ABSTRACT

The richness of web content has also made it progressively more difficult to leverage the value of information. Web being the large repository of data people often rely on web content for information gathering and knowledge enhancement. Typically the information present in web pages lack predetermined structure. With large number of hyperlinks between pages users find it difficult to extract required data efficiently.

Users of websites are often in need of fetching information by spending minimum effort and time. Mining knowledge about the usage of a website can be used effectively for user personalization and that can facilitate search information very fast and efficiently. Identifying users’ topic of interest, recommending content to a user based on past behaviour without much restructuring of the site is a major challenge.

Users aim at fast and easy traversal through websites. Browsing needs of individuals may not be the same. Even the same user may require different pages at different times, or may have some regular browsing patterns. Mostly applications follow page/link ranking based on prestige (high standing achieved through success or influence) that do not ensure personalization. The proposed system is a solution to user personalization which is demanded by most of the web users.

The proposed study ensures effective personalization and fast navigation to users by mined data from server logs. The study contributes to enhanced personalization of site by three key aspects. First approach is that of
classification of data for better ranking of pages. Secondly the link structure visualization of a website is aimed which can facilitate faster traversal and thus better personalization. Link hierarchy of navigation can be visualized through link structure. As a third approach, prediction based on past visits is proposed. This is achieved by segmenting database logs to clusters based on sequence of traversal patterns.

First is to select the top categories mostly visited by user and then make classification accordingly. Classified data enables ranking of pages. In traditional ranking algorithms occurrence of record relevancy is ranked rather than user centric record ranking. In the proposed Easy User Navigation (EUN) algorithm the author recommend a methodology based on Bayes approach for classification of user preferences. This is a weight centric approach. When representing pages in session as associated weights, for weight calculation in link and out link factors are considered.

The importance of taking recent repeated visits of user for better setting of user priorities is studied. If the total login count of a user is taken as the likelihood only log in and not visiting pages or crawler possibilities may result in not getting exact personalization. In order to avoid this visited category count is taken. The proposed methodology had set a threshold value of the count of hits of web pages. Values above the threshold are ranked in order for usage classification.

Through the visualized structures users can easily sketch relationships between the pages they have already visited. We studied the relevance of ranking pages based on weight centric approach. The importance of weights on inlinks and outlinks of a node are considered for ordering pages for personalizing website. We observed the fact that in a user centric perspective
forward traversal has to be given more weightage than out links. In links are assigned higher constant value which gives better user navigation supporting results than the related work. The system visualized user based link structure which automatically change by learning from web usage data. This also helps much in personalization.

User visits of a particular category is predicted by segmentation of the log data first and then by clustering. Server logs are segmented in different period of time to study the change of user behaviour over time. For better personalization in addition to prefix, a user based scan is performed in the novel USP (user span pattern) algorithm. The system had first sorted the data and then a threshold based approach is adopted. Future possibility of visit for a particular user is based on this threshold. The factors such as navigation pattern, usage styles and changes in navigation over time are taken in to account for prediction.

An integration and experimental analysis of the proposed methodologies is done in our virtual learning platform. VL Schools, the virtual learning system, has a total of 1846 pages with 2409 links. The system was developed using Java and Apache Tomcat as server. 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.

Analysed the time reduction with link view and without link view in the system and found that our proposed link view method is efficient. It is observed that rate of success of a search is found to be very high and failure
rate is negligibly less. On analysis of the most probably to be visited page for the next attempt also, success rate is found to be very high. Comparative study with similar systems prove the proposed system more effective in achieving enhanced user personalization and faster traversal results.