# An Approach for Improved Website Structure using **Smart K-Means**

Shukla Bharati<sup>1</sup>, Wakhare Sagar<sup>2</sup>, Khule Sujit<sup>3</sup>, Deokate Vanita<sup>4</sup>, Prof. Deepti Varshney<sup>5</sup> <sup>1, 2, 3, 4, 5</sup>Department of Computer Science

<sup>1, 2, 3, 4, 5</sup>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, Websitedesign, Usernavigation, 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.

#### **II. 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.

#### **III. MOTIVATIONS**

Data Security is the science and study of methods of protecting data from unauthorized disclosure and modi\_cationAs per the technology upgraded, there is need to secure data which is transmitted over the network. Unsecured networks can be hacked intoeasily, and hackers can do lots of things in short amounts of time. A hacker can searchthe hard drive of the average PC user in less than a minute. In this short time perioda search can be conducted on spread sheets or databases that contain user names and Passwords

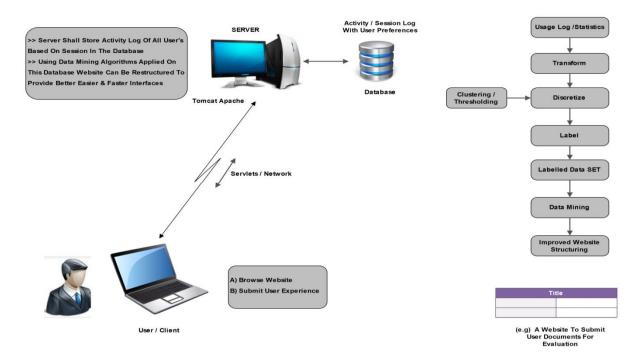
#### **IV. OBJECTIVES**

To provide the Improved Web Site Structure Using K-means Algorithem.

#### **Existing System**

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, WebPagesshould be structured in a way that generally matches the user's model of how pages should be organized

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 substantially. Specifically, we develop a mathematical programming (MP) model that simplifies user navigation on a website with slight changes to its present structure. Our model is mostly suitable for informational websites whose matters are static and quite stable over time. Examples of informational websites are universities, hospitals, tourist attractions, federal agencies, and sports organizations. Our model, however, may not be appropriate for websites that purely use dynamic pages or have volatile contents. This is because a steady state might never be reached in user access patterns in such websites, so it may not be possible to use the weblog data to improve the site structure. The relevancy of web page can be attained by considering the amount of inlinks and out-links existing in a particular web page. When the web page has more number of out-links to a pertinent page, then that page can be treated as a central page. From this central page, all remaining web pages are compared for similarity and the most similar pages are grouped together. The grouping of most similar pages together is known as clustering. Clustering can be done based on different algorithms such as hierarchical, k-means, partitioning, etc. The simplest unsupervised learning algorithm that solve clustering problem is K- Means algorithm. It is a simple and easy way to classify a given data set through a certain number of clusters.



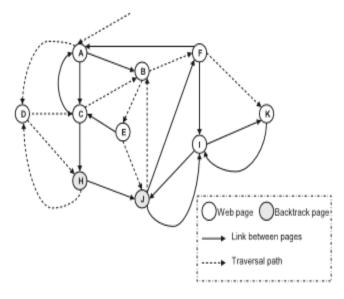


Fig3 . Example of a mini session.

Our model allows Webmasters to specify a goal for user navigation that the improved structure should meet. This goal is associated with individual target pages and is defined as the maximum number of paths allowed to reach the target page in a mini session. We term this goal the path threshold for short in this paper. In other words, in order to achieve the user navigation goal, the website structure must be altered in a way such that the number of paths needed to locate the targets in the improved structure is not larger than the path threshold.

In the example shown in Fig. 2, the user has traversed three paths before reaching the target. An intuitive solution to help this user reach the target faster is to introduce more links . There are many ways to add extra links. If a link is added from D to K, the user can directly reach K via D, and hence reach the target in the first path. Thus, adding this link "saves" the user two paths. Similarly, establishing a link from B to K enables the user to reach the target in the second path. Hence, this saves him one path. We could also insert a link from E to K, and this is considered the same as linking B to K. This is because both B and E are pages visited in the second path, so linking either one to K saves only one path. Yet, another possibility is to link C to F, a nontarget page. In this case, we assume that the user does not follow the new link, because it does not directly connect a page to the target. While many links can be added to improve navigability, our objective 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. There are several reasons that we should insert minimal links. First, minimizing changes to the current structure can avoid disorienting familiar users. Second, adding unnecessary links can lead to pages having too many links, which increases users' cognitive loads and makes it difficult for them to read and comprehend . Third, since our model improves site structures on a regular basis, the number of new links should be kept at minimum such that the links in a website in the whole course of maintenance do not expand in a chaotic manner.

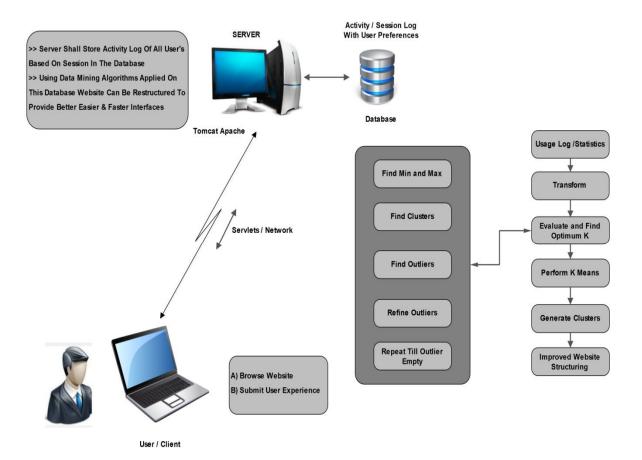
#### V. PROPOSED SYSTEM

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 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.

#### **VI. SYSTEM ARCHITECTURE**

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.



# 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 theobjective 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)</li>
  - 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)

#### VII. RESULT COMPARISON OF EXISTING AND PROPOSED SYSTEM

#### EXISTING SYSTEM

#### Normal link on website:

# WELCOME TO EFFECTIVE USER NAVIGATION WELCOME TO EFFECTIVE USER NAVIGATION Joint Colspan="2">Joint Colspan="2"/Joint Colspan="2"/J

#### Normal link on website1:

# WELCOME TO EFFECTIVE USER NAVIGATION

## WELCOME TO EFFECTIVE USER NAVIGATION

#### **0.Albert Einstein**

#### Early life and education



Albert Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879 His parents were Hermann Einstein, a salesman and engineer, and Pauline Koch. In 1880, the family moved to Munich, where his father and his uncle founded Elektrotechnische Fabrik J. Einstein & Cie, a company that manufactured electrical equipment based on direct current 2.Isaac Newtor

#### Academic career

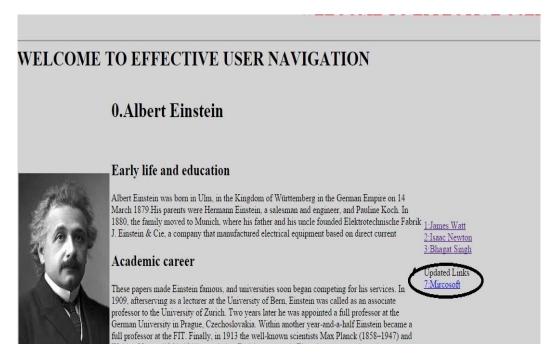
These papers made Einstein famous, and universities soon began competing for his services. In 1909, afterserving as a lecturer at the University of Bern, Einstein was called as an associate professor to the University of Zurich. Two years later he was appointed a full professor at the German University in Prague, Czechosłovakia, Within another year-and-a-half Einstein became a full professor at the FIT. Finally, in 1913 the well-known scientists Max Planck (1858–1947) and With a View and Called and Ca



Updated Links <u>7:Mircosoft</u>

#### **PROPOSED SYSTEM**

#### Final updated link:



#### First page screenshot:

# WELCOME TO EFFECTIVE USER NAVIGATION WELCOME TO EFFECTIVE USER NAVIGATION O.Albert Einstein D.Albert Einstein Alber Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879 His parents were Hermann Einstein, a salesman and engineer, and Pauline Koch In Silo. the family moved to Munch, where his father and has uncle founded Elektrotechansche Fabrit 1: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt 2: Einstein & Cie, a company that manufactured electrical equipment based on direct current 1: Linnes Watt

Academic career These papers made Einstein famous, and universities soon began competing for his services. In 1909, afterserving as a lecturer at the University of Bern. Einstein was called as an associate professor to the University of Zurich. Two years later he was appointed a full professor at the German University in Prague, Czechoslovakia. Within another year-and-a-half Einstein became a full professor at the FIT. Finally, in 1913 the well-known scientists Max Planck (1858–1947) and

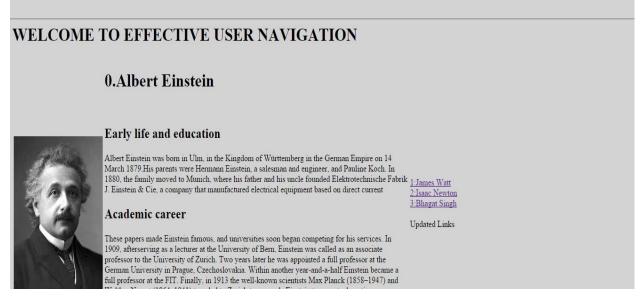
Page | 480

## VIII. RESULT ANALYSIS

#### Screenshot 1:

Here we can see the normal web page with some link and there one link i.e. Updated link But it is initially no link present.

# WELCOME TO EFFECTIVE USER NAVIGATION



#### Screenshot 2:

Here in this window we can see number of hits. That mean when visit any link on website like we dick

On link that is the hit for that link so that count of that hit store on this module.

					ADM	IN	MC	DU	LE						
	Sessin ID	10	Page		Updated	Links	centroid Ca	Rotation							
1 2	80E 049CF 15415364844725 80E 049CF 15415366844725	7	1	÷	-				Calc	ulate	e Hit	8			
3	IBE080F15415369844725. IBE080F15415369844725.	1	1			Page 2	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Pagel
5	BDE006CF15415366844725.		1		Page 0		0	4	0		23	19			26
6	C28A8D4D8202F2A8ECECA.		. 1		Page 1	10	0	15	.0	7	22	18		10	24
7	C28ABD6D8202F2A6ECECA.	1	1		Page 2	10	11		28	0	0	18			23
4	C28ABOKD8200F2A8ECECA.	3	1		Page 3	.11	0		6		23	19	11	10	0
2	C2BABD4D6262F2A6ECECA.				Page 4	4	7		12		0	11	3	3	12
10	E703ETA97E50F8CDC447E		1		Page 5			13	25		0	0			24 26
11	E7D3E7A97E50F8CDC447E	3	3		Page 6		0		0		0	0			26
12	E703E7A07E50F8C0C447E		3		Page 7	5	0	12	0	y .	0	13		. 0	12
13	E7D3E7A97E50F9CDC447E		1		Page 2	. 6	10	12	0	.7	18	13	0	. 0	0
14	E703ETA97E50F8C0C447E				Page 9		12	. 14	.0	.0	. 9	19	11	15	
35	E703ETA87E50FBCDC447E		1.1		1-1203/1										
11	OB20AFFACTOFAB3796F774 OB20AFFACTOFAB3796F774					_	_	_	_	_	_	_	_	_	_
18	OBDOAFFACTDFAB3796F774	-		1.4						pdate Li	in her				
75	OB20AFFACTOFAB3796F774			F	1.00	_	_		0	possible in	and a		_		_
20	0820AFFACTDFAE3T96F774	1													
21	OBZOAFFACTOFAB3796F774														
22	OR20AFFACTDFAE3796F774	4													
22	9820AFFAC7DFA83796F774_		1.0												
24	0820AFF/sC/TDFA83796F774														
100	SETTINE AND IN NETTINE ACON														

# Screenshot 3:

This is the first window from the admin side only admin can see it .if admin use this with help of user name and password for security purpose.

ects Admn UserNavigation	6 <sup>3</sup> 1	4	😻 NetBeanside	
		scover	My NetBeans	What's New
		LOGIN		es and technologies by installing plugins from the
	USERNAME PASSWORD	user ****		javr
	Message Log	n successfully!	LOGIN	

#### Screenshot 4:

In previous screen we see the no of hit store in that module, here we used this no of hit as input to

The k-means algorithm. With help of that algorithm we find centroid of that data set.

##	Sessin ID	To	Page	Updated	LINKS	Centroid Ca	Iculation							_
1	8BE086DF15415366B4A725	1	2											
2	8BE086DF15415366B4A725	7	2					Calc	ulate	Hit H	S			
3	8BE086DF15415366B4A725	3	1								1.000			
4	8BE086DF15415366B4A725	8	1	##	Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page
5	8BE086DF15415366B4A725	9	1	Page 0	0	0	0	0	8	23	19	9	9	26
3	C2BABD6D8202F2A6ECECA	0	1	Page 1	10	0	15	0	7	22	18	0	10	24
7	C2BABD6D8202F2A6ECECA	1	1	Page 2	10	11	0	25	0	0	18	9	9	23
3	C2BABD6D8202F2A6ECECA	3	1	Page 3	11	0	0	0	8	23	19	11	10	0
9	C2BABD6D8202F2A6ECECA	2	1	Page 4	4	7	0	12	0	0	11	3	3	12
10	E7D3E7A97E50FBCDC447E	0	1	Page 5	8	9	13	25	8	0	0	9	9	24
11	E7D3E7A97E50FBCDC447E	3	3	Page 6	0	0	0	0	8	0	0	9	9	26
12	E7D3E7A97E50FBCDC447E	9	3	Page 7	5	0	12	0	7	0	13	0	0	12
13	E7D3E7A97E50FBCDC447E	4	1	Page 8	5	10	12	0	7	18	13	0	0	0
14	E7D3E7A97E50FBCDC447E	5	1	Page 9	0	12	14	0	0	0	19	11	10	0
15	E7D3E7A97E50FBCDC447E	6	1											
16	0B20AFFAC7DFA83796F774	0	2											
17	0B20AFFAC7DFA83796F774	3	3		_	_	_				_	_	_	_
18	0B20AFFAC7DFA83796F774	9	2					U	pdate Li	nks				
19	0B20AFFAC7DFA83796F774	5	3											
20	0B20AFFAC7DFA83796F774	6	3	Centroid:	0 - * 0									
21	0B20AFFAC7DFA83796F774	2	2	 and subserve										
22	0B20AFFAC7DFA83796F774	1	1	Centroid:	14 - * 14									
23	0B20AFFAC7DFA83796F774	7	6											
24	0B20AFFAC7DFA83796F774	8	6	Centroid:	0 - * 0									
25	98776EAADE3F1E2DE4C2C	0	3											
26	98776EAADE3F1E2DE4C2C	3	9	Centroid:	0 - * 0									
27	98776EAADE3F1E2DE4C2C	9	9											
28	98776EAADE3F1E2DE4C2C	5	15	Centroid:	19 - * 19									
29	98776EAADE3F1E2DE4C2C	6	10											
30	98776EAADE3F1E2DE4C2C	2	10											

# Screenshot 5:

With help of k-mean algorithm we find the cluster with centroid .According to cluster here reconstruct the link of web pages or automatically update the link.

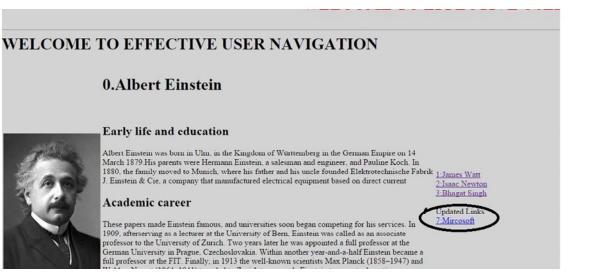
				1	
##	Sessin ID	To	Page		Updated Links Centroid Calculation
1	8BE086DF15415366B4A725	1	2	A	
2	8BE086DF15415366B4A725	7	2		Final updated Links:
3	8BE086DF15415366B4A725	3	1		Page 0 - (7,7:Mircosoft),
4	8BE086DF15415366B4A725	8	1		Page 1 - (7,7:Mircosoft),
5	8BE086DF15415366B4A725	9	1		Page 2 - (7,7:Mircosoft),
6	C2BABD6D8202F2A6ECECA	0	1		Page 3 - (7,7:Mircosoft),
7	C2BABD6D8202F2A6ECECA	1	1		Page 4 - (4,4:Sukdev),
8	C2BABD6D8202F2A6ECECA	3	1		Page 5 - (1,1:James Watt),
9	C2BABD6D8202F2A6ECECA	2	1		Page 6 - (1,1:James Watt),
10	E7D3E7A97E50FBCDC447E	0	1		Page 7 - (7,7:Mircosoft)
11	E7D3E7A97E50FBCDC447E	3	3		Page 8 - (6,6:Apple) ,
12	E7D3E7A97E50FBCDC447E	9	3		Page 9 - (7,7:Mircosoft),
13	E7D3E7A97E50FBCDC447E	4	1		
14	E7D3E7A97E50FBCDC447E	5	1		Message
15	E7D3E7A97E50FBCDC447E	6	1		
16	0B20AFFAC7DFA83796F774	0	2		
17	0B20AFFAC7DFA83796F774	3	3		Updated Sucessfully!
18	0B20AFFAC7DFA83796F774	9	2		Opulated Sucessicily
19	0B20AFFAC7DFA83796F774	5	3		
20	0B20AFFAC7DFA83796F774	6	3		
21	0B20AFFAC7DFA83796F774	2	2		OK
22	0B20AFFAC7DFA83796F774	1	1		
23	0B20AFFAC7DFA83796F774		6		
24	0B20AFFAC7DFA83796F774		6		
25	98776EAADE3F1E2DE4C2C		3		
26	98776EAADE3F1E2DE4C2C		9		

In previous screen we see the no of hit store in that module, here we used this no of hit as input to The k-means algorithm. With help of that algorithm we find centroid of that data set.

##	Sessin ID	To	Page		Updated	LINKS	Centroid Ca	iculation [							
1	8BE086DF15415366B4A725	1	2	A	1			_	100			-			
2	8BE086DF15415366B4A725	7	2						Calc	ulate	Hit	S			
3	8BE086DF15415366B4A725	3	1		and the second s										
4	8BE086DF15415366B4A725	8	1		##	Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page
5	8BE086DF15415366B4A725	9	1		Page 0	0	0	0	0	8	23	19	9	9	26
6	C2BABD6D8202F2A6ECECA	0	1		Page 1	10	0	15	0	7	22	18	0	10	24
7	C2BABD6D8202F2A6ECECA	1	1		Page 2	10	11	0	25	0	0	18	9	9	23
8	C2BABD6D8202F2A6ECECA	3	1		Page 3	11	0	0	0	8	23	19	11	10	0
9	C2BABD6D8202F2A6ECECA	2	1		Page 4	4	7	0	12	0	0	11	3	3	12
10	E7D3E7A97E50FBCDC447E	0	1		Page 5	8	9	13	25	8	ő	0	9	9	24
11	E7D3E7A97E50FBCDC447E	3	3		Page 6	0	0	0	0	8	Ő	0	9	9	26
12	E7D3E7A97E50FBCDC447E	9	3		Page 7	5	0	12	0	7	0	13	0	0	12
13	E7D3E7A97E50FBCDC447E	4	1		Page 8	5	10	12	0	7	18	13	0	0	0
14	E7D3E7A97E50FBCDC447E	5	1		Page 9	0	12	14	0	0	0	19	11	10	0
15	E7D3E7A97E50FBCDC447E	6	1		( of a										
16	0B20AFFAC7DFA83796F774	0	2												
17	0B20AFFAC7DFA83796F774	3	3		1				12.00						
18	0B20AFFAC7DFA83796F774	9	2						U	pdate Li	inks				
19	0B20AFFAC7DFA83796F774	5	3		and the second second	1000									
20	0B20AFFAC7DFA83796F774	6	3		Centroid:	0-*0									
21	0B20AFFAC7DFA83796F774	2	2												
22	0B20AFFAC7DFA83796F774	1	1		Centroid:	14 -* 14									
23	0B20AFFAC7DFA83796F774	7	6												
24	0B20AFFAC7DFA83796F774	8	6		Centroid:	0 - * 0									
25	98776EAADE3F1E2DE4C2C	0	3												
26	98776EAADE3F1E2DE4C2C	3	9		Centroid:	0 - * 0									
27	98776EAADE3F1E2DE4C2C	9	9												
28	98776EAADE3F1E2DE4C2C	5	15		Centroid:	19 - * 19									
29	98776EAADE3F1E2DE4C2C	6	10												
30	98776EAADE3E1E2DE4C2C	2	10	7											

After the updating link these updated link which is present on first page of web site. This is automatically

Construction of link is done with help of k-means algorithm.



##	Sessin ID	То	Page	Updated Links Centroid Calculation
1	8BE086DF15415366B4A725	1	2	
2	8BE086DF15415366B4A725	7	2	Final updated Links:
3	8BE086DF15415366B4A725	3	1	Page 0 - (7,7:Mircosoft),
4	8BE086DF15415366B4A725	8	1	Page 1 - (7,7:Mircosoft),
5	8BE086DF15415366B4A725	9	1	Page 2 - (7,7:Mircosoft),
6	C2BABD6D8202F2A6ECECA	0	1	Page 3 - (7,7:Mircosoft),
7	C2BABD6D8202F2A6ECECA	1	1	Page 4 - (4,4:Sukdev),
8	C2BABD6D8202F2A6ECECA	3	1	Page 5 - (1,1:James Watt),
9	C2BABD6D8202F2A6ECECA	2	1	Page 6 - (1,1:James Watt) ,
10	E7D3E7A97E50FBCDC447E	0	1	Page 7 - (7,7:Mircosoft) ,
11	E7D3E7A97E50FBCDC447E	3	3	Page 8 - (6,6:Apple) ,
12	E7D3E7A97E50FBCDC447E	9	3	Page 9 - (7,7:Mircosoft),
13	E7D3E7A97E50FBCDC447E	4	1	
14	E7D3E7A97E50FBCDC447E	5	1	Message
15	E7D3E7A97E50FBCDC447E	6	1	
16	0B20AFFAC7DFA83796F774	0	2	
17	0B20AFFAC7DFA83796F774	3	3	Updated Sucessfully!
18	0B20AFFAC7DFA83796F774	9	2	Opulated Successfullys!
19	0B20AFFAC7DFA83796F774	5	3	
20	0B20AFFAC7DFA83796F774	6	3	
21	0B20AFFAC7DFA83796F774	2	2	ОК
22	0B20AFFAC7DFA83796F774	1	1	
23	0B20AFFAC7DFA83796F774	7	6	
24	0B20AFFAC7DFA83796F774		6	
25	98776EAADE3F1E2DE4C2C	0	3	
26	98776EAADE3F1E2DE4C2C	3	9	

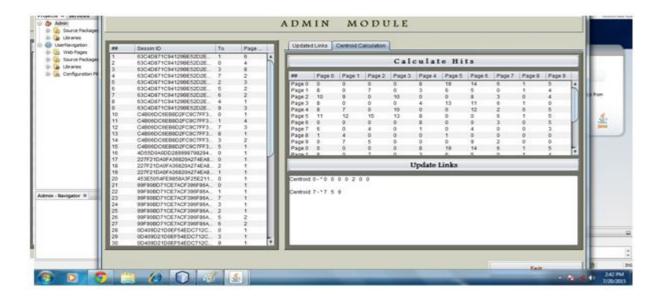
# Steps to run software:

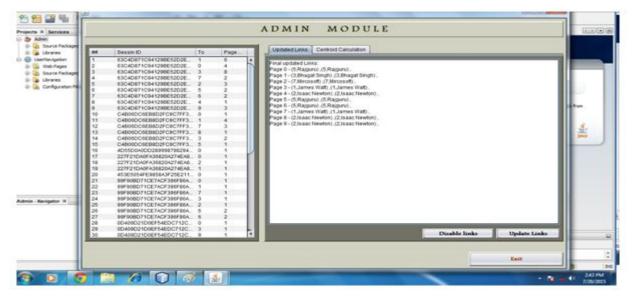
# 1. Login procedure of admin module :

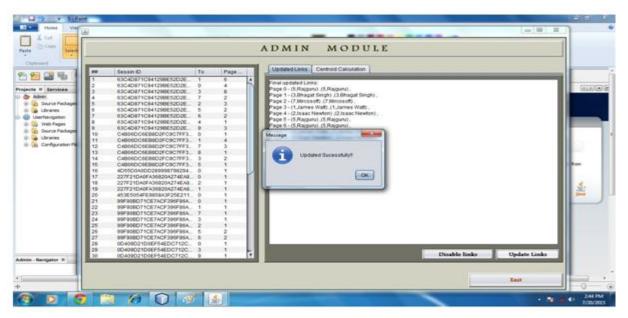
ojects Admin UserNavigation		5 *	S NetBeanside	
	( <u>s</u>	cover	Mv NetBeans	What's New
		LO	GIN	es and technologies by installing plugins from the
	USERNAME PASSWORD	user ****		java
	Message	Login successfully!	LOGIN	

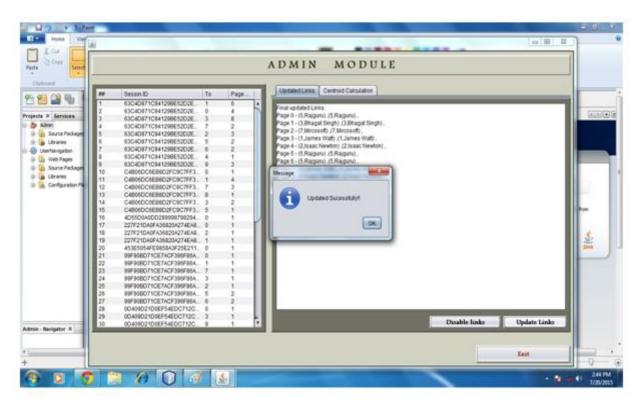
	Sessin (D	To	Page .	10.0	UpdatedL	inks Centrol	d Calculation	6		_	_				
1	63C4D871C941298E52D2E	1	6	4	-		_	-	<i>c</i> .	lculate					
2	63C4D871C941298E52D2E	0	4						C a	rcurate	nits				
2	63C4D871C841298E52D2E	3	8			TAWARA		Test State		10000000		1.000			/
4	63C4D871C941298E52D2E	7	2			Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 6	Page 8
5	63C4D671C941298E52D2E	2	3		Page 0	0		.0	0	7	15	12	. 6	1	. 5
6	63C4D671C941298E52D2E	5	2		Page 1		0	7	0	3	6	5	0	1	4
7.	63C4D871C941298E52D2E	- 6	2		Page 2			0	10	0	0	6	- 3	0	4
8	63C4D871C941298E52D2E	4	1		Page 3		0	0	0	4	13	11		1	0
B	63C4D871C941298E52D2E		. 3		Page 4	7	7.	0	10		0	10	2	0.	5
10	C4806DC6EB802FC9C7FF3.		1		Page 5	9	12	11	13	7	0	0	6	1	5
11	C4806DC6E88D2FC9C7FF3.	1	. 4		Page 6	0		0	0	7	0	0	3	0	5
12	C4B06DC6EB8C2FC9C7FF3	7	3		Page 7	6	0		0	4	0	4	0	0	3
13	C4806DC6EB802FC9C7FF3.		- 1		Page 8	1	4	0	0	0	1	0	0	0	0
14	C4B06DC6EB802FC9C7FF3	3	2		Page 9	0	7	5	0	0	0		2	0	0
15	C48060C6EB802FC9C7FF3.				10000							22			
16	4D55D0A0DD209990708294	0	1.8		- Contract										
17	227F21DA0FA36820A274EA8	0	1		_	_	_	_					_	_	
18	227F21DA0FA36820A274EA8	- 2	1							Update Li	inks				
19	227F21DA0FA36820A274EA8	1	1	14	line and the second sec						-				
20	453E5054FE9858A3F25E211.	0		1.1											
21	99F908D71CE7ACF396F86A	0	1												
22	\$9F908D71CETACF396F86A	. 1													
23	99F90BD71CE7ACF396F86A	7	1												
24	99F908D71CE7ACF396F86A	3													
25	99F90BD71CETACF396F86A		1												
26	99F90BD71CETACF396F86A		2												
27	99F90ED71CE7ACF39EF86A		2												
28	0D409D21D0EF54EDC712C	0													
29	00409021D0EF14EDC712C	3	1												
30	0D409021D0EF54EDC712C			12											

۴.	Sessin ID	To	Page .	UpdaledL	inks Cerdiol	d Calculation								
	\$3C4D871C941298E52D2E	3	0 4					C a	lculate	Hite				
	63C4D871C941298E52D2E 63C4D871C941298E52D2E	2	: IN						rearant					_
	63C4D871C941298E5202E	2			Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page 9
	63C4D871C941298E52D2E	6	1		1.986.0	Page 1	Freger		r age a			Fage 1	r age e	r sge
	63C4D871C941298E52D2E	2	2	Page 0	0		0	0	7	15	12		1	
	63C4D871C941298E52D2E	6	-	Page 1 Page 2		-	1	0	3	0		0	1	
	63C4D671C941298E52D2E			Page 2 Page 3			0	10		13			0	
	\$3C4D871C941298E52D2E		1.1	Page 4			0	10		12	11	-	1	0
	C4B06DC6EB802FC9C7FF3	0	1	Page 5		12	11	10		0	10	-		
	C4806DC6EB8D2FC9C7FF3	4	10.00	Page 5		12		0						2
	C4806DC6EB802FC9C7FF3	7	1	Page 5 Page 7			0	0		0	0	3		
	C4B05DC8EB802FC9C7FF3		1	Page 8				0	-					
	C4B06DC6EB802FC9C7FF3	3	2	Page 9				0						
	C4806DC6EB8D2FC9C7FF3	5	1	1.694.1	0		2					-	<u>.</u>	
	4D55D0A0DD289998798294		1	1.000										
	227F21DA0FA36826A274EA8	0	1											
	227F21DA0FA36820A274EA8	2	. 1						Update Li	inks				
	227F21DA0FA36820A274EA8	1	1		_	_	_	_			_	_	_	_
	453E5054FE9858A3F25E211	0	1											
	99F908D71CE7ACF396F86A.	0	1											
	99F908D71CE7ACF396F86A	1	1											
	99F908D71CE7ACF396F86A.	7	1											
	99F908D71CE7ACF396F86A	3	1											
	10F908D71CE7ACF386F86A.	2	1											
	99F908D71CE7ACF386F86A	5	2											
	99F90BD75CE7ACF396F86A.	.6	2											
	6D409D21D0EF54EDC712C	0	1											
	8D409D21D0EF54EDC712C	3	1											
	0D409D21D0EF54EDC712C	0	1 1											









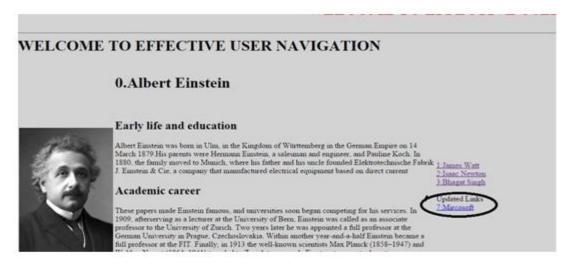
#### 2. Calculate no. of hit:

				A D M	IN	MC	DU	LE						
##	Sessin ID	To	Page	Updated	1 Links	Centroid Ca	lculation							
1	8BE086DF15415366B4A725	1	2					a 1	1.4					
2	8BE086DF15415366B4A725	7	2					Calc	ulate	Hit	S			
3	8BE086DF15415366B4A725	3	1		30		101		AV.	-	112	W.	- N.S.	-
4	8BE086DF15415366B4A725	8	1	##	Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page
5	8BE086DF15415366B4A725	9	1	Page 0	0	0	0	0	8	23	19	9	9	26
6	C2BABD6D8202F2A6ECECA	0	1	Page 1	10	0	15	0	7	22	18	0	10	24
7	C2BABD6D8202F2A6ECECA	1	1	Page 2	10	11	0	25	0	0	18	9	9	23
8	C2BABD6D8202F2A6ECECA	3	1	Page 3	11	0	0	0	8	23	19	11	10	0
9	C2BABD6D8202F2A6ECECA	2	1	Page 4	4	7	0	12	0	0	11	3	3	12
10	E7D3E7A97E50FBCDC447E	0	1	Page 5	8	9	13	25	8	0	0	9	9	24
11	E7D3E7A97E50FBCDC447E	3	3	Page 6	0	0	0	0	8	0	0	9	9	26
12	E7D3E7A97E50FBCDC447E	9	3	Page 7	5	0	12	0	7	0	13	0	0	12
13	E7D3E7A97E50FBCDC447E	4	1	Page 8	5	10	12	0	7	18	13	0	0	0
14	E7D3E7A97E50FBCDC447E	5	1	Page 9	0	12	14	0	0	0	19	11	10	0
15	E7D3E7A97E50FBCDC447E	6	1	, age a			1000		-					
16	0B20AFFAC7DFA83796F774	0	2											
17	0B20AFFAC7DFA83796F774	3	3	-						727				
18	0B20AFFAC7DFA83796F774	9	2					U	odate Li	nks				
19	0B20AFFAC7DFA83796F774	5	3		_	_	_			_	_	_	_	
20	0B20AFFAC7DFA83796F774	6	3											
21	0B20AFFAC7DFA83796F774	2	2											
22	0B20AFFAC7DFA83796F774	1	1											
23	0B20AFFAC7DFA83796F774	7	6											
24	0B20AFFAC7DFA83796F774	8	6											
25	98776FAADE3E1E2DE4C2C	0	3											

#### **3.** Calculate centroid:

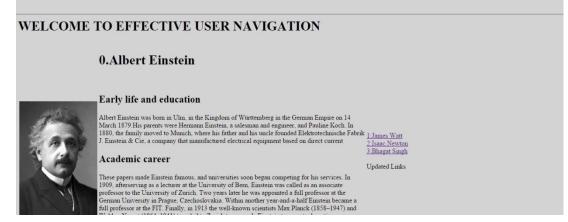
	Sessin ID	To	Page		Updated	ILINKS S	Centroid Ca	iouuson [	_						
1	8BE086DF15415366B4A725	1	2	4	-	_						_			
2	8BE080DF15415366B4A725_	3	2						Calc	ulate	Hit.	5			
3	8BE086DF15415366B4A725	3	1											_	
4	8BE086DF15415366B4A725	8	1		**	Page 0	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page 9
5	8BE086DF15415366B4A725	8	1		Page 0	0	0	0	0	8	23	19	9	9	26
6	C2BABD6D8202F2A6ECECA_	D	1		Page 1	10	0	15	0	7	22	10	0	10	24
7	C2BABD6D8202F2A6ECECA.	1	1		Page 2	10	11	0	25	0	0	10			23
	C28ABD6D8202F2A6ECECA.	3	1		Page 3	11	0	0	0		23	19	11	10	0
9	C28ABD6D8202F2A6ECECA.	2	.1		Page 4	4	7	0	12	0	0	11	3	3	12
10	E7D3E7A97E50FBCDC447E	0	1		Page 5		.0	13	25	8	0	.0	.9		24
11	E7D3E7A07E50FBCDC447E	3	3		Page 6		0	0	0	8	0	0			26
12	E7D3E7A97E50FBCDC447E	9	3		Page 7	5	0	12	0	7	0	13	0	0	12
13	E7D3E7A97E50FBCDC447E	4	1		Page 8	6	10	12	0	7	18	13	0	0	0
14	E7D3E7A97E50FBCDC447E	5	1		Page 9	8	12	14	0	0	0	19	11	10	0
15	E7D3E7A97E50FBCDC447E	6	1				-			-	-				-
16	0B20AFFAC7DFA83796F774_	0	2												
17	0B20AFFAC7DFA83796F774	3	3		-	_	_	_				_	_	_	_
18	0B20AFFAC7DFA83796F774_	9	2						U	pdate Li	inks				
19	0B20AFFAC7DFA83796F774	5	3		-						_				
20	0B20AFFAC7DFAB3796F774_	6	3		Centroid.	0-*0									
21	0820AFFAC7DFA83796F774_	2	2												-
22	0820AFFAC7DFA83796F774	1	1		Centroid.	14-* 14									
23	0820AFFAC7DFA83796F774_	7	6												
24	0820AFFAC7DFA83796F774_	8	6		Centroid.	0-*0									
25	98776EAADE3F1E2DE4C2C	0	3												
26	98776EAADE3F1E2DE4C2C	3	.9		Centroid	0-*0									
27	98776EAADE3F1E2DE4C2C	9	9												
28	98776EAADE3F1E2DE4C2C	5	15		Centroid	10-110									
29	98776EAADE3F1E2DE4C2C.	6	10												
30	98776EAADE3F1E2DE4C2C	2	10												

#### 4. Final updated link:



#### 5. First page screenshot;

#### WELCOME TO EFFECTIVE USER NAVIGATION



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.

#### X. FURTHER WORK

The limitation with proposed system if we select the number containing zero then result produced is not proper . As a future scope we will work to remove this limitations.

#### ACKNOWLEDGEMENTS

We are presenting this paper with great pride. There are too much efforts ofgardener to yield the beautiful owners. So we should not forget him while praising owners. It is a matter of gratification for we to pay our respects and acknowledgments to allthose who have imparted knowledge and helped we to complete my paper. Wewould first like to acknowledge the great contribution and support we have received inthis endeavor from my Guide Prof. DeeptiVarshney.We are very grateful and indebted to the project guide for providing her enduring patience, guidance and invaluable suggestions. she was the one who neverlet our moral down and always supported we through us thick and thin.

We would also like to thank all the Staff members for their invaluable co-operation and permitting we to work in the Computer labs of SRCOE.

#### REFERENCES

- 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, ArtificialIntelligence,vol.118,pp.245- 275,2000.
- [3] Y.Fu,M.Y.Shih,M.Credo, andC.Ju, "Reorganizing WebSites Based on UserAccess Patterns," Intelligent Systemsin Accounting, Financeand Management, vol.11,no.1, pp.39-53,2002.

- 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, ACMTrans. 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 Initialization 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 YoungU.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 Removalof Dependency on K" 2014 International Conference on Reliability, Optimization and Information Technology