

Semantic based Automated Service Discovery

Mr.Likshesh N. Kolhe¹, Prof. Sachin M. Bojewar²

¹(PG Scholar, Alamuri Ratnamala Institute of Engineering & Technology, Mumbai University, India)

²(Associate Professor, Vidyalkar Institute of Engineering & Technology, Mumbai University, India)

Abstract: A vast majority of web services exist without explicit associated semantic descriptions. As a result many services that are relevant to a specific user service request may not be considered during service discovery. In this paper, we address the issue of web service discovery given no explicit service description semantics that match a specific service request. Our approach to semantic based web service discovery involves semantic-based service categorization and semantic enhancement of the service request. We propose a solution for achieving functional level service categorization based on an ontology framework. Additionally, we utilize clustering for accurately classifying the web services based on service functionality. The semantic-based categorization is performed offline at the universal description discovery and integration (UDDI). The semantic enhancement of the service request achieves a better matching with relevant services. The service request enhancement involves expansion of additional terms (retrieved from ontology) that are deemed relevant for the requested functionality. An efficient matching of the enhanced service request with the retrieved service descriptions is achieved utilizing Latent Semantic Indexing (LSI). Our experimental results validate the effectiveness and feasibility of the proposed approach.

Keywords: Semantic Matching, Semantic Web services, Service-oriented Computing, SVM, UDDI

I. INTRODUCTION

The need for information technologies which are able to support agile organizations and fast-changing business processes, has led to the wide propagation of Service-oriented Computing (SOC). Today, SOC is a multi-level approach, ranging from the engineering and operation of IT infrastructures to the usage in small Web-based applications called software mashups. One particular application area, which has heavily influenced the computer science research community, as well as the software industry, in recent years, is the Service-oriented Architecture (SOA) paradigm, where services are deployed in order to organize and implement IT architectures and, eventually, realize Business/IT alignment. Independent of the actual application area, SOC is based on services. Services are self-describing encapsulations of functionalities offered by software components. As services are loosely coupled and self-contained, it is possible to dynamically invoke and substitute services, e.g., in a business process, even across the borders of a single company or organization. Hence, one particular application area is the usage of services in workflows, i.e., IT-supported business processes.

Many use cases for SOC are based on the invocation of services through the Internet using Web service standards like the Web Service Description Language (WSDL) and the Simple Object Access Protocol (SOAP). In fact, Web service technologies are currently the most common way to implement service oriented concepts and have led to the vision of an "Internet of Services", where services for all areas of life and business will be offered on the Web. Regardless of whether using services on a large or small scale, the discovery of services, which offer a distinct functionality under certain non-functional requirements is one of the vital steps in service invocation and is hence deemed one of the grand challenges in Web service and SOA research. For the discovery of Web services, it is necessary to describe the offered and requested services' elements and capabilities as precisely as possible. Ideally, a service consumer would find a range of ready-to-use services, which are offered by different service providers.

Today's Web service standards are primarily motivated by the interoperability of software components over the Web and rely on the Extensible Markup Language (XML). As there is no further framework for the accurate description of service functionalities and properties, services are usually described using natural language which is often too imprecise. As a result, WSDL-based service descriptions cannot be interpreted without human intervention. Accordingly, further functionalities such as automated Web service discovery, execution, or composition are very difficult to achieve. However, a certain degree of automation is necessary in order to achieve the application of Web service technologies on a larger scale.

In order to overcome the shortcomings of syntax-based service descriptions, several researchers have proposed the usage of semantic information in Web services, resulting in the concept of semantic Web services (SWS). Here, it seems reasonable to apply methods, tools, and technologies from the Semantic Web, which is an extension of the current Web as envisioned by Berners-Lee. The comprehensible Semantic Web activities of the

World Wide Web Consortium (W3C) aim to augment information on the Web with a well-defined meaning and provide a layer of machine-interpretable data. The ultimate goal is to accomplish tasks automatically by supplying machines with adequate information.

Today, SWS are a prominent field of research and have resulted in a number of different approaches and standards such as the Web Ontology Language for Web Services (OWL-S), the Web Service Modelling Language (WSML), or Semantic Annotations for WSDL and XML Schema (SAWSDL), i.e., formalisms which explicitly make use of semantic technologies in different parts of a service description.

One of the primary application areas of SWS is service discovery, which is essentially affected by three steps: (i) The ability of service providers to describe their services, (ii) the ability of requesters to describe their requirements towards services, and (iii) the effectiveness of the service matchmaker, i.e., an algorithm that takes into account a request and finds the best fitting services from a set of service offers. Service matchmaking that considers semantic information is contemplated by a very agile research community, with a large number of different approaches having been proposed in recent years. A lot of experimentation is conducted concerning the selection of elements from a service description, similarity metrics, and the combination of the resulting similarity values. State-of-the-art matchmakers are mostly quite inflexible towards differing service domains or need to be adapted manually. This is rather inappropriate as single services as well as service domains might differ to a very large degree regarding basic assumptions towards semantic descriptions of distinct service components or even the availability of a semantic-based domain model. Furthermore, semantic-based matchmaking is mostly based on rather coarse-grained Degrees of Match (DoMs), which allow only a very basic ranking of matchmaking results.

As a consequence, it is necessary to complement these values in order to facilitate a sophisticated ranking. If non-logic techniques like Cosine similarity or the Jaccard coefficient, which are well-known in the field of Information Retrieval (IR) research, are applied, there is always some degree of uncertainty as linguistic based methods are, for example, unaware of different meanings a term might have.

II. RELATED WORK

The challenges pertaining to automatic classification of web services have been addressed in prior work [7], [11], [8], [26]. In [11], Heb and Kushmerick propose an approach of using the information contained in the service description to dynamically create the categories for service classification, comparing five clustering algorithms. The classification process has similarities to our approach in terms of construction of term vectors with relevant words and utilizing a hierarchical clustering approach for achieving the best results. Our approach builds on this by 1) including relevant semantic concepts based on semantic relationship ranking for expanding the domain coverage, 2) deletion of nonrelevant terms resulting in the reduction of noise and increase in the purity of the clusters.

Bruno et al. [7] propose a classification approach utilizing Support Vector Machines (SVM) to classify the term vectors. Bruno et al. [7] also make use of concept lattice created using Formal Lattice Analysis to identify concepts for a specific domain as well as the relationships between services belonging to a class. This approach is the closest to our approach. Our approach, however, is based on gleaning of semantic utilizing a domain ontology hierarchy. Additionally, from our point of view, this approach does not address the issue of SVM mapping training data to higher dimensional space, and then finding the maximal marginal hyperplane to separate the data.

One of the approaches for enhancing the training time of SVM, specifically when dealing with large data sets, recommends hierarchical clustering analysis. Also ontologies can be used to improve Formal Concept Analysis (FCA) applications. In standard FCA, the set of attributes does not carry any structure. By considering this set as a set of ontology concepts, we can model relations and dependencies between the attributes. Although this does not increase the complexity of the resulting lattices (as concept lattices cover, up to isomorphism, the whole class of complete lattices), it enriches the conceptual structure and provides new means of interaction and analysis. FCA may also complement our approach by facilitating ontology merge and linking to provide a better depth and span in terms of the domain concepts coverage.

In [26], Oldham et al. propose a framework to semi automate the semantic annotation of web services for classification-based on matching web service data types and domain ontology concepts making use of schema matching. The main drawback of this is that it is not simple to find similarities with domain ontology concepts as no single domain concept contains the complete structure of a complex schema containing all service parameters.

Existing approaches to web service matching address either syntactic and/or semantic matching, e.g., Sajjanhar et al. [29] have studied LSI to acquire the semantic associations between short textual web service descriptions, Corella et al. [8] describe a heuristic approach for semiautomated web services' classification based on a previously classified services corpus. These approaches utilize the initial web services' descriptions advertised by service provider and functionality request specified by the service requestor. These initial

descriptions do not include any semantic augmentation. Our approach extends this work by adding semantics to the service request. As validated by the experimental results, this helps us achieve improved results for appropriate service discovery. The most widely used IR technique constitutes the Vector-Space Model [31]. VSM, however, considers the syntactic aspect of term association and does not account for the underlying semantic structure. Kokash et al. [12] address the inadequacy of VSM approach by expanding both the service query and the WSDL descriptions. A Hybrid matching approach is proposed that may combine various matching methods (e.g., syntactic and semantic) into a composite algorithm. This enables ad hoc composition of several (pre-existing) matching approaches based on predefined criteria. In principle, this is similar to our work. However, although it may provide flexibility, it also increases the human intervention for selection of a composite algorithm applicable to a set of services for specific application.

The usage of synonyms does not capture the overall semantics of the domain and application functionality. However, our approach utilizes concepts extracted from domain ontology. These extracted concepts account for relationships between the domain objects and provide a comprehensive coverage for the underlying semantics for both the domain and the application functionality. Our approach appends the syntactic service description with relevant semantic terms. This enables uniform combination of syntactic and semantic matching rendering our approach more generalizable for overall service matching and requiring minimal human interaction.

Our approach has similarities to existing approaches [33], [1], [2], [19] of natural language processing techniques that address the text part of the challenge in content-based image retrieval (CBIR). These approaches, however, were used in isolation to one another. Our approach, on the other hand, combines both these techniques using concept lists, distance within an ontological structure and latent semantic indexing.

In [13], Sassen et al. describe the SeCSE approach for architecture time service discovery that is based on ontologies that are validated, easy to use, complete, and widely accepted in domains. In contrast to this, our approach begins with the description of an ontology framework that includes upper ontologies, e.g., SUMO and more descriptive domain and application related ontologies. We propose a linked ontology structure for a wide-ranging description of domain semantics. Our approach for service discovery initiates service request enhancement with concepts extracted from related domain ontologies and reduces the space of service request and WSDL specification term vectors utilizing LSI to reduce the dimensions to be considered.

III. EXISTING SYSTEM

A majority of the current approaches for web service discovery call for semantic web services that have semantic tagged descriptions through various approaches, e.g., OWL-S, Web Services Description Language (WSDL)-S. However, these approaches have several limitations. First, it is impractical to expect all new services to have semantic tagged descriptions. Second, descriptions of the vast majority of already existing web services are specified using WSDL and do not have associated semantics. Also, from the service requestor's perspective, the requestor may not be aware of all the knowledge that constitutes the domain. Specifically, the service requestor may not be aware of all the terms related to the service request. As a result of which many services relevant to the request may not be considered in the service discovery process.

Existing service discovery approaches often adopt keyword-matching technologies to locate the published web services. This syntax-based matchmaking returns discovery results that may not accurately match the given service request. As a result, only a few services that are an exact syntactical match of the service request may be considered for selection. Thus, the discovery process is also constrained by its dependence on human intervention for choosing the appropriate service based on its semantics.

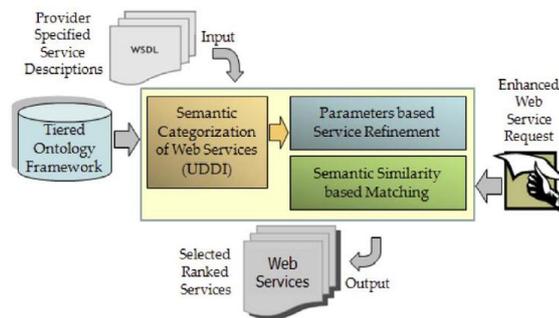
IV. PROPOSED SYSTEM

The limitations of existing approaches, an integrated approach needs to be developed for addressing the two major issues related to automated service discovery: 1) semantic-based categorization of web services; and 2) selection of services based on semantic service description rather than syntactic keyword matching. Moreover, the approach needs to be generic and should not be tied to a specific description language. Thus, any given web service could be described using WSDL, OWL-S, or through other means. Semantic-based categorization of web services is performed at the UDDI that involves semantics augmented /classification of web services into functional categories. The semantically related web services are grouped together even though they may be published under different categories within the UDDI. Service selection then consists of two key steps: 1) parameters-based service refinement; and 2) semantic similarity-based matching.

In order to address the limitations of existing approaches, an integrated approach needs to be developed for addressing the two major issues related to automated service discovery: 1) semantic-based categorization of web services; and 2) selection of services based on semantic service description rather than syntactic keyword matching. In this paper, we present a novel approach for semantic based automated service discovery. Specifically, the proposed approach focuses on semantic-based service categorization and selection as depicted

in Fig. 1. In our proposed approach, semantic-based categorization of web services is performed at the UDDI that involves semantics augmented classification of web services into functional categories.

V. METHODOLOGY



User Registration

This module explains the design and implementation of user registration via web based services. This module will also communication established between client and web based service.

Service Categorization

The semantic categorization of UDDI wherein we combine ontologies with an established hierarchical clustering methodology, following the service description vector building process. For each term in the service description vector, a corresponding concept is located in the relevant ontology. If there is a match, the concept is added to the description vector. Additional concepts are added and irrelevant terms are deleted based on semantic relationships between the concepts. The resulting set of service descriptions is clustered based on the relationship between the ontology concepts and service description terms. Finally, the relevant semantic information is added to the UDDI for effective service categorization.

Service Refinement

The next step is service selection from the relevant category of services using parameter-based service refinement. Web service parameters, i.e., input, output, and description, aid service refinement through narrowing the set of appropriate services matching the service request. The relationship between web service input and output parameters may be represented as statistical associations. These associations relay information about the operation parameters that are frequently associated with each other. To group web service input and output parameters into meaningful associations, we apply a hyper clique pattern discovery. These associations combined with the semantic relevance are then leveraged to discover and rank web services.

Semantic Matching

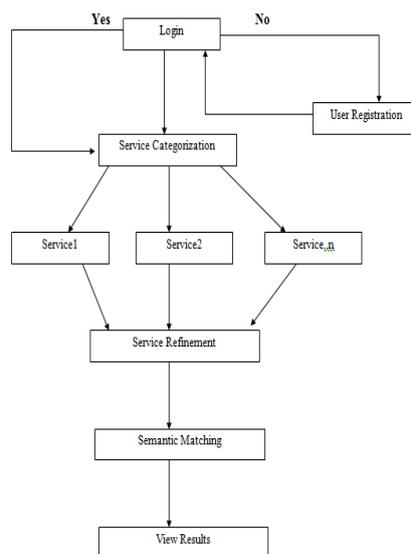
The parameter-based refined set of web services is then matched against an enhanced service request as part of Semantic Similarity-based Matching. A key part of this process involves enhancing the service request. Our approach for web semantic similarity-based service selection employs ontology-based request enhancement and LSI based service matching. The basic idea of the proposed approach is to enhance the service request with relevant ontology terms and then find the similarity measure of the semantically enhanced service request with the web service description vectors generated in the service refinement phase.

Feasibility Analysis

Preliminary investigation examine project feasibility, the likelihood the system will be useful to the organisation. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited recourses and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

- ✓ Technical Feasibility
- ✓ Operational Feasibility
- ✓ Economical Feasibility

System Flowchart



VI. CONCLUSION

In this project, we present an integrated approach for automated service discovery. Specifically, the approach addresses two major aspects related to semantic-based service discovery: semantic-based service categorization and semantic-based service selection. For semantic-based service categorization, we propose an ontology guided categorization of web services into functional categories for service discovery. This leads to better service discovery by matching the service request with an appropriate service description. For semantic-based service selection, we employ ontology linking (semantic web) and LSI thus extending the indexing procedure from solely syntactical information to a semantic level. Our experiments show that this leads to increased precision levels, recall levels, and the relevance scores of the retrieved services.

REFERENCES

- [1] J. Adcock, A. Girgensohn, M. Cooper, T. Liu, L. Wilcox, and E. Rie, "FXPAL Experiments for TRECVID," Proc. TRECVID, 2004.
- [2] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, 1993.
- [3] E. Al-Masri and Q.H. Mahmoud, "Investigating Web Services on the World Wide Web," Proc. 17th Int'l Conf. World Wide Web (WWW '08), Apr. 2008.
- [4] M.-L. Antonie and O.R. Zaane, "Text Document Categorization by Term Association," Proc. IEEE Int'l Conf. Data Mining (ICDM '02), 2002.
- [5] K. Anyanwu, A. Maduko, and A. Sheth, "SemRank: Ranking Complex Relationship Search Results on the Semantic Web," Proc. 14th Int'l Conf. World Wide Web (WWW '05), 2005.
- [6] P. Baldi, P. Frasconi, and P. Smyth, "Modeling the Internet and the Web," Probabilistic Methods and Algorithms, Wiley, 2003.
- [7] M. Bruno, G. Canfora, M.D. Penta, and R. Scognamiglio, "An Approach to Support Web Service Classification and Annotation," Proc. IEEE Int'l Conf. E-Technology, E-Commerce and E-Service (EEE '05), 2005.
- [8] M.A. Corella and P. Castells, "Semi-Automatic Semantic-Based Web Service Classification," Proc. Int'l Conf. Business Process Management Workshops (BPM '06), 2006.
- [9] P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.
- [10] E. Han, G. Karypis, and V. Kumar, "Scalable Parallel Data Mining for Association Rules," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '97), 1997.
- [11] A. Heb and N. Kushmerick, "Automatically Attaching Semantic Metadata to Web Services," Proc. IJCAI Workshop Information Integration on the Web, 2003.
- [12] http://dit.unitn.it/~kokash/documents/WS_matching-hybrid.pdf, 2012.
- [13] <http://idcrue.dit.upm.es/biblioteca/mostrat.php?id=2154>, 2012.
- [14] XMethods, <http://www.xmethods.net>, 2012.
- [15] <http://reliant.tekknowledge.com/DAML/SUMO.owl>, 2008.
- [16] <http://www.uddi.org/specification.html>, 2012.
- [17] <http://www.few.vu.nl/~andreas/projects/annotator/ws2003.html>, 2012.
- [18] H.L. Johnson, K.B. Cohen, W.A. Baumgartner Jr., Z. Lu, M. Bada, T. Kester, H. Kim, and L. Hunter, "Evaluation of Lexical Methods for Detecting Relationships Between Concepts from Multiple Ontologies," Proc. Pacific Symp. Biocomputing, 2006.
- [19] M. Kher, D. Brahma, and D. Ziou, "Combining Visual Features with Semantics for a More Efficient Image Retrieval," Proc. 17th Int'l Conf. Pattern Recognition (ICPR '04), 2004.
- [20] M. Klusch and X. Zhing, "Deployed Semantic Services for the Common User of the Web: A Reality Check," Proc. IEEE Int'l Conf. Semantic Computing (ICSC), 2008.
- [21] J. Lu, Y. Yu, D. Roy, and D. Saha, "Web Service Composition: A Reality Check," Proc. Eighth Int'l Conf. Web Information Systems Eng. (WISE '07) Dec. 2007.

- [22] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGunneess, B. Barsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, and K. Sycara, "Bringing Semantics to Web Services: The OWL-S Approach," Proc. First Int'l Workshop Semantic Web Services and Web Process Composition, July 2004.
- [23] S. McIlraith, T. Son, and H. Zeng, "Semantic Web Services," IEEE Intelligent Systems, vol. 16, no. 2, pp. 46-53, Mar. 2001.
- [24] S. McIlraith and D. Martin, "Bringing Semantics to Web Services," IEEE Intelligent Systems, vol. 18, no. 1, pp. 90-93, Jan. 2003.
- [25] I. Niles and A. Pease, "Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology," Proc. IEEE Int'l Conf. Information and Knowledge Eng. (IKE '03), 2003.
- [26] N. Oldham, C. Thomas, A. Sheth, and K. Verma, "METEOR-S Web Service Annotation Framework with Machine Learning Classification," Semantic Web Services and Web Process Composition, vol. 3387, pp. 137-146, Jan. 2005.
- [27] A.V. Paliwal, N. Adam, and C. Bornhoevd, "Adding Semantics through Service Request Expansion and Latent Semantic Indexing," Proc. IEEE Int'l Conf. Services Computing (SCC), July 2007.
- [28] A.V. Paliwal, N. Adam, H. Xiong, and C. Bornhoevd, "Web Service Discovery via Semantic Association Ranking and Hyperclique Pattern Discovery," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence, 2006.
- [29] A. Sajjanhar, J. Hou, and Y. Zhang, "Algorithm for Web Services Matching," Proc. Asia-Pacific Web Conference (APWeb), pp. 665-670, 2004.
- [30] G. Salton and C. Buckley, "Improving Retrieval Performance by Relevance Feedback," J. Am. Soc. for Information Science, vol. 41, no. 4, pp. 288-297, 1990.
- [31] G. Salton, A. Wong, and C.S. Yang, "A Vector Space Model for Automatic Indexing," Comm. ACM, vol. 18, pp. 613-620, Nov. 1975.
- [32] K. Verma, K. Sivashanmugam, A. Sheth, A. Patil, S. Oundhakar, and J. Miller, "METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services," Information Technology and Management J., vol. 6, pp 17-39, 2005.
- [33] http://www.musclenoe.org/research/sci_deliv_pub/D5.1_WP5_SoA_RevisedVersion_sept05.pdf, 2012.
- [34] H. Xiong, P. Tan, and V. Kumar, "Mining Strong Affinity Association Patterns in Data Sets with Skewed Support Distribution," Proc. IEEE Third Int'l Conf. Data Mining (ICDM), 2003.