FINAL REPORT ON THE MINOR RESEARCH PROJECT

"An Efficient Info-Gain Algorithm For Finding Frequent Sequential Traversal Patterns From Web Logs "

Principal Investigator- Sheetal Sahu (Department of Computer Science Gandhi P.R. College, Bhopal)

Web Usage Mining is a popular research area comprises: Internet and Data Mining. By examining the possible rules invisible in web logs. It helps to improve web design and personalize the web page content, customer satisfaction and user navigation through caching and pre-fetching. Finding frequent sequential traversal patterns from web log is a very important application of web usage mining. Various pattern mining approach has been developed that mines the set of frequent subsequences pattern satisfying a min. support constraint on a session database. On the other hand, previous sequential pattern mining methods manage sequential traversal patterns uniformly whereas the pages in sequential patterns have different significance and weights.

This research presents a novel approach Frequent Sequential Traversal Pattern Mining with Self Organizing Map for mining patterns. The major confines of the traditional approach for mining patterns are that weight of every page is updated manually, but by proposing method it is updated automatically using web services. And another limitation is full database is scanned for finding the next item. The proposed method clustered the items, so that only clustered items scan, not the whole database, and using min-max weight and support of every page, so that every page having different importance. Because of this it is significant enough to carry out enormously calculation costly operations in a relatively short amount of time for finding next page prediction.