CHAPTER 7

CONCLUSION

7.1 INTRODUCTION

Browsing pattern of users’ change from time to time to accommodate their need. Based on the fact that different user’s need different pages and their interest may vary from time to time, we proposed few methodologies to enhance traversals. Implemented with a virtual learning environment, our approaches effectively personalizes the system to the preferred course faster. The link structure visualization and the automatically changing structure of the site based on user needs enhances personalization and thus reduces traversal and search time. Rest of the chapter is organized as follows. Section 7.2 gives a summary of the work, including the findings and methodologies used. Future enhancements to the study is suggested in section 7.4.

7.2 SUMMARY OF WORK

Knowledge seekers rely on virtual learning environments and virtual labs for continuous knowledge updation. Regular learners may need continuity of topic which is not provided by most of the virtual learning environments today. The methodologies adopted to address the needs of such users are the focus of this thesis.

We addresses the issue of personalization by applying methodologies for classification, link hierarchy display and sequence analysis based on mined data from usage log sessions. The methodologies adopted are
elaborated in Chapter 3, 4 and 5. Chapter 6 briefs the evaluation of our implemented personalized virtual learning system VL Schools.

Chapter 3 deals with the classification of categories and probabilistic approach for personalization. Our study focuses on user personalization rather than web personalization. We have done the process of classifying the usage of the site and finally user personalization by 1) Finding the associated weight on a web page from sessions for a particular period of time 2) calculate the top visited categories by user and 3) personalization based on his/her visit interest. We observed that there is a direct relationship between the usage data, which is a reflection from session and server logs and the weight associated with each page. Sessions are identified from server log data segmented over specific period of time. Then the total weight on a particular page is calculated as the total number of hits on that page. The proposed methodology sets a threshold value for the number of visits. Lower and upper bounds of these threshold and distance from lower to upper bound is considered for finding the desired page accurately.

In the proposed Easy User Navigation (EUN) algorithm a novel probabilistic method based on Bayes approach for classification of user preferences is put forward. When represented session as associated weights, for weight calculation in link and out link factors are considered. Number of mouse clicks (N) on a page is taken as the key factor for weight calculation. For finding the probability of a page in a session the author observed the need of an additional constant value and termed it as user navigation constant. This is based on the observation that continuous interest shown by a user plays an important role in personalization. The user on his login will be getting a personalized page based on the factors of weight, priorities considering UNC. This personalized page could easily satisfy his browsing needs.
Chapter 4 is about the link hierarchy visualization for enhancing user personalization. For reaching a desired page user may traverse through many hyper links. The link structure visualization enable user achieve faster and efficient traversal. The system represented the graph structure of a web page to a weight tree format. The study followed a particular traversal pattern over the tree to sort the user preference reach the desired page. The methodology ranks and arranges pages in user preferred way which is found to be faster and efficient than BFS and the other weighted path methods.

In our approach of ranking pages, the count of number of visited pages are mined from usage log data rather than taking the number of logins. Sometimes users may login and can logout the session without browsing any page. Users may also login, do the necessary works on the site and leave the system without a proper logout. By taking the count of visited pages for weight calculation false login and crawlers possibility which may cause error is avoided. We proposed a user centric page ranking algorithm for ranking pages. Understanding the fact that in hyper link traversals, in link has to be given priority, we considered this in our algorithm. The UPR gives easily identifiable values of page rank.

Chapter 5 deals with sequence matching for future prediction. We have followed an extraction of data from session logs. Sessions considered at different intervals are matched to find a relationship of user traversal pattern. We considered many factors which can cause change in the patterns identified. Uncertain occurrence of items in sequence and uncertain database patterns are two among them.

A pattern growth which is more user centric, termed USP is adopted. We had first sorted the data and when the count of repeated patterns
for predicting possibility of future visit is considered a threshold based approach is adopted.

The author integrated all the above mentioned methods in the proposed virtual learning platform. VL Schools, the virtual learning system, has a total of 1846 pages with 2409 links. The usage log file is 350 MB in size. The system was developed using Java and Apache Tomcat as server. Users of the system are Computer Science students. The system was tested with 250 student user logins. The system was found to give efficient solutions to the browsing need of users. It gives enhanced personalization in accordance with user preference which improves the course interest. Automatic rearrangement of tabs and loading of previous page on login enhances continuity in topic even after days.

7.3 CONTRIBUTIONS OF THE WORK

The main contributions this research work propose are as follows.

7.3.1 Improvement of navigation based on user traversal patterns

We propose a method based on weight, more weightage to in link factor and mouse clicks for user centric page ranking. Our method ranks pages in the desired way of user based on usage log data. In HITS algorithm used by google for ranking pages total number of hits on a page alone contributes to ranking. This results in older pages being ranked high. Our algorithms rank pages on session relevance and thus avoids older pages getting high rank than the desired page. We observed the direct relationship between the usage data, which is a reflection from session and server logs and the weight associated with each page. This user centric approach enhances user personalization and thus achieves faster traversal results.
7.3.2 Enhanced knowledge base to users

The system provides knowledge about their previous visit patterns to users by link hierarchy display and listing previous site visits. This additional knowledge of visits pattern makes future traversals easy. Visualized structure ensures faster reach of the desired page. Compared to BFS and weighted path methods our link hierarchy visualization method achieves desired pages faster and with lesser memory utilization. The displayed structure tells the user about the organization of his desired site’s structure. This avoids unnecessary crawling over hyper links for extraction of information. This also reduces search time. Knowledge about his recent activity is displayed on left side of the page to enhance personalization. If the user needs a page from the list he can easily be directed to the page from the current location.

7.3.3 An integrated platform of personalization and search

We propose an efficient model for personalization and search. The key inspiration is Jianhan Zhu et al. (2004) method of conceptual link hierarchies for navigation and search. We enhance personalization by automatically arranging user preference as the first item. In the proposed study user on login will get a personalized page with his priority as the first item which is a novel approach. Much restructuring is not done to achieve this. The key aim of the study was to develop a personalized structure for web page to aid user in getting desired page with minimal effort. The proposed EUN implementation, VL Schools also provides previous visited page on login itself which ensures the page listing with no mouse clicks. Other desired pages are obtained in lesser mouse clicks in the proposed system. The system is a better solution to faster and efficient traversal results.
7.4 SUGGESTIONS FOR FUTURE WORK

In this section we list directions towards future enhancement of the system.

The proposed methodology for user centric page ranking giving more weightage to inlinks, gives easily distinguishable values of ranking. The method looks at many factors mainly click through data, in link factor and weight assignment based on user preferences from usage log data for ranking pages. The methodology tried to display the link structure for enhancing traversal. As a future work, the authors suggestion is to extend the proposed study to find influence of this User Centric Page Ranking on other websites. The study also wish to design a less complex algorithm for personalization.

Most of the web sites have highly complicated link structure and this makes extraction usage traversal patterns difficult. The proposed method predicts future behaviour of user in a site based on similarity generated sequences. Though comparison with different tree structures is done, sequence generation is not the core of this work. Frequent sequence is found in this study only to predict future visit pattern of user. This is to enhance personalization for user.

Thorough study on sequence generation and clustering is a future scope of the work. Clustering similar usage groups promotes the possibility of group personalization and more studies in this area can be done as future enhancement. Further study could be extended on recommendations for sites with large clutter. Another future scope is to extend the work to other fields such as Market watch and Bioinformatics.