

A P2P Platform for Socially Intelligent Web Service Publication and Discovery

Michael Pantazoglou and Aphrodite Tsalgatidou
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens, Greece
{michaelp, atsalga}@di.uoa.gr

Abstract

The ability of service providers to adequately describe and publish their Web services is a critical factor to the success in their discovery by interested parties. To date, most of the efforts made towards this direction have mainly focused on the use of ontologies. However, the resulted approaches impose centralized control of the ontology vocabularies and/or schemas, suffer from considerable interoperability issues and require complex, heavyweight technologies. In this paper we present our ongoing work on a peer-to-peer platform supporting publication and discovery of Web services in a socially intelligent manner. The novelty of our approach is that it combines the p2p architecture with the principles of the Social Web to enable autonomous, decentralized semantic annotation and maintenance of Web service descriptions, through the collective intelligence emerging by the social networking interactions between service providers and requesters.

1. Introduction

Service publication and discovery are two tasks of significant importance to any Web service-oriented architecture. Indeed, according to the service-oriented computing paradigm, service providers publish their offered Web services in appropriate registries, where service requesters may search to discover them. A critical factor in the success of such model of interaction is the expressiveness of the corresponding service descriptions. The more adequate and detailed a service has been described, the easier it becomes to a service requester to understand its capabilities and match them against his/her requirements.

Web services are traditionally described with the use of the Web Services Description Language (WSDL) [1], which unfortunately cannot adequately

represent their actual semantics, i.e. capabilities, inputs and outputs. To address this shortcoming, a number of ontology-based approaches in Web service description have emerged over the years, with OWL-S [2], WSMO [3], and SAWSDL [4] being the most prominent ones. Also, various research approaches in Web service publication and discovery were built on top of these specifications (e.g. see [5], [6], [7]).

Despite the significant research challenges that the aforementioned developments brought forth to the Web services community, they still fail in being widely deployed and used, due to a number of considerable bottlenecks they produce. Ontologies adopt a top-down modeling approach which often differs from the way experts view their knowledge domains. Moreover, the existence of multiple ontologies describing the same domain produces significant interoperability problems, which can only be avoided through the enforcement of centralized control. Alas, this in turn makes ontology maintenance (a constant task that needs to be performed to keep the ontology consistent with the real world) cumbersome and resource demanding. Finally, most of the technologies involved in processing ontologies and applying logical reasoning over their contents are complex and rather heavyweight.

Recently, a new computing paradigm has emerged along with the advent of the Web 2.0 technologies, namely the Social Web. According to the principles of this new paradigm, users form communities (social networks) to communicate and collaborate. Content in such communities is usually organized by means of user-defined descriptive keywords, also known as *tags*. The sets of categories that result from tagging are commonly referred to as *folksonomies*. Compared to ontologies, folksonomies offer greater flexibility and adaptability, as users need not agree on a detailed hierarchy or vocabulary [8]. Despite their distributed nature and lack of central control, folksonomies tend to

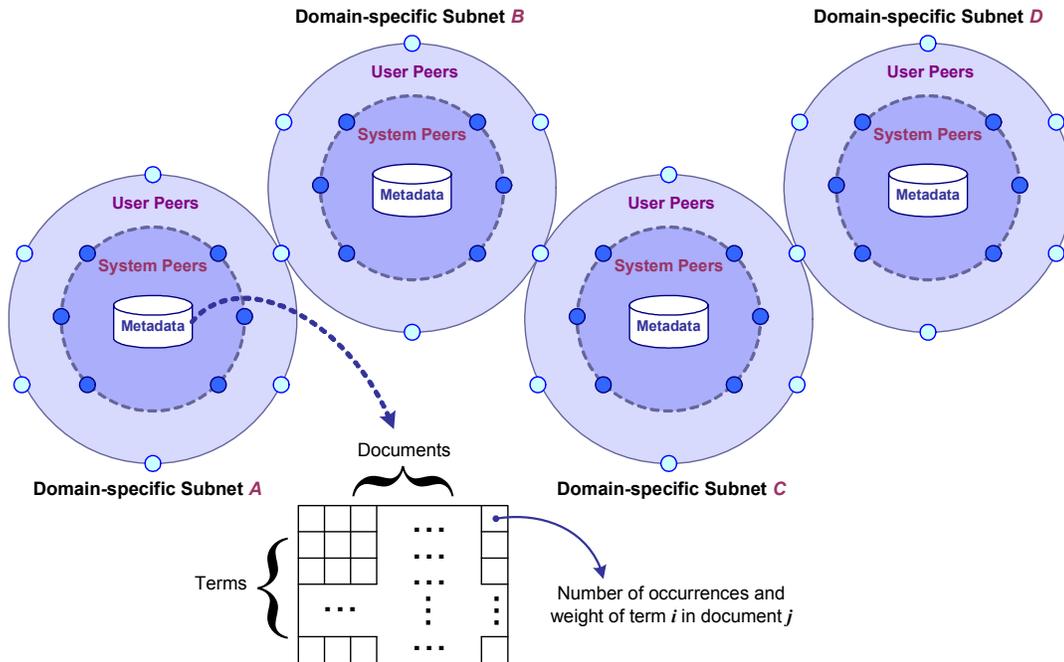


Figure 1. Domain-specific social networks of service providers and requesters organized in a Peer-to-Peer architecture

reach a stable form and represent the collective intelligence of their social network community under certain conditions, regarding the number of users and the duration of the tagging process.

In the light of these developments, we believe that Web service description, publication and discovery could be greatly improved if it was performed in a more autonomic, flexible, and decentralized manner, ensuring interoperability among service providers and requesters. To meet these challenges, we present in this paper an ongoing research on a peer-to-peer platform which enables socially intelligent Web service discovery and publication. Our approach is inspired by the Social Web principles. Specifically it exploits the collective intelligence that is derived by social networking interactions among service providers and requesters, with the aim of dynamically formulating folksonomies of Web services, which are distributed, organized in domains, and self-maintained.

Briefly, the remainder of this paper is structured as follows: Section 2 details our approach by presenting the p2p platform in terms of its architecture and delivered functionality; in Section 3 we present some early results in the implementation of the proposed platform; Section 4 contains a comparison of our work to similar efforts whereas Section 5 concludes the paper with a brief discussion on our future work plans.

2. Platform Architecture and Functionality

The platform for Web service publication and discovery that is proposed within this paper is built on top of a semi-decentralized p2p infrastructure, such as the one prescribed by the JXTA specification [9]. We took the decision to adopt a p2p approach for our platform inspired by the promising results that have become available in the literature within the last years. Indeed, p2p architectures have been extensively applied to various service discovery platforms (e.g. [14], [17]) with the aim of enhancing their overall performance, scalability, and fault tolerance. Moreover, there is a natural mapping between the Social Web paradigm, where humans interact and collaborate directly with each other without any supervising authority, and the p2p philosophy.

As it can be seen in Figure 1. the underlying p2p network is subdivided into domain-specific peer groups, or subnets, where service providers and requesters activating in the same application domain may join and interact with each other. Subnets may overlap in the sense that there may be peers (i.e. service providers and/or requesters) activating in more than one application domains, thereby being members of the corresponding peer groups.

At the heart of each domain-specific peer group lays an internal database where metadata of the Web

services published within the group are stored and maintained. The architecture of the internal database could be centralized or decentralized; however, this topic falls out of the scope of the work reported here and therefore is not examined any further.

The metadata stored in the internal database of each peer group take the form of a two-dimensional matrix, where rows correspond to individual terms, or tags, and columns represent the different documents. In our approach, we treat service, operation, and message elements of the WSDL descriptions as separate documents to enable their independent tagging and matchmaking during service discovery. Hence, the matrix holds the weight (quantified significance) and number of occurrences of each individual term within each document in the corpus. Such information is utilized for appropriately ranking the popularity of each term, and ultimately establishing consented semantic annotations for the published Web services, as we will see in the following paragraphs.

2.1. User Peers and Services

The proposed platform defines two main peer categories: user peers and system peers. The latter are described in the next paragraph, whereas the former are further split in two subcategories: service provider peers and service requester peers.

Service provider peers implement a set of services offering the required functionality for WS publishing and un-publishing. On the other hand, service requester peers implement a set of services enabling the execution of WS queries, as well as supporting the tagging of their Web services of interest, along with their operations, and input/output messages.

Let us briefly describe the aforementioned user peer services.

WS Publish. This service accepts as input the URL pointing to the WSDL description document of the Web service, and responds with a boolean indicator of success or failure for the publication success. The WSDL URL and the service provider name are appended to an advertisement document, the structure of which depends on the underlying p2p technology implementation. This advertisement is then periodically multicasted within the subnet, depending on its TTL. It should be noted that, under the hood, the publish service makes use of core services provided by the p2p infrastructure, and also activates the WSDL parsing service (described later in this section) provided by the system peers of the peer group.

WS Un-publish. This service un-publishes a previously published Web service based on its WSDL URL. Un-publishing is performed through interrupting the periodical scheduled multicast of the corresponding advertisement, followed by the deletion of all related documents in the database.

WS Query. Service requesters utilize this service to search for services which meet their requirements towards functionality, inputs and outputs. Queries and their responses are constructed with the use of the *Unified Service Query Language (USQL)*, an XML-based grammar which is part of our previous work [13]. With the USQL, users may properly set their search criteria towards the functionality and interface of the requested Web services, in the form of free text, or keyword-based, descriptions. The following snippet illustrates a sample USQL query for weather reporting Web services:

```
<USQLRequest ...>
<SearchCriteria ...>
  <Functionality>
    <Description>
      Get weather report for a given city or area code
    </Description>
  </Functionality>
  <Interface>
    <Input>
      <Description>the city/area code</Description>
    </Input>
    <Output>
      <Description>weather conditions</Description>
    </Output>
  </Interface>
</SearchCriteria>
</USQLRequest>
```

Upon discovery and matchmaking, the functionality, input and output descriptions are extracted from USQL queries, they are tokenized, and the resulting terms are used to search within the metadata matrix. The terms deriving from the desired functionality description are first used to search for service and operation documents, whereas subsequently the input/output ones are used for searching among the message documents that correspond to the input and output of the discovered operations. Next, interface (i.e. input/output) based matchmaking is performed. Being an instantiation of the Assignment Problem (AP), interface matchmaking has been implemented within the WS Query service with the use of the Hungarian method [18].

Finally, the matching services are assessed in a USQL response document which is returned to the

service requester. The following example illustrates such a response to our sample query.

```

<USQLResponse ...>
  <MatchedServices>
    <Service degreeOfMatch="1.0" ...>
      <Provider>Foreca</Provider>
      <Name>WeatherInfoService</Name>
      <WSDL>...</WSDL>
      <Port>WeatherInfoServiceSoap</Port>
      <Operation>GetWeatherReport</Operation>
    </Service>
  </MatchedServices>
</USQLResponse>

```

WS Tag. This service enables users to append their own descriptions on a published Web service. Upon submitting a description about a specific Web service, operation, or message, the following natural language processing tasks are performed: First, the description content is appropriately tokenized; next, each token is stemmed with the use of WordNet [11], an electronic English dictionary and stop words are eliminated; following that, each stemmed token is further expanded with synonym terms extracted from WordNet. The resulted list of tags is then properly used to update the respective document entry in the metadata matrix of the database.

As user peers stand for service providers and requesters, the p2p platform needs to support the necessary security mechanisms and policies, to avoid malfunctioning caused by malicious users. Nevertheless, such issues are not covered in this paper, as they are not part of our work. As far as user authorization and authentication are concerned, we rely on the existing implementations provided by the underlying p2p infrastructure.

2.2. System Peers and Services

System peers constitute the backbone of each domain-specific subnet in the p2p platform and are exclusively responsible for generating, processing, and maintaining the metadata stored in the internal database. In a sense they act as a middleware by providing the following set of core services:

WSDL Parsing. This service is activated each time a service provider publishes a new Web service within the peer group. It accesses the corresponding WSDL description document through the submitted URL and parses its content in order to extract and concatenate the name attribute and documentation element values

of the included service, operation, and message elements. Subsequently this information is analyzed in a way similar to the natural language processing that is applied to the tags submitted by user peers (see description of the WS Tag service). The analysis results are finally used to generate the documents that correspond to the processed services, operations, and messages in the WSDL document.

Hence, WSDL parsing produces a set of service, operation, and message documents consisting of term lists, which are finally used to populate the metadata matrix in the internal database.

Metadata Processing. This service is activated upon insertion of a new document or term in the metadata matrix. Its purpose is to calculate and update the weight and number of occurrences of each affected by the insertion term, per document. In performing this task, the metadata processing service implements the *Term Frequency – Inverse Document Frequency* (TF-IDF) algorithm [10], a popular heuristic that has been extensively applied in various information retrieval applications. Briefly, the TF-IDF algorithm is defined as follows:

Let $D = \{d_1, d_2, \dots, d_n\}$ be a collection of n documents. For each term t_j appearing in the collection, let n_{ij} be its number of occurrences in document d_i . Also, let n_j be the number of documents which contain the term t_j at least once. Then, the TF-IDF weight of t_j in d_i is computed as

$$w_{ij} = TF_{ij} \cdot IDF_j = \frac{n_{ij}}{|d_i|} \cdot \log\left(\frac{n}{n_j}\right)$$

where $|d_i|$ is the length (i.e. total number of terms) of document d_i .

Intuitively, the TF-IDF algorithm tells us that, the semantic importance of a term within the document not only depends on its number of occurrences within that document, but also on the frequency in which it is met throughout the whole corpus of documents. Hence, the higher a TF-IDF weight value is, the more semantically important the corresponding term is to the specified document.

2.3. Reaching Consensus on WS Semantics

Recent studies have evidenced that, the magnitude of social networking communities gradually transforms them into containers of collective intelligence. More specifically, the frequency distribution of tags in folksonomies tends to stabilize over time into power law distributions [12]. In our approach we exploit this

characteristic of social networks in order to let Web service semantics be autonomously shaped and maintained.

The TF-IDF heuristic presented in the previous paragraph provides the platform with a reliable indicator of the collective intelligence that gradually emerges within a peer group, expressed as consensus on the semantics of the advertised Web services, their offered operations and input/output messages. Indeed, the more users have used a specific term to tag a service, operation, or message, the higher the TF-IDF weight value of that term is for the corresponding document. Hence, for each document stored in the database of a domain-specific peer group, its top- k weighted terms formulate its semantics. Subsequently, the domain folksonomy is dynamically constructed by accumulating the top- k terms of all documents in the corpus. Note that, variable k can be parameterized for optimal performance and precision.

3. Early Implementation Results

We are currently developing the proposed p2p platform and its algorithms in a Java prototype, on top of the JXTA 2.5 implementation. We have utilized the JXTA peer group mechanisms to construct the various domain-specific social networks, and all services described in the previous section have been implemented as JXTA peer group service modules.

Within each peer group, Web services are published by appropriately constructing and multicasting Module Spec Advertisements (MSA). Specifically, the WSDL URL is set as value of the *SpecURI* sub-element, whereas the service provider name is set as value of the *Crtr* sub-element. Finally, the database lying in the middle of each peer group has been modelled by a relational schema and implemented in MySQL technology.

4. Related Work

The p2p architecture has been extensively applied in a number of related research efforts in Web services publication and discovery. In [14], the authors present the METEOR-S framework, which the ontology-based organization of registries and classification of Web services over a scalable p2p infrastructure. Although this approach shares some common ideas with our proposed platform in terms of domain-based organization, it makes the assumption that registry operators define and maintain their own domain ontologies, which are used to manually or semi-automatically annotate the published services. In our approach, semantic annotations are applied

autonomously, whereas ontologies are substituted by dynamically constructed folksonomies based on the input of service requesters and providers.

The notion of p2p registry organization for optimized service publication and discovery has also been applied in other research results, such as [17], which builds a distributed discovery service utilizing DAML-S semantic service descriptions to provide semantically-enhanced search capabilities. Our approach departs from such frameworks, mainly due to the fact that they all depend on existing domain ontologies, whilst our platform is capable of producing semantic metadata out of folksonomies populated by users in a distributed and flexible fashion.

In [16], another p2p-based approach to Web service discovery is proposed. The presented framework provides a p2p indexing system and associated p2p storage mechanisms to support large-scale and decentralized discovery of Web services. Searches are performed with the use of queries made of keywords, which, similar to the Chord [15] data lookup protocol, are directly forwarded to the right peer for execution. Our work can be considered complementary to that approach, as it is more focused on the way semantics are created for the published services within the p2p network. Moreover, in our platform, queries are not merely sets of keywords. Instead, they are well structured documents enabling requesters to express their requirements on service capabilities and interface either with the use of keywords or even free text descriptions.

5. Discussion

We presented a novel approach for socially intelligent Web services publication and discovery. The proposed platform exploits the p2p architecture and effectively combines multidisciplinary principles and techniques from areas such as social networking, information retrieval and natural language processing, to enable the autonomic and decentralized formulation of Web service semantics. Compared to the majority of related ontology-based approaches, our work differentiates in that Web service semantics are dynamically specified by the outcome of the collective intelligence that emerges through the online social interactions among service providers and requesters.

We would like to note that the work reported in this paper is at early stages and still ongoing. In this regard, we intend to finalize our prototype, which will be subsequently used in a number of simulations and experiments. The assessed results will allow us to further evaluate the efficiency and viability of our approach.

6. Acknowledgement

The work reported in this paper has been partially funded by the Special Account of Research Funds (ELKE) of the National and Kapodistrian University of Athens under contract 70/4/5829, and the European Commission under contract FP6-IST-511680 for the SeCSE project.

References

- [1] Christensen, E., Curbera, F., Meredith, G., and Weerawarana, S. 2001. Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001, available online at <http://www.w3.org/TR/wsdl> (accessed March 2008).
- [2] Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N., and Sycara, K. 2004. OWL-S: Semantic Markup for Web Services. W3C Member Submission, 22 November 2004, available online at <http://www.w3.org/Submission/OWL-S/> (accessed March 2008).
- [3] Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., and Fensel, P. 2005. Web Service Modeling Ontology. *Applied Ontology*, 1 (1), 2005, IOS Press, 77-106.
- [4] Farrell, J. and Lausen, H., Eds. 2007. Semantic Annotations for WSDL and XML Schema. W3C Recommendation 28 August 2007, available online at <http://www.w3.org/TR/sawSDL/> (accessed March 2008).
- [5] Naveen Srinivasan, N., Paolucci, M., and Sycara, K. 2004. An Efficient Algorithm for OWL-S Based Semantic Search in UDDI. In *Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004)*, San Diego, CA, USA, July 6, 2004, 96-110.
- [6] Keller, U., and Lausen, H. Semantic Web Service Discovery in WSMO. *Chapter 12 in Semantic Web Services: Theory, Tools and Applications*. Edited by Jorge Cardoso, Idea Group Inc. ISBN 978-1599040455. March 2007.
- [7] Dogac, A., Kabak, Y., and Laleci, G.B. 2004. Enriching ebXML registries with OWL ontologies for efficient service discovery. In *Proceedings of the 14th International Workshop on Research Issues on Data Engineering: Web Services for e-Commerce and e-Government Applications*, March 2004, 69-76.
- [8] Mikroyiannidis, A. 2007. Toward a Social Semantic Web. *IEEE Computer*, November 2007, Vol. 40 (11), pp. 113-115.
- [9] Li, G. 2001. JXTA: a network programming environment. *IEEE Internet Computing*, May/June 2001, Vol. 5 (3), 88-95.
- [10] Baeza-Yates, R. and Ribiero-Neto, B. *Modern Information Retrieval*. Addison Wesley, 1999.
- [11] Miller, G. Wordnet: A Lexical Database for English. *Communications of the ACM*, November 1995, Vol. 38 (11), 39-41.
- [12] Halpin, H., Robu, V., and Shepherd, H. 2007. The Complex Dynamics of Collaborative Tagging. In *Proceedings of the 16th International Conference on World Wide Web*, 211-220.
- [13] Pantazoglou, M., Tsalgatidou, A., and Athanasopoulos, G. 2006. Discovering Web Services and JXTA Peer-to-Peer Services in a Unified Manner. In *Proceedings of the 4th International Conference on Service Oriented Computing (ICSOC 2006)*, December 2006, Chicago, USA, 104-115.
- [14] Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S., and Miller, J. 2005. METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Information Technology and Management*, Springer, Vol. 6, 17-39.
- [15] Stoica, I., Morris, R., Karger, D., Kaashoek, F., and Balakrishnan, H. 2001. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of ACM SIGCOMM*, San Diego, CA, 2001, 149-160.
- [16] Schmidt, C. and Parashar, M. 2004. A Peer-to-Peer Approach to Web Service Discovery. *World Wide Web: Internet and Web Information Systems*, Kluwer Academic Publishers, Vol. 7, 2004, 211-229.
- [17] Thaden, U., Siberski, W., and Nejdli, W. 2003. A Semantic Web based Peer-to-Peer Service Registry Network. Technical Report 2003.
- [18] Kuhn, H.W. and Res, N. 1956. Variants of the Hungarian method for assignment problems. *Logistics Quarterly*, 3 (1956), pp. 95-110.