International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Comparative Study of Web Structure Mining Techniques for Links and Image Search

Rashmi Sharma[1], Kamaljit Kaur[2]

[1]Student of M.Tech in computer Science and Engineering, Sri Guru Granth Sahib World University, Fatehgarh Sahib - 140406, Punjab, India

[2]Assistant Professor, Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, Fatehgarh Sahib - 140406, Punjab, India

Abstract: ***Web mining is a vast area and growing very rapidly. It employs text, audio, video, contents and images from World Wide Web. The World Wide Web contains huge number of web pages and a lot of information available within web pages. When a user put a query to the search engine, it generally returns a large amount of information in response to user's query. To retrieve relevant information from Web pages, web mining performs various web structure mining techniques. This paper proposed the hybrid technique of Weighted PageRank based on Visit of Links and Fuzzy K-Means algorithms which are applied on the search result. Fuzzy K-Means algorithm is used to group the given data into clusters and Weighted PageRank is used to re-rank the data according to the visit of links to taken in the account. We have extracted the relevant information such as image links, images and total hyperlinks from a web links using the hybrid approach. In an existing work, previous approach PageRank with K-Means has done on only text or URL not on links and has several limitations as compared to Weighted PageRank and Fuzzy K-Means algorithm. In this paper, we have search on links and images and provide quality search to users and applied valid parameters like execution time, recall, precision and f-measure on our proposed method and compared with previous approach and conclude the result. The principle idea of overall our method is to provide the fast and more relevant results in response to user requirements and can be seen as future of web mining.***

Keywords: PageRank, Weighted PageRank based on Visit of Links, K-Means and Fuzzy K-Means**.**

1. Introduction

Web mining is an application of data mining. It defined as the process of extracting useful information from World Wide Web data. Two different approaches are taken to define the web mining. (1) "Process-centric view" which defines web mining as sequence of task. (2) "Data-centric view" which defines web mining in terms of the types of web data that is being used in the mining process [4].

Web mining can be broadly divided into three distinct categories, according to the types of data to be mined, Web Content Mining, Web Structure Mining and Web Usage Mining [7]. Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page contents. Content mining is the scanning and mining of text, pictures and graphs of a Web page to determine the relevance of the content to the search query [5]. Web usage mining is the process of extracting useful information from server logs i.e. users' history. Web usage mining is the process of finding out what users are looking for on Internet. Web usage mining process involves the log time of pages. The world's largest portal like yahoo, msn etc., needs a lot of insights from the behavior of their users' web visits [5] and Web structure mining, one of three categories of web mining for data, is a tool used to identify the relationship between Web pages linked by information or direct link connection. This structure data is discoverable by the provision of web structure schema through database techniques for Web pages [5].

In this paper our approach will extend to image web links, the proposed work has done on web links search and including hyperlinks and image links using Weighted PageRank based on visit of links along with Fuzzy K-Means clustering. In this research the problem is formulated on efficient search of Web Links along with images, Weighted PageRank and Fuzzy K-means will be followed in link and image search. This approach holds simultaneously two problems of link and image search. Mostly when user requires the image some non relevant images are also got extracted which is poor knowledge discovery in data mining and inaccurate approach. The main objective of our algorithms is to provide the relevant links which the user wants. At the end we compared our approach to the previous approach PageRank and K-Means. PageRank has its own limitations as compared to Weighted PageRank such as PageRank is less efficient, slow in speed and provide less quality result. Similarly K-Means has some disadvantages as small in size, more complex, problems with outliers and empty clusters. But in contrast, Weighted PageRank based on Visit of Links with Fuzzy K-Means overcomes the problem of previous approach.

The rest of this paper is organized as follows: a brief summary of previous approach is given in section 2. Section 3 described the proposed algorithms in detail. The methodology of proposed work is illustrated in section 4. Section 5 demonstrated the results, screenshots of proposed method and experiments applied on hybrid technology and illustrate a general conclusion in section 6.

2. Existing Work

Web mining is an application of data mining that discovers and extracts the required information from web services and documents. Web structure mining techniques play main role to fetch the relevant data from web pages. In this paper we have focused on hybrid approach such as ranking algorithm with clustering technique. In an existing work, various

Licensed Under Creative Commons Attribution CC BY

ranking methods like PageRank and Weighted PageRank has been used with K-Means algorithm to fetch the contents from web pages. But they did not use on links and image search. At the end of our paper we have taken common data set and done some experiment with PageRank with K-Means and our proposed hybrid technique and conclude the result. Ranking techniques and clustering mechanisms plays a vital role in web search engine to re-rank the web pages and group the similar data into clusters.

Brin and Page [1, 3] developed PageRank algorithm at Stanford University based on the hyper link structure. PageRank algorithm is used by the famous search engine, Google. PageRank algorithm is the most frequently used algorithm for ranking billions of web pages. During the processing of a query, Google's search algorithm combines precomputed PageRank scores with text matching scores to obtain an overall ranking score for each web page. Functioning of the Page Rank algorithm depends upon link structure of the web pages. The PageRank algorithm is based on the concepts that if a page surrounds important links towards it then the links of this page near the other page are also to be believed as imperative pages. The Page Rank imitate on the back link in deciding the rank score. Thus, a page gets hold of a high rank if the addition of the ranks of its back links is high.

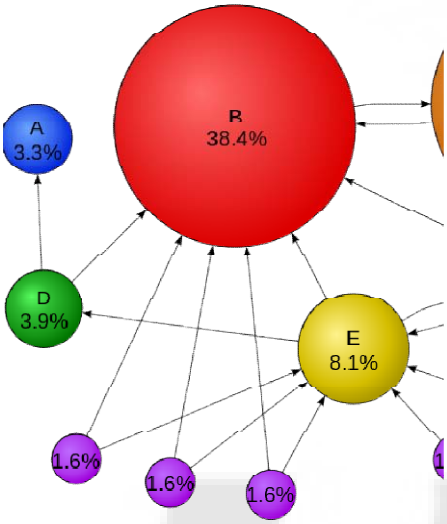

Figure 1: A Principle of PageRank Algorithm [14]

A simplified version of PageRank is given as:

$$PR(u) = (1-d) + d\sum_{v\varepsilon B(u)}\left(\frac{PR(v)}{Nv}\right)$$

Where u represents a web page, $B(u)$ is the set of pages that point to u, $PR(u)$and $PR(v)$ are rank scores of page u and v respectively, Nv indicates the number of outgoing links of page v, d is a damping factor.

Supreet Kaur et al. [9] have studied that Clustering is an essential task in Data Mining process which is used for the purpose to make groups or clusters of the given data set based on the similarity between them. This paper is intended to give the introduction about K-means clustering and its algorithm. The experimental results of K-means clustering and its performance in case of execution time are discussed here. But there are certain limitations in K-means clustering algorithm such as it takes more time for execution. So in order to reduce the execution time we are using the weighted page rank with k means clustering and also shown that how clustering is performed in less execution time as compared to the traditional method. This work makes an attempt at studying the feasibility of K-means clustering algorithm in data mining using the weighted page rank with k means clustering. K means with page rank algorithm gave results with better result set of various numbers of data-sets. In this research the work is going on k means clustering of database with weighted page content rank algorithm.

Amar Singh et al. [10] have described the work represents ranking based method that improved K-means clustering algorithm performance and accuracy. In this paper, the analysis of K-means clustering algorithm, one is the existing K-means clustering approach which is incorporated with some threshold value and second one is ranking method which is weighted page ranking applied on K-means algorithm, in weighted page rank algorithm mainly in links and out links are used and also compared the performance in terms of execution time of clustering. He has proposed ranking based K-means algorithm which produced better results than that of the existing k-means algorithm.

K-Means Clustering Algorithm [8] K-means clustering is a well known partitioning method. In this objects are classified as belonging to one of K-groups. The results of partitioning method are a set of K clusters, each object of data set belonging to one cluster. In each cluster there may be a centroid or a cluster representative. In case where it considers real-valued data, the arithmetic mean of the attribute vectors for all objects within a cluster provides an appropriate representative; alternative types of centroid may be required in other cases.

Steps of K-Means Algorithm: K-Means Clustering algorithm is an idea, in which there is need to classify the given data set into K clusters, the value of K (Number of clusters) is defined by the user which is fixed. Euclidean Distance is used for calculating the distance of data point from the particular centroid [8].

This algorithm consists of four steps:

- **Initialization:** In this first step data set, number of clusters and the centroid that we defined for each cluster.
- **Classification:** The distance is calculated for each data point from the centroid and the data point having minimum distance from the centriod of a cluster is assigned to that particular cluster.
- **Centroid Recalculation**: Clusters generated previously, the centriod is again repeatly calculated means recalculation of the centriod.
- **Convergence Condition**: Some convergence conditions are given as below:
  - Stopping when reaching a given or defined number of iterations.
  - Stopping when there is no exchange of data points between the clusters.
  - Stopping when a threshold value is achieved.

Licensed Under Creative Commons Attribution CC BY

- If all of the above conditions are not satisfied, then go to step 2 and the whole process repeat again, until the given conditions are not satisfied.

Limitations of K-means:
- K-means has problems when clusters are of differing sizes, densities and non-globular shapes
- Problems with outliers
- Empty clusters

3. Proposed Hybrid Approach

Here we proposed hybrid approach. In this hybrid technique two well known web mining algorithms is used such as Weighted PageRank based on Visit of Links and Fuzzy K-Means. This proposed method is implemented in two steps: Firstly Fuzzy K-Means is used to group the data set into clusters and secondly Weighted PageRank is implemented on cluster to rank the data in it according to the visit of links. These two algorithms are explained in detail as following:

3.1 Fuzzy K-Means

The fuzzy k-means clustering algorithm partitions data points into k clusters Sl (l = 1,2, …, k) and clusters Sl are associated with representatives (cluster center) Cl. The relationship between a data point and cluster representative is fuzzy. The major process of FKM is mapping a given set of representative vectors into an improved one through partitioning data points. It begins with a set of initial cluster centers and repeats this mapping process until a stopping criterion is satisfied. It is supposed that no two clusters have the same cluster representative. In the case that two cluster centers coincide, a cluster center should be perturbed to avoid coincidence in the iterative process. The fuzzy k-means clustering algorithm is now presented as follows[6].

Algorithm: There are following steps of algorithm give as [6]:

Pre-processing Phase
Step 1: Select informative genes by using Entropy filtering approach. The Genes with low entropy value is re-moved.
Step 2: Fuzzify feature vector values using S-shaped membership function or Z-shaped membership function.

Clustering Phase
Input: K-No of clusters.
n- Total number of Genes.
m- Number of samples.
W_{lower}, W_{upper} and threshold (ε).
Output**:** K-Gene clusters.
Step 1: Randomly assign each object into exactly one lower Approximation C_k, the objects also belongs to upper approximation C_k of the same cluster. Boundary region isC_k^B.
Step 2: Compute cluster centroids.
$i = 1, 2 \ldots, n, j = 1, 2 \ldots m \; and \; h = 1, 2, \ldots, k$
If $C_K^B = \overline{C_k} . \underline{C_k} \neq \emptyset$

$$Z_k = \left(W_{lower} \times \frac{\sum X \varepsilon \underline{C_k} X_i}{|\underline{C_k}|}\right) + \left(W_{lower} \times \frac{\sum X \varepsilon \underline{\underline{C_k}} X_i}{|\overline{C_k} - \underline{C_k}|}\right)$$

Else
Compute new centroids,

$$Z_k = \sum x \varepsilon c \; \frac{X}{|C_k|}$$

End
Step 3: Find Similarity Si, Here, i-Represents genes,

$$S_i(\hat{X}, \hat{Z}) = 1 - \frac{\sum_{j=1}^{n} |\hat{x}_{ij} - \hat{z}_{hj}|}{\sum_{j=1}^{n} (\hat{x}_{ij} + \hat{z}_{hj})}$$

Step 4: Compute $Pi = \frac{Max_{Si}}{Min_{Si}}$

and normalize the Pi values
If $\geq (\varepsilon)$, insert i^{th} object in K^{th}Cluster.
Step 5: Update centroids. Repeat the steps 2 to step 5, until New centroid = Old centroid.

3.2 Weighted PageRank based on Visit of Links

In Weighted PageRank algorithm, we assign more rank value to the outgoing links which is most visited by users and received higher popularity from number of inlinks. The user's browsing behavior can be calculated by number of hits (visits) of links. The modified version based on WPR (VOL) is given as [3].

$$WPR_{vol}(u) = (1 - d) + d \sum_{v \varepsilon B(u)} \frac{L_u WPR_{Vol}(v) W_{in}(v,u) TL(v)}{TL(v)}$$

Algorithm: The various steps of the proposed algorithm are given below [3].
Step 1: Finding a Website: Find a website which has rich hyperlinks because WPR (VOL) methods rely on the web structures.
Step 2: Building a Web Map: Then generate the web map from the selected website
Step 3: Calculate W^{in} (v,u): Then calculate the W^{in} (v,u) for each node present in web graph by applying the equation as below.

$$W^{in}(m,n) = \frac{I_n}{\sum_{p \varepsilon R(m)} I_p}$$

Where
- $W^{in}(v,u)$ is the weight of link (v,u) calculated based on the number of inlinks of page u and the number of inlinks of all reference pages of page v.
- In and Ip are the number of incoming links of page n and page p respectively.
- $R\,(m)$ denotes the reference page list of page m.

Step 4: Apply proposed formula: Now calculate the PageRank value of the nodes present in web graph by using the proposed formula

$$WPR_{vol}(u) = (1 - d) + d \sum_{v \varepsilon B(u)} \frac{L_u WPR_{Vol}(v) W_{in}(v,u) TL(v)}{TL(v)}$$

Notations are:
- d is a dampening factor ,
- u represents a web page,
- $B(u)$ is the set of pages that point to u,

Licensed Under Creative Commons Attribution CC BY

- $WPR_{vol}(u)$ and $WPR_{vol}(v)$ are rank scores of page u and v respectively,
- L_u is the number of visits of link which is pointing page u from v.
- $TL(v)$ denotes total number of visits of all links present on v.

Step 5: Repeat by going to step 4: final step will be used recursively until the values are to be stable.

4. Implementation

We have done analysis on well known URLs using both proposed approach and previous approach and at last compare the results. Methodology describes the web based image search as user extract the relevant information in the form of image links, images and hyperlinks from the URL and compute the number of visits using Weighted PageRank from the URL. At the end we compute various parameters such as Execution Time, Recall, Precision and F-Measure. The whole process of implementation as described in steps:

- Firstly, in proposed method Fuzzy K- Means is used to group the given data set into clusters whereas in previous approach K-Means is used to group the data into clusters.
- Secondly, Weighted PageRank is applied on clusters to re-rank the data according to number of Visits of Links and popularity of inlinks to taken into account and in previous approach PageRank is applied and re-rank the data according to count the backlinks to taken into account.
- Now, User refers a keyword as input and URL is come to the top from database.
- After that, user able to extract the relevant image links, images and hyperlinks from the URL.
- In proposed approach we compute Weighted PageRank based on Visit Links and in previous approach PageRank is computed.
- User can also check the user availability in other URLs that is to find the URL in other URLs.
- At the end we compute various parameters described in next section and compare the parameters of both project.

5. Results and Comparisons

We have run our proposed approach and base approach on various well known URLs such as tanishq, yahoo, WhatsApp, WeChat, BBM, homeshop18, twitter etc and extract the image links, images and hyperlinks from them. At the end we compute various parameters to conclude our results and results sets are compared using a graph and table that presents the actual research work with in comparison to previous approach. We have done analysis on validation measurements as following:

Execution Time: We define the execution time in our project as to time taken to retrieve the image links and hyperlinks from the URL. The unit of time taken is milliseconds (ms). The result set is shown in table and compared by graph.

Table 1: Comparisons of Execution Time

ID	*URLs*	*Time Taken (ms)by Proposed Approach*	*Time Taken (ms)by Previous Approach*
1.	http://www.naturewallpaper.net/	5746	28000
2.	http://www.whatsapp.com/	6636	15000
3.	http://www.wechat.com/en/	4159	16000
4.	http://all-free-download.com/free-photos/rose-flowers.html	7012	26000
5.	http://www.bbm.com/bbm/en.html	5421	24000
6.	http://www.homeshop18.com/	8069	34000
7.	http://adaptive-images.com/	4885	18000
8.	https://twitter.com/	6159	33000
9.	https://in.yahoo.com/?p=us	8717	30000
10.	http://jewellery.picturesklix.com/	8182	26000
11.	http://tanishq.co.in/Home	6740	24000

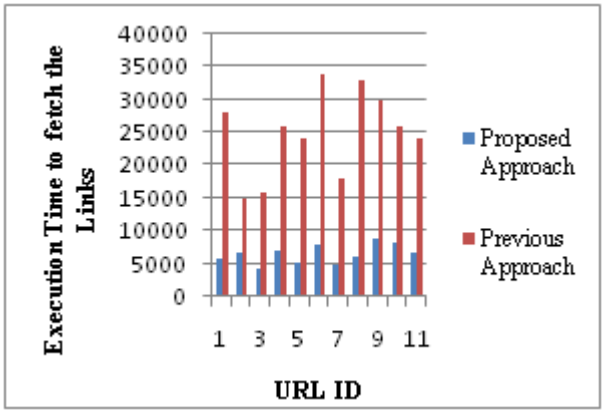

Figure 2: Graph illustrates comparison in terms of Time

Precision: The fraction of retrieved documents that is relevant to find [12]:

$$Precision = \frac{Total\ Relevant\ documents}{Total\ Retrieved\ documents}$$

Precision is compared by proposed technique and previous technique as following in the form of table and graph.

Table 2: Comparisons of Precision

ID	*URLs*	*Precision of Proposed Approach*	*Precision of Previous Approach*
1.	http://www.naturewallpaper.net/	0.4285	0.3953
2.	http://www.whatsapp.com/	0.3333	0.2631
3.	http://www.wechat.com/en/	0.1208	0.0967
4.	http://all-free-download.com/free-photos/rose-flowers.html	0.2086	0.1880
5.	http://www.bbm.com/bbm/en.html	0.6190	0.4782
6.	http://www.homeshop18.com/	0.3478	0.2916
7.	http://adaptive-images.com/	0.6071	0.5
8.	https://twitter.com/	0.09615	0.0555
9.	https://in.yahoo.com/?p=us	0.6847	0.6666
10.	http://jewellery.picturesklix.com/	0.3596	0.3362
11.	http://tanishq.co.in/Home	0.1458	0.1224

Licensed Under Creative Commons Attribution CC BY

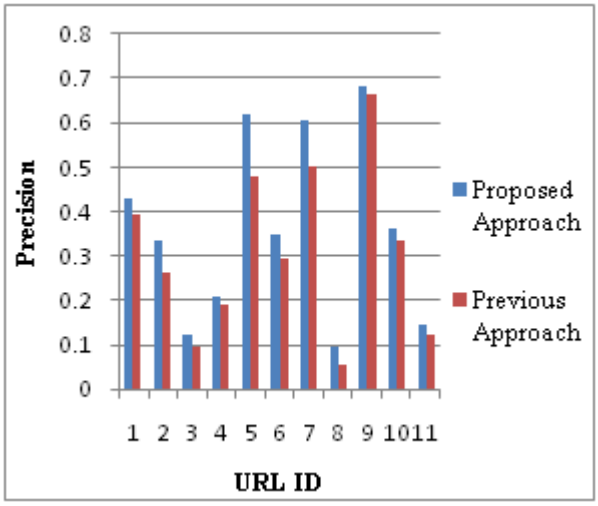

Figure 3: Graph shows comparison in terms of Precision

Recall: The fraction of the documents that is relevant to the query that is successfully retrieved [12]:

$$Recall = \frac{Retrieved\ Relevant\ documents}{Total\ Relevant\ documents}$$

Recall is calculated over both approaches and their comparison is shown in the form of table and graph.

Table 3: Comparisons of Recall

ID	*URLs*	*Recall of Proposed Approach*	*Recall of Previous Approach*
1.	http://www.naturewallpaper.net/	0.8888	0.8205
2.	http://www.whatsapp.com/	0.6666	0.5333
3.	http://www.wechat.com/en/	0.5454	0.4285
4.	http://all-free-download.com/free-photos/rose-flowers.html	0.8333	0.4074
5.	http://www.bbm.com/bbm/en.html	0.6923	0.5625
6.	http://www.homeshop18.com/	0.75	0.6315
7.	http://adaptive-images.com/	0.7647	0.65
8.	https://twitter.com/	0.2	0.125
9.	https://in.yahoo.com/?p=us	0.9365	0.8955
10.	http://jewellery.picturesklix.com/	0.9024	0.8409
11.	http://tanishq.co.in/Home	0.7142	0.5882

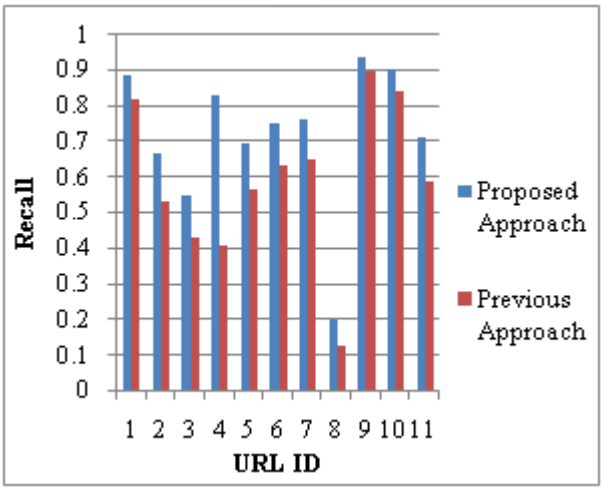

Figure 4: Graph demonstrates comparison in terms of Recall

F-Measure: The harmonic mean of combination of precision and recall:

$$F - Measure = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

F-Measure is compared between both approaches and their comparison is shown in the form of table and graph.

Table 4: Comparison of F-Measure

ID	*URLs*	*F-Measure of Proposed Approach*	*F-Measure of Previous Approach*
1.	http://www.naturewallpaper.net/	0.5763	0.5335
2.	http://www.whatsapp.com/	0.4444	0.3524
3.	http://www.wechat.com/en/	0.1979	0.1578
4.	http://all-free-download.com/free-photos/rose-flowers.html	0.3337	0.2573
5.	http://www.bbm.com/bbm/en.html	0.6536	0.5169
6.	http://www.homeshop18.com/	0.4752	0.3990
7.	http://adaptive-images.com/	0.6767	0.5652
8.	https://twitter.com/	0.1298	0.0769
9.	https://in.yahoo.com/?p=us	0.7911	0.7643
10.	http://jewellery.picturesklix.com/	0.5143	0.4803
11.	http://tanishq.co.in/Home	0.2422	0.2027

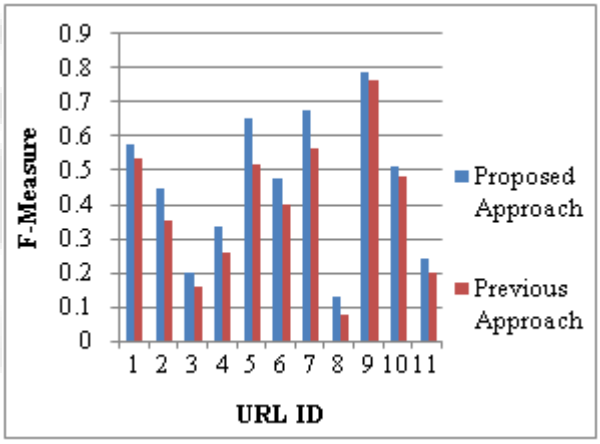

Figure 5: Graph reveals comparison in term of F-Measure

G-Measure: The geometric mean of combination of precision and recall:

$$G - Measure = \sqrt[2]{Precision \times Recall}$$

G-Measure is compared between both approaches and their comparison is shown in the form of table and graph.

Table 5: Comparison of G-Measure

ID	*URLs*	*G-Measure of Proposed Approach*	*G-Measure of Previous Approach*
1.	http://www.naturewallpaper.net/	0.6172	0.5695
2.	http://www.whatsapp.com/	0.4714	0.3746
3.	http://www.wechat.com/en/	0.2567	0.2038
4.	http://all-free-download.com/free-photos/rose-flowers.html	0.4170	0.2767
5.	http://www.bbm.com/bbm/en.html	0.6546	0.5186
6.	http://www.homeshop18.com/	0.5107	0.4291
7.	http://adaptive-images.com/	0.6813	0.5700
8.	https://twitter.com/	0.1386	0.0833
9.	https://in.yahoo.com/?p=us	0.8008	0.7726
10	http://jewellery.picturesklix.com/	0.5697	0.5317
11	http://tanishq.co.in/Home	0.3227	0.2683

Licensed Under Creative Commons Attribution CC BY

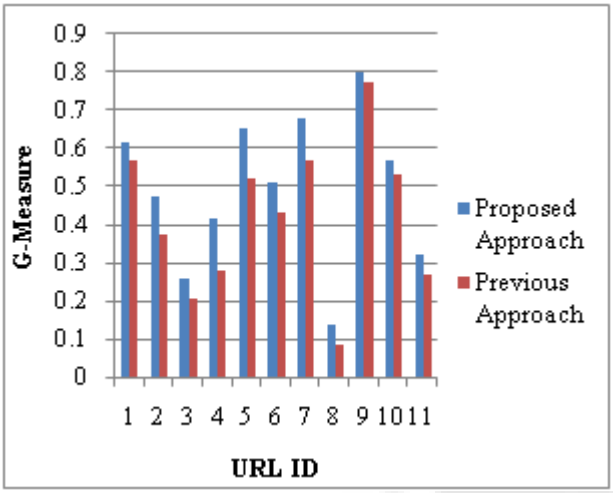

Figure 6: Graph represents comparison in terms of G-Measure

6. Conclusions

The rapid growth of the Web and its vast accessibility has created the need for better and more accurate search capability. In this paper, we have proposed a hybrid technique of Fuzzy K-Means with Weighted PageRank based on Visit of Links for image and links search. User refers a query and in response provide with quality search. To conclude all the results, we analyzed our method is very fast to retrieve the records in the form of image links and hyperlinks whereas in an existing work, PageRank with K-Means is very slow to retrieve the data from webpage. Same as an optimized hybrid technique returns more relevant information and quality search results to the user as compared to previous approach. We have test our proposed method and previous method on same well known URLs and extract the links and images and compute various factors upon them and our technique at last proved itself is better than previous approach. The main goal of our project the paper described is to provide the more reliable and relevant search results in response to user in the form of images and links. That can see our hybrid approach as future of web structure mining.

7. Future Scope

In this paper, our cluster based ranking scheme i.e. Weighted PageRank based on Visit of Links with Fuzzy K-Means is discussed which is more target oriented than previous approach. Our proposed hybrid technique calculates the weighted pagerank values based on visit of incoming links as well as popularity of incoming links and cluster the data using fuzzy k-means. We analyzed user spend a lot of surfing to find the more relevant information from webpage or web link. This technique provides more relevant result and return fast response to the user. So we can say our approach can be combined with search engine for optimizing it can be efficiently used by search engines to retrieve the required relevant links and images to the user.

References

[1] S. Brin, and Page L., "The Anatomy of a Large Scale Hypertextual Web Search Engine", Computer Network and ISDN Systems, Vol. 30, Issue 1-7, pp. 107-117, 1998.
[2] Miguel Gomes da Costa Júnior and Zhiguo Gong, "Web Structure Mining: An Introduction", In Proceedings of the 2005 IEEE International Conference on Information Acquisition, July 3, 2005
[3] Neelam Tyagi and Simple Sharma, "Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012
[4] Preeti Chopra and Md. Ataullah, "A Survey on Improving the Efficiency of Different Web Structure Mining Algorithms", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-2, Issue-3, February 2013
[5] Monika Yadav and Mr. Pradeep Mittal, "Web Mining: An Introduction", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013
[6] Gurpreet Kaur and Shruti Aggarwal, "Improving the Efficiency of Weighted Page Content Rank Algorithm using Clustering Method", International Journal of Computer Science & Communication Networks,Vol 3(4),231-239.
[7] Seifedine Kadry and Ali Kalakech, "On the Improvement of Weighted Page Content Rank", Journal of Advances in Computer Networks, Vol. 1, No. 2, June 2013
[8] Manpreet Kaur and Usvir Kaur, "Comparison Between K-Means and Hierarchical Algorithm Using Query Redirection", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3,Issue 7, July 2013
[9] Supreet Kaur and Usvir Kaur, "An Optimizing Technique for Weighted Page Rank with K-Means Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.
[10] Amar Singh and Navjot Kaur, " To Improve the Convergence Rate of K-Means Clustering Over K-Means with Weighted Page Rank Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3,Issue 8, August 2013.
[11] Jaideep Srivastava, Prasanna Desikan and Vipin Kumar, Web Mining— Concepts, Applications, and Research Directions
[12] Precision, Recall and F-measure Definition, Available: http://en.wikipedia.org/wiki/Precision_and_recall
[13] Web Mining Definition, Available: http://en.wikipedia.org/wiki/Web_mining
[14] PageRank Image, Available: http://en.wikipedia.org/wiki/PageRank

Licensed Under Creative Commons Attribution CC BY

Author Profile

Rashmi Sharma received the B.Tech. degree in Information Technology from Punjab Technical University, Jalandhar, Punjab, India in 2012. Currently, she is pursuing M. Tech in Computer Science and Engineering at Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India.

Kamaljit Kaur received the B.Tech. and M.Tech degree in Computer Science and Engineering. She is pursuing PhD in Data and Web Mining. Currently, she is an Assistant Professor at Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India. She has over 10 year's job experience and she is guiding various M.Tech. Thesis in the field of Database Security and Data Mining.

Licensed Under Creative Commons Attribution CC BY