Improving service discovery for goal oriented system using SOAP architecture

Miss Heena Dabherao, Prof A.P.Bodkhe

Abstract— **Mining the web is defined as discovering knowledge from hypertext and World Wide Web. The World Wide Web is one of the longest rising areas of intelligence gathering. Now a day there are billions of web pages, HTML archive accessible via the internet, and the number is still increasing. However, considering the inspiring diversity of the web, retrieving of interestingness web based content has become a very complex task. The large amount of data heterogeneity, complex format, high dimensional data and lack of structure of web, knowledge mining is a challenging task. In this paper, it is proposed to introduce a new framework generated to handle unstructured complex data.. Based on the research of web knowledge mining, XML is used to create well-structured data. Here we are trying to exploit the semantic web discovery using SOAP architecture in order to meet goals like type of Service, delay in service operation & service discovery in comparison with RESTfull architecture & analyse the comparative results. This work while integrating all our earlier proposed frameworks aims to evaluate this integrated framework on existing metrics to explore its applicability in real world applications.**

***Index Terms*—Semantic Web, service-oriented discovery, agents,SOAP architecture.**

I. INTRODUCTION

The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. To achieve this functionality of semantic web,agents play an important role. The effectiveness of such software agents will increase exponentially as more machine-readable Web content and automated services (including other agents) become available. Many automated Web-based services already exist without semantics, but other programs such as agents have no way to locate one that will perform a specific function. This process, called service discovery, can happen only when there is a common language to describe a service in a way that lets other agents "understand" both the function offered and how to take advantage of it. Services and agents can advertise their function by, for example, depositing such descriptions in directories analogous to the Yellow Pages.
Semantic annotations describe the semantic of the resources so that software agents can reason about them in order to reach a predefined goal. The goals of the agents vary from application to application, but they all rely on the operation of finding and using the resources necessary to perform the goal. To allow the deployment of the Semantic Web, technology is being developed for representing semantic annotations, for finding them, for reasoning about them and for using the resources that they annotate. Service orientation is a major trend in distributed computing. Relying on the web service core technology, IT systems in the future shall be composed of web services as the building block instead of proprietary solutions. Although there are reports on successful implementation of service-oriented architecture (SOA) in industry [1], there are still many problems and challenges facing full scale adoption [2] and enterprises maturity towards SOA is still in its early phases [3]. Indeed, the technologies underlying SOA are very complex, and service-oriented engineering (SOE) find its origin in diverse disciplines that are woven together in an intricate manner.

Heena P Dabherao, *Information Technology, S.G.B.A.U,Amravati, P.R.I.T & R, ,Amravati,India, 9970547529.*

Prof.A.P.Bodkhe,, *Information Technology, S.G.B.A.U,Amravati, P.R.M.T & R,, Amravati, India*

II. LITERATURE REVIEW

In this section we describe the relevant contributions to the field of Semantic web Service discovery and their relationship to our proposed approach. There are many studies that have highlighted the importance of semantics in today's world of fast growing intelligent applications.

There has already been plenty of research work on service and content discovery on the web, with several approaches that usually have some differences from the one presented here. At the content level, scraping techniques have been used to extract data, in a similar way, as other information extraction approaches. We can point to the system Reform [4], Thresher [5] and Marmite [6], Chickenfoot [7], or Denodo[8]. These approaches propose techniques that facilitate the scraping process, and which can complement our discovery rules. Also, gleaning resource descriptions from dialects of languages (GRDDL) [9] is the standard technology for extracting RDF from HTML documents. Unlike these approaches, ours provides a semantic framework that enables dereferencing the scraped data and reasoning about the scraping process. This enables enhancements in the scraping process, such as reasoning about visual aspects of the web resource, distributed scraping using multiagent systems, or focussed scraping and reasoning about the scraped resources.WADL[10] defines a format for building and publishing RESTful semantic service descriptions that can be discovered and processed by automatic agents.
Web content mining is performed by extracting useful information from the content of a web page/site. It includes extraction of structured data/information from web pages, identification, match, and integration of semantically similar data, opinion extraction from online sources, and concept hierarchy, ontology, or knowledge integration. A lot of work

ISSN: 2278 – 1323

has been carried out in this field which enables to improve the content by adding metadata.

J. Han and C. Chang [11] author a paper on data mining for web intelligence that claims that "incorporating data semantics could substantially enhance the quality of keyword-based searches," and indicate research problems that must be solved to use data mining effectively in developing web intelligence.

Barsagade[12] provides a survey paper on web mining usage and pattern discovery. Kolari and Joshi[13] provide an overview of past and current work in the three main areas of web mining research-content, structure, and usage as well as emerging work in semantic web mining.

Another research work by Sharma et. al [14] highlighted use of cloud computing in web mining , Kosala et. al [15] focused on scope of agent technology in it whereas Eirinaki et. al [16] provided details on web personalization through web mining.All of this provided detailed review on web mining focusing on different dimensions of this field. Meirong et. al [17] proposed an agent based web mining model for e-buisness.

Over the time,agent based software system research developed with the objective of enabling self-decribed software components to automatically cooperate and combine to perform user-specified tasks without the need to pre-program how tasks would be completed.Diffrent languages were developed to describe the capabilities that agents offered and requests that a user of an agent-system could make.The languages have formal semantics & can be unambiguously reasoned over using logical reasoning engines.thus,on this notion of an agent-facilitator or agent-broker was devised to bring scalability to agent systems.

III. Analysis of Problem

In the present generation of WWW, the user is more interested in getting useful, relevant and knowledge oriented contents from the WWW. The paradigm is shifting from demand of information to demand for knowledge. The WWW is transforming into Semantic web which is knowledge oriented. Still we are away from complete realization of semantic web; web content mining when applied on semantic web contents can lead to discovery of knowledge that could be provided to end users to better serve their requirements. This work aim to propose agent based framework for mining contents of semantic web, which would provide query relevant knowledge using clustering technique.

In order to define the semantics for digital content, it is necessary to formalize the ontologies by using specific languages as Resource Description Framework (RDF) and Web Ontology Language (OWL). An ontology is considered as an explicit representation of a shared understanding of the important concepts in some domain of interest.

Ontology-matching results, called alignments, can thus express the relations between the ontologies under consideration with various degrees of precision.

When data is marked up using ontologies, software agents can better understand the semantics and therefore more intelligently locate and integrate knowledge for a wide variety of tasks.

As described in a dissertation by Roy Fielding, REST is an "architectural style" that basically exploits the existing technology and protocols of the Web, including HTTP (Hypertext Transfer Protocol) and XML.

SOAP is the standard messaging protocol used by Web services. SOAP's primary application is inter application communication. SOAP codifies the use of XML as an encoding scheme for request and response parameters using HTTP as a means for transport. Service discovery and semantic web technologies have received tremendous attention in the academic research community. However, to our knowledge, only a limited set of service oriented engineering approaches are clearly based on the goal concept.

IV. Proposed Work

There is an existing architecture that proposes a framework for the discovery of services and contents in the web. The agent architecture allows implementing agents that intelligently crawl and use services for retrieving contents in the web that correspond to some top goals, usually, stated by the user. The agent searches the web and combines different information sources to carry out effective reasoning that determines its decisions and behaviour during the crawling.

Although browsing is extensively used for searching and finding information on the web, it is proposed in a goal based manner by only one approach (Stollberg & al.). It relies on visualizing a hierarchical graph structure in which nodes are goal templates and edges are generalization links. Nodes are associated with available semantic web services. Navigation in the space of available services can significantly help in better understanding the problems that can be solved by them. An open question is the structure of complex goals, and how browsing can exploits different categories of links between goals. This is tightly related to goal and service composition.

REST is a set of principles that define how Web standards, such as HTTP and URIs, are supposed to be used. REST (Representational State Transfer) [26] is a simple stateless architecture that generally runs over HTTP. REST involves reading a designated Web page that contains an XML file. The XML file describes and includes the desired content. REST is often used in mobile applications, social networking Web sites, mashup tools and automated business processes. The REST style emphasizes that interactions between clients and services is enhanced by having a limited number of operations (verbs). Flexibility is provided by assigning resources (nouns) their own unique universal resource indicators (URIs). Because each verb has a specific meaning (GET, POST, PUT and DELETE), REST avoids ambiguity.

Discovery can be defined as the process of identification and construction of an element's semantically meaningful description at some particular level. The framework is shown in Fig. 1. It is stacked on top of the representational stateless transfer (REST) architectural style [27], the architectural style on which the World Wide Web is based. This framework uses three levels of abstraction on the content and services that are available on the web.

ISSN: 2278 – 1323

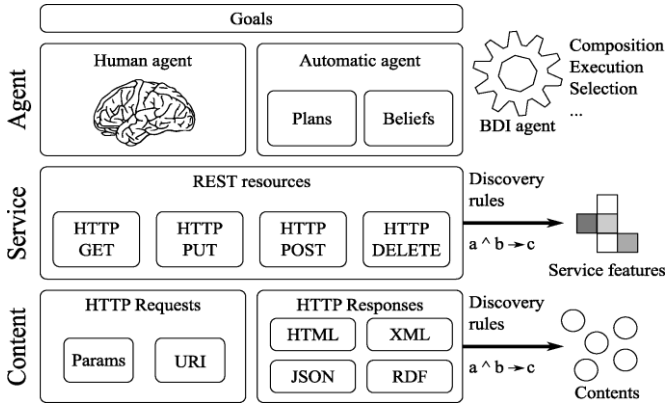

Fig: Discovery Framework

The agent level comprises the orchestration of services for fulfilling a user goal. Searching blogs to obtain relevant information about a product is one of the examples.

The Service layer comprises the services that the agents are able to use on the web which vary from services to services. They make use of contents in the lower layer by exchanging requests and responses with representations of resources that are present on the web.

The Content level comprises the requests and responses that are exchanged between clients and servers when interacting with web resources.

Discovery will take place at these levels. At the service level, REST resources would be the discoverable elements to be analyzed, with semantic service descriptions being obtained. At the Service level, once the services are discovered, we would check if the service goals are met or not. If the goals are not met by implementing Restfull architecture then we would again perform another search till the goals are properly met. Goals can be, type of service, delay in service operation, delay in service discovery and more. If the goals are not met even by Restfull architecture, we would try to improve the discovery of goals by implementing SOAP architecture (Simple Object Access Protocol)which would improve the service discovery and reduce the time needed for service discovery as well.Goals will be predefined.

V. CONCLUSION

Matching user requirements with service capabilities is a fundamental issue in service-oriented engineering. Our analysis shows clearly that goals are becoming an important element in approaches to service discovery, and that it can be the core building block for a fully recognized SOC engineering method.

This work has proposed agent based solution for mining semantic web contents, with the aim to provide context based knowledge oriented results to the user. The next generation of WWW will be knowledge oriented and to satisfy the customers web mining is a promising solution. The amalgamation of web mining techniques with agent technology will lead to improved performance, reduced network traffic, and better results.

Thus we would be checking different parameters like type of service, delay in service operation, delay in service discovery etc using Restfull architecture & compare the results obtained by using SOAP architecture to that of Restfull architecture in order to optimize it if needed.

VI. REFERENCES

[1] J. Eckert, M. Bachhuber, N. Repp, and R. Steinmetz, « The Implementation of Service-Oriented Architectures in the German Banking Industry - A Case Study », in AMCIS 2009, 2009, Paper 473

[2] R. Hirschheim, R. Welke, and A. Schwarz, « Service-Oriented Architecture:Myths, Realities, and a Maturity Model »,MIS Quarterly Executive, vol. 9, no. 1, p. 37–48, 2010.

[3] R. Welke, R. Hirschheim, and A. Schwarz, « Service-Oriented Architecture Maturity », IEEE Computer, vol. 44, no. 2, p. 61–67, 2011.

[4] M. Toomim, S. M. Drucker, M. Dontcheva, A. Rahimi, B. Thomson, and J. A. Landay ttaching UI enhancements to websites with end users," in Proc. Conf. Human Factors Comput. Syst., pp. 1859–1868, 2009

[5] A. Hogue, "Thresher: Automating the unwrapping of semantic content from the World Wide Web," in Proc. 14th Int. World Wide Web Conf., 2005, pp. 86–95

[6] .J. Wong and J. I. Hong, "Making mashups with marmite: Toward end-user programming for the web," in Proc. Conf. Human Factors Comput. Syst., 2007, p. 1435.

[7] M. Bolin, M. Webber, P. Rha, T. Wilson, and R. C. Miller, "Automation and customization of rendered web pages," in Proc. Symp. User Interface Softw. Technol., 2005, p. 163.

[8] A. Pan, J. Raposo, M. ´Alvarez, P. Montoto, V. Orjales, J. Hidalgo, L. Ardao, A. Molano, and A. Vi˜na, "The Denodo data integration platform," in Proc. 28th Int. Conf. Very Large Data Bases (VLDB), Aug. 2002, pp. 986–989

[9] D.Connolly.(2007).Gleaning resource descriptions from dialects of languages[Online].Available:http://www.w3.org/TR/grddl/

[10] .M. J. Hadley. (2006). web application description language [Online]. Available: https://wadl.dev.java.net/wadl20061109.pdf

[11] J.Han and C.Chang, Data mining for web intelligence, Computer (November2002),pp.54–60,http://www-faculty.cs.uiuc.edu/hanj/pdf/computer02.pdf

[12] B. Hay, G. Wets and K. Vanhoof, Mining navigation patterns using a sequence alignment method, Knowledge Inform. Syst. 6(2) (2004) 150–163.

[13] A. Joshi, Web/data mining and personalization, University of Maryland Baltimore County (UMBC) eBiquity Research Area (2001), http://ebiquity.umbc.edu/project/html/id/17/Web-Data-Mining-and-Personalization.

[14] Sharma K., Shrivastava G. & Kumar V., 'Web Mining: Today and Tommorrow'. In Proceedings of the IEEE 3rd International Conference on Electronics Computer Technology, 2011.

[15]] Kosala R. & Blockeel H., 'Web Mining Research: A Survey'. Published in ACM SIGKDD, Vol. 2, Issue1,July 2000.

[16] Eirinaki M. & Vazirgiannis M., 'Web Mining for Web Personalization'. Published in ACM Transactions on Internet Technology, Vol.3 , No. 1, February 2003, pp. 1-27.

[17] Meirong T. & Xuedong C. , 'Application of Agent Based Web Mining in E-business'. Published in 2010 IEEE Second International Conference on Intelligent Human-Machine Systems and Cybernetics, pp.192-195.

AUTHOR INFORMATION

Heena P Dabherao, *completed her B.E(IT)and currently pursuing ME(IT) from S.G.B.A.U. Amravati.*

Prof.A.P.Bodkhe,*, He is the Principal & Professor at prof Ram Meghe institute of technology and research Amravati,*

ISSN: 2278 – 1323