A New Approach to Ontology Evolution

Sushma Sonarkar, Prof Sumit Bhattacharjee

Abstract

Ontologies have become the effective modeling for various applications and significantly in the semantic web. One of the latest methods for construction of Ontology is ontology evolution which focuses on automatic generation of concepts and their relations in a given area. The difficulty with numerous, largely dissimilar concepts must be handled by the construction of ontology which depends on the collection of already defined textual sources like web services.
In this paper we propose a new approach that bootstraps domain ontologies from WSDL descriptions. The approach uses the technique of information retrieval and machine learning

The web service consists of two components the web service descriptor and the free text descriptors. The WSDL descriptor is evaluated using two methods basically term frequency and inverse document frequency (TF/IDF) and web context generation methods. It also develops the ontology hierarchy by using TreeGenerator function.

Keywords: Ontology, semantic web, web service, WSDL.

I. Introduction

Ontologies are used in large no of applications, remarkably the Semantic web, and essentially have become the preferred modeling tool, but the design and maintenance of ontologies is a formidable process. Ontology construction is a significant technology for the construction of ontology. It includes automatic recognition of concepts significant to a specified area and the relations between the concepts [13].

Ontology is the task of building computable models of some domain for some purpose and it is a specification of conceptualization. To be successfully used within information systems, ontologies should provide clear semantics for their concepts and a rich formalization of their semantics. These requirements are particularly important in the area of semantic web services whose ontological descriptions should be used for reasoning tasks.The Artificial-Intelligence literature contains many definitions of ontology; many of these contradict one another. For the purposes of this guide an ontology is a formal explicit description of concepts in a domain of discourse (classes sometimes called concepts), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties), and restrictions on slots (facets (sometimes called role restrictions)). Ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins. Classes are the focus of most ontologies. Classes describe concepts in the domain.

ISSN: 2278 – 1323

A. *Motivation*

The basic problem in web service is largely unrelated concepts. Literature works on ontology bootstrapping done with only limited domains. Since UDDI registries are not based on lim- ited domain, it has dynamic registration of web services by various business concerns in world. The second problem is that ontologies are created by expanding the existing ontolo- gy. Due to which new concepts cannot be identified and also memory is wasted. The already created concepts may be out dated but still it is not deleted. Some concepts may need only little updates based on advanced concepts. But the previous work will create new ontology itself instead of little update in existing ontology.

II. Literature Survey

H. Davulcu, S. Vadrevu, presented automated techniques for bootstrapping and populating specialized domain ontologies by organizing and mining a set of relevant Web sites provided by the user [1]. They develop algorithms that detect and utilize HTML regularities in the Web documents to turn them into hierarchical semantic structures encoded as XML. Next, they present tree-mining algorithms that identify key domain concepts and their taxonomical relationships

.Noy and Klein defined a set of ontology-change operations and their effects on instance data used during the ontology evolution process [2].

A context-based semantic approach to the problem of matching and ranking web services for possible service composition is suggested by A. Segev and E. Toch [8]. They present an analysis of different methods for classifying Web services for possible composition and present a context-based semantic matching method for ranking these possibilities.

A.Patil, S. Oundhakar, A. Sheth, and K. Verma "METEOR-S Web Service Annotation Framework,"[3] presents a combined approach toward automatic semantic annotation of web services. They have developed algorithms to match and annotate WSDL files with relevant ontologies. They use domain ontologies to categorize Web services into domains. The approach relies on several matchers (e.g., string matcher, structural matcher, and synonym finder), which are combined using a simple aggregation function.

The work proposed in [10] uses a clustering algorithm, Quality Threshold (QT), to cluster Web services into functionally similar service groups. The similarity between two services is measured by comparing the elements in WSDL documents, including service names, complex data types, messages, portTypes. as well as terms.

In [12], a co-clustering approach is proposed to generate Web service communities based on WSDL descriptions. The approach improves the precision and recall of community generation by clustering Web services and operations together. It builds up a service matrix and an operation matrix based on their term TF/IDFs.

ISSN: 2278 – 1323

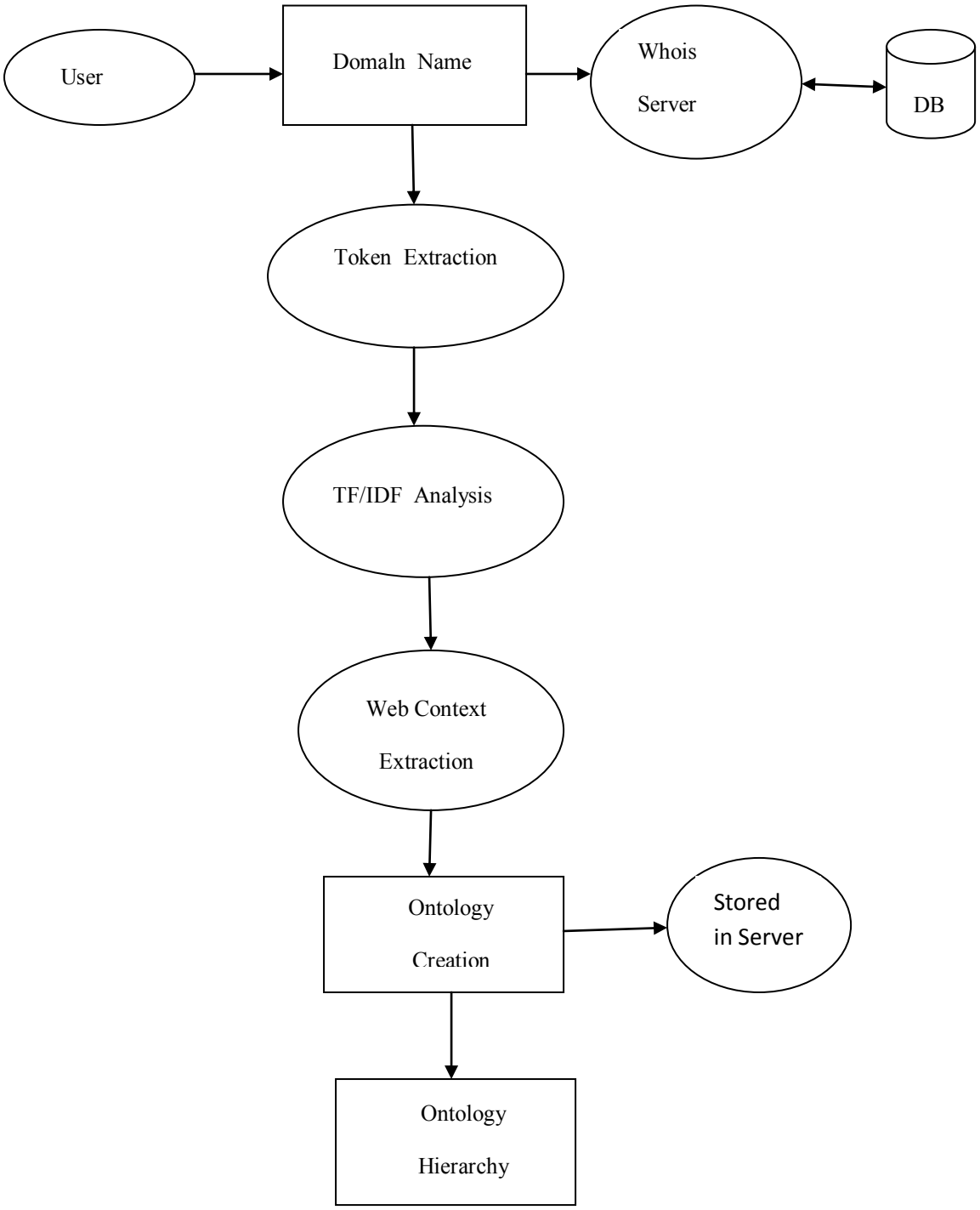

Fig 1. Ontology Construction Process

ISSN: 2278 – 1323

III. The Proposed System Architecture

We propose a new approach that bootstraps domain Ontologies from WSDL descriptions. It first constructs the ontology and then develop the internal structure of the ontology using a TtreeGenerator function .Fig 1 shows the Ontology Construction process.

A. *Token Extraction*

First step to construct domain ontology is token extraction where the *descriptor* represents the set of tokens for a web service S. The tokens extracted are textual terms from the WSDL document. The extracted tokens serves as the baseline.

B. *TF/DF Analysis*

In TF/IDF analysis process, the most common terms appearing in each web service document and appearing less frequently in other documents are analyzed. TF/IDF weight is a statistical measure used to evaluate how important a word is to a document in a collection.
To formally define TF/IDF, we start by defining freq(ti,Di) as the number of occurrences of the token ti within the document descriptor Di.

$$tf(ti) = \frac{freq\,(ti,Di)}{|Di|} \qquad (1)$$

The inverse document frequency is computed as the ratio between the total number of documents and the number of documents that contain the term:

$$idf(ti) = \frac{\log |D|}{|\{Di : ti \in Di\}|} \qquad (2)$$

Using the token weight the descriptor tokens can be ranked. The filtering of the tokens is done by ranking using a threshold value.

C. *Web Context Extraction*

The context extraction process is defined as tokens extracted from the web service WSDL descriptor. The set of tokens is then sent to a web search engine and a set of descriptors is extracted by grouping the web pages search results for each token set.

D. *Ontology Evolution*

On refining the possible identified concepts, new concept can be build. The evocation of a concept in the previous step does not guarantee that it should be integrated with the current ontology. Instead, the new possible concept should be analyzed in relation to the current ontology.
The external web service textual descriptor serves as a moderator if there is a conflict between the current ontology and a new concept. New concepts can be checked against the free text descriptors to verify the correct interpretation of the concept.
The OWL file gets created once the Ontology creation is over Web Ontology language (OWL)is a family of knowledge representation languages or ontology language for authoring ontologies. The languages are characterized by formal semantics and RDF/XML based serializations for the semantic web.

E. *Ontology Hierarchy*

The TreeGenerator function is used to create the internal structure of the Ontology. The TreeGenerator accepts the Datatable in the constructor, with the node list to create the structure of the tree. The

ISSN: 2278 – 1323

data in the Datatable is obtained from the database.

F. Experimental Analysis

The user here can enter any domain name and then the token extraction is done for that particular website .After the TokenExtraction the term frequency is calculated by using TF/IDF method. Then by using both the TF/IDF and WebExtraction method the Ontology is evolved and the OWL file gets created which represents the Ontology. We have tested this for many different domains and the Ontology as well as their Ontology hierarchy's are generated successfully. Our experiment results indicate high precision. Thus our approach leads to dynamic ontology creation.

IV. Conclusion

We present a new approach to bootstrap domain ontologies for Web services. First the ontology construction is done which is based on TF/DF and web context extraction methods. Then the ontology hierarchy is displayed using TreeGenarator function. The hierarchical ontology framework captures more information than a flat ontology structure.. Higher accuracy can be achieved using this ontology hierarchy based approach rather than keywords based approach. Ontology construction and maintenance effort can be substantially reduced.

References

[1] H.Davulcu, S.Vadrevu, S.Nagarajan, and I. Ramakrishna, "OntoMiner: Bootstrapping and Populating Ontologies from Domain Specific Web Sites," 2003.

[2] N.F. Noy and M. Klein, "Ontology Evolution: Not the Same as Schema Evolution," 2004.

[3] A.Patil, S. Oundhakar, A. Sheth, and K. Verma, "METEOR-S Web Service Annotation Framework," 2004

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

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

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

[7] R. L. Cilibrasi and P. Vitanyi. The google similarity distance. IEEE Trans. on Know!. and Data Eng., 19:370-383, March 2007.Composition" 2009

[8] Aviv Segev and Eran Toch, "Context-Based Matching and Ranking of Web Services for Composition" 2009

[9] P. Plebani and B. Pernici. URBE: Web service retrieval based on similarity evaluation. IEEE Transactions on Knowledge and Data Engineering, 21:1629-1642, 2009.

[10] K. Elgazzar, A.E. Hassan,and P. Martin, Clustering wsdl documents to bootstrap the discovery of web services. In ICWS 2010, pages 147-154,2010

[11] F. Liu, Y. Shi, J. Yu, T. Wang, and 1. Wu. Measuring similarity of web services based on wsdl. In ICWS 2010, pages 155 -162, 2010.

[12] Q. Yu and M. Rege. On service community learning: A co-clustering approach. In ICWS 2010, pages 283-290, 2010.

[13] Aviv Segev, Member, IEEE, and Quan Z. Sheng, Member, IEEE, "Bootstrapping Ontologies for Web Services", IEEE Transactions On Services Computing, Vol. 5, No. 1, January-March, 2012.

ISSN: 2278 – 1323

Authors
Sushma G. Sonarkar
M. E Student
Computer science & engineering,
MGM's College of engineering
&Technology, Navi Mumbai.

Prof. Sumit Bhattacharjee
Computer science & engineering,
MGM's College of engineering
&Technology, Navi Mumbai.

ISSN: 2278 – 1323