A Novel Framework for Semantic Discovery of Web Services using Integrated Semantic Model

Shailja Sharma, Jagdeep Singh Lather, Mayank Dave

***Abstract* — Semantic web technology plays a very critical role in the automatic web service discovery by assigning formal semantics to the service descriptions. Practically, it is not feasible to explicitly annotate the formal semantics to millions of existing services. Further, in user context, the request formation for services in semantic web is a complex process as it requires the user to be technically aware of the underlying technologies of the web services, discovery frameworks, description languages and implementation details. In this paper, we propose a semantic framework that enables Web service discovery based on the combination of semantic and syntax information contained in the service profiles. This novel approach for automatic discovery of Web services employs measures of semantic relatedness, Natural Language Processing techniques and information retrieval based statistical models to match a user request. Additionally, we present an efficient semantic matching technique to compute the intra service semantic similarity scores which further facilitates semantic ranking of services. The efficiency of the proposed approach has been demonstrated through experimental evaluations which clearly show that high degree of automation can be achieved with high precision.The results have been further authenticated by providing comparisons with other Information Retrieval based methods.**

Keywords: Semantic Web Service Discovery, Measures of Semantic Relatedness, Machine learning, Text Mining, OWL-S.

I. INTRODUCTION

Web services provide standard mechanisms for integration of distributed application over heterogeneous platforms [1]. Major challenge in the current service oriented architecture is to have automatic discovery process for the desirable services.

Manuscript received March 26, 2015, revised June 10, 2015.
Shailja Sharma is a Ph.D student in the Department of Computer Applications at National Institute of Technology, Kurukshetra, India (email: shailjakaushik@gmail.com)
Jagdeep Singh Lather is working as a Professor at Department of Electrical Engineering at National Institute of Technology, Kurukshetra (email: jslather@gmail.com)
Mayank Dave is working as Professor at Department of Computer Engineering at National Institute of Technology, Kurukshetra, India (email: mdave67@gmail.com)

The existing service standards like Extensible Markup Language (XML) [2], Simple Object Access Protocol (SOAP) [3], Web Service Description Language (WSDL) [4], Universal discovery and description Integration (UDDI) [5] are confined to syntax based matching scheme, where matching of service profiles is purely syntactic. This type of matching considers only those services whose syntactic descriptions exactly match the query keywords irrespective of semantic relatedness between the terms. Therefore, the outcome is either no results or a list of irrelevant services. Due to usage of different keywords, several services having semantically similar terminology are excluded from the result set, although they are potentially good candidates for the user's request. Therefore, human intervention is required in the service discovery process to filter the relevant services out of the resultant list. The numerous approaches for web service discovery range from Information Retrieval based similarity measures to semantic logic-based inference rules [6].

The current semantic approaches for web services discovery presume the services to be described in semantic description languages like OWL-S [7], WSMO [8], WSDL-S [9], SAWSDL [10], *etc*. Although, the semantic web technology is a promising way towards the realization of automatic discovery of web services but practical limitations of the semantic based approaches is that it is not possible to expect all service requestors and service providers to have same understanding of context and to use same ontological concepts.

Frameworks like OWL-S [7], WSMO [8] and WSDL-S [9] assume request as a web service and ask the user to express the request in a formal specified language which requires that web service users be friendly with these technologies and frameworks. Existing Web Services, whose descriptions are written in syntax based languages like WSDL, do not have explicitly associated semantics in their descripttions. Further, it is not feasible to have semantic tagged descriptions for all the new services. Thus, it can be concluded that annotating semantics to the existing services is time consuming, cumbersome and nearly impractical.

In order to address some of the cited limitations of existing approaches, this paper proposes an easy to use semantic framework for the discovery of services described in Web Service Description Language. The proposed approach generates concrete similarity score which facilitates the ranking of existing services. Our framework supports formulation of easy user request, keeping the user technical expertise independent of the technical knowhow of the services. Service provider's comments written in natural language as documentation have been utilized to match the user requirements in a better way. The proposed solution for semantic web service discovery integrates the syntactic as well as semantic information of a web service and tries to utilize the hidden semantics of the existing services. The rest of the paper is divided into following sections: section 2 provides the literature review and section 3 illustrates the proposed framework for describing new semantic matching algorithm; evaluations and results have been presented in section 4 and finally conclusion and future scope are covered in section 5.

II. RELATED WORK

It has been analyzed that most of the existing service discovery approaches adhered strictly to single service description languages and standards like DAML-S, OWL-S, WSMO, WSDL, SAWSDL, *etc*. These approaches vary on I/O matching vs. IOPE matching. DAML-S based approach [17] and OWL-S based approaches [12, 18, 19, 20, 21 and 22] and SAWSDL based approach [14] perform Input-Output (IO) matching on service profiles whereas approaches in [13, 23] perform IOPE matching. Research has been focused on addition of the semantics into the frameworks like WSDL using SAWSDL and WSDL-S. In [24], a WSDL-S based discovery technique over federated registries using the METEOR-S infrastructure has been proposed.

Some hybrid approaches [12, 13, 15 and 14] have also been proposed which considers semantic as well as syntactic description of the services. OWLS-MX [12] is an OWL-S based hybrid matchmaker that gives semantic as well as hybrid degree of matching.

In comparison to input and output (I/O) parameter matching by Klusch *et al.*[12], ITL based LARKS [15] perform IOPE matching. LARKS do not supports logical subsumes and hybrid nearest neighbor and has never been evaluated experimentally. Similar to OWL-MX, WSMO-MX [13] is also a hybrid matchmaker but it matches IOPE's of profiles instead of (I/O) and is based on WSMOframework. SAWSDL-iMatcher [14] annotates semantics to the existing profiles and supports user customizable matching strategies according to different application requirements. One of the major shortcomings of OWL-S based approaches is that it does not support mapping. Compared to OWL-S, WSMO is capable of modeling mediation for handling ontology heterogeneities in ontologies. Further, it has been observed that hybrid approaches based on logical reasoning are computationally expensive.

Among the various data mining based approaches, majority of them use classification and clustering to find semantic similarity between the services [25, 26, and 27]. SWSC [25] method uses Jaccard coefficient and hierarchical agglomerative clustering whereas in comparison Wen *et. al* [27] has modified the K-Mediod clustering mechanism to sort out the problem of web service discovery.

Batra and Bawa [26] propose a classification based approach that uses the Normalized semantic score MSR for semantic web service discovery.

In [28], García *et al.* use SPARQL-based repository filter for improving the semantic web service discovery.

III. SEMANTIC MATCHING APPROACH FOR WEB SERVICE DESCRIPTIONS

The proposed discovery framework aims to provide efficient discovery of services by combining the semantic and syntactic information from the service profiles. The complete discovery framework and algorithm is depicted in Figure 1 and Figure 2. Initially, the services are parsed using the text miner. This involves removal of markup(s), translation of upper case characters into lower case, punctuation and white space removal, followed by the stop word removal, *etc*. The textual documentation, arguments, operations and service name from the service profiles are extracted out. After preprocessing, the statistical weights are assigned to the terms using four different widely used weighing schemes of information retrieval i.e. TFIDF scheme, Binary scheme, Term Occurrence and Term Frequency. Further, a semantic relatedness matrix is generated using WordNet based measure of semantic relatedness. In next step, the syntactic vectors are transformed into the semantically enriched service vectors through semantic integration engine. The query is also transformed to the semantic query vector using the semantic integration engine.

The similarity match engine calculates the $semantic\_similarity$ between the semantic query vector $Query$ with the semantic description vectors of services present in the four kernels. Based on user specified threshold value $User\_Threshold$, a ranked list of relevant services having $semantic\_similarity$ greater than $User\_Threshold$ are returned to the user.

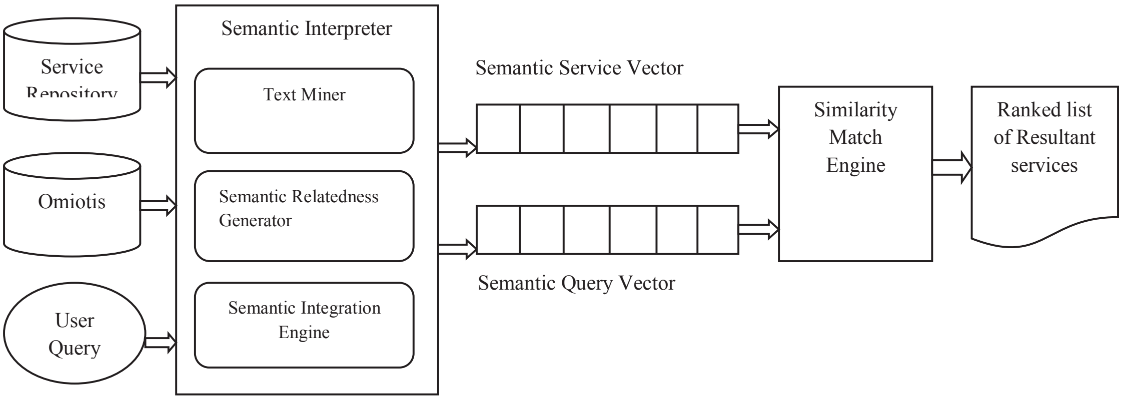


FIGURE 1: THE PROPOSED ARCHITECTURE

FIGURE 2: PROPOSED ALGORITHM

Input: Service set, $Query$: user query and $User\_Threshold$:Threshold value;

Output: Resultant service set $results$;

1. Extractthe terms from the textual documentation, arguments, operations and service name from the service profiles of the m WSDL profiles through Text mining.
2. Assign weight to all the terms in the $Terms(X)$ and generate
   a) $TF-IDF(m,X)$;
   b) $Binary(m,X)$;
   c) $Term\_Occurance(m,X)$;
   d) $Term\_Frequency(m,X)$;
3. Calculate semantic relatedness matrix $Omiotis(X,n)$ for all the terms in the $Terms(X)$ vector.
4. Generate semantic kernel for all four weight representations by merging semantic information in them:
   a) $Omiotis\_TF-IDF(m,n)$;
   b) $Omiotis\_Binary(m,n)$;
   c) $Omiotis\_Term\_Occurance(m,n)$;
   d) $Omiotis\_Term\_Frequency(m,n)$;
5. Generate user query semantic vector $Query$;
6. For all the semantic description vectors within a kernel
   i) Calculate cosine angle between the $Query$ vector and semantic description vectors of semantic kernel.
   ii) If $(semantic\_similarity \geq User\_Threshold)$ Then append service no. and $semantic\_similarity$ to the $results$ vector
   else drop the service.
7. Sort the $results$ and return ranked list of services $results$;

The proposed approach is divided into three phases:

i. *Parsing and weight generation of WSDL service profiles*
ii. *Calculation of integrated Semantic vectors for WSDL profiles*
iii. *Semantic matchmaking module for services*

i. *PARSING AND WEIGHT GENERATION OF WSDL SERVICE PROFILES*

In the proposed framework, in first stage service profiles are pre-processed to generate a set of tokens where relevant information under <element name> and <documentation> tags of the WSDL service profiles is extracted. All these extracted terms are stored in $Terms(X)$ vector. These terms are assigned weight using the four different schemes used in Information Retrieval [44]:

1. Term Frequency–Inverse Document frequency weighing ($TF-IDF$) scheme: The $TF-IDF$ is used for knowing the rare or important features in the corpus:

$$TF-IDF_{ij} = TF_{ij} * (\log \frac{N}{\{T_j \in D_n\}} \;; \qquad (1)$$

2. Binary Weightage: This kind of weightage scheme counts only the presence of a term within a WSDL profile. It does not consider the frequency of the term. It is calculated as:

$$Binary_{ij} = 1,\ if\ TF_{ij} > 0;$$
$$Binary_{ij} = 0, if\ TF_{ij} = 0; \qquad (2)$$

Here, TF_{ij} is the number of times that term j appears in WSDL profile for service i .

3. Term Occurrence wise weightage: This scheme counts the number of times a term has appeared in the WSDL profile.
$Term\_Occurance_{ij} = TF_{ij};$ (3)
4. Term Frequency wise weightage [43]: In addition to the term frequencies counted in Term Occurrence weighing scheme; this approach also normalizes the normal term frequencies by the square root of the sum of squares of all frequencies of the terms present in the WSDL profiles.

$$Term\_Frequency_{ij} = \frac{TF_{ij}}{\sqrt{\sum_{k=1 toX}(TF_{ik})^2}}; \quad (4)$$

The numerator is the frequency of the word being considered and the denominator is the square root of the sum of the squares of the frequency of each unique word. The term weighing (1-4) of all the services have been calculated using Rapidminer, a free available statistical tool.

After the four weighting matrices have been generated, these are forwarded to next phase of the framework to merge them with the semantic information of the services.

ii. *CALCULATION OF INTEGRATED SEMANTIC VECTORS FOR WSDL*

The semantic relatedness values of all the terms of the WSDL profiles are calculated with the dimension vector to generate a semantic matrix i.e. $Omiotis(X,n)$ for X terms present in the WSDL corpus. Here, dimension vector $Dim(n) = \{Dim_1, Dim_2, Dim_3, \cdots Dim_n\}$ contains different domains for the service set, as many times, it is not possible to allocate a single category to any service as services may potentially belong to more than one category. For calculating semantic relationship value between the terms and the dimensions, WordNet based Omiotis measure of semantic relatedness has been used [15]. Omiotis is the first measure of semantic relatedness between texts that considers all three factors for measuring the pair-wise word-to-word semantic relatedness scores. For weighting the semantic path Omiotis considers three key factors: (a) the semantic path length, (b) the intermediate nodes specificity denoted by the node depth in the thesaurus' hierarchy, and (c) the types of the semantic edges that compose the path. Omiotis is based on a sense relatedness measure, called SR. Semantic relatedness for a pair of terms $T(t_1,t_2)$ is calculated as follows [15]:

Definition 1 Given a word thesaurus O, let $T(t_1,t_2)$ be a pair of terms for which entries exist in O, let X_1 be the set of senses of t_1 and X_2 be the set of senses of t_2 in O. Let $S_1, S_2, \cdots, S_{|X_1||X_2|}$ be the set of pairs of senses, $S_k = (s_i, s_j)$, , with $s_i \in X_1$ and $s_j \in X_2$. The semantic relatedness of $T(SR(T,S,O)$ is defined as $max_{S_k}\{max_P\{SCM(S_k,O,P) \cdot SPE(S_k,O,P)\}\}$ $= max_{S_k}\{SR(S_k,O)\}$ for all $k = 1..|X_1| \cdot |X_2|$ (5)

Semantic relatedness between two terms t_1,t_2 where $t_1 \equiv t_2 \equiv t$ and $t \notin O$ is defined as 1. Semantic relatedness between t_1,t_2 when $t_1 \in O$ and $t_2 \notin O$, or vice versa, is considered 0.

The experimental evaluation have proved that Omiotis measure of semantic relatedness approximates human understanding of semantic relatedness between words better than previous related measures [15].The main benefit of integrating Omiotis with $TF-IDF$ is that it improvesthe discovery accuracy as semantic information contained in the Omiotis is merged with the statistical information present in the $TF-IDF$. Details about the Omiotis and various terms used here are availablein [15].

The term dimension semantic vectors for all the terms of the WSDL profiles are calculated. Further, the semantic syntactic integration is applied to generate four semantic kernels using different weighing schemes. The semantic relatedness values of different terms of the services contained in $Omiotis(X,n)$ is merged with the different weighing schemes based matrices viz. $TF-IDF$, $Binary$, $Term\_Occurance$ and $Term\_Frequency$ thus generating semantic integration kernels of web service description vectors.

Four kernels are:

1. Omiotis-TFIDF Integration Kernel: The $TF-IDF(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix i.e. $Omiotis(X,n)$ to generate a Omiotis-TFIDF Integrated Kernel:
$Omiotis\_TF\text{-}IDF = TF\text{-}IDF \times Omiotis$ (6)
2. Omiotis-Binary Integration Kernel: The $Binary(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix i.e. $Omiotis(X,n)$ to generate a Omiotis-Binary Integrated Kernel:
$Omiotis\_Binary = Binary \times Omiotis$ (7)
3. Omiotis-Term Occurrence Integration Kernel: The $Term\_Occurance(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix $Omiotis(X,n)$ to generate a Omiotis- Term Occurance Integration Kernel:

$$Omiotis\_Term\_Occurance = Term\_Occurance \times Omiotis \quad (8)$$

4. Omiotis-Term Frequency Integration Kernel: The $Term\_Frequency(m, X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix $Omiotis(X, n)$ to generate a Omiotis- Term Frequency Integration Kernel:

$$Omiotis\_Term\_Frequency = Term\_Frequency \times Omiotis \quad (9)$$

Finally, each row of these semantic kernels represents the semantic description vector for each service.

iii. *SEMANTIC MATCHMAKING MODULE FOR SERVICES*

In this phase, the semantic query vector for the users' query is calculated using Omiotis measure of semantic relatedness and the query is matched with the semantic web service description vectors of the services using cosine similarity measure. Thus, a ranked list of semantically similar services is generated as the output of the algorithm. The services with values less than the user specified threshold values are eliminated from the output list and remaining services are returned to the service consumer.

IV. IMPLEMENTATION OF THE APPROACH FOR WSDL SERVICE DESCRIPTIONS

The proposed approach was implemented on 98 services described in WSDL language; these service descriptions were downloaded from the Internet [45, 46]. These services were chosen from various domains like weather, stock, vehicle, food etc. The proposed algorithm has been evaluated using Precision, Recall and Fscore. Accordingly, set of relevant services for each query were defined and the framework was tested using ten queries with different user specified threshold values. The textual documentation, arguments, operations and service name from the service profiles were extracted out. After pre-processing, 562 terms were extracted from the WSDL profiles and stored in $Terms(562)$ vector. Further, for all the 562 terms in the $Terms(562)$ vector, the $TF-IDF(98,562)$ matrix, $Binary(98,562)$, $Term\_Occurance(98,562)$ and $Term\_Frequency(98,562)$ matrices were generated. Next, the semantic relatedness value between all the extracted terms of the services and all dimensions set $Dim(10)$ was calculated for in the dataset. The dimension vector constitutes the following 10 domains: Dim = {"Automobile", Book", "Film", "Weather", "Food", "Hospital", "Hotel", "SMS", "Stock", "Missile"}

TABLE 1.
COMPARISON OF PROPOSED APPROACH ON WSDL DATA SET WITH LATENT SEMANTIC ANALYSIS AT DIMENSION 10

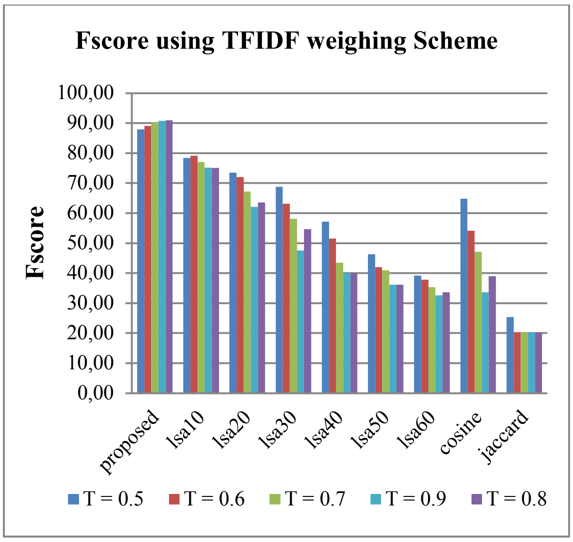
Weightage scheme	Threshold Value	Proposed Approach			LSA Approach		
		Precision	Recall	Fscore	Precision	Recall	Fscore
TFIDF	0.5	84.45	92.75	87.93	75.45	92.32	78.36
	0.6	87.33	91.63	89.09	78.00	91.61	79.16
	0.7	92.09	89.06	90.27	79.00	88.27	77.01
	0.8	96.39	87.39	90.95	79.00	85.73	75.02
	0.9	97.50	85.91	90.71	80.16	84.62	75.15
Binary	0.5	66.09	92.75	73.78	40.64	71.98	45.88
	0.6	76.55	92.03	81.39	40.97	66.14	43.74
	0.7	86.48	88.24	86.94	39.72	57.59	40.57
	0.8	93.75	83.29	87.78	40.54	54.60	40.66
	0.9	98.75	75.91	84.15	54.67	53.83	49.81
Term Occurrence	0.5	85.16	92.75	88.31	68.62	92.71	75.42
	0.6	86.66	91.84	88.90	69.96	91.81	76.14
	0.7	91.55	90.73	90.88	74.17	90.18	78.46
	0.8	96.39	87.39	90.95	81.76	77.30	77.17
	0.9	97.50	85.85	90.64	89.00	65.89	71.69
Term Frequency	0.5	85.16	92.75	88.31	68.62	92.71	75.42
	0.6	86.66	91.84	88.90	69.96	91.81	76.14
	0.7	91.55	90.73	90.88	74.17	90.18	78.46
	0.8	96.39	87.39	90.95	81.76	77.30	77.17
	0.9	97.50	85.85	90.64	89.00	65.89	71.69

The semantic relatedness score between the terms within each WSDL profile and all the dimensions of the dimension vector was calculated using the Omiotis measure of semantic relatedness and stored in $Omiotis(562,10)$ matrix. Further, the semantic information contained in the $Omiotis(562,10)$ matrix was merged with the statistical information of the $TF-IDF(98,562)$ matrix, $Binary(98,562)$, $Term\_Occurance(98,562)$ and $Term\_Frequency(98,562)$ matrices. Thus, four semantic kernels viz. $Omiotis\_TF-IDF$, $Omiotis\_Binary$, $Omiotis\_Term\_Occurance$ and $Omiotis\_Term\_Frequency$ based on different weighing schemes were calculated. Each row of these semantic kernels represents the semantic description vector for each service.

A total of ten queries were run to test the proposed approach on service profiles written in WSDL language. The semantic description vector of each query was matched against all the semantic description vectors for services in each of the four kernels and $semantic\_similarity$ was calculated between the queries and the service vectors. The proposed approach allows user to specify a degree of similarity matching threshold, i.e. $User\_Threshold$. Therefore, the services having $semantic\_similarity$ higher than the user specified $User\_Threshold$ will be returned to the user. The proposed approach was run for different threshold values of 0.5, 0.6, 0.8 and 0.9 for filtering the semantic value of match. The result of the proposed approach on WSDL profiles for ten queries in terms of Macro average precision, Macro average recall and Macro average F-Score were calculated.

The results of our approach were compared with information retrieval based Latent Semantic Analysis (LSA), $TF-IDF$ cosine similarity and Jaccard similarity. Latent Semantic Analysis, $TF-IDF$ cosine similarity and Jaccard similarity are all standard prevalent approaches which are normally considered for comparisons in IR research. The detailed results in terms of precision, recall and Fscore of top two performing approaches for ten user queries are provided in Table 1. The results were compared with the LSA (dimensions varying from 10 to 60).By looking at the results in Table 1, it can be clearly observed that the Macro average precision and Macro average Recall of proposed approach are considerably high than the precision of LSA10 approach.From the results, it can be analysed that our approach resulted in more relevant results in the final resultant set as the Macro average precision and macro average F-Measure of the proposed approach are much better than the results obtained from LSA (dimensions varying from 10 to 50).

The cosine similarities of the query vectors with the semantic description vectors of services for all the four kernels based on four weighing schemes were calculated. The comparative output of proposed approach with other approaches for all the four weighing schemes is presented in Figure 3 and Figure 4.



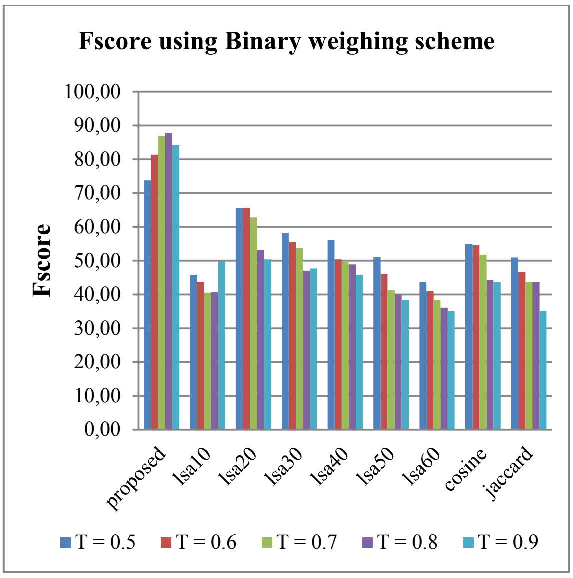


FIGURE 3: FSCORE COMPARISON FOR TFIDF AND BINARY WEIGHING SCHEME

Fscore comparison using TFIDF scheme: The proposed approach performed better than all other methods with Fscore value 87.9% at threshold 0.5. The accuracy improved on increasing the threshold value and attained its peak value of 90.9% at threshold value 0.8.Latent Semantic Analysis approach for dimension 10 performed next to the proposed approach at different thresholds. The Fscore of cosine similarity was found to be 64.7% at threshold 0.5 and the results of jaccard similarity were found to be the lowest.

Fscore comparison using Binary scheme: Again, the system performed best for the proposed approach at all threshold values. For binary scheme, LSA approach at dimension 20 performed next to the proposed approach. All other approaches (except theproposed and LSA 10) gave better results at threshold value 0.5 whereas the proposed attained peak results at threshold 0.8 with Fscore 87.7% and outperformed the other approaches.

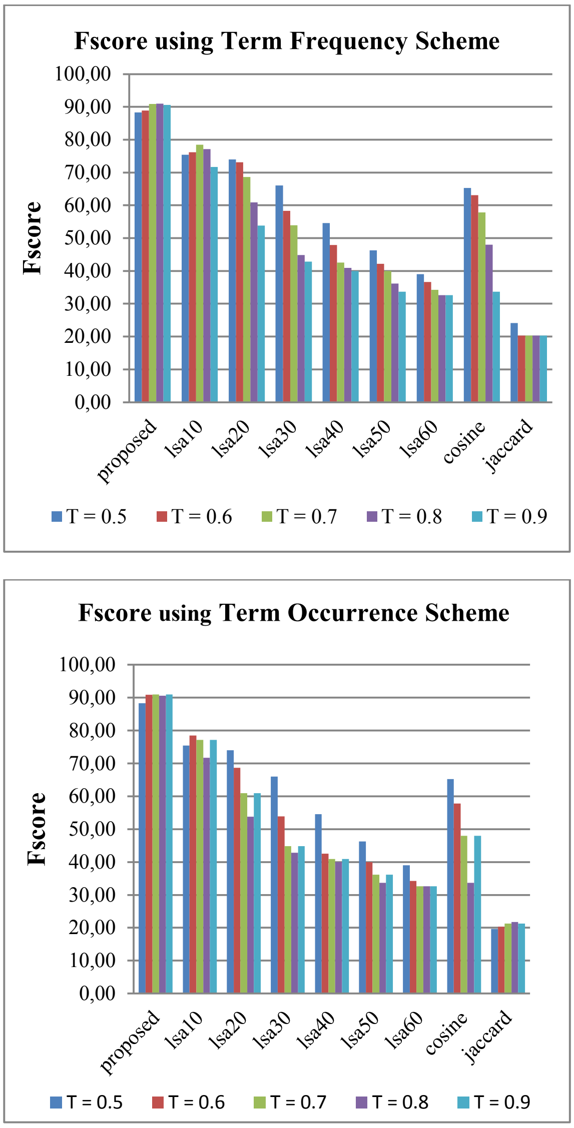


FIGURE 4: FSCORE COMPARISON FOR TERM OCCURRENCE AND TERM FREQUENCY WEIGHING SCHEME

Fscore comparison for Term Occurrence weighing scheme: The results of the proposed approach were quite promising for all threshold values in this scheme. The highest Fscore of 90.9% was achieved at threshold 0.8. The LSA approach performed better at threshold 0.7 at dimension 10 and for rest of the dimensions and approaches, better values were attained at threshold 0.5.

Fscore comparison for Term Frequency weighing scheme: In this scheme too, the results of the proposed approach were again found to be of highest values. Fscore values for this kernel were similar to the Term Occurrence weighing scheme. No major variations were seen in this weighing scheme.

During implementation, it was found that the Fscore of the proposed approach outperforms all the other methods for all the four weighing schemes. Based on the comparative analysis of all the results, it was found that the proposed framework is efficient and easy to use discovery approach, suitable for all kind of users. A novice user, who is not familiar with concerned ontologies, technology and implementation details can easily discover the existing services over the internet without any technical overhead. Ambiguity, polysemy and synonymy issues of the terms used in the service profiles are dealt through Omiotis measure of semantic relatedness. This approach matches the user query to the semantically similar service vectors of the semantic kernels according to the hybrid similarity score and finally returns a ranked list of semantically similar Web services along with their corresponding similarity scores.

V. CONCLUSION AND FUTURE SCOPE

Keeping in view the fact that the users have very less knowledge regarding the underlying technologies of the web services, discovery frameworks, description languages and implementation details, we have proposed a semantic framework that enables the web service discovery based on the combination of semantic and syntactic information of the service profiles. A novel approach has been presented which takes advantages from measures of semantic relatedness and statistical models for providing efficient means to automate the discovery process of Web services. The proposed approach is implemented on WSDL services and the experimental evaluations have shown that the performance of the discovery process can be significantly improved by combining the information retrieval techniques with the measures of semantic relatedness. The proposed approach proved to be effective in retrieving more relevant results for the user requirements.

REFERENCES

[1] S. Robak, B. Franczyk, Modeling web services variability with Feature diagrams, In: Proceedings of web services and database systems, 2002, pp. 20-128.

[2] E. Dashofy, M. van der Hoek, A. Taylor, A highly-extensible, XML-based architecture description language Software Architecture, In: Proceedings of Working IEEE/IFIP Conference, Amsterdam, Netherlands, 2001.

[3] F. Curbera, M. Duftler, R. Khalaf, W. Nagy, N. Mukhi, S. Weerawarana, Unraveling the Web Services, Web: An introduction to SOAP, WSDL, and UDDI, IEEE Internet Computing, 6(2), April 2002.

[4] World Wide Web Consortium (W3C), Web Services Description Language (WSDL) 1.1., 2001. http://www.w3.org/TR/wsdl/.

[5] UDDI Technical White Paper, 2000. http://www.uddi.org/pubs/Iru-UDDI-Technical-White-Paper.pdf

[6] N. Shadbolt, T. B. Lee, W. Hall, The Semantic web revisited, IEEE Intelligent Systems, 21(3), 2006, pp. 96-101.

[7] D.Martin, M.Burstein, J.Hobbs, O.Lassila, K.Sycara, OWL-S: Semantic markup for web services, 2004. http://www.w3.org/submission/OWL-S/.

[8] D. Fensel, H. Lausen , A.Pollers, J.D. Bruijn, M. Stollberg, D. Roman, J. Domingue, Enabling Semantic Web Services: The web service modeling Ontology, Springer Verlag, ISBN 978-3-540-34520-6, 2007, pp. 38-61.

[9] R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, T. Schmidt, A. Sheth, K. Verma, Web service semantics-WSDL-S,2005. http://www.w3.org/submission/WSDL-S/.

[10] J. Farrell, H. Lausen, Semantic Annotations for WSDL and XML Schema, 2007. http:/www.w3.org/TR/SAWSDL/.

[11] C. Kiefer, A. Bernstein, The creation and evaluation of ISPARQL strategies for matchmaking, In: Proceeding of Fifth European Semantic Web Conference (ESWC), volume-5021, LNCS, Springer Verlag Berlin Heidelberg, 2008, pp. 463-477.

[12] M. Klusch, B. Fries, K. P. Sycara, OWL-MX: A Hybrid semantic Web Service matchmaker for OWL-S services, Web Semantics 7(2), 2009, pp.121-133.

[13] M. Klusch, F. Kaufer, Wsmo-mx: A hybrid semantic web service matchmaker, Web Intelligence and Agent Systems, 7(1) 2009, pp. 23-42.

[14] D. Wei, T. Wang, J. wang, A. Bernstein SAWSDL-iMatcher: A customizable and effective Semantic Web Service Matchmaker, Web Semantics: Science, Services and Agents on the World Wide Web 9(4), 2011, pp. 402-417.

[15] G. Tsatsaronis, I. Varlamis, M. Vazirgiannis, Text relatedness based on a word thesaurus, Artificial Intelligence Research ,Volume-37, 2010, pp. 1-39.

[16] J.A. Nasir, A. Karim, G. Tsatsaronis and I. Varlamis, A Knowledge-based Semantic Kernel for Text Classification, In: Proceedings of the 18th international conference on String processing and information retrieval, 2011, pp. 261-266.

[17] M. Paolucci, T. Kawamura, T.R Payne, K.P Sycara, and Semantic matching of Web Services capabilities In: Horrocks, I., Hendler, J. (eds.) ISWC, LNCS, vol. 2342, Springer, Heidelberg, 2002, pp. 333-347.

[18] D. Bruijn, R. Lara, A. Polleres, D. Fensel, and OWL- DL vs. OWL light: conceptual modeling and reasoning for the semantic Web, In: Proceeding of WWW conference, Chiba, Japan, 2005.

[19] T. Kawamura, J.A.D Blasio, T.Hasegawa, M.Paolucci, K.Sycara, Public deployment of semantic Service Matchmaker with UDDI business Registry, LNCS, Volume 3298, Springer, Heildelberg, 2004, pp.752-766.

[20] R. Gronmo, M.C. Jaeger, Model-Driven Semantic Web Service Composition, In: APSEC December 2005, pp. 79–86.

[21] L. Li, I. Horrocks, A software framework for matchmaking based on Semantic Web technology, In: Proc. of WWW, Budapest, May 2003, pp. 331–339.

[22] H. Wang, Z. Li, A Semantic Matchmaking Method of Web Services Based On SHOIN(D)*, In: Proc. of IEEE APSCC, Guangzhou, China, December 2006, pp. 26–33.

[23] K. Sycara, S.Widoff, M.Clutch, J.Lu, LARKS: Dynamic Matchmaking among Heterogenous Software Agents in Cybersace,Autonomous Agents and Multi agents Systems, Volume 5, 2002, pp.173-203.

[24] K. Verma, K. Sivashanmugam, A. Sheth, A. Patil, S. Oundhakar, J.Miller, METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services, Information Technology and Management, 6(1),2005, pp. 17-39.

[25] K. Sycara, S.Widoff, M.Clutch, J.Lu, LARKS: Dynamic Matchmaking among Heterogenous Software Agents in Cybersace,Autonomous Agents and Multi agents Systems, Volume 5, 2002, pp.173-203.

[26] S. Batra, S. Bawa, Semantic discovery of Web Services using principal component analysis, Physical Sciences, 6(18), 2011, pp. 4466-4472.

[27] T. Wen, G. Sheng, Y. Li, Q. Guo, Research on Web Service Discovery with Semantics and Clustering, In: Proceedings of IEEE sixth joint international Conference on Information technology and Artificial Intelligence (ITIAC), 2011, pp. 62-67.

[28] J. M. García, D. Ruiz, A. Cortes, Improving semantic Web Services discovery using SPARQL-based repository filtering, Web Semantics: Science, Services and Agents on the World Wide Web, Elsevier, 2012, pp.12-24.

[29] R. Thiagarajan, W. Mayer, and M. Stumptner, Semantic Service Discovery by Consistency-Based Matchmaking, Advances in Data and web Management, Springer,2009, pp. 492-505.

[30] M. Senvar, A. Bener, Matchmaking of Semantic Web Services Using Semantic-Distance Information, *ADVIS, volume 4243, LNCS, Springer, 2006, pp. 177-186.*

[31] L. Chen, Z. Song, Y. Zhang, and Z. Miao, Wordnet Enhanced Dynamic Semantic Web Services Discovery, Emerging Research in Artificial Intelligence and Computational Intelligence,CCIS volume-237, 2011, pp. 529-536.

[32] A. Grintsvayg, V. D. Veksler, R. Lindsay, W. D. Gray, Vector generation of explicitly defined multidimensional semantic space, In: Proceedings of ICCM Eight international conference on Cognitive Modeling, Oxford, UK, 2007, pp. 231-232.

[33] I. Kaur, A. J. Hornof, A Comparison of LSA wordnet and PMI for predicting User Click Behaviour, In: Proceedings of the conference on human factors in Computing,CHI, pp. 51-60.

[34] Normalized (Point wise) Mutual Information in Collocation Extraction http://www.ling.uni-potsdam.de/~gerlof/docs/npmi-pfd.pdf

[35] R. L. Cilibrasi, P.M.B. Vitanyi, The Google Similarity Distance, ArXiv.org or Clustering by Compression, IEEE Journal of Transactions in Information Theory, 51(4), 2004, pp. 1523 - 1545.

[36] E. Gabrilovich, S. Markovitch , Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis, In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, 2007.

[37] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, In: Proceedings of the 14th International joint conference on Artificial intelligence,Volume-1, Morgan Kaufmann Publishers Inc., San Francisco CA,USA,1995, pp. 448-453.

[38] D. Lin, An Information-Theoretic Definition of Similarity, In: Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98), Jude W. Shavlik (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998, pp. 296-304.
[39] J. Jiang, W. Conrath, Semantic similarity based on corpus statistics and lexical Taxonomy, In: Proceedings of International conference on Research on Computational Linguistics, 1997, pp. 90-98.
[40] M. Crasso, A. Zunino, M. Campo, AWSC: An approach to Web Services classification based on machine learning techniques, CONICET, Inteligencia Artificial, RevistaIberoamericana de Inteligencia Artificial. 12(37), 2008, pp. 25-36.
[41] http://projects.semwebcentral.org/projects/owls-tc/.
[42] http://Omiotis.hua.gr/WebSite/wsinfo.html.
[43] http://rapid-i.com/rapidforum/index.php?action=printpage;topic=3728.0
[44] R. Cummins, O. Colm, Evolving general term-weighting schemes for information retrieval: Tests on larger collections, Artificial Intelligence Review, 24.3-4, 2005, 277-299.
[45] www.service-repository.com
[46] www.xmethod.com

Shailja Sharma is working as System Analyst in Kurukshetra University, Kurukshetra. She did B.Tech from Kurukshetra University, Kurukshetra and M.E in Computer Sc. and Engineering from Thapar University, Patiala. She is a research scholar in the area of Semantic Web Services from the Department of Computer Applications, National Institute of Technology, Kurukshetra. Her research interests are focused on Semantic Web, Web Services, Data Mining and Machine Learning.

Jagdeep Singh Lather is working as a Professor at Department of Electrical Engineering and Computer Applications, National Institute of Technology, Kurukshetra. He holds a B.Tech, M.Tech and doctorate degree in Electrical Engineering. He has over 15 years of experience in teaching and research. His areas of interests are Robust Control, Flexible AC Transmission systems, Mining Techniques and Decision Trees, ANN and Fuzzy Logic. He is life time member of Scientific and Professional Societies, ISTE, India.

Mayank Dave is working as Professor in the Department of Computer Engineering at National Institute of Technology, Kurukshetra, since 1991. He holds a PhD in Computer Science and Technology from Indian Institute of Technology, Roorkee, India. He has twenty four years of experience in research in the area of Computer Networks. His areas of interests are Computer Networks, Database Systems, QoS in Mobile Adhoc and Sensor Networks, Software Engineering, Web Technologies. He is a member of the IEEE, IEEE Computer Society, IETE, Institution of Engineers (I), CSI and ISTE.