

# A New Improved Clustering Algorithm based Diversified Web Page Recommendation

Meghna Guru  
M.E. Student, Dept. of CSE  
Meenakshi College of Engineering  
Chennai-600078

S. Anitha Angayarkanni  
Asst. Professor, Dept. of CSE  
Meenakshi College of Engineering  
Chennai-600078

J.C. Kavitha  
HOD, Dept. of CSE  
Meenakshi College of Engineering  
Chennai-600078

## ABSTRACT

The tremendous growth of internet over the years, has given rise to the large number of web services, containing lot of information. Due to this information overload, it has become difficult to get the correct information. Web Service Recommendation system focuses on satisfying the user's potential interests. Most of the existing recommendation approaches focus only on missing QoS values only, assuming that the result contains independent web services, which might not be true. As a result redundant web services appear in the list. The existing system takes into consideration active user's QoS preferences as well as diversification of the web services list. First, the active user's usage history is mined, and then the experiences of other service users are collected through collaborative filtering approach. Scores are computed for the web service candidates by measuring their relevance with historical and potential user interest and the QoS utility. Web Service graph is constructed based on the functional similarity of the web service candidates. Finally, the diversity-aware web service ranking algorithm is applied on the web service candidates based on the scores calculated and the diversified degree derived from the web service graph.

## Keywords

Web service recommendation, diversity, user interest, potential interest, QoS preference

## 1. INTRODUCTION

Web services have been rapidly developed in recent years and played important roles in E-Commerce and other organizations. With this information overload, the web service discovery has become a critical problem. Many web services are available with similar functionality but non-quality requirements. Web service selected should be high in quality, and must satisfy user's functional and non-functional requirements. [1]

Web Service Recommendation is a process of proactively discovering and recommending web services to the user. Most of the recommendation system uses collaborative filtering approach, content-based approach, or a mix of both i.e., hybrid approach, for recommending potential active user's interest. Most of the systems focus only on missing QoS preferences. But, they miss to exploit the user's own service usage history for predicting QoS preferences. [10][12][18]

Another major drawback of existing systems is that, they assume that the top-k recommendation list is dissimilar, which may not be true. They do not feel the need to make top-k recommendations dissimilar. As a result, user's satisfaction degree is reduced. In the top-k recommendation list, the service with higher QoS appears and because of higher QoS, that web service is redundant in the list. However, the user's interest is more focused on the functionality of the web

service then the higher QoS. In order to remove the redundancy and at the same time to maintain the quality, diversity should be considered in the recommendation systems. It is desirable for a recommender system to return a diverse set of cases in order to provide the user with the optimal coverage of the information space. [12]

In this paper, the top-k recommendation list is given to the user with the diversity of web services. The system takes into consideration functional and non-functional requirements of the user with the diversified recommendation list. User's usage history is mined, through the text similarity and operation similarity. Then active user's potential interest is calculated through the collaborative filtering approach. QoS preferences are mined from user's usage history. Scores are calculated for each web service candidate through interest relevance and QoS utility. Ranking of optimal web services is done through diversity-aware web ranking algorithm. [18]

## 2. PREVIOUS WORKS

A literature survey on the Web Service Recommendation has number of approaches for discovering web service, but there are few drawbacks on QoS parameters, diversified ranking.

### 2.1 Diversifying Web Service Recommendation

In this technique, Guosheng Kang, Mingdong Tang, Jianxun Liu, Xiaoqing Liu and Buqing Cao, projected a web service recommendation approach incorporating user's potential QoS preferences and diversity feature of user interests on web services. User's interests and QoS preferences on web services are initial mined by exploring the web service usage history. User's potential interests are collected through cooperative filtering approach. Then the scores are computed for web service candidates by measuring their relevance with historical and potential user interests, and their QoS utility. Web service graph is built based on the functional similarity between web services. Then diversity-aware web service ranking algorithmic rule is employed to rank the web service candidates based on their scores, and diversity degrees derived from the web service graph. [1]

### 2.2 Combining Collaborative Filtering with Content-based Features

This technique is proposed by L. Yao, Q. Z. Sheng, A. Segev, proposes a system where web services are dynamically recommended to fit user's interests. The hybrid approach is used, which is the combination of collaborative filtering and content-based recommendation. Collaborative filtering is used to recommend web services to a user by considering other similar user's ratings on their web services. Content-based method recommends web services based on the similarity of user preferences and content of web services. The approach

considers simultaneously both rating data and content data of web services using a three-way aspect model. [2]

### 2.3 Location-Aware Collaborative Filtering

M. Tang, Y. Jiang, J. Liu, and X. Liu, proposed a new method where location-aware collaborative filtering is used to recommend web services to users by incorporating locations of both users and services. Different from other user-based collaborative filtering for finding similar users for a target user, instead of searching entire set of users, concentration is on users physically near to the target user. Service similarity measurement is modified and used as service location information. After finding similar users and services, similarity measurement is used to predict missing QoS values based on a hybrid collaborative filtering technique. Web service candidates with the top QoS values are recommended to users. [3]

### 2.4 Clustering Web Services to Facilitate Service Discovery

J. Wu, L. Chen, Z. Zheng, M. R. Lyu, and Z. Wu, proposed a system where web services are clustered by utilizing WSDL documents and tags. To handle the clustering performance limitation caused by uneven tag distribution and noisy tags, a hybrid web service tag recommendation strategy, WSTRec, which employs tag co-occurrence, tag mining and semantic relevance measurement for tag recommendation. The advantage of WSDL-based clustering approach is taken by extracting five features from WSDL documents i.e., content, type, message, port and service name and compute the WSDL-level similarities among web services. Then the tag-level similarities are computed between web services. Then the WSDL-similarity and tag-similarity are merged to be a composite similarity, which is used to cluster web services. [4]

### 2.5 History Record-Based Service Optimization Method

W. Lin, W. Dou, X. Luo, and J. Chen, planned a history-record based service optimization methodology is proposed in this research work. This methodology aims at enhancing the credibility of service composition plan, by using the records of service's performance quality related to their past executions to rate service's performance quality next time within the future, instead of using the tentative QoS values publicized by the service provider. Some is enabled by 3 constituent components i.e., data assets, an IP-based solver Some-solver, algorithmic rule for computing contribution score. Higher the web service's contribution score is, the more contribution it's done to the final optimal composition plan. Web service that owns the highest contribution score, within the service pool, will be chosen as the final service to compose the final composition plan. [5]

### 2.6 QoS-Based Service Ranking and Selection

S. S. Yau, and Y. Yin, presents a "QoS-based service ranking and selection" approach, to assist users to select the service that best satisfies user's QoS requirements from a set of services having already satisfied user's functionality requirements. To determine how well a service satisfies user's involved QoS requirements, a set of functions is conferred to normalize service's QoS on numerous QoS aspects with different metrics and scales, compute service's satisfaction scores on every QoS aspect, and combine every service's

satisfaction scores on all QoS aspects along as an overall satisfaction scores. Service is chosen so best satisfies users QoS requirements instead of the service with the best QoS which can be overqualified for the user's QoS requirements. Prospect theory is employed to more accurately model the relation between service's QoS and their satisfaction scores. [6]

### 2.7 Personalized Collaborative Filtering

Y. Jiang, J. Liu, M. Tang, and X. Liu, presents an effective personalized collaborative filtering method. Different from the Pearson Correlation Coefficient (PCC) for similarity measurement, the personalized influence of services is taken into consideration, when computing similarity measurement between users and personalized influence of services. Based on the similarity measurement model of web services, an effective Personalized Hybrid Collaborative Filtering technique is developed, by integrating personalized user-based algorithm and item-based algorithm. User-based approach recommends to an active user items collected by other users sharing similar tastes. Item-based approach recommends to an active user those items similar to the ones which the active user has preferred in the past. [7]

### 2.8 Region KNN

X. Chen, X. Liu, Z. Huang, and H. Sun, presents Region KNN, a hybrid collaborative filtering algorithm designed for large scale web service recommendation. The method employs the characteristics of QoS by building an efficient region model. Based on this model, web service recommendations will be generated quickly by using modified memory-based collaborative filtering algorithm. Users are clustered based on their regions and historical QoS similarities. Then region-sensitive services are identified. Nearest neighbor-based approach predicts QoS of web services for an active user by leveraging historical QoS information gathered from users of highly correlated regions. Based on the prediction, the service with the best predicted QoS will be recommended to the active user. [8]

## 3. PROPOSED ALGORITHM

The proposed module implements the selection of web services through functional and non-functional evaluation using clustering.

### 3.1 Query Processing

*Step 1* – Usage history dataset is mined to get the relevant web services related to the query

This is based on terms in WSDL (Web Service Description Language) documents of the available web service candidates can be looked upon as a corpus and therefore we employ the TF/IDF (Term Frequency/Inverse Document Frequency) algorithm. TF/IDF is a statistical measure to evaluate how important a word is to a document in the corpus. The term count is the number of times the word occurring in the document. This count is usually normalized to prevent a bias towards longer documents to give a measure of the importance of the term. Thus, the term frequency  $tf(t_j, WSDL_i)$  of the  $j^{th}$  term of the WSDL document  $WSDL_i$  in the corpus is calculated as follows:

$$tf(t) = \frac{\text{freq}(t_j, WSDL_i)}{|WSDL_i|} \quad (1)$$

Where  $t_j$  is the  $j^{th}$  term in the corpus;  $WSDL_i$  is the WSDL document of the  $i^{th}$  web service  $WS_i$ ;  $\text{freq}(t_j, WSDL_i)$  is the occurrence number of  $t_j$  in  $WSDL_i$ ;  $|WSDL_i|$  is the total occurrence number of all meaningful terms in  $WSDL_i$ .

**Step II** – Potential user interest is collected through collaborative filtering

Web service recommendation system not only takes user's usage history in consideration, but experiences of other web service users are also considered. Experience of other web service users can be used to predict the potential interest of the active user. Collaborative filtering approach is used to predict the potential interest of active user. In collaborative filtering approach, user similarity is calculated based on the web service invocation records of a set of users. Similar users share the common interests, so likely to use the web services with same functionality. The more commonly invoked web services two users have in their invocation records, the larger the user similarity between them. User similarity is calculated as follows:

$$\text{userSim}(u_i, u_j) = \frac{2 \times C_{ij}}{|S_{u_i}| + |S_{u_j}|} \quad (2)$$

Where  $S_{u_i}$  and  $S_{u_j}$  are the sets of web services used by user  $u_i$  and  $u_j$  respectively,  $C_{ij}$  is the set of web services used by both user  $S_{u_i}$  and  $S_{u_j}$ . If  $|C_{ij}| = 0$ , then  $\text{userSim}(u_i, u_j) = 0$ . In collaborative filtering approach, web services used by similar users are recommended to the active user.

**Step III** – Clusters of web services which are closely related and similar to each other are created from historical user interest and potential user interest.

**Step IV** – Hierarchical association between the web services is calculated through the attribute based clustering.

### 3.2 Clustering Algorithm

**Input:** Query to be searched.

**Step 1:** Mine the dataset.

**Step 2:** Get the [url, mined words] related to the query.

**Step 3:** Stemming algorithm is applied to remove unwanted words.

**Step 4:** Group the [(url, mined words), c1] into clusters of closely-related and similar clusters.

**Output:** Clusters of web services.

### 3.3 Hierarchical Association

**Input:** Historical interest and collaborative filtered clusters.

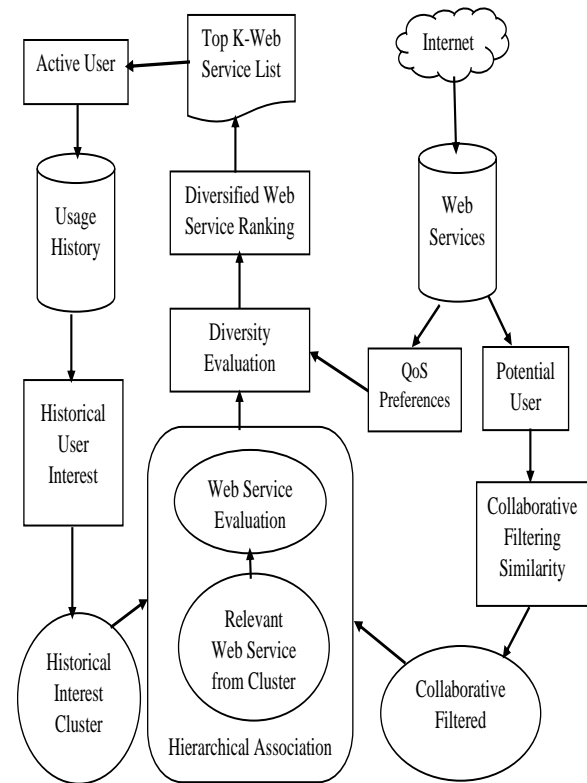
**Step 1:** Select the clusters which are closely related.

**Step 2:** Checked the association between web services based on the attributes.

**Step 3:** Associated and most relevant web services are selected.

**Output:** Relevant web services which have some association between them.

### 3.4 Proposed System Architecture



**Fig. 1: Flow diagram of Proposed Method**

The data flow can be explained as following:

**Step 1:** User inputs the query.

**Step 2:** Usage history is checked.

**Step 3:** Historical user interest is collected.

**Step 4:** Clusters of historical interest data are created.

**Step 5:** Internet is searched to get the QoS preferences and potential user interest.

**Step 6:** Potential interest is calculated through collaborative filtering.

**Step 7:** Clusters of web services are created from collaborative filtered web services.

**Step 8:** Historical interest cluster and collaborative filtered cluster are now combined in hierarchical association to get associated web services.

**Step 9:** Diversity is evaluated of the collected web services taking QoS preferences into consideration.

**Step 10:** Diversified web service ranking is calculated.

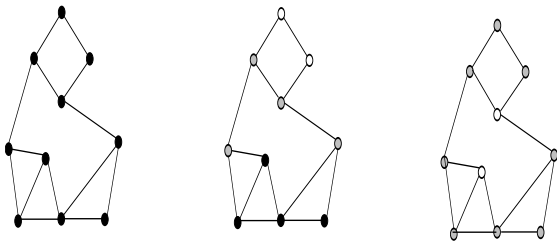
**Step 11:** Top-K diversified web service list is returned to the user.

### 3.5 Diversified Web Service Ranking

This Module Implements the Diversity Evaluation and Diversified Web Service Ranking On Web Services Obtained Through Clustering and Hierarchical Association.

Consider the Web service graph in Figure 2 and select two nodes from it. Figure 2(b) and Figure 2(c) are two different cases, where white nodes represent the nodes selected. There

are six nodes (white nodes and gray nodes) in the 2-hop expanded set for the case in Figure 2(b), and ten nodes in the 2-hop expanded set for the case in Figure 2(c). Then, the 2-hop expansion ratio of the selected nodes in Figure 2(b) and Figure 2(c) are 0.6 and 1.0, respectively. The selected nodes in Figure 2(b) are well connected, thus they are probably similar to one another. As a contrast, there is no edge between the two selected nodes in Figure 2 (c).



(a) A Web Service Graph (G) (b)  $\sigma=0.6$  (c)  $\sigma=1.0$

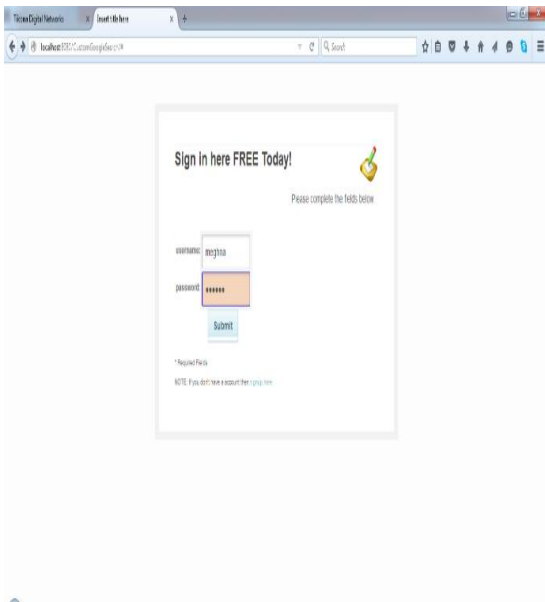
**Fig. 2: Illustration of 2-hop Expansion Ratio vs. Diversity**

Based on the k-hop expansion, we define the generalized diversified ranking measure  $F_k(S)$  as follows –

$$F_k(S) = (1 - \lambda) \sum_{v \in S} \text{Score } v + \lambda \frac{\|N_k(S)\|}{K} \quad (3)$$

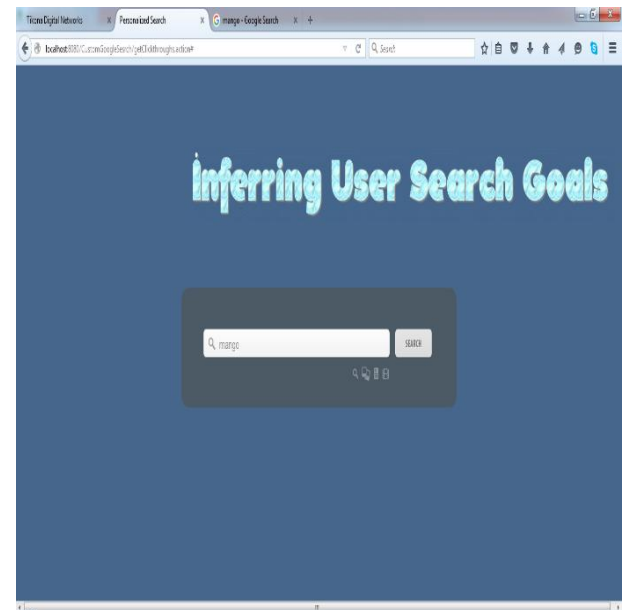
#### 4. EXPECTED OUTCOMES

The simulated outcome of new proposed improved clustering algorithm based diversified web page recommendation is shown in below. Figure 3 shows the login page of new proposed method that is designed in JAVA script language.

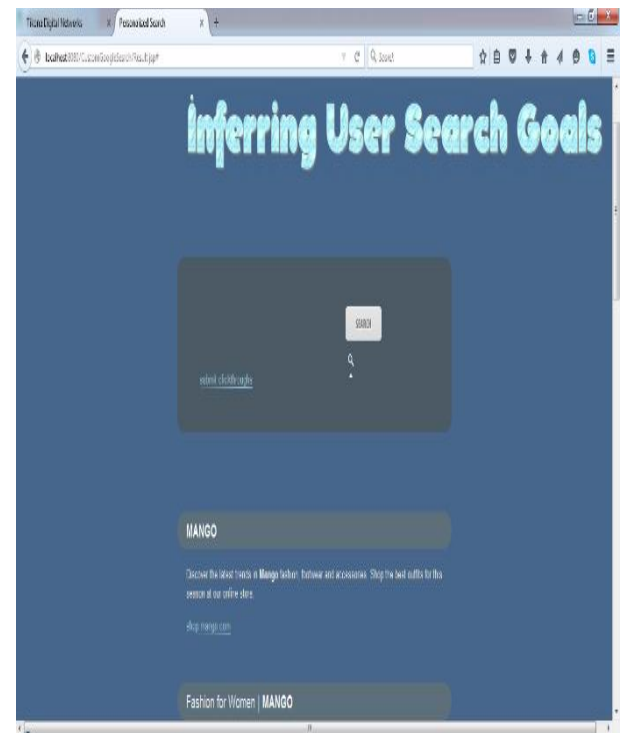


**Fig. 3: Login Page**

Figure 4 shows the inferring user search goals that is query input page. In this page of proposed method design for giving the input of query.

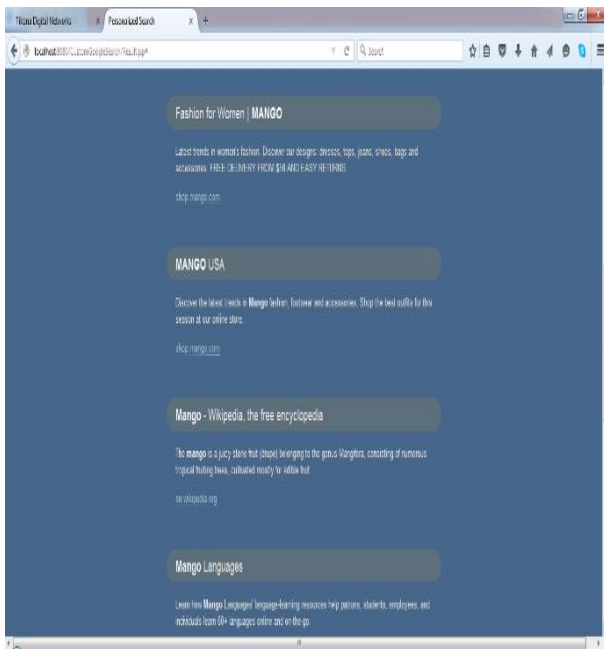


**Fig. 4: Query Input**



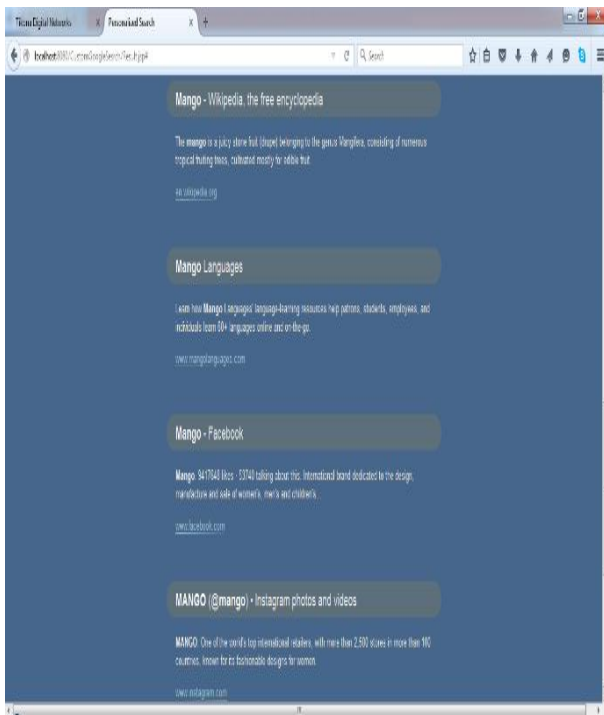
**Fig. 5: Diversified Output 1**

In figure 5 shows the query output of diversified output method that is shown in figure. The output of proposed method of for first query is shown in figure.



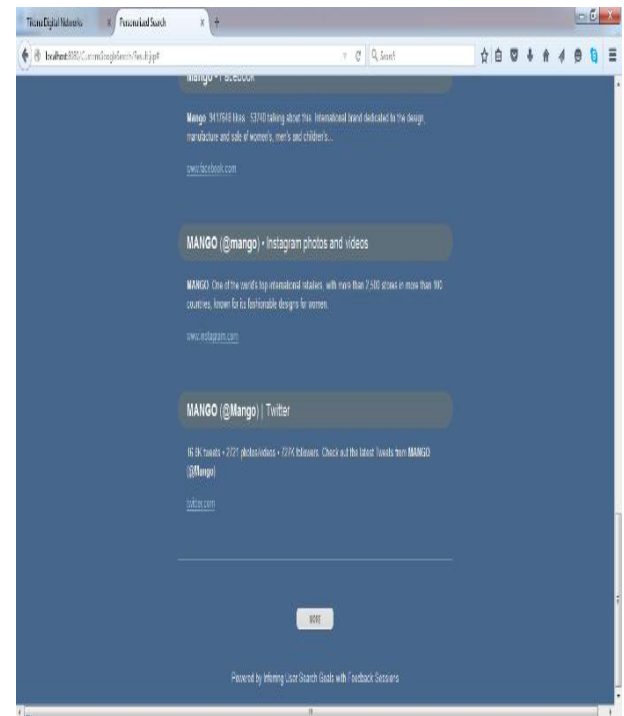
**Fig 6: Diversified Output 2**

In figure 6 shows the query output of diversified output method that is shown. The output of proposed method of for second query. In similar also give the query of base method that is Diversifying Web Service Recommendation Results via Exploring Service Usage History [1].



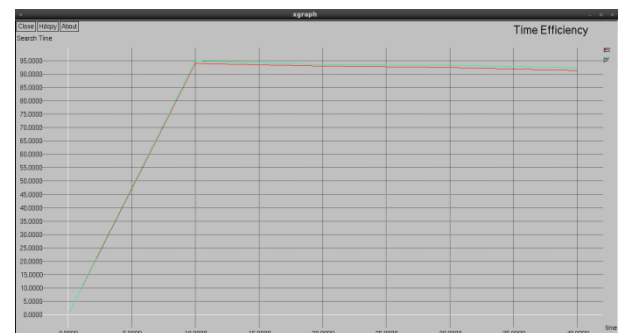
**Fig. 7: Diversified Output 3**

In figure 7 shows the query output of diversified output method that is shown in figure 7. The output of proposed method of for third query. In similar also give the query of base method that is Diversifying Web Service Recommendation Results via Exploring Service Usage History [1].



**Fig 8: Diversified Output 4**

The Graphical Representation of expected outcome of proposed method is shown in the both figure 9 and figure 10. The proposed simulated graph is calculated on the network simulator tool (NS-2).



**Fig. 9: Graphical Representation of Expected Outcome**



**Fig. 10: Graphical Representation of Expected Outcome**

The new proposed improved clustering algorithm based diversified web page recommendation shown in graphs green color as compare to diversifying web service recommendation [1] denoted by red. The improvement of the proposed method better as compare to diversifying web service shown in figure

9 and 10. The outcome compression is based on accuracy and precision that is shown in figure 9 and 10.

## 5. CONCLUSION

In the existing system, a Web service recommendation approach is presented, with diversity to find desired Web services for users. Here incorporate functional interest, QoS preference, and diversity feature for recommending top-k diversified Web services. A diversified Web service ranking algorithm is proposed to find the top-k diversified Web service ranked list based on their functional relevance including historical user interest relevance and potential user interest relevance, non-functional relevance such as QoS utility, and diversity feature.

In the proposed system, the concept of clustering is employed for improving selection of the web services. The clustering is combined with the hierarchical association. In hierarchical association, the web services are associated with each other, to check the association between the input query and the selected web services.

## 6. REFERENCES

- [1] Guosheng Kang, Mingdong Tang, Jianxun Liu, Xiaoqing Liu and Buqing Cao, "Diversifying Web Service Recommendation Results via Exploring Service Usage History," *IEEE Transactions on Service Computing*, 2015.
- [2] L. Yao, Q. Z. Sheng, A. Segev, and Yu, "Recommending Web Services via Combining Collaborative Filtering with Content-based Features," *Proc. of Int. Conf. on Web Services*, pp. 42-49, 2013.
- [3] M. Tang, Y. Jiang, J. Liu, and X. Liu, "Location-Aware Collaborative Filtering for QoS-based Service Recommendation," *Proc. of Int. Conf. on Web Services*, pp. 202-209, 2012.
- [4] J. Wu, L. Chen, Z. Zheng, M. R. Lyu, and Z. Wu, "Clustering Web Services to Facilitate Service Discovery," *Knowledge and Information Systems*, pp. 1-23, 2012.
- [5] W. Lin, W. Dou, X. Luo, and J. Chen, "A History Record-Based Service Optimization Method for QoS-Aware Service Composition," *Proc. of Int. Conf. on Web Services*, pp. 666-673, 2011.
- [6] S. S. Yau, and Y. Yin, "QoS-Based Service Ranking and Selection for Service-Based Systems," *Proc. of Int. Conf. on Service Computing*, pp. 56-63, 2011.
- [7] Y. Jiang, J. Liu, M. Tang, and X. Liu, "An Effective Web Service Recommendation Based on Personalized Collaborative Filtering," *Proc. of Int. Conf. on Web Services*, pp. 211-218, 2011.
- [8] X. Chen, X. Liu, Z. Huang, and H. Sun, "RegionKNN: A Scalable Hybrid Collaborative Filtering Algorithm for Personalized Web Service Recommendation," *Proc. of Int. Conf. on Web Services*, pp. 9-16, 2010.
- [9] Z. Zheng, H. Ma, M. R. Lyu, I. King, "Wsrec: a collaborative filtering based web service recommender system," *Proc. of Int. Conf. on Web Services*, pp. 437-444, 2009.
- [10] Mohammad Alrifai, Dimitrios Skoutas, Thomas Risse, "Selecting Skyline Services for QoS based Web Service Composition", *WWW2010*, April 26-30, 2010, Raleigh, North Carolina.
- [11] Jian Wu, Liang Chen, Zibin Zheng, Michael R. Lyu, Zhaoxun Wu, "Clustering Web services to facilitate service discovery," © Springer-Verlag London 2013, *Knowl Inf Syst* (2014) 38:207–229.
- [12] Yutu Liu, Anne H.H. Ngu, Liangzhao Zeng, "QoS Computation and Policing in Dynamic Web Service Selection", *WWW2004*, May 17–22, 2004, New York, New York, USA. ACM 1-58113-912-8/04/0005.
- [13] Lina Yao and Quan Z. Sheng, Aviv Segev, Jian Yu, "Recommending Web Services via Combining Collaborative Filtering with Content-based Features", 2013 *IEEE 20th International Conference on Web Services*, 978-0-7695-5025-1/13 © 2013 IEEE.
- [14] Lu Qin, Jeffrey Xu Yu, Lijun Chang, "Diversifying TopK Results", 38th *International Conference on Very Large Data Bases*, *Proceedings of the VLDB Endowment*, Vol. 5, No. 11 © 2012 VLDB Endowment.
- [15] Rong-Hua Li, Jeffrey Xu Yu, "Scalable Diversified Ranking on Large Graphs", 2011 11th *IEEE International Conference on Data Mining*, © 2011 IEEE.
- [16] Karen Spärck Jones, "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*, Volume 60 Number 5 2004 pp. 493-502 Copyright © MCB University Press ISSN 0022-0418.
- [17] Yechun Jiang, Jianxun Liu, Mingdong Tang, Xiaoqing (Frank) Liu, "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering", 2011 *IEEE International Conference on Web Services*, 978-0-7695-4463-2/11 © 2011 IEEE.
- [18] Mingdong Tang, Yechun Jiang, Jianxun Liu, Xiaoqing (Frank) Liu, "Location-Aware Collaborative Filtering for QoS-Based Service Recommendation", 2012 *IEEE 19th International Conference on Web Services*, 978-0-7695-4752-7/12 © 2012 IEEE.
- [19] Xi Chen, Xudong Liu, Zicheng Huang, and Hailong Sun, "Region KNN: A Scalable Hybrid Collaborative Filtering Algorithm for Personalized Web Service Recommendation", 2010 *IEEE International Conference on Web Services*, 978-0-7695-4128-0/10 © 2010 IEEE.
- [20] Neil Hurley, Mi Zhang, "Novelty and Diversity in Top-N Recommendation – Analysis and Evaluation", *ACM Transactions on Internet Technology*, Vol. 10, No. 4, Article 14, Publication date: March 2011.