Keywords: Cloud Services, Discovery, Semantic Web, Preconditions and Effects, Information Systems.

Abstract: Nowadays, most of companies deploy their services as a Web service format over the Internet. Additionally, cloud environments are dealing with web services published on the Internet. The use of the services deployed is subject to its discovery. Many techniques are being developed to increase the description of a web service, thus automating their discovery and providing more significant results. Such techniques are based on semantic concepts to provide more unambiguous information on describing such services. This paper proposes to discover cloud services based on the semantic Web, especially on preconditions and effects. Several works have been carried out on preconditions and effects, however we have achieved more accurate results using service descriptions based on conditions. We present our approach examining three main aspects: (i) relevant retrieved services (ii) complexity and (iii) execution time and we compare with other closely related work in an attempt to position our work and our results.

1 INTRODUCTION

A great number of business processes in information systems have been developed in a web service format because of their standard and loosely coupled features. Most of these web services can be published over a distributed environment, such as the Internet, an enterprise (as a local network) or into a cloud.

The proliferation of such services is enabling the creation of various types of web services. From a global view, we can classify two types of services: (a) a client-side service, developed in JavaScript, using technologies such as AJAX and HTML 5 and handled by customers in their respective browsers. We call these services a light-weight service (b) a server-side service, where connections are made between web servers and their development environments. This kind of service (b) can usually be used through a proxy and allows interoperability between heterogeneous applications and can be considered as a heavy-weight service. Both types of services are loosely coupled and are immerged in a distributed environment, such as a cloud (Zhang et al., 2010).

Some of the services on the Internet are incorporated into cloud environments as a SaaS (Software as a Service) (Höfer and Karagiannis, 2011). Sometimes, a single SaaS does not achieve a user request, thus it has to be composed. These services are combined on-the-fly inside their clouds as a CaaS (Composition as a Service)(Höfer and Karagiannis, 2011). Nowadays, GoogleAppEngine (Höfer and Karagiannis, 2011) is a platform (PaaS) that can be used for free of charge. Some other platforms need a valid credit card number, even if it is not charged. This was our main motivation for using this platform within an academic research project. However, GoogleAppEngine only supports heavy-weight services, that is a server-side approaches. Thus in this work, we carried out only heavy-weight services. As our SaaS is being treated as heavy-weight services, our experiments were carried out over a server-side service approach.

With the continuing increase in service availability, a major challenge has been the automated discovery for new services closer to user/machine requirements. Among other features that can be incorporated into an automation process, the addition of a semantic web is widely used. A semantic description of web services can minimize the ambiguities and consequently facilitate such automated discovery (Berners-Lee et al., 2001).

One way to incorporate semantics into web ser-
services is through the use of semantic description languages, such as SAWSDL (Kopecky et al., 2007), OWL-S (Burstein et al., 2004) and WSMO (Roman et al., 2005). Among these languages, OWL-S can be coupled to an OWL (Mcguinness and van Harmelen, 2004) domain ontology, which makes inferences through its classes, properties and axioms, thus reducing the gap existing between a machine/user request and the offered service.

As we are dealing with composite services (CaaS), it is important to analyze the semantic links between each service in a composition. In the literature, several approaches have focused on service discovery, especially on semantic links. Some approaches deal with inputs and outputs (Paolucci et al., 2002b; Amorim et al., 2011; Lécué, 2011). Others have demonstrated that it is necessary to perform an additional analysis on conditions in order to retrieve more useful services (Bener et al., 2009; Bellur and Vadodaria, 2008). In addition to these semantic-based approaches, several authors have incorporated a syntactic analysis complement of the semantic analysis, such as (Amorim et al., 2011; Paolucci et al., 2003; Klusch et al., 2006; Junior et al., 2009). Our previous work proposed a hybrid approach, combining a semantic analysis based on inputs and outputs and a syntactic analysis based on the class structure. This work aims to extend our previous approach (Amorim et al., 2011) to incorporate preconditions and effects on semantic link analysis.

This paper is structured as follows: section 2 presents our proposed approach; section 3 presents our experiments and results; section 4 compares our work with related ones; section 5 shows our tool; and, finally, section 6 discusses our conclusions and future directions.

2 PROPOSED APPROACH

The basic idea behind our approach is to match the set of IOPE (input, output, preconditions and effects) of a request (user/machine) to the IOPE of a service. In this work, we have concentrated our efforts in OWL-S profile to describe services and requests because of its relationship with a domain ontology and its ability to be classified into a heavy-weight service, i.e. a cloud service.

2.1 Problem Definition

Given a set of finite services \( S = \{s_1, \ldots, s_n\} \) and a machine/user request, the semantic matching between the request \( r \) and a service \( s_i \in S \) can be defined based on semantic similarities between its inputs, outputs, preconditions and effects. The semantic similarity is analyzed based on a matching function \( \text{match}(C_a, C_b) = d \) within two concepts \( (C_a, C_b) \) inside the same ontology or knowledge representation. The matching result is given by a similarity degree \( d \), where \( d \in D \) and \( D = \{\text{Exact}, \text{Plugin}, \text{Subsumes}, \text{Sibling}, \text{Fail}\} \). Ideas of this similarity degree were kept track of (Paolucci et al., 2002b; Samper et al., 2008).

Each service \( s_i \), where \( i = [1..n] \) has a set of inputs \( (I_i) \), outputs \( (O_i) \), preconditions \( (P_i) \) and effects \( (E_i) \). Each set of inputs \( (I_i) \) has atomic inputs \( y_j \), where \( y_j \in I_i \). Each atomic input \( y_j \) is matched through the function \( \text{match}(C_a, C_b) = d \) as a concept \( C_a \) inside an OWL ontology. The concept \( C_b \) corresponds to the user/machine request’s input value. For the output set the process is similar.

Concerning the preconditions and effects, some minor differences need to be analyzed. In SWRL (Horrocks et al., 2004), each precondition \( p_k \), where \( p_k \in P \) of a service \( s_i \) is composed of a predicate and attributes. SWRL rules extend OWL axioms to include some Horn rules, with an antecedent and a consequent. For instance, the SWRL rule \( \text{hasCard(User, Account)} \) means that if \( \text{User} \) and \( \text{Account} \) are true, then \( \text{hasCard} \) is true. In SWRL, the predicate is the consequent of a rule and the attributes are the antecedents, i.e: \( \text{hasCard} \) is a predicate in SWRL and \( \text{User} \) and \( \text{Account} \) are attributes in SWRL.

Our approach analyzes separately each predicate and attributes. Thus, our function \( \text{match}(C_a, C_b) = d \) firstly performs the comparison between a service predicate \( (p_j[predicate]) \) and a request predicate \( (q[predicate]) \) as \( C_a \) and \( C_b \) respectively. Thus, we can have a function similarity of predicates as follows:

\[
\text{simPredicate}(p_j, q) = \text{match}(p_j[predicate], q[predicate])
\]

As regards attribute matching, for each precondition \( (p_j) \) two attributes are analyzed. The semantic similarity between these attributes is the maximum degree of matching between the service attribute and request attribute. For instance, if an attribute has a \( \text{Subsumes} \) degree \( (d = \text{Subsumes}) \) and another semantic similarity analyzes another attribute and it returns \( d = \text{Exact} \), thus the result attribute value is those with an \( \text{Exact} \) degree of matching. Our semantic similarity function of the attributes can be described as follows:

\[
\text{simAttribute}(p_j, q) = \max(p_j[attribute] \times q[attribute])
\]

The function similarity precondition is given by the minimum degree of similarity between these func-
tions. For instance, if the function simPredicate returns \( d = \text{Subsumes} \) and the function simAttribute returns \( d = \text{Plugin} \) and \( d = \text{Sibling} \) respectively, thus the similarity function returns the minimum degree \( d = \text{Plugin} \). With this mechanism, we can ensure that a minimum similarity degree between our semantic approach occurs.

\[
\text{simPrec}(p_i, q) = \min(\text{simPredicate}(p_i, q) \times \text{simAttribute}(p_i, q))
\]

Finally, the overall semantic similarity function between a service and a request is given by the minimum degree given by the above mentioned function as follows: The function similarity analyzes the minimum degree between a service \( (s_i) \) and a request \( (q) \) as follows:

\[
\text{similarity}(s_i, q) = \min(\text{simIn} \times \text{simOut} \times \text{simPrec} \times \text{simEff})
\]

In order to illustrate, suppose a service precondition \( P_1 \) and a request precondition \( R_1 \) within the following structure predicate \( (\text{attribute}, \text{attribute}) \).

**Service:** \( P_1(A_1, A_2) \), where \( P_1 \) is the predicate and \( A_1 \) and \( A_2 \) are two attributes of a service.

**Request:** \( P_1(A_1, A_2) \), where \( P_1 \) is the predicate and \( A_1 \) and \( A_2 \) are two attributes of a request

Each function simPredicate and simAttribute is executed separately. There is no comparison between predicates and attributes, only within a service predicate and request predicate. The simPrec function is executed and returns a set of 3-tuple with a term of a service, a request and the degree of match \((G_i)\), as described below:

\[
\{(P_1, P_1, G1), (A_1, A_1, G2), (A_1, A_2, G3), (A_2, A_1, G4), (A_2, A_2, G5)\}
\]

Suppose \( G2 > G3 \) and \( G4 > G5 \), this means \( G2 \)’s similarity degree is higher than \( G3 \), i.e. \( G2 \) is Exact and \( G3 \) is Subsumes. Thus the precondition function returns \( (A_1, A_1, G2) \) and \( (A_2, A_1, G4) \), because of their major degree.

It is important to understand that an attribute could be used twice if it has a higher degree of similarity than other attributes.

\[
\forall P_i, \exists (P'_i, P'_i, G_i) \neq \text{Fail}, \text{ where } P'_i \text{ and } P'_i \text{ are the preconditions of a request and a service respectively.}
\]

Thus, we can ensure that for the execution of a service, it is important that at least one degree of similarity be different from \text{Fail}.

### 3 EXPERIMENTS AND RESULTS

Despite the rapid growth of semantic services, there has been little work in experimental tests and comparison among discovery methods. A public collection of currently available tests is called OWL-S Test Collection 2, but it does not include pre-conditions nor effects. Authors in (Bener et al., 2009) modified this collection by adding preconditions and effects described in SWRL (Horrocks et al., 2004). Despite the fact that they incorporate SWRL preconditions and effects into this dataset, the version used is not compatible with OWL-S API 1.2, thus no preconditions or effects can be actually retrieved. Thus, we enhanced this test set adapting it to the latest version of OWL-S 1.2 files that can be read automatically by an algorithm. This means that a machine can automatically interpret each condition described.

#### 3.1 Relevant Retrieved Services

In order to compare our approach with the algorithm proposed by (Bener et al., 2009) we developed two case studies:

**First:** Reuse of preconditions and effects.

**Second:** Predicate of service preconditions and the predicate of request preconditions are not exactly equal.

In the first test, suppose we have a service and a request with the following preconditions:

**Service:** \( \text{buyAutomobile}(\text{User}, \text{Model}) \) and \( \text{buyCar}(\text{User}, \text{Model}) \)

**Request:** \( \text{buyAutomobile}(\text{User}, \text{Model}) \) and \( \text{buyFood}(\text{User}, \text{Type}) \)

Suppose as well an ontology with \( \text{buyCar} \) and \( \text{buyBus} \) as descendant of \( \text{buyAutomobile} \).

The algorithm proposed by (Bener et al., 2009) performs the matching between the service precondition \( \text{buyAutomobile}(\text{User}, \text{Model}) \) and the given request. However, if it tries to find a request precondition that matches the service precondition \( \text{buyCar}(\text{User}, \text{Model}) \), such an algorithm returns \text{Fail} because the request precondition \( \text{buyFood}(\text{User}, \text{Type}) \) is not similar to \( \text{buyCar}(\text{User}, \text{Model}) \), unlike our approach that re-uses preconditions. Thus, our proposed algorithm matches the service precondition \( \text{buyAutomobile}(\text{User}, \text{Model}) \) with a similar request precondition \( \text{buyCar}(\text{User}, \text{Model}) \). Additionally, trying to
match a similar precondition \( \text{buyCar}(User, Model) \) our approach retrieves the same precondition once again \( \text{buyAutomobile}(User, Model) \), as we have similar preconditions \( \text{buyCar} \) and \( \text{buyAutomobile} \) as presented in the ontology. In this case our approach is able to retrieve a similar service precondition where the algorithm proposed by (Bener et al., 2009) does not retrieve the service; their approach fails.

The second experiment carries on the service precondition \( \text{buyCar}(User, Model) \) and the request precondition \( \text{buyBus}(User, Model) \). Suppose the existence of the predicate \( \text{buyAutomobile} \). The algorithm proposed by (Bener et al., 2009) defines the similarity degree between such preconditions as \( \text{Fail} \) because it does not analyze the semantic similarity between predicates through an ontology. Thus, they conclude that \( \text{buyCar}(User, Model) \) and \( \text{buyBus}(User, Model) \) are different, because these preconditions are not the same, thus the algorithm discards this useful service.

In our approach, we analyze each predicate and attributes, thus these predicates match as \( \text{Sibling} \) and can return the potential service.

The behavior of our approach is advantageous because it retrieves more useful services, avoiding the wrong elimination of services. In some cases, it is better to retrieve more and leave the decision to a final user, instead of wrongly eliminating a potentially interesting service. For instance, in criminal investigations suspicious, it is better to retrieve more suspects, instead of wrongly discarding a criminal.

### 3.2 Complexity Analysis

We analyzed the complexity of our algorithm within semantic filters on preconditions and effects. We considered that for each service precondition, our algorithm compares each request precondition, and for each request effect, our algorithm compares each service effect.

Based on (Cormen et al., 2001), we can conclude that, in the worst case, the complexity involved in our algorithm is \( O(n^2) \). Our algorithm analyzes the effects that work just like the pre-conditions, then its complexity is also \( O(n^2) \). On the other hand, our algorithm analyzes inputs and outputs and has complexity \( O(n^2) \) as stated by (Amorim et al., 2011).

The flow of the algorithm has the following order:

1. Input analysis \( O(n^2) \)
2. Preconditions analysis \( O(n^2) \)
3. Effects analysis \( O(n^2) \)

Thus, the total complexity of our algorithm is:

\[
O(n^2) + O(n^2) + O(n^2) + O(n^2) = 4(O(n^2))
\]

### 3.3 Execution Time

In order to evaluate the execution time of our algorithm, we carried out an experiment within 20 requests and a repository with 100 services adapted from the Test Collection 2 (Bener et al., 2009). All 20 requests were created within the same input and output sets so as to ensure that preconditions and effects were the only variable factors between requests. Thus, we can concentrate in a single factor to analyze, i.e. runtime.

We performed two experiments. The first experiment was to analyze inputs and outputs, calculating the average execution time. The second experiment dealt with inputs, outputs as well as preconditions and effects. With these two measures, it was possible to calculate the overhead between our new approach (with IOPE) and other works with only (IO).

Considering the requests \( r3, r4, r7 \) and \( r8 \) specifically on Table 1, we can observe that the relation between the execution time (in ms) values of IO and IOPE taking the requests \( r3 \) and \( r4 \) (i.e. 16.86 and 16.39 respectively) is considerably larger than the relation of execution time values between the requests \( r7 \) and \( r8 \) (i.e. 7.95 and 8.08 respectively).

Despite the fact that all four requests contain the same total number of preconditions and effects, the difference between the requests \( r3 \) and \( r4 \) is that they have preconditions but no effects. On the other hand, we can observe that requests \( r7 \) and \( r8 \) have an effect but no precondition. The difference between the computational time of \( r3 \), \( r4 \), \( r7 \) and \( r8 \) can be justified by the algorithm flow analysis of preconditions and effects. Suppose a service \( s1 \) has a precondition and an effect. To make the comparison between \( s1 \) and \( r3 \), the algorithm first examines whether \( r3 \) has preconditions. If there is a positive result, an analysis is performed on the preconditions between \( s1 \) and \( r3 \). Then the algorithm verifies that \( r3 \) has effects, obtaining a negative result. Thus, the algorithm stops its execution. The same applies to the request \( r4 \). Now suppose the comparison between \( r7 \) and \( s1 \). First the algorithm checks that \( r7 \) has preconditions. If the result is negative, the algorithm stops its execution immediately. The same happens with the request \( r8 \). Thus,
we can conclude that for requests $r_3$ and $r_4$, the algorithm performed the comparison on preconditions, but fails to compare the effects. In the case of requests $r_7$ and $r_8$, the algorithm does not compare the preconditions nor the effects. This explains why the runtime for requests that have pre-conditions is greater than the runtime for requests that do not have preconditions.

Figure 1 depicts the behavior of the runtime variation by the number of preconditions and effects.

Concerning our execution time, as we are using preconditions and effects from an OWL ontology, our algorithm does not have good results to scale up when there is a large number of preconditions and effects inside a service. We also analyzed empirically that in almost all tests sets experiments found in the literature, authors usually dealt with two preconditions. Within two preconditions and effects we can consider that our algorithm is working pretty well as other related approaches. However, we are working on an improvement to our algorithm to scale up better in the presence of more than two preconditions and effects.

4 ANALYSIS OF RELATED WORK

The wide use of service has motivated some researchers to develop discovery algorithms. This section presents some related work and makes a comparative analysis of these works and our approach.

The use of semantic filters were first introduced by (Paolucci et al., 2002b) with input and output parameters. This is a flexible approach based on the similarity degree, taking into account the distance between two nodes in a tree (ontology). This work proposed four degrees called exact, plugin, subsumes and fail.

Authors in (Amorim et al., 2011) have studied IO (input and output) on semantic filters. However, they proved that only semantic features were not sufficient to achieve good accuracy; thus they proposed an hybrid approach in two steps: (i) functional semantic based on semantic filters (Paolucci et al., 2002b) over their inputs and outputs parameters and (ii) structural semantic that analyzes each neighbor concept inside an ontology structure. The second step is only carried out in case of failure on the first step due to performance analysis. The major problem here is tackling only inputs and outputs parameters.

As regards some closely related work, SAM+ (Bener et al., 2009) is an algorithm that deals with preconditions and effects by using Paolucci filters (Paolucci et al., 2002b; Paolucci et al., 2002a). Firstly, their algorithm analyzes inputs and outputs and then their preconditions and effects described in SWRL (Horrocks et al., 2004). Their proposed approach uses three methods: subsumption that uses semantic filters and gives an intermediate weight; semantic distance that calculates the semantic distance between two concepts inside an ontology and the third method, wordnet, that uses the wordnet as a database of synonyms. Thus, a final weight is given for the overall matching.

Work proposed by (Bellur and Vadodaria, 2008) uses semantic filters (Paolucci et al., 2002b) to analyze the compatibility of web service parameters. They also divide their approach into three steps: parameter compatibility that uses the semantic filters within IOPE parameters; condition equivalence that proposes a structural analysis between two conditions; condition evaluation that evaluate if such conditions really satisfy each other.

Table 2 summarizes and compares previously related work with our proposed approach.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Paolucci</th>
<th>Amorim</th>
<th>Bener</th>
<th>Bellur</th>
<th>Our</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of OWL-S 1.1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Use of OWL-S 1.2</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Analysis of preconditions and effects</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Independency of the quantity of parameter</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Re-use of preconditions and effects</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of semantic predicate</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first column is the feature analyzed and the following columns depict the presence (mark as X) and absence (empty column) of each feature within each approach: (Paolucci et al., 2002b; Amorim et al., 2011; Bener et al., 2009; Bellur and Vadodaria, 2008) and our approach.
5 OUR TOOL IMPLEMENTATION

In order to spread and socialize our work, we developed a tool OWL-S Composer\(^1\). This tool was developed in Java language, with some additional features such as Jena 2.6.3 (JENA, 2000), pellet 2.2.2 (Clark-Parsia, 2010) and OWL-S API 3.1-SNAPSHOT (OsirisNext, 2010), which was the lastest available version of OWL-S API. These additional features were necessary to read and analyze OWL and OWL-S description of services.

In this project, Jena API was used for retrieving OWL elements. Jena API also includes the Pellet inference engine that was used to make semantic inferences concerning the axioms inside an ontology.

The developed system also makes use of OWL-S API 3.1-SNAPSHOT. This API transforms the elements of OWL-S document into Java objects and was used to obtain the set of inputs, outputs, preconditions and effects. This version of that API deals with OWL-S 1.2, its latest version.

6 CONCLUSIONS AND FUTURE WORK

The cloud environment is facing some new challenges for discovering SaaS (Software as a Service). In this work, we can retrieve more relevant services using preconditions and effects, thus minimizing the recovery of useless services and consequently the time configuring and finding services in a cloud. We also analyzed the execution time with preconditions and effects to ensure that our approach is feasible.

We have incorporated this solution into a OWL-S Composer plugin so as to discover in Google cloud environments.

As future work, we are working on the scalability of our solution and other PaaS, such as Amazon, SalesForce.

ACKNOWLEDGEMENTS

Some of the authors would like to acknowledge the Brazilian Government by CNPq (Grant 560231/2010-5).

REFERENCES


\(^{1}\)http://homes.dcc.ufba.br/dclaro/tools.html#owls3
service matchmaker combined with a uddi business registry. In *1st International Conference on Service Oriented Computing (ICSOC 2003)*, Trento, Italy.


