



A semantic similarity measure integrating multiple conceptual relationships for web service discovery



Fuzan Chen^a, Chenghua Lu^{a,b}, Harris Wu^{c,*}, Minqiang Li^{a,d}

^a College of Management and Economics, Tianjin University, Tianjin, China

^b Department of Information Technology and Decision Sciences, Old Dominion University, Norfolk, VA 23529, USA

^c College of Pearl River, Tianjin University of Finance and Economics, Tianjin, China

^d State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, China

ARTICLE INFO

Article history:

Received 21 December 2015

Revised 16 September 2016

Accepted 17 September 2016

Available online 19 September 2016

Keywords:

Web service

Web service discovery

Semantic similarity

Ontology

Conceptual relationship

ABSTRACT

The process of Web service discovery identifies the most relevant services to requesters' service queries. We propose a new measure of semantic similarity integrating multiple conceptual relationships (SIMCR) for Web service discovery. The new measure enables more accurate service-request comparison by treating different conceptual relationships in ontologies such as is-a, has-a and antonymy differently. Each service or request is represented by vectors of terms (or words) that characterize both the interface signature and textual description. The overall semantic similarity is computed as a weighted aggregation of interface similarity and description similarity. The experimental results confirm the effectiveness of the proposed semantic similarity measure. As demonstrated in this study, the semantic Web service discovery method based on the proposed similarity measure outperforms existing state-of-the-art discovery methods in terms of precision, recall and F-measure. The proposed semantic similarity measure has wider applications such as to improve document classification or clustering, and to more accurately represent and apply knowledge in expert and intelligent systems.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The paradigm of service-oriented computing (SOC) has received a lot of attention in support of dynamic, low-cost, and cross-organizational construction of distributed applications in heterogeneous environments (Joshi, Yesha, & Finin, 2014). SOC encourages enterprises to build distributed applications, such as business analytics and stock trading, using externally offered Web services which can be described, published, located, and invoked via the Web. Web service discovery is to identify the most relevant services in response to a requester's query by comparing Web services and the requester's requirement. Service composition and service recommendation highly rely on the accuracy of service discovery. The success of SOC relies on the assumption that the appropriate Web services are found to accomplish the user's requests (Zhao, Tanniru, & Zhang, 2007). As popularity of SOC and availability of Web services increase, to retrieve most relevant Web services for a specific service request has triggered a considerable amount of

research effort (Crasso, Zunino, & Campo, 2011; Farrag, Saleh, & Ali, 2013; Narock, Yoon, & March, 2014; Sangers, Frasincar, Hogenboom, & Chepin, 2013; Wang, De, Cassar, & Moessner, 2015).

In the early stage, service registries were built on Web Services Description Language (WSDL) and Universal Description, Discovery, and Integration (UDDI) protocol; the corresponding service searching engines usually depended on a syntactic keyword matchmaking mechanism. Unfortunately, they suffer from a lack of depth. As an example, customers send queries that include keywords relevant to their requirements and receive a result with Web services advertised with the same keywords that are assumed to be the 'right' services. It is impractical, however, to expect that providers and requesters share an identical understanding of an application domain or that the appropriate services can be exactly matched to the requirement.

To support knowledge-based intelligent service matching, the Semantic Web Service approach (Burstein et al., 2005) utilizing an ontology, which is a model of knowledge from domain experts about the business process, has been proposed. Several Semantic Web Service description frameworks have also been developed, such as Ontology Web Language for Services (OWL-S), and Web Service Modeling Ontology (WSMO). With well-defined semantic associations existing between words organized in the ontology, the semantic Web service approach is capable of

* Corresponding author.

E-mail addresses: fzchen@tju.edu.cn (F. Chen), iamluchenghua@sina.com (C. Lu), Hwu@odu.edu (H. Wu), mqli@tju.edu.cn (M. Li).

autonomous publication, discovery, and execution of Web services at a semantic rather than syntactic level. For example, with ontologies of financial and economic knowledge, new component of semantic Web services support business intelligence (BI) and executive information system (EIS) utilizing financial concepts (e.g. income, cash flow, cost, sale, profit, and revenue.) and semantic associations between them. Recently, some efforts (Meditskos & Bassiliades, 2011; Mezni, Chainbi, & Ghedira, 2014) have been done to extend the original UDDI repositories to support semantic service matchmaking. For instance, the semantically extended JUDDI (<http://juddi.apache.org/>), MashUDDI (Meditskos & Bassiliades, 2011) and UDDI4J (<http://uddi4j.sourceforge.net/>) provide semantic annotation model of Web services. Di Modica, Tomarchio, and Vita (2011) and Wang, Zhu, Sun, Zou, and Yang (2013) adopt a distributed service registry paradigm to improve the scalability, robustness of service discovery. Nowadays, the semantic discovery mechanism has become an inevitable trend (Bener, Ozadali, & Ilhan, 2009; Deng, Yin, Li, Wu, & Wu, 2008; García, Ruiz, & Ruiz-Cortés, 2012; Dong & Hussain, 2014; Farrag et al., 2013; Zapater, Escrivá, García, & Durá, 2015).

Aiming at a comprehensive framework of utilizing semantic information for intelligent service discovery, this paper proposes a semantic similarity measure that can be applied to both textual description and interface signature of Web services. The proposed similarity measure integrating multiple conceptual relationships (SIMCR) is developed on the basis of relational semantic distance among concepts in WordNet and other ontologies. Similarity measures for nouns/verbs organized in a hierarchical structure and adjectives organized in a bipolar structure are defined and differentiated. The proposed SIMCR takes account of multiple conceptual relationships in order to offer a refined representation and matching of the service's capabilities. The proposed approach measures similarity among terms in interface signature and textual description differently. The evaluation in this study shows that the proposed approach outperforms other state-of-the-art semantic Web service discovery methods in terms of precision, recall and F-measure. The performance improvement can be mainly attributed to the proposed semantic similarity measure SIMCR, and the differential treatment of SIMCR in textual description and interface signature.

This paper is organized in six sections. Section 2 summarizes the related work in previous research. Semantic similarity measure among concepts and semantic similarity among Web services and requests are presented in Section 3 and Section 4, respectively. Evaluation and experiments are presented in Section 5. Section 6 concludes this study and presents the future work and directions.

2. Related work

Web Service Discovery is the process to retrieve appropriate Web services to accomplish the users' request. Providers publish their Web services (service description documents) according to open standards such as WSDL, OWL-S, and WSMO, etc. The published services are discovered by a service search engine and organized by the composition engine (Al Ridhawi & Karmouch, 2015; Joshi et al., 2014). The service search engines use metadata, an abstract representation of service or request, to match services to requests. Some examples of service metadata include keywords of interfaces, terms associated with textual descriptions, Petri-nets representing service processes. The most widely used metadata are terms from the interface signature (i.e., the "Input" section and "Output" section) and/or the textual description of a Web service's capabilities. As interface signature provides a derived and concise description of the functions offered by the Web service, a straightforward mapping approach has been to extract a set of keywords

or terms from the interface signature, then the service discovery system evaluates the relevance between sets associated to service and request (Al Ridhawi & Karmouch, 2015; Bener et al., 2009; Farrag et al., 2013; Ou, Zeng, Xiang, Peng, & Li, 2011; Plebani & Pemici, 2009). For instance, Plebani and Pemici (2009) evaluated similarity among service/request interfaces by using a bipartite graph matching approach. Ou et al. (2011) proposed a method of service query that matches the concept sets of the considered I/O interfaces. The idea of concept relaxation was combined with a unidirectional bipartite graph matching approach to measure the similarity of concept sets. Farrag et al. (2013) measured the minimum semantic distance between I/O interface concepts of the request and available services. Interface matching, however, has the inherent limitation of low recall due to the insufficient information contained in the interface signature.

In real applications, providers offer a textual description (the 'textDescription' section in a service profile) of functional capabilities of the Web service. It provides more specific information than the interface signature provides. Consequently, some researchers utilize textual descriptions for the service matching process (Kapitsaki, 2012; Sangers et al., 2013; Wu, Chen, Zheng, Lyu, & Wu, 2014; Yang, Zhang, & Chen, 2008). Yang et al. (2008) and Sangers et al. (2013) abstracted a term set from the section 'textDescription', and then used a semantic approach to evaluate the degree of relevance of sets for service and request. Kapitsaki (2012) proposed a semantic Web service matching approach where the context of the service was constructed by keyword retrieval in WSDL interface signature and textual service description. Additionally, process metadata describing how the service implements the 'Process' part of the advertisement can be used to improve the searching performance, especially for the retrieval of composite services (Paulraj, Swamynathan, & Madhaiyan, 2011; Shafiq et al., 2014). However, a normative framework for this type of service process extraction and translation is still immature.

Web service discovery methods can be syntactic or semantic. The service search engines are quite different depending on the discovery methods and the underlying Web service description standards utilized. WSDL is a well-adopted standard for non-semantic service description, whereas OWL-S and WSMO use more advanced protocols which include semantic annotation of Web services. The syntactic or non-semantic methods retrieve services in a syntactic matching level where keywords in a service request are matched to keywords in a WSDL description (Crasso, Zunino, & Campo, 2008; Wu, 2012). Most of syntactic methods use information retrieval (IR) techniques to measure the similarity between request and service, such as Term Frequency-Inverse Document Frequency (TF-IDF) and Vector Space Model (VSM) (Crasso et al., 2008). As a simple keyword-matching approach, the syntactic service discovery method may fail to retrieve relevant services due to the heterogeneity of Web services and requests.

In contrast to the syntactic Web service discovery, semantic methods use ontology as a key mechanism to globally define or reference a common understanding and annotate Web Services with the common understanding. Semantic methods identify semantic similarities (i.e., the relevance) between the request and the available services utilizing the power of ontologies. When compared to syntactic methods, semantic methods are capable of automated Web Service Discovery as well as identifying the similarities more accurately between the functionalities of Web service and requests (Crasso et al., 2011; García et al., 2012). Currently, semantic matchmaking services can be similarity-based or logic-based. Similarity-based approaches use text similarity to measure the degree of relevance between the service and the request. Referencing logical concepts and rules in ontologies, logic-based approaches match requests and services using either description logic (DL) based subsumption reasoning or taxonomic structural reason-

ing on service/request semantics (Dasgupta, Aroor, Shen, & Lee, 2014; Klusch, Fries, & Sycara, 2009). This study focuses on the similarity-based methods.

The foundation of semantic Web service discovery is to estimate the semantic similarity of terms using concepts. There are mainly two types of semantic similarity measurement methods, ontological methods and web search engine-based (e.g., Google) methods. Ontological method is the most popular way to measure semantic similarity among terms (Li, Bandar, & McLean, 2003; Liu, Quan, Feng, & Qiu, 2010). WordNet (Miller, 1995) is the most common ontology in semantic Web service discovery. In contrast, Cilibrasi and Vitanyi (2007) used a search engine method to examine results retrieved by the Google search engine, and the concept similarity was measured by a normalized Google distance. Similarly, Bollegala, Matsuo, and Ishizuka (2011) calculated similarity among concepts by examining page counts and text snippets retrieved from a Web search engine. Unfortunately, the search engine method has large inconsistency and the information retrieved by the search engine contains much noise, resulting in inaccuracy of the similarity measurement.

When using a hierarchical ontology, two types of approaches are usually used to estimate the similarity of concepts, information content approaches and path-based approaches. As an example of information content approach, Li et al. (2003) measured concept similarities by estimating the probability of each concept according to a corpus database. Path-based approaches compute similarities according to a network structure rather than using statistical information of words. For instance, Rada, Mili, Bicknell, and Blettner (1989) measured the semantic similarity of concepts by the minimum number of edges connecting them. Liu et al. (2010) estimated the similarity among concepts based on the sum of weighted lengths in a path.

Although the hierarchical WordNet covers ISA, HASA, and ANT relationships, only the ISA relationship is commonly considered for the concept similarity measure (Farrag et al., 2013; Li et al., 2003; Rada et al., 1989). Furthermore, most Web service discovery research extracts nouns and verbs and omits adjectives from a Web service description (Deng et al., 2008; Kapitsaki, 2012; Sangers et al., 2013). Adjectives, such as 'highest' and 'lowest' that are contained in the Web service's descriptions, may provide useful information for a similarity measure of request and service. In the domain of text classification, Wiemer-Hastings (2004) stated that different segments (i.e., various syntactic roles or parts of speech) in a text made different contributions to the similarity measure among texts. Oliva, Serrano, del Castillo, and Iglesias (2011) demonstrated that adjectives had an important influence on the computation of text similarity. Consequently, we can conclude that the consideration of multiple conceptual relationships and adjective terms contribute to the effective estimation of similarity between services and requests.

To our knowledge, no existing semantic service discovery methods have considered multiple conceptual relationships, including is-a (ISA), part-whole (has-a, or HASA), and antonymy (ANT), in combination with adjectives and adverbs in measuring the textual similarity. For an accurate similarity measurement, this study develops two similarity measurement formulas, one for nouns/verbs and another for adjectives/adverbs, utilizing multiple conceptual relationships from WordNet. Moreover, the discovery method in this study utilizes both textual description and interface signatures of Web services and service requests.

3. Similarity measure between terms

In this study, the most widely used ontology of English vocabulary, i.e., WordNet (Miller, 1995), is used for the annotation of Web services. WordNet represents each concept as a set of synonymous

words, also called synset, and explicitly encodes related concepts as semantic relationships (i.e., edges between concepts) to indicate synonyms, antonyms, or also-see references as offered in any dictionary or thesaurus. In WordNet, concepts of nouns and verbs are organized in a hierarchical structure while concepts of adjectives and adverbs are grouped in bipolar structures. The semantic relationships formed between nouns and verbs cover is-a, part-whole (has-a), and antonymy, whereas the semantic relationships of adjectives and adverbs include similar-to and antonymy.

The terms in Web services' interface signature and functional description cover three kinds of senses, i.e., nouns, verbs, and adjectives/adverbs. As the foundation of our proposed similarity measure, similarity of nouns and verbs and similarity of adjectives are defined differently by a semantic distance approach. For nouns and verbs organized in a hierarchical structure, in contrast to conventional approaches utilizing only the ISA relationship, we examine three kinds of semantic relationships including ISA, HASA, and ANT. For adjectives grouped in a bipolar structure, the possible semantic relationships similar-to and ANT are both considered in the similarity measurement. A notation of semantic distance, essentially the length of a weighted path, is used to estimate the similarity of concepts.

3.1. Similarity measure for nouns and verbs

As mentioned previously, nouns and verbs in WordNet are organized in a hierarchical structure where generic concepts are located at the upper levels and specific concepts are located at the lower levels. An edge connecting parent-child concept denotes either an ISA relationship or a HASA relationship; meanwhile an alternative edge between-siblings concept denotes an ANT relationship.

Fig. 1 shows a segment of WordNet. For instance, 'publication' is a general concept which includes three sub-concepts: 'book', 'periodical', 'magazine'. The semantic relationship between 'book' and 'publication' is an ISA relationship, the relationship between 'song-book' and 'chapter' is a HASA relationship, and an ANT relationship exists between 'professional' and 'amateur'.

Path-based methods are commonly used in the measurement of semantic similarity for words in a hierarchical or taxonomy structure (Liu et al., 2010; Rada et al., 1989). A path-based method uses the length of the shortest path connecting the two concepts to measure their semantic similarity. However, the link strength of adjacent concepts, often referred to as semantic distance or edge weight, may be different. Some aspects affecting the semantic distance include the type of semantic relationship, depth, and local density (number of children to a parent concept). This study uses a weighted path to calculate the similarity or semantic distance of concepts.

3.1.1. Weighting the edges among concepts

Semantic distances between adjacent concepts in a hierarchical structure vary with semantic relationships among them. This section explains how to weight the edge in the hierarchical structure.

Generally, for the parent-child concepts with an ISA relationship or a HASA relationship, concepts in a lower level have more specific semantics than concepts in an upper level. Thus the semantic distance between lower-level concepts is less than that of concepts in upper levels. Furthermore, the overall semantics mass for a given concept is of a certain amount. If a concept contains more sub-concepts than a different concept in the same level, it may indicate that the concept is more adequately defined than the concept containing fewer sub-concepts. In other words, the semantic distance between concept and its child concept is also dependent on the local density. Therefore both the depth and local density of the considered concepts should be examined to measure the weight of an ISA or a HASA edge.

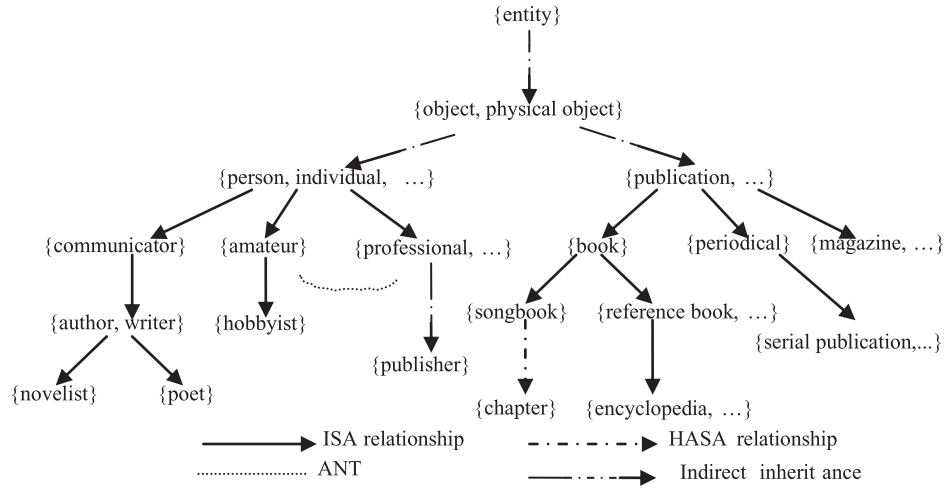


Fig. 1. A segment of WordNet.

Below are details of how edges of different conceptual relationships are weighted. For an ISA edge, a negative exponential function is used to calculate the weight (Eq. 1). Therein, e denotes the considered edge, h and n are the depth of the upper concept and the number of its sub-concepts, respectively. Eq. (1) indicates that the semantic distance of concept-pair with an ISA relationship in the upper level is farther than one in the lower level. The equation also shows that a concept is farther from its insufficiently defined child concepts than adequately defined child concepts. The weight of an ISA edge scales to $[0, 0.5]$. When either h or n increases, the weight of an ISA edge decreases and it decreases to zero as h increases to infinity.

$$W_{ISA}(e) = \frac{1}{2 + \lg n} e^{-\lg(h+1)} \quad (1)$$

A HASA edge has different implications than one from an ISA edge. An ISA relationship implies that a child concept inherits the semantics of the parent concept. For instance, a novelist is an author who concentrates specifically on a novel rather than a different type of text. However, a HASA relationship implies that a parent concept comprises of a child concept, i.e., the child concept is a part of the parent concept. For instance, a chapter is a part of a songbook. If an ISA edge and an ASA edge are associated with the same depth and local density, the semantic distance of concepts in a HASA relationship should be longer than that in an ISA relationship. From this perspective, we use Eq. (2) to estimate the weight of a HASA edge. The weight of a HASA edge scales to $[0, 1]$. The weight of a HASA edge decreases with increasing h and decreasing n . It decreases to zero as h increases to infinity.

$$W_{HASA}(e) = \left(1 - \frac{1}{2 + \lg n}\right) \cdot e^{-\lg(h+1)} \quad (2)$$

An ANT relationship is a special semantic relationship between sibling concepts. The semantic distance associated with an ANT edge is affected by the depth rather than the local density. The antonymy degree decreases with increasing depth. Eq. (3) is used to estimate the weight of ANT edge.

$$W_{ANT}(e) = f + e^{-\lg(h+1)} \quad (3)$$

In Eq. (3), f is a penalty factor that increases the semantic distance between concepts with an ANT relationship. The weight of an ANT edge scales to $[f, f+1]$. We use $f=1$ so that the semantic distance of concepts in an ANT relationship is always greater when compared to an ISA or a HASA relationship.

As shown in Fig. 1, the weight of 'chapter-songbook' edge with HASA relationship is different from the weight of edge among

'encyclopedia-reference book' with ISA relationship, even though they are in the same hierarchy. According to the previous estimation, the weight of edge between 'encyclopedia' and 'reference book' is less than that of 'chapter' and 'songbook' due to the effect of semantic relationship. In addition, the three concepts of 'communicator', 'amateur', and 'professional' have sibling relationships, but there exists an obvious ANT relationship between 'amateur' and 'professional'. The weight of edge between 'amateur' and 'professional' is much larger than that of 'communicator' and 'amateur' which has common sibling relationship.

3.1.2. Estimation of the semantic similarity for nouns and verbs

The semantic similarity of concepts is measured by a notion of semantic distance keeping the shortest path (SP) connecting them. The semantic distance between concepts in the hierarchy is the sum of weights on the edges included in the shortest path connecting them, as shown in Eq. (4). In Eq. (4), e_i , e_j , and e_k are the ISA, HASA and ANT edges respectively, these are included in shortest path connecting values, c_1 and c_2 . Note that the semantic distance is 0 when two words belong to the same synset. For example, 'person' and 'individual' in Fig. 1 belong to the same synset and their semantic distance is 0.

$$d_H(c_1, c_2) = \sum W_{ISA}(e_i) + \sum W_{HASA}(e_j) + \sum W_{ANT}(e_k) \quad (4)$$

The semantic distance is significantly different from the steps in the shortest path. For example, the steps in the shortest path between 'novelist' and 'communicator' (i.e., 'novelist-author-communicator') is the same as the steps between 'novelist' and 'poet', but the semantic distances are different. However, 'novelist' is more similar to 'poet' than 'communicator'.

Semantic distance and semantic similarity are inversely related measurements. The shorter the semantic distances between concepts, the greater the semantic similarity values and therefore the more similar they are. To scale the similarity value from 0 to 1, we use a negative exponential function of semantic distance to compute the similarity between concepts, as shown in Eq. (5) where d_H is the semantic distance of c_1 and c_2 in the hierarchical structure calculated by Eq. (4).

$$SimNV(c_1, c_2) = e^{-d_H} \quad (5)$$

Note that semantic distance between words within the same synset is 0, i.e., the similarity is 1 when two words share the same meaning. For example, 'person' and 'individual' in Fig. 1 belong to the same synset and their semantic similarity is 1.

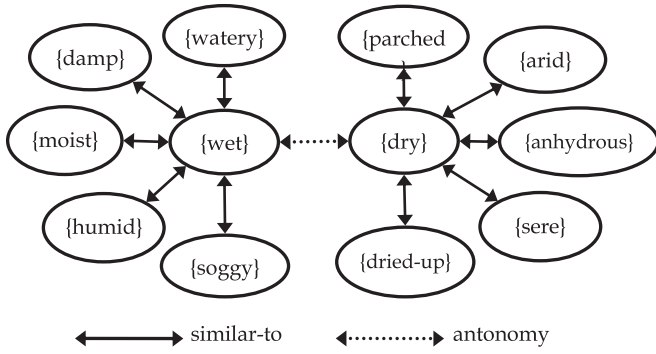


Fig. 2. A bipolar structure of adjectives in WordNet.

3.2. Similarity measure for adjectives

Adjectives in WordNet are organized in bipolar structures where two opposing words have an ANT relationship. Antonyms like wet and dry reflect strong semantic contrast. Each of these polar adjectives in turn is linked to a number of “semantically similar” ones: dry is linked to parched, arid, desiccated, and bone-dry; and wet to soggy, waterlogged, etc. Semantically similar adjectives are “indirect antonyms” of the central member of the opposite pole. Fig. 2 shows a bipolar structure of adjectives in WordNet. The semantic relationship of ‘wet’ and ‘dry’ is an ANT relationship represented by a broken line. The similar-to relationship exists between ‘watery’ and ‘wet’ represented by a solid line.

We present a gloss-based method to measure the semantic similarity between. The gloss of an adjective is a short textual definition that briefly describes the meaning of a concept. The gloss-based method uses the gloss of adjectives provided in WordNet or other dictionaries to estimate the semantic similarity of adjectives. For instance, the gloss of ‘wet’ is ‘covered or soaked with a liquid such as water’. Note that two adjectives in one synset share the same gloss.

The Vector Space Model (VSM) (Billhardt, Borrajo, & Maojo, 2002; Chen & Chiu, 2012) is used to estimate the semantic distance of adjectives in this study. It represents texts or documents as vectors of identifiers; each dimension corresponds to a separate term, such as words, keywords or phrases. In particular, the dimensionality of the vector is the number of distinct terms and each value indicates the associated term’s weight. We initially extract index keywords, also known as terms, from the gloss of each adjective, and then generate vectors where each value is the term frequency (TF), i.e., the number of times a term occurs in a gloss. The TF measurement is used to weigh the term. Common words (such as ‘the’) are removed because they tend to incorrectly emphasize style of the texts.

Given adjectives a_1 and a_2 , the corresponding index keywords are denoted by $G_1 = \{g_{11}, g_{12}, \dots, g_{1m}\}$ and $G_2 = \{g_{21}, g_{22}, \dots, g_{2n}\}$, respectively. $G = G_1 \cup G_2 = \{g_1, g_2, \dots, g_l\}$, i.e., terms occur in either G_1 or G_2 . The weight vector corresponding to G is $v_j = (v_{j1}, v_{j2}, \dots, v_{jl})$ ($j=1,2$) where v_{ij} ($1 \leq i \leq l$) is the TF of the i th term in G occurring in g_j . The semantic distance of a_1 and a_2 can be estimated by the cosine value of the angle between vectors v_1 and v_2 . A penalty factor is used to represent that semantic similarity among adjectives with ANT relationship is less than those with similar-to relationship. The similarity for adjectives is calculated by Eq. (6), where r is the semantic relationship of a_1 and a_2 , $\text{Cos}(v_1, v_2)$ indicates the cosine value of the angle between weight vectors v_1 and v_2 corresponding to G_1 and G_2 , and f is a penalty factor, the same meaning

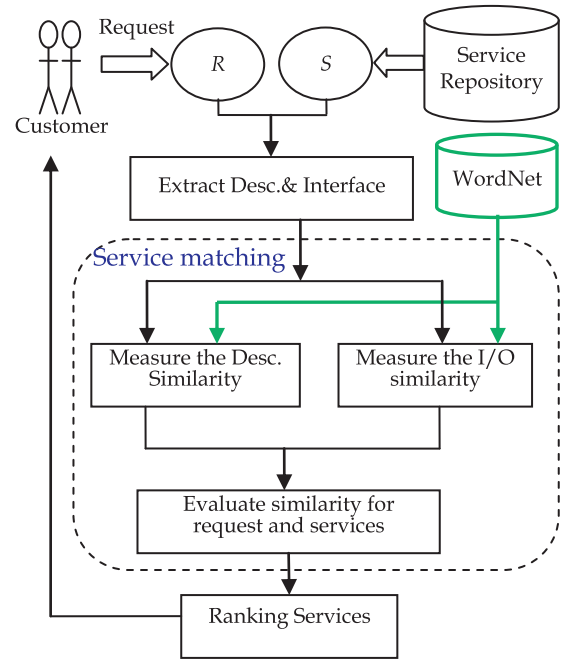


Fig. 3. Processes of Web services discovery approach.

as in Eq. (3).

$$\text{SimA}(a_1, a_2) = \begin{cases} \text{Cos}(v_1, v_2)/f + 1, & \text{if } r \text{ is an ANT relation} \\ \text{Cos}(v_1, v_2), & \text{otherwise} \end{cases} \quad (6)$$

The semantic similarity among adjectives with similar-to relationship scales to $[0, 1]$, whereas the semantic similarity between adjectives with ANT relationship scales to $[0, 1/(f+1)]$. The semantic similarity is 1 when two vectors are equal, i.e., the two adjectives belong to the same synset and share the same gloss.

Note that the proposed similarity measures of nouns and verbs as well as adjectives can be used for other domain ontologies with structures similar to WordNet’s.

4. Semantic similarity measure for web service discovery

This section describes the proposed SIMCR for requests and services based on the underlying proposed concept similarity measure in Section 3. Description similarity and interface similarity are evaluated with consideration of multiple conceptual relationships, and then integrated into an overall semantic similarity.

4.1. Overview of semantic web service discovery approach

As illustrated in Fig. 3, when a customer submits a request (R), available services in the repository are compared to the request using SIMCR. The results are properly ranked services.

4.2. Extracting term vectors of description and interface

This study uses integrated information, i.e., term vectors from the sections of ‘textDescription’, ‘Input’, and ‘Output’ in the service profile, respectively. Three parts of speech are used to abstract the textual description and only noun vector is used in the representation of the interface signature.

Generally the textDescription section in a service profile presents a specific description of functionalities of the service. It

usually consists of several sentences including nouns, verbs, adjectives, and so on. When evaluating the similarity between the request and services, most studies consider only the ISA relationship among nouns and verbs and omit adjectives. A flaw in that type of review is that adjectives may have provided additional useful information. For instance, there may be services named 'bookHighestPrice' and 'bookLowestPrice'. The textDescription sections in the profiles of the considered services are 'this service searches the highest price for a given book title' and 'this service searches the lowest price for a given book title', respectively. The distinct word in the sentences is an adjective describing the 'price', i.e., 'highest' and 'lowest'. In this case, overlooking adjectives would lead to completely opposite results. This study uses adjectives and adverbs in addition to nouns and verbs, to accurately evaluate the similarity between a request and a service. In the cases where adverbs exist in the textDescription part, we insert the corresponding adjective form of the adverb. For instance, the word 'automatically' can be transformed into 'automatic'.

We extract 'textDescription' section into several vectors of terms being compared by a service search engine. A natural language processing technique that includes part-of-speech tagging CLAWS POS tagger, (<http://ucrel.lancs.ac.uk/claws/>) and Porter's stemming algorithm (Porter, 2006) are applied to the abstraction. First, a CLAWS POS tagger is used to extract nouns vectors, verbs vectors, and adjectives vectors from the given textual document. The stem of each term, especially an inflected verb or adjective (such as 'ed', 'ly' or 'ing' forms), is obtained by Porter's stemming algorithm. Some high-frequency but useless terms, or so-called stop words such as 'service' and 'they', are removed. For instance, the textDescription section of a service 'bookmaxprice' is 'this service searches the maximum price for a given book title'. The corresponding noun vectors {service, price, book, title}, verb vectors {searches, given}, and adjective vectors {maximum} can be returned by CLAWS POS tagger. These vectors of terms are pruned to {price, book, title}, {search, give}, and {maximum}.

It is straightforward to get the term vectors associated with the input and the output. The sections of 'Input' and 'Output' in the profile of services enumerate the input parameters and outcomes. Generally, they are all nouns rather than verbs or adjectives.

For convenience, term vectors associated with input and output in interface signature, and nouns, verbs, and adjectives in textual description are denoted as I , O , CN , CV , and CA , respectively. Without generalization, we denote the abstract of a service or a request as a five-tuple representation, i.e., $S = \langle I, O, CN, CV, CA \rangle$.

4.3. Semantic similarity measurement for request and service

As mentioned in Section 4.2, the description of a service can be summarized into a few vectors of terms with various parts of speech. Thus, the estimation of the semantic similarity for a request and service is the similarity evaluation of term vectors with various sizes. This section presents how to estimate the similarity measure between term vectors, and then defines a semantic similarity measurement for request and service.

4.3.1. Estimate the semantic similarity for term vectors

We apply structure similarity matrix (SM) (Huang, Yin, & Hou, 2011), a widely used approach in clustering tasks, to compute the similarity between term vectors. Given term vectors $T_1 = \{t_{11}, t_{12}, \dots, t_{1m}\}$ and $T_2 = \{t_{21}, t_{22}, \dots, t_{2n}\}$, S is the similarity matrix associated to them (as shown in Eq. (7)), therein $s_{ij} \in S$ ($1 \leq i \leq m$ for m terms in T_1 , $1 \leq j \leq n$ for n terms in T_2) is the semantic similarity of $t_{1i} \in T_1$ and $t_{2j} \in T_2$. Note that s_{ij} can be estimated by Eqs. (5) and (6)

with respect to the parts of speech for terms, respectively.

$$S = [s_{ij}]_{m \times n} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{m1} & s_{m2} & \dots & s_{mn} \end{bmatrix} \quad (7)$$

Based on the SM approach, the semantic similarity of T_1 and T_2 is defined as the average of similarity of T_1 over T_2 and T_2 over T_1 , i.e., an average of the maximal values in rows and columns, as indicated in Eq. (8) where m and n are the sizes of T_1 and T_2 . Specifically, for each term t_{1i} in T_1 , the similarity between t_{1i} and T_2 is the maximal similarity of t_{1i} to each element included in T_2 as denoted by v_{1i} in Eq. (8), i.e., maximal values in i^{th} row. The similarity of T_1 over T_2 is the average of similarities for each term in T_1 to T_2 , i.e., the former part of Eq. (8); and vice versa for T_2 over T_1 , i.e., the later part of Eq. (8). The similarity of term vectors is the average of similarities for each set over the other.

$$\begin{aligned} \text{SimT}(T_1, T_2) &= \frac{1}{2} \left(\frac{1}{m} \sum_{i=1}^m v_{1i} + \frac{1}{n} \sum_{j=1}^n v_{2j} \right) \\ V_1 &= \{v_{1i} | v_{1i} = \max\{s_{ik}, 1 \leq k \leq n\}, 1 \leq i \leq m\}, \\ V_2 &= \{v_{2j} | v_{2j} = \max\{s_{kj}, 1 \leq k \leq m\}, 1 \leq j \leq n\} \end{aligned} \quad (8)$$

In order to explain why we choose the SM approach, an alternative bipartite graph approach (Ou et al., 2011; Plebani & Pemici, 2009) is presented. The bipartite graph represents elements in different sets as separate groups of vertices and each edge just connects vertices included in separate sets. Compared to the SM bi-directionally matching the considered vectors, the bipartite graph is an incomplete graph. Particularly, Plebani et al. (2009) developed a linear programming formulation for assignment in bipartite graphs (LPABG) in order that the maximum similarity between the terms in the two sets can be compared. Assume that T_1 is the smaller size set, i.e. $m \leq n$; the line programming formulation is defined in Eq. (9).

$$\begin{aligned} \text{SimT}_{bg}(T_1, T_2) &= \frac{1}{m} \cdot \max \sum_{i \in I} \sum_{j \in J} s_{ij} \cdot x_{ij} \\ \sum_{j \in J} x_{ij} &\leq 1, \quad \forall i \in I \\ \sum_{i \in I} x_{ij} &\leq 1, \quad \forall j \in J \\ x_{ij} &\in \{0, 1\}, \quad I = [1 \dots m], \quad J = [1 \dots n] \end{aligned} \quad (9)$$

For example, for the given $T_1 = \{t_{11}, t_{12}, t_{13}\}$ and $T_2 = \{t_{21}, t_{22}, t_{23}, t_{24}\}$, Fig. 4(a) shows a graphical representation of the corresponding bipartite graph. Each vertex on the left side corresponds to a term included in T_1 and one on right side in T_2 . The edge is weighted by the similarity of two terms separately contained in T_1 and T_2 .

The best matching caught by LPABG is $\{(t_{11}, t_{22}), (t_{12}, t_{21}), (t_{13}, t_{23})\}$. Thus the similarity of T_1 and T_2 computed by shown in Eq. (9) is 0.823. The similarity calculated by the SM approach shown in Eq. (8) is 0.84. It is a little bit larger than using the bipartite graph approach. Details are shown below.

$$\begin{aligned} \text{SimT}_{bg}(T_1, T_2) &= \frac{1}{3} \cdot (0.82 + 0.96 + 0.69) \approx 0.823 \\ \text{SimT}(T_1, T_2) &= \frac{1}{2} \left(\frac{0.82 + 0.96 + 0.94}{3} + \frac{0.96 + 0.82 + 0.69 + 0.62}{4} \right) \approx 0.840 \end{aligned}$$

LPABG approach ignored the impact of unmatched terms, e.g. the term t_{24} in T_2 . Thus, it may be incapable to distinguish services with small difference. Suppose that the term t_{24} is removed

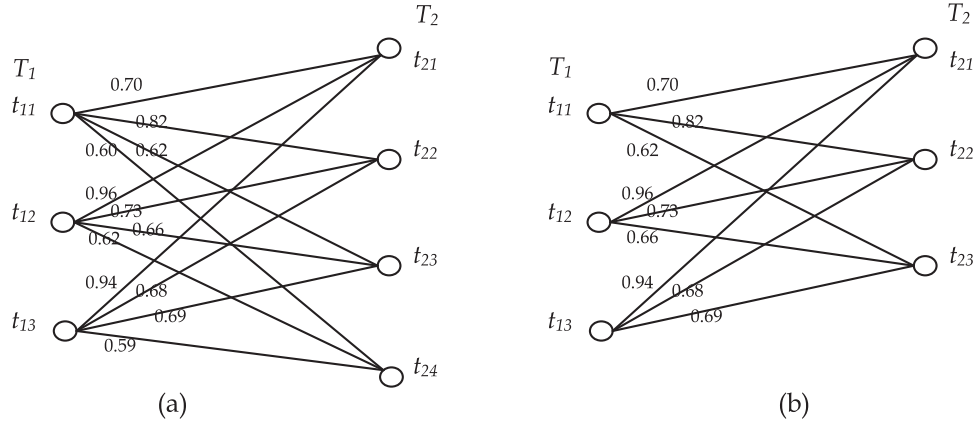


Fig. 4. Graphical representation of bipartite graph for two term sets.

from T_2 which may indicate another service, as shown in Fig. 4(b). As a result, the similarity of T_1 and T_2 would change somewhat. Precisely, the similarity should grow up as the similarity between t_{24} and terms in T_1 is smaller than others in T_1 . At that time, T_1 is completely matched with the truncated T_2 . But the LPABG reaches the same similarity as that reserving t_{24} in T_2 . Fortunately, the SM approach is capable to recognize such change, i.e. the similarity increases from 0.840 to 0.865, e.g.

$$\begin{aligned} \text{SimT}(T_1, T_2) &= \frac{1}{2} \left(\frac{0.82 + 0.96 + 0.94}{3} + \frac{0.96 + 0.82 + 0.69}{3} \right) \\ &= 0.865 \end{aligned}$$

In summary, the proposed SM approach with bidirectional matching outperforms the bipartite approach. It can be more responsive to the size changes of term sets. Moreover, it is a little efficient than LPABG since the linear programming for assignment in bipartite graphs costs more time especially for large sets.

4.3.2. Semantic similarity measure for request and service

To estimate the similarity for request and service, this section develops a SIMCR by considering both the textual description and the interface signature. The proposed SIMCR is defined by the weighted sum of description similarities and interface similarities.

As mentioned, the textual description includes nouns, verbs, and adjectives in this study. In fact, the impact of the terms varies with their part-of-speech (Oliva et al., 2011). The nouns make the greatest contribution to the Web services advertisement, followed by verbs, then adjectives. The description similarity can be calculated by a weighted average of the similarity of noun vectors, verb vectors, and adjective vectors, as defined by Eq. (10). The weights of noun, verb, and adjective vectors are denoted as n , v , and a .

$$\begin{aligned} \text{SimC}(R, S) &= n \cdot \text{SimT}(CN_R, CN_S) + v \cdot \text{SimT}(CV_R, CV_S) \\ &\quad + a \cdot \text{SimT}(CA_R, CA_S) \end{aligned} \quad (10)$$

The similarity of interfaces is the average of input similarity and output similarity, as defined by Eq. (11).

$$\text{SimIO}(R, S) = \frac{1}{2} (\text{SimT}(I_R, I_S) + \text{SimT}(O_R, O_S)) \quad (11)$$

The proposed SIMCR integrates the similarities among textual description and interface signature. It is the weighted sum of description similarity and interface similarity, as defined by Eq. (12), where α is an impact coefficient indicating the importance of description aspect:

$$\text{SIMCR}(R, S) = \alpha \text{SimC}(R, S) + (1 - \alpha) \text{SimIO}(R, S) \quad (12)$$

5. Evaluation and experiment studies

The proposed SIMCR is evaluated in this section. The experiments were carried out using a PC 2CPU (2.4 GHz) and 2 G of memory, the development platform was Matlab 2010a for Windows. Additionally, the latest version (3.0) of WordNet was used in this study as WordNet has been widely used in previous research within the field.

The effectiveness of the proposed similarity measure of concepts is examined first. Then the performance of the proposed SIMCR is demonstrated by comparing the results to other relevant research studies. For term similarity in SIMCR in those experiments, we have chosen to use weights for noun, verbal and adjective vectors as $n=0.40$, $v=0.35$, and $a=0.25$. For impact weight of description similarity, we chose to use $\alpha=0.6$. Finally, the impact of the textual description and interface signature on SIMCR (i.e., α) is investigated.

5.1. Evaluation of semantic similarity measure for terms

We compared the proposed semantic similarity measure for terms defined in Section 3 with human judgment and three additional concept similarity measurement methods: Li et al. (2003); Liu et al. (2010), and Bollegala et al. (2011). The three compared methods are denoted as Li's, Liu's and Bollegala's (shown in Table 1), respectively. Li et al. (2003) measured the concept similarity by information content in a lexical taxonomy, only the ISA relationship was examined. Liu et al. (2010) estimated the similarity of concepts by means of the weighted length of the shortest path between two concepts. They considered the ISA relationship and the HASA relationship in weighting the path but without the ANT relationship. Bollegala et al. (2011) developed a similarity measure by means of page counts and text snippets retrieved from a Web search engine. The similarity values of the three compared methods were taken from the corresponding literature.

We used the word pair set of Miller and Charles (Miller & Charles, 1991), referred to as the MC dataset, to evaluate the similarity measure of words. The MC dataset contains 30 word pairs with a human rating score. The scores scale from 0 to 4 for 'semantically unrelated' to 'highly synonymous'. Twenty-eight word pairs (Table 1), the same as Li's and Bollegala's, were selected as 2 pairs were not covered by WordNet. The human rated similarity scores are shown in the column of 'Human scores' in Table 1. Additionally, the columns two through the proposed in Table 1 indicate the semantic relationships covered by the shortest path between two words in each pair. A semantic relationship exists in the path labeled with a '✓' mark. Words in each of the last 3 pairs belong

Table 1
Similarities and comparisons for various methods on the MC dataset.

Word pairs	ISA	HASA	ANT	Human scores	Li's	Liu's	Bollegala's	SIMCR
Chord-smile	✓			0.13	0.01	0.00	0.01	0.21
Rooster-voyage	✓		✓	0.08	0.00	0.00	0.05	0.09
Noon-string	✓		✓	0.08	0.01	0.00	0.00	0.09
Glass-magician	✓			0.11	0.04	0.00	0.05	0.12
Monk-slave	✓		✓	0.55	0.26	0.12	0.24	0.11
Coast-forest	✓			0.42	0.10	0.00	0.15	0.14
Monk-oracle	✓			1.10	0.10	0.00	0.80	0.22
Lad-wizard	✓			0.42	0.26	0.10	0.23	0.56
Forest-graveyard	✓			0.84	0.04	0.00	0.44	0.22
Food-rooster	✓			0.89	0.00	0.00	0.02	0.22
Coast-hill	✓			0.87	0.19	0.00	0.36	0.24
Car-journey	✓		✓	1.16	0.00	0.00	0.17	0.18
Crane-implement	✓			1.68	0.25	0.07	0.06	0.62
Brother-lad	✓			1.66	0.26	0.10	0.13	0.65
Bird-crane	✓	✓		2.97	0.38	0.44	0.85	0.61
Bird-cock	✓			3.05	0.72	0.80	0.87	0.90
Food-fruit	✓	✓		3.08	0.02	0.00	0.94	0.52
Brother-monk	✓			2.82	0.72	0.80	0.27	0.88
Asylum-madhouse	✓			3.61	0.72	0.80	0.79	0.84
Furnace-stove	✓			3.11	0.05	0.00	0.88	0.44
Magician-wizard	✓			3.50	0.98	1.00	1.00	1.00
Journey-voyage	✓			3.84	0.72	0.80	1.00	0.90
Coast-shore	✓			3.70	0.59	0.75	0.97	0.84
Implement-tool	✓			2.95	0.68	0.77	0.50	0.89
Boy-lad	✓			3.76	0.71	0.79	0.96	0.88
Automobile-car				3.92	1.00	1.00	0.92	1.00
Midday-noon				3.42	0.99	1.00	0.99	1.00
Gem-jewel				3.84	0.96	1.00	0.82	1.00
PCC	–				0.891	0.850	0.870	0.903
SRCC	–				0.751	0.750	0.850	0.886

to the same concept. Thus, the corresponding cells in these lines are all blank.

Pearson's correlation coefficient (PCC) in Eq. (13) and Spearman's rank correlation coefficient (SRCC) in Eq. (14) (Bollegala et al., 2011) were used to evaluate the concordant degree between the human judgment (i.e. the MC scores) and the compared similarity measures. In Eq. (13), n is the number of instances, s_i^C denotes the similarity of each word-pair achieved by the compared approach and \bar{s}^C is the mean of s_i^C , s_i^H and \bar{s}^H and represents the same variables of the human judgment. In Eq. (13), r_i^C and r_i^H denote the rank of each instance qualified by the compared approach and the human judgment.

$$PCC = \frac{\sum_{i=1}^n (s_i^C - \bar{s}^C)(s_i^H - \bar{s}^H)}{\sqrt{\sum_{i=1}^n (s_i^C - \bar{s}^C)^2} \sqrt{\sum_{i=1}^n (s_i^H - \bar{s}^H)^2}} \quad (13)$$

$$SRCC = 1 - \frac{6 \sum_{i=1}^n (r_i^C - r_i^H)^2}{n(n^2 - 1)} \quad (14)$$

The PCC and the SRCC reflect the degree of the correlation between two variables. Here the PCC and the SRCC denote the agreement degree between each option and the human judgment. The PCC is used for the normalized similarities correlation of each approach and human institution. Considering approaches may vary with ranges of similarity, the SRCC is used to measure the rank correlation between each approach and human institution. The larger the rank correlation, the compared option gets closer to the human institution.

Table 1 shows the similarities calculated by 4 different methods, and the PCC/SRCC between the human scores and similarities yielded by each method (the last two lines in Table 1).

Table 1 indicates that both the PCC and the SRCC of the proposed similarity measure are superior to those of the compared approaches by a wide margin. Li's and Liu's SRCC values are much lower than SIMCR's although they have acceptable PCC values. The

SRCC is actually more significant than PCC, because it emphasizes the ranks while the PCC focus on ratings.

As discussed in Section 3.1, several factors are useful for the word's similarity measure including the shortest path length of two words, semantic relationship covered by the path, the depth of the subsumption and the local density. We argue that all of the information should be taken into account in order to achieve an accurate similarity measure.

The MC dataset covers three kinds of semantic relationships between concepts, ISA, HASA, and ANT. As shown in the columns two through four in Table 1, an ISA relationship exists in almost all paths associated with word pairs except those that belong to the same synset, e.g., 'automobile-car'. Some pairs also cover the HASA or ANT relationship, e.g., 'bird-crane' covers HASA relationship and 'monk-slave' covers ANT. Although Lin's study computes the concept similarity by means of weighted paths, an ISA relationship is taken into account while the HASA and ANT relationships are not addressed. We believe that the performance of our measure thrives on the consideration of distinct semantic relationships, especially the ANT relationship not considered by extant studies.

Note that adjectives are seldom considered in the concept similarity measurement. So far we are not aware of any existing datasets with an adjective similarity measurement. The consideration of adjectives is a novel direction that will improve the semantic similarity measurement in Web service matching results.

5.2. Dataset and experiment setup for the evaluations of SIMCR

5.2.1. Dataset

SIMCR is experimentally evaluated using the well-known service repository of OWLS-TC v3.0 test collection (http://semwebcentral.org/frs/?group_id=89). There are 1007 semantic Web services advertised with OWL-S 1.1 and 29 requests. There are seven service domains: Education, Medical, Food, Travel, Com-

Table 2
Details for OWLS-TC v3.0.

Domains	Num. of services	Num. of requests	Num. of matched service/request pairs
Economy	395	12	83
Communication	59	2	27
Medical care	73	1	5
Travel	197	6	30
Education	286	6	39
Weapon	40	1	16
Food	34	1	5

munication, Economy, and Weapon. Table 2 shows the details of OWLS-TC v3.0.

A discrepancy between the textual description and the interface signature of the request may disturb the performance of the service matchmaking. Thus requests whose textual description is close to the interface signature are selected in various domains. We use a binary relevance set to evaluate the performance of the proposed SIMCR. Each service matched with a candidate's request is manually remarked with a 'relevant' label. The number of remarked service/request (S/R) pairs is shown in Table 2.

5.2.2. Metric for evaluation

Precision, recall, and F-measure were used to evaluate the performance of the Web service discovery methods (Farrag et al., 2013). The precision value is a function of relevancy and is equal to the fraction of the retrieved services which are relevant, whereas the recall value is a function of retrieval results and is the fraction of the relevant results which have been retrieved, as shown in Eqs. (15) and (16). There is an inverse relationship between the equations, where it is possible to increase one value at the cost of reducing the other. Generally, they are not discussed in isolation. The F-measure integrates precision and recall into a single equation, i.e., composite harmonic mean, and mitigates the impact of large outliers and intensifies the impact of small ones, as shown in Eq. (17). As a weighted harmonic metric, the F-measure is much more meaningful than precision or recall individually.

$$\text{Precision} = \frac{\text{Relevant} \cap \text{Retrieved}}{\text{Retrieved}} \quad (15)$$

$$\text{Recall} = \frac{\text{Relevant} \cap \text{Retrieved}}{\text{Relevant}} \quad (16)$$

$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

5.2.3. Labeling relevant services according to SIMCR

We use a binary relevance set to compute the metrics of evaluation.

A parameter ε is introduced to represent the threshold of similarity for services to be considered as relevant to a request. In order to find a suitable ε value for each domain, we investigated the results of F-measure, precision, and recall with ε ranging from 0.60 to 0.90 with an interval of 0.05. Fig. 5 shows the F-measure, precision, and recall with various ε values. In these experiments, α is set to 0.6. Fig. 5(a) indicates that each domain has a different appropriate ε value that will be used in the later experiments.

5.3. Comparisons of the SIMCR to other methods

To validate the effectiveness of SIMCR, we compared SIMCR with various α values ($\alpha=0.6, 0, 1.0$) to several state-of-the-art Web service discovery methods including URBE (Plebani & Pemici, 2009), SimRSP (Ou et al., 2011), SDMA (Farrag et al., 2013), OWLS-MX (Klusch et al., 2009), SMARTSPACE (Dasgupta et al., 2014), and

Table 3
Comparisons of different methods.

Metrics	F-measure			
	Precision	Recall	Value	Ratio
SIMCR ($\alpha=0.6$, both)	0.844	0.896	0.869	1.000
SIMCR ($\alpha=1$, only Context)	0.680	0.706	0.693	1.254
SIMCR ($\alpha=0$, only I/O)	0.607	0.675	0.639	1.360
URBE (only I/O)	0.700	0.535	0.590	1.473
SimRSP (only I/O)	0.775	0.530	0.625	1.390
SDMA (only I/O)	—	—	0.837	1.038
UltiMatch-NL (both)	0.760	0.705	0.731	1.189
OWLS-MX (Logic-based)	0.740	0.557	0.636	1.366
SMARTSPACE (Logic-based)	0.805	0.900	0.850	1.022

UltiMatch-NL (Mohebbi, Ibrahim, & Zamani, 2014). Of these compared methods, four are similarity measurement methods: URBE, SimRSP, SDMA, and UltiMatch-NL. URBE, SimRSP, and SDMA use only interface signature. UltiMath-NL uses both interface signature and textual description. In addition, two logic-based semantic methods, OWLS-MX and SMARTSPACE, are compared against the proposed method.

All these methods used the OWLS-TC dataset but different versions. The OWLS-TC v2.2 data were used in OWLS-MX, SMARTSPACE and UltiMatch-NL tests. In addition to our method, the other methods used the OWLS-TC v3.0. These two datasets contain the same services and queries, but WSDL documents are included in the OWLS-TC v3.0.

Precision, recall, and F-measure were used to evaluate the performance of the service discovery methods. The measures of each of the compared methods were evaluated with reference to the corresponding literature. Note that Farrag et al. (2013) reported F-measure without measuring precision and recall. We investigated the SIMCR with $\alpha=0.6$ against the other methods. Two extreme cases, $\alpha=0$ and $\alpha=1.0$, are also examined, these cases indicate variations of the SIMCR, i.e., the results when either only interface signature or only textual description is considered. Furthermore, the ratio of the F-measure of the proposed SIMCR with $\alpha=0.6$ to the compared methods was investigated. The comparisons are shown in Table 3 and Fig. 6.

Fig. 6 indicates that the F-measure of the SIMCR with $\alpha=0.6$ outperforms all the other approaches: semantic similarity measurement and logic-based matches. The proposed SIMCR is better than the logic-based semantic approaches including OWLS-MX and SMARTSPACE. The F-measure is 33.6% higher than OWLS-MX's and 2% higher than SMARTSPACE's, respectively. Especially, the comparison in Table 3 demonstrates that SIMCR is superior to UltiMatch-NL in terms of precision, recall and F-measure, even though they all combine interface similarity and the description similarity. The F-measure of SIMCR is 18.9% higher than the UltiMatch-NL. However, the UltiMatch-NL focuses on weighting and combining the results of signature-based filter (i.e., interface matcher) and description-based filter to predict the relevance of requests and Web services rather than a semantic similarity among terms. In

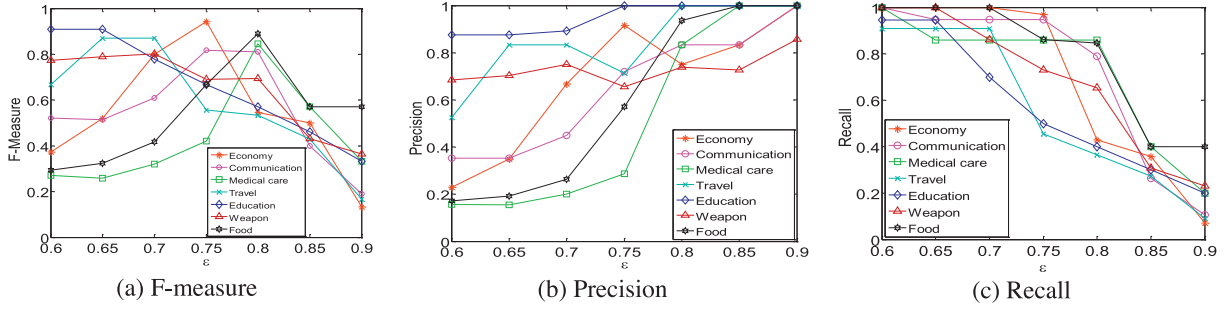
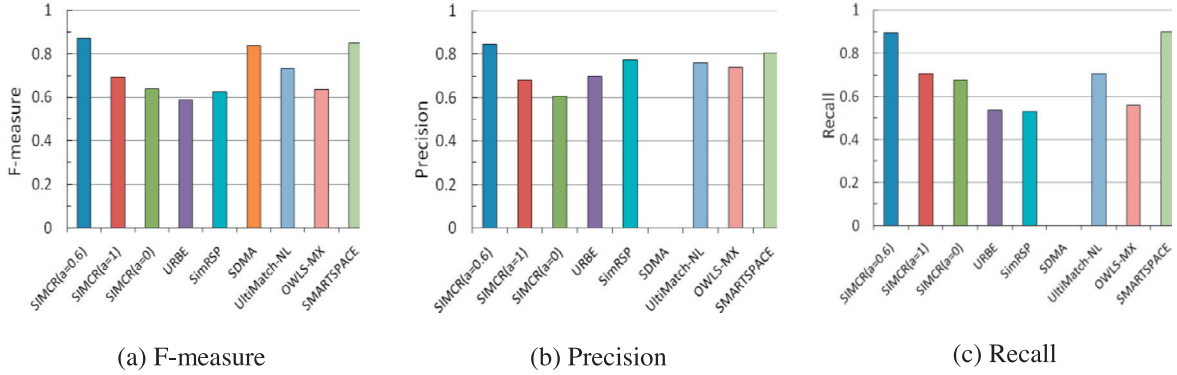
Fig. 5. Performance estimations with various ε .

Fig. 6. Comparisons of different methods.

contrast, the SIMCR aims a new semantic similarity measure between terms taking advantage of multiple conceptual relationships and an effective semantic similarity among requests and Web services. We argue that the outstanding performance of SIMCR can be attributed to the accuracy of the proposed semantic similarity measure of terms.

With respect to semantic similarity service matchmakers, we compared URBE, SimRSP, SDMA and UltiMatch-NL. They varied in semantic similarity measure of terms and textual similarity measure between term vectors associated to services/requests. In terms of the matching service methods, both URBE and SimRSP used a bipartite graph approach to measure the similarity among term vectors extracted from the interface section, and SDMA represented the matching degree of Web services as the least distance between term vectors. With respect to the semantic similarity of terms, URBE, SimRSP and SDMA just examined the ISA relationship between concepts. URBE measured the similarity by information content approach. Meanwhile SimRSP and SDMA used the shortest path approach.

In the specific case of SIMCR with $\alpha=0$, i.e., considering only the interface signature, the F-measure is 0.639, higher than that of URBE (0.590) and SimRSP (0.625) but lower than that of the SDMA (0.837). The reason is that there are some differences in the concept similarity measurement between the proposed SIMCR and the SDMA method. SDMA creates a new concept tree that supports the concept similarity measure and considers the multi-inheritance between concepts. Furthermore, the shortest path information between two concepts involves the number of up-steps and the number of down-steps. Compared to the SDMA, the proposed SIMCR uses common ontologies such as WordNet. We believe SIMCR is easier to apply in practice as some common ontologies have been well-defined.

Based on our findings, we believe that the performance of SIMCR benefits from both utilizing multiple conceptual relationships among concepts and the integration of interface signature and textual description.

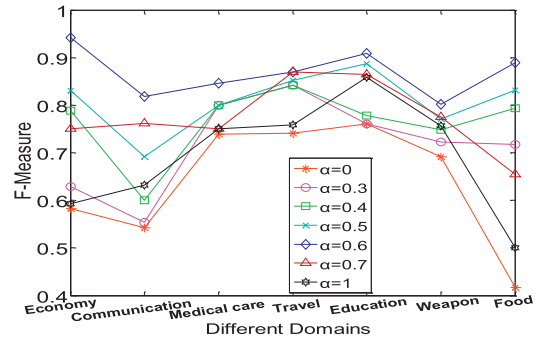


Fig. 7. F-measure for various domains.

5.4. Impact of description aspect in SIMCR

The proposed SIMCR is the weighted sum of description similarity and interface similarity, and coefficient α represents the importance of description aspect offering more specific information of functionalities than the general interface signature providing. This section investigates the influence of α on matching services with requests.

We conducted experiments with α ranging from 0.30 to 0.70 at intervals of 0.05. In addition, two extreme cases, $\alpha=0$ and $\alpha=1$, were tested which means that SIMCR is attributed to either interface aspect or description aspect. Note that the ε in each domain is set to the value that maximizes F-measure shown in Fig. 5(a). Table 4 shows the F-measure performance with various α values. Bold figures in Table 4 indicate the best F-measure among ones with various α in each domain. Fig. 7 shows the corresponding results of F-measure with certain α for various domains. For explanation in details, Fig. 8 illustrates the corresponding results of F-measure corresponding to Table 4, precision, and recall with various values of α for each domain.

Table 4
F-measure scores with various α .

α	0.00	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	1.00
Economy	0.583	0.629	0.788	0.788	0.831	0.831	0.831	0.942	0.800	0.750	0.593
Communication	0.543	0.554	0.562	0.600	0.654	0.692	0.706	0.818	0.818	0.762	0.632
Medical	0.739	0.800	0.800	0.800	0.800	0.800	0.800	0.846	0.785	0.750	0.750
Travel	0.741	0.842	0.842	0.842	0.842	0.852	0.852	0.872	0.872	0.872	0.759
Education	0.761	0.761	0.778	0.778	0.836	0.887	0.898	0.912	0.912	0.865	0.859
Weapon	0.692	0.723	0.748	0.748	0.771	0.771	0.802	0.802	0.775	0.775	0.757
Food	0.417	0.718	0.718	0.794	0.804	0.832	0.832	0.889	0.698	0.655	0.500
Avg.	0.639	0.718	0.748	0.764	0.791	0.809	0.817	0.869	0.809	0.776	0.693

Table 5
Details of services selected from economy domain.

Service name	Interface terms	Description terms	Similarity
Bookpersoncreditcardaccount	Credit, card, account, book, person	Book, purchase, credit, card, account, person, user, stock, cart, use, check, select, add, shop, available	0.588
EBookOrder3	e-book, request, user, account	e-book, order, request, title, user, account, cart, availability, stock, check, place, verify	0.616
Tizonbook_recommendedpriceindollar	Book, price, dollar	Company, book, provider, purchase, price, dollar, provide, recommend, rival	0.621
BookpersonOptional_price	Book, person, price	Book, title, person, price, user, history, return, depend, land, optional	0.654
Bookpersoncreditcardaccount_price	Person, book, credit, card account, price	Account, person, purchase, price, book, title, user, history, verify, current, different	0.695
Book_authorprice_Novel	Book, author, price	Author, name, purchase, price, book, short-story, novel, title, first, last	0.767
Book_pricereviewbook	Book, review, price	Book, review, price, purchase, delegate, recommend, suitable	0.794
Book_pricesizebook-type	Book, size, price	Book, size, price, type, affordable	0.803
Book_price	Book, price	Book, title, purchase, price, version, include, new	0.822
Author_bookprice	Author, book, price	Author, price, book, write, frequent, usable	0.921
Avg. of similarity			0.728

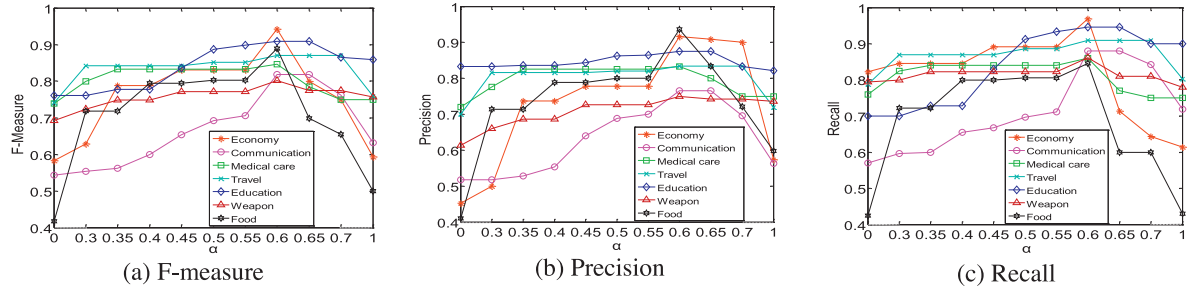


Fig. 8. Performance with various α .

In order to confirm that the increased F-measure owes to the incorporation of textual description rather than interface signature alone, we compared the F-measure values with various α as shown in Fig. 7.

Fig. 7 indicates that the F-measure values when $\alpha > 0$ increase on 7 domains compared to the lowest F-measure achieved with $\alpha = 0$. Furthermore, we did tests with $\alpha = 1$ which means the SIMCR is fully determined by the description similarity; and tests with $\alpha = 0$ which means the SIMCR is fully determined by the interface similarity. The fact that F-measures with $\alpha = 1$ are higher than ones with $\alpha = 0$ in all domains indicates that description offers more useful information than the interface. Therefore, we conclude that the description aspect may present more specific information than the simple interface signature.

As shown in Fig. 8, all the F-measure scores reach the best value in the case of $\alpha = 0.6$. As stated, both the interface aspect and description aspect contribute to the SIMCR. Furthermore, Fig. 8 indicates that ranges for F-measure vary in domains, as well as the precision and recall. The F-measure in three domains (i.e., 'Economy', 'Communication', and 'Food') fluctuates severely with changes in α , fluctuates gently in two domains (i.e., 'Education' and 'Travel'), and lightly in the other two domains (i.e., 'Weapon' and 'Medical'). As stated, SIMCR is the weighted sum of descrip-

tion similarity and interface similarity. The range of fluctuation is mainly attributed to the consistency between description aspect and interface aspect. In the domain where most services provide a consistent textual description with interface signature, the range is narrow; otherwise the range is much wider.

In order to confirm that the SIMCR is affected by the degree of consistency between textual description and interface signature, we looked closely into two domains, Economy and Weapon, and randomly retrieved 10 services from Economy domain and Weapon domain, respectively. Details including interface terms, description terms and the similarity of interface and description for the selected service are shown in Tables 5 and 6, respectively. The results demonstrate that the similarity of interface and description in the Economy domain is much lower than that in the Weapon domain. Therefore the F-measure of Economy fluctuates more widely than that of the Weapon domain due to the margin of interface signature and textual description.

In this section, we conducted a series of experiments to evaluate the proposed service discovery method. We compared the proposed semantic similarity measure for terms (i.e., the foundation of the SIMCR) with three popular similarity measures on the MC dataset. The result indicated that the proposed measure outperforms the others. Then we evaluated SIMCR for Web service

Table 6
Details of services selected from Weapon domain.

Service name	Interface terms	Description terms	Similarity
Missile_funding_Pak	Missile, funding	Missile, type, amount, government, Funding, spend	0.817
Governmentmissile_funding	Missile, government, funding	Missile, funding, type, government, possess, available	0.826
Government_funding_ABomb	Government, funding	Funding, government, atomic, bomb, special,	0.865
Ballisticmissilegovernment	Government, ballistic, missile,	Government, ballistic, missile, user, financing, range, target	0.879
_financingrange	financing, range		
Ballisticmissilegovernmentorganization	Government, organization,	Government, organization, user, ballistic missile, funding, target, range	0.901
_fundingrange	ballistic missile, funding,		
	range		
Governmentmissile_funding	Missile, government, funding	Missile, government, funding, possess, available	0.915
Weaponmissile_funding_Iraq	Weapon, missile, destruction,	Weapon, missile, government, destruction, funding, use	0.927
	government, funding		
Ballisticmissilegovernment	Government, ballistic, missile,	Government, user, ballistic, missile, target, lending, range,	0.934
_lendingrange	lending, range		
Missile_financing_China	Missile, financing	Missile, amount, financing, government,	0.948
Missilegovernment_giving	Missile, government, giving	Missile, technology, giving, aid, government	0.980
	Avg. of similarity		0.899

discovery by using the well-known service repository of OWLS-TC v3.0 test collection. Experiments show that SIMCR is superior to the compared state-of-the-art Web service discovery methods. The experiments demonstrated that the effectiveness of SIMCR could be attributed to the consideration of both multiple conceptual relationships among concepts and thorough use of information, integrating both interface signature and textual description. Experiment results indicate the SIMCR is affected by the degree of consistency between textual description and interface signature.

6. Conclusion and future work

We have proposed a novel semantic similarity measure, SIMCR, for semantic Web service discovery. The similarity measures takes into account different conceptual relationships in ontologies such as is-a, has-a and antonymy. The proposed similarity measure leads to a refined estimation of similarity among terms, web services and requests. Furthermore, in contrast to the state-of-the-art Web service discovery studies (Deng et al., 2008; Kapitsaki, 2012; Narock et al., 2014; Sangers et al., 2013; Wang et al., 2015) which failed to thoroughly extract information from a Web service description, the proposed method utilizes much richer information including both interface signature and textual description. Both the novel semantic similarity measure and the integration of overall Web service description contribute to effectiveness. Experiments show that the proposed SIMCR performs better than existing state-of-the-art methods in Web service discovery, in terms of precision, recall and F-measure. We believe that an effective Web service discovery approach will help customers build their distributed applications more efficiently. Furthermore, the proposed SIMCR can be used to facilitate classification or clustering on a semantic level, to help represent and apply knowledge in expert and intelligent systems.

However, the proposed method has some limitations. The semantic similarity of terms is measured on the basis of generic WordNet ontology. Disambiguation methods can be applied to improve the accuracy of similarity measure. The similarity measure can be improved upon specific application domains. The proposed method matches a specific request to each Web service in the service pool. Future work will address the design of scalable service discovery methods, such as cluster-based service searching.

Acknowledgments

The work was supported by the General Program of the National Science Foundation of China (No.71101103).

References

- Al Ridhawi, Y., & Karmouch, A. (2015). Decentralized plan-free semantic-based service composition in mobile networks. *IEEE Transactions on Services Computing*, 8(1), 17–31.
- Bener, A. B., Ozadali, V., & Ilhan, E. S. (2009). Semantic matchmaker with precondition and effect matching using SWRL. *Expert Systems with Applications*, 36(5), 9371–9377.
- Billhardt, H., Borrajo, D., & Maojo, V. (2002). A context vector model for information retrieval. *Journal of the American Society for Information Science and Technology*, 53(3), 236–249.
- Bollegala, D., Matsuo, Y., & Ishizuka, M. (2011). A web search engine-based approach to measure semantic similarity between words. *IEEE Transactions on Knowledge and Data Engineering*, 23(7), 977–990.
- Burstein, M., Bussler, C., Finin, T., Huhns, M. N., Paolucci, M., Sheth, A. P., et al. (2005). A semantic web services architecture. *IEEE Internet Computing*, 9(5), 72–81.
- Chen, Y. L., & Chiu, Y. T. (2012). Vector space model for patent documents with hierarchical class labels. *Journal of Information Science*, 38(3), 222–233.
- Cilibrasi, R. L., & Vitanyi, P. M. (2007). The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3), 370–383.
- Crasso, M., Zunino, A., & Campo, M. (2008). Easy web service discovery: a query-by-example approach. *Science of Computer Programming*, 71(2), 144–164.
- Crasso, M., Zunino, A., & Campo, M. (2011). A Survey of approaches to web service discovery in service-oriented architectures. *Journal of Database Management*, 22(1), 102–132.
- Dasgupta, S., Aroor, A., Shen, F., & Lee, Y. (2014). SMARTSPACE: multiagent based distributed platform for semantic service discovery. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(7), 805–821.
- Deng, S. G., Yin, J. W., Li, Y., Wu, J., & Wu, Z. H. (2008). A method of semantic web service discovery based on bipartite graph matching. *Chinese Journal of Computers*, 31(8), 1364–1375.
- Di Modica, G., Tomarchio, O., & Vita, L. (2011). A P2P based architecture for semantic web service discovery. *International Journal of Software Engineering and Knowledge Engineering*, 21(07), 1013–1035.
- Dong, H., & Hussain, F. K. (2014). Self-adaptive semantic focused crawler for mining services information discovery. *IEEE Transactions on Industrial Informatics*, 10(2), 1616–1626.
- Farrag, T. A., Saleh, A. I., & Ali, H. A. (2013). Semantic web services matchmaking: semantic distance-based approach. *Computers & Electrical Engineering*, 39, 497–511.
- García, J. M., Ruiz, D., & Ruiz-Cortés, A. (2012). Improving semantic web services discovery using SPARQL-based repository filtering. *Journal of Web Semantics*, 17, 12–24.
- Huang, C. H., Yin, J., & Hou, F. (2011). A text similarity measurement combining word semantic information with TF-IDF method. *Chinese Journal of Computers*, 34(5), 856–864.
- Joshi, K. P., Yesha, Y., & Finin, T. (2014). Automating cloud services life cycle through semantic technologies. *IEEE Transactions on Services Computing*, 7(1), 109–122.
- Kapitsaki, G. M. (2012). Web service matchmaking for the development of context-aware applications. *IET Software*, 6(6), 536–548.
- Klusich, M., Fries, B., & Sycara, K. (2009). OWLS-MX: a hybrid semantic web service matchmaker for OWL-S services. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(2), 121–133.
- Li, Y., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 871–882.
- Liu, W., Quan, X., Feng, M., & Qiu, B. (2010). A short text modeling method combining semantic and statistical information. *Information Sciences*, 180(20), 4031–4041.

- Meditskos, G., & Bassiliades, N. (2011). A combinatory framework of Web 2.0 mashup tools, OWL-S and UDDI. *Expert Systems with Applications*, 38(6), 6657–6668.
- Mezni, H., Chainbi, W., & Ghedira, K. (2014). Extending policy languages for expressing the self-adaptation of web services. *Journal of Universal Computer Science*, 20(8), 1130–1151.
- Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and cognitive processes*, 6(1), 1–28.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Mohebbi, K., Ibrahim, S., & Zamani, M. (2014). UltiMatch-NL: a web service match-maker based on multiple semantic filters. *Plos One*, 9(8), e104735.
- Narock, T., Yoon, V., & March, S. (2014). A provenance-based approach to semantic web service description and discovery. *Decision Support Systems*, 64, 90–99.
- Oliva, J., Serrano, J. I., del Castillo, M. D., & Iglesias, Á. (2011). SyMSS: a syntax-based measure for short-text semantic similarity. *Data & Knowledge Engineering*, 70(4), 390–405.
- Ou, W. J., Zeng, C., Xiang, X. M., Peng, Z. Y., & Li, D. Y. (2011). Efficient Web service query approach based on concept relaxation. *Chinese Journal of Computers*, 34(12), 2381–2390.
- Paulraj, D., Swamynathan, S., & Madhaiyan, M. (2011). Process model ontology-based matchmaking of semantic web services. *International Journal of Cooperative Information Systems*, 20(04), 357–370.
- Plebani, P., & Pernici, B. (2009). URBE: Web service retrieval based on similarity evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 21(11), 1629–1642.
- Porter, M. F. (2006). An algorithm for suffix stripping. *Program*, 40(3), 211–218.
- Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, 19(1), 17–30.
- Sangers, J., Frasincar, F., Hogenboom, F., & Chepegin, V. (2013). Semantic web service discovery using natural language processing techniques. *Expert Systems with Applications*, 40(11), 4660–4671.
- Shafiq, O., Alhaji, R., & Rokne, J. (2014). Log based business process engineering using fuzzy web service discovery. *Knowledge-Based Systems*, 60(2014):1–9.
- Wang, S., Zhu, X., Sun, Q., Zou, H., & Yang, F. (2013). Low-cost web service discovery based on distributed decision tree in P2P environments. *Wireless Personal Communications*, 73(4), 1477–1493.
- Wang, W., De, S., Cassar, G., & Moessner, K. (2015). An experimental study on geospatial indexing for sensor service discovery. *Expert Systems with Applications*, 42(7), 3528–3538.
- Wiemer-Hastings, P. (2004). All parts are not created equal: SIAM-LSA. In *Proceedings of 26th annual conference of the cognitive science society*.
- Wu, C. (2012). WSDL term tokenization methods for IR-style web services discovery. *Science of Computer Programming*, 77(3), 355–374.
- Wu, J., Chen, L., Zheng, Z., Lyu, M. R., & Wu, Z. (2014). Clustering web services to facilitate service discovery. *Knowledge and Information Systems*, 38(1), 207–229.
- Yang, S. J., Zhang, J., & Chen, I. Y. (2008). A JESS-enabled context elicitation system for providing context-aware web services. *Expert Systems with Applications*, 34(4), 2254–2266.
- Zapater, J. J. S., Escrivá, D. M. L., García, F. R. S., & Durá, J. J. M. (2015). Semantic web service discovery system for road traffic information services. *Expert Systems with Applications*, 42(8), 3833–3842.
- Zhao, J. L., Tanniru, M., & Zhang, L. J. (2007). Services computing as the foundation of enterprise agility: overview of recent advances and introduction to the special issue. *Information Systems Frontiers*, 9(1), 1–8.