AntiPattren-based Cloud Ontology Evaluation

Faiza Loukil
University of Sfax, Tunisia
Email: faiza.loukil2011@gmail.com

Molka Rekik
University of Sfax, Tunisia
Email: molka.rekik@gmail.com

Khouloud Boukadi
University of Sfax, Tunisia
Email: khouloud.boukadi@gmail.com

Abstract—Nowadays, cloud computing is an emerging technology thanks to its ability to provide on-demand computing services (hardware and software) with less description standardization effort. Multiple issues and challenges in discovering cloud services appear due to the lack of the cloud service description standardization. In fact, the existing cloud providers describe their similar offered services in different ways. Thus, various existing works aim at standardizing the representation of cloud computing services while proposing ontologies. However, since the existing proposals were not evaluated, they might be less adopted and considered. Indeed, the ontology evaluation has a direct impact on its understandability and reusability.

In this paper, we propose an evaluation approach to validate our proposed Cloud Service Ontology (CSO), to guarantee an adequate cloud service discovery. This paper contribution is threefold. First, it specifies a set of patterns and anti-patterns in order to evaluate CSO. Second, it defines an anti-pattern detection method based on SPARQL queries which provides a set of correction recommendations to help ontologists revise the ontology. Finally, some experiment tests were conducted in relation to: (i) the method efficiency and (ii) anti-pattern detection of design anomalies as well as taxonomic and domain errors within CSO.

I. INTRODUCTION

Over the last years, cloud computing has become an attractive strategy for the users thanks to its on-demand computing services provisioned and released with minimal management effort. Cloud computing offers its benefits through three types of services, namely, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). However, the cloud is hindered by the lack of a standard language for describing providers’ offers making it difficult for users to discover appropriate services. Indeed, cloud providers describe similar services in different manners (different pricing policies, different quality of services, etc.). Thus, various studies are proposed to standardize the representation of cloud computing services while introducing ontologies [1][2][3][4][5]. However, all the existing proposals were not evaluated, therefore they might be less adopted and considered by the users. Indeed, while developing the ontology, ontologists may commit errors, such as redundancy, missing information, inconsistency, etc. Moreover, while populating the ontology, by automatically extracting information from the providers’ catalogues and the cloud registries, some errors and redundant instances can be inserted. Consequently, it is necessary to evaluate the ontology concept to ensure the consistency maintenance and the redundancy elimination.

Usually, the ontology evaluation is an integral part of its development lifecycle. A set of tools known as reasoners have been proposed in the literature to get rid of errors in ontologies. According to Dentler et al. [6], "a reasoner is a program that provides automated support for reasoning tasks, such as classification, debugging and querying". Researchers in [7][8] identified new error types not covered by the traditional evaluation techniques. In fact, reasoners, such as FaCT++ [9], HermiT [10], Pellet [11], etc. can not deal with the inconsistency, the incompleteness and the redundancy errors. Evaluating such errors is a very difficult task [12].

In this paper, we define a novel ontology evaluation approach which considers the cloud domain to enhance the previously proposed CSO quality [5]. We denote that our approach is the building block for an adequate service discovery.

To do so, we take profit from the existing reasoners as well as cover their limitations by introducing a set of patterns and anti-patterns. In fact, the introduction of patterns represents a good solution to the recurrent design problems as they create and maintain ontologies [13]. Moreover, patterns help to create rigorous ontologies with the least effort. In opposition to the patterns, anti-patterns check the errors in order to correct them and consequently achieve the CSO’s high quality.

The rest of the paper is organized as follows. Section 2 overviews ontology evaluation techniques. Section 3 briefly describes CSO. Section 4 defines the ontology evaluation approach by demonstrating the limitations of the existing reasoners and presenting our anti-pattern detection method. Section 5 describes the evaluation of both the anti-pattern detection method and CSO and discusses the experimental results. Finally, we summarize the presented work the paper and present some future endeavors.

II. RELATED WORK

Several techniques for ontology evaluation exist in the literature among which we can cite the use of DL reasoners and anti-pattern techniques.

Gómez-Pérez et al. [7] proposed an evaluation framework of ontolingua ontologies. The authors identified a set of taxonomic errors that can occur within an ontology. Furthermore, they supposed that the presence of partition problems lead to an inconsistent ontology. Some researchers, such as [8][12], revised these errors and added new ones, for instance, disjointedness error, functional property omission for a single valued property and redundancy of disjoint relation error.

Fahad et al. [12], criticized the DL reasoner performance by evaluating their consistency in a case study of automobile ontology. They proved that the following reasoners, Racer [14],
Pellet [11] and FaCT++ [9], could not detect various types of errors, such as circulatory errors, semantic inconsistency errors and some types of redundancy errors like disjoint relation and identical formal definition.

That is why several techniques have been elaborated to detect these errors. In this field, the anti-patterns represent a good solution to evaluate ontologies [12]. In [15], Roussey and Zamazal identified an anti-pattern catalogue which is defined with DL expressions. In another work [16], the authors used the anti-patterns instead of the reasoners in order to evaluate an ontology. They proposed an anti-pattern detection approach based on SPARQL queries for the OIL anti-pattern.

In their work, Roussey and Zamazal defined a detection method based on ontology transformations to simulate the reasoner’s work [16].

In summary, reasoners can not detect the taxonomic errors as well as the domain errors. In addition, starting from scratch is not a good solution. To optimize the effort, it is better to reuse and benefit from the existing reasoners and cover their limitations by a set of anti-patterns.

In order to evaluate CSO [5], we combine, in this work, the two evaluation techniques, the reasoner and anti-patterns, to identify and correct the redundancy, the incompleteness and the domain knowledge missing of the ontology.

III. CLOUD SERVICE ONTOLOGY: CSO

CSO improves the interoperability problem among multiples Cloud infrastructures, platforms and applications within a Cloud federation environment [5]. The CSO building process was based on cloud standards, such as NIST [17], OCCI [3], CIMI [2], TOSCA [1]. For more details about CSO, we refer the reader to [5] and it is available on [4]. In this section, we present the important CSO’s concepts and axioms.

Among the main concepts in CSO, we cite: Actor, Service, Essential_Characteristic and CloudFederation, as shown in Figure 1.

The Actor class presents the different actors which participate in the cloud. In Figure 2, we distinguish between: the user, the provider and the broker. The Provider class includes three subclasses, Provider_IaaS, Provider_PaaS and Provider_SaaS.

Each cloud provider offers three different service types which are:
- **IaaS**: it includes the Storage and the Virtual_Machine resources.
- **PaaS**: it covers the Platform resource.
- **SaaS**: it encompasses the Software resource as CRM, mailing list, video gaming, etc.

Table I identifies the set of axioms which cover the specialization, the disjoint and the union relations between the important concepts within CSO.

Each Virtual_Machine (VM) has specific characteristics depending on its type. The object property hasVMType links between the two concepts Virtual_Machine and VMType, as shown in Table II.

The Provider_IaaS offers a wide range of virtual machine types, such as **small**, **medium** and **large**. Each VMType has an essential characteristic which is either on demand self-service (OnDemand_Service), which is a prime feature of most cloud offerings where the user pays for use, or on reserved service (Reserved_Service) where the user will get a discount after a service is used for a specific amount of time. Moreover, each VMType should have a compute capability which includes two characteristics: CPU and memory. We admit that the two characteristics could not exist without a compute capability. The object properties cpuIsPartOf and memoryIsPartOf are the inverse of the object properties ComputeCapability.ContainsCpu and ComputeCapability.ContainsMemory respectively, as shown in the two last lines in Table II.

The Figure 3 shows the CloudFederation concept. In a cloud federation, the services are provisioned by a group of cloud providers that collaborate together to share resources. The property interconnected represents the different members of a cloud federation. Each federation has an architecture hasFederationArchitecture, a type hasFederationType and a network hasNetwork, as shown in Table III. The federation architecture can be installed with broker, FederationArchitecture_Centralized, where there is a central broker that performs and facilitates the resource allocation or without broker, FederationArchitecture_Decentralized. The federation type can be vertical which spans multiple levels, horizontal which takes

<table>
<thead>
<tr>
<th>Axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broker ⊑ Actor; User ⊑ Actor; Provider ⊑ Actor;</td>
</tr>
<tr>
<td>User ⊑ Provider;</td>
</tr>
<tr>
<td>IaaS ⊑ Provider;</td>
</tr>
<tr>
<td>Provider ⊑ PaaS;</td>
</tr>
<tr>
<td>PaaS ⊑ Service;</td>
</tr>
<tr>
<td>Service ⊑ SaaS;</td>
</tr>
<tr>
<td>IaaS ⊑ SaaS;</td>
</tr>
<tr>
<td>Storage ⊑ IaaS;</td>
</tr>
<tr>
<td>Virtual_Machine ⊑ IaaS;</td>
</tr>
<tr>
<td>Platform ⊑ PaaS;</td>
</tr>
<tr>
<td>Software ⊑ SaaS;</td>
</tr>
<tr>
<td>OnDemand_Service ⊑ Essential_Characteristic;</td>
</tr>
<tr>
<td>Reserved_Service ⊑ Essential_Characteristic;</td>
</tr>
<tr>
<td>OnDemand_Service ⊑ Reserved_Service;</td>
</tr>
<tr>
<td>ComputeCapability.ContainsCpu ≡ cpuIsPartOf;</td>
</tr>
<tr>
<td>ComputeCapability.ContainsMemory ≡ memoryIsPartOf;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE I: CSO’s important axioms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axiom</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Broker ⊑ Actor; User ⊑ Actor; Provider ⊑ Actor;</td>
</tr>
<tr>
<td>User ⊑ Provider;</td>
</tr>
<tr>
<td>IaaS ⊑ Provider;</td>
</tr>
<tr>
<td>Provider ⊑ PaaS;</td>
</tr>
<tr>
<td>PaaS ⊑ Service;</td>
</tr>
<tr>
<td>Service ⊑ SaaS;</td>
</tr>
<tr>
<td>IaaS ⊑ SaaS;</td>
</tr>
<tr>
<td>Storage ⊑ IaaS;</td>
</tr>
<tr>
<td>Virtual_Machine ⊑ IaaS;</td>
</tr>
<tr>
<td>Platform ⊑ PaaS;</td>
</tr>
<tr>
<td>Software ⊑ SaaS;</td>
</tr>
<tr>
<td>OnDemand_Service ⊑ Essential_Characteristic;</td>
</tr>
<tr>
<td>Reserved_Service ⊑ Essential_Characteristic;</td>
</tr>
<tr>
<td>OnDemand_Service ⊑ Reserved_Service;</td>
</tr>
<tr>
<td>ComputeCapability.ContainsCpu ≡ cpuIsPartOf;</td>
</tr>
<tr>
<td>ComputeCapability.ContainsMemory ≡ memoryIsPartOf;</td>
</tr>
</tbody>
</table>

2. [https://www.dmtf.org/standards/cloud](https://www.dmtf.org/standards/cloud)
3. [http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.0/TOSCA-Simple-Profile-YAML-v1.0.html](http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.0/TOSCA-Simple-Profile-YAML-v1.0.html)
4. [https://sites.google.com/site/molkarekklaadhar/home/cso-owl](https://sites.google.com/site/molkarekklaadhar/home/cso-owl)
place on one level of the cloud stack, hybrid which covers both vertical and horizontal or delegation.

It is worth noting that tables I and II present the axioms which are useful in the evaluation process, particularly, in the anti-pattern detection phase.

IV. Ontology Evaluation Approach

The evaluation is an important phase in the ontology life-cycle. According to Gómez-Pérez, ontology evaluation "is a technical judgment of the content of the ontology" [7]. The main goal is to detect the errors within an ontology and then correct them in order to increase the ontology quality and consequently to guarantee its utilization. In this section, we present an overview of the CSO evaluation approach (see Figure 4).

Our approach is composed of two phases:

- **Evaluation process**
  In order to validate CSO’s consistency, we choose the DL reasoner FaCT++, developed at the University of Manchester and designed as a platform for experimenting through tableaux algorithms and optimization techniques. It is an open source, free and available with Protégé.
Besides, in order to validate the CSO cloud standard conformity, we define a set of patterns and anti-patterns. The detection of ontology anti-patterns contributes to ontology quality assessment \[16\]. This phase, as shown in Figure 4, includes three steps:

- **Consistency Verification**: we use the FaCT++ reasoner in order to validate the consistency of CSO. According to \[7\], "a given definition is consistent if and only if the individual definition is consistent and no contradictory knowledge can be inferred from other definitions and axioms". If ontology is inconsistent, then the correction phase should be applied and the reasoner’s application should be repeated. Otherwise, the next step, namely anti-pattern detection, will be triggered.

- **Anti-pattern detection**: this step consists in detecting the errors undiscovered by the FaCT++ by using a set of proposed anti-patterns (see Sect. IV-B). If no error is detected, ontology is considered as evaluated. Otherwise, the recommendations for the detected errors and anomalies will be triggered.

- **Recommendations**: in this step, we use the anti-pattern catalogue which associates correction recommendations for each anti-pattern. The proposed recommendations can be taken into consideration by ontologists, and in this case, the correction phase will be triggered. After that, we restart the evaluation process. If the recommendations are ignored by the ontologists, who consider that the errors are not relevant, the evaluation process is stopped.

**Ontology correction**

This manual phase aims at changing the ontology in order to correct the errors detected by either the reasoner or by following the recommendations proposed by the anti-patterns.

**A. Pattern Definition**

As defined in \[13\], the Ontology Design Pattern (ODP) is a modeling solution to a recurrent ontology design problem. In our work, the pattern definition is mainly based on taxonomy. In general, taxonomy hierarchically organizes classes and instances within an ontology \[7\]. In the cloud domain, taxonomy can be identified based on standards, such as NIST \[17\], OCCI \[1\], CIMI \[2\], TOSCA \[3\] which hierarchically present cloud service characteristics. We define the following decomposition pattern in order to guarantee cloud computing taxonomy.

1) The decomposition pattern: To respect the cloud standards, the hierarchy definition should meet one of the following decomposition manners.

- **Disjoint decomposition**: this pattern defines a set of disjoint subclasses of class C. This classification implies that each instance can belong directly to class C or to one of the subclasses of C, as shown in Figure 5. This pattern is expressed through DL as follows:

\[
C_1 \sqsubseteq C; C_2 \sqsubseteq C; C_3 \sqsubseteq C; C_1 \sqcap C_2 \sqcap C_3 \sqsubseteq \bot; \quad (1)
\]

- **Exhaustive decomposition**: this pattern defines a complete classification of subclasses of class C. These subclasses are not necessarily disjoint. Indeed, each instance of class C should be an instance of one or more of the subclasses as shown in Figure 6. This pattern is expressed through DL as follows:

\[
C_1 \sqsubseteq C; C_2 \sqsubseteq C; C_3 \sqsubseteq C; C_1 \sqcup C_2 \sqcup C_3 \equiv C; \quad (2)
\]

- **Partition**: this pattern defines a set of disjoint subclasses of class C. This classification is also complete and class C is the union of all the subclasses. Each instance of class C should be an instance of only one subclass, hence no shared instances are allowed. Figure 7 shows the partition pattern. This pattern is expressed through DL as follows:
\[ C_1 \sqsubseteq C; C_2 \sqsubseteq C; C_3 \sqsubseteq C; C_1 \cap C_2 \cap C_3 \sqsubseteq \perp; C_1 \sqcup C_2 \sqcup C_3 \equiv C; \] (3)

B. Anti-Pattern Definition

The anti-pattern is similar to a pattern, “except that instead of a solution, it gives something that looks superficially like a solution, but it is not” [13].

In our work, we propose three anti-pattern categories: taxonomic errors, design anomalies and domain errors. Each category includes a set of anti-patterns. The reason behind the proposal of taxonomic and domain errors is to cover, respectively, the errors related to the ontology concepts as well as to instances. Besides the errors, we propose the design anomaly category. In fact, anomalies point out to badly designed areas in ontology and their removal is necessary to improve ontology usability [12]. We present in what follows, the anti-pattern definitions using the DL expressions.

1) Taxonomic Errors: In this category, we aim at detecting errors specify:

- Inconsistency
  The inconsistency errors occur when there are external instances in a complete classification or when the disjoint relation is defined between classes with different hierarchies.

- External instances in exhaustive decomposition and partition:
  these errors occur when there are one or more instances of the class that do not belong to any subclass. This anti-pattern is expressed through DL as follows:

\[
C(\text{Individual}_1); C_1 \sqsubseteq C; C_2 \sqsubseteq C; C_3 \sqsubseteq C; \]

\[
C_1 \sqcup C_2 \sqcup C_3 \equiv C; \quad (4)
\]

- Partition Error: this error appears when a class is disjoint with the sibling of its super class instead of being disjoint with its sibling classes. This anti-pattern can be expressed through DL as follows:

\[
C_1 \sqsubseteq C; C_2 \sqsubseteq C; C_3 \sqsubseteq C_1; C_2 \cap C_3 \sqsubseteq \perp; \quad (5)
\]

- Incompleteness
  This kind of error occurs when the ontologists omit the definition of disjoint relation or the union relation. In fact, the information about the super class or the subclasses is missing.

- Decomposition knowledge Omission:
  this error covers two types of omission whenever identifying a decomposition of a concept: (i) the omission of which subclasses are disjoint or (ii) the omission, which it is the union of all its subclasses. This anti-pattern is expressed through DL as follows:

\[
C_1 \sqsubseteq C; C_2 \sqsubseteq C; \quad (6)
\]

- Redundancy
  A redundancy error appears when two object properties have the same formal definition or when the disjoint relation is duplicated. The redundancy error covers:

  - Identical formal definition of object properties: this error occurs when different object properties have the same formal definition, i.e. they have the same couple (domain, range), although they have different names. This anti-pattern is expressed through DL as follows:

\[
\exists R. \top \sqsubseteq C_1; \top \sqsubseteq \forall R.C_2;
\]

\[
\exists R_1. \top \sqsubseteq C_1; \top \sqsubseteq \forall R_1.C_2; \quad (7)
\]

- Redundancy of Disjoint Relation: it is the fact of defining a concept as disjoint with other concepts more than once, i.e. classes that have more than one disjoint relation.

The DL expression shows that the disjoint relation between subclasses C3 and C4 is repeated since they already inherit it from their base classes C1 and C2.

\[
C_3 \sqsubseteq C_1; C_4 \sqsubseteq C_2; C_1 \cap C_2 \sqsubseteq \perp; C_3 \cap C_4 \sqsubseteq \perp; \quad (8)
\]

2) Design Anomalies: Likewise, we identify the design anomalies within ontology. The correction of these anomalies guarantees the usability of an ontology [12]. Actually, these anomalies include:

- Lazy Concept: is an instantiated concept of which all the datatype properties are not defined. This anti-pattern is expressed through DL as follows:

\[
\neg \exists T.d \equiv \forall T. \neg d; \quad (9)
\]

- Weak Lazy Concept: is an instantiated concept of which the datatype properties are not defined. This anti-pattern can be expressed through DL as follows:

\[
\neg \forall T.d \equiv \exists T. \neg d; \quad (10)
\]

- Weak Concept Definition: this refers to the situation when an instantiated concept has all or some of its object properties not defined. This anti-pattern is expressed through DL as follows:

\[
\neg \forall R.C; \neg \exists R.C; \quad (11)
\]

3) Domain Errors: The previous cited anti-patterns are generic, which can be useful in different domains. However, our objective is to evaluate a cloud service description ontology. For this reason, the following set of anti-patterns focuses on cloud service description errors. It is worth noting that due to space limitations, we choose only the following set of anti-patterns and we can refer the reader to the [3] in order to find the complete list illustrated with examples.

- Invalid VMType Characteristic: each VM type must have at least an essential characteristic and a compute capability. Such anti-pattern occurs, as shown in the

\[ \text{https://sites.google.com/site/csoevaluation/home} \]
DL expression, when these two characteristics are not defined.

\[
\neg \exists \text{hasEssentialCharacteristic.}
\]

\[
\text{Essential Characteristic;}
\]

\[
\neg \exists \text{Compute Capability Attachment.}
\]

\[
\text{Compute Capability;}
\]

(12)

- **Invalid Provider Description**: each provider follows a deployment model, has a hosting type, a pricing model and a set of service capabilities. The omission of one or more of the information leads to an invalid provider description.

\[
\neg \exists \text{hasDeploymentModel. Deployment Model};
\]

\[
\neg \exists \text{hasHostingType. Hosting Type};
\]

\[
\neg \exists \text{hasPricingModel. Pricing Model};
\]

\[
\neg \exists \text{Provider Service Capabilities.}
\]

\[
\text{Provider Capability;}
\]

(13)

- **Invalid Service Capability Description**: each service must have a set of capabilities. For example, the storage and network capacities must be provided by the IaaS service. The following DL expression shows the capabilities that should be offered by each service type.

\[
\neg \exists \text{Platform Capabilities. Functional Capability};
\]

\[
\neg \exists \text{Software Capabilities. Functional Capability};
\]

\[
\neg \exists \text{Storage Capabilities. Functional Capability};
\]

\[
\neg \exists \text{Network Capabilities. Functional Capability};
\]

\[
\text{(14)}
\]

- **Invalid Compute Capability Definition**: each virtual machine type must have a compute capability. This capability includes two characteristics: CPU and memory. We admit that these two characteristics could not exist without a compute capability. The following DL expression shows the verification in both directions.

\[
\neg \exists \text{cpu Is Part Of. Compute Capability};
\]

\[
\neg \exists \text{Compute Capability ContainsCPU. CPU};
\]

\[
\neg \exists \text{memory Is Part Of. Compute Capability};
\]

\[
\neg \exists \text{Compute Capability ContainsMemory.}
\]

\[
\text{Memory;}
\]

(15)

- **Faulty Value**: the previous set of domain errors considers object properties. In the following part, we focus on the datatype properties. The values used in the two following anti-patterns are extracted from the three giant cloud provider catalogues, namely AWS Cloud\(^6\), Google Cloud\(^7\) and Microsoft Azure\(^8\).

- **Faulty Value_CPU_RAM**: each cloud provider offers a specific capability of a core number and memory size. This anti-pattern occurs, as shown in the DL expression, when one or more datatype properties violate the restrictions of each provider.

\[
\text{Provider IaaS(Microsoft_Azure);}
\]

\[
\exists \text{Cores. < 1}; \exists \text{Cores. > 32};
\]

\[
\exists \text{MemorySize. < 0.75}; \exists \text{MemorySize. > 448};
\]

\[
\text{Provider IaaS(Google_Cloud);}
\]

\[
\exists \text{Cores. < 1}; \exists \text{Cores. > 32};
\]

\[
\exists \text{MemorySize. < 0.6}; \exists \text{MemorySize. > 208};
\]

\[
\text{Provider IaaS(AWS);}
\]

\[
\exists \text{Cores. < 1}; \exists \text{Cores. > 40};
\]

\[
\exists \text{MemorySize. < 0.5}; \exists \text{MemorySize. > 244};
\]

(16)

C. Consistency Verification

In order to investigate how FaCT++ handles various errors, we load CSO in Protégé and introduce some errors, such as:

- we define an instance "IaaS_1" and link it directly to class IaaS instead of IaaS’s subclasses. This introduces an external instance in exhaustive decomposition and partition error,
- we add a disjoint relation between the class Broker and the subclass User_IaaS in order to make a partition error,
- we omit the union relation of the class ProviderCapability. Due to the absence of disjoint relation between the subclass Functional_Capability and the subclass Non_Functional_Capability, a decomposition knowledge omission error will occur,
- in order to introduce an identical formal definition object properties error, we define a new object property named hasArchitectureFederation that has the same couple (domain, range) of hasFederationArchitecture which is (FederationCloud, ArchitectureFederation) and
- we add a disjoint relation between the two Platform and the Software subclasses, yet, we have already one between PaaS and SaaS. This new disjoint relation introduces redundancy of disjoint relation.

Thereafter, we execute the FaCT++ reasoner which treats our definition of the instance "IaaS_1" to be a normal one, and gives no error or warning. Furthermore, it does not detect any errors from these, previously, introduced ones.

In summary, the FaCT++ reasoner is not able to detect the problems related to the taxonomic errors defined in Sect. [IV-B]

To the best of our knowledge, any existing reasoner is not able to detect the domain errors from these, previously, introduced ones.

D. Anti-Pattern Detection and Recommendations

According to the results obtained through the FaCT++ reasoner, it seems necessary to apply the anti-patterns defined in Sect. [IV-B] We propose an anti-pattern detection method based on SPARQL queries.

Due to space limitations, we present only the SPARQL query of the invalid VMType characteristic anti-pattern. A complete description of the queries is available in.\(^{[5]}\)
Invalid VMType Characteristic: the query in Listing 1 returns each type of virtual machine (?VMt) that is instantiated, but its definition is missing an essential characteristic (?E_C) and a compute capability (?compC).

Listing 1: SPARQL Query: Invalid VMType Characteristic
1: SELECT ?VMt ?E_C ?compC
2: WHERE
3: {?VMt rdf:type ns:VMType
4: OPTIONAL
5: {?VMt ns:hasEssentialCharacteristic ?E_C}
6: OPTIONAL
7: {?VMt ns:Compute_Capability_Attachment ?compC}
8: FILTER ( ! bound(?E_C) || ! bound(?compC))
9: }

1) The anti-pattern detection method: The proposed method in Listing 2 which is responsible for performing the anti-pattern detection, takes as input CSO and the list of SPARQL queries. As an output, it shows the detected errors and guides the ontologists to correct these errors by providing a set of recommendations.

During execution, the following cycle happens: (i) select a query from the list; (ii) execute the query and (iii) once the anti-pattern is detected, the error appears with a list of correction recommendations.

Listing 2: Anti-pattern detection method
1: Input: CSO.owl, SPARQL_query_antipattern_list
2: Output: Detected Errors and Recommendations
3: for each anti-pattern
4: Begin
5: for each query in
6: SPARQL_query_antipattern_list do
7: {
8: result=Execute(query)
9: if( Exist(result) ) then
10: Display("Recommendation_Text!")
11: }
12: End

To verify the anti-pattern detection method efficiency, we implement it and we test its performance on a test data (i.e. the CSO modified by the taxonomic error introduction). The method is implemented using the apache jena9.

V. EXPERIMENTATION: METHOD EVALUATION AND DETECTING ANTI-PATTERNS IN CSO

The CSO modified with several taxonomic errors, which are presented in Sect. IV-C, is taken as a test data. First, we run the method on this ontology to verify its efficiency (Sect. V-A). Then, we test the method on the real CSO (Sect. V-B).

A. Method evaluation

We qualify the method as efficient when it successfully detects the introduced errors. Figure 8 shows that the proposed method detects the taxonomic errors as well as the design anomalies and the domain errors within the modified CSO.

Table III presents the three anti-pattern categories used for our evaluation. For each category, this table indicates the number of the detected errors by the anti-pattern detection method. The precision measures the ratio of correctly found anti-patterns over the total number of detected anti-patterns. The last column indicates the recall which infers the ratio of correctly found anti-patterns over the total number of proposed anti-patterns.

Results from Table III are encouraging. In fact, the anti-pattern detection method covers the limitations of FaCT++ reasoner. Indeed, the set of SPARQL queries is sufficient for detecting the three anti-pattern categories, such as, taxonomic errors, design anomalies and domain errors.

B. CSO’s evaluation

After the efficiency verification, the method was applied on CSO to detect the existing errors. We notice that the taxonomic errors are not detected in CSO which is mainly based on the patterns (see Sect. IV-A).

Besides the design anomalies, the anti-pattern detection method found some domain errors, as shown in Table IV. These results prove that the classification of CSO concepts is well-defined since the absence of taxonomic errors. However, the presence of the design anomalies and the domain errors can be explained by several reasons. The main reason is related to the ontology population process. In fact, CSO population is based on the providers’ catalogues and the cloud registries. Due to standard unconformity, these data resources can include some errors on their textual content. Moreover, they do not provide a complete cloud service description. This is the reason why concepts, such as VMType and Service_Capability, have some missing information within CSO.

9https://jena.apache.org/
TABLE IV: Errors detected within CSO per category

<table>
<thead>
<tr>
<th>Errors</th>
<th>Number of Detected Errors</th>
<th>Percentage of CSO’s Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomic Errors</td>
<td>0/5</td>
<td>0%</td>
</tr>
<tr>
<td>Design Anomalies</td>
<td>3/3</td>
<td>100%</td>
</tr>
<tr>
<td>Domain Errors</td>
<td>2/8</td>
<td>25%</td>
</tr>
<tr>
<td>Total</td>
<td>5/16</td>
<td>31.25%</td>
</tr>
</tbody>
</table>

Anti-Pattern Definition: Invalid VMType Characteristic

The virtual machine type is invalid: its essential characteristic and its compute capability are omitted.

Recommendation

Please complete the description of the VM type class by defining these two object properties: hasEssentialCharacteristic and ComputeCapabilityAttachment.

Fig. 9: Detected Invalid VMType Characteristic Error

Figure 9 shows the anti-pattern detected errors, namely Invalid VMType Characteristic, and its correction recommendations. A complete demonstration of the CSO evaluation can be found in [10].

VI. CONCLUSION AND FUTURE WORK

Ontology evaluation is an active research domain which is mainly necessary to improve the ontology quality. In fact, only a well defined ontology is utilized. In this paper, we proposed an ontology evaluation approach composed of two phases: evaluation process and ontology correction. We applied the FaCT++ reasoner to validate the consistency of the previously proposed CSO. After that, we applied the anti-pattern detection method based on SPARQL queries in order to detect errors and anomalies not covered by this reasoner. To help ontologists revise CSO, the method provided a set of correction recommendations.

The experimental results showed that our method is consistent and efficient as it can detect errors within CSO and present correction recommendations.

As future work, we plan to extend the anti-patterns evaluation to other research groups. Depending on their observations, we will extend our proposed patterns and anti-patterns. In addition, we intend to integrate CSO in a cloud federation environment and then interrogate it with the users’ requirements.

REFERENCES