Efficient Web Service Discovery and Selection Model

Saravanan Chandrasekaran
Research scholar
Dept. of CSE
Annamalai University
Tamil Nadu, India
availableathere@gmail.com

Vijay Bhanu Srinivasan
Assistant Professor
Dept. of CSE
Annamalai University
Tamil Nadu, India
vbhanu22@yahoo.in

Latha Parthiban
Assistant Professor
Department of Computer Science
Pondicherry University
Pondicherry, India
lathaparthiban@yahoo.com

***Abstract*—**Selection of an optimal web service is a challenging task due to the uncertainty of Quality of Service, which is the deciding factor to identify the accurate web service. Several discovery mechanisms have proposed but most of the research work does not consider the non-functional characteristics called Quality of service. The proposed model for web service selection combines two techniques. First, with Skyline method reduce the search space by filtering the redundant service and secondly to calculate the Relevancy function to normalize the skyline services. The experimental results show that the proposed technique outperforms the existing method.

Keywords-*Web service, Quality of service, Skyline.*

*****

I. INTRODUCTION

In the recent times, web services have developed as the most innovative software on the internet; and many researchers have contributed to make it more open and generally utilized. Web service is a software product framework intended for interoperable machine-to-machine interaction over a system. Web services are an implementation of the service oriented architecture (SOA) that permits loosely-coupled, reusable, and composable services.

The significant issue in SOA is to focus on web service selection mechanism which to choose the appropriate candidate service to the user. The SOA comprises of service registry, service provider, and service consumer. The service provider registers the new service in the service registry and the service consumer uses the service available in the repository. Several web service providers are available in the market satisfying the same functional characteristics. There should a mechanism to find the appropriate service from the number of candidate service [1].

Fig. 1 depicts the web service selection process. In the first step of the selection process, the user specifies the service constraints.

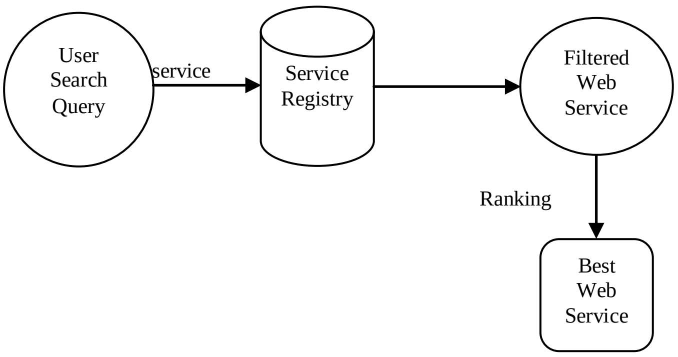

Figure 1. Generalized web service selection

In the second step; based on the user requirements that's been met by the web services are discovered that are within the service registry, which holds the information about the available concrete web services. At this stage, there is a chance of more than one candidate service satisfying the basic functional requirements to be be discovered, however different Quality of Service (QoS) attribute values may be offered, i.e. Execution Time, Reputation, Availability, Cost, etc. For selection purpose, Multi Criteria Decision Making (MCDM) approaches are applied to find the best services. This is achieved by comparing and ranking services according to the client requirement. Several researchers have proposed algorithms on QoS attribute to find the optimal web service.

The paper proposes a new technique for Web service selection. To achieve this; first, we exploit the Skyline [3] method to reduce the search space of the candidate service. More specifically, skyline allows to eliminate the dominated elements and to keep only the dominant ones. It is described in [3] that, when a optimal solution exists within a population, it is necessarily amongst its dominant elements. Because of its low computation effort, skyline is considered as a first step towards web service selection.

The introduction of skyline is more than a decade ago [2], the skyline operator has played a vital role in the database research community. Skyline queries introduce a retrieval model that selects all optimal choices for any monotone preference function. Second, the skyline services are ranked using Relevance function [6] to find the appropriate service for the user. The rest of this paper is organized as follows. Section II discusses the related work. The Skyline method and Web service ranking mechanism is presented in Section III. Results are discussed in Section IV, and section V discusses about the conclusion.

II. RELATED WORKS

Web services are utilized as a part of the advancement of various web applications in the current scenario. The existing searching techniques for web services mostly concentrate on keyword matching. But these methodologies and scheme are now being replaced by newer methods because of the lack of efficient search in keyword matching. Moreover it does not yield effective results according to the user's requirements. A multiobjective query optimization algorithm is discussed in [4] which maintain the optimum cost-delay trade-off.

A list of web service shortlisted from the large space to the user is the necessary criteria. There are different web services which offer comparative functionalities, yet may have distinctive non-functional or quality of service parameters. Most of the research is on ranking the web service, and several methods are proposed for ranking the web service. The matrix method approach is the common one [7]. This strategy takes the QoS parameters of different web services as contribution to a matrix, standardizes it and determines a Web Service Relevancy Function (RF) based on which a rank is given for the services.

In the context based method [8], the WSDLs are scrutinized semantically to separate a more accurate outcome to the client's query. The proximity of the context of comparable web services are resolved to give the finalized rank list. But choosing the specific sites for context extraction is the disadvantage of this method. The quality enabled approach [9] makes use of a quality constraint tree (QCT) which takes the functionally similar web services as input. The traversal algorithm does filtering, scaling and ranking to arrive at the result.

Skyline is developed to solve the problem by selecting the best services from the available services; proposing the skyline as the best service model as well as novel algorithm to maintain dynamic skyline [10]. As in [11] the problem is to focus on exactly to optimize the response time of frequent or near to frequent skyline queries raised against the static and the update intensive dataset. The SkyQUD framework is proposed in [12] to efficiently calculate the probability of datasets in uncertain dimensions. The constant change and uncertainty of data makes the query process difficult and would require more computations, making it crucial to have an effective skyline query process in terms of time and space over uncertain data streams. In [13] discusses Efficient Probabilistic Skyline Update (EPSU), an approach in which, using a new data structure by augmenting the R-tree structure is used to rank the web service. As in [14] a different method called ProMiSH (Projection and Multi Scale Hashing) that uses random projection and hash-based index structures, and achieves high scalability and speedup are used to predict the suitable web service to the client.

III. QoS MODEL

A. *QoS criteria for a service*

QoS is the overall performance of the particular system. The criteria for various domains may be different. In our model; to be more generic and precise, we consider six criterias: Response Time, Throughput, Reliability, Best Practices, Documentation and Cost.

- Response Time (RT): The RT is the time length (turn round time) between a demand being sent from client, and the outcome received.
- Throughput (TP): The number of request satisfied by the web service for the particular time.
- Reliability: The probability that the web service do not fail for a specified interval time.
- Best Practices (BP): The document is compliance with the necessary standard.
- Documentation (Doc.): Describes whether the WSDL is potential to advertise the service.
- Cost: Money incurred to generate the service requested by client.

B. *Skyline services*

Intuitively, a Skyline function selects the best points or the most interesting points in all dimensions. We explore the dominance relationship between the web services and their quality of service factors. Here the number of services is reduced by the dominated service in the same class. Determining Skyline services of a class of service requires pair-wise comparisons of QoS vectors of web services. This procedure can be expensive in calculation time if the quantity of administrations is imperative, but it reduces the searching space to a significant level. More efficient algorithms have been proposed for calculating Skyline. The Branch and bound skyline algorithm [16] is used for large amount of dataset and the same algorithm is used to reduce the searching space. This data filtering is used to put the sorted services in central memory and those rejected in a buffer file.

1) *skyline operation*

As in [2] proposes few approaches about the Skyline Queries implementation and about the existing (object-oriented, relational or object relational) database system's extension with a new logical operator which we refer to as the Skyline operator. The Skyline operator encapsulates the implementation of the SKYLINE OF clause. The Skyline operator embodies the execution of the SKYLINE OF condition. The implementation of other operators (e.g., join) need not be changed. If the query has an ORDER BY clause; in accordance to the semantics of Skyline queries, the Skyline operator is typically executed after scan, join, and group-by operators and before a final sort operator.

Considering the web service as points and the Quality of service as dimension; with a given set of points of a d-dimensional space, a Skyline function chooses points which are dominated by other points. A point a_i dominates another point a_j; if a_i is lower or equal to a_j in all dimensions and strictly less in at least one dimension. In this n-dimensional space, all points which are not dominated by other points combined to a set named skyline. The skyline query optimizes to find the best data for the user. The application of skyline queries is applied in the field of multi-criteria decision making problem.

The skyline operator [5] is based on the relation of dominance, defined as follows: The relation of dominance is denoted as ($\prec$) : Let us consider the functionally equivalent web services(WS) are defined as $\{WS_1, WS_2,\ldots,WS_m\}$ and the set of QoS properties (P) for the web services are denoted as

ISSN: 2454-4248
01 – 05

$\{P_1, P_2, \dots, P_j\}$. Here, we say WS_1 dominates WS_2 as $WS_1 \prec WS_2$ if WS_1 is better or equal to WS_2 for all QoS properties and strictly better at least for one QoS property. The skyline query displays all the web services and it is not dominated by other services in the dataset.

2) Normalization using Relevancy function

The QoS properties of a web service vary in measurement unit based upon the metric. The QoS properties of a web service should be normalized to a common metric. The Relevancy function (RF) computes the relevancy value and then ranks the web services [15]. Among the functionally equivalent web service, there will be a service with the highest relevancy value and is treated as most suitable for that web service request.

The RF values are calculated for all the filtered services. It is assumed that RF values are computed for Web services that are in the same domain. A Web service with the highest calculated RF value is considered to be the most desirable and relevant to a client based on preferences. Using j criteria for evaluating a given Web service, we obtain the following RF matrix in which each column represents QoS property P_j and the row represents a single Web service WS_m. The RF matrix is represented as in (1).

$$E = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} & \cdots & P_{1,j} \\ P_{2,1} & P_{2,2} & P_{2,3} & \cdots & P_{2,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{m,1} & P_{m,2} & P_{m,3} & \dots & P_{m,j} \end{bmatrix} (1)$$

Due to the fact that QoS properties vary in units and magnitude, E $(P_{i,j})$ values must be normalized to be able to perform RF computations and perform QoS-based ranking. Normalization gives a more uniform appropriation of QoS measurements that have distinctive units. Normalization allows fine-tuning their QoS search criteria in an effective way and also providing to associate the weight with the QoS properties. In order to calculate $RF(WS_m)$, we need the maximum normalized value for each P_j column. . Let K be an array where $K = \{K_1, K_2, K_3, \dots K_m\}$ with $1 \le m \le i$ such that:

$$K(j)=\textstyle\sum_1^i P_{m,j} \qquad (2)$$

Where $P_{m,j}$in (2) represents the actual value from the RF matrix in (1). Each element in the RF matrix is compared with the largest QoS value in the same column based on the following equation:

$$H_{m,j}=\frac{P_{m,j}}{\max(K(j))} \; (3)$$

In Equation (3), the difference of $P_{m,j}$from the largest normalized value in the corresponding QoS property group is denoted as $H_{m,j}$. The preference of the user on the QoS property may vary on their perspective. Taking into account of the diverse conditions, there is an evident need to weight every QoS property with respect to the significance or extent that it invests after positioning Web services based on QoS parameters.

In this way, we have to characterize an array that represents to the weights commitment for each P_j where $W = \{W_1, W_2, W_3, \dots, W_j\}$. Every weight in this array represents to the level of significance or weight factor related with a particular QoS property. The value of weight is in fraction and range from 0 to 1. In addition, all weights must add up to 1. Every weight is corresponding to the significance of a specific QoS parameter to the general Web service relevancy ranking. The maximum weight of a particular parameter, the more vital that parameter is to the client. The weights are acquired from the client by means of a client interface. Assigning different weights to equation (3) results as:

$$H_{i,j}=W_j\left[\frac{q_{i,j}}{\max(N(j))}\right](4)$$

Applying (4), in each QoS value, we get a new weighted matrix defined in (5).

$$E' = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} & \cdots & H_{1,j} \\ H_{2,1} & H_{2,2} & H_{2,3} & \cdots & H_{2,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H_{m,1} & H_{m,2} & H_{m,3} & \dots & H_{m,j} \end{bmatrix} (5)$$

Once every Web service QoS value is compared with its corresponding group of other QoS values in the same group, we can calculate the RF for each Web service as shown below in (6).

$$RF(WS_m)=\textstyle\sum_{i=1}^{N} H_{m,j} \qquad (6)$$

Where, N denotes the number of Web services from a set. To show how RF works, we will consider a straightforward case in which a customer allots weights to QoS properties discussed earlier as follows: W1 = 0.1, W2 = 0.8, W3 = 0, W4 = 0, W5 = 0, and W6 = 0.1. From the weights assigned, plainly the weight that speaks to the most essential QoS property to this customer is W2. The significance level assigned to each QoS parameter shifts since QoS properties fluctuate in units. Because of the fact that each QoS property selected by clients has an associated unit that is different from other properties, it is mandate to clear that every weight represents a different degree of significance which must be optimized. For instance, a customer that sets all weights to zero with the exception of cost demonstrates that RF ought to limit cost since it speaks to 100% essential to the customer. Therefore, RF calculation will yield the least expensive Web service for this request.

IV. RESULTS AND DISCUSSIONS

The data used is based on the web service implementation [17]. The resultant web service is from the same domain which shares the same functionality of phone service. The average QoS values for the resultant web service are displayed in Table 1. The skyline approach is introduced to reduce the search space of web service. Keeping in mind the end goal to locate the most appropriate Web service, it is vital to streamline the qualities for each QoS parameter. For example, having higher likelihood for accessibility rate is ideal than having a Web service with lower likelihood for accessibility. In this case, RF will maximize accessibility.

TABLE 1. QOS METRICS FOR VARIOUS AVAILABLE PHONE WEB SERVICES

Service Provider	RT (ms)	TP (hits/ sec)	Reliability (%)	BP (%)	Doc. (%)	Cost ($/req-uest)
DOTSGeoPhone	126.2	12.3	78.7	80	86	1.4
Phone	150.45	7.4	82.1	82	37	1.1
DOTSPhoneAppend	140.5	0.7	70.2	80	90	1
PhoneVerify	131	1.6	65.9	72	41	2
PhoneNotify	437.62	1	68.4	69	93	6
PhoneService	133	1.4	64.7	82	10	3
Phonebook	464	3.1	43.2	80	2	5

A comparison is done about the web services based on the parameter Relevancy Function (RF). The tabulation of the comparison is given in Table 2. Applying RF in (6) without any associated weights, the following results are obtained:

TABLE 2. RESULTS OF RF WITHOUT WEIGHTS

Sl. No.	Service Provider	RF
1	DOTSGeoPhone	4.125
2	Phone	3.334
3	DOTSPhoneAppend	3.08
4	PhoneVerify	2.528
5	PhoneNotify	3.696
6	PhoneService	2.3
7	Phonebook	2.7

From the Table 2 we can interpret that, web services with the highest RF value proves to be the one that has the best QoS. For this example, RF determines that the best Web service without having any dependency on any specific QoS parameter (i.e. keyword-based search) is Web service number one.

To demonstrate the effectiveness of RF ranking technique and how it outperforms other discovery methods that merely depend on keyword-based technique, we consider three test scenarios in which each scenario represents a different combination of QoS requirements.

A Test is conducted by giving more importance to Response Time (RT). The result shows the web service number 1 has the highest RF and having the response time of 126.2ms. Another test conducted emphasizing on Throughput (TP), it results that again the web service number 1 dominates the other service because it shows 12.3hits/second. When running RF with more emphasis on the maximum TP, it ranks the best web service relevant to the TP.

A Test was conducted with varying the weights for RT and cost. In this test, response time has given more weightage than cost which yields web service number 3 is the best service. The graph plotted against the Execution Time and No. of Web Services in Fig. 2 shows that, as the number of web services increases, the execution time with Skyline is lesser when compared to the methods without the Skyline. Execution time is increased when the search space is gradually raised. The proposed model with skyline Search shows better results. Due to this the overall Execution time is decreased for the client request.

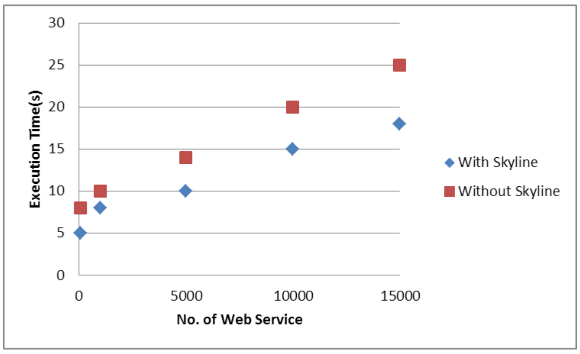

Figure 2. Comparison of the Execution time

Fig. 3 shows how efficient are the technique with Skyline when compared to the one without Skyline. From Fig.3 we can see that as the Web Services keeps increasing the efficiency Without Skyline is drastically decreasing, while With Skyline the efficient is above 85%, and as the Web Services keeps increasing, the efficiency gradually increases to 90% and proceeds further above it. For all the ranges of web services the proposed method results in better efficiency when compared to the conventional scheme.

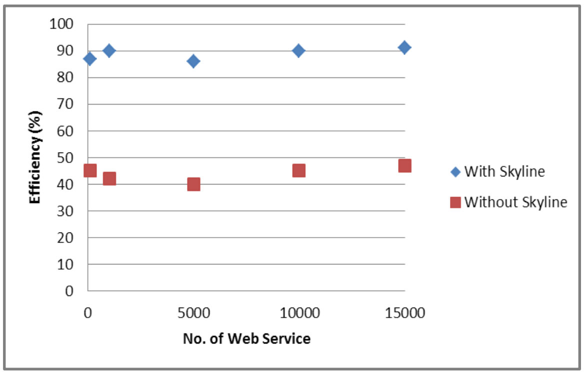

Figure 3. Comparison of Efficiency

V. CONCLUSION

The web service ranking system is proposed by combing both the skyline algorithm and relevancy function. With the inclusion of skyline, the large available web services are reduced to certain level and with the limited number of web service the relevancy function is implemented. The overall computational time is reduced by introducing the skyline algorithm with relevancy calculation. The inclusion of the QoS parameter Best practices and Documentation helps to find the best web services which are following the web service standard. The proposed method has demonstrated the adequacy of consolidating QoS parameters as a component of the search criteria and in distinguishing Web services from one another during the discovery process. Moreover, the capacity to discriminate on selecting appropriate Web services depends on the customer's ability to distinguish proper QoS parameters. The results prove that the skyline based relevancy function shows better results in discovering the web services. Our future work is to initially cluster the web service based on the domain so that the overall computational time is reduced.

References

[1] M. Moghaddam, and J. G. Davis, “Service selection in web service composition: a comparative review of existing approaches”, Web Services Foundations, pp. 321-346, 2013.

[2] S.Borzsonyi, D.Kossmann, K.Stocker, “The skyline operator”, In: Proc. IEEE Int’l Conf. on Data Engineering, pp. 421–430, 2001.

[3] M. Alrifai , D. Skoutas, T. Risse ”Selecting skyline services for qos- based web service composition” ACM WWW, pp. 11-20, 2010.

[4] C. H. Papadimitrious and M. Yannakakis, “Multiobjective query optimization” in Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 52–59, 2001.

[5] W.Serrai ,A.Abdelli , L.Mokdad and Y.Hammal, “An efficient approach for web service selection” , IEEE Symposium on Computers and Communication (ISCC) , pp. 167-172, 2016.

[6] M.Suchithra ,M.Ramakrishnan, “Efficient discovery and ranking of web services using nonfunctionalqos requirements for smart grid applications”,ProcediaTechnology,SMART GRID TECHNOLOGIES, Vol. 21, Pages 82–87, 2015.

[7] R.Manoharan, A. Archana and SiddhikaNitinCowlagi, “Hybrid web services ranking algorithm”, IJCSI International Journal of Computer Science issues, vol. 8, issue 3, No. 2, pp. 452-460, 2011.

[8] A.segev ,E.toch, “Context based matching and ranking of web services for composition”, IEEE Transactions on Services Computing, Vol.2, No.3, pp. 210-222, 2009.

[9] D. D’Mello and V.S. Ananthanarayana, “Quality driven web service selection and ranking”, Fifth International Conference on Information Technology: New Generations, pp. 1175-1176, 2008.

[10] J.Wu ,L.Chen and T.Liang, “Selecting dynamic skyline services for qos-based service composition”, Appl. Math. Inf. Sci. 8, No. 5, 2579-2588, 2014.

[11] R.D. Kulkarni, B.F. Momin, “Skyline computation for frequent queries in update intensive environment”, Journal of King Saud University - Computer and Information Sciences, Vol. 28, Issue 4, pp. 447–456, 2016.

[12] N.HusnaM.Saada, ,H.Ibrahima, , A.A.Alwanb, , F.Sidia, R.Yaakoba, “A framework for evaluating skyline query over uncertain autonomous databases”, Procedia Computer Science, Vol. 29, pp. 1546–1556, 2014.

[13] C.Liu, S.Tang, “An effective probabilistic skyline query process on uncertain data streams”, Procedia Computer Science, Vol. 63, pp. 40–47, 2015.

[14] V.Singh, B.Zong, A. K. Singh, “Nearest keyword set search in multi-dimensional datasets”, IEEE Transactions on Knowledge and Data Engineering , Vol. 28, Issue: 3, pp. 741 – 755, 2016.

[15] E.Al-Masri, Q.H. Mahmoud, “Qos-based discovery and ranking of web services” , 16th International Conference on Computer Communications and Networks, pp. 529 – 534, 2007.

[16] D. Papadias, Y.Tao, G. Fu, B. Seeger, “An optimal and progressive algorithm for skyline queries”, Proceedings of the 2003 ACM SIGMOD international conference on Management of data, pp. 467-478, 2003.

[17] http://www.uoguelph.ca