AN EFFICIENT ONTOLOGY BOOTSTRAPPING APPROACH FOR VALIDATING WEB SERVICES

Kodhai E[1], Yuvaraj B[2], Thirumal K[3], Abijith V[4]

[1]*Associate Professor, Sri Manakula Vinayagar Engineering College, Pondicherry-605107, India*

[2, 3, 4] *Dept. Of Information Technology, Sri Manakula Vinayagar Engineering College, Pondicherry-605104, India*

[1]kodhaiej@yahoo.co.in
[2]yuva7790@gmail.com
[3]thiruthiru19@gmail.com
[4]abijith.blaze@gmail.com

Abstract— **Ontology defines the relationship between the entities. Bootstrapping ontology is the generation of concepts and connection between the services for the domain. Web services usually consist of both WSDL and free text descriptors. The WSDL descriptor is validated using two techniques, namely Term Frequency/Inverse Document Frequency (TF/IDF) and web context generation. Our proposed bootstrapping ontology process integrates both methods and applies a third method to validate the concepts using the service free text descriptor.**

Keyword- **Term Frequency (TF), Inverse Document Frequeny(IDF), Web services discovery, service-oriented relationship modeling**

I. INTRODUCTION

Ontologies are used in many applications mainly in semantic web and it is preferred modeling tool. However the design, implementation and maintenance of ontologies is a difficult process. Ontology bootstrapping which is a recently developed technology for ontology creation, involves in automatic generation of concepts and association between the concepts for a domain. Past works on ontology bootstrapping gives attention to a limited domain or expanding an existing ontology. In web sevices, the registries such as universal Description, Discovery and integration (UDDI) is created for interoperability of web services. But UDDI registries have some errors. It stores only the partial description of available services. But creating a ontology gives a solution. The web service ontology Bootstrapping process proposed in this paper is based on the Web Service Description Language (WSDL) describing “how” the service should be used and a textual description of the web service in free text describing “what” the service does .This method analyzes the web services using three methods, each method describes in different point of view of web services. As a result, this process gives a more precise description of ontology and gives better results. And by using a three methods we are going to validate the web services. In particular, the Term Frequency/Inverse Document Frequency (TF/IDF) method analyzes the web service from an inner point of view. The Web Context Extraction method describes the WSDL document and returnes proper description for the web search queries. Atlast the the free text description verification method is used to solve the inconsistencies with current ontology. On a conceptual level, we introduce an ontology bootstrapping model, a model for automatically generating the concepts and relations “from scratch”. On an algorithmic level, we provide an implementation of the model in the web service domain using integration of two methods for implementing the ontology construction and a Free Text Description Verification method for validation using a different source of information. On a practical level, we validated the implementation and benefits of our method using a set of real-world web services. Given that the task of designing and maintaining ontologies is still so difficult, our approach presented in this paper can be valuable in practice. Finally the ultimate aim of this paper is to validate the web services and these validation helps the user to know which service is updated .

II. RELATED WORKS

A. *Web Service Annotation*

Several works have been done providing assistance to our research in the field of automatic annotation of web services.In the beginning a combined approach was used by integrating several matchers on which the web services depend on[1]. The matchers which are used in that combined approach are string , structural matcher with that of synonym finder, these all are combined by a simple aggregation function[6]. Then a technique has been evolved to perform for semantic annotation which integrated all the results into WSDL. Next came the Naive Bayesian Classifier ,which is a Machine Learning(ML) technique improved service annotation a more accurate manner than previous approaches[9].ML technique is also further improvised for new annotations of semantic web services, called Assam. Existing ontology was altered by categorising, mapping and ranking those web services according to the service compositions that are made. However, the altered one had failed to provide clear semantic mapping to existing ontologies..

B. Ontology Creation And Evolution

The works undergone recently were on creation and evolution of ontology , that too particularly on schema matching. Many theoretical models were proposed which involved mapping relationships using upper ontologies , by calculating automatic semantic reconciliation and evolving relationships between the ontologies itself[16]. The Information Science domain has provided a wide role in ontology creation. Then many approaches evolved to make ontology management and its learning such as DOGMA project(which gives an engineering approach), Text-To-Onto , Thematic Mapping, and TexaMiner. Later the researchers evolved the interoperability of the ontology system which lead to the introduction of two systems : Chimaera and Prote`ge`[3]. But these works had a drawback of requiring manual help for ontology construction. With the help of domain specific websites and digital library collections, the ontology evolution has been shaped up[4]. Ontology has been evolved by the bootstrapping approach of knowledge acquisition in the domain media and visual media and another work used reusing ontologies and language components. The instance data's effects were defined in the sets of ontology-change operations[11]. The previous works were totally dependent on ontology or domain specific where as our recent work evolves the ontology from the beginning.

C .Ontology Evolution On Web Service

Ontology evolution has been suggested as one among the future directions of research based on the surveys on ontology techniques implementations to the semantic web and service discovery approaches[12]. Based on Natural Language Processing (NLP), the ontology learning tools for semantic web service have been constructed.In order to overcome the drawbacks due to NLP, the research has been moved to next level by context on directed ontology learning.A survey based on the state-of the-art from which the information is retrieved has suggested that analysis of the web service textual description in addition to the WSDL description will be more useful than analysis of each separate descriptor[13].The above survey also pointed out the drawback of low recall in the existing ontology, which has been overcome in our recent work by usage of web context recognition.

D. Ontology reuse

For many years ontology reuse has been examined. Reusable ontology repository has been focused and studied in the prior analysis."Start -to- finish process " is applied in small scale application. The author concludes that re-using the 0ntology is very far process when compare to automated process[17,5]. Ontnology is mainly based on semantic web research, many ontology has been created and applied in real time applications. In ontology there are two types of study and they are theoretical study and practical study[19]. The management of heterogeneous ontology is difficult by simple repositories.

E. Information Retrieval:

In this paper the formal concept analysis (FCA) and a notion of information anchors is clubbed to increase the efficiency of the delivery to the end user[20]. Here attribute concepts uses ranked objects to facilitate topical queries. The domain specific Concept is bootstrapped using formal concept analysis.

F. Ontology-Driven Information Systems

Information system (IS) has ontology of its own. It refers to the symbols used to view the world in a particular way. All aspects and all components of an IS are profitably Driven by the ontology[18]. An IS consists three different types of components namely application programs and information resources similar to databases, user interfaces. In order to accomplish a strong business purpose the three components are integrated[19] .In IS we can distinguish two orthogonal Dimensions namely a temporal dimension, which concerns whether an ontology is used at development time or at run time and a more structural dimension, which concerns the particular way an ontology can affect the main IS components[20].

III. THE BOOTSTRAPPING ONTOLOGY MODEL APPROACH

A.Introduction

Our proposal is to provide validation to the web service, to check that the particular domain is updated or not with the help of registries available in the document descriptor available for each web service. The first step Data Extraction retrieves the registries for the given domain. The second step Token Extraction extracts tokens from the WSDL document. The next process Term frequency/Inverse document frequency analyze the document and shows most commonly used terms in web service and appearing less frequently in other service documents. The next method web context Extraction use the tokens as a query to a web search engine and returns the results based on descriptors.The last step Ontology Evolution identifies the newly arrives concepts and changes the relation .And provides the relationship between the concepts. Our Bootstrapping ontology model approach is described in fig.1.There are five major steps in the process.

Our proposal is carried out with five modules , and they are as follows,

1. Data Extraction.
2. Token Extraction.
3. Term Frequency/IDF analysis.
4. Web Context Extraction.
5. Ontology Evolution.

B. Data Extraction:

In this module we develop the data extraction process using Whois. Whois is a Web service that allows domain details to be identified based on the domain name .It maintains a web services related with operations and services.

C .Token Extraction:

In this module we develop the token extraction process using WSDL (Web Service Description Language).Each token is textual document. It is extracted by documentation of the web service.Each token must contain a meaning. The extracted token list serves as a base. These tokens are extracted from the WSDL descriptor of a Web service Whois. The service is used as a starting process for developing the ontology. Additional more services are used later to explain the process of expanding and defining the ontology. The tokens are extracted based on capital letter of each word. The tokens are filtered based upon stop words.

D. Term Frequency/IDF Analysis:

Term Frequency/Inverse Document Frequency analysis is for generating a set of representative keywords from a mass set of documents. This method is applied for WSDL descriptor. In this inappropriate terms are more unique and it can be put away with a higher confidence. The TF/IDF is defined by freq($t_{i,}d_i$),it is calculated by number of times the token t_i is occurred in the document descriptor $d_{i.}$ The inverse document frequency is calculated by the ratio between the number of documents that contains the term and the total number of documents.The weight of the document is calculated by both Term frequency and inverse document frequency(w=tf*idf) which is illustrated in Fig.1.The token weight is used to give ranking for the descriptor tokens.This ranking filters out the tokens that have higher value.The tokens are removed by filtering process and it is identified by the tokens that contains high IDF value is removed.

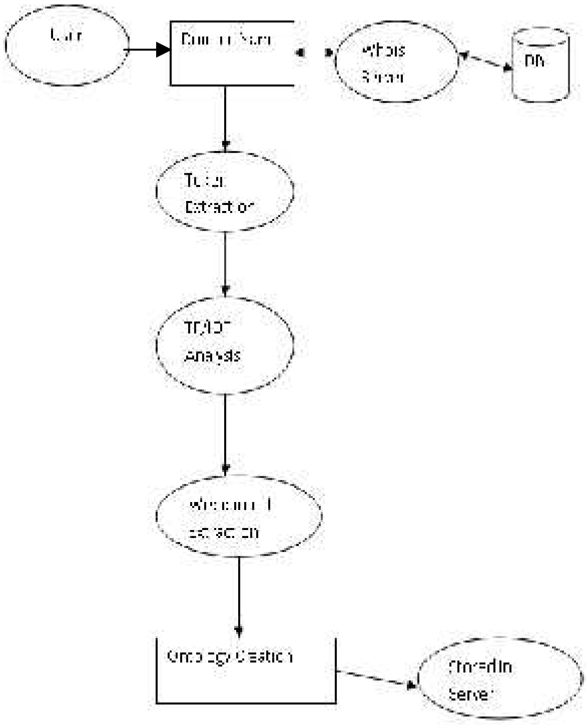

Fig.1. Architecture of Bootstrapping ontology model approach

E. Web context extraction:

In this module, we develop the web context extraction process This module identifies the record of information from a web services. It contains word, phrase or alphanumerical term. A weight w finds the significance of descriptor in the web service. By finding all the information and descriptors we find the context. The input to this method is set of tokens retrieved from the web service WSDL descriptor. The web pages clustering algorthim is used. The algorithm then calculates the number of web pages that finds the same descriptor and the sum of references in the WSDL.

F. Ontology Evolution:

There are four steps in the ontology evolution, they are building new concepts, determining the concept relations, identifying relation types and resetting the process for the next WSDL document. The possible identified concept pays way for building a new concept. There is no guarantee for the evocation of a concept should be integrated with previous step and current ontology. The textual service descriptor is used for validating the descriptor. The are two separate descriptions in web service they are WSDL description and textual description which is taken as an advantage for the analysis. In the verification process the matching of concept descriptors against all the service textual descriptors is done. TF/IDF is used for identifying the descriptor domain. The web context result are matched with a textual descriptor. Overlapping context descriptors between different concepts is used for analyzing the relation between the concepts.TF/IDF and the web context methods are used for extracting descriptors that were included in the union of descriptors.

```
For each domain
Search Who is Server
Extract tokens using filtering based on WSDL
Apply IDF algorithm to $\mathbf{D}_{\mathbf{WSDL}}$
Apply TF algorithm
Result=Apply Webcontext algorithm
Apply Wt to WSDL
Result$_{\mathbf{Wt}}$=TF*IDF
```

Fig.2 Concept Generating Algorithm

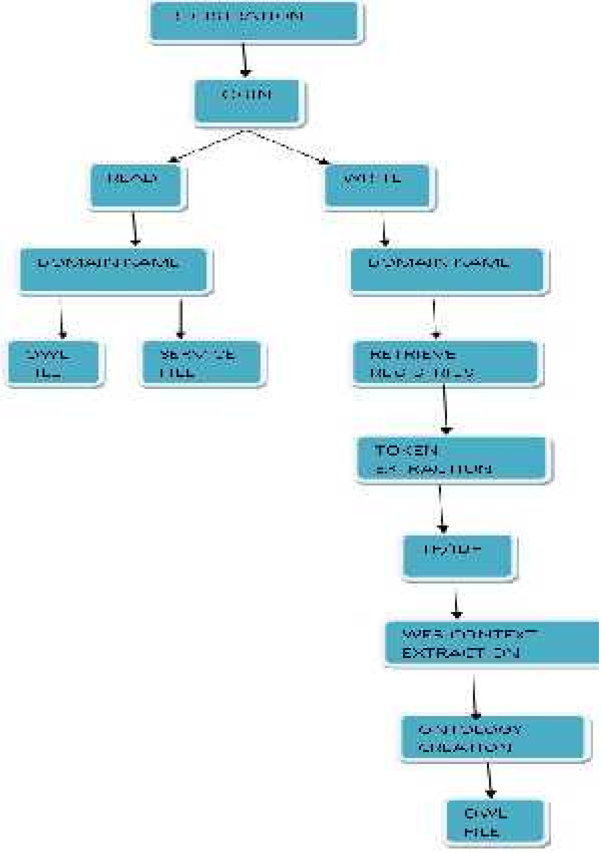

Fig.3. Work flow of Bootstrapping ontology model approach

IV. CONCLUSION

In this paper by calculating the TF/IDF value and web context extraction and by applying the bootstrapping ontology we validate the web services. This makes the user to know about the services whether it is updated or not. we validated the web services by integrating the results of all methods. Our work flow which carries our problem and working of our project is illustrated in figure .3. This approach develops the automatic construction of ontology This method is applicable for all domains. Our further idea in ontology bootstrapping is to increase the efficiency and also to implement automatic generation of results.

REFERENCES

[1] Aviv Segev, Member, IEEE, and Quan Z. Sheng, Member, IEEE," Bootstrapping Ontologies for Web Services" March 2012.

[2] D. Kim, S. Lee, J. Shim, J. Chun, Z. Lee, and H. Park, "PracticalOntology Systems for Enterprise Application," 10th Asian Computing Science Conf. (ASIAN '05), 2005.

[3] M. Ehrig, S. Staab, and Y. Sure, "Bootstrapping Ontology Alignment Methods with APFEL," Fourth Int'l Semantic Web Conf. (ISWC '05), 2005.

[4] G. Zhang, A. Troy, and K. Bourgoin, "Bootstrapping Ontology Learning for Information Retrieval Using Formal Concept Analysis and Information Anchors," 14th Int'l Conf. Conceptual Structures (ICCS '06), 2006.

[5] S. Castano, S. Espinosa, A. Ferrara, V. Karkaletsis, A. Kaya, S. Melzer, R. Moller, S. Montanelli, and G. Petasis, "Ontology Dynamics with Multimedia Information: The BOEMIE Evolution Methodology," Int'l Workshop Ontology Dynamics (IWOD '07), held with the Fourth European Semantic Web Conf. (ESWC '07), 2007.

[6] C. Platzer and S. Dustdar, "A Vector Space Search Engine for Web Services," Third European Conf. Web Services (ECOWS '05), 2005.

[7] L. Ding, T. Finin, A. Joshi, R. Pan, R. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, "Swoogle: A Search and Metadata Engine for the Semantic Web," 13th ACM Conf. Information and Knowledge Management (CIKM '04), 2004.

[8] A. Patil, S. Oundhakar, A. Sheth, and K. Verma, "METEOR-S Web Service Annotation Framework," 13th Int'l World Wide Web Conf. (WWW '04), 2004.

[9] Y. Chabeb, S. Tata, and D. Belad, "Toward an Integrated Ontology for Web Services,". Fourth Int'l Conf. Internet and Web Applications and Services (ICIW '09), 2009.

[10] Z. Duo, J. Li, and X. Bin, "Web Service Annotation Using Ontology Mapping," IEEE Int'l Workshop Service-Oriented System Eng. (SOSE '05), 2005.

[11] N. Oldham, C. Thomas, A.P. Sheth, and K. Verma, "METEOR-S Web Service Annotation Framework with Machine Learning Classification," First Int'l Workshop Semantic Web Services and Web Process Composition (SWSWPC '04), 2004.

[12] A. Heß, E. Johnston, and N. Kushmerick, "ASSAM: A Tool for Semi-Automatically Annotating Semantic Web Services," Third Int'l Semantic Web Conf. (ISWC '04), 2004.

[13] Q.A. Liang and H. Lam, "Web Service Matching by Ontology Instance Categorization," IEEE Int'l Conf. on Services Computing (SCC '08), pp. 202-209, 2008.
SEGEV AND SHENG: BOOTSTRAPPING ONTOLOGIES FOR WEB SERVICES 43

[14] A. Segev and E. Toch, "Context-Based Matching and Ranking of Web Services for Composition," IEEE Trans. Services Computing, vol. 2, no. 3, pp. 210-222, July-Sept. 2009.

[15] J. Madhavan, P. Bernstein, and E. Rahm, "Generic Schema Matching with Cupid," Int'l Conf. Very Large Data Bases (VLDB), pp. 49-58, Sept. 2001.

[16] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web," 11th Int'l World Wide Web Conf. (WWW '02), pp. 662-673, 2002.

[17] A. Gal, G. Modica, H. Jamil, and A. Eyal, “Automatic Ontology Matching Using Application Semantics,” AI Magazine, vol. 26, no. 1, pp. 21-31, 2005.

[18] J. Madhavan, P. Bernstein, P. Domingos, and A. Halevy, “Representing and Reasoning about Mappings between Domain Models,” 18th Nat’l Conf. Artificial Intelligence and 14th Conf. Innovative Applications of Artificial Intelligence (AAAI/IAAI), pp. 80- 86, 2002.

[19] V. Mascardi, A. Locoro, and P. Rosso, “Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation,” IEEETrans. Knowledge and Data Eng., doi:10.1109/TKDE.2009.154, 2009.

[20] A. Gal, A. Anaby-Tavor, A. Trombetta, and D. Montesi, “A Framework for Modeling and Evaluating Automatic Semantic Reconciliation,” Int’l J. Very Large Data Bases, vol. 14, no. 1, pp. 50-67, 2005