

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Relational Model for Parameter Description in Automatic Semantic Web Service Composition

Paul Diac^{a*}, Liana Țucăr^a, Andrei Netedu^a^a*Alexandru Ioan Cuza University of Iași
Faculty of Computer Science, Iași, Romania*

Abstract

Automatic Service Composition is a research direction aimed at facilitating the usage of atomic web services. Particularly, the goal is to build workflows of services that solve specific queries, which cannot be resolved by any single service from a known repository. Each of these services is described independently by their providers that can have no interaction with each other, therefore some common standards have been developed, such as WSDL, BPEL, OWL-S. Our proposal is to use such standards together with JSON-LD to model a next level of semantics, mainly based on binary relations between parameters of services. Services relate to a public ontology to describe their functionality. Binary relations can be specified between input and/or output parameters in service definition. The ontology includes some relations and inference rules that help to deduce new relations between parameters of services. To our knowledge, it is for the first time that parameters are matched not only based on their type, but on a more meaningful semantic context considering such type of relations. This enables the automation of a large part of the reasoning that a human person would do when manually building a composition. Moreover, the proposed model and the composition algorithm can work with multiple objects of the same type, a fundamental feature that was not possible before. We believe that the poor model expressiveness is what is keeping service composition from reaching large-scale application in practice.

© 2019 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of KES International.

Keywords: web service composition; semantic web; relational concepts; automatic composition algorithm; knowledge-based systems

1. Introduction

Semantic models used for Automatic Service Composition were developed some years ago, but their popularity slightly decreased more recently. We believe this is motivated by the lack of a model that is both expressive enough to allow the reasoning involved in manual composition, and simple enough to be easily adopted by service developers. One of the main shortcomings of previous models is that they do not define relationships over service parameters, which are intuitive for users creating manual compositions. Our contribution intends to take this next step.

E-mail address: paul.diac@info.uaic.ro

The proposed model is designed in the context of stateless, information-providing services; and in applications where services parameter information is well structured. Services are known ahead of the composition, so *discovery* is not needed and providers adhere to the same ontological model. Particularly, service parameters are defined over a known ontology's set of concepts enhanced with a set of *relations* specified in service definitions. Possible parameter *relations* are declared in the common reference ontology. Relations between inputs are restrictions or preconditions, and relations between output parameters are generated after calling the service; similar to postconditions. The ontology also includes a set of *inference rules*, conceptually similar with rules described in fundamental works like [1] or [2]. Rules are defined at ontology level only, and services cannot declare new rules. Rules are generally available and applicable anywhere as long as their *premises* hold. These rules also help to achieve the desired expressiveness and model non-trivial examples.

Our main contribution is adding relations to the model, with properties and inference rules. Therefore we extended the semantics in a natural manner increasing the expressiveness of classic models first described in [3] and used in many other works, with variations such as [4], [5] or [6]. Addressing the relational aspect of parameters was not tackled before in any work we know of, for example it is not mentioned in survey [7] on semantic services. There are a few works like [8] that use relational databases to resolve the composition queries, however, the "relations" are only at services level. A more recent formalism presented in [9] extends the semantic model with elaborate constructs, that seem hard to be implemented in practice, and yet does not introduce our type of parameter relations. Our paper also describes the implementation and synthetic evaluation of a proposed algorithm that searches a valid composition on this model. We motivate our approach with examples where classical composition fails to model human reasoning.

Section 2 presents such a motivating example. The problem definition constitutes the following Section, 3, that first describes our model textually and then formally states the problem requirements. In Section 4 we describe the standards we use to serialize services. The algorithm that computes the composition on the defined model is presented in Section 5 and performance is evaluated in Section 6. The last Section 7 concludes the paper and suggests some ideas for continuation.

2. Motivating Example

We analyze the following use case where **relational semantic** composition can model more complex interactions than simple or syntactic models based for example just on a hierarchy of concepts. More precisely, our addition of *relationships* and *inference rules* has the most important role, enabling the required expressiveness.

Assume a researcher is trying to schedule a meeting at some collaborating university. We are provided with the following information: the person's name, the name of the university where the researcher works, and the name of the university to visit. We consider the person and the two universities as instances of two concepts: **Person** and **University**. As we can already see, there is a need to distinguish between the two different instances of **University** and we are going to model that by the use of two different types of relations between our **Person** and each **University**: **isEmployeeOf** and **hasDestination**. Finally, we use a third relation: **isLocatedIn** and two inference rules that can help to expand relations. For example, if **X** is an employee of **Y** and **Y** is located at **Z** then **X** is located in **Z** (this may not be true in any situation, but it is a reasonable assumption that works for our example). In the composition model and in the further presented algorithm we can handle inference rules in the same manner as web services that "return" only new *relations* defined on already known "parameters", for simplicity. If the services cost, execution time, throughput or other Quality of Service is considered, rules are services with zero cost or instant runtime, unlimited throughput.

More precisely, we have the following web services and inference rules (with names ending in *Rule*). We specify in a service definition the required relations between input parameters and generated relations between output and possibly input parameters:

$$\begin{array}{ll}
 \text{getUniversityLocation} & \begin{array}{l} \text{input} = \{\text{univ} : \text{University}\} \\ \text{output} = \left\{ \begin{array}{l} \text{city} : \text{City}, \\ \text{isLocatedIn}(\text{univ}, \text{city}) \end{array} \right\} \end{array} \\
 \text{locatedAtWorkRule} & \begin{array}{l} \text{input} = \left\{ \begin{array}{l} X, Y, Z, \\ \text{isEmployeeOf}(X, Y), \\ \text{isLocatedIn}(Y, Z) \end{array} \right\} \\ \text{output} = \{\text{isLocatedIn}(X, Z)\} \end{array} \\
 \text{getAirplaneTicket} & \begin{array}{l} \text{input} = \left\{ \begin{array}{l} \text{pers} : \text{Person}, \text{source}, \text{dest} : \text{City}, \\ \text{isLocatedIn}(\text{pers}, \text{source}), \\ \text{hasDestination}(\text{person}, \text{dest}) \end{array} \right\} \\ \text{output} = \{\text{airplaneTicket} : \text{Ticket}\} \end{array} \\
 \text{destinationGenRule} & \begin{array}{l} \text{input} = \left\{ \begin{array}{l} X, Y, Z, \\ \text{hasDestination}(X, Y), \\ \text{isLocatedIn}(Y, Z) \end{array} \right\} \\ \text{output} = \{\text{hasDestination}(X, Z)\} \end{array}
 \end{array}$$

The solution for a composition is a list of services that can be called in the order from the list, and after all calls, the information required by the user is known. We also specify how to pass the output of a service to the input of another. This was trivial on previous models, but now as multiple instances of the same concept can be known, we need to distinguish between them, based on their *relations* with other objects.

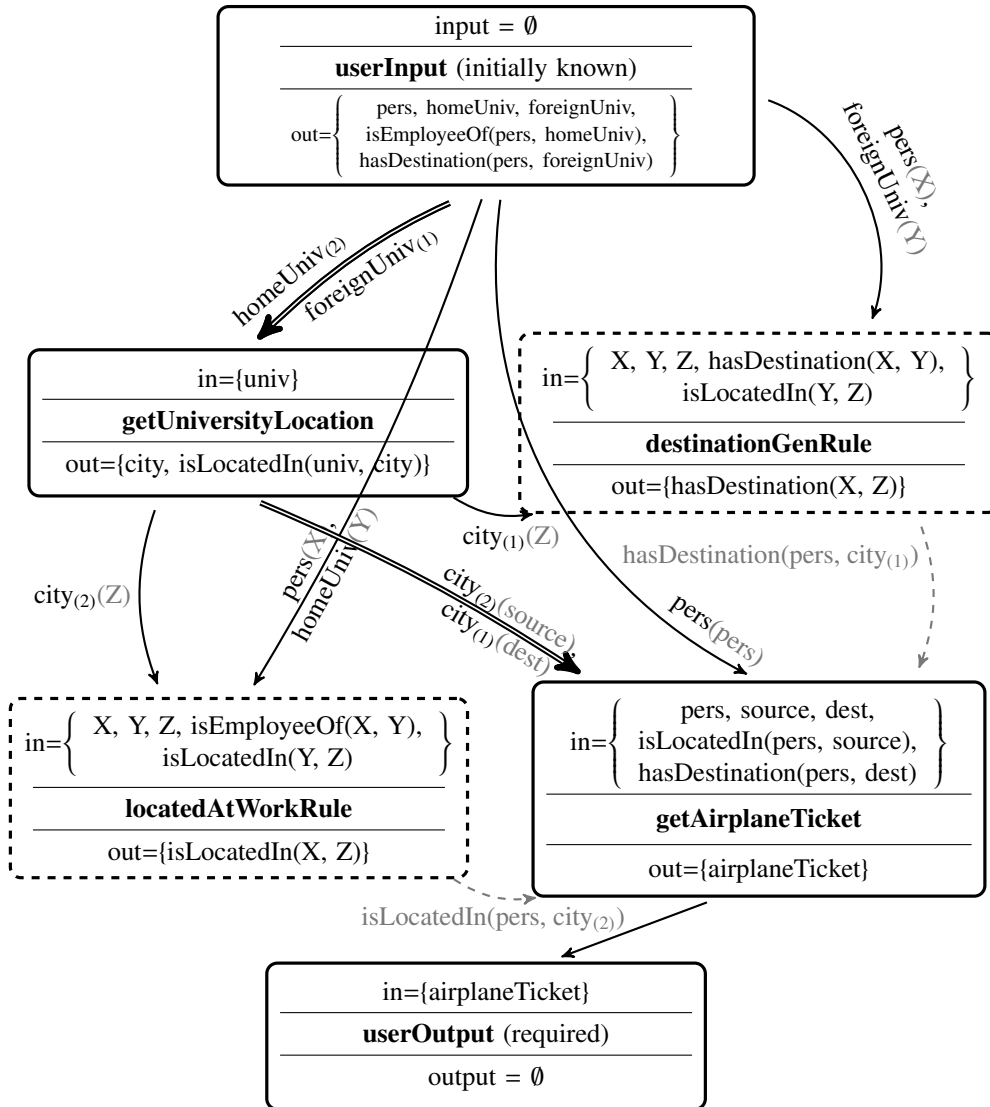


Fig. 1. Example motivating the need for relations and rules in composition language.
Service parameter types are not shown for simplicity, see Section 3 for types.

In Figure 1, solid boxes represent services and the user request (one for input - the initially known concepts and their relations, and one for the required output). Dashed boxes represent inference rules. Edges show information flow: solid edges - parameters and the dashed edges - relationships among them. Not all possible relations are used. Parameters are matched to rule variables (rule "parameters"), or other service parameters, based on the specification in gray in parenthesis. Multiple calls to the same service can be handled and are shown with double edges.

One composition representing a valid solution for the above instance would be the following list of service invocations, in order:

getInput(\emptyset) \Rightarrow pers, homeUniv, foreignUniv, isEmployeeOf(pers, homeUniv), hasDestination(pers, foreignUniv);
getUniversityLocation(homeUniv) \Rightarrow homeCity, isLocatedIn(homeUniv, homeCity);
getUniversityLocation(foreignUniv) \Rightarrow foreignCity, isLocatedIn(foreignUniv, foreignCity);

The two cities: *homeCity* and *foreignCity* are differentiated based on their relations, the names are not relevant for the composition (there is no restriction on what names they get if they would be automatically created; i.e. any distinct strings would work).

locatedAtWorkRule(pers, homeUniv, homeCity, isEmployeeOf(pers, homeUniv), isLocatedIn(homeUniv, homeCity)) \Rightarrow isLocatedIn(pers, homeCity);

destinationGenRule(pers, foreignUniv, foreignCity, hasDestination(pers, foreignUniv), isLocatedIn(foreignUniv, foreignCity)) \Rightarrow hasDestination(pers, foreignCity);

getAirplaneTicket(pers, homeCity, foreignCity, isLocatedIn(pers, homeCity), hasDestination(pers, foreignCity)) \Rightarrow airplaneTicket;

We can immediately notice the usefulness of semantic relations as the pairs *homeUniv* and *foreignUniv*, as well as *homeCity* and *foreignCity* are essentially indistinguishable between themselves otherwise: if knowledge would consist only of known types as before.

If we did not use semantic relations, to get the same desired functionality we would copy same services separating them for each possible instance of input parameters, for example **getUniversityLocation** could be split into two different services: **getForeignUniversityLocation** and **getHomeUniversityLocation**. We can imagine cases where this workaround raises the number of services provided at input by a high amount for each group of different input parameters.

In this example we did not use hierarchical concepts. However, *subsumption* can be used together with presented additions to the model of composition, as defined in the next Section 3. Overall, the simple hierarchy model proves to be easier to design and solve than our proposal.

3. Problem Model Formulation

3.1. Informal problem description

The first method of adding semantics to Web Service Composition was done by mapping parameters of services to a common *ontology* of concepts. More precisely, this was a simple *taxonomy* or tree of concepts, like in the first Semantic Composition Challenge [4]. Parameter matching, in this case, is based not just on name - string equality, as in the previous model; but on the *subsumes* relation. Intuitively, this means that one output of a service can be used as input for another service if the latter is a superconcept of the former. A subconcept can then replace a more generic concept, e.g., a species of a tree like *oak* can be used as an input of a service that requires any type of *plant*.

The proposed approach is increasing the expressiveness of the classic model by introducing *binary relations*, defined over instances of concepts. *Relations* are identified by a unique name that is a string. Within the ontology, there are also a set of relation *properties* and *inference rules*. The properties and rules are important because they enable the inference of new relations over parameter instances that are not from within a single service definition.

From one service definition, some relations between parameters are easily deduced by a human person because they are semantically implied by the name and description of the service and the name of parameters. A person would then naturally reason with the supposed relations when composing services. This must be considered by automatic composition as well.

3.2. Formal problem definition

Concept. A concept *c* identified by a *conceptName* is an element of the given set of concepts *C*. Concepts are arranged in a hierarchy, by the use of inheritance. Specialization concepts (or sub-types) can replace their more generic version (or super-types).

Object. An object is defined by a name *objectName* and a type *objectType* $\in C$. Intuitively, objects exist in the dynamic context of a structure of services that have been called. It is similar to an instance of a *Concept*, for which we know how it was produced and used in a workflow, or a series of service calls with their inputs matched by objects. Let *O* be the set of all objects known in the current composition.

Relation. A binary relation over the objects is identified by a unique *relationName* and is a subset of $O \times O$. Relations are not restricted on types.

Relation properties. A relation can have none, one or both of the properties: *symmetry* and *transitivity*. A symmetric relation *rel* it is not oriented from the first object to the second, i.e., if $rel(o1, o2)$ holds then $rel(o2, o1)$ holds as well. Transitivity implies that if $rel(o1, o2)$ and $rel(o2, o3)$ hold then $rel(o1, o3)$ holds.

Inference rules. An inference rule is composed of a *premise* and a *conclusion*. Both are a set of conjunction connected relations over variables of the set of objects *O*. Variables use a local name that has relevance only within the definition of this rule.

Inference rules can specify properties of relations, but we highlighted the frequently used *symmetry* and *transitivity* as properties, for simplicity. This also helps to optimize the composition search algorithm.

Ontology. An ontology is constituted by concepts *C*, a set of *relations* names with *properties* and *rules*. Only names and properties of *relations* are part of the ontology; the *objects* that are in a given *relation* depend on the context of the further defined composition.

Web Service. A Web Service is a triple (*input*, *output*, *relations*). The *input* and *output* are **disjoint** sets of **parameters**. One parameter is identified by a *parameterName* and has a type $\in C$. *relations* are restricted to local service parameters: $\subseteq (input \cup output) \times (input \cup output)$, i.e. a service defines relations only for its set of parameters. Similar to *rule* variable names, *service* parameter names have relevance only within the current service definition. The relations between input parameters of a service have to hold before the service is called, or, more precisely added to the composition and can be understood as *preconditions*. All the rest of the relations within a service definition hold after the service is called, they can be between objects created by this service call and are similar to *postconditions*.

The **user request** structure is similar to a web service. The *input* specifies what the user initially knows, that are objects with types and possible relations between them, and the *output* is the user's desired objects with required relations. Like in a service, there can be relations between inputs and outputs, specifying restrictions on what the user required outputs relative to his initially known object types.

Relation based Web Service Composition Problem. Given an ontology defining *concepts*, *relations* with *properties* and *inference rules*, a user request and repository of services, defined on the ontology; find an ordered list of services that are callable in that order, that solves the user request, starting with the initial information. For each service in the composition, the source of its input parameter must be clearly specified: resolved or bound to an output of a specified previously called service or user query. **Further clarifications** on the proposed model:

3.2.1. Services and Rules

Services and inference rules are structurally similar, with the distinction that: services must output at least some new parameter and cannot generate only new relations. Services do not "generate" relations between their input ($input \times input$) parameters. If their definition specifies relations between input parameters, they represent restrictions, i.e., conditioning the calling of the service. Rules however never "generate" any parameters but only new relations based on outputs of previously called services. Rules variables are not restricted by type i.e., they do not have types.

3.2.2. Parameters and Types

Each parameter of a service has a type, but unlike previous models, objects of the same type can be differentiated by their relations. This is fundamental in manual compositions, or other types of workflows, and we believe the impossibility to express this was the main flaw of previous models. This greatly increases the problem computational complexity, but as we will see in Section 5, our proposed algorithm is still able to find compositions on non-trivial instances of significant size, that a human user could not work with.

3.2.3. Objects

Objects are defined by their types, and all relations they are in. An object is similar to an instance of that type but is not an actual value or an instantiation. Objects are at the abstract level, they are dynamic elements that pass through the composition stages, and we keep information about their semantic context, through relations. On the current model, it is useless to have multiple objects of the same type with the same set of relations, so we keep only one for each possible context. Even with this reduction, the total possible number of objects is exponential compared to the number of relations (any subset of relations can define a different object). Moreover, when considering a relation of the current object, the type of the pair object on which the relation is defined, is also relevant to the current object state. This is motivated by parameter matching mechanism, without considering this the model (or algorithm) might fail to find valid compositions even if they would exist.

4. Standards

The information required to work with data on this model is serialized by a set of WSDL and JSON-LD files and an XML extension file. In order to add rules over parameters, the WSDL is extended with a type of field that defines relations. JSON-LD is used to describe types, and for inference rules an additional XML file is added. In short there are only three files: the semantic description of the concepts (types and relations), the inferences and the repository. The user query is serialized exactly as a service.

Ontology. Below we show the serialization of *Person*, as subtype of *Thing*. The JSON-LD file describes both types and relations. A relation has two fields that describe if it has the properties *transitivity* or *symmetry*.

Inference Rules are similar to a service since they have input and output variables, equivalent of parameters. Since it makes no sense to use WSDL, the input and output XML nodes are not necessary `<message>` nodes.

Repository: below is the *getAirplaneTicketService*. The WSDL description is extended in order to incorporate rules in the definition. The first message to appear in the service contains the input parameters for the service, while the next contains the output parameters. Lastly, a relation is described using a source, a target and name.

Ontology Format

```
["@graph": [
  { "@id": "Person",
    "@type": "rdfs:Class",
    "rdfs:subClassOf": {
      "@id": "Thing"
    }
  },
  { "@id": "IsLocatedIn",
    "@type": "rdfs:Class",
    "rdfs:subClassOf": {
      "@id": "Relation"
    },
    "isTransitive": true,
    "isSymetric": false
  }, ...
], ]
```

Inference Rules Format

```
<inference
  name = "locatedAtWorkRule">
  <input>
    <part name = "X"/>
    <part name = "Y"/>
    <part name = "Z"/>
    <relation name = "IsEmployeeOf"
      source = "X"
      target = "Z">
    <relation name = "IsLocatedIn"
      source = "Y"
      target = "Z">
  </input>
  <output>
    <relation name = "IsLocatedIn"
      source = "X"
      target = "Z">
  </output>
</inference>
```

Repository Format

```
<service name = "getAirplaneTicket">
  <message
    name = "getAirplaneTicketInput">
    <part name = "pers" type = "Person"/>
    <part name = "source" type = "City"/>
    <part name = "dest" type = "City"/>
  </message>
  <message
    name = "getAirplaneTicketOutput">
    <part name = "airplaneTicket" =
      "xsd:Reservation"/>
  </message>
  <relation source = "pers"
    target = "source"
    name = "IsLocatedIn"/>
  <relation source = "pers"
    target = "dest"
    name = "HasDestination"/>
</service>
```

5. Composition algorithm

The proposed algorithm is inspired from our previous work [10]. Previously, the focus was on the computational efficiency. The classic model from [4] was solved, with a simple hierarchy of types. To solve the composition on the new model, we prioritized less the efficiency, and more the inclusion of the extra functionality of the much more complicated semantic inference stage and parameter matching. Another choice was to avoid reducing the instance to any other known problem like *planning* or other solvers with reasoning but directly solve the instance by building the composition step by step and expanding information.

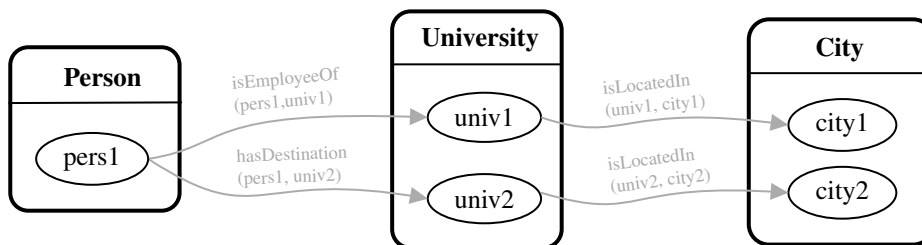


Fig. 2. **Knowledge:** rectangles show types with their instance objects in circles and gray arrows represent relations.

During the construction of the composition, a **knowledge base** is kept in memory, that is constituted of **objects**. An object has a known type, that is a concept from the ontology, and also a set of *relations* defined on the object. For example objects in Figure 2 represent a knowledge base. For each relation, the relation type and the pair object with

which the relation is defined are kept in memory. Based on this information, the next service is chosen together with a matching between known objects and service inputs. The matching must satisfy the defined relations between service input parameters. After this fictive "call" (conceptually, the call is done by adding the service to the composition), the knowledge base is updated with the output of the service. New objects can be created after a call, and for known objects, new relations can be added. After calling services, all inference rules are processed to see if they can be applied based on new objects or relations.

To check if in a given **knowledge** state some service can be validly called, a backtracking procedure is implemented. For each service input parameter in order, all *objects* of type or sub-types of the parameter type are iterated. The algorithm checks all the relations with objects that have already been matched with service parameters like shown from left to right in Figure 3. Branches that do not satisfy any relation between completed levels are dropped. If finally, the last parameter could be matched with some object, then the match is complete, and the service can be called. Also, to avoid repeated loops, a history of calls is kept for every service, that includes matched objects (by using a hash value computed over the objects). Using the history of calls, calling services with parameters that have been used before is avoided.

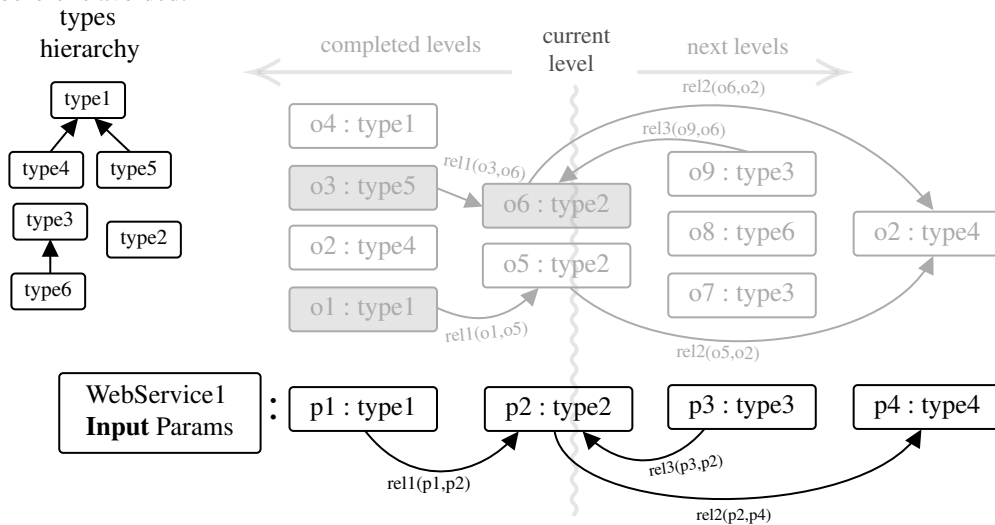


Fig. 3. Objects matching to input parameters, iterated from left to right. Parameter **p1** can match objects of **type1**, or subtypes **type4** or **type5**, but only **o1** and **o3** have the necessary relations, and similarly **o6** for **p2** on the current level.

For *inference rules*, the only distinction is that variables do not have types, so any level of backtracking can take any object. Rules with large premises and few relation restrictions would significantly slow down the search. This is another reason for disallowing provides to add rules.

The problem of matching input parameters to objects is **NP-Complete** on the defined model, both for services and for rules. This is because the problem is equivalent with *labeled subgraph isomorphism*, that is known to be **NP-Complete** as stated, for example in [11]. The equivalence is obvious: graph nodes are objects and directed labeled edges are relations with their types as labels. There are four main structures used in the algorithm: **ontology**, **repository**, **knowledge**, and **query**. First of all, the ontology is loaded, then the repository and the query that are verified to use the correct terms of the ontology. With the user query, the knowledge is instantiated with initially known objects and relations, and inference rules are applied for the first time.

The high-level view of the main Algorithm 1 is simple: iterate all services and verify for each if the service can be called with the current knowledge, and do so if possible. After looping all services, the inference rules are applied, similarly. If at any time the user required output is satisfied with all specified relations, the composition is complete, thus returned and the algorithm ends. If in some loop no new service or rule could be applied, then the algorithm is blocked in the current state; thus the instance is unsolvable. The use of *query.out* and *query.in* as input and, respectively, output parameters of services is an implementation shortcut that can be done based on the structural similarity: $(\emptyset, query.in)$ is a fictive service that has \emptyset as input and *query.in* as output. $CALLSERVICE(\emptyset, query.in, \emptyset)$ is just adding initially known objects to the knowledge, using the same method that adds service outputs later. The $APPLYINFERENCERULES()$ is very similar to the core part of the composition search, Algorithm 1. It iterates all rules and

applies any applicable rule until no rule can be applied for objects it was not already applied. Also, at the end the algorithm checks if there are any services generating only unused parameters and relations and removes them.

Algorithm 1 Main composition search loop

```

1: function SEARCHCOMPOSITON(query)                                ▷ // ontology and repository are global objects
2:   CALLSERVICE(( $\emptyset$ , query.in),  $\emptyset$ );
3:   APPLYINFERENCERULES( );
4:   newCall  $\leftarrow$  true;
5:   while (newCall  $\wedge$  FINDMATCH((query.out,  $\emptyset$ )) = null) do
6:     newCall  $\leftarrow$  false;
7:     for all service  $\in$  repository do
8:       matchObjects[ ]  $\leftarrow$  FINDMATCH(service);
9:       if (matchObjects  $\neq$  null) then
10:        newCall  $\leftarrow$  true;
11:        CALLSERVICE(service, matchObjects);
12:      APPLYINFERENCERULES( );
13:   if (FINDMATCH((query.out,  $\emptyset$ ))  $\neq$  null) then return called services in order;    ▷ // useless services are dropped
14:   else return null;

```

CALLSERVICE() - Algorithm 2 takes as parameters the *service* to call and the already found matching objects. Output objects are created with their name generated from the joined service name with the parameter name and also the number of times the service was called in total. This can help for debugging and getting provenance [12] information about objects later (though this is not yet implemented). After objects are created, all relations between input and output, or between output parameters, defined in the service are added to the matched or new objects according to the order known by their names.

Algorithm 2 Call service for known matched objects

```

1: function CALLSERVICE(service, matchObjects)
2:   hash  $\leftarrow$  HASHVALUE(matchObjects);
3:   service.history  $\leftarrow$  service.history  $\cup$  {hash};   newObjects[ ]  $\leftarrow$   $\emptyset$ ;
4:   for all parameter  $\in$  service.out do
5:     newObjects.add(new object of type parameter.type);
6:   for all relation  $\in$  (service.relations  $\setminus$  service.in2) do
7:     ADDRELATION(newObjects[x], newObjects[y]);      ▷ // for the objects matching parameters in relation
8:   knowledge  $\leftarrow$  knowledge  $\cup$  newObjects;

```

The most runtime consuming method is FINDMATCH() for services and its equivalent for rules that have exponential complexity. Searching for objects that match variables in rule premises is similar to services, and simpler to implement, as variables are not typed like parameters. The matching was defined on problem model, exemplified in Figure 3 and is implemented in Algorithm 3. For each level of the backtracking, all possible objects of the type are tried, and object relations are checked to match any parameter relation from the current level to any previous level. Both possible orientations of relations are checked. Reaching one level higher than the number of input parameters indicates a complete match, for which we also lookup in service history. If the match is also new, *matchFound* is set to true, and the search stops for all branches. There are a few intuitive special structures and methods used to get information faster. For example, *ontology.subTypes(...)* returns all subtypes for a type including itself and *knowledge.objectsOfTypes(...)* returns all objects known of specified types.

The algorithm presented is more a proof of concept implementation for the composition model. There are many possible enhancements and extra features that could be useful, some of which will be discussed in Section 7.

6. Evaluation

In order to evaluate the algorithm presented in Section 5, we implemented a test generator. The generator produces a problem instance: the repository, ontology and user query that has a high probability of being solvable. In the

Algorithm 3 Backtracking search for matching objects

```

1: function FINDMATCH(service)
2:   matchFound  $\leftarrow$  false;   matchObjects  $\leftarrow$  new Object[ ];
3:   BKTPARAM(0, service, matchObjects);
4:   if ( $\neg$  matchFound) then return null;
5:   else return matchObjects;
6: function BKTPARAM(level, service, matchObjects[ ])
7:   if (level = service.in.size) then
8:     hash  $\leftarrow$  HASHVALUE(matchObjects);
9:     matchFound  $\leftarrow$  (hash  $\notin$  service.history);
10:  else
11:    type  $\leftarrow$  service.in[level].type;   subTypes  $\leftarrow$  ontology.subTypes(type);
12:    candidates  $\leftarrow$  knowledge.objectsOfTypes(subTypes);
13:    for all candidate  $\in$  candidates do
14:      if ( $\neg$  matchFound  $\wedge$  RELATIONSMATCH(0...level, candidate, ...)) then
15:        matchObjects.add(candidate);
16:      BKTPARAM(level + 1, service, matchObjects);

```

first phase, it generates a set of services from which a composition can be extracted. To generate services with this property, we use this observation: as services are added in the composition, the knowledge is enhanced: more objects are obtained with their corresponding relations. So, to generate the repository, we start by incrementally generating the knowledge i.e., objects and relations, and saving all intermediate stages of this process. To generate the services we consider that every intermediate stage in the knowledge is the information gained during composition, thus, we need to create services that can generate these stages. Between each two stages K_i , K_{i+1} we generate a "layer" of services with each service having as input parameters objects and relations between them from K_i and output parameters in a similar way from K_{i+1} , as shown in Figure 4. The user query is generated as following: the input parameters and relations are a subset of the first stage, while the output parameters and relations are a subset of the last stage of the knowledge. "Noise" is added in the repository and ontology, i.e., random concepts and services with the aim of hiding the composition.

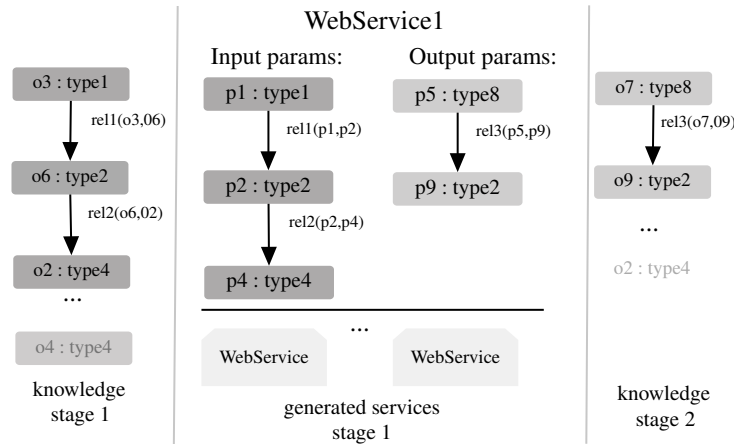


Fig. 4. Consecutive stages of knowledge and services generated between the stages.

The algorithm provided the expected composition on the example in Section 2. We evaluated its efficiency on tests generated as above, with and without inference rules (and also, ignoring them). Results in Table 1 show that the use of inference rules improves the composition, that becomes shorter on two tests; and that the algorithm is efficient. The slowest run is observed on a testcase where more rules are applied, which is expected.

To compare our solution with others, we also tested on the converted benchmark from the composition challenge [4]. These tests are obviously without any relations or rules, and these results are shown in Table 2. Even if our

Table 1. Results on **generated tests**. Headers: number of services or repository size, composition length, number of applied rules, running time, and composition length if rules are ignored.

repository	solution	number of rules applied	run time (seconds)	solution (ignoring rules)
63	11	0	0.07	13
30	14	74	0.30	15
30	8	3	0.04	13
46	4	0	0.02	4

Table 2. Results on **composition challenge tests** [4]. Headers: repository size; composition length and run time of Algorithm 1, composition length and run time of the winning solution of the composition challenge.

repository	solution	run time (seconds)	solution in [4]	run time in [4] (seconds)
1041	38	0.03	10	0.31
1090	62	0.04	20	0.25
2198	112	0.07	46	0.40

algorithm is designed for the extended model, it finds a composition of size comparable with the challenge winners (relative to the size of the repository) and in shorter amount of time. This also shows that the algorithm is compatible with the previous models.

7. Conclusion and Future Work

The essence of our contribution is the definition of the new composition model. Without the contextual-aware representation of elements involved in a composition at the parameter level, the composition is limited and cannot allow automation of the reasoning involved in manual composition. A simple hierarchy of types is insufficient. The implemented algorithm proves the problem is still solvable even on larger cases with many relations and rules.

There are many paths of continuation. The obvious is to mature the algorithm, for example by improving the strategy to reduce composition length or detection of useless services, improve service selection, prune useless paths of parameter matching or composition, by adding more conditions; or to consider more elaborate Quality of Service metrics. Also, it would help to develop ideas for aiding service providers to adhere to the model, including to convert definitions of existing services to the relational model. This process should be automated as well at least partially.

References

- [1] T. Berners-Lee, J. Hendler, O. Lassila et al., “The semantic web,” *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.
- [2] M. Sintek and S. Decker, “Triplea query, inference, and transformation language for the semantic web,” in *International Semantic Web Conference*. Springer, 2002, pp. 364–378.
- [3] S. A. McIlraith, T. C. Son, and H. Zeng, “Semantic web services,” *IEEE intelligent systems*, vol. 16, no. 2, pp. 46–53, 2001.
- [4] A. Bansal, M. B. Blake, S. Kona, S. Bleul, T. Weise, and M. C. Jaeger, “Wsc-08: continuing the web services challenge,” in *10th Conference on E-Commerce Technology and the Fifth Conference on Enterprise Computing, E-Commerce and E-Services*. IEEE, 2008, pp. 351–354.
- [5] T. Weise, S. Bleul, D. Comes, and K. Geihs, “Different approaches to semantic web service composition,” in *2008 Third International Conference on Internet and Web Applications and Services*. IEEE, 2008, pp. 90–96.
- [6] S. Bansal, A. Bansal, G. Gupta, and M. B. Blake, “Generalized semantic web service composition,” *Service Oriented Computing and Applications*, vol. 10, no. 2, pp. 111–133, 2016.
- [7] M. Klusch, P. Kapahnke, S. Schulte, F. Lecue, and A. Bernstein, “Semantic web service search: a brief survey,” *KI-Künstliche Intelligenz*, vol. 30, no. 2, pp. 139–147, 2016.
- [8] D. Lee, J. Kwon, S. Lee, S. Park, and B. Hong, “Scalable and efficient web services composition based on a relational database,” *Journal of Systems and Software*, vol. 84, no. 12, pp. 2139–2155, 2011.
- [9] W. Viriyasitavat, L. Da Xu, and Z. Bi, “The extension of semantic formalization of service workflow specification language,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 741–754, 2019.
- [10] L. Tucăr and P. Diac, “Semantic web service composition based on graph search,” *Procedia Computer Science*, vol. 126, pp. 116–125, 2018.
- [11] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, “A (sub) graph isomorphism algorithm for matching large graphs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 10, pp. 1367–1372, 2004.
- [12] W. Ding, J. Wang, and Y. Han, “Vipen: A model supporting knowledge provenance for exploratory service composition,” in *2010 IEEE International Conference on Services Computing*. IEEE, 2010, pp. 265–272.