GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
AN APPROACH FOR IMPROVED WEBSITE STRUCTURE USING SMART K-MEANS

(Shukla Bharati ,Wakhare Sagar,Khule Sujit ,Deokate Vanita)[*1] and Prof.Deepti Varshney[2]

[*1]BE student , Department of Computer Science , Shree Ramchandra College Of Engineering , India
[2]Assistant Professor , Department of Computer Science , Shree Ramchandra College Of Engineering , India

ABSTRACT
Development of websites to facilitate effective user navigation is the challenging task observed these days. Because the way web developers think and design the system is quite different from that of the user. Different methods have been projected to re-link WebPages in order to recover navigability using user direction-finding data. The fully reorganized emerging structure can be highly impulsive, and the cost of disorienting users after the changes remains unanalyzed. The proposed system presents architecture to cluster the usage statistics of all the users to re-link WebPages. The re-ordering or reforming will mostly be based on clusters generated. Hence an optimal selection of clusters is significant step in implementation of the system. Hence system uses an enhanced K means clustering algorithm where in the number of clusters (optimal) can be routinely designed and clusters are generated consequently.

The system also develops a arithmetical programming model to recover the user navigation on a website. The system is imagined the deliver the functionality of a test bench website for data collection and then reorder it based on statistics collected to present the effectiveness of our model.

Keywords:*clustering,Website design,user navigation,web mining*.

I. INTRODUCTION
Clustering is a data mining technique which helps in grouping or making clusters of data having similar values of Some of the data attributes. Clustering can be used in various Fields like in Health sector for grouping patients with similar Symptoms of the disease, in banking sector to group customers who have dues in their credit card payments, in Market analysis to identify the customers having similar buying patterns. Currently, researchers are exploring the application of this technique in the field of education to better understand students' academic performance and the academic framework in which they learn. Nowadays there has been increasing investments in website design but it is still exposed, however, that finding necessary material in a website is relatively problematic. Designing effective websites is cumbersome task. Palmer indicated that poor website design has been a key element in a number of high profile site letdowns. McKinney et al. also discover that users having difficulty in pinpointing the targets are probably to leave a website even if its information is of good quality. Earlier studies on website has concentrated on a diversity of issues, such as understanding web structures, locating related pages of a given page, mining useful structure of a news website, and removing template from web pages. This work is related to the literature that observes how to recover website navigability through the use of user navigation data. Different works have made an effort to address this question and they can be usually categorized into two types: to help a particular user by animatedly reconstructing pages based on his contour and traversal paths, often denoted as personalization, and to adapt the site structure to simplify the navigation for all users, often stated as transformation. A principal cause of poor website design is that the web developers understanding of how a website should be organized can be considerably diverse from those of the users. Such variances result in cases where users cannot certainly trace the preferred information in a website. This problem is hard to escape because when forming a website, web developers don't have a perfect understanding of users likings and can only form pages based on their own verdicts. However, the degree of website effectiveness should be the approval of the users rather than that of the developers. Thus, Webpages should be structured in a way that generally matches the users model of how pages should be organized. This paper proposed a modified K-means algorithm which classifies the input data set into appropriate clusters without taking number of clusters K as input, as it was required in the case of K-means. The proposed algorithm does not require the number of clusters K as input distinguishing the shortcomings of website reorganization tactics, proposed system addresses the question of how to recover the organization of a website rather than reorganize it

1

II. METHOD & MATERIAL
ALGORITHM USED:

Smart K-mean:

1) Compute sum of the attribute values of each tuple (to find the points in the data set which are farthest   apart)
2) Take tuples with minimum and maximum values of the sum as initial centroids.
3) Create initial partitions (clusters) using Euclidean distance between every tuple and the initial centroids.
4)Find distance of every tuple from the centroid in both theinitial partitions. Take d=minimum of all   distances.
(other than zero)
5)Compute new means (centroids) for the partitions created
in step 3.
6)Compute Euclidean distance of every tuple from the new means (cluster centers) and fmd the outliers depending
on the following objective function:
7)If Distance of the tuple from the cluster mean<d then not an Outlier.
8)Compute new centroids of the clusters.
9)Calculate Euclidean distance of every outlier from the new cluster centroids and fmd the outliers not satisfying the
objective function in step 6.
10)Let B={Yl,Y2,. ....Y p) be the set of outliers obtained in step 8 (value of k depends on number of outliers).
Repeat unti I (B==<D)
a) Create a new cluster for the set B, by taking mean value of its members as centroid.
b) Find the outliers of this cluster, depending on the objective function in step 6.
c) If no.of outliers = p then
i) Create a new cluster with one of the outliers as its member and test every other outlier for the objective
function as in step 6.
ii) Find the outliers if any
d) Calculate the distance of every outlier from the centroid of the existing clusters and adjust the outliers in the
existing which satisfy the objective function in step 6.
e) B={ ZI,Z2 .... Zq}be the new set of outliers.(value of q depends on number of outliers)

III. OTHER SECTIONS

1.MOTIVATIONS

        Data Security is the science and study of methods of protecting data from unautho-rized disclosure and
modification As per the technology upgraded, there is need to secure data which is transmitted over the network.
Unsecured networks can be hacked into easily, and hackers can do lots of things in short amounts of time. A hacker
can search the hard drive of the average PC user in less than a minute. In this short time period a search can be
conducted on spread sheets or databases that contain user names and passwords.

2.PROBLEM STATEMENT

        The network security is the main problem now a days.To provide a data security there isneed of some extra
security primitives. We are tackling this problem by using Armstrong numbers and color method.

3.PROPOSED SYSTEM

     When User want Surfing on Internet that time user did not get actual information he want He/she has to spend
a   lot of time on that particular web site .In this Paper we propose a new Data mining algorithm k means i.e.
improved k-means algorithm . This Improved K-means Work on at database of web server .this algorithm take input
as session log with preferences And then transform these input into the number of clusters. The cluster is depends on
the no of input so the total no link and the relinking of that all pervious links of particular website. With the help of

2

relinking and linking we find the priority of that particular link and these link come on the very front page of web site .This way we can reduce time complexity over web.
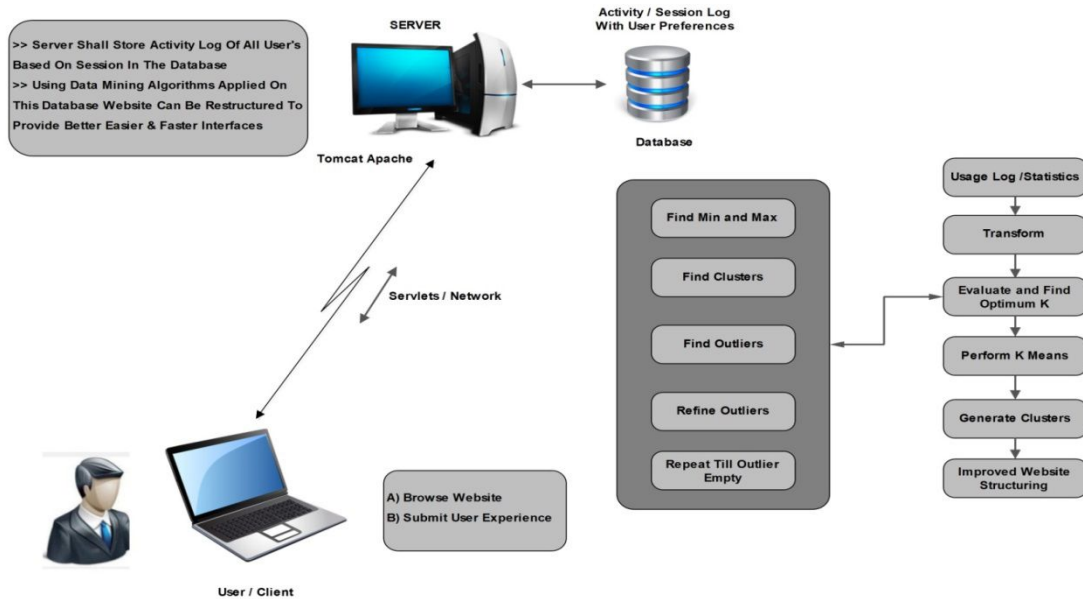

Fig. Architecture of propose system

The diagram shown above shows the proposed system. The system consists of end users/client, a server for storing the data.
• **USER**: The end user is the actual user who searches for the relevant data on the web using browser installed in his system.
• **Web Server**: It tracks the search request made by the user. Also server maintains the activity session log with the user preferences and stores it in the database. Using Improved k means clustering algorithm it improves the navigation of the website so that it provides better, easier and fast interface.

IV. RESULT & DISCUSSION
The proposed system is useful where user wants to send his data over the internet.
The system can also useful where security is concern. The proposed system can also useful in the research center to protect the confidential data on that center.
Personal information and official information is secure with the use of this system. The widest application is in Military operations and sensitive data Tran's reception. There are wide range and shades of various colors and color combinations. It becomes very difficult to guess the user unique color also the underlying computation of matrix is so efficient that guessing the keys is near about impossible.

V. CONCLUSION
The Proposed system is efficient system for navigating website structure. Using the data mining algorithm i.e. Improved K means algorithm it helps to restructure the link of website. It provides the easier, fastest and better interface to retrieve information from the website. The major benefit of this system is that it performs the process of retrieving information in minimum span of time so that we can say it is time efficient

VI. ACKNOWLEDGEMENTS

REFERENCE

[1] T.Nakayama,KatoAndY. Yamane,Discovering the Gap between WebSite Designers Expectations and Users Behavior, Computer Networks,vol.33,pp.811-822,2000.

[2] M.Perkowitz and O.Etzioni,Towards Adaptive WebSites:Conceptual Framework and CaseStudy,Artificial Intelligence,vol.118,pp.245- 275,2000.

[3] Y.Fu,M.Y.Shih,M.Credo, andC.Ju, "Reorganizing WebSites Based on UserAccess Patterns,"IntelligentSystems in Accounting,Finance andManagement, vol.11,no.1, pp.39-53,2002.

[4] J.HouandY.Zhang,Effectively Finding Relevant Web Pages from Linkage Information, IEEE Trans. Knowledge and Data Eng.,vol.15,no.4,pp.940-951,July/Aug.2003.

[5] M.Eirinaki and M.Vazirgiannis, Web Mining for Web Personalization,ACM Trans.Internet Technology,vol.3, no.1, pp.1-27,2003.

[6] Shehroz, Ahmad, "Cluster center initiation algorithm for K-means clustering", Pattern Recognition Letters 25, pages 1293-1302,2004.

[7] Mohammed EI Agha, Wesam M. Ashour, " Efficient and Fast Initializtion Algorithm for K-means Clustering" Intelligent Systems and Applications, vol. 4, issue 1,pages 21-31, 2012.

[8] Facilitating Effective User Navigation through Website Structure Improvement, MinChenand Young U.Ryu,Knowledge an dData Engineering, Vol.25,No.3, March2013

[9] 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

[10] Anupama Chadha, Suresh Kumar" An Improved K-Means Clustering Algorithm: A Step Forward for Removal of Dependency on K" 2014 International Conference on Reliability, Optimization and Information Technology - ICROIT 2014, India, Feb 6-8 2014.