Natural Language Processing And Clustering Based Service Discovery

R.Gunasri, R.Kanagaraj

ABSTRACT: Web service used in Service Oriented Architecture creates a wide network of services that work together in order to execute complex task. Web services are defined throughweb pages having data and details only in natural languages it cannot use by computer to automatically process the information. To support automation of web service discovery there are many semantic based languages used for describing the process of the whole web service for the particular domain. The semantic web service description use ontologies to describe the performance of a web service. Natural Language techniques can help in better defining the background of a web services. This project Develop a Semantic web service discovery framework for finding semantic web services by making use of natural language processing techniques and clustering method. By make use of clusterTerms we can improve the optimization and eliminating irrelevant services and gives accurate service discovery.

Keywords:Web service discovery, Semantics, OWL-S,WSDL, ontology.

1. INTRODUCTION

Web services has become a important in service oriented architecture business components are decoupled with each other and used to implemented a expensive task. A huge number of this available web services gives importance to the accurate service discovery to satisfy user request. Services are discovered and published only through the web pages which are written in natural language human readable format. It cannot understand by machine to automatically process the data for search engine like Google, Yahoo. So we use Semantic Web Services (SWS) has semantically described the services and it has machine interpretable form. World Wide Web consortium(W3C) is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web related standards. But this standards only has syntax so the agents cannot understand the meaning of information exchange through this web page. Semantic web is a sharing information through the web.. Web services is a sharing application through web. Semantic webservice is a combination of semantic web and web services it uses ontologies to describe the web services.The semantic Web service descriptions use ontologies to describe the behavior of a Web service by applying reasoning over Web service semantics.

- *Mrs. R. Gunasri, PG Scholar, Sri Ramakrishna Engineering College, Tamil Nadu, India rgunasri2it@gmail.com*
- *Mr. R. Kanagaraj, Assistant Professor, Sri Ramakrishna Engineering College, Tamil Nadu, India anandkana@gmail.com*

Ontologies are created using human so it has natural language format. Service composition should be driven by people who know business processes and not by technicians. End user can able to discover these web services based on keyword written in human language. Service requester should discover these Web services based on keywords written in ordinary human language. A discovery process must be developed in such a way that gives link between keywords written in a natural language and Web service descriptions provided using semantically developed languages. This project we developed semantic web service discovery framework by make use of ClusterTerms. Here we are using hierarchal clustering to give optimization and accurate service discovery by eliminating irrelevant services. In this paper, the semantic web service discovery framework have developed it gets input of user request keyword and web service description files for web services to annotation are gathered. In first step the web service description files are extracted using service annotation reader to find the context of web services. This context are described using set of keywords by using wordnet this keyword are sensed. Each word assigned with different sense. It can be clustered using agglomerative hierarchal clustering. Semantic matching algorithm used to match this cluster with user request keyword cluster and then the web services are ranked according to the matching degree. Web service collection given to the user for the further process.

2. RELATED WORKS

The web services are automatically processed their interfaces are commonly defined in languages. In this section, a general survey is conducted for the different types of web service description languages presented and the different types of semantic web service discovery engines presented.

2.1 Web Service Description Languages

This section discussed with two type of mainly used ontology languages there are OWL-S, WSMO.OWL-S can allow individual software agents to dynamically discover, invoke, composeand maintaining Web services with a high degree of automation. OWL used for automated service discovery. OWL-S ontology consists of three main components: the services profile, the process model and the grounding. The services profile is for advertising and discovering Web services. The process model is used to describe detailed operations of services and define composite Web services. The grounding is used to map the abstract definition of services to concrete specifications of how to access the services.WSMO defines four models that support the development of semantic Web services namely Ontologies, Goals, Mediators and Web services [7]. Ontologies define a common terminology by providing concepts and relationships among the set of services for world domain. Goals are defined the expectations of a user when finding for a service on the following aspects functionality, approach and quality of service. Mediators coordinate the heterogeneity problem that occurs between descriptions at different levels [8] like data level – different terminologies, protocol level – different communication behavior between services, and process level—different business processes.

2.2 Web Service Discovery Engine

Two types of approaches for web service discovery, discovery based on clustering operation parameters and rich semantics language based approach. Clustering operation find the similarity among different web services. It search comparable and comprisable Web services as similar operations and services can be discovered based on operation parameters. Woogle (Dong, Halevy, Madhavan, Nemes, & Zhang, 2004) is a one of the cluster based Web service search engine that employs grouping operation parameters, and for a given query, it searches for similar andcomprisable Web service operations. Woogle uses semantics to automatically define the discovery process using operational parameters. This technique also applied in another search engine Seekda. It extract semantics similar service from WSDL file, It have runtime exchange of similar and composable services. Seekda is piece of the Service-Finder (Anonyms, 2009) framework, which is a platform for service discovery where information about services is gathered from various sources like Web pages and blogs. Another approach for semantic Web service discovery is the use of predefined ontologies. By identifying semantic similarities between ontologies, related semantic Web services can be discovered. GODO (Gomez, Rico, Garcia-Sanchez, Bejar, & Bussler, 2004)does not employ search for similar Web services, but instead employs a goal-driven approach. GODO has a repository with WSMO Goals and analyzes a user-specified goal (in natural language). The WSMO Goal with the highest match will be sent to WSMX (DERI Galway, 2008), which is an execution environment for WSMO service discovery and composition. Paulraj and Swamynathan (2012) proposed a methodfor content-based semantic Web service discovery. In different with general approach where user queries are matched againstOWL-S inputs, outputs, preconditions, and effects (IOPE), theframework allows users to submit free text as input. And it find the related services using wordnet disambiguation process to find the similar word search presented in aweb service.

3. OVERVIEW OF SERVICE DISCOVERY FRAMEWORK

Semantic web services are dependent on the availability of high quality ontologies. The SWSD framework requires a set of Web services that are described in semantic languages (e.g., WSMO (de Bruijnet al., 2005), WSMO-Lite (Vitvar et al., 2007), or OWL-S (Martinet al., 2004)). We have taking the two important semantic languages WSMO and OWL-S which has syntax in the format of WSML and OWL.The descriptions are subsequently analyzed, resultingin the extraction of words that could represent the context of theWeb services (i.e., the names of the operations, and nouns andverbs stated in non-functional descriptions of concepts or conditions).Next, the extracted words are disambiguated, as multiplesenses can be assigned to the same words. The disambiguated words are cluster using parameter ClusterTerms method then user query terms also clustered. Last this clusterTerms are matched using semantic matching algorithm, resulting in a ranked list of Web services.Ranking used to rank the web service according to the similarity matching degree.Find the web service which has the highest priority and given to the user.

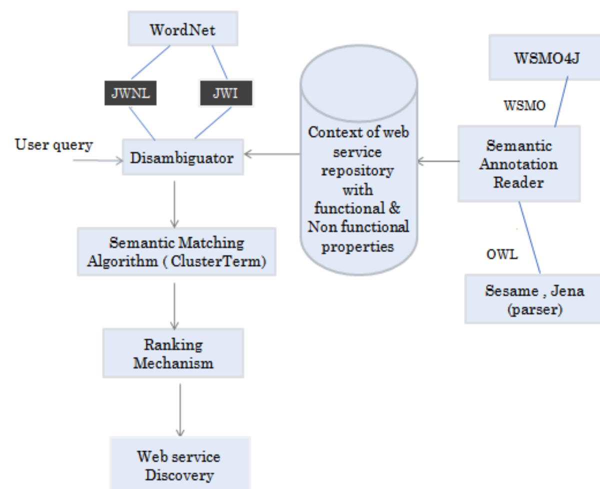


Fig 1: SWSD Architecture

3.1 SEMANTIC WEB SERVICE ANNOTATION READER

The readers should be able to parse a description and extract concepts, attributes, and relations from WSMO, OWL-S, etc. Therefore the first step in the model of searching for semantically described Web services is to implement readers for different languages and formats. A semantic Web service reader must be able to extract various elements out of a Web service description and its used ontologies. In the case of a WSMO Web service, names and non-functional descriptions of elements such as the capabilities and their sub elements, conditions and effects, help in understanding the context of the Web service. In the case of OWL-S file class names, subclass nonfunctional properties such as concepts has to be extract.By extracting these description on can thus establish the context of the web service description. Context is the words that defined the concepts of the web services. ForWSMO a WSML parser like WSMO4J

(Anonyms, 2008) can help in this process and for OWL-S and WSMO-Lite, which are written using an OWL (Bechhofer et al., 2004) or RDF (Klyne & Carroll, 2004) language, a parser like Sesame (Anonyms, 2009) or Jena (HP LabsSemantic Web, 2009) can be used.

3.2 WORD SENSE DISAMBIGUATION

This module has two type,

- Disambiguated sense of each word
- Sense Similarity

3.2.1 Disambiguated sense of each word

A user can represent its goal by defining a sets of words (nouns and verbs). Because many words have associated multiple meanings (e.g., mouse can be used to represent either an animal or a computer device), disambiguation of word senses helps in finding the correct context. The disambiguation will be applied to the set of words gathered from the user input and from a semantic Web service description, resulting in two sets of disambiguated senses that can be employed for matching. To find the sense this system uses wordnet. WordNet is like a dictionary it stores words and meanings. words in WordNet are arranged semantically instead of alphabetically. Synonymous words are grouped together to form synonym sets, or synsets. Each such synset therefore represents a single distinct sense or concept.It specify the senses of the word is being used in the given context. Lesk algorithm applied for finding word sense disambiguation. Given a word to disambiguate, the dictionary definition of each of its senses is compared to the senses of every other word in the phrase. A word is assigned that sense whose definition shares the largest number of words in common with the other words.

Algorithm:

1. Retrieve from MRD all sense definitions of the words to be disambiguated
2. Determine the definition overlap for all possible sense combinations
3. Choose senses that lead to highest overlap

3.2.2 Sense Similarity

This formula is proposed by Wu & Palmer, the measure takes into account both pathlength and depth of the least common sub-summer, the concept tree for a particular sysnset find from wordnet.

$$\mathrm{sim}(s,t) = 2 * \frac{\mathrm{depth(LCS)}}{[\mathrm{depth}(s) + \mathrm{depth}(t)]}$$

- where s and t: denote the source and target words being compared.
- Depth(s): is the shortest distance from root node to a node S on the taxonomy where the synset of S lies .
- LCS: denotes the least common sub-submer of s and t.

3.4 CLUSTERTERM

The words find in the previous steps are cluster using parameter clustering. The key ingredient of this technique is to cluster parameter names in the collection of web services into semantically meaningful concepts. We utilize the heuristic as the basis of our clustering, in that parameters tend to express the same concept if they frequently occur together. This allows us to cluster parameters by exploiting the conditional probability of their occurrences in the requests and responses of web service resources. Specifically, we are interested in the association rules of the form R:t1⇒t2, where t1 and t2 are two terms. Our clustering algorithm is a refinement of a classical agglomerative hierarchical clustering.

Algorithm 1: Finding cluster terms

1. clusterTerms = Ø
2. **for each**C_i in clusterSet
3. **if**Ci includes term **then**
4. **for all** terms in C_i
5. terms are added to clusterTerms
6. **endfor**
7. **endif**
8. **endfor**
9. return clusterTerms

3.5 SEMANTIC MATCHING ALGORITHM

We use a linear combination to combine the similarity of each component. Each type of similarity is assigned a weight based on the user's confidence. The overall similarity is calculated by using $\mathrm{Similar}_{des}$,$\mathrm{Similar}_{req}$, and $\mathrm{Similar}_{res}$:

$$\mathrm{similarit} = \frac{\alpha(\mathrm{similar}_{des}) + \beta(\mathrm{similar}_{req}) + \gamma(\mathrm{similar}_{res})}{\alpha + \beta + \gamma}$$

Where $\mathrm{Similar}_{des}$,$\mathrm{Similar}_{req}$, and $\mathrm{Similar}_{res}$ are the description, request, and response similarity, respectively. These components return a real value between 0 and 1, indicating the degree of similarity. The weights α, β and γ are real values between 0 and 1; they indicate the degree of confidence. High weight values indicate the user's confidence. Please note that the sum of the weights does not have to add up to 1; in our experiments we use equal weights. Our semantic matching algorithm based on the clustering technique can improve the recall performance of the search engine by introducing semantically meaningful concepts. All matches whose similarity scores exceed 0 are assigned to a candidate set. To select the best matches from the candidate set, however, an additional ontological pruning step is required. Since the parameter clustering technique considers all terms in a cluster as an equivalent concept and ignores hierarchicalrelationships between the terms, matches might exist that are irrelevant to the user's intention (i.e., false positives). A query is matched against all resources stored in the repository using pre defined matching rules. Our algorithm proceeds in a Greedy fashion. If two concepts are matched to the rules, the weight is set to 1. If two concepts are not matched to the rules, then the weight is set to 0 and they are removed from the results. The core procedure for the semantic matching algorithm is shown in Algorithm 2.

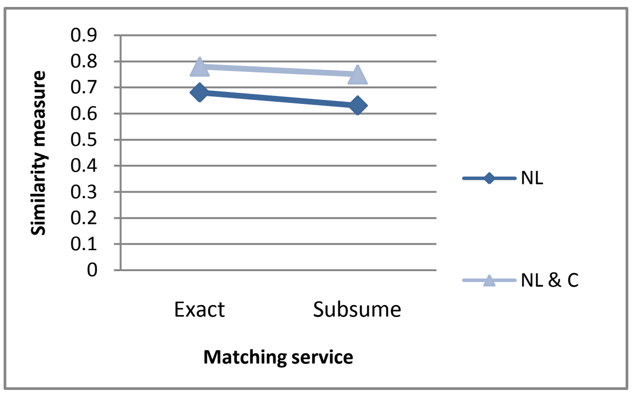
Algorithm 2: Semantic matching

1. **for each** resource S in repository
2. Compute $\mathrm{Similar}_{des}$=Sim(Q_{des}, S_{des})
3. **for each** term in query request/response

4. Get clusterTerms from Algorithm 1
5. Replace term with culsterTerms
6. **endfor**
7. Perform pruning process
8. Compute Similarreq and Similarres
9. Compute Similarity
10. **endfor**

3.6 RANKING AND WEB SERVICE DISCOVERY
User query are matched with web service cluster using semantic matching algorithm if it matched it will ranking the web services according to the matching weight degree. Then the accurate web service is given to the user.

4. EXPERIMENTAL RESULT
To test the performance of natural language processing with clustering based method we using 20 types of user queries matched with 30 type of different web service domain. Each queries can matched with minimum of one web services. Below graph gives the performance evaluation when we applied natural language techniques only it gives low performance. If this model used with clustering it gives the high performance.



According to the above graph the performance is based on accurate matching of Web Services.

Exact – Exact matching of web service.
Subsume – Some of the operations of particular web service domain matched with user request keyword.

NL - Natural Language Technique
NL & C - Natural Language Technique with Clustering

5. CONCLUSION
Compare with related service discovery methods clustering based natural language processing techniques gives a high precision and recall. Natural language processing used keyword matching with context of service description. It has accurate matching because Wordnet gives an exact sense for a particular web service domain. Clustering techniques gives optimization. It can eliminated the irrelevant services.

6. REFERENCES
[1] Jordy Sangers, Flavius Frasincar, Frederik Hogenboom,Vadim Chepegin, Semantic Web service discovery using natural language processing techniques, Expert Systems with Applications 40(2013) 4660–4671.

[2] Bener, A.B, Ozadali, V., & Ilhan, E. S. (2009). Semantic matchmaker with precondition and effect matching using SWRL. Expert Systems with Applications, 36(5), 9371–9377.

[3] Dietze,S., Gugliotta, A.,& Domingue, J.(2009). Exploiting metrics for similarity- based semantic Web service discovery. In IEEE seventh international conference on Web services(ICWS2009)(pp.327–334). IEEE Computer Society.

[4] Navigli,R.,& Velardi,P.(2005).Structural semantic interconnections: A knowledge- based approach to word sense disambiguation. IEEE transactions on pattern analysis and machine

[5] intelligence(Vol. 27,pp.1075–1086). IEEE Computer Society.

[6] Dr.Yongju Lee (2013).Semantic Matching and Resource DiscoveryAlgorithms for RESTful Web Services. International Journal of Innovative Research in Computer and Communication Engineering(Vol. 1, Issue 9),2320-9801.

[7] D. Roman, U. Keller, H. Lausen, J. de Bruijn, R. Lara, M. Stollberg, A. Polleres, C. Feier, C. Bussler, D. Fensel, Web service modeling ontology, Applied Ontology 1 (1) (2005) 77–106.

[8] I. Toma, D. Foxvog, M.C. Jaeger, Modeling QoS characteristics in WSMO, in: The Proceedings of the 1st Workshop on Middleware for Service Oriented Computing, MW4SOC'06, ACM, New York, NY, USA, 2006, pp. 42–47.

[9] Papazoglou M.P, Georgakopoulos D, (2003), "Service oriented computing", Proc. Communications of the ACM 46 (10), pp. 25–28.

[10] Shi Zhong-Zhi, JIANG Yun-Cheng, ZHANG Hai-Jun, DONG Ming-Kai, (2004), "Agent service matchmaking based on description logic", Proc. Chinese Journal of Computers 27 (5), pp. 625–635.

[11] Yoo Donghee, (2012), "Hybrid query processing for personalized information retrieval on the Semantic Web", Proc. Knowledge-Based Systems 27, pp. 211–218.

[12] Ying-Long Ma, Bei-Hong Jin, Yu-Lin Feng, (2005), "Dynamic discovery for semantic web services based on evolving distributed ontologies", Proc. Chinese Journal of Computers 28 (4), pp. 603–614.

[13] Zhou C, Chia L.T, Lee B.S, (2005), "Service discovery and measurement based on DAML-QoS ontology", Proc. International World Wide Web Conference, pp. 1070–1071.