A Proteus Configuration to Respond to the Various Web Services

K.S.Kannan, C.Agjelia Lydia

Abstract— **The discovery of the Web services leads to the enhanced service categorization and service request process. The clustering for accurate web services classification based on service functionality validates the effectiveness and feasibility of the proposed approach. Already there are service discovery mechanisms which lacks in the elasticity and scalability of the web services. A new model, proteus generic query model for the discovery of operations is offered by heterogeneous services. The need for such a model is because it unifies the task of service discovery through abstractions, which allows for the technology-independent formulation of stages like service advertisements, queries, and query response. The query and its response documents may contain string values, numeric values, semantics, data types, qualifiers and sub property groups. These are used in the query evaluation for providing the response to the query. It is also called as similarity measures. Frequency rate mechanism helps in overcoming the problem of precision and recall in the previous keyword based matching mechanism. Correlation similarity adds advantage to the semantic matching and Minkowski distance are used to enhance the results.**

Index Terms— **Web service discovery,service discovery process, web service publish.**

I. INTRODUCTION

Discovery of patterns from the web is undertaken by utilizing the data mining techniques. Three different types of web mining are Web usage mining, Web content mining and Web structure mining. It succors in extracting server log information to analyze what users explore over internet.

Users seek various forms of data like textual, multimedia and image data. Node evaluation and structure connection of the website is done by graph theory to perform the web structure mining. This mining splits into hyperlinks that link the web pages and pattern extractions from hyperlinks. The tree-like page structures for illustrating HTML and XML tags are used for researching the document structure.

The process of putting together the knowledge from the web page content is known as the Web content mining. Applications that can be published and invoked throughout the web are executed by the Web services. In the service registry, the Service providers publish the web services with varied classifications. The Web Services description is maintained in registries and one such registry is the Universal Description, Discovery and Integration. The job of defining the interface is performed by the specification which helps in facilitating the web pages for Local-networked service connection over HTTP. The home network devices supply access to the content and services.

A series of related application functions named as the Web service in the internet is invoked using the programs. A recognized, computer-readable description of Web services is offered by WSDL. An interface records the groups of operations which can be accessed using the standard XML messaging by network is done in the latter. The needs of an application in the area od information is present in the description like the structures of messages , responses and the binding information. The Architecture of web service helps the business applications that are heterogeneous to operate together. The layers from the protocol stack can cause interoperability issues. There are some main actors in the process of service-oriented engineering.

Service provider is the actor which provides the services and delivers the publication of their narration to the respective brokers. The suitable structure to assist the available services for publication and discovery is made by these brokers. Service consumers are the frontend who receive these services. The Heterogeneity in service description causes the need of service requesters to supports the shaping of their queries separately from multiple models of services and their formats. The Heterogeneity in service discovery mechanisms has detailed information over the low-level areas which should be known to the service requester. The other issues are the Multi-dimensional query formation with evaluation and Technological volatility.

II. RELATED WORK

The related work overview focuses over the query languages and engines proposed and their limitations are analyzed. In [2], the advertisements and constraints of the services are matched based on the semantics. Similar web services are grouped together and get better service discovery. Addition of semantics to Web services mostly aims in automating the tasks that must be performed with services before or during messaging and communications. Based on various efforts in SWS and service-oriented computing communities (such as OWL-S and WSMO), the generally usual tasks are detection, concession, filtering, assortment, and invocation. The limitation of semantic annotations for WSDL and XML schema does not recommend annotating interfaces with nonfunctional properties.

In [3], an incorporated move toward the automated service discovery addresses two main characteristic that is associated to semantic-based service discovery: service categorization process the WSDL description of a service is assigned to its corresponding domain. Another approach is semantic-based service selection which uses ontology linking and latent

Manuscript received May 20, 2014.

K.S.KANNAN,Assistant Professor,Computer science and Engineering, NPR College of Engineering and Technology Tamilnadu, India

C.AGJELIA LYDIA,PG Student, Computer science and Engineering, NPR College of Engineering and Technology Tamilnadu, India

semantic indexing thus expanding the indexing process from only syntactical in order to a semantic intensity and Selection process of the WSDL entities are associated with pre-existing domain ontology concepts. To quantify the perfection to the maximum weight matching that is useful to evaluate the similarity of services. Thus the result shows that better performance than the original algorithms in both the service classification and query. It focuses only on semantics data rather than heterogeneous web services data.

III. SYSTEM ARCHITECTURE DESIGN

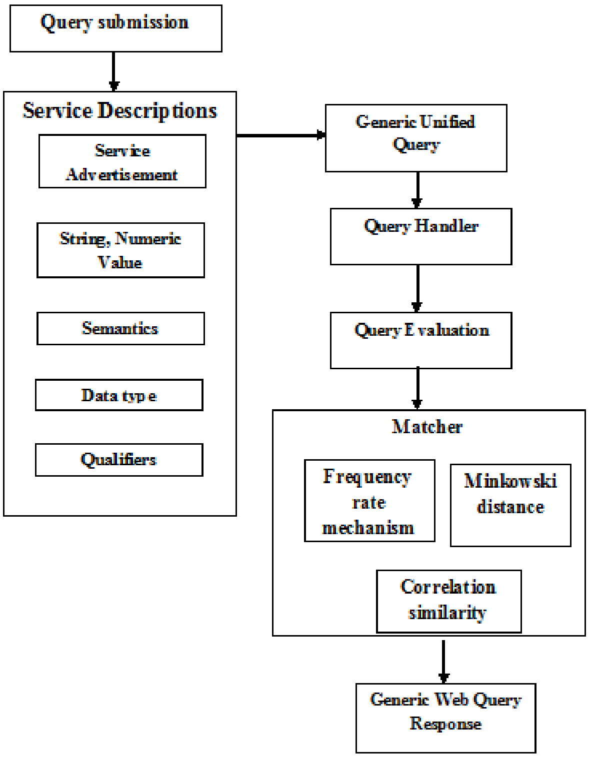

Fig 1: System Architecture

IV. QUERY EVALUATION METHODS

Proposal of the unified service discovery model called Proteus which provides the independent and appropriate abstraction for open set of properties in the heterogeneous services framework. Proteus framework involves the query and response documents by means of following generic structures as follows for

String value constraint:

The process for the String value input is done by the Keyword based matching mechanism. This mechanism checks using the following methods String equals, starts With, ends With and contains. Their query evaluation mechanisms are mostly limited and hence yields poor results in terms of precision and recall. So an alternate mechanism, frequency rate mechanisms is use to find the weights of attributes and ranking for improving the Search scenario. The frequency rate mechanism is calculated by,

Frequency=no of occurrence/total occurrence

Numeric value constraint:

To calculate for the numeric value of a service property, the one dimensional Euclidean distance(Ed) metric is used in its normalized form. Thus the degree of match between numeric value v and numeric value check c(v)=E is calculated by,

$$(c(v),v)=1-(1-Ed).\min_{\alpha,\beta}.\mathrm{Max}_{\alpha,\beta} \quad (1)$$

where $\alpha=|v-v_{min}|$, $\beta=|v-v_{max}|$,Ed is the indicator function of the range accepted values for v, $Ed=[v_{min},v_{max}]$.Euclidean distance is used for matching the numeric value constraints. Euclidean distance formula is calculated by,

$$d(a,b)=\sqrt{\sum_{1=1}^{m}(a_i-b_i)^2} \quad (2)$$

Certain drawbacks occur by using the distance method (2). They are sensitive to outliers and they have larger magnitudes which would create larger similarity scores. There is a high noise-to-signal ratio and unconstructive spikes in correlation are hard to found.

To overcome all the above stated drawbacks, the Minkowski distance is used. The Minkowski distance is a metric in the Euclidean space which can be considered as a generalization of both the Euclidean distance and the Manhattan distance. Rather than basic Euclidean distance, it produces better relevancy measures. The Minkowski formula is represented by,

$$d(a,b) =\left(\sum_{i=1}^{m}|a_i-b_i|^q\right)^{1/q} \quad (3)$$

Where q is a query. The advantages of distance method (3) provides a brief parametric reserve function that generalizes the reserve function and the user can also adapt the reserve function to suit the needs of the application by implementing the modification of the minkowski parameter. This results in better Accuracy with respect to outliers

Semantics value constraint :

In proteus, a semantic can be expressed through a textual values and an appropriate ontology concept. Semantic property p=(s,t) and . Thus the degree of match between a semantic check $c(p)=(s_r,t_r)$ calcualted by,

$$d(c(p),p) = max\left(x_s.d(s_r,s),x_t.d(t_r,t)\right)$$

Where $x_s+x_t>0$, $d(s_r,s)\in[0,1]$ measure the similarity between textual values s_r and s and $d(t_r,t)\in[0,1]$ enumerates the similarity between ontology values t_r and t. For finding semantic similarity measurements, it computes $d(s_r,s)$ by using Vector Space model. Generally, it uses search engines based on natural language. The core concept is uncomplicated. A document is broke up into keywords. All keywords include measurement in a n-dimension vector space. Thus a document can be seen as a vector within this "term space". The arithmetical method is used to calculate how similar two documents are to each other and correspondingly match a given query. A common method is to evaluate a cosine value for them and express the result as a percentage rating. This method produces very good results for natural language but it is not limited to this field alone.

Weights are calculated with the use of the Term Frequency-Inverse Document Frequency (TF-IDF). The degree of match between two given term documents having TF-IDF vectors, which corresponds to the textual description of a semantics check , and its respective semantics feature is calculated with the use of the correlation coefficient and dice co-efficient. In semantics scenario the correlation similarity is introduced effectively for Semantic data matching. Comparison among two items s_r *and* s is calculated by Pearson-r correlation $d(s_r, s)$. The correlation computation accurate by isolate the co-rated cases. Let the set of users who both rated s_r and s are denoted by U then the correlation similarity is calculated by,

$$d(s_r, s) = corr_{s_r,s} = \frac{\sum_{u\in U}(R_{u,s_r} - \overline{R_{s_r}})(R_{u,s} - \overline{R_s})}{\sqrt{\sum_{u\in U}(R_{u,s_r} - \overline{R_{s_r}})^2}\sqrt{\sum_{u\in U}(R_{u,s} - \overline{R_s})^2}} \quad (4)$$

Here R_{u,s_r} denotes the rating of user u on item i, $\overline{R_s}$is the average rating of the s-th item. Then the distance between the two ontology concept is in the ontology graph to calculate $d(t_r, t)$. Dice coefficient of similarity defined as S_{dice} calculated by

$S_{dice} = 2.\frac{|X \cap Y|}{(|X| + |Y|)}$ Where x and y are the is-a links sets between the root of ontology and concepts. The direction and hierarchy semantics between the advertised concept t_r and requested one t by introducing a numeric variable Q which quantifies the widely used semantic relations: exact(Q=3);plug in(Q=2) ;subsume (Q=1) ;fail (Q=0).Thus the similarity assessment between the requested ontology concept t_r in a semantic check and corresponding one in the semantic feature t is performed by the following equation:

$$d(t_r, t) = \frac{1}{3}.[Q + (3 - Q).S_{dice}]$$

Data type constraint :

Data type matcher plug-ins is used by the Query Processor to calculate the degree of match among requested and advertised data types. They are specifically involved in the matchmaking of requirements towards the input/output messages of a service, where the requester may have specified the desired data type for each constituent message element and the resource of a grid service, where the desired data types of the constituent resource properties may be specified.

Qualifiers constraint:

Qualifier Constraint is used by the Query Processor to calculate the degree of match among requested and advertised types. They are specifically involved in the matchmaking of requirements towards the input/output messages of a service where the desired types of the constituent resource properties may be specified.

Sub properties constraint:

Besides qualifiers, a service property may contain grouped sub properties of the same type.

V. Conclusion

Each and every Generic query model has distinct techniques to find the similarity measures and to calculate the relevancy between the entities. Optimized query relevancy and its performance evaluation is implemented by the proteus crawler for the proper translation of heterogeneous service descriptions into Service advertisements. Parsing and match-making mechanisms used to bring out the optimal query relevancy from the heterogeneous data sources. Performance is evaluated in terms of similarity measures by means of framing generic framework model and to meet the functional and non-functional requirements.

References

[1] Michael Pantazoglou and Aphrodite Tsalgatidou "A Generic Query Model for the Unified Discovery of Heterogeneous Services ",IEEE Transactions On Services Computing, Vol. 6, No. 2, April-june 2013.

[2] J. Farrell and H. Lausen, "Semantic Annotations for WSDL and XML Schema," http://www.w3.org/TR/sawsdl/, Aug. 2007.

[3] Aabhas V. Paliwal, Student Member, IEEE, Basit Shafiq, Member, IEEE,Jaideep Vaidya, Member, IEEE, Hui Xiong, Senior Member, IEEE, and Nabil Adam, Senior Member, IEEE "Semantics-Based Automated Service Discovery" IEEE Transactions On Services Computing, Vol. 5, No. 2, April-June 2012.

[4] Fang fang Liu, Lei Wang, and Jie Yu "Context-aware similarity search of web service" 2012 IEEE International Conference on Information Science and Technology Wuhan, Hubei, China; March 23-25, 2012.

[5] Dimitrios Skoutas, Dimitris Sacharidis, Alkis Simitsis, and Timos Sellis, Fellow, IEEE "Ranking and Clustering Web Services Using Multicriteria Dominance Relationships" IEEE Transactions On Services Computing, Vol. 3, No. 3, July-September 2010.

K.S.Kannan was born in Dindigul, Tamil Nadu (TN), India, in 1979. He received the Bachelor of Technology(B.Tech.)degree from the Pondicherry University (PU), Pondicherry, TN, India, in 2003 and the Master of Engineering (M.E.) degree from the Annamalai University, Annamalai Nagar, TN, India, in 2005,He is currently pursuing the Ph.D. degree with the Department of Information & Communication Engineering in Anna University,Chennai. His research interests include GeoSpatial Database,Distributed Systems.

C.Agjelia Lydia was born in Dindigul, Tamil Nadu (TN), India, in 1991. She received the Bachelor of Engineering (B.E.) degree from the Anna University (PU), Chennai, TN, India, in 2012 . she is currently pursuing the Master of Engineering (M.E) degree from the Anna University, Chennai , TN, India. Her research interests include Web mining , Network Security.