

Approaches for Automated Web Service Discovery

Gajanan V. Bhole , Prof. Prakash. R. Devale

^{1,2}Department of Information Technology,

Bharati Vidyapeeth University College of Engineering, Pune-43, India

Abstract— Paper highlights the previous work on web service discovery and techniques; also describes the proposed work for automated web service discovery with combination of different approaches. The approaches enable users to form queries and to search and discover the services based on different requirements. The proposed system uses methods such as probabilistic machine-learning techniques and logic based search to extract latent factors from user query and match with probable web service by searching in ranked and clustered web service repository. Otherwise it discovers the related web service from web and adds it in web service repository for further use.

Keywords— Machine learning, Logic based search, Web service discovery, Probabilistic matchmaking

I. INTRODUCTION

Due to an increasing count of web services on the web it is important to have an reliable, accurate and effective way to manage and retrieve web services in such a way that to search the best result for required user query. Web services are the software components which are loosely coupled and published and invoked over the web. Due to large and increasing number of web services available on the web or within the organization creates new search problem that is finding or locating required web services [1]. In current situation, many companies in the world are going to make the use of external services and trying to search and invoke the web services is becoming more popular. So that, service discovery has a vital role in web based application using web services. How to search the required web service? Retrieved web services are related to the subject or user query? Which web service is satisfying the user query? Answer to these problems or questions require lots of human efforts for checking web services one by one but this approach is not feasible in current situation problem because e.g. more than 1 million web services.

In current scenario to locate or search published web services there are approaches of web service discovery that often adopt keyword-matching technologies. The search results returned by the service registries are effectively inadequate because lack of semantic descriptions of the Web services. Searching the similarity for web services is difficult or challenging task because lack of descriptions of web services and their operations and also names of the input and output parameters cannot convey completely semantics of the operation [1]. Internet Services are software components that support interoperability machine to machine interaction in network using industry standard protocols such as SOAP (Simple Object Access Protocol) on Hyper Text Transfer Protocol (HTTP). Simple Object

Access Protocol is a lightweight protocol for exchange of information between Web services [2]. Internet services are described by WSDL (Web Service Description Language).

Whenever a Web service is published, to serve all required data or information to other developers or programmers to invoke the service a WSDL file will be created [3]. Even if the development tool automatically generates the description, it still contains some important information regarding data types assigned by developer or programmer.

Registries of web services like UDDI (Universal Description, Discovery, and Integration) are developed to encourage adoption and interoperability of web services [3]. UDDI registries are require registration or made either available publicly and contain many obsolete entries [4]. In either case for available services a registry may store only a limited description. As a technology concern web services are a self describing, self containing and modular applications, which are defined by WSDL (Web Services Description Language) and these web services are published at registries of services such as UDDI (Universal Description, Discovery and Integration) [5]. Universal Description, Discovery and Integration (UDDI) in the service-oriented architecture bridge the link between the service providers and the service to applicants.

This paper defines the different approaches for web service discovery and proposed model for an automated web service discovery and computing with the help of probabilistic matchmaking and combination of different approaches. The rest of this paper is described such as: In Section 2 contains the related work, section 3 discuss the proposed system architecture in detail. Section 4 contains overview of proposed system, approaches and methods in detail. Finally, conclusion is made in section 5.

II. RELATED WORK

The work in which a new web service-based distributed search engine on the vector space model for information retrieval. An overview of the basic technologies and expanded the existing methods with a technique to make the concept of distributed environments. Finally, they have examined their prototype implementation and shown that the method presented also works for the large WSDL repositories [3].

A non logic based matchmaking approach which uses techniques of machine learning in particular PLSA (Probabilistic Latent Semantic Analysis) and LDA (Latent Dirichlet Allocation) to get or extract the latent factors from the semantic description of web service and search the web

services in latent factor space. Also this paper defines use of an probabilistic machine learning for service matching and ranking and these LDA and PLSA unsupervised probabilistic machine learning methods used for the service description data and generates a lower dimensional vector model for representation of services[11].

A. Semantic approaches

The useful techniques to access and investigate large service repositories to extend the scope of web service utilization. They specified the method for sake of categorization by labeling the web services automatically using semantic concepts [4].

To extract the semantic topics hidden behind the words in the query in services they use PLSA (Probabilistic Latent Semantic Analysis) so that the matching of services will be done at concept level [6].

The method discussed in [7], Semantic Web Service Classification (SWSC) analysis the WSDL and checks its structure and configuration for future processing.

OWL-S standard is considered for defining the web services semantically, so that every input and output of the web service is defined semantically by concept of ontology [8].

The approach is used as an effective reduction of dimension techniques; they are able to acquire semantic relations between word-item and item-document interpreted in terms of probability distributions [9].

Probabilistic models Latent Semantic Analysis (LSA) and probabilistic Latent Semantic Analysis (PLSA) item were originally developed and used for modeling and item removal in information retrieval [9]. With the help of these LSA and PLSA topic models They described a new approach for learning terminological ontology's based on the "Information Theory Principle for Concept Relationship" and topic hierarchy learning algorithms.

B. Clustering approaches

Dong [1] presents clustering approach for searching the web services where search operation consists of two stages. A user gives text input the service search engine, for required services. After that, based on the starting web services returned, given approach extracts semantic topics from the natural language descriptions provided in the web services.

To extend the groups, the similar Web services and semantic representation of services for the improvement of the service discovery they presented the Semantic Web services Clustering (SWSC) method [7]. The improvement in service discovery with the use of SWSC is shown by the empirical analysis. Also the three methods such as a) use of the WSDL information for basic keyword search; b) use of the clusters which are derived from the WSDL for keyword search; c) use of WSWC method for keyword search, shows the effectiveness of the Semantic Web Services clustering (SWSC).

C. Latent factor approaches

To extract latent factors from the concepts or data in the service descriptions they proposed a latent factor approach

for service ranking which is depend on the probabilistic machine learning techniques [10]. To assign a probability to every link among the latent factor and a service description this approach is located in the inherent property of probabilistic machine learning and this is advantage or versatility of this approach.

D. Ranking approaches

In [5] they presented a rank approach for the matching of web services that uses logic based reasoning along with approximate matching, which ranks the service descriptions matching degree as per the OWL-S. In this approach of matching service is based on latent semantic indexing and logic based reasoning. Ranked results give criteria or facility to select particular service from the large set of results.

New approach for semantic web services ranking [8], this approach will help to users to search needed web service in precise format. This approach presents an algorithm for ranking depend on the VSM (Vector Space Model). In this model the user query and retrieved services related to that user query will be considered as vector and based on that which service vector is very close to the user query vector will considered as higher rank. The rank of services are calculated automatically whenever the user fire the required query to system and rank values are not predefined to the web services.

III. SYSTEM ARCHITECTURE

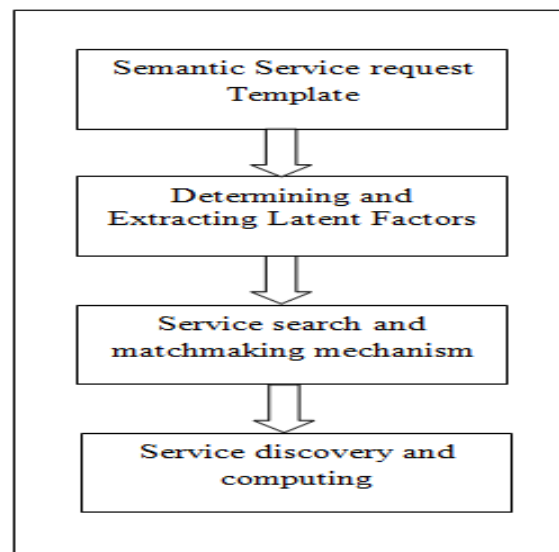


Fig. 1 Proposed system architecture of automated web service discovery and computing model.

The Fig. 1 depicts the process flow for automated web service discovery and computing model. The semantic service request template or the user query from which the latent factors are get extracted for matchmaking; after that service search and matchmaking from the local registry or web according to latent factor and then required service is discovered and after discovering the web service, model will invoke or compute the web services.

IV. PROPOSED WORK

The proposed work for automated web services discovery and computing has different approaches and methods for automatic web service discovery and invoking. In model there are probabilistic matching, logic based search and combination of different approaches are used for getting good results of web service discovery. The model has some functional blocks or discovery models such as

A. Detection of Hidden Topics in Service Description

Here, when a user gives service description or required query, the Latent Factors are used to find out hidden topic behind the service descriptor or user query. Proposed model will use a latent factor approach. In model the latent factor in the sense the meaningful or required part from the user query.

B. Registry Check and Retrieval of Service Information

After completion of extracting Latent Factors from service description or user query; model will check whether that service is present in Registry or not. If service is present, the description or the related information is returned back to the user. If service is not present in the Registry, Folding-In approach is used.

C. Folding-In

After the model is trained Folding-in method is used for fitting new web service description into the latent factor model. A folding-in approach is used search the required or needed web services and then add new services into the model [11]. In proposed work our model will search the service in local registry, if not then search services on web and then add that new service to the local registry for future use, for this purpose our model will use this folding-In approach.

D. Web service discovery

UDDI registries are designed as the central point to register Web services and to make them publicly available. Also they are possible to use data directly from UDDI registries [3], but since the data is not required to be in XML but can be only a textual description too. Retrieving enough WSDL Files from the Internet to form a satisfying repository was a particularly hard task. In our proposed model, after searching the web services according to user query and the probabilistic matchmaking these required services are discovered.

E. Service computing

After discovering the service and reading WSDL provide input message to web service and process output message after computing or invoking the services.

V. CONCLUSION

The paper defines the approaches and the previous work regarding the web service discovery. Also it shows the proposed model for automatic web service discovery and computing. The proposed work will enable user to make the query for required web service and model will automatically discover the related web services and shows the result to the user. The proposed model will use a logic based search, probabilistic matchmaking methods as well as combination of different approaches for better and fast web service discovery.

REFERENCES

- [1] X. Dong, A. Halevy, J. Madhavan, E. Nemes and J. Zhang, "Similarity Search for Web services", Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004.
- [2] A Comparative Evaluation of Semantic Web Service Discovery Algorithms and Engines by Savitha Emani.
- [3] C. Platzer and S. Dustdar, "A Vector Space Search Engine for Web Services," Proc. Third European Conf. Web Services (ECOWS '05), 2005.
- [4] A. Segev and E. Toch, "Context-based matching and ranking of web services for composition," IEEE Transactions on Services Computing, vol. 99, no. PrePrints, pp. 210–222, 2009
- [5] S.-L. Pan and Y.-X. Zhang, "Ranked web service matching for service description using owl-s" Web Information Systems and Mining, 2009. WISM 2009. International Conference on, pp. 427 – 431, nov. 2009.
- [6] Jiangang Ma, Jinli Cao, Yanchun Zhang, "A Probabilistic Semantic Approach for Discovering Web Services", WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. ACM 978- 1-59593-654-7/07/0005
- [7] R. Nayak and B. Lee, "Web service discovery with additional semantics and clustering," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 555–558.
- [8] O. Mola, P. Emamian, and M. Razzazi, "A vector based algorithm for semantic web services ranking," Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on, pp. 1 –5, apr. 2008.
- [9] W. Wei, P. Barnaghi, and A. Bargiela, "Probabilistic topic models for learning terminological ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1028–1040, 2010.
- [10] G. Cassar, P. Barnaghi, and K. Moessner, "A probabilistic latent factor approach to service ranking," in *Intelligent Computer Communication and Processing (ICCP), 2011 IEEE International Conference on*, aug. 2011, pp. 103–109
- [11] Gilbert Cassar, Payam Barnaghi, Klaus Moessner, "Probabilistic Matchmaking Methods for Automated Service Discovery" *IEEE Transactions on Services Computing*, 2013.