Cluster based Approach for Service Discovery using Pattern Recognition

Arunachalam N[1], MohanaSundaram P[2], Santhan R[3], SarathKumar R[4]

[1]*Assistant Professor, Dept. of Information Technology, Sri Manakula Vinayagar Engineering College, Pondicherry-605104, India*

[234]*Dept. of Information Technology, Sri Manakula Vinayagar Engineering College, Pondicherry-605104, India*

[1]narunachalam85@gmail.com
[2]mohanasundaram.vpm@gmail.com
[3]santhnyogesh@gmail.com
[4]sarathkumar.rit@gmail.com

Abstract— **Web services that are appropriate to a user specific request are usually not considered in discovering the exact service since they are present without explicit related semantic descriptions. In our approach, we deal with the issue of service discovery provided non-explicit service description semantics that match a particular service request. We propose a system that involves semantic-based service categorization which is performed at the UDDI with a key for achieving the service categorization at functional level based on an ontology skeleton. Also, clustering is used for literally systemizing the web services based on functionality which is achieved by using analytic algorithm. An efficient matching for the relevant services is achieved by the enhancing the service request semantically and involves expanding the additional functionality (obtained from ontology) that are related for the requested service. The pattern recognition algorithm is used to select appropriate service from the cluster formation of related (grouped) web services**.

Keywords—web service; ontology ;clustering ; semantic categorization; pattern recogniion;

I. INTRODUCTION

The need for B2B Integration and the limitations of conventional middleware in B2B Integration raised the need for a novel technology – a need addressed by web services. Web services are components with specific functionalities that can be integrated into more complex real time distributed applications. The description of the services is stored in the registry which allows the designers to register a new service and allow service users to search for and locate services. The service registry contains web services which include predefined types that are generally specified by service providers. The UDDI registry specifications have two main goals with respect to service discovery: first, to support developers in finding information about services, so that they know how to write clients that interact with those services. Second, enable the dynamic binding, by allowing clients to query the registry and obtain references to services of interest. Service discovery can be done at also in design-time, by surfing the directory and identifying the most relevant services, and at run-time, for these dynamic binding techniques are being used. The proper discovery, configuration, and communication between devices and services with each other is enabled by Service discovery process. The relevant services which are discovered do not fulfill user needs all-time. The number of web services continues to grow and side by side the business environment keeps demanding newer applications that have to be rolled out according to very tight schedule. Relatively large number of web services and the distribution of similar services might be listed in different categories in the registry infrastructure which makes it difficult to find appropriate services. Therefore, rather than classifying the services based on their providers, they must be categorized on their functional semantics. Services categorized on their functional semantics will facilitate the organization of similar services together. Majority of the service description that exists so far are syntactic in nature. The same syntax might be used at different place for various purposes. Actually when a service is requested only small amount of services that are an exact syntactical match of the request is selected. And the selected services may perform different functions rather than the requested functions and the discovery process is limited by its dependence of human involvement for selecting right service based on its functionality. The description of web services should be described using WSDL in order to make the approach to be generic[4]. In this paper, we address a clustered approach for service discovery based on their functionality and providing similar service using pattern matching, thus satisfying user needs.

II. RELATED WORKS

The end-user have special interests on web service and its functionality, several logics and techniques were proposed by researchers for discovering the web service that satisfy the user needs. In such way, researchers proposed the service discovery process using the concepts of Dual Clustering, Service Matching and etc.

A. Service Matching

Similar web services identification process is now becoming increasingly an important issue to make sure the accomplishment of dynamically integrated Web-service-based applications. Ontology-based models are applied for the improvement of the searching capabilities in Agent Systems. Matchmaking troubles happen when a service is being requested and it comprises the distance calculation between the required service description and also from the service registry[1]. The problems in service and resource matching is

being keenly discussed at present as one of important new challenging job for the next generation of semantic discovery approaches for Web agents and Web services. Correct ontology-based matching tools to be used to effectively integrate Agent, Grid Services, and Web Service technologies with each other[1] .A categorization-based scheme is used to identify the identical Web services that could function on diverse domain ontologies. By using ontology instance categorization concept, the matching system states whether a given Web service is a possible replacement and then it also adapts itself by the process of enhancing with the recognized ontologies by means of the newly determined ontology instances[1].

B. Dual Clustering

Increasingly, service requestors look for the ability to seek for existing Web services in large Internet-based warehouse and the main objective is to retrieve services that match the user's requirements. With the repositories increasing in number of services and the challenges in rapid finding the accurate ones, the call for clustering associated services becomes evident to improve search engine results with a record of similar services for each hit. Clusters typically comprise terms in the same industry or knowledge area that share. Harvard Business School professor Michael Porter (1990) popularized the concept. Clusters are always dynamic. Spatial clustering has various applications. In most conventional clustering issues, the geometric attributes are taken into consideration for the similarity measurement. Users are generally concerned about the non-geometric attributes in many real applications[7]. The input data set is partitioned into numerous compacted regions in conventional spatial clustering, and data points which are comparable to one another in their non-geometric attributes may be scattered over various regions, hence making the corresponding objective complicated to reach. Dual clustering is remedy to this. Constraint domain specifies the application dependent and attributes present in the optimization domain are those intricate in the optimization of the objective function. The information in both domains is combined by the ICC algorithm and clustering algorithm on the optimization domain is iteratively performed[7].

III. PROPOSED SYSTEM

The previous works have the advantages like optimization of domains and replacement of the web services but they lack in the important properties such as the functionality based categorization and providing the appropriate service to the user. Our proposed work is to categorize the services based on their functionality and they are clustered based on their category along with related ontology concepts. The pattern recognition algorithm is used for retrieving the exact service that matches the user request.

A. Semantic Categorization

Here in our approach we begin with the semantic categorization of services in the UDDI where we use the ontology concepts. The semantic categorization is achieved by adding a user-defined tag in WSDL file, so for a particular search keyword we can make the services to fall under a given category. A single service can be made to appear in different categories by implementation of the ontology concepts that identifies the relationships[3]. This user defined tag is given by the service provider and it is based on the functionality of the service. Web service description vectors are built and the markups and index entries are removed. The web service vector development generally includes the parsing of the WSDL file forms part of the initial WSDL set and its parallel description and also the related parameters[11]. Web Service Vector Modification is being done by enhancing the service vectors with the concepts from the core ontology resolves issues related to synonyms and induces domain related concepts that provide the context[3][8]. First, the initial service vector is added with relevant concepts of ontology. The architecture design of our proposed system as follows.

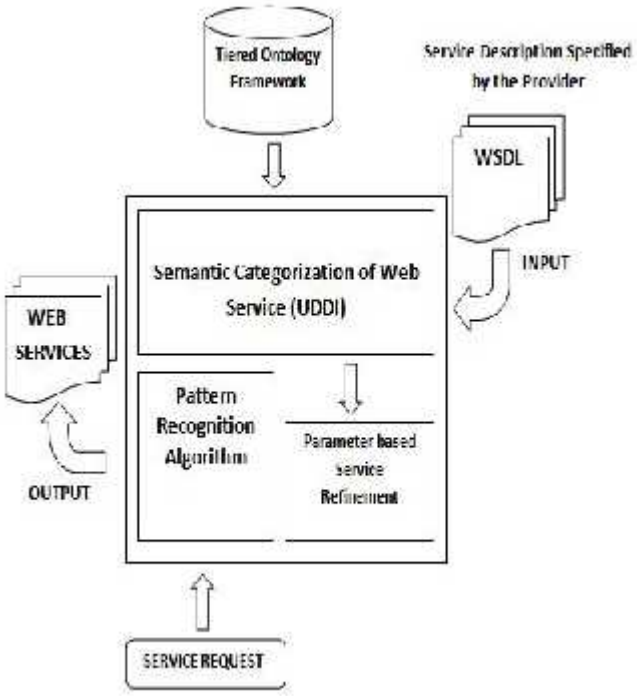

Fig.1 Architectural Design of the Proposed System

B. Clustering

The next step of this phase involves deleting irrelevant terms based on the ranking of semantic relationships among the terms. Grouping of functionally similar services together is performed by the process of clustering of service vectors. Clustering is actually done in hierarchical manner facilitates classification of all the services[6], such that each secondary cluster and the combinations of secondary clusters create a hierarchy—a structure that is more informative than the set of

clusters which are unstructured. The service categorization is performed manually and only service request enhancement is performed during runtime, thus increase in the timing delay will not be significant[10]. The relevant ontology concepts are associated to clusters during the creation. The association of concepts to each cluster help web service discovery by mapping to functional categories. A cluster is defined as $\Theta_i = c_j$ where, c_j is the corresponding ontology concept. The ontology concepts deliver semantic for web service categorization. We build a set which contains all concepts that exist in at least one service description and remove the concepts that makes again[11]. This is followed by locating the places of the remaining concepts in the concept hierarchy H_C. Each concept is checked for subsumes or subsumed relationship with the elements of the set. The resultant super concept is then mapped to the cluster[11]. The process of association of the ontology concept to the cluster extends semantic information in UDDI is done by the creation of tModels for the associated web services of the cluster within the registry. The relateOntologyCluster algorithm is given above.

ALGORITHM

relateOntologyCluster
Input: Web Service Description clusters
set $\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_n\}$,
Min. Term Frequency Threshold μ
Output: Modified UDDI tModels

1. begin
2. For each Web Service cluster Θ_i do
3. Retrieve modified Web Service vector $ws_m \in \Theta_i$ do
4. Calculate term frequency (t_j , x_j) where $t_j \subseteq \Theta_i$
5. if $x_j < \mu$
6. delete t_i
7. Map $t_j \approx c_j$
8. Traverse H_c for upper ontology concept C
8. if the term concept is subsumed by the
upper concept $c_j \subseteq C$
9. $C_\Theta = C$
10. else
11. $C_\Theta = c_i$
12. Map C_Θ to Θ_i
13. for each Web Service $w_k \in \Theta_i$
14. Update $tModel_m$ to include C_Θ
15. end for
16. end

C. Parameter-Based Service Refinement

This phase actually used to select service from the related group of services. Input, output, and the description, of web service help service refinement process through narrowing the set of appropriate services matching the service request[12]. The statistical associations are generally used to represent the relationship between the web service input and output parameters[12]. The parameters relationship pattern item set for all the web services within the cluster is being built. Then the corresponding WSDL document is processed to retrieve the relevant service parameters[12]. The weights are assigned for each of the parameters by user to refine the request. The ranking process is made more flexible by this. The binary values are actually assigned to the ranking parameters which is an important task. The association pattern mining phase generates a large number of association patterns and the patterns having unrelated information that will negatively influence the service discovery process have to be discarded.

D. Service Search

The next phase for our proposed system is to search the appropriate service based on the user request. The service has to be selected from the cluster containing number of services. For searching the appropriate service we use pattern recognition algorithm[9]. The search cluster is given below.

ALGORITHM

searchClusters(hN,S,FS_i)

- hN - a head node intended for the linked list containing of a collection of sibling clusters.
- S – the service to be searched in the cluster
- FS_i – feature section to be used for comparing similarity

1: hPtr = hN → Next;
2: σ = α;
3: while hPtr ≠ Null
4: if σ > Dist(hPtr → data,S|FS_i)
5: σ = Dist(hPtr → data,S|FS_i);
6: C_l = hPtr → data;
7: end if
8: hPtr = hPtr → Next;
9: end while
10: return C_l

The above algorithm provides the method for searchClusters, which has to iterate the linked list which is being headed by hN and the cluster having the minimum distance with S is also found.

- Dist(hPtr → data,S|FS_i) – distance between S and cluster
- hPtr → data – cluster is pointed using this

IV. CONCLUSION

In this paper, we provide a novel approach which deals with the service discovery and the approach addresses two of the major aspects namely: categorization of services based on their functional semantics and the clustering of the service using related ontology concepts. The pattern recognition algorithm is implemented here to identify the appropriate service from the cluster. Thus the proposed system satisfies the user needs by providing the appropriate service as requested by the user.

V. REFERENCES

[1] Anton Naumenko, Sergiy Nikitin, Vagan Terziyan, 2006, "Service matching in agent systems", Springer Science+Business Media, LLC.

[2] Hui Xiong, Junjie Wu, and Jian Chen, 2009, "K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective", Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 39, No. 2.

[3] M.A .Corella and P.Castells, 2006, "Semi-Automatic Semantic-Based Web Service Classification," Proc. Int'l Conf. Business Process Management Workshops (BPM '06).

[4] www.ibm.com/developerworks/tutorials/wesws511pt1/wesws511pt1-pdf.pdf

[5] Liang and Herman Lam, July 2008"Web Service Matching by Ontology Instance Categorization", IEEE International Conference, Services Computing.

[6] Jian Wen, Zhoujun Li and Xiaohua Hu, 2007, "Ontology Based Clustering for Improving Genomic IR", International Symposium on Computer-Based Medical Systems, 2007. CBMS.

[7] Cheng-Ru Lin, Ken-Hao Liu, and Ming-Syan Chen, 2005, "Dual Clustering: Integrating Data Clustering over Optimization and Constraint Domains", IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 5.

[8] Alireza Zohali, DR.Kamran Zamanfiar, 2005 – 2009, "Matching Model for Semantic Web Services Discovery", Journal of Theoretical and Applied Information Technology, JATIT.

[9] Liang-Jie Zhang, Shuxing Cheng, Carl K. Chang, and Qun Zhou, 2012, "A Pattern-Recognition-Based Algorithm and Case Study for Clustering and Selecting Business Services", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, Vol. 42, No. 1.

[10] Tsang-Hsiang Cheng and Chih-Ping Wei, March, "A Clustering-Based Approach for Integrating Document-Category Hierarchies", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, Vol. 38, No. 2.

[11] Aabhas V. Paliwal, Basit Shafiq, Jaideep Vaidya, Hui Xiong, and Nabil Adam, 2012, "Semantics-Based Automated Service Discovery", IEEE Transactions on Services Computing, Vol. 5, No. 2.

[12] Michael Forbes, Jim Lawrence, Yu Lei, Raghu N. Kacker and D. Richard Kuhn, 2008, "Refining the In-Parameter-Order Strategy for Constructing Covering Arrays", Journal of Research of the National Institute of Standards and Technology, Volume 113, Number 5.

[13] Fei You, Qingxi Hu, Yuan Yao, Gaochun Xu, Minglun Fang, 2009, "Study on Web Service Matching and Composition Based on Ontology", WRI World Congress on Computer Science and Information Engineering, Volume 4.

[14] Jie Liu, Jigui Sun and Shengsheng Wang, 2006, "Pattern Recognition: An overview", IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.6.