Semantic Approach for Classification of Web Services Using Unsupervised Normalized Similarity Measure

Shailja Sharma
Department of Computer Applications, National Institute of Technology, Kurukshetra, India
Email: shailjakaushik@gmail.com

J.S. Lather and Mayank Dave
Department of Computer Engineering, National Institute of Technology, Kurukshetra, India
Email: jslather@gmail.com, mdave67@gmail.com

Abstract—**Automatic or semiautomatic categorization of web services facilitates the relevant service retrieval as well as it helps the administrators in attaining globally consistent classification decisions that are independent of the administrator's knowledge of application domain, organization of taxonomies and service characteristics. Lack of automatic mechanisms to help service publishers in the classification task, irrelevant and huge number of services returned by the UDDI, and lack of standard mechanisms that helps in the discovery of desired web services are some of the issues that need focus. In this paper, an automatic approach for service categorization is proposed that uses a lexical semantic network constructed from the web snippets as a knowledge base for semantic similarity calculation between the service profiles and the categories. Our approach involves mapping of service profiles to a category based dimension vector by using the notion of semantic similarity and aims at alleviating the administrator's job by automatically providing them with a set of categories ranked based on the degree of semantic similarity. Empirical evaluation on a set of OWL-S services shows that the proposed approach helps in better decisions for relevant classification of services by giving an ordered set of categories based on the similarity scores.**

Index Terms—**Web Service Discovery, Semantic similarity, Semantic web, Information Retrieval, Text Mining, OWL-S.**

I. Introduction

The emergence of Web services provides flexibility for the interoperation and integration of applications through standard protocols for the next level of evolution of e-business. Classification of web services facilitates the service retrieval through the (semi)automatic discovery mechanisms. Decisions for assigning proper category to a service needs complete knowledge regarding multilevel hierarchal taxonomies which tend to be extremely large as they contain thousand of categories. This also requires the sufficient expertise and knowledge regarding the characteristics of the web service, domain and the complete organization of the UDDI [1]. Involvement of different administrators in decentralized repositories takes their knowledge based classification decisions which may vary from one to other. As a result, the classification decisions get dependent on the knowledge and expertise of the repository administrators. Apart from the service discovery and retrieval, assessing the relevant domain for a service is also vital at the time when semantic annotations need to be provided. Semantic Web Services [2] annotates the different aspects of Web services using machine-understandable semantics that enables the automatic discovery of Web Services. Various ontology based frameworks like OWL-S [3] (Ontology Web Language for Services), WSMO [4] (Web Service Modeling ontology) etc. have been proposed in past to facilitate the description and discovery of semantic Web Services. Similarly efforts have been put on to add semantics to the existing WSDL standard specifications leading to the introduction of WSDL-S [5] and the W3C recommendation SAWSDL [6] (Semantically Annotated WSDL) etc. Practically, most of the services found on web do not have explicit semantic information in terms of ontological concepts. In order to annotate these services with relevant ontologies, one needs to know the appropriate domain that a service belongs to. Automatic categorization also serves this purpose by telling the most relevant domain so that the ontologies from the recommended domain can be used for annotation. For adding semantic concepts to already existing services, manual annotation is required. But as the number of services and ontologies are growing exponentially day by day where each ontology contains thousand of concepts, it is really time consuming and cumbersome task to manually discover the appropriate domain and thereafter, the concepts from the relevant ontologies and annotate all existing services with the ontological concepts. In order to annotate the terms, the service profiles must be matched to the ontological concepts and accordingly appropriate and relevant

doi:10.4304/jetwi.6.3.364-372

ontology needs to be identified from the ontology store. For determining the suitable ontology, the web service must be classified to a relevant domain and accordingly the most appropriate ontology must be selected. Selection of the correct domain and then discovery poses the problems during semantic service discovery as well because in semantic web service frameworks again the user request needs to be formulated using semantic concepts from the different ontologies. Moreover, a service may have semantic relationship with two or more categories. In order to get a standardized classification decision system irrespective of the administrator knowledge; semiautomatic or automatic mechanisms need to be proposed.

Taking these problems into consideration an effective method for service classification is required which can use the hidden semantics of the existing services and can identify the most appropriate domains for the service profiles. We aim at automatically providing the service administrator with a set of likely appropriate taxonomy categories ranked by semantic relationship between the service and the categories, when a new service has to be registered in the repository. The proposed approach for service classification uses the implicit semantic relationship between the terms of the service profile and the category set for the task of classification of services. Our approach utilizes the semSim[7] corpus which is based on a semantic lexical network which has been created using the snippets downloaded from the web using the IND relation. The service profiles are mapped to category based semantic vectors using the normalized similarity scores. These vectors are merged with the IR based $TF-IDF$ metric. We consider such a combination vital since the $TF-IDF$ based syntactic similarity is unable to capture service semantics. If we have two profiles where web service1 is to know the price of an automobile and web service 2 provides information regarding the cost of a vehicle. As terms in both of the service profiles share common context, they need to be classified to a common domain based on their semantic relations. Therefore, we like to utilize the semantic of the profiles for better classification of the services. By using the proposed approach the services can be ranked based on their similairy to the categories, enabling better classification decisions, better retrieval of services and selection of appropriate ontologies from the recommended domain so that the terms present in the service profiles can be annotated using these explicit semantic concepts.

The remainder of this paper is organized as follows. Section 2 reviews main approaches for service classification. Section 3 introduces the proposed approach. Section 4 evaluates the data, tools, implementation and results of the proposed approach. Finally, the conclusions and future work are summarized in Section 5 followed by the references.

II. LITERATURE REVIEW

Various approaches [8, 9, 10, 11, 12, 13, 14, 15 and 16] for service classification in literature differ on the basis of matching syntactic or semantic concepts of argument definitions [8, 9, and 12] or document classification techniques [10, 11]. Among the various classification approaches MWSAF [8] matches the argument definitions for classifying the WSDL documents. MWSAF is based on graph matching. It converts the argument definitions into various graphs and then matches it with the ontological concepts for categories. [9] Converts the service definitions into the ontological concepts and then uses an ontology mapping is done for classifying various services into their relevant class. METEOR-S [10] presents improved version of MWSAF and considers each web service as a document and applies the document classification in the area of web services. In comparison, ASSAM considers the natural language description. It combines the SVM and Naïve Bayes algorithms to the set of WSDL documents. Batra and Bawa et al. [13] proposed a NSS based approach for semantic web service classification. By using the Normailized Similarity Score (NSS) Measure of Semantic Relatedness, the similarity score between the terms of a service and all the categories is found. In [16], an approach to automatic classification of web services has been proposed by using several vector based representations models for web services. By combining the textual descriptions and the Input /Output signature for syntax /semantic annotations, the services are classified using different machine learning classifiers. AWSC[14] approach for automatic classification of service is based on the Rocchio algorithm where the each service is considered as a separate document and Text mining and machine learning techniques have been used for the service classification. Major observations of these approaches are: first, Classification approach proposed by MWSAF and ASSAM has shown low accuracy and are implemented on small dataset. Second, Meteor-S, the machine learning based enhanced version of MWSAF is fast and shown to have better accuracy, but it do not make include the service documentation or comments written in natural language for classification. Third, Assam has the problem of low accuracy. The major disadvantage of the above said approaches is that most of these do not takes into account the semantic information hidden in the service profiles for the classification purpose which can be considerably beneficial in improving accuracy. In comparison to the approaches presented above we have implemented our approach including the natural language comments and semantics. This automatic classification helps in the better classification and retrieval of services.

III. OVERVIEW OF THE PROPOSED METHODOLOGY

The proposed approach for automatic classification of services uses a lexical semantic network constructed from the web snippets as a knowledge base for the calculation of semantic similarity between the service profiles. In order to better utilize the hidden semantics of service profiles, we have tried to exploit the semantic relation between the services and the categories. Our method involves mapping of service profiles to a category based dimension vector by using the notion of semantic similarity which is further merged with the IR based techniques of weight generation and is used for calculating the semantic degree of similarity between the services. It considers web services in the service repository i.e. $Service\_set$, and a $Category\_vector$ = $\{Category_1, Category_2, Category_3, \cdots Category_n\}$ having n domains used to classify the services in the repository. For a new service $Service_{New}$ that has to be registered and classified in the service registry, our proposed method will automatically recommend a service from the $Category\_vector$ that best suit the new service i.e. $Service_{New}$. For service classification, it uses the classification of services. Our approach utilizes the semSi implicit semantic relationship between the terms of the service profile and the category set for the task of m[7] corpus which is based on a semantic lexical network which has been created using the snippets downloaded from the web using the IND relation. In SemSim[7] the terms or concepts present in the lexical network are linked on their context based similarity. For calculating semantic similarities between the service terms and dimensions, an unsupervised normalized similarity values derived from a linear query complexity has been used. Further based on the notion of semantic neighborhood, the semantic similarities between the terms are calculated and finally the symmetric global normalization has been applied. The proposed approach can be divided into two phases, involving Parsing, Similarity calculation, and classification and ranking.

A. Parsing the Profiles for Service Classification

Many a times the classification based on the brief information in terms of input and output concepts provided by the service profiles may mislead the system. Two services intended for different purposes may have syntactically similar interfaces, therefore they will exhibit high similarity values despite they share different context. Since the similarity measure between services is a crucial aspect in classification approach, these cases would lead to misclassification. Keeping in view the practical feasibility and independence from any specific language or framework like OWL-S[3], WSMO[4] etc., we propose an approach that takes into consideration the minimum information that is present in each profile irrespective of its language or framework i.e. the signatures and the textual description written in natural language. We have used the $TF-IDF$ representation of service profiles as a base vectors which has been further modified to the dimension vector based on semantic relationship with the different domains. The preprocessing process starts with extracting the relevant, non-trivial and quality information from the service profiles. In our approach the functional parameters like input, output, service names and service description are extracted from each service document. Pre-processing includes tokenization, split of combined words, Stop words removal and stemming. Further, the $TF-IDF$ Matrix was generated. This measure combines both term-frequency and inverse-document-frequency and can be calculated as:

$$TF-IDF = TF * IDF \tag{3.1}$$

Where TF is the term frequency and IDF is the Inverse document frequency and is represented as:

$$IDF = \log \frac{N}{\{|Tj \in Dn|\}} \tag{3.2}$$

Term-frequency inverse-document-frequency ($TF-IDF$) is the most widely adopted measure for finding the important feature or terms in a collection.

B. Mapping of Service Profiles to Category Based Dimension Vectors

In this phase the data extracted from the service profiles are mapped to a category based dimension vector using the normalized scores. The category dimension vector constitutes the following ten categories:

$Category\_vector$ = {“Library”, “Automobile”, “Food”, “Country”, “Weapon”, “Entertainment”, “Book”, “Education”, “Medicine”, “ Weather”}

Further, the semantic similarity between all the terms of the service profile is calculated with all the categories of the $Category\_vector$ using the SemSim [7] based similairty. In SemSim [7], for each term, individual query has been formulated and 1000 top web snippets were retrieved using the Yahoo search API. After downloading the IND based web snippets, a contextual window of size 2H + 1 words is centered on the word of interest and lexical features are extracted [7].

$$[f_{H,l} \cdots f_{2,l}\ f_{1,l}]\ w_i\ [f_{1,r}\ f_{2,r} \cdots f_{H,r}] \tag{3.3}$$

For every instance of w_i in the corpus, the H words left and right of w_i as shown in (3.3) are taken into consideration. Further, the computation of semantic similarity is done using (3.4)

$$S^H(w_i, w_j) = \frac{\sum_{k=1}^{Z} t_{w_i,k}\ t_{w_j,k}}{\sqrt{\sum_{k=1}^{Z} t_{w_i,k}^2}\ \sqrt{\sum_{k=1}^{Z} t_{w_j,k}^2}} \tag{3.4}$$

The non-zero feature value $t_{w_i,k}$ indicates the occurrence of vocabulary word t_k within the left or right context of w_i . Further, a lexical network is generated. Here the semantic neighborhood is calculated for all the terms in the corpus. At last, the context similarities are normalized according to the symmetric Global normalization scheme (Z normalization) given at (4). The statistics of similarities, i.e., mean and variance, were computed across the entire network. Contextual window size H = 5 was used for calculating Similarity scores between various terms.

$$S_Z^M(n_1, n_2; H) = \max\{s_Z(n_1, n_2; H), s_Z(n_2, n_1; H)\} \quad (3.5)$$

Once Semantic similarity has been calculated between the category dimension vector and the terms present in the service profile using the normalized scores, the resultant vectors are merged with the $TF-IDF$ values to give a combined semantic vector representation of the service profile. Finally, this combined service vector is taken as a measure value for ranking the categories according to their predicted similarity for $Service_{New}$. The categories are ranked as per the order of higher semantic similarity and this ordered list is returned to the user.

IV. Expermental Evalution And Results

For implementing our approach we used 80 services from the Owls-TC v2 dataset. The collection Owls-TC v2 is available as open source at [17]. For calculating the semantic similarity between various term and dimensions, we used SemSim Corpus which is publically available at [20]. SemSim[7] corpus is build form the 8, 752, 000 snippets downloaded from the web that contained 199, 510, 174 tokens.

Table 1:
Semantic Similarity Calculator Batween Service Profile And Category Vector

	Library	Automobile	Food	Country	Weapon	Entertainment	Book	Education	Medicine	Weather
Hotel	1.5529	1.5358	1.8151	1.8543	1.1682	2.3087	1.7776	1.6400	1.1229	2.1356
City	2.1059	1.4744	2.5917	2.8274	1.6722	2.5700	2.6802	2.4257	1.6748	2.4809
country	2.1409	1.3841	2.6591	4.7017	1.6812	2.4406	2.7143	2.5448	1.8600	2.3475
portal	1.6069	1.3611	1.1624	1.1914	1.0615	1.6057	1.1254	1.4065	1.2515	1.3235
travel	2.1235	1.5774	2.4580	2.4055	1.5242	2.5165	2.4348	2.2496	1.7561	2.5784

For implementation, we have used Rapid Miner [18], and Matlab[19] tools. The dataset contains services from different domains like entertainment, library, weather, vehicles, food, weapon etc. For our approach we used the value of H = 5 for context window. To illustrate the approach we parse all the input, output and textual description of services. After parsing the semantic similarity between the services and all ten dimensions was calculated using the semSim[7] similarity corpus and merged with the $TF-IDF$ data. A service may belong to more than one category and depending upon the domain knowledge and expertise, different administrator will assign different category to a service. In order to resolve the issue of finding the most semantically relevant category for a service, automatic ranking based on the semantic similarity degree is proposed in this paper.

The approach can be described with the help of an example, where a new service $Service_{New}$ is a travel portal meant for giving information about a hotel in a city of a country. After extracting the terms from the input, output and text description of this service the similarity values are calculated with the $Category\_vector$ which is represented in Table 1 above.

Table 2:

Sr. No.	Terms	TFIDF
1	Hotel	0.3638
2	City	0.3462
3	country	0.3462
4	portal	0.1513
5	Travel	0.5457

TFIDF Values For The Above Service

We do not consider each term as equally important therefore, the similarity scores are adjusted as per the importance of the term in the corpus. The similarity value is adjusted with the syntactic information present in the profile in terms of the $TF-IDF$ values. The $TF-IDF$ values of the terms present in the service $Service_{New}$ are presented in Table 2. These values are merged with the term to category based similarity scores. In order to do this the $TF-IDF$ values are multiplied to the term to category similarity scores as calculated earlier in Table 1.

TABLE 3:
CATEGORY BASED INTEGRATED DIMENSION VECTOR FOR THE SERVICE 1

	Library	Automobile	Food	Country	Weapon	Entertainment	Book	Education	Medicine	Weather
Hotel	0.5649	0.5587	0.6603	0.6746	0.4250	0.8399	0.6466	0.5966	0.3638	0.7769
City	0.7291	0.5105	0.8973	0.9789	0.5790	0.8898	0.9280	0.8399	0.3462	0.8590
country	0.7413	0.4792	0.9206	1.6279	0.5821	0.8450	0.9398	0.8811	0.3462	0.8128
portal	0.2431	0.2059	0.1758	0.1802	0.1606	0.2429	0.1702	0.2128	0.1513	0.2002
travel	1.1587	0.8607	1.3412	1.3126	0.8317	1.3731	1.3286	1.2275	0.5457	1.4069
	3.4370	2.6150	3.9953	4.7741	2.5783	4.1907	4.0132	3.7578	1.7532	4.0558

After merging the $TF-IDF$ values, the integrated category based dimension vector is calculated. The similarity score of each term present in the service is combined to get a cumulative value for the category. Similarly, the service to category similarity score is calculated and the service profile is finally mapped to a category based dimension vector. The integrated dimension vector for the service in consideration is depicted in Table 3. Finally, the cells of this integrated service vectors is sorted to know the ordered list of categories and then depending upon the highest value, the category is suggested to the service. In this case, by using the proposed approach the category “Country” having the highest similarity score of 4.7741 will be assigned to the service $Service_{New}$.

TABLE 4:
LIST OF CATEGORIES ORDERED ON SIMILARITY DEGREES FOR SERVICE 1

Categories	Similarity Degree	Ranking
Country	4.7741	1
Entertainment	4.1907	2
Weather	4.0558	3
Book	4.0132	4
Food	3.9953	5
Education	3.7578	6
Library	3.4370	7
Medicine	2.7799	8
Automobile	2.6150	9
Weapon	2.5783	10

One more scenario is presented to better justify the results, where incoming services will serve the purpose of providing information regarding the books, novels, author and price etc. In this case the input, output and service description extracted from the different service profiles of the data set are depicted in Table 5. In case, if categorization is done manually during the registration of service, then different administrators may place these services under different categories like Library, Education, Entertainment, Book but this will create an ambiguity for the service users during the service retrieval as users have to search these services again based on the categories and despite its relevance to a query and presence in a different category, one will not be able to find and reutilize the service. Our proposed approach will alleviate the problem by automatically providing a list of categories ranked based on the semantic degree of similarity. The category having highest similarity will be assigned to the service.

TABLE 5:
SNAPSHOT OF INPUT, OUTPUT AND DESCRIPTION EXTRACTED FROM THE SOME OF THE SERVICE PROFILES

SR. NO.	Service Name	INPUT	OUTPUT	Description
1	book_author_EncSSservice	book	author	This service returns author of a certain book such as short story or encyclopedia.
2	book_author_service	book	author	This service returns author of the given book.
3	book_authorbook-type_service	book	author booktype	BAT service is one of most reliable service to returns author and type of the given book.
4	book_authorprice_Novelservice	book	author price	This service returns author and price of a book, short-story or novel.
5	book_authorprice_service	book	author price	This service returns author and price of a book, short-story or text book (but no novel).
6	book_authortext_service	book	author text	This service returns author of the book, novel or short story, and its personal notes as a text for the book
7	book_Cheapestprice_service	book	price	A Service that searches the cheapest price for a book
8	book_person_Publisherservice	book	person	This service informs you for a person who works as co-publisher of a certain book.
9	book_price_service	book	price	return price of a book
10	bookpersoncreditcardaccount__BShopservice	person credit account book	price	adds the selected book in his shopping cart.
11	bookpersoncreditcardaccount_price_service	creditcardaccount person book	price	adds the selected book in his shopping cart.
12	BookSearchService	title	book	returns recommended price of the book.
13	BookPrice	book	price	return price of a book
14	novel_author_price_service.owls	novel	author, price	This service returns author and price of a given novel.
15	novel_price_service	novel	price	This service returns price of a novel.
16	title_pricebook_service	title	price , book	KAHN is a recommended service to find high valuable books
17	science-fiction-novel_authorprice_service	Science_Fiction novel	author, price	This service returns author and price of a given science-fiction-novel.
18	author_sciencefictionbookprice_service	author	Science_Fiction_ book, price	This service returns science fiction books written by the given author and their price as well.

TABLE 6:
SNAPSHOT OF SEMANTIC SIMILARITY OF FEW TERMS WITH CATEGORIES

	Library	Automobile	Food	Country	Weapon	Entertainment	Book	Education	Medicine	Weather
Book	2.4439	1.0721	2.7734	2.8490	1.6372	2.4978	4.4500	2.5896	1.8710	2.3280
author	2.0388	0.8176	1.9314	2.0619	1.2765	1.9945	2.5802	2.0596	1.6338	1.6306
Price	2.1629	1.4827	2.6011	2.5194	1.5720	2.2764	2.7592	2.1380	1.6826	2.3662
Text	2.3294	1.2081	1.9175	1.9482	1.5944	1.8493	2.2779	2.0481	1.6336	1.7650
review	2.3703	1.1692	2.6274	2.6404	1.5747	2.2932	3.0351	2.4169	1.8827	2.2506
publisher	1.1655	0.8516	0.7242	0.8264	0.7868	1.2436	0.9842	0.9668	0.7375	0.7825
reader	1.5336	0.9241	1.1706	1.1818	1.1911	1.4047	1.3425	1.3086	1.2147	1.2097
recommendation	0.5678	0.6567	0.2275	0.3215	0.5276	0.2981	0.1979	0.5408	0.6272	0.2798
Novel	1.6662	0.8337	1.6648	1.6614	1.3837	1.5798	2.0596	1.5884	1.5657	1.4431
Title	2.1474	1.1901	2.0675	2.2541	1.5818	2.1013	2.5792	2.0813	1.6273	1.8944
science	2.1325	1.3359	2.1151	2.0856	1.6497	1.9470	2.2787	2.4403	2.2356	1.8425
fiction	1.2994	0.8574	1.0102	1.0833	1.1493	1.4694	1.3450	1.1608	1.1829	0.9891
Tax	1.7465	1.7786	1.7760	1.9711	1.5825	1.7157	1.7449	2.1178	1.5244	1.7423
Cost	2.0626	1.6906	2.3907	2.2919	1.6476	2.0099	2.2421	2.2783	1.9438	2.2493
encyclopedia	1.1101	0.4622	1.0580	1.0847	0.7013	0.9993	1.3002	0.8912	0.7813	1.0394
Story	2.1748	1.0309	2.4600	2.6430	1.6495	2.3191	3.0115	2.3008	1.6824	2.1508
person	2.1383	1.1873	2.5280	2.6279	1.7646	1.9870	2.8115	2.4493	2.0535	2.0381
Type	2.2735	1.3277	2.7175	2.5175	1.6763	1.9978	2.7582	2.2851	2.0861	2.2174

Table 6 represents a snapshot of the semantic relationship between terms of services present in Table 5 and set of categories. For all the terms in a service profile the semantic relationship between the terms and the domains is calculated using the normalized similarity value and the service profile is mapped to a category dimension vector. Further, the $TF-IDF$ weights are merged with these vectors so that the syntactic as well as well as semantic information of the service can be utilized for its better classification. In applications where non semantic service profiles are mapped to the semantic concepts from the ontologies, the resultant ordered set of categories is returned to the user so that he can find the most relevant domain and further the appropriate ontologies.

Table 7 gives a snapshot of some of the integrated category based dimension vectors. Now each cell of this vector consists of the semantic degree of similarity with the service profile. Finally, the cells of this integrated service vectors is sorted to know the ordered list of categories and the depending upon the highest value, the category is suggested to the service.

For service number 48 in Table 7 the ranked list of category is calculated and Table 8 shows ranked list of categories based on the semantic similarity scores. Here this service is most semantically close to the category "Book", then "Education", then "Library" and then "Entertainment" and so on. Thus by following the proposed approach better decisions regarding the classification can be taken.

TABLE 8:
RANKED LIST OF CATEGORIES WITH SIMILARITY DEGREES

Category	Similarity degree	Ranking
Book	4.6408	1
Education	4.1407	2
Library	4.0376	3
Entertainment	3.9844	4
Country	3.9659	5
Food	3.9137	6
Medicine	3.7481	7
Weather	3.4708	8
Weapon	3.1059	9
Automobile	2.3302	10

By observing the results, it has been analyzed that by merging the cognitive information through similarity scores from web based lexical network with the $TF-IDF$

scheme, better ordered set of categories based on semantic similarity degree is attained which helps in achieving effective automatic classification.

This automatic approach of classification reduce the manual efforts and helps in globally consistent decisions even in the presence of multiple users and reduces the ambiguities in the service classification area and results in better search options with in a class based on the semantic relatedness. Empirical evaluation have also shown that the proposed approach will also help in the identification of the relevant domain so that proper ontology can be discovered and hence semantic annotations to non semantic service profiles can be done in better and effective manner

TABLE 7:
MAPPING OF SERVICE PROFILES TO CATEGORY BASED DIMENSION VECTORS

Service No.	Library	Automobile	Food	Country	Weapon	Entertain ment	Book	Education	Medicine	Weather
37	1.2511	0.5739	1.3061	1.3532	0.8200	1.2551	1.7884	1.2774	0.9846	1.1168
38	2.9928	1.2586	3.1258	3.2661	1.9417	2.9956	4.6485	3.0999	2.3428	2.6304
39	1.9423	1.0245	2.2515	2.2636	1.3409	2.0064	3.1559	2.0085	1.4956	1.9490
40	4.1330	1.9885	4.4718	4.6216	2.9056	4.2023	5.9576	4.2331	3.2939	3.8584
41	4.8083	2.4794	5.0020	5.1277	3.4106	4.6505	6.5191	4.8949	3.8563	4.3845
42	2.9302	1.4645	3.3156	3.3418	1.9784	2.9296	4.3429	3.0188	2.2892	2.8449
43	2.8065	1.4803	3.2533	3.2707	1.9375	2.8991	4.5600	2.9020	2.1611	2.8161
44	3.0353	2.0697	3.3826	3.5022	2.3014	3.0870	4.3343	3.3091	2.4354	3.0401
45	3.5879	2.1341	3.4660	3.5381	2.8015	3.6285	4.1399	3.5545	3.2099	3.1429
46	3.7425	1.8932	3.8568	3.9228	2.6663	3.7026	4.7145	3.6972	3.1263	3.3593
47	2.5496	1.4802	2.7721	2.7275	2.0029	2.5327	3.1954	2.4695	2.2172	2.4599
48	4.0376	2.3302	3.9137	3.9659	3.1059	3.9844	4.6408	4.1407	3.7481	3.4708

V. CONCLUSION AND FUTURE WORK

With emergence of semantic web technologies, it is imperative that existing services be annotated with the semantic concepts from the suitable ontologies. Prior to the semantic annotation of services, the service profiles needs to classified appropriately to know the most appropriate domain a service belong to so that the requisite ontologies may be known. There is great need for the automatic classification mechanisms that can utilize the implicit semantic information of the services and classify them accordingly. This will help in the identification of the correct ontology and thus semantic annotations. This will also facilitate the retrieval, composition and interoperation of these services. In this paper, an approach for classification of web services is proposed which uses a lexical semantic network constructed from the web snippets as a knowledge base for the calculation of semantic similarity between the service profiles. For calculating semantic similarities between the service terms and dimensions, an unsupervised normalized similarity calculation approach having linear query complexity has been used. By considering the semantic relations from the lexical network, and merging them with the IR based $TF - IDF$ metric; we tried to propose an efficient mechanism for the classification of services based on the semantic similarities. Empirical evaluation has shown that this cognitive web based approach model gives ordered result set of domains based on semantic similarity scores which helps in better service classification and globally consistent classification decisions. This will facilitate effective and efficient retrieval of services. Apart from the service retrieval it can be used for the selecting the relevant ontologies from the domain so that non semantic service profiles can be annotated and mapped to the semantic concepts.

REFERENCES

[1] UDDI Technical White Paper, 2000. http://www.uddi.org/pubs/Iru-UDDI-Technical-White-Paper.pdf
[2] N. Shadbolt, T. B. Lee, W. Hall, The Semantic web revisited, IEEE Intelligent Systems, 21(3), 2006, pp. 96-101.
[3] 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/.
[4] D. Fensel, H. Lausen , A. Pollers, J. 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.
[5] 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/.
[6] J. Farrell, H. Lausen, Semantic Annotations for WSDL and XML Schema, 2007. http:/www.w3.org/TR/SAWSDL/.

[7] E. Iosif, A. Potamianos, 'SemSim: Resources for
[8] Lexical Networks', In: Proceedings of the Eight International Conference on Language Resources and Evaluation, 2012, Istanbul, Turkey.
[9] A. Patil, S. Oundhakar, A. Sheth, and K. Verma, METEOR-S Web service annotation framework, In Proc. of the 13th international conference on WWW. ACM Press, 2004.
[10] Z. Duo, L. Zi, and X. Bin, Web service annotation using ontology mapping, In IEEE International Workshop on Service-Oriented System Engineering, 2005, pp. 235–242.
[11] N. Oldham, C. Thomas, A. Sheth, and K. Verma, METEOR-S Web service annotation framework with machine learning classification, In Semantic Web Services and Web Process Composition, vol. 3387 LNCS, San Diego, CA, USA, 2004,Springer, pp.137-146.
[12] A. Heß, N. Kushmerick, Learning to attach semantic metadata to Web services. In Dieter Fensel, Katia P.Sycara, and John Mylopoulos, editors, International Semantic Web Conference, volume 2870 of Lecture Notes in Computer Science, Springer, 2003, pp. 258–273.
[13] M. Corella, P. Castells, Semi-automatic semantic-based Web service classification, In Business Process Management Workshops, volume 4103 of LNCS, Vienna, Austria, Springer, 2006, pp. 459–470.
[14] S. Batra, S. Bawa, Semantic discovery of Web Services using principal component analysis, Physical Sciences, 6(18), 2011, pp. 4466-4472.
[15] M. Crasso, A. Zunino, M. Campo, AWSC: An approach to Web Services classification based on machine learning techniques, CONICET, Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. 12(37), 2008, pp. 25-36.
[16] R. Mohanty, V. Ravi, M. Patra, Classification of Web Services Using Bayesian Network, In Journal of Software Engineering and Applications, 2012(5), pp. 291-296.
[17] I. Katakis, G. Meditskos, G. Tsoumakas, N. Bassiliades and I. Vlahavas, On the Combination of Textual and Semantic Descriptions for Automated Semantic Web Service Classification , Artificial Intelligence applications and innovations III,IFIP, Vol. 296, 2009, pp. 95-104.
[18] http://projects.semwebcentral.org/projects/owls-tc/.
[19] http://rapid-i.com/
[20] http://www.mathworks.in/
[21] http://www.telecom.tuc.gr/˜iosife/downloads.htm

Shailja Sharma is working as System Analyst in Kurukshetra University, Kurukshetra. She has done her M.E degree in Computer Sc. & Engineering from Thapar University, Patiala in 2009 and is pursuing Ph.D 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.

Mayank Dave holds a PhD in computer science and technology from IIT, Roorkee. In 1991, he joined the Computer Science and Engineering Department at National Institute of Technology, Kurukshetra, where he is presently working as a Professor. He has more than two decades of experience in teaching and research in the area of Normalized Semantic Similarity Computation using 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, the IETE, and the Institution of Engineers (India).

J.S. Lather is working as a professor at department of Electrical Engineering and Computer applications, National Institute of Technology, Kurukshetra. He holds a 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.