ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 10, October 2016

Clustering Web services for effective service discovery

M.Chandrasekar, K.JayaShree

***Abstract*- A web service is a collection of open protocols and standards used for exchanging data between applications or systems. Software applications written in several programming languages and running on many platforms can use web services to interchange data over computer networks like the Internet in a way similar to inter-process communication on a single computer. Web services are used for various applications. As it grows in numbers, it is difficult to retrieve the correct web services by user and other applications. Hence it is necessary to cluster the web services to recognize the specific web service and make the service discovery much more effective. The objective of the proposed work is to cluster various web services.**

***Index Terms*--Clustering, Service discovery, Web service, XML response**

I. INTRODUCTION

A web service is any piece of software that makes itself available over the internet and uses a consistent Extensible Markup Language (XML) messaging system. Interoperability of web service is due to the use of open standards. Web services describes a standardized way of integrating Web-based applications using the Universal Description, Discovery and Integration (UDDI) specification, XML, Simple Object Access Protocol (SOAP) and Web Service Description Language (WSDL) open standards over an Internet protocol backbone. XML is used to tag the data and XML is used to encode all communications to a web service, SOAP is used to transfer the data, WSDL is used for describing the services available and UDDI is used for listing what services are available. Used primarily as a means for businesses to communicate with each other and with clients, web services allow organizations to communicate data without intimate knowledge of each other's IT systems behind the firewall. It has been in various applications such as Hotels, Transport services, Health care Services, Financial Services. Due to the large amount of web services, application of data mining is required to improve the performance in the service identification process. Data mining is a process used to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers and develop more effective marketing strategies as well as increase sales and decrease costs. Data mining depends on effective data collection and warehousing as well as computer processing. Data mining algorithms are applied to web services to cluster it into the functional similar groups hence the service can be retrieved certainly.

Manuscript received Oct, 2016

***M.Chandrasekar**, PG Student, Department of Computer Science and Engg, Rajalakshmi Engg College, Chennai,India*
***K.JayaShree**, Associate professor, Department of Computer Science and Engg, Rajalakshmi Engg College, Chennai,India*

II. RELATED WORK

A lot of research has been carried out in the field of clustering web services. Text mining approach is automatically classify services to specific domains and identify key concepts inside service textual documentation. Classification approach can be used to focus user queries to a refined set of web service categories [1]. The method discussed in [2] mine WSDL documents and cluster them into QoS similar Web service groups. QoS mediator agent based Web Service Selection Model is anticipated where QoS Consultant acts as a Mediator Agent between clients and service providers [3]. The approach is used as an effective reduction of dimension techniques; they are able to acquire semantic relations between word-item and item-document interpreted in terms of probability distributions [4]. To extend the groups, the similar web services and semantic representation of services for the enhancement of the service discovery they presented the Semantic Web services Clustering (SWSC) method [5]. New approach for semantic web services ranking [6], this approach will help to users to search needed web service in precise format. This approach presents an algorithm for ranking depend on the VSM (Vector Space Model). In this model the user query and reclaimed services related to that user query will be considered as vector and based on that which service vector is very close to the user query vector will considered as higher rank. The rank of services are calculated automatically whenever the user fire the required query to system and rank values are not predefined to the web services. In order to extend the scope of web service utilization, accessing and investigating large service repositories by useful techniques. By using semantic concepts, they specified the method for the sake of categorization by labeling the web services automatically [7]. In order to get or extract the latent factors from the semantic explanation of web service and search the web services in latent factor space, techniques such as machine learning in particular PLSA(Probabilistic Latent Semantic Analysis) and LDA (Latent Dirichlet Allocation) are used by a non- logic based matchmaking approach. This paper also defines generation of a lower dimensional vector model for demonstration of services and use of an probabilistic machine learning for service matching and ranking and

these LDA PLSA unsupervised probabilistic machine learning methods used for the service description data[8]. An improved web services clustering method which uses Piano Space filling curve. The work argues that an effective clustering technique is needed to support web services selection and retrieval processes. The technique must have the abilities to consider all the possible parameter matches and include multiple criteria in the processes of parameter matching. The technique proposed by the work is able to meet these requirements, and is compared with the previous work that employs Hilbert space filling curve. The results show that the proposed technique is better than the previous work in terms of fairness, scalability and irregularity [9]. Ref. [10] proposes the clustering of web services for discovery purposes. The clustering method proposed by the work employs web services ontology. That means, web services matching is calculated based on their functions and processes using an precise semantic similarity concept of the domain ontology. Ref. [11] argues that web services discovery which is based only on registries is not adequate in assuring good process delivery. In addition, the increasing number of web services has also caused bad performance and low efficiency of discovery process. Hence, they propose the discovery method that is based on semantics and clustering. In this work, the web services descriptions similarity is used for clustering purpose. Particle Swarm Optimization is proposed as a method for web services clustering. The method applies the similar concept to cluster services, which is by evaluating the similarity of services semantic descriptions. Moreover, the work also defines a set of metrics which is able to perform this similarity assessment [12]. Ref. [13] argues that web services matching processes are slow as a result of complex semantic calculations. Thus, they propose a clustering method that is based on user preferences and ontology. The proposed method is capable of removing unrelated web services in its process, hence producing semantic calculation results with less time. Ref. [14] suggests semi-supervised clustering algorithm for web services composition. The algorithm performs the clustering based on tags and constrictions, and uses a huge number of unlabeled data in order to support the supervised learning procedure. Ref. [15] presents a web services clustering method called WTCluster, which employs both WSDL documents and tags annotation in its operation. The work argues that considering information that is gathered only from WSDL for web services clustering can actually limit its accuracy. Hence, the WT Cluster pools the WSDL information with the web services functional descriptions, and contextual and semantic information. This additional information is sought from the tags that are provided by web services search engines. Ref. [16] argues that clustering of web services is important especially in the current trend of the growing number of web services application development. Hence, this work presents a similarity model to evaluate the similarity among web services for clustering purpose. The model employs a preprocessing approach that takes into account the programming style and naming rules of services information. This approach also employs the Structural Clustering Algorithm for Networks (SCAN) that can further improve the similarity calculation. The model emphases on evaluating the functional similarity of web services. Ref. [17] proposes a web services recommendation system which ranks web services based on their predicted QoS. The QoS prediction is performed based on its patterns. Due to the dynamic nature, each pattern does not have an particular cluster. Hence the work argues that fuzzy clustering can produce more correct prediction results than hard clustering since the former does not assign each of the QoS patterns with an exact cluster. As an alternative, each pattern has more than one cluster with different degrees of membership. The work however, focuses on improving fuzzy clustering method to produce a better QoS prediction for web services recommendation system. There is no description on categorizing web services QoS into fuzzy clusters as proposed by this paper. Ref. [18] proposes an algorithm that uses clustering technique to cluster a huge number of web services into a number of groups according to QoS properties. The algorithm is able to reduce computational time and produce near-optimal web services selection process. The clustering is performed by employing the OPTICS algorithm and does not consider fuzzy clustering as proposed by this paper. Ref. [19] argues that web services must also be able to support a variety of mobile devices in mobile ad-hoc network (MANET), the environment in which mobile devices move independently. So, they propose a method for services discovery in MANET. The proposed method performs the clustering based on the devices mobility. The services discovery is carried out based on the clusters characteristics. Ref. [20] presents an algorithm for web services clustering that utilizes graph theory and a corresponding algorithm of web services discovery. The proposed algorithms are designed for semantic web services, hence, ontology is used to describe the input and output parameters of the services. Ref. [21] proposes however another semantic clustering of web services that can resolve the problem of semantic web services discovery in the server center. The clustering method is performed based on semantic descriptions of web services functionalities. A major interference in exploiting external information sources for service clustering is that irrelevant information may be included that is outside of context of the original services. This may negatively affect the clustering accuracy, which has been observed in [22]. Ref. [23] proposes a system that uses methods such as probabilistic machine-learning techniques and logic based search to extract latent factors from user query and match with possible web service by searching in ranked and clustered web service repository. Otherwise it discovers the related web service from web and adds it in web service repository for additional use.

III. Proposed Architecture

The proposed architecture is shown in Fig 1.The proposed architecture consists of three components namely service client, service registry and service provider. Service providers creates and publishes the web services in the service registry. It uses publish operation to publish the services. A service registry is an important part of any service-oriented architecture. It designed to allow developers to register their services and developer or end-user, to locate

useful services. Service client consumes the web service available in the service registry. Service client find the web service using keywords in the service registry and obtain WSDL file from registry. Service client application consumes WSDL and sends a service request message to a service provider, service provider returns a response message to the service consumer.

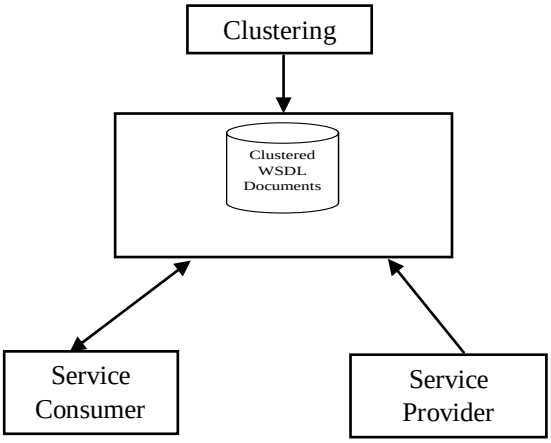

Fig.1 Proposed architecture of web service clustering

Clustering is a process of partitioning a set of data or objects into a set of meaningful sub-classes, called clusters. It help users understand the natural grouping or structure in a data set and used either as a stand-alone tool to get insight into data distribution or as a preprocessing step for other algorithms. Service registry consists of massive amount of web services so discovery of particular service is a critical task. Hence clustering is applied to the web services and resultant clustered web services can be placed in the service registry. Various algorithm such as K-means, Hierarchical clustering, Density Based Clustering can be used for clustering. Since the web service in the service registry is clustered into several groups it is easy to get the relevant web service efficiently.

IV. CONCLUSION

Web services are used in enormous number of applications. Hence there is necessity to retrieve correct web service from enormous number of web services. Therefore mining technique is applied to web services, to mine the required web service. So in our proposed system clustering technique is applied to cluster the services into the several groups. Henceforth the efficiency in service discovery can be improved as well as the user can also get the correct web service as according to their requirements.

REFERENCES

[1] Rozina Nisa, Usman Qamar, "A text mining based approach for web service Classification", *Information Systems and e-Business Management*, Vol 13, Issue 4, Nov 2015.

[2] R.Karthiban, "A QoS-Aware Web Service Selection Based on Clustering", *International Journal of Scientific and Research Publications*, Vol 4, Issue 2,Feb 2014.

[3] D Sachan, "A formalized model for semantic Web service selection based on QoS Parameters", *International Journal of Software and Web Sciences*, pp. 265-283, 2014.

[4] W. Wei, P. Barnaghi, and A. Bargiela, "Probabilistic topic models for learning terminological ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1028-1040, 2010.

[5] R. Nayak and B. Lee, "Web service discovery with additional semantics and clustering," in Proceedings of the IEEE/WIC/ACM *International Conference on Web Intelligence. Washington*, DC, USA: IEEE Computer Society, pp. 555-558, 2007.

[6] O. Mola, P. Emamian, and M. Razzazi, "A vector based algorithm for semantic web services ranking," *Information and Communication Technologies: From Theory to Applications*, 2008.ICTTA 2008. 3rd International Conference on, pp. 1-5, Apr.2008.

[7] A.Segev and E. Toch, "Context-based matching and ranking of web services for composition," *IEEE Transactions on Services Computing*, vol. 99, no. Pre Prints, pp. 210-222, 2009.

[8] Gilbert Cassar, Payam Barnaghi and Klaus Moessner, "Probabilistic Matchmaking Methods for Automated Service Discovery" *IEEE Transactions on Services Computing*, 2013

[9] A. S. Sukumar, J. Loganathan, and T. Geetha, "Clustering Web services based on multi-criteria service dominance relationship using Piano Space filling curve,"*International Conference on Data Science and Engineering* , pp. 13-18, 2012.

[10] L. Xie, F. Chen, and J. Kou, "Ontology-based semantic web services clustering," in Proc. *IEEE 18th International Conference on Industrial Engineering and Engineering Management*, pp. 2075-2079, 2011.

[11] T. Wen, G. Sheng, Y. Li, and Q. Guo, "Research on Web service discovery with semantics and clustering," in Proc. *6th IEEE Joint International Information Technology and Artificial Intelligence Conference*, pp. 62-67, 2011

[12] A. Nagy, C. Oprisa, I. Salomie, C. B. Pop, V. R. Chifu, and M.Dinsoreanu, "Particle swarm optimization for clustering semantic web services," in Proc. *10th International Symposium on Parallel and Distributed Computing*, pp. 170-177,2011.

[13] M. Li and Y. Yang, "Efficient clustering index for semantic Web service based on user preference," in Proc. *International Conference on Computer Science and Information Processing*, pp. 291-294,2012.

[14] Q. Zheng and Y. Wang, "The application of semi-supervised clustering in web services composition," *Advances in Intelligent and Soft Computing*, vol. 169, pp. 683-688, 2012.

[15] L. Chen, L. Hu, Z. Zheng, J. Wu, J. Yin, Y. Li, and S. Deng, "WTCluster: Utilizing tags for web services clustering," *Lecture Notes in Computer Science*, vol. 7084, pp. 204-218, 2011.

[16] Z. Zhu, H. Yuan, J. Song, J. Bi, and G. Liu, "WS-SCAN: An effective approach for Web services clustering," in Proc. *International Conference on Computer Application and System Modeling*, pp. V5-618-V5-622, 2010.

[17]M. Zhang, X. Liu, R. Zhang, and H. Sun, "A Web service recommendation approach based on QoS prediction using fuzzy clustering," in Proc. *IEEE 9th International Conference on Services Computing*, pp. 138-145,2012.

[18] Y. Xia, P. Chen, L. Bao, M. Wang, and J. Yang, "A QoS-aware Web service selection algorithm based on clustering," in Proc.*IEEE 9th International Conference on Web Services*, pp.428-435, 2011.

[19] Y. Kim, Y.S. Shim, and K.H. Lee, *Mobile Information Systems*, vol. 7, no. 4, pp. 299-315, 2011.

[20] Y. Li, "Algorithm for semantic web services clustering and discovery," in Proc. *WRI International Conference on Communications and Mobile Computing*, pp. 532-536, 2010.

[21] L. Wang, J. Si, and X.-B. Hu, "Research on the clustering and composition of P2P-based web services," in Proc. *2nd International Conference on Biomedical Engineering and Informatics*, pp. 1-5, 2009.

[22]Q. Yu and J. Kang. Integrating user invocation data and extended semantics for service community discovery. *International Journal of Next-Generation Computing*, 3(2), 2012.

[23] Gajanan V. Bhole , Prof. Prakash. R. Devale, "Approaches for Automated Web Service Discovery", *International Journal of Computer Science and Information Technologies*, Vol. 5, 2014.