



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

A Survey on Effective User Navigation through Website Using Optimal Links

Piyusha P. Dhalpe, Deeksha Bhardwaj

ME Student, Dept. of Computer Engineering G.H. Raisoni College of Engineering & Technology, Savitribai Phule
university, Pune, India

Assistant Professor, Dept. of Computer Engineering G.H. Raisoni College of Engineering & Technology, Savitribai
Phule university, Pune, India

ABSTRACT Website navigation has been seemed as one of the most important design features across many domains including finance, e-commerce, entertainment, education, government, and medical. But to facilitate effective website navigation for user by designing a well structured website has long been a challenge. A main reason is web developers' understanding of website structure is considerably different from that of the users. There are so many methods have been discovered to relink the webpages in order to improve the navigability website structure. The completely reorganized new structure can be highly unpredictable, and the cost of disorienting users after the changes remains unanalyzed. This literature survey work describes various approaches that has been used to improve a website without introducing substantial changes. In this survey, we have described a mathematical programming model to improve the user navigation on a website while minimizing alterations to its current structure. Also we have described PageGather approach, Web sites reorganization based on user access patterns and ant colony system with their pros and cons. Also we have described Automatic Personalization of website Based on Web Usage Mining

KEYWORDS: Website Design, User Navigation, Web Mining, Mathematical Programming, PageGather.

I. INTRODUCTION

A basic reason of poor website design is that the web developers' understanding of how a website should be structured can be considerably different from those of the users. Due to the such cases users cannot easily locate the desired information in a website.

This problem is quite difficult to avoid because when web developers creating a website, he may not have a clear understanding of users' preferences and web developer can only organize pages based on his own judgments. However, satisfaction of the user is the main measure of the website effectiveness rather than that of the developers. Thus, webpages should be organized in a such way that generally it matches the user's model of how pages should be organized.

There are several differences between web transformation and personalization approaches.:

1. Web Transformation approaches generally, creates or modify the structure of a website used for all the users. While, personalization approaches dynamically reconstitute pages for individual users. Hence, there is no predefined or built-in web structure for personalization approaches.

2. In order to understand the preference of individual users, personalization approaches need to collect information associated with these users. This computationally difficult and time-consuming process is not required for transformation approaches.

3. Transformation approaches basically using a aggregate data from weblog files and it does not require to track the past usage for each user while dynamic pages are typically generated based on the users' traversal path. Thus, personalization approaches are more suitable for dynamic websites whose contents are more volatile and

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

transformation approaches are more appropriate for websites that have a built-in structure and store relatively static and stable contents.

In this work ,author mainly focuses on transformations approaches that develop methods to completely reorganize the link structure of a website.

Drawbacks:

1. A complete reorganization of website could radically change the location of familiar items, the new website may disorient users.
- 2.The reorganized website structure is highly unpredictable, and the cost of disorienting users after the changes remains unanalyzed. This is because a website structure is typically designed by experts and bears business or organizational logic, but this logic may no longer exist in the new structure when the website is completely reorganized.
3. Website reorganization approaches could dramatically change the current structure, they cannot be frequently performed to improve the navigability.

In order to overcome this drawbacks of website reorganization approaches, author develop a mathematical programming (MP) model that allow user navigation on a website with minimal changes to its current structure. Mathematical programming (MP) particularly used for informational websites whose contents are static and relatively stable over time. However, this model, may not be appropriate for websites that purely use dynamic pages or have volatile contents.

II. SYSTEM IMPLEMENTATION

A.Mathematical Programming (MP):

A Mathematical programming[1] model is used to improve the navigation efficiency of a website while reducing changes to its original website structure. Here in this work , in order to analyze the interaction between users and a website, the log files must be broken up into different user sessions. A session, it is nothing but a group of activities that are performed by a user during his visit to a site. A session may include one or more target pages, as a user may visit several targets during a single session. The metric used for analysis is the number of paths traversed to find one target, author used a different term mini session to refer to a group of pages visited by a user for only one target. Thus, a session may contain one or more mini sessions, each of which comprises a set of paths traversed to reach the target.

In this approach , author used the page-stay timeout heuristic. We can easily identify whether a page is the target page by evaluating if the time spent on that page is greater than a timeout threshold. The basic idea here is that a user generally spends more time reading on the documents that they find relevant than those they do not . Though it is impossible to identify user sessions from weblog files.

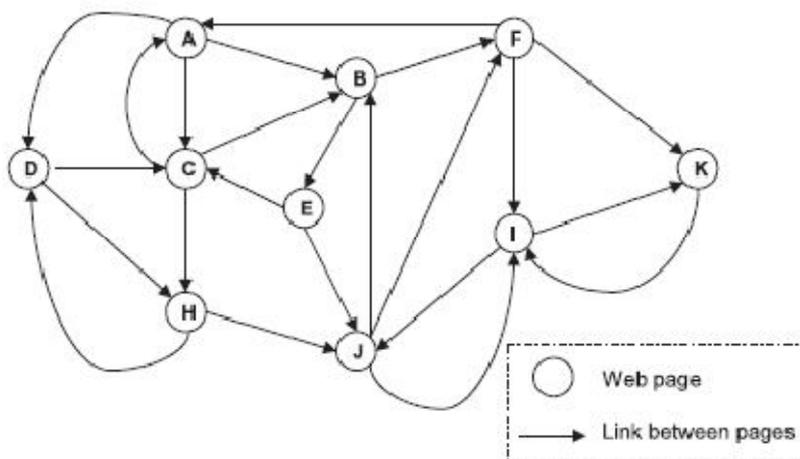


Fig.2.1. General Structure of website pages

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

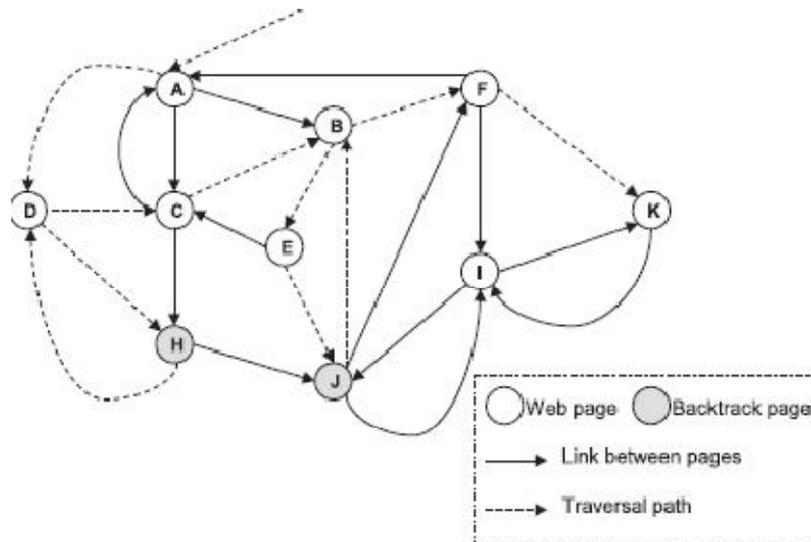


Figure 2.2. Mini Session

Figure 2.1 describes a structure of website that has 10 pages. Figure 2. 2 describes a mini session, where a user starts from A, browses D and H, and backtracks to D, from where he visits C, B, E, J, and backtracks to B. Then, this user goes from B to F and finally reaches the target K. We formally denote the mini session by $S = \{ \{A, D, H\} \{C, B, E, J\} \{F, K\} \}$ where an element in S represents a path traversed by the user. Here, mini session S has three paths as the user backtracks at H and J before reaching the target K.

In Figure 2.2 as we can see, user has been traversed three paths before reaching the target. In order reach to the target faster, we need to introduce more links and there are many so ways to add extra links. Suppose if a link is added from D to K, then user can directly reach to the target webpage K via D, and hence user can reach to the target in the first path itself. Thus, by adding this link saves the user two paths. Similarly, establishing a link from B to K allows the user to reach the target in the second path. Hence, this saves him one path.

Also, we can easily add a link from E to K, and this is considered the same as linking B to K.

While many links can be added to improve navigability, our objective here is to achieve the specified goal for user navigation with minimal changes to a website. We measure the changes by the number of new links added to the current site structure

Advantages

1. Mathematical programming improves the user navigation on a website with minimal changes to its current structure.
2. Mathematical programming (MP) model which not only successfully, performs its task but it also generates the optimal solutions fast.
3. Here, out degree can be considered as a cost term in the objective function instead of as hard constraints. This allows a page to have more links than the out-degree threshold if the cost is reasonable and hence offers a good balance between minimizing alterations to a website and reducing information overload to users.

III. LITERATURE SURVEY

A. The PageGather Algorithm:

For a large access log, here task is to find collections of pages that tend to co-occur in visits. With the help of clustering we can find out the collections of pages that are related to each other. In clustering, documents are



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

represented in an N-dimensional space . Generally , a cluster is a collection of documents close to each other and relatively distant from other clusters. Standard clustering algorithms partition the documents into a set of mutually exclusive clusters.

Cluster mining is a variation on traditional clustering that is well suited to our task. Instead of attempting to partition the entire space of documents, here author try to find a small number of high quality clusters. Traditional clustering is concerned with placing each document in exactly one cluster, cluster mining may place a single document in multiple overlapping clusters.

The PageGather algorithm[2] uses cluster mining to find collections of related pages at a web site. Generally, It takes a web server access log as input and maps it into a form ready for clustering it then applies cluster mining to the data and produces candidate index-page contents as output. The PageGather algorithm has five basic steps:

1. Process the access log into visits.
2. Compute the co-occurrence frequencies between pages and create a similarity matrix.
3. Create the graph corresponding to the matrix, and find maximal connected components in the graph.
4. Rank the clusters found, and choose which to output.
5. For each cluster, create a web page consisting of links to the documents in the cluster, and present it to the webmaster for evaluation.

1.Process the access log into visits.

A visit can be defined as an ordered sequence of pages accessed by a single user in a single session. An access log, is a sequence of page views, or requests made to the web server. Each request generally includes the time of the request, the URL requested, and the machine from which the request originated.

2.Compute the co-occurrence frequencies between pages and create a similarity matrix.

For each pair of pages p_1 and p_2 , we compute $P(p_1/p_2)$, the probability of a visitor Visiting p_1 if user has already visited p_2 and $P(p_2/p_1)$, the probability of a visitor visiting p_2 if user has already visited p_1 . The co-occurrence frequency between p_1 and p_2 is the minimum of these values.

Here, goal is to find clusters of related but currently unlinked pages. Therefore, it is necessary to avoid finding clusters of pages that are already linked together. Author prevent this by setting the matrix cell for two pages to zero if they are already linked in the site.

The similarity matrix could be viewed as a graph, which allow us to apply graph algorithms to the task of identifying collections of related pages. However, a graph corresponding to the similarity matrix would be completely connected. In order to reduce noise, we need to apply a threshold and remove edges corresponding to low co-occurrence frequency.

3.Create the graph corresponding to the matrix, and find maximum connected components in the graph.

Here, in this step author have created a graph in which each page is a node and each nonzero cell in the matrix is an arc. Next, he apply graph algorithms that efficiently extract connectivity information from the graph . By creating a sparse graph, and using efficient graph algorithms for cluster mining, it is possible to identify high quality clusters substantially faster than by relying on traditional clustering methods .

PG_{CLIQUE}

It is used to find all maximal cliques i.e connected componenet in the graph. A clique is a subgraph in which every pair of nodes has an edge between them a maximal clique is not a subset of any larger clique.

PG_{CC}

It finds all connected components subgraphs in which every pair of pages has a path



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

of edges between them. When the threshold is high, the graph will be sparse, and we can find few clusters, which will tend to be of small size and but with high quality. When the threshold is lower, we can find more, larger clusters.

4. Rank the clusters found, and choose which to output.

In Step (3) we may get many clusters, but we may wish to output only a few cluster. For example, the site's webmaster might want to see no more than a handful of clusters every week and decide which to turn into new index pages. Here, all clusters found are rated and sorted by averaging the co-occurrence frequency between all pairs of documents in the cluster.

5. For each cluster found, create a web page consisting of links to the documents in the cluster, and present it to the webmaster for evaluation.

The PageGather algorithm finds all the candidate link sets and presents them to the webmaster. The webmaster is prompted to accept or reject the cluster, to name it, and remove any links that are inappropriate. Links are labeled with the titles of the target pages and ordered alphabetically by those titles. The webmaster is responsible for placing the new page at the site.

Advantages And Limitation Of PageGather:

1. In PageGather PG_{CLIQUE} 's clusters perform much better than the existing human-authored index pages.
2. Here, cluster mining approach finds genuine regularities in the access logs that carry over from training data to test data. PageGather is geared toward finding these regularities and is apparently successful. a high proportion of links when there are so many on the page.
3. Although creating index pages with high click-through rates is desirable, the webmaster of a site has other considerations in mind.

Limitation:

PageGather approach can produce useful index pages, it does not address issues of purity or completeness.

B. Reorganizing Web Sites Based on User Access Patterns:

Here [3], author describe an approach to reorganize webpages so as to provide users with their desired information in just only few clicks. However, this approach considers only local structures in a website rather than the site as a whole, so the new structure may not be necessarily optimal.

This proposed approach consists of three steps:

1. Preprocessing
2. Page Classification,
3. Site Reorganization.

a. Preprocessing

In preprocessing, pages on a web site are processed in order to get an internal representation of the site. There are three tasks in preprocessing. 1. Web site preprocessing to obtain the current structure of a Web site, i.e., how the pages are linked. 2. Server log preprocessing to organize access records into sessions. 3. Collect access information for the pages from the sessions.

1. Web Site Preprocessing

The main purpose of web site preprocessing is to create an internal data structure to represent the Web site. Generally, the Web site is represented as a directed graph in which each page consist of a node and a link is an arc. Each page of the Web site is parsed sequentially and the links in the page are extracted. Here, each page has a unique page identifier (PID). For each page, PIDs of pages which has a link to it (parent) and pages which it links to (children) are stored.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

2. Server Log Preprocessing

As so much irrelevant information is recorded in the server log, it need to be processed first. A number of preprocessing algorithms and heuristics can be used for better Web mining. Author has described steps that are involved in preprocessing of the server log are as follows.

1. Records about image files (.gif, .jpg, etc) are filtered as well as unsuccessful requests
2. Requests from the same IP address are grouped into a session. A timeout of 30 minutes is used to decide the end of a session, i.e., if the same IP address does not occur within a time range of 30 minutes, the current session is closed.
3. The time spent on a particular page is determined by the time difference between two consecutive requests.

3. Access Information Collection

In this step, author is going to collect the access statistics for the pages that are collected from the sessions. The data obtained will later be used to classify the pages as well as to reorganize the site.

The sessions obtained in server log preprocessing are scanned and the access statistics are computed. The statistics are stored with the graph that represents the site obtained in Web site preprocessing.

Here, author have computed following statistics for each page.

- Number of sessions in which the page was accessed;
- Total time spent on the page;
- Number of times the page is the last requested page of a session.

b. Page Classification

In this page classification phase, the pages on the Web site are classified into two categories: index pages and content pages .

An index page is a page that is used by the user for navigation of the Web site. It normally contains little information except links.

A content page is a that contains the information that user would be interested in. Its content offers something other than links. The classification provides clues for site reorganization.

c. Site Reorganization

In site reorganization [3], the Web site is examined to find the better ways to organize and arrange the pages on the site. Web site can be reorganized based on the access information. The goal of this phase is to reorganize the Web site such that its users will spend less time searching for the information they want. More specifically, we want to reorganize the pages so that users can access the information they desire with fewer clicks. Although other factors such as page layout affects navigation, the number of clicks a user has to go through is the dominant factor for navigation since every click requires active rather passive effort from users and often involves a request to and an reply from the server. Here, the general idea of reorganization is to cut down the number of intermediate index pages a user has to go through. To achieve this, we need to place the frequently accessed pages higher up in the Web site structure, i.e., pages that are closer to home page, while pages that are accessed infrequently should be placed lower in the structure.

d. Advantages And Limitation:

1. This approach is helpful to build Web sites that provide its users the information they want with less clicks. By analyzing the usage of a Web site and the structure of the Web site, modifications to the Web site structure are found to improve the structure of the Web site.
2. This proposed approach has been implemented and tested on a real data set from an actual Web site. The results demonstrate a high accuracy in page classification and it decrease the number of clicks the user must perform to get interested information. So this particular approach is very promising for adaptive Web sites.

C. Ant Colony System And Local Search Based Hybrid Approach For Web Site Optimization.

The proposed methodology [4] is divided into two stages.

First, the ant colony based algorithm is applied to find the initial improved website link structure and then the resultant website structure is used as initial solution for the local search algorithm in order to improve the solution.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

a. Generation Of Solution Subgraph

In the first stage, the Ant Colony based model is applied to find the initial optimal link structure. In this method, the m numbers of ants are used one by one to generate the initial solution. This stage is further divided into two steps which is described in following section:

Step 1: Spanning Tree Generation

Here, the Ant Colony System (ACS) approach is used to generate a spanning tree satisfying the outdegree and required level constraints. Ant colony algorithm chooses edges repeatedly in the construction of the spanning tree. The ant chooses a single edge every time to develop the tree. The ants select edges from a set of candidate edges. Candidate edge set consists of edges having starting node in the constructing tree and an ending node that does not belong to the constructing tree.

Step 2: Solution Subgraph

After generating a spanning tree, the edge with the largest frequency is consecutively selected from remaining edges to add to the constructing tree to form a subgraph that satisfies the specified level and degree constraints. Every ant generates a solution subgraph that satisfies the required constraints and the best link structure out of m solutions that are generated by the m ants is passed to local search stage.

b. Local Search Application

The website structure obtained from the ant colony method act as the initial solution in the local search method. The local search is applied on the structure repeatedly to find if the objective can be improved. The solution formation of the local search employs different neighborhood than ACS which increases the probability of getting an improved website structure.

1. Initialization

In website structure improvement problem, the decision variables have value of either 0 or 1 and candidate solutions are represented by a matrix having 0 or 1 as elements. As described in the previous section the webgraph structure generated by the ACS is taken as initial solution.

2. Neighborhood structures

There are two operations which are used to alter the website structure:

1. Link Deletion
2. Link Insertion.

These operations are used to generate the neighborhood structures to analyze if some structure improves the objective function.

Firstly, a candidate edge list is generated which is having only edges which exist in the original link structure but are excluded from the initial structure that are generated by the ant colony system. Every link in the candidate edge list has the source and destination nodes in the constructing tree. An edge from the candidate list is taken and inserted in the website structure.

This insertion of an edge is also accompanied by an edge deletion operation so that the outward degree constraint is not violated. The edge to be deleted should be the edge of the lowest weight originating from a source webpage. Suppose a link with source node S and destination node D is picked from the candidate list for insertion, then out of all the outgoing edges from node S , the edge with the lowest weight should be removed to maintain the specified outdegree.

In this approach, a link insertion move should be made only if the weight of the edge to be inserted is greater than the edge to be removed. In the link deletion and insertion operations, the connectedness of the structure should remain intact. The level constraint of the nodes should also be maintained during the neighborhood generation. Suppose the deletion of link from A to B makes the level of node B more than the maximum allowed level then this deletion operation must be cancelled.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

3. Memory structures

The proposed local search approach uses a memory structure named insert tabu list to boost the search efficiency of the algorithm. In this work author used tabu list to respond to link delete operations. The insert tabu list contains the links which cannot be inserted in the future moves. When a link is removed from the website structure then that link is added to the insert tabu list so that the link is not added in future steps.

4. Advantages And Limitations:

1. Experiments with artificial graphs reveal that the proposed ant colony system can outperform the 0-1 programming approach by providing very good solutions in a short computation time.
2. The ant colony system exhibits low time complexity, and can therefore be applied to reorganizing large websites.
3. The ant colony system can be successfully applied to real-world websites.

D. Automatic Personalization Based on Web Usage Mining

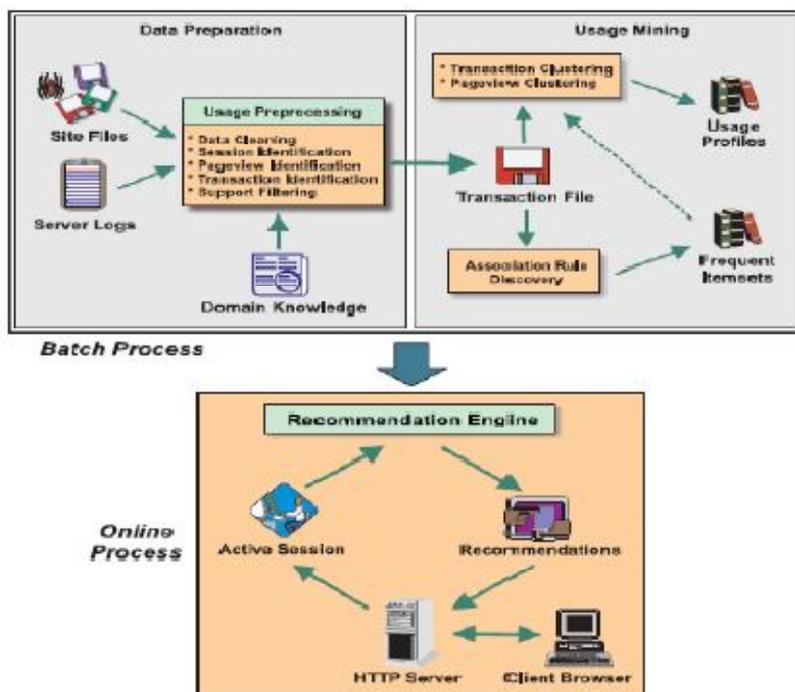


Figure 3.3: System Architecture of Automatic Personalization with Web Usage Mining.

The above figure 3.3 describe the system architecture in which offline component of usage-based Web personalization divided into two separate stages.

1. Preprocessing and Data Preparation- it includes data cleaning, filtering, and transaction identification.

2. Mining – In this stage usage patterns are discovered via methods such as association-rule mining and clustering.

1. Preprocessing Tasks

Here, the first important task is identification of a set of user sessions from the raw usage data which is provided by the Web server. Ideally, each user session gives an exact accounting of who accessed the Web site, what pages were requested and in what order, and also how long each page was viewed by the particular user. Here, local caching and proxy servers are the biggest impediments to forming accurate user sessions. In order to improve performance and minimize network traffic, most Web browsers cache the pages that have been requested by the user during the session.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

As a result, when a user hits the "back" button, the cached page is displayed and the Web server is not aware of the repeat page access.

In addition to identifying user sessions, the raw log must also be cleaned, or transformed into a list of page views. Due to the stateless connection properties of the HTTP protocol, several file requests (HTML, images, sounds, etc.) are often made as the result of a single user action. The group of files that are sent due to a single click are referred to as a page view.

Cleaning [5] the server log involves removing all of the file accesses that are redundant, leaving only one entry per page view. This includes handling page views that have multiple frames, and dynamic pages that have the same template name for multiple page views. It may also be necessary to filter the log files by mapping the references to the site topology induced by physical links between pages.

Each user session in a user session file can be thought of in two ways; either as a single transaction of many page references, or a set of many transactions each consisting of a single page reference. The goal of transaction identification is to dynamically create meaningful clusters of references for each user.

In this work, author have assumed that each user session is viewed as a single transaction. Finally, the session file may be filtered to remove very small transactions and very low support references. This type of support filtering can be important in removing noise from the data, and can provide a form of dimensionality reduction in clustering tasks where URLs appearing in the session file are used as features.

2. Discovering Frequent Itemsets and Association Rules

The association rule discovery methods such as the Apriori algorithm, initially find groups of items occurring frequently together in many transactions. Such groups of items are referred to as frequent item sets. Given a set $I = \{I_1, I_2, \dots, I_k\}$ of frequent itemsets, the support of I_i is defined [5] as follows

$$\sigma(I_i) = \frac{|\{t \in T: I_i \subseteq t\}|}{|T|}$$

The itemsets returned by the algorithm satisfy this minimum support threshold.

Also support is considered as downward closed that is if an item set does not satisfy the minimum support criteria, then neither do any of its supersets.

Association rules capture the relationships among items based on their patterns of co-occurrence across transactions. In the case of Web transactions, association rules capture relationships among URL references based on the navigational patterns of users. An association rule r is an expression of the following form [5].

$$X \Rightarrow Y \quad (\sigma_r, \alpha_r)$$

Clustering:

Transaction clustering [5] will result in a set $C = \{c_1, c_2, \dots, c_k\}$ of clusters, where each c_i is a subset of user transaction say T . Here, each cluster represents a group of users with "similar" access patterns. However, transaction clusters by themselves are not an effective means of capturing an aggregated view of common user profiles. Each transaction cluster may potentially contain thousands of user transactions involving hundreds of URL references.. The URL clusters associated with each transaction cluster will serve as an aggregated view and a representative of users whose behavior is captured in the transaction clusters.

IV. CONCLUSION

In this paper, author have proposed a mathematical programming model to improve the navigation effectiveness of a website while minimizing changes to its current structure. This model is appropriate for informational websites whose contents are relatively stable over time. It improves a website rather than reorganizes it and hence it is suitable for website maintenance on a progressive basis. Mathematical programming model can provide significant improvements to user navigation by adding only few new links. Optimal solutions were quickly obtained, suggesting



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

that the model is very effective to real world websites. The MP model was observed to scale up very well, optimally solving large-sized problems in a few seconds in most cases on a desktop PC.

REFERENCES.

- [1] Min Chen and Young U. Ryu, "Facilitating Effective User Navigation through Website Structure Improvement", IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 3, March 2013
- [2] M. Perkowski and O. Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study", Artificial Intelligence, vol. 118, pp. 245-275, 2000.
- [3] Y. Fu, M.Y. Shih, M. Creado, and C. Ju, "Reorganizing Web Sites Based on User Access Patterns," Intelligent Systems in Accounting, Finance and Management, vol. 11, no. 1, pp. 39-53, 2002.
- [4] Harpreet Singh, Parminder Kaur, "Website Structure Optimization Model Based on Ant Colony System and Local Search", I.J. Information Technology and Computer Science, 2014, 11, 48-53.
- [5] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining", Comm. ACM, vol. 43, no. 8, pp. 142-151, 2000.
- [6] Y. Fu, M.Y. Shih, M. Creado, and C. Ju, "Reorganizing Web Sites Based on User Access Patterns," Intelligent Systems in Accounting, Finance and Management, vol. 11, no. 1, pp. 39-53, 2002.
- [7] M.D. Marsico and S. Levialdi, "Evaluating Web Sites: Exploiting User's Expectation", Int'l J. Human-Computer Studies, vol. 60, no. 3, pp. 381-416, 2004.
- [8] R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages", INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.
- [9] C.C. Lin, "Optimal Web Site Reorganization Considering Information Overload and Search Depth", European J. Operational Research, vol. 173, no. 3, pp. 839-848, 2006.
- [10] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization", ACM Trans. Internet Technology, vol. 3, no. 1, pp. 1-27, 2003.
- [11] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization", Data Mining and Knowledge Discovery, vol. 6, no. 1, pp. 61-82, 2002.
- [12] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining", Comm. ACM, vol. 43, no. 8, pp. 142-151, 2000.
- [13] B. Mobasher, R. Cooley, and J. Srivastava, "Creating Adaptive Web Sites through Usage-Based Clustering of URLs", Proc. Workshop Knowledge and Data Eng. Exchange, 1999.