



## Controversy Corner

## Web service discovery based on goal-oriented query expansion

Neng Zhang<sup>a</sup>, Jian Wang<sup>a,\*</sup>, Yutao Ma<sup>a</sup>, Keqing He<sup>a</sup>, Zheng Li<sup>b</sup>, Xiaoqing (Frank) Liu<sup>c</sup><sup>a</sup>School of Computer Science, Wuhan University, Wuhan, China<sup>b</sup>School of Computer and Information Engineering, Henan University, Kaifeng, China<sup>c</sup>Computer Science and Computer Engineering Department, University of Arkansas, Fayetteville, USA

## ARTICLE INFO

## Article history:

Received 13 July 2017

Revised 9 March 2018

Accepted 18 April 2018

Available online 21 April 2018

## Keywords:

Web service

Service discovery

Service-based system (SBS)

Service goal knowledge

Query expansion

## ABSTRACT

With the broad adoption of service-oriented architecture, many software systems have been developed by composing loosely-coupled Web services. Service discovery, a critical step of building service-based systems (SBSs), aims to find a set of candidate services for each functional task to be performed by an SBS. The keyword-based search technology adopted by existing service registries is insufficient to retrieve semantically similar services for queries. Although many semantics-aware service discovery approaches have been proposed, they are hard to apply in practice due to the difficulties in ontology construction and semantic annotation. This paper aims to help service requesters (e.g., SBS designers) obtain relevant services accurately with a keyword query by exploiting domain knowledge about service functionalities (i.e., service goals) mined from textual descriptions of services. We firstly extract service goals from services' textual descriptions using an NLP-based method and cluster service goals by measuring their semantic similarities. A query expansion approach is then proposed to help service requesters refine initial queries by recommending similar service goals. Finally, we develop a hybrid service discovery approach by integrating goal-based matching with two practical approaches: keyword-based and topic model-based. Experiments conducted on a real-world dataset show the effectiveness of our approach.

© 2018 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Recent years have witnessed a software development paradigm transformation from component-based to service-based due to the broad adoption of service-oriented architecture (SOA) and its related technologies (He et al., 2014). The service-based paradigm can not only reduce the cost, time, and efforts required for software development, but also promote the reusability, agility, and quality of the resulting systems (Bano et al., 2014). A large number of software systems have been developed by discovering and composing loosely-coupled Web services provided by different organizations (He et al., 2016).

Many companies like Google, Facebook, and Amazon have encapsulated some of their capabilities as Web services and published them through Web service registries such as ProgrammableWeb<sup>1</sup> (PW) and Mashape.<sup>2</sup> For example, as of June 1, 2017, more than 17,600 and 6,000 services have been regis-

tered at PW and Mashape, respectively. These services can be used to create value-added services (referred to as composite services or mashups) or other service-based systems (SBSs). Generally speaking, the process of building an SBS consists of three stages (He et al., 2016): 1) *system planning*, which determines tasks needed to implement the functionalities of the SBS; 2) *service discovery*, which identifies a set of candidate services for each task; and 3) *service selection*, which selects one service from each set of candidate services to fulfil the constraints of system quality, e.g., reliability and throughput. In this research, we focus on the service discovery stage, i.e., *how to retrieve a set of candidate services that can perform a specific task from a service registry?*

There are many semantically similar functionalities provided by different services in a registry. The keyword-based search technology adopted by existing service registries is insufficient to discover services whose functionalities are similar to a query (e.g., the keyword description of a task) (Chen et al., 2013; Wang et al., 2017). For example, many services in PW provide similar functionalities like “book hotel,” “book lodge,” and “book accommodation.” Using any one of these functional keywords as a query, services that contain similar functionalities are hard to be retrieved by PW. As an example, we submitted “book lodge” to the PW service search engine, but the search engine returned only six services that ex-

\* Corresponding author.

E-mail addresses: [nengzhang@whu.edu.cn](mailto:nengzhang@whu.edu.cn) (N. Zhang), [jianwang@whu.edu.cn](mailto:jianwang@whu.edu.cn) (J. Wang), [ytma@whu.edu.cn](mailto:ytma@whu.edu.cn) (Y. Ma), [hekeqing@whu.edu.cn](mailto:hekeqing@whu.edu.cn) (K. He), [zhengli\\_hope@whu.edu.cn](mailto:zhengli_hope@whu.edu.cn) (Z. Li), [frankliu@uark.edu](mailto:frankliu@uark.edu) (X. Liu).<sup>1</sup> <https://www.programmableweb.com/>.<sup>2</sup> <https://www.mashape.com/>.

explicitly contain words *book/booking* and *lodge/lodging* in their descriptive data.

There are two essential elements in service discovery: the queries specified by service requesters and the service discovery approaches adopted by service registries. Accordingly, to address the drawback of keyword-based technology, two research directions are dedicated to discovering similar services.

One research direction is to expand queries with relevant concepts extracted from lexical databases (e.g., WordNet<sup>3</sup> (Miller, 1995)) or domain ontologies. The primary limitation of existing query expansion approaches (Kokash et al., 2006; Aljoumaa et al., 2011; Paliwal et al., 2012; Ma et al., 2013) is that they purely rely on external knowledge bases and do not leverage the local knowledge about service registries. Therefore, the expanded queries may be ineffective for service discovery in a specific registry. For example, the highly relevant concepts (including the synonyms, hypernyms, and hyponyms) of *book* and *hotel* that can be extracted from WordNet are

*book*: reserve, record, enter, hold;

*hotel*: lodge, hostel, building, edifice.

Many of the extracted concepts are irrelevant, e.g., *record*, *enter*, *hold*, *building*, and *edifice*, while they do not include some desired concepts, e.g., *accommodation*, which will result in retrieving many unnecessary services and missing some relevant ones. Similarly, some concepts extracted from domain ontologies may also be irrelevant while some relevant ones are missing. Moreover, ontology-based query expansion approaches are limited by the fact that there is usually no suitable domain ontology at hand.

The other research direction is to develop semantics-aware service discovery approaches. The basic idea is to describe services and queries using ontology-based semantic Web service description languages, e.g., SAWSDL<sup>4</sup> (Semantic Annotations for WSDL and XML Schema), OWL-S<sup>5</sup> (Ontology Web Language for Services), and WSMO<sup>6</sup> (Web Service Modeling Ontology), and to design logic-based reasoning algorithms for retrieving similar services (Crasso et al., 2011; Klusch et al., 2016). This principle has developed a large number of methods and techniques (Klusch and Kaufer, 2009; Wei et al., 2011; García J et al., 2012; Roman et al., 2015; Rodríguez Mier et al., 2016; Chen et al., 2017a), which have been shown to be more useful than the keyword-based technology. However, these approaches are difficult to apply due to several factors (Crasso et al., 2011; Aznag et al., 2014; Cassar et al., 2014): 1) constructing and maintaining ontologies may be difficult, 2) it requires considerable efforts in manually annotating services and queries using ontology-based description languages, and 3) logic-based reasoning algorithms usually lead to high complexity. Alternatively, there are many non-logic-based semantics-aware service discovery approaches (Wang et al., 2017; Aznag et al., 2014; Cassar et al., 2014; Li et al., 2014; Naim et al., 2016) proposed by leveraging latent topic/factor models, e.g., Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999), Latent Dirichlet Allocation (LDA) (Blei et al., 2003), and Bi-term Topic Model (BTM) (Cheng et al., 2014). These approaches can help retrieve similar services; however latent topics learned by these models are too coarse to precisely match services to queries, which limits the performance.

In Zhang et al. (2017), we proposed an approach to extracting service goals (i.e., service functionalities), e.g., <book, hotel, null> and <retrieve, pricing information, {for booking hotel}>, from textual descriptions of services. In this paper, we propose a service discovery approach by utilizing the service goals, aiming to help service requesters obtain similar services accurately with a simple

keyword query. The key part of our approach is a goal-oriented query expansion approach. Given a query, a set of semantically similar service goals is recommended. Informed by the recommendations, service requesters can gain a better understanding of service functionalities relevant to their functional requirements and select some desired goals as an expanded query. The contributions of this research are outlined below:

1. We propose a method for service goal clustering. Similarities between service goals are measured based on the semantic similarities of words in WordNet. Moreover, we distinguish the importance of three different types of words in service goals, which contributes to more accurate similarities.
2. We propose a goal-oriented query expansion approach based on service goal clusters. Faced with a query, service goals assigned to the clusters that are similar to the query are recommended, from which the service requester can expand the query by selecting some appropriate goals. Note that we do not adopt an automatic strategy that directly uses all recommended goals as an expanded query without the requester's involvement because some of the recommendations may not be appropriate for the requester, which will lead to returning irrelevant services.
3. We propose a hybrid service discovery approach based on the expanded query. The approach integrates a goal-based service matching mechanism with two widely adopted approaches: a keyword-based approach and an approach based on the LDA topic model.

We conducted experiments on a real-world service dataset crawled from PW. Fourteen subjects were recruited to build a set of 21 queries and evaluate the recommended service goals and retrieved services for each query. According to the evaluation results, the proposed goal-oriented query expansion approach can efficiently recommend semantically similar service goals for queries; and the proposed service discovery approach achieves better performance than several existing approaches.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 gives an overall framework of the proposed service discovery approach. In Section 4, we briefly introduce our previous service goal extraction approach and describe the method for service goal clustering. Section 5 describes the process of service discovery based on goal-oriented query expansion. Section 6 presents the experiments and evaluations. Section 7 discusses the contributions, limitations, and threats to validity. Section 8 concludes this work and introduces our future work.

## 2. Related work

### 2.1. Web service discovery

Web Service discovery refers to finding services that can satisfy functional requirements specified by a user query. As one of the core techniques in SOA, it has been studied extensively in the past two decades. Most existing works on service discovery can be categorized into two groups: syntactic approaches (Dong et al., 2004; Wang and Stroulia, 2003; Cong et al., 2015) and semantics-aware approaches (Wang et al., 2017; Aznag et al., 2014; Klusch and Kaufer, 2009; Wei et al., 2011; García J et al., 2012; Roman et al., 2015; Rodríguez Mier et al., 2016; Cassar et al., 2014; Li et al., 2014; Naim et al., 2016; Lu et al., 2016; Chen et al., 2017b).

*Syntactic approaches* mainly discover services through matching the keywords of services with those of queries using information retrieval (IR) techniques, e.g., vector space model (VSM) and term frequency-inverse document frequency (TF-IDF) (Manning et al., 2009). Although syntactic approaches can benefit from IR techniques and have incorporated several enhancement methods, e.g.,

<sup>3</sup> <http://wordnet.princeton.edu/>.

<sup>4</sup> <https://www.w3.org/TR/sawSDL/>.

<sup>5</sup> <https://www.w3.org/Submission/OWL-S/>.

<sup>6</sup> <https://www.w3.org/Submission/WSMO/>.

Download English Version:

<https://daneshyari.com/en/article/6885293>

Download Persian Version:

<https://daneshyari.com/article/6885293>

[Daneshyari.com](https://daneshyari.com)