A Semantic web: Intelligence in Information Retrieval

Jenice Aroma R, Mathew Kurian
Department of Computer Science and Engineering,
Karunya University, India
jenicearoma@karunya.edu, mathewk80@karunya.edu

***Abstract*— The rapid growth of web resources lead to a need of enhanced Search scheme for information retrieval. Every single user contributes a part of new information to be added to the web every day. This huge data supplied are of diverse area in origin being added, without a mere relation. Hence, a novel search scheme must be applied for bringing out the relevant results on querying web for data. The current web search scheme could bring out only relevant pages to be as results. But, a Semantic web is a solution to this issue through providing a suitable result on understanding the appropriate need of information. It can be acquired through extending the support for databases in machine readable form. It leads to redefinition of current web into semantic web by adding semantic annotations. This paper gives an overview of Semantic mapping approaches. The main goal of this paper is to propose the steps for bringing out a new Semantic web discovery algorithm with an efficient Semantic mapping and a novel Classification Scheme for categorization of concepts.**

***Index Terms*— Semantic mapping, OWL-S, WSDL**

I. INTRODUCTION

THIS paper clearly illustrates the purpose of Semantic web in bringing out an efficient result for web search schemes. The current web admits a syntactic mode of discovery in searching results. It is based on the keyword query being supplied to search the data in search engines, [3]. Normally, the keywords supplied with the query could show only the relevant pages as results. These relevant pages carry only a spotted keyword in query to be as a key for relevance. Most of the pages retrieved as results are of different sense and can never be the expected results.

It brings a notion in the mind of Tim Berners Lee, Director of World Wide Web Consortium (W3C) to frame an innovative idea over Semantic web. Though many years have gone, still semantic web is in a growing phase. It is impractical to frame Semantic web to be as an entirely new web. The conventional web today is full of unstructured content and It doesn't favour machine interaction. Hence, a feasible solution of redefining the conventional web through addition of Semantic Annotations be done.

This act of redefining the semantic web is termed for Semantic mapping and it must be done as an automated task. Since, manual mapping of huge concepts in web is highly tedious. It is due to the consumption of huge resources of manpower and time to accomplish this task.

An Efficient Semantic discovery algorithm includes mapping of concepts from conventional web services to Semantic web services, a classification scheme for categorization of concepts and a ranking algorithm, [1]. This paper proposes a novel scheme of Semantic discovery, through adding the measure of Semantic similarity of words in the discovery process. It extends the focus of Semantic discovery with detailed overview of Semantic Mapping and Classification of concepts.

II. NEED FOR SEMANTIC WEB

A. Limits on HTML

The Semantic web is familiarly renowned as web3.0 and Web of data etc-. The word 'Semantic' is termed for "Meaning" or "Understanding". The major difference between Semantic web and conventional web is the meaning and structure of data. The goal of Semantic web is to add 'semantics' to the current web services for achieving machine interoperability.

The Conventional web lacks machine understanding due to underlying HTML. This leads to the evolution of Semantic Web. It is a tedious task to develop an entirely new web service framework, [2]. Hence, the choice of redefining the web resources from syntactic discovery to semantics based discovery gets prevailed all over the world.

The current web services hold the issues over uncertainty and denial. It is due to the complete keyword based discovery of concepts. In this mode of search, the supplied keywords in queries are matched to the keywords in meta tags of HTML design of web pages,[6]. This task of matching is done by web crawlers. On matching these keywords, only related pages could be acquired as relevant results.

978-1-4673-5036-5/13/$31.00 © 2013 IEEE

B. Evolution of the web

The Ontology model in semantic web can be compared with that of Human Brain. For example: Child gains knowledge from the environment through matching their experiences gained day by day. The use of ontology supports Semantic web through offering an efficient data model, [4]. In 2003, Semantic Web Research began, but still there is a lack of advance improvement. It is due to complexities like huge data model and dotcom burst.

The redefinition of web from existing web resources to Semantic web resources is exhibited through addition of Semantic Annotations, [14]. These annotations can be added to the web services by relating and tagging their descriptions with standard concepts in ontologies.

To add annotations to the web services, the knowledge of Semantic Markup Languages is must. Hence, the following section of this paper details the notes on Semantic Markup Languages.

III. Semantic Definition Languages

The Semantic Definition Languages like WSDL, OWL-S define the type of Services and the Service Provider information. The following section details the different types of Semantic Definition Languages. The key features of these languages on being compared clearly shows that, the original features of earlier languages be extended in the upcoming language types.

A. WSDL

The Web Services Description Language is used for describing the underlying functionality of a Web service. It is otherwise, termed for a WSDL file which provides a machine-readable description of service invocation, expected parameters and return of data structures,[11].

B. DAML-S

A DAML-S Process Model is an ordered collection of processes, where each process produces a state of information exchange with the web service clients. The processes are classified into atomic and composite processes, [9]. The DAML-S Grounding is termed for a direct mapping from atomic processes to WSDL descriptions.

C. OWL-S

The OWL-S is an upper ontology for describing the properties and capabilities of web services in OWL. The process of discovery mechanism depends with the OWL-S Service profile component for redefining the web services from WSDL file descriptions, [12]. These descriptions are applied with mapping algorithm, to bring out OWL-Service profile on extraction.

The Basic components of the Semantic Languages detailed above are listed with their features in the following table.

TABLE I
Basic Components Of WSDL

Components	Functions
Service	It is a container for a set of system functions being exposed to the web based protocols.
Port	It defines the address or connection that points a web service.
Binding	It specifies the interface and brings out the SOAP binding definition.
Port Type	The <portType> element has been renamed to <interface> in WSDL
Operation	Every Operation can be compared with a method or function call in programming.
Types	It describe the type of data

TABLE II
Basic Components Of OWL-S

Components	Functions
Service Profile	It describes the service information which is necessary for the discovery process.
Service Model	It details the composition, execution of process.
Service Grounding	It is the coordinator of Service usage.

IV. Semantic Mapping Approaches

An Efficient mapping of web services can be achieved through automation of this process. But, the WSDL files naturally doesn't support automatic mapping. The following Section offers an overview of different mapping schemes applied for redefining the conventional web and the related works over Semantic mapping of tools using different approaches on semi automation of mapping algorithms for redefinition of web services.

A. ASSAM: A Tool for Semi-automatic Semantic mapping

The two machine learning algorithms like iterative relational classification algorithm and a Schema Mapping Algorithm is used. The Iterative relational classification algorithm is for semantically classifying Web Services and the other algorithm is applied with String Distance Metrics, [5].

Usually, vendors who want to step into web3.0 apply integration of their existing web services through addition of Semantic annotations, according to some shared Ontology. These tasks can be accomplished through annotating ontological descriptions along with standard classification scheme.

This Semantic mapping approach consists of two parts, a WSDL annotator application, and OATS, (Operation Aggregation Tool for Web Services). The WSDL annotator enables the user to semantically annotate a Web service using an easier point-and-click user interface. The WSDL annotator has a key feature to suggest which ontological class can be

used to annotate each element in the WSDL Descriptions. It can be applied to automatically aggregate the resulting heterogeneous data into some coherent structure. It is a schema matching algorithm that is specifically suited for aggregating data from web services.

B. METEOR-S

This tool mainly focuses with generation of Schema graphs on matching them with ontologies. It is a semi –automatic tool that favours optimization on Semantic mapping of concepts. The generated files from mapping called .daml files are support ontology modeling.

It consists of an Ontology store and a Translator library. The use of ontology store is to reserve these ontologies. It allows user over adding new ontologies, [15]. The Translator Library generates the schema graph representations. The main purpose of this approach is the generation of schematic graphs for matching their common properties. This generation of visual representation can be achieved through two different translators like Ontology2graph and wsdl2graph. The 'wsdl2graph' favours conversion of WSDL files into schema graphs and 'Ontology2graph' favours conversion of ontology files into schema graphs, [8]. The matcher library depends with two types of matching algorithms, Such as Element level matching and Schema matching algorithms.

The Element level matching algorithm is applied in order to compute measures on names and to sort their relevance, [7]. It applies a semantic measure for similarity, where the Schema matching algorithm is used for finding structural similarity among two different concepts.

C. WSDL to DAML-S

It converts the web services from WSDL to DAML-S. It is termed for Semi-automated mapping of concepts. Since, the supplied WSDL definitions are converted to the complete specification of Grounding. But, only an incomplete specification of the Process Model and Profile can be generated in this mapping, [9]. This incompleteness is due to the limited feature of WSDL. It doesn't provide any information on process composition. The service capability description also is not available with WSDL descriptions. Hence, it must be done manually.

D. Semantic Mapping using Standardization

This mapping algorithm extends the purpose of Semantic mapping of concepts on applying effective standardization, using concepts of ontology, [13]. It admits use of a novel scheme called Ontology Search and Standardization engine for standardization of concepts.

This Standardization can reduce the complexity of huge concepts underlying with the applied ontology, [10].

It offers standardization through three stages of refining. Like, Linguistic Refining, Structural Refining and Statistical Refining. In linguistic refining, the more related words are analyzed to compute Term Frequency. In Structural Refining, the concept-to-concept relationships are derived. The relative measures of concepts are found similar to be ranked to reorder the ontologies list. Thus, standardization of concepts is an added advantage over refining the better results on Semantic mapping.

V. PROPOSED APPROACH

The Semantic Discovery of web services includes different phases of mapping, classification, measure of Semantic Similarity and ranking. The following section details these various phases of Semantic Discovery Algorithm. The Initial phase is to apply mapping of concepts through Semantic mapping. It can be done to turn the conventional web services to Semantic web services which could favour reasoning. It is followed by a simple classification that organizes the mapped concepts into the specified categories. The classification can be done based on a pre-defined order. Now, the web services are arranged in an order. In order to check the semantic discovery of results from the underlying concepts that are standardized, a query interface be designed.

When queries are applied, the semantic similarity measure between words supplied in the query is observed. On this Similarity observed, the relevant results matching the given query are retrieved. For more efficient results, these matched results are sorted in an order of ranking applied to them.

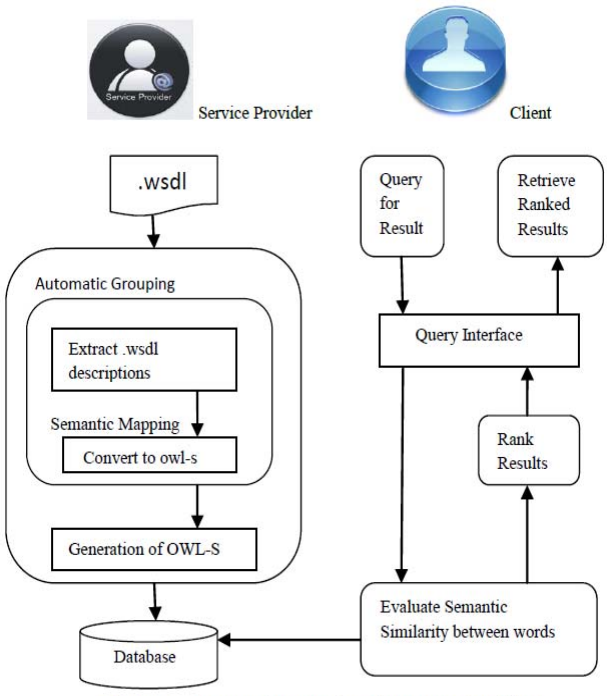

Figure I – Proposed Semantic Discovery Algorithm

A. Mapping of Conventional web services

The Service Providers submit their own WSDL files for the purpose of Semantic Discovery. They register their own web services using URLs. The Semantic Mapping of Web services can be done by extracting the WSDL descriptions from this .wsdl file. The WSDL files of Service Providers can be acquired from their registered URLs. These extracted definitions are converted to OWL-S descriptions.

B. Classification of Concepts for categorization

To achieve standardization over these mapped OWL-S descriptions, classification of concepts be applied. On applying simple classification, concepts are arranged into relative domain as groups. This grouping of concepts could favour easier retrieval of results on querying. Initially, groups are defined for categorization of concepts. A simple classifier is applied for checking the relativeness of the new concept to the pre-defined concepts, for being admitted to the pre-defined groups. It could achieve better classification which reduces the complexity of huge concepts. On classification of concepts, the rate of speed in accessing the results can be increased to some extent. It achieve less execution time for search.

C. Measure of Semantic Similarity

A query is normally, a keyword being supplied for retrieving the required results from the database. On measuring the semantic similarity between the supplied query and the mapped OWL-S file Service Profile, a meaning of the required result can be found. It can be achieved through matching the key words from query to the similar words (i.e) Synonyms from the Library being added. For Example: If the key word from query supplied be 'Money'. It can be matched to the similar words like 'Currency'. Hence, more related results for the required information can be found. It can achieve a best rate of Search performance for achieving relevant results. It can be analyzed using metrics like Precision and Recall.

D. Ranking

The related results found with the measure of Semantic similarity can be ranked using a simple ranking algorithm. Like, more relative results and most viewed results for that supplied query. It can also be ordered, according to their computed value of importance. The importance of a concept can be ranked depending on the number of sub-concepts that follow this concept.

VI. Conclusion

This paper illustrates a brief note on the importance of Semantic web and the need for redefinition. The related approaches on existing Semantic Mapping schemes are listed with their key features, which are extended in purpose of efficient mapping. The proposed discovery algorithm combines the method of semantic similarity measure between words to be applied with the queries supplied, in order to retrieve the semantically matched results. To optimize the retrieved results, ranking be applied over the matched results. It brings more relevant results to be ranked highest. Thus, Intelligence on Information retrieval for achieving more relevant results can be implemented on applying this proposed semantic discovery algorithm with semantic similarity measure.

References

[1] M. Burstein, C. Bussler, M. Zaremba, T. Finin, M. N. Huhns, M.Paolucci, A. P. Sheth, and S. Williams, "A Semantic Web Services Architecture", *IEEE Internet Comput.* , vol. 9, no. 5, pp.72–81, 2005.

[2] Karin K.Breitman, Marco Antonio, Walter Truszkowski , Semantic web: Concepts, Technologies and applications, NASA Monographs in systems and software Engineering.Springer,2007.

[3] Brahmananda Sapkota, Dumitru Roman and Dieter Fensel, "Distributed Web Service Discovery Architecture", *Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services* (AICT-ICIW'06), 2006

[4] Dragan Gasevi and Marek Hatala," Searching Web Resources Using Ontology Mappings", *School of Interactive Arts and Technology*.

[5] Andreas Heß, Eddie Johnston and Nicholas Kushmerick,"ASSAM: A Tool for Semi-Automatically Annotating Semantic Web Services", *12th International Conference on Web Technologies*, pages 470–475, 2008.

[6] Jenice Aroma R, Mathew Kurian, 2012. A Survey on need for Semantic web. International Journal of Scientific Research and Publications.

[7] Abhijit A. Patil, Swapna A. Oundhakar, Amit P. Shethand Kunal Verma, "METEOR-S Web Services annotation framework", *proceedings of the 13th international conference on WWW*.ACM Press, 2004.

[8] Nicole Oldham, Christopher Thomas, Amit Sheth and Kunal Verma, "METEOR-S Web service Annotation Framework with Machine Learning Classification" , *Semantic Web Services and Web Process Composition*, Volume 3387 of LNCS, pp. 137–146, San Diego, CA, USA, Springer,2004.

[9] Massimo Paolucci, Naveen Srinivasan, Katia Sycara, Takuya Nishimura, "Towards a Semantic Choreography of Web Services: from WSDL to DAML-S", *In Proceedings of the First International Conference on Web Services (ICWS'03)*, Las Vegas, Nevada, USA, pp 22-26, June 2003.

[10] Tamer A. Farrag, A. I. Saleh and H. A. Ali," Towards SWSs discovery: mapping from WSDL to OWL-S based on ontology search and standardization engine", *IEEE Transactions on Knowledge and Data Engineering, 2012.*

[11] Web Services Description Language, *W3C Note*,2001,[online] http://www.w3.org/TR/wsdl

[12] Web ontology language for services (OWL-S), *W3C Member Submission*,2004,[online] http://www.w3.org/Submission/OWL-S/

[13] David Martin, Massimo Paolucci, Sheila McIlraith, "Bringing Semantics to Web Services: The OWL-S Approach".

[14] Juan Bernabé Moreno, "Semantic Web Services, State of the art". *University of Granada*, *Department of Computer Science and Artificial Intelligence*.

[15] Dr. Waralak V. Siricharoen, "Ontology Modeling and Object Modeling in Software Engineering", *International Journal of Software Engineering and Its Applications (IJSEIA)*, Vol.3 No.1 2009, ISSN 1738-9984, January 2009.

First Author – Jenice Aroma R, is doing her Master degree in Computer science and Engineering in Karunya University. She is interested in emerging trends on information retrieval approaches of Data Mining technologies. She did her undergraduate research work based on an efficient content based video detection scheme for video search. Now, she had extended her research work in bringing out an efficient scheme of search, on applying semantics. Email: jenicearoma@karunya.edu

Second Author – Mathew Kurian, He has finished his M.E in computer science and engineering from Jadavpur University, Kolkata and currently he is working as Assistant professor in Department of Computer Science and Engineering in Karunya University. Previously, he worked as a Software Engineer with Aricent Technologies. He is doing his PhD degree in Data Mining. Email: mathewk80@karunya.edu