to your website. The website's navigation system is similar to a road map of all the different areas and information contained within the website. Using a consistent navigation plan from page to page helps the website user learn your website navigation system. To be effective, the website navigation system needs consistency, ease of use

throughout the website. Website navigation has two main functions to tell the user wherethey areandto enable the user to go somewhere else**.**Creating the website navigation network at theplanning stage of the website will affect the overall design of the web page layout and help establish the overall plan for the website.

II.   RELATED STUDY

A. An Integrated Theory of Minds

In the paper an integrated theory of mind [1], a concept called Adaptive control of thought-rational (ACT-R) is explained. According to this concept the data will be divided into many small modules and finally all the modules are integrated together to produce the expected result. This theory contains three modules namely perceptual-motor modules which explains ACT-R focuses on accuracy rather than perception. The next module is goal module which explains that the humans have the advantage of abstracting the contents than other species. The third module is the declarative memory module that contains the information obtained from the goal module. These three modules help to provide data as per the human's interest. The disadvantage of this theory is it is a complicated process.

B. Building on the usability Study: Two Explorations on How to Understand an Interface

In this paper two studies are implemented to better understand the interest of the users. The first study deals with combining expert and beginner's performance data to yield a measure of spontaneity. According to Building on the usability study: Two explorations on how to understand an interface [2], the second study contains a method that combines both verbal and nonverbal emotion to understand the impact of users on our products. It contains a long procedure to obtain the performance data of the users.

C. Identifying Web Sessions Using Simulated Annealing

Providing efficient services by a website according to the user's interest is the most difficult thing to do. It can be done by a web log that records the pages that are visited by the users. The reconstruction of web pages visited by the users is known as the web sessionization problem. This is overcome by an integer programming model. According to this model, web log is reconstructed over a period of time. But it requires a long computational time. This paper identifying web sessions using simulated annealing [3], provides an approach called simulated annealing which reduces the time up to 166 times compared to the previous model. The method cannot

tell whether it has found an optimal solution. Some other complimentary method is required to do this.

D. Research Based Web Designs and Usability Guidelines

This paper research based web designs and usability guidelines [4], provides the guidelines toassist who develops websites to base their decisionson the most current and best available evidence. The guidelines are mostly for the information-oriented sites but also applicable to other commercial sites. These guidelines were to create better and more usable websites and to translate the information in the websites to practical and easy-to-use format. These guidelines are applicable to the following persons: designers, usability specialists, researches and managers.

E. Internet Web Services: Workload Characterization and Performance Implications.

This paper presents a method to study the workload characterization of the internet servers. Six different data sets are used for the study. This data set represents three different orders of magnitude in server activity and two different orders of magnitude in time duration, ranging from one week of activity to one year. The workload characterization depends on the type of the document, size of the document, behavior of the document and the geographic location from where the server got the request. This study gives importance to the useful workload characteristics that are common to all the data sets. The disadvantage of this paper internet web services: workload characterization and performance implications [5], is caching that is storing and retrieving the data and it is not sure whether the observed workload characters will enhance the performance.

F. Defining and Testing EMR Usability

EMR (Electronic medical record) is a computer system which holds the health information documents and store patient information. Patient safety is concern in this paper. It deals with the usability principles such as simplicity, efficient interaction and consistency. In this principle it achieves the efficiency of use. And some usability evaluation and rating methods are done such as effectiveness, efficiency and safety as the object methods to evaluate EMR usability. And 5-star rating system which helps the EMR purchasers about an EMR's usability. Limitations of the paper defining and testing EMR usability [6], are Performance testing, Expert review, Risk assessment.

G. ACT-R/PM and Menu Selection: Applying a Cognitive Architecture to HCI

In this paper ACT-R/PM and menu selection: applying a cognitive architecture to HCI [7], it tells us the interaction with the user GUI devices such as three components: the cognitive, perceptual, and motor capabilities of the user. The Human Computer Interaction (HCI) is created for human performances on the menu selection, where the final modules provides zero-parameter fit. One of the system is ACT-R/PM which is the architecture with a set of perceptual-motor modules, done asinteractive task with those 3 components. The application of this architecture is simple, and rapid tasks takes place for more complex high performance tasks such as random menu selection will be considered. The advantages are avoidance of free parameters and the performance will be high.

H. Web Usability Probe: a Tool for Supporting Remote Usability Evaluation of Web Sites

WUP, a tool is used here that allows the users to decide what tasks they should perform, and gather many types of data related to user interaction. The tool provides some graphical representations, which allows the user to analyze the data collected from a usability perspective. The limitations of this paper, web usability probe: a tool for supporting remote usability evaluation of web sites [8] is, it does not allow the user to know exactly the elements referenced by the trigged events if any usability problem is identified, which complicates its correction. The cost is also high.

I. CODEIN- a new notation for GOMSto handle evaluations of reality-based interaction style interfaces

In this, a new diagrammatic notation system for Goals, Operators, Methods, Selection rules (GOMS), called Codein, with extensions to support the evaluation of Reality Based InteractionStyles is used. Theproposed notation of this paper CODEIN- a new notation for GOMS to handle evaluations of reality-based interaction style interfaces [9], gives added power to GOMS to model and evaluate the task completion time of parallel actions during the performance of a task, which was possible only with the use of CPM-GOMS in the previous model. The limitation is it is complicated to use.

J. Data Preparation for Mining World Wide Web Browsing Patterns

This paper presents several Data Preparation techniques in order to identify unique user and user sessions. Also, a method to divide sessions into meaningful transactions is defined and tested. The sessions are useful in the reconstruction of websites as per the user interest. But the disadvantage of this paper data preparation for mining World Wide Web browsing patterns [10], is that they are not fully automated and needs assistance. There is a chance of missing information as a resulting of client or proxy site caching. These data can be inferred from the server logs.

K. Supporting Activity Modelling From Activity Traces

This paper supporting activity modelling from activity traces [11], it helps the analyst to build ontology of the activity, that is, a set of symbols and hierarchical semantics that supports the construction of activity models. By constructing the activity models it is possible to trace the user's activity. But the disadvantage is the inability to trace the previoussessions of the users. There are some tools and diagrammatic notations to improve interaction and performance.

L. Designs for Explaining Intelligent Agents

This paper designs for explaining intelligent agents [12], describes the designs that can be reused to create intelligent agents capable of explaining themselves. The designs include ways to provide ontological, mechanistic and operational explanations. Some of the software engineeringtechniques have been applied to develop intelligent agents and some cognitive models.
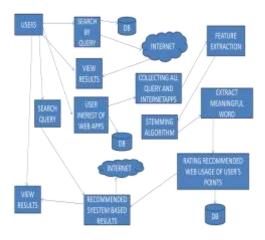
M. The Value of Web Log Data in Use-Based Design and Testing

In the paper, The Value of Web Log Data in Use-Based Design and Testing [19], the Web-based logs that contain potentially useful empirical data with which World Wide Web (Web) designers and design theorists can assess usability and effectiveness of design choices. Most Web design guidelines from artistic or usability principles feature no empirical validation, while empirical studies of Web use typically rely on observer ratings. Web server logs and client-side logs can provide naturally-occurring, unobtrusive usage data, partially amenable to normative use assessments but particularly useful in experimental research comparing alternative Web designs. Identification of types of Web server logs, client logs, types and uses of log data, and issues associated with the validity of these data are enumerated.

N. Simplifying the Development and the Analysis of Cognitive Models

Usability of complex dynamic human computer interfaces can be evaluated by cognitive modeling to investigate cognitive processes and their underlying structures. Even though the prediction of human behavior can help to detect errors in the interaction design and cognitive demands of the future user the method is not widely applied. The time-consuming transformation of a problem "in the world" into a "computational model" and the lack of fine-grained analysis of simulation data are mainly responsible for this in this paper Simplifying the Development and the Analysis of Cognitive Models [20].

III. ARCHITECTURE DIAGRAM

Initially, the user will enter the query that needs to be searched. The search will be stored in the database and the result will be retrieved from the internet. According to the searches made by the user the data is collected and stored in the database which is later retrieved from it. The data are extracted based on the searches by the user. An algorithm called 'Stemming' is used here to extract the meaningful word from the given data where it gives the exact, meaningful data based on the user's interest. It is also stored in the database. The results are shown based on the previous frequent searches.

IV. MODULES

Creating Search history: Any personal documents such as browsing history and emails on a user's computer could be the data source for user profiles. This focus on routine terms limits the dimensionality of the document set, which further provides a clear description of users' interest. This module allows the search engine to better understand a user's session

and potentially tailor that user's search experience according to her needs. Once query groups have been recognized, search engines can have a good depiction of the search context behind the current query using queries and clicks in the corresponding query group.

Query clustering: User's queries can be classified into different query clusters. Concept-based user profiles are hired in the clustering process to achieve personalization effect. The most similar pair of concept intersection, and then, merge the most similar pair of query nodes, and so on. Each individual query submitted by each user is treated as an individual node and each query with a user identifier. We perform the grouping in a similar dynamic fashion, whereby we first place the current query and clicks into a query group

Query reformulation: To ensure that each query group contains closely related and relevant queries and clicks, it is important to have a suitable relevance between the current query groups. We assume that users generally issue very identical queries and clicks within a short period of time. The search history of a huge number of users contains signals regarding query relevance, such as which queries likely to be issued closely together. This captures the relationship between queries frequently leading to clicks on similar URLs. Query reformulation graph and the query click graph from search logs, and how to use them to determine relevance between queries or query groups within a user's history.

History grouping: Query groups is to first treat every query in a user's history as a query group, and then merge these query groups in an iterative fashion (in a k-means). However, this is impractical in our scenario for two reasons. First, it may have the unappealing effect of changing a user'sexisting query groups, possibly undoing the user's own physical efforts in organizing the history. Next, it involves a high-computational cost, since we would have to redo a large number of query group similarity computations for every new query.

CONCLUSION

We have implemented the personalized recommendation system which improves the web navigation by simulating or predicting human behavior or by their performance on tasks. Web server logs which maintains a history of page results are used to retrieve the appropriate results. As a result, our method provides usability improvement in effectiveness (higher task completion rate) and efficiency (less time for given task) when a user successively iterates the same query. The results from the query are efficient. And we are currently implementing this model for overall quality improvement for web applications.

REFRENCES

[1] J.R. Anderson, D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere& Y. Qin, "An integrated theory of the mind," Psychol. Rev., 111(2004):1036-1060.

[2] A. Agarwal & M. Prabaker, "Building on the usability study: To explorations on how to better understand an interface," in Human Computer Interaction Springer, (2009):385-394.

[3] T. Arce, P.E. Roman, J.D. Velasquez, & V. Parada, "Identifying web sessions with simulated annealing," Expert Syst. Appl., 41(2014):1593-1600.

[4] S.J. Koyani, R.W. Bailey, & J.R. Nall, "Research-Based Web Design and Guidelines," Health Human Services, (2004).

[5] M.F. Arlitt& C.L. Williamson, "Internet Web Servers: Workload Characterization and Performance Implications," IEEE/ACM Trans. Netw., 5(1997):631-645.

[6] J.L. Belden, R.Grayson, &J.Barnes, "Defining and Testing EMR Usability: Principles and proposed methods of EMR usability evaluation and rating," Healthcare Information and Management Systems Society, (2009).

[7] M.D. Byrne, "ACT-R/PM and Menu Selection: Applying a cognitive architecture to HCI," Int. J. Human-Comput. Stud., 55(2001):41-84.

[8] T. Carta, F. Paterno, & V.F.D. Santana, "Web usability probe: A tool for supporting remote usability evaluation of web sites," in Human Computer Interaction Springer, (2011):349-357.

[9] G. Christou, F.E. Ritter, & R.J. Jacob, "CODEIN- A new notation for GOMS to handle evaluations of reality-based interaction style interfaces," Int. J.Human-Comput. Interaction, 28(2012):189-201.

[10] R. Cooley, B. Mobasher, & J. Srivastava, "Data preparation for mining World Wide Web browsing patterns," Knowl. Inf. Syst., 1.1(1999):5-32.

[11] O.L. Georgeon, A. Millle, T. Bellet, B. Mathern, & F.E. Ritter, "Supporting activity modelling from activity traces," Expert Syst.,29.3(2012):261-275.

[12] S.R. Haynes, M.A. Cohen, & F.E Ritter, "Design for explaining intelligent agents," International Journal of Human-Computer studies, 67.1(2009):99-110.

[13] M. Rauterberg, "AMME: An automatic mental model evaluation to analyse user behaviourtraced in a finite, discrete state space," Ergonomics, 36.11(1993):1369-1380.

[14] P.M. Sanderson & C. Fisher, "Exploratory sequential data analysis: Foundations," Human-Computer. Interaction, 9(1994):251-317.

[15] P.M. Sanderson, M.D. McNeese, & B.S. Zaff, "Handling complex real-world data with two cognitive engineering tools: COGENT and Mac-SHAPA," Behaviour Res. Methods, Instrum. Comput, 26.2(1994):117-124.

[16] J. Sant, A. Souther, & L. Greenwald, "An exploration of statistical models for automated test case generation," ACM SIGSOFT Softw.Eng.Notes, 30(2005):1-7.

[17] L. Tausher, & S. Greenberg, "Revisitation patterns in World Wide Web navigation, "Human Factors Comput. Syst., (1997):399-406.

[18] D.E. Kieras, "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction,"Human-Computer. Interaction, 12(1997):391-438.

[19] M. C. Burton &J. B. Walther, "The value of Web log data in use-based design and testing," J. Comput.-Mediated Commun., vol.6. 3(2001).

[20] M. Heinath, J. Dzaack, &A. Wiesner, "Simplifying the development and the analysis of cognitive models," in Proc. Eur. Cognitive Sci. Conf., Delphi, Greece, (2007):446–451.