

Received September 7, 2020, accepted September 20, 2020, date of publication September 24, 2020, date of current version October 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3026397

Assessing Ontology Mappings on a Level of Concepts and Instances

MARCIN PIETRANIK¹, ADRIANNA KOZIERKIEWICZ¹, AND MATEUSZ WESOŁOWSKI

Faculty of Computer Science and Management, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

Corresponding author: Marcin Pietranik (marcin.pietranik@pwr.edu.pl)

This work was supported by the National Science Centre, Poland under Grant 2017/26/D/ST6/00251.

ABSTRACT In recent years, the number of domain ontologies on the Internet has been steadily increasing. Many ontologies describe overlapping universes of discourse in various ways, therefore, the need for an efficient ontology alignment method is required. Currently, there are many solutions for this problem. However, the only known way to evaluate their output is to confront it with some pre-prepared reference alignment, therefore making it impossible to incorporate in real-world applications where no reference alignment is given. This paper presents some innovative methods of evaluating ontology alignments which allows assessing their quality without the aforementioned reference alignment. The main contribution are formal foundations of such methods, algorithms developed based on those foundations, and an experimental verification of their usefulness.

INDEX TERMS Knowledge-based systems, knowledge management, ontology alignment.

I. INTRODUCTION

In recent years, the number of domain ontologies has been steadily increasing. It is caused by the fact that they provide a convenient and expressive way of describing some universe of discourse. Their foundation is a set of well-defined concepts, which represent classes of objects from the real world, along with relationships that occur between them. Additionally, class descriptions can be enriched with definitions of their instances, which represents specific materializations of classes. Such freedom in expressing some assumed domain entails the high level of heterogeneity between independently developed ontologies.

When communication of two independently developed information systems that utilize ontologies is expected, some kind of a bridge between them is obviously required. This task can be described as designating which elements of ontologies express the same parts of a modeled universe of discourse to eventually create a set of mappings between ontologies. In the literature, this task is called ontology alignment.

One of the examples of practical appearances of this issue is providing communication between two medical systems which utilize different ontologies within their knowledge bases ([1]). It is not uncommon that systems operating

different therapeutic devices (e.g. MRI or CT) incorporate different clinical healthcare terminologies (e.g. SNOMED-CT or ICD10). Asserting their interoperability requires to identify which parts of these terminologies refer to the same concepts in the real world. In other words - an unequivocal communication between two such systems must be preceded by designating a mapping between contents of utilized ontologies ([16]).

As easy to expect, many systems offering ontology alignment. Their number and diversity naturally entail a necessity of comparing ontology mappings they generate to judge which is better. The most common way of performing such a comparison is using an approach developed by the Ontology Alignment Evaluation Initiative ([6]). It is based on a benchmark dataset, which contains a large number of ontologies along with prepared, reference alignments between them. An alignment provided by some system is confronted with such reference alignment, by calculating values of Precision, Recall, and F-measure. In consequence, deciding about its quality is made possible.

The described procedure is very straightforward and convenient. However, it has two disadvantages. First, it requires that for every pair of ontologies in the dataset their reference alignment needs to be prepared before any evaluation can take place. This task may be very difficult and cumbersome. Secondly, the requirement of such pre-prepared reference

The associate editor coordinating the review of this manuscript and approving it for publication was Justin Zhang¹.

alignment makes the evaluation impossible when two production-ready ontologies need to be mapped. In such a case it is obvious that no reference alignment is available. Thus, there is no way to judge the quality of the designated alignment.

The described situation may be very problematic in the context of the aforementioned communication of medical systems. Imagine that in some hospital, two medical health-care systems operate. Due to new business requirements, an exchange of information about patients must be provided. One of the steps in implementing the solution for the given requirements involves utilizing some ontology alignment methods. However, if the obtained results are different from one another, how a system architect should decide which alignment is the best? Obviously, there is no pre-existing reference alignment to compare to. Moreover, if such reference alignment there would be no need to create new ones.

Therefore, we believe that there is a necessity of creating alternative methods of assessing the ontology alignments quality without any reference alignment. They could be used to support deciding which of two or more competing alignments of two real-world ontologies should be deployed in a production environment. Moreover, the proposed methods should be based only on information available in aligned ontologies. In this paper, we address this issue, by developing some innovative methods of evaluating ontology alignments, which allow comparing them without the flaws of the previous approaches to this task. At first, we identified that the alignments of two ontologies usually contain mappings of concepts and instances. Thus, in our research, we focused on these two elements available in ontologies.

Both concepts and instances are characterized by the concepts' taxonomy. Obviously, the deeper a concept is placed within a hierarchy, the more detailed knowledge it and its instances express. Beginning with the most general classes that categorize entities at a high level of abstraction, more and more complications may appear as classes go deeper into ontology. This remark was a foundation of our research on assessing ontology mappings. We claim that the mappings of classes located lower in the hierarchy are more important than mappings of more general classes that represent less detailed knowledge. Such specific alignments represent more precisely the interoperability potential of two ontologies.

The second factor that influences concepts and instances is the way they are related to other concepts and other instances. The purpose of mapping ontologies is merging knowledge bases of related fields, therefore, it can be expected that similar subtrees of ontologies will be repeated within two ontologies that are aligned. When the ontology alignment includes mappings of classes placed in the same subtree of taxonomy it is more likely that such alignment can score high values of accuracy and high if confronted with a reference alignment. From a user perspective, a smaller, more focused alignment can be more valuable than a large set of mappings of unrelated classes taken from the entire ontologies. Therefore, we claim that a set of mappings of closely related concepts with a high

level of continuity between them are more important than mappings of classes dispersed throughout whole ontologies.

The presented observations allowed us to formulate the four main Research Goals that are eventually addressed in this article. Let's assume that \tilde{O} is a finite set of ontologies, and A represents some alignment between two ontologies (which is a set of correspondences between two elements from these two ontologies) taken from \tilde{O} . Then the Research Goals can be defined as follows:

- 1) To develop a function $\lambda_D : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{R}^+$ that can be used as a criterion of assessing ontology mappings on the level of concepts based the depth of the mapped classes.
- 2) To develop a function $\lambda_C : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{N}$, that can be used as a criterion of assessing ontology mappings on the level of concepts based the continuity of mapped classes.
- 3) To develop a function $\sigma_D : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{R}^+$ that can be used as a criterion of assessing ontology mappings on the level of instances based on the depth of their classes.
- 4) To develop a function $\sigma_C : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{N}$ than can be used as a criterion of assessing ontology mappings on the level of concepts based the continuity of their classes.

The above functions can be treated as novel methods for assessing the quality of ontology alignment based on the content of mapped ontologies and independent from the aforementioned pre-prepared reference alignments. These methods are the main contribution of this article, which is organized as follows. In Section II a state-of-the-art in the considered field is given. Section III contains basic definitions that are a mathematical foundation for our research. In Section IV we describe inconsistencies that may appear in ontology mappings, that are further used in Section V. This part contains the main contribution of the following article - the definitions of novel ontology alignment evaluation function aforementioned earlier. Section V-A includes solutions for Research Goal 1 and 2, while in Section V-B solutions for Research Goals 3 and 4 can be found. The experimental verification is described in Section VI. A summary is given in the last part of the paper.

II. RELATED WORKS

A. ASSESSMENT THE ONTOLOGY MAPPINGS BASED ON THE REFERENCE ALIGNMENT

The number of systems for ontology alignment determination increases each year. Users of such systems expect speed and, in particular, efficiency and correctness of generating mappings. The choice of the best tools for ontology alignment is supported by an international organization OAEI - Ontology Alignment Evaluation Initiative [20]. Its goal is to assess the strengths and weaknesses of alignment and matching systems, comparing the performance of techniques and providing benchmark datasets: the ontologies and correct mapping between them. Since 2004, OAEI organizes evaluation campaigns aiming at evaluating ontology matching technologies.

The assessment criteria are execution time, the number of correspondences, precision, recall, recall+, F-measure, and consistency [6].

The precision and recall are typical measures used in the retrieval information field [12] which are widely accepted. Precision and completeness are based on comparing the resulting alignment (A) with the reference alignment (R). In the context of ontology alignment, precision is defined as the percentage of correctly generated mappings ($|R \cap A|$) among all founded by the proposed algorithm ($|A|$). The recall is the ratio of the number of correctly found mappings ($|R \cap A|$) to all connections taken from reference alignment ($|R|$). However, these measures are not ideal in the context of assessing ontology mappings. The main problem is their binary characteristics [3], [5] - both of those criteria only compare two sets of correspondences without considering if these correspondence are semantically equivalent.

The next measure, called the F-measure, is a harmonic mean of the precision and recall values. Considering only precision it is easy to create an empty set of mappings and obtain the highest precision. Alignment containing all possible correspondences would have a 100% recall. F-measure is free from the mentioned drawbacks and it allows us to assess the quality of the determined alignment. It is calculated as $F_\beta(1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$ where variable β allows to establish balance between precision and recall. In OAEI competition the most popular is F_1 , however for *Conference* track the analysis are made also for $F_{0.5}$ and F_2 [6], [20].

The modifications of the mentioned measures are proposed in [3]. The first and the most intuitive way to extend precision and completeness is to attach the distance between elements in the ontology structure. In this approach, the correspondences which are close to the reference element are not assessed as clearly erroneous. The second modification relies on counting necessary operations to correct correspondences.

The most important generalization of precision and recall are their semantic version [4], [5], [7]. New measures use full knowledge contained in the mapped ontologies. For this purpose, the set of α -consequences have been applied. The first experiments [5] demonstrated that semantic measures better reflect the correctness of mappings. However, further work [4] showed that the semantic measures are dependent on mapping syntax and as a consequence, they can assign different values to semantically equivalent solutions. The subsequent modification does not solve all appearing problems, therefore, OAEI still uses standard versions of precision and recall.

In 2009 OAEI organization proposed a new measure called recall+ [2]. This measure is specially dedicated to the OAEI campaign and consider only the non-trivial mappings. Recall+ is calculated as the number of correct and non-trivial correspondences founded by tested method ($|R \cap A - S|$) divided by the number of all non-trivial references mappings ($|R - S|$, where S -alignment determined by comparing labels.

Another form of assessing ontology alignment is based on the detection of conservativity and consistency principles violations [18], [19]. The consistency principle requires that correspondences should not lead to unsatisfiable classes in the merged ontology, conservativity principle proposes that correspondences should not introduce new semantic relationships between concepts from one of the input ontologies. Typically, it is presented as the number of incoherent connections and their percentage share.

Both standard and modified measures have one significant disadvantage - they all require reference alignment. Creating a reference mapping is a cost- and time- consuming process because it requires involving human experts. If the task at hand is assessing tools dedicated to determining ontology alignment it is not a big problem since OAEI provides benchmark datasets. Despite, assessing the ontology mappings by comparing them with reference alignment has some drawbacks, it is the most reliable evaluation (supported by OAEI). Thus, our developed methods will be verified in relation with recall, precision and F-measure calculated based on reference mappings.

B. ASSESSMENT THE ONTOLOGY MAPPINGS IF REFERENCE ALIGNMENT DOES NOT EXIST

In real applications (i.e. in expert systems or to provide interoperability of medical systems) there is a need to integrate ontologies between which alignment has been designated. Moreover, in such practical scenarios, one cannot expect that any kind of reference alignment exists.

In the first step of ontology integration (determining mappings between input ontologies) it is possible to use one of the available ontology alignment tools. Depending on the size, subject matter, or complexity of the ontology structures, these applications can create mappings of different qualities. In such a context, a big question arises - which of the held alignments should be chosen as the final one?

The most popular approach to this issue is user-centered. The alignments are evaluated by domain experts or groups of users [11]. The ontology alignment validation process consists of asking one or more users to classify the mappings in an ontology alignment as correct or incorrect, as well as potentially replace incorrect mappings with correct alternatives, or even add new mappings. These activities serve as a basis to decide how good the obtained alignment is and thus to compare alignment tools and algorithms. However, involving users for the alignment evaluation process is error-prone, time, and cost-consuming.

A different strategy has been proposed in [8], where the quality of alignment is estimated by the number of the most frequently used correspondences. Selecting the most frequently used correspondences for evaluation is beneficial in two situations. First, if there is a difference in quality between the frequently used correspondences and the infrequently used correspondences, the frequency-weighted precision will give a more reliable estimate of the performance of the application using the alignment. Second, if one intends

a semi-automatic matching process in which suggested correspondences are manually checked and corrected by an expert, the frequency provides an ordering in which to check. However, this evaluation method requires end-user support (which filled a prepared query scenario) and is not fully automatic.

In [13] authors introduced quality measures that are based on the notion of mapping incoherence, which can be used without a reference mapping. Measuring the incoherence of mappings was motivated by the idea that it will hinder its sensible use even though it might contain a significant amount of correct correspondences. The provided measure of incoherence gives a strict upper bound for the precision of mapping and can, therefore, be used as a guideline for estimating the performance of matching systems. However, any experimental verification of the proposed measure has not been conducted, so the real applicability of this method is unknown.

To the best of our knowledge, there is a lack of reliable methods that allow evaluating the quality of alignment and the formulated problem has not been investigated widely. Therefore, easy methods for assessing the correctness of the obtained alignment without comparing it's with reference mapping are desired.

Providing semantic interoperability based on ontology alignment in the context of, for example, the Internet of Things [14] where the diversity of communicating systems makes it impossible to prepare reference alignments beforehand. In consequence, it is crucial to evaluate and assess the quality of particular alignments that are intended to be used. Other practical applications where the methods presented in this paper may be useful are related to smart spaces [17].

The need to compare the quality of different mappings also exists in the process of maintaining several integrated knowledge bases. Ontologies evolve which entails the fact that mappings can become obsolete. Changes like adding, editing, or deleting some elements of ontology influence directly on alignment. Maintaining the current mappings is an important task of ontology developers [15] and none of the classical measures presented in this section provides an easy way for continuously assessing the correctness and the quality of alignments. The solutions presented in this paper are free of the mentioned drawbacks. Such approach may be found useful especially in practical applications which require exchanging domain knowledge.

III. BASIC NOTIONS

The formal model of ontology is based on our previous publications (for example [10]). It distinguishes three main elements. **Concepts** aggregate real-world entities, that have a common set of attributes with an indefinite value. We assume the existence of an abstract class *Thing* from which all of the other classes inherit. Such approach is consistent with a model enforced by OWL ontology representation format. **Relations** which are various types of connections that occur between classes that depend on each other in some way.

In particular, the inheritance relation stands out, which is used when it is possible to aggregate a set of classes into one more general one. **Instances** which represent specific objects in the domain and have specific attribute values. Formally, an ontology is defined as a quintuple:

$$O = (C, H, R^C, I, R^I) \quad (1)$$

In the equation above C denotes set of concepts; H is a concepts' hierarchy; R^C is a set of relations between concepts $R^C = \{r_1^C, r_2^C, \dots, r_n^C\}$, $n \in N$, such that every $r_i^C \in R^C$ ($i \in [1, n]$) is a subset of $C \times C$; I denotes a set of instances'; $R^I = \{r_1^I, r_2^I, \dots, r_n^I\}$ denotes a set of relations between concepts' instances.

The aforementioned inheritance relation H makes it possible to consider an ontology as a taxonomy of concepts. Its root is an abstract class *Thing*, from which all classes of the analyzed ontology inherit. In this context, the classes are arranged in a hierarchy from the most general to the most specific. This approach allows to define the depth properties.

Definition 1: For a given ontology, O , the depth (denoted as $Depth(O)$) of its hierarchy H is the number of subconsumption relationships that divides the parent class *Thing* from the most distant inheritance hierarchy in the class belonging to that ontology.

Definition 2: For a given ontology O and one of its classes $c \in C$, the depth of this class (denoted as $Depth(O, c)$) in the hierarchy H is the number of subsumption relationships in the O 's hierarchy between this class and the root class *Thing*.

Thanks to the above definition we also define a series of auxiliary functions and properties that will be further used in assessing ontology alignments:

- $Root(O)$ - denotes a set of classes in the ontology O which are direct children of the abstract class *Thing*. In other words- this is a set of classes which are in the highest level in the taxonomy;
- $Sub_A(O, c)$ - denotes a set of classes in the ontology O which are descendants of the given class c in the taxonomy H ;
- $Sub_D(O, c)$ - denotes a set of classes in the ontology O which are direct descendants (children) of the given class c in the taxonomy H ;
- $Sup_A(O, c)$ - denotes a set of classes in the ontology O which are predecessors of the given class c in the taxonomy H ;
- $Sup_D(O, c)$ - denotes a set of classes in the ontology O which are direct predecessors (parents) of the given class c in the taxonomy H ;
- $Type(O, i)$ - a function which for given instance i returns a class to which this instance belongs within ontology O ;
- $Ins(O, c)$ - a function which returns a set of all instances belonging to the class c within ontology O .

Designating a mapping between two ontologies involves finding their common parts. The result of this process is a set of corresponding elements from compared ontologies, connected by some relationship at the some confidence level.

These relationships may be equivalence ($=$), refinement ($<$), generalization ($>$) and incompatibility ($\%$).

Following the formal definition of ontologies from Equation 1 one could expect that there are three levels of mapping due to the type of component elements: concepts, instances and relationship. However, both OAEI organization and most ontology mapping systems focus only on the first two levels. Only some of the systems described in the literature, such as SORAL (*Supervised Ontology Relation ALignment*), attempt to efficiently map relations, but the results of these research are not publicly available [9]. For this reason, in this paper we focused only on the ontology alignment on the level of concepts and instances.

Definition 3: For given two ontologies O^1 and O^2 , an alignment A between them is a finite set of quadruples $\langle e^1, e^2, r, n \rangle$ (further referred to as correspondences). Elements respectively from O^1 and O^2 are denoted as respectively; a relationship $r \in \{=, <, >, \%\}$ is a semantic connection describing a type of correspondence. $n \in [0, 1]$ is a value representing a confidence level of a particular correspondence. If $e^1 \in C^1 \wedge e^2 \in C^2$ then such correspondences occurs on the level of concepts, and if $e^1 \in I^1 \wedge e^2 \in I^2$ then such correspondences occurs on the level of instances. Note that both e^1 and e^2 must refer to elements from the same level, therefore, there is no possibility that a correspondence between a concept and an instance exists.

IV. INCONSISTENCIES IN ONTOLOGY MAPPINGS

Ontology alignment should not contain mappings that lead to violating target ontology restrictions. Such relations between ontologies elements can cause malfunctioning of reasoning engines, but most importantly they can be, with high probability, just incorrect. As part of this work, two cases are defined in which restrictions arising from basic taxonomic relationships (inheritance and disjointness) are violated.

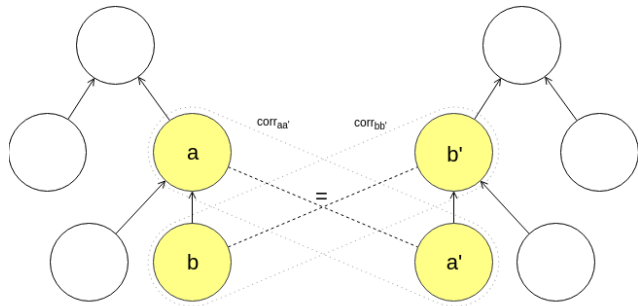


FIGURE 1. An example case of circular inheritance in aligned ontologies.

A. CIRCULAR INHERITANCE

Ontology cannot be considered consistent if it contains classes that inherit from each other. Assuming that alignment is the reason for a circular inheritance in target ontology, it means it has inconsistent correspondences. This case, as an example, is shown in Figure 1.

Graphs represent integrated ontologies, nodes - their classes, and edges - inheritance relationships. The yellow nodes are linked by dotted lines which symbolize correspondences ($corr_{bb'}$ and $corr_{aa'}$) of a certain alignment. Class b is subclass of a and has been linked with b' class. While b' is superclass of a' which is linked with a . As a result of these mappings, it is possible to conclude that a is both a superclass and a subclass of b .

Definition 4 (Circular Inheritance Inconsistency): Assuming the existence of ontologies O^1 and O^2 , classes $a, b \in C^1$, $a', b' \in C^2$ and alignment A , correspondences $corr_{aa'} = \langle a, a', =, n_{aa'} \rangle$, $corr_{aa'} \in A$ and $corr_{bb'} = \langle b, b', =, n_{bb'} \rangle$, $corr_{bb'} \in A$ are inconsistent because of circular inheritance when:

$$a \in Sup_A(O^1, b) \wedge b' \in Sup_A(O^2, a'), \quad (2)$$

To denote this inconsistency \neq_C symbol is used and therefore a statement $corr_{aa'} \neq_C corr_{bb'}$ is true for the given example.

B. INHERITANCE FROM DISJOINT CLASS

For classes OWL provides a *disjointWith* relation which is denoted in this work by an operator \neq . This means that there cannot be an instance that belongs to both classes connected by such relations. Furthermore, the subclasses of these classes also inherit this restriction. Inconsistency, that may be a result of a violation of the disjoint restriction, is shown in Figure 2.

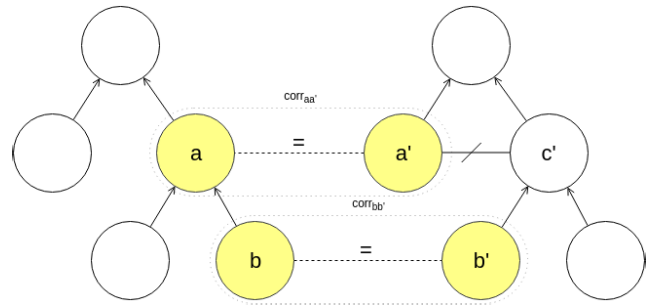


FIGURE 2. An example case of inheriting from disjoint class in aligned ontologies.

Class a is aligned to a' class ($corr_{aa'}$) and its subclass b is aligned to b' class ($corr_{bb'}$). The a' and c' classes are disjoint, which is symbolized in the diagram by the crossed out line that links them. It means that b' is also disjoint with a' . This alignment is inconsistent because in the reasoning process the b' class will be considered as descendant of the a' class with which it is disjoint.

Definition 5 (Inheritance From Disjoint Class Inconsistency): Assuming the existence of ontologies O^1 and O^2 , classes $a, b \in C^1$, $a', b' \in C^2$ and alignment A , correspondences $corr_{aa'} = \langle a, a', =, n_{aa'} \rangle$, $corr_{aa'} \in A$ and $corr_{bb'} = \langle b, b', =, n_{bb'} \rangle$, $corr_{bb'} \in A$ are inconsistent because of inheritance from disjoint class when:

$$\exists_{c' \in Sup_A(O^2, b')} c' \neq a' \wedge b < a \quad (3)$$

To denote this inconsistency $\not\equiv_R$ symbol is used and therefore a statement $corr_{aa'} \not\equiv_R corr_{bb'}$ is true for the given example.

C. INCONSISTENT CORRESPONDENCE

The definition in subsections IV-A and IV-B were used to define inconsistent correspondences. When two pairs of aligned classes conflict with each other, a correspondence that is lower in the hierarchy is referred to as an inconsistent correspondence.

Definition 6 (Inconsistent correspondence): Assuming the existence of ontologies O^1 and O^2 , classes $a \in C^1$, $a' \in C^2$ and an alignment A , a correspondence $corr_{aa'} = \langle a, a', =, n_{aa'} \rangle$, $corr_{aa'} \in A$ is inconsistent when:

$$\exists_{corr_{bb'} \in A} (b \in Sup_A(O^1, a) \vee b' \in Sup_A(O^2, a')) \wedge (corr_{aa'} \not\equiv_C corr_{bb'} \vee corr_{aa'} \not\equiv_R corr_{bb'}) \quad (4)$$

To denote this, $\not\sim$ symbol is used. Therefore, a following statement: $corr_{aa'} \not\sim A$ is true for defined elements and can be understood as 'a correspondence that links classes a and a' is inconsistent in A alignment'.

V. METHODS OF ASSESSING ONTOLOGY MAPPINGS

In this paper, we propose some novel approaches to evaluating ontology mappings. The first one is developed on the criterion based on the depth of classes in the hierarchy of integrated ontologies, and the second one on a criterion based on a continuity of a mappings. Both can be used to assess mappings of concepts and instances. The following section contains definitions of those criteria and methods of incorporating them to assess ontology mappings.

A. ASSESSING ONTOLOGY MAPPING ON THE LEVEL OF CONCEPTS

1) CRITERION BASED ON THE DEPTH OF THE MAPPED CLASSES

The deeper class in the hierarchy is, the more detailed knowledge can be learned from it. Beginning with the most general classes that categorize entities at a high level of abstraction, more and more complications may appear as classes go deeper into ontology. Depending on how a domain has been modeled, differences in the way entities are described may cause difficulties in designating mapping between them. It is for this reason that during finding such corresponding elements the algorithms come across many issues. Therefore, we claim that if these difficulties are solved, mappings of classes located lower in the hierarchy are more important. This remark is a backbone of the criterion based on the depth of the mapped classes presented in this section, which addresses the Research Goal 1 defined in Section I. The solution of this goal can be treated as one of the proposed methods for assessing an ontology alignment quality.

The proposed measure will assess with more points mappings that connect classes that are located deep in the class hierarchy. Formally, λ_D gives points for every found cor-

respondence that is not inconsistent, while the value of γ_D increases with the depth in the hierarchy of mapped classes. For some selected correspondence $corr_{cc'} = \langle c, c', =, n \rangle$ the value of γ_D is calculated according to the formula:

$$\gamma_D(O^1, O^2, A, corr_{cc'}) = \begin{cases} \frac{1}{Depth(O^1) - Depth(O^1, c) + 1} \\ + \frac{1}{Depth(O^2) - Depth(O^2, c') + 1} & \text{if } corr_{cc'} \sim A \\ 0 & \text{if } corr_{cc'} \not\sim A \end{cases} \quad (5)$$

Ontology alignments evaluation function $\lambda_D : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{R}^+$, based on a depth of mapped classes in the ontology hierarchy, gets two ontologies and an alignment between them and returns sum of all correspondences' γ_D . \tilde{O} symbolizes set of all ontologies and \tilde{A} - set of all ontology alignments. The formula for this function is:

$$\lambda_D(O^1, O^2, A) = \sum_{corr_{cc'} \in A} \gamma_D(O^1, O^2, A, corr_{cc'}) \quad (6)$$

Figures 3 and 4 present in the form of graphs two different alignments A^1 and A^2 of the same ontologies from the perspective of one of them. Mapped classes of this ontology are marked in yellow. Each of these alignments consists of the same number of correspondences: 8, but the second example maps classes from lower parts. According to Equation 6, the mapping A^1 was assessed with 2.42 and A^2 with 4.25. Despite the same number of correspondences, A^2 was rated higher because it relates to more accurate entities.

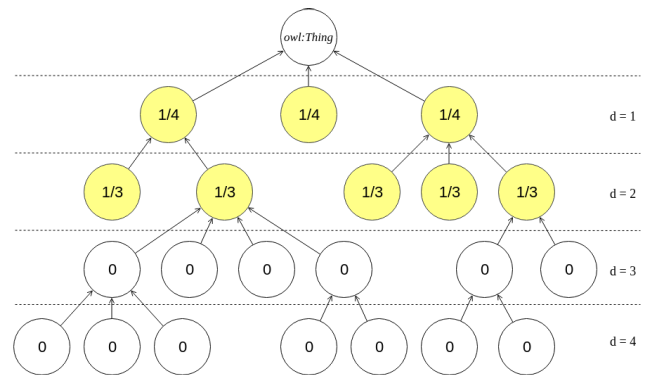
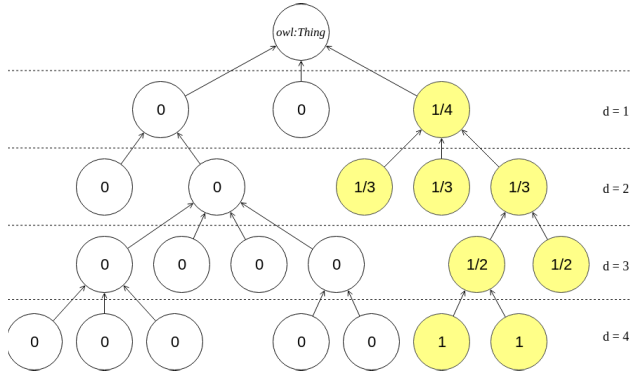


FIGURE 3. $\lambda_D(A^1) = 2.42$.

The way of calculating the value of λ_D for the alignment A of ontologies O^1 and O^2 is presented step by step on the Algorithm 1. In the first step, the measure λ_D is initiated with 0. Then, for each correspondence of A alignment, value γ_D is calculated. With each loop iteration λ_D is increased by γ_D and eventually is returned as result.

In the example of medical systems communication from Section I, the method presented in the current section will favor alignments of concepts from two ontologies that allow to exchange the detailed knowledge about specific medical

FIGURE 4. $\lambda_D(A^2) = 4.25$.

Algorithm 1 Evaluation Based on the Depth of the Mapped Classes

Require: O^1 , O^2 and A

Ensure: λ_D

```

1:  $\lambda_D \leftarrow 0$ 
2: for all  $corr_{cc'} \in A$  do
3:   if  $corr_{cc'} \sim A$  then
4:      $\gamma_D \leftarrow \frac{1}{Depth(O^1) - Depth(O^1, c) + 1} + \frac{1}{Depth(O^2) - Depth(O^2, c') + 1}$ 
5:   else
6:      $\gamma_D \leftarrow 0$ 
7:   end if
8:    $\lambda_D \leftarrow \lambda_D + \gamma_D$ 
9: end for
10: return  $\lambda_D$ 

```

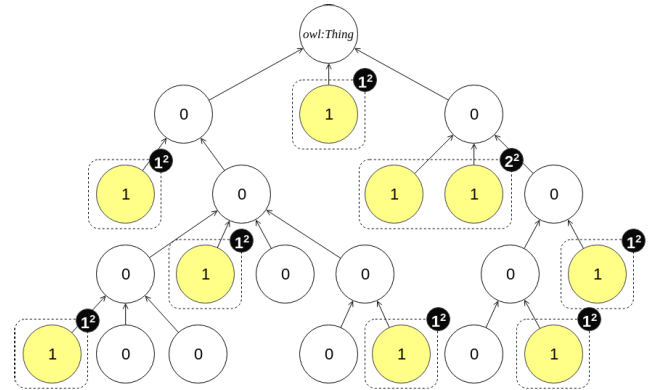
conditions of patients. Such alignments are treated as far more valuable mappings of some generic information stating that a patient is sick. Therefore, the proposed method of evaluating ontology alignments based on the depth of the mapped classes (which is a straight realization of Research Goal 1) is very intuitive and can be proved useful in practical applications.

2) CRITERION BASED ON THE CONTINUITY OF MAPPED CLASSES

The most common way of storing correspondences between two ontologies is a dedicated RDF document which contains tuples describing subsequent correspondences (consistent with the Definition 3) of elements taken from aligned ontologies. Comparing only such correspondences without a context of ontologies they connect (based on the number of correspondences in the RDF document) does not require any complex analysis and can indicate alignment which is more complete. However, as aforementioned, such an approach required knowledge about correct alignment. This requirement is obviously unrealistic in practical applications. Therefore, having two competing, similarly-sized alignments of ontologies with no information about the correctness of correspondence, it is worth analyzing their structures.

If correspondences refer to classes that are not related in any way, then the structure of the alignment looks chaotic and incoherent. Based on this observation, it can be concluded that in methods described in the literature during designating ontology mapping, the knowledge about entities was used superficially. Moreover, there is a chance that such correspondences are incorrect and the context of the classes has not been understood. One example would be linking two different entities with the same names (homonyms). In this case, both their parent and child classes of these entities are unlikely to be mapped.

In the opposite situation, when ontology alignment contains correspondences of classes that, from a taxonomic point of view, are part of the same subtrees, one can expect from such alignment accuracy and high precision. Especially since mapping ontologies aim to merge knowledge bases of related fields, so it is assumed that both entities and relationships between them will be repeated. Also from the user perspective, a smaller alignment that focuses on ontologies fragments can bring more benefits than bigger alignment that maps different, unrelated classes across the entire ontologies. In this case, assessing alignments depends on how close taxonomically their correspondences are.

FIGURE 5. $\lambda_C(A^1) = 11$.

Presented observations were the inspiration to work on the second assessment criterion, which is based on the continuity of mapped classes and addresses the Research Goal 2 defined in Section I. The outcome of this goal is a second of the proposed methods for assessing a ontology alignment quality. This time, instead of individual correspondences, groups of classes that form consistent subtrees are scored. The example of scoring alignments in this method is presented in Fig. 5 and 6 where diagrams presenting two different alignments from the perspective of one of two integrated ontologies. Subtrees that bring points consist only of consistent, mapped classes with a common ancestor. The alignment A^1 (Fig. 5) contains nine correspondences, but only two mapped classes in the presented ontology have the same direct ancestor. In contrast, the second alignment A^2 has two correspondences less, but it maps classes closely related to each other. In that case, one seven-element subtree was created.

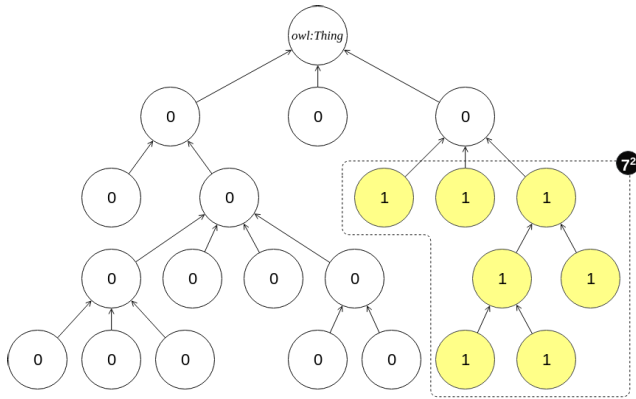


FIGURE 6. $\lambda_C(A^2) = 49$.

For this criterion, the rating function γ_C which determines the number of points for a given subtree t was defined as follows:

$$\gamma_C(t) = |t|^2, \quad (7)$$

where $|t|$ is the number of classes in subtree t .

The values of γ_C calculated for every subtree are presented in black circles congruent to them. By adding these numbers it is easy to conclude that despite the smaller number of mapped classes, alignment A^2 represents higher quality.

Ontology alignments evaluation function $\lambda_C : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{N}$, based on the continuity of mapped classes, gets two ontologies and alignment and returns sum of all subtrees' γ_C . \tilde{O} denotes a set of all ontologies and \tilde{A} - set of all ontology alignments. The formula for this function is:

$$\lambda_C(O^1, O^2, A) = \sum_{t \in \text{SubTrees}(O^1, A)} \gamma_C(t) + \sum_{t \in \text{SubTrees}(O^2, A)} \gamma_C(t), \quad (8)$$

where $\text{SubTrees}(O^1, A)$ and $\text{SubTrees}(O^2, A)$ are collections of subtrees determined on the basis of O^1 and O^2 respectively. The subsequent steps to calculate λ_C for the alignment A of ontologies O^1 and O^2 are presented on the Algorithm 2.

As part of this algorithm, the auxiliary recursive function *getConsistencyScore* (line 5) was defined. It is called for a set of classes inheriting directly from the *Thing* class, and then, recursively, in a loop for the direct ancestors of these classes (line 9). The variable t stores information about the classes that are part of the currently processed subtree, and λ_C^O the current number of points. If the class is mapped and its correspondence is consistent (line 10), the subtree t is increased by this class (line 11) and its ancestors (line 12) that meet the same requirements. Then λ_C^O increases by the points gained by the already closed subtrees of ancestors (line 13). Otherwise, the subtree is closed and scored, and λ_C^O is increased by the number of these points and points scored by the ancestors. Recursive calls of this function return t and λ_C^O , and the originally called function scores

Algorithm 2 Evaluation Function Based on the Continuity of Mapped Classes

Require: $O^1, O^2 \in A$

Ensure: λ_C

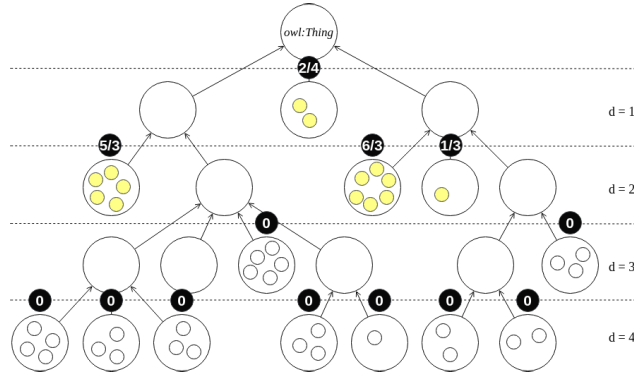
```

1:  $\lambda_C^{O^1} \leftarrow \text{getConsistencyScore}(A, O^1, \text{Root}(O^1), \text{true})$ 
2:  $\lambda_C^{O^2} \leftarrow \text{getConsistencyScore}(A, O^2, \text{Root}(O^2), \text{true})$ 
3:  $\lambda_C \leftarrow \lambda_C^{O^1} + \lambda_C^{O^2}$ 
4: return  $\lambda_C$ 
5: procedure getConsistencyScore( $A, O, \text{directSubclasses}, \text{isRoot}$ )
6:    $t \leftarrow \text{empty tree}$ 
7:    $\lambda_C^O \leftarrow 0$ 
8:   for all  $\text{subclass} \in \text{directSubclasses}$  do
9:      $cs \leftarrow \text{getConsistencyScore}(A, O, \text{Sub}_D(O, \text{subclass}), \text{false})$ 
10:    if  $\exists \text{corr}_{cc'} \in A (\text{subclass} = c \vee \text{subclass} = c') \wedge \text{corr}_{cc'} \sim A$  then
11:      add  $\text{subclass}$  to  $t$ 
12:      add all  $cs.t$  to  $t$ 
13:       $\lambda_C^O \leftarrow \lambda_C^O + cs.\lambda_C^O$ 
14:    else
15:       $\gamma_C \leftarrow \text{size}(cs.t)^2$ 
16:       $\lambda_C^O \leftarrow \lambda_C^O + cs.\lambda_C^O + \gamma_C$ 
17:    end if
18:  end for
19:  if  $\text{isRoot}$  then
20:     $\gamma_C \leftarrow \text{size}(t)^2$ 
21:    return  $\lambda_C^O + \gamma_C$ 
22:  else
23:    return  $\{t, \lambda_C^O\}$ 
24:  end if
25: end procedure

```

last processed subtree and returns the sum of all points. The λ_C value consists of ratings calculated for both ontologies by *getConsistencyScore* (line 3) and it is returned as final result.

It is obvious to claim an alignment containing mappings of concepts that are closely clustered within ontology is better than an alignment of single concepts dispersed throughout aligned ontologies. Mappings of such closely related concepts, which concern similar topics, may become more important especially in the context of presented in Section I example of medical ontologies alignments. Such an approach to assessing ontology alignment would allow exchanging between medical systems knowledge that is more focused on certain medical conditions. Sharing only knowledge from single random concepts taken from ontologies may not prove useful in such an application. Therefore, the proposed method of evaluating ontology alignments based on the depth of the continuity of mapped classes (which is a straight realization of Research Goal 2) may become very valuable.

FIGURE 7. $\sigma_D(A^1) = 4.5$.

B. ASSESSING ONTOLOGY MAPPINGS ON THE LEVEL OF INSTANCES

Ontology classes are elements that classify, at a different level of detail, entities with the same properties. Actual entities are in the form of instances and are assigned to a given class with *rdf:type* relation. In order to assess ontology alignments, which in addition to classes contain also instances, previously proposed criteria were adapted to consider this type of element too. Proposed methods take into account only those instances that are classified into the classes defined within the analyzed ontology.

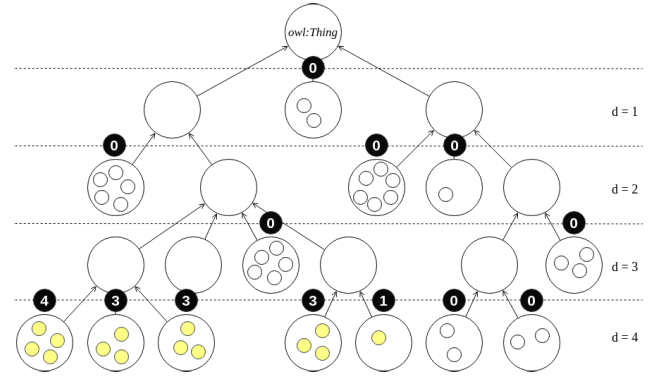
1) CRITERION BASED ON THE DEPTH OF THE MAPPED CLASSES

This section addresses Research Goal 3 defined in Section I. We claim that mappings of classes that are deeper in the taxonomical hierarchy may be more important, instance mappings of these classes may also be considered more valuable. As such classes are more specialized, also their instances can provide knowledge that is more detailed for a given field. Hence, the measure σ_D was defined, which is analogous to the proposed in the previous chapter measure λ_D , but focuses on instances. The final form of this measure can be used to assess the quality of some ontology alignment.

The γ_D equivalent for depth-based measure at instance level is ψ_D . The ψ_D takes into account classes that instances belong to. If classes of mapped instances are disjoint, the value of ψ_D is equal to zero. Otherwise, ψ_D is calculated based on the depth of these classes in the ontology taxonomical hierarchy. Definition of the ψ_D for correspondence $corr_{i,i'} = \langle i, i', =, n \rangle$, where $c = Type(O^1, i)$ and $c' = Type(O^2, i')$ is:

$$\psi_D(O^1, O^2, A, corr_{i,i'}) = \begin{cases} \frac{1}{Depth(O^1) - Depth(O^1, c) + 1} \\ + \frac{1}{Depth(O^2) - Depth(O^2, c') + 1} & \text{if } c \equiv c' \\ 0 & \text{if } c \not\equiv c' \end{cases} \quad (9)$$

Figures 7 and 8 present two different alignments A^1 and A^2 from the perspective of one of the two integrated ontologies.

FIGURE 8. $\sigma_D(A^2) = 14$.

As in the previous diagrams, the nodes represent ontology classes, but this time instances of these classes are also included. They are represented by smaller circles located inside the nodes. Instances with yellow ones have been mapped. The value of ψ_D for each class is placed on diagrams above them. Both alignments have the same number of instance correspondences, but according to the defined criterion, A^2 is more valuable as it may contain more important mappings.

Ontology alignments evaluation function $\sigma_D : \tilde{O} \times \tilde{O} \times \tilde{A} \rightarrow \mathbb{R}^+$, based on the depth of the classes of mapped instances, gets two ontologies and alignment and returns sum of all instance correspondences' ψ_D . \tilde{O} symbolizes set of all ontologies and \tilde{A} - set of all ontology alignments. The formula for this function is:

$$\sigma_D(O^1, O^2, A) = \sum_{corr_{i,i'} \in A} \psi_D(O^1, O^2, A, corr_{i,i'}), \quad (10)$$

Based on this definition, the Algorithm 3 has been developed, which contains subsequent steps to calculate a value of the σ_D function.

Algorithm 3 Evaluation Based on the Deep of Classes of Mapped Instances

Require: O^1, O^2 i A

Ensure: σ_D

```

1:  $\sigma_D \leftarrow 0$ 
2: for all  $corr_{i,i'} \in A$  do
3:    $c \leftarrow Type(O^1, i^{corr_{i,i'}})$ 
4:    $c' \leftarrow Type(O^2, i'^{corr_{i,i'}})$ 
5:   if  $c \neq c'$  then
6:      $\psi_D \leftarrow 0$ 
7:   else
8:      $\psi_D \leftarrow \frac{1}{Depth(O^1) - Depth(O^1, c) + 1} + \frac{1}{Depth(O^2) - Depth(O^2, c') + 1}$ 
9:   end if
10:   $\sigma_D \leftarrow \sigma_D + \psi_D$ 
11: end for
12: return  $\sigma_D$ 

```


Algorithm 4 Evaluation Based on the Continuity of Mapped Instances**Require:** $O^1, O^2 \text{ i } A$ **Ensure:** σ_C

```

1:  $\sigma_C^{O^1} \leftarrow \text{getInstanceConsistencyScore}(A, O^1)$ 
2:  $\sigma_C^{O^2} \leftarrow \text{getInstanceConsistencyScore}(A, O^2)$ 
3:  $\sigma_C \leftarrow \sigma_C^{O^1} + \sigma_C^{O^2}$ 
4: return  $\sigma_C$ 
5: procedure  $\text{getInstanceConsistencyScore}(A, O)$ 
6:    $\sigma_C^O \leftarrow 0$ 
7:   for all  $c \in C$  do
8:      $s \leftarrow \text{Ins}(O, c)$ 
9:      $s_A \leftarrow \text{empty collection}$ 
10:    for all  $ins \in s$  do
11:      if  $\exists_{corr, i' \in A}(ins = i \vee ins = i')$  then
12:        add  $ins$  to  $s_A$ 
13:      end if
14:    end for
15:     $\psi_C = \text{size}(s_A)^2$ 
16:     $\sigma_C^O \leftarrow \sigma_C^O + \psi_C$ 
17:  end for
18:  return  $\sigma_C^O$ 
19: end procedure

```

variable (line 18). The final value is the sum of the scores calculated based on both ontologies (line 3).

The proposed method is similarly useful to the one presented in Section V-A2. Intuitively, when there are two alternative alignments, one including mappings of instances of concepts that are closely related (and therefore, related to a single topic of interest), and the other including mappings of random concept instances, the first should be treated as more expressive, thus better. In the context of a medical system communication from Section I, more practical effect should have a detailed communication about a single topic, rather than a communication concerning multiple different topics, but treated more sketchy.

VI. EXPERIMENTAL VERIFICATION OF DEVELOPED METHODS

According to the literature (reviewed in Section II), the most common method of evaluating ontology alignments is based on calculating Precision, Recall, and F-measure. To the best of our knowledge, there is no other widely accepted method of assessing ontology alignments, especially when a reference alignment is not provided. There is very little research (which can be found in Section II-B) devoted to the given problem and most of the found publications assume the existence of a reference alignment (please refer to Section II-A).

What is worth emphasizing is the fact that the Ontology Alignment Evaluation Initiative (OAEI), the most widely known organization that focuses on ontology alignment research, in their evaluation campaigns aiming at evaluating ontology matching technologies also omits the issue

addressed in the submitted article ([20]). However, all of the available publications devoted to ontology alignment utilize the datasets they provide. These are the reasons why we have based our experiment on comparison with the experimental methodology designed by OAEI.

Chapter V presents four innovative methods of alignment assessment at two levels - classes and instances. What distinguishes the proposed criteria is no need to have a reference alignment. The purpose of our experiment was the evaluation of the usability of these methods and to compare them with commonly used alignment assessment measures, which are the F-measure, precision, and recall.

A. EXPERIMENTAL PROCEDURE

The following assumptions were defined: alignment is in RDF format, and ontologies, consisting of many elements with a multilevel taxonomic hierarchy, are in OWL format. For these reasons, for the evaluation of the defined algorithms, the data provided by the OAEI organization will be used. Due to space limitations, we will only evaluate algorithms developed for the concept level within Research Goal 1: the criterion of assessing ontology mappings on the level of concepts based the depth of the mapped classes, and Research Goal 2: the criterion of assessing ontology mappings on the level of concepts based the continuity of mapped classes. Evaluation of methods for instance-level can be performed in an analogous way using different OAEI datasets.

OAEI provides ontologies grouped thematically along with reference alignments between them. Subsequently, outcomes of a variety of alignment tools for ontologies from each track are provided. These include sizes of created mappings and their comparison with reference alignments in terms of basic assessment measures: precision, recall, and their harmonic mean F-measure. For every alignment tool in each track, the values of the proposed assessment measures λ_D and λ_C of individual alignments will be calculated. Then, the obtained results will be compared with the aforementioned standard measures.

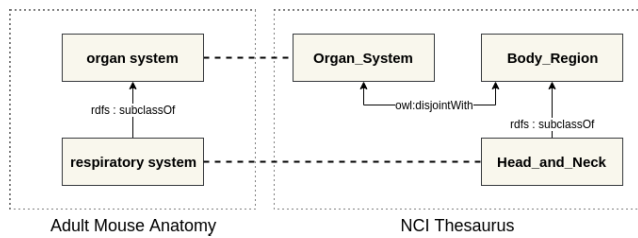
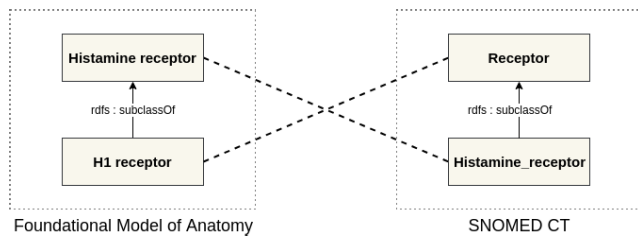
Table 1 contains selected OAEI ontology matching tracks from three categories: *Anatomy* (Adult Mouse Anatomy and NCI Thesaurus Ontology), *Biodiversity and Ecology* (AFlorea Phenotype and Plant Trait Ontology) and *Large Biomedical Ontologies* (Foundational Model of Anatomy, SNOMED CT, NCI Thesaurus Ontology). Each of the selected ontologies is composed of more than 1500 classes, with a minimum taxonomical hierarchy depth equals 9.

To analyze the intraclass correlation between assessments across the entire data set, regardless of the alignment belonging to the track, the determined measures λ_D and λ_C will be normalized to the reference alignment measures in given tasks. For the purpose of this experiment, a Java software application was implemented. *Apache Jena*¹ library and

¹<https://jena.apache.org/index.html>

TABLE 1. Selected OAEI ontology matching tracks.

No.	Category	Ontology	Classes	Depth
1	Anatomy	Adult Mouse Anatomy	2744	9
		NCI Thesaurus	3304	14
2	Biodiversity and Ecology	Flora Phenotype Ontology	28932	13
		Plant Trait Ontology	1504	14
3	Large Biomedical Ontologies	Foundational Model of Anatomy	10157	20
		SNOMED CT	13412	34
4	Large Biomedical Ontologies	Foundational Model of Anatomy	3696	20
		NCI Thesaurus	6488	14

**FIGURE 11.** FCAMapX alignment inconsistency.**FIGURE 12.** KEPLER alignment inconsistency.

SPARQL² language it supports were used to read and navigate through RDF graph.

B. EXPERIMENT RESULTS

The experiment began with calculating values of λ_D and λ_C measures, which are methods developed within Research Goals 1 and 2 defined in Section I. During this process, information on the found inconsistencies was collected. It was observed that some of the found conflicts were caused by different ways of conceptualizing domains. However, most of them pointed to real errors. Figures 11 and 12 show sample inconsistencies found in the alignments. FCAMapX linked the classes *respiratory system* and *head and neck* as equivalent. This conflict was successfully detected thanks to the disjoint relationship between the classes *organ system* and *body area* in *NCI Thesaurus*. The second example could be easily fixed by replacing the equivalence with the subsumption relationship. *H1 receptor* is type of *histamine receptors*. The error is to match a specific receptor to the general one, probably because of the high similarity of their names.

The next task was to process designated assessments. For ontologies O^1 and O^2 and their reference alignment R , a normalized measure λ_D which is a method of assessing ontology alignment based on the taxonomic depth of alignment A (developed during Research Goal 1) is defined by a formula:

$$\hat{\lambda}_D(O^1, O^2, A) = \frac{\lambda_D(O^1, O^2, A)}{\lambda_D(O^1, O^2, R)} \quad (13)$$

A normalized measure λ_C (method of assessing ontology alignment based on the taxonomic continuity of the alignment A developed within Research Goal 2) is described by a formula:

$$\hat{\lambda}_C(O^1, O^2, A) = \frac{\lambda_C(O^1, O^2, A)}{\lambda_C(O^1, O^2, R)}, \quad (14)$$

The denominator in Equation 13 and Equation 14 (which contains values of the proposed measures calculated for reference alignments) is introduced to normalize obtained values that can acquire values higher than 1. It is done to perform a statistic comparison with ontology alignment quality measures taken from the literature.

All data (including normalization) results are shown in Table 2. These data consist of 5 different measures, each of which is a certain percentage in relation to the ‘gold standard’ in a given track. Using the nonparametric Kolmogorov–Smirnov test with a significance level of $\alpha = 0.01$ it was shown that each of the studied sets comes from the normal distribution. Therefore, it was possible to use the intra-class correlation coefficient (ICC) for compliance analysis. According to ICC theory, these methods will be seen as judges who give marks to each matcher. The purpose of the tests is to calculate the degree to which their assessments are consistent. The results are shown in table 3.

Compliance of the proposed methods measures with recall is at a very high level, which is indicated by $ICC(3, 1)$ values, which are close to 1 (maximum value). It is different in the case of compliance with precision, where the results are completely divergent. There, the $ICC(3, 1)$ value is below zero. It is also worth noting that the criterion based on taxonomic depth more closely matches the recall and F-measure than the criterion based on continuity. F-Measure used for analysis is a measure of F1, where the share of recall and precision is equal. Despite this, the compliance of the proposed assessments with the F-measure is at a high level as evidenced by the

²<https://www.w3.org/TR/rdf-sparql-query/>

TABLE 2. Size and measures of individual matchers.

Track	Matcher	Size	Precision	Recall	F-Measure	$\hat{\lambda}_D$	$\hat{\lambda}_C$
1	ALIN	928	0.998	0.611	0.758	0.630	0.430
1	ALOD2Vec	987	0.996	0.648	0.785	0.667	0.477
1	AML	1493	0.950	0.936	0.943	0.984	1.076
1	DOME	935	0.997	0.615	0.761	0.632	0.433
1	FCAMapX	1274	0.941	0.791	0.859	0.843	0.712
1	Holontology	456	0.976	0.294	0.451	0.282	0.131
1	KEPLER	1173	0.958	0.741	0.836	0.783	0.605
1	LogMap	1397	0.918	0.880	0.846	0.918	0.746
1	LogMapBio	1550	0.888	0.908	0.898	1.026	0.994
1	LogMapLite	1147	0.962	0.728	0.828	0.770	0.566
1	POMAP++	1446	0.919	0.877	0.897	0.957	0.945
1	SANOM	1450	0.888	0.844	0.865	0.957	0.948
1	XMap	1413	0.929	0.865	0.896	0.924	0.883
2	AML	233	0.880	0.840	0.860	0.942	0.895
2	Lily	176	0.813	0.586	0.681	0.729	0.547
2	LogMap	235	0.817	0.787	0.802	0.935	0.982
2	LogMapBio	239	0.803	0.787	0.795	0.947	1.007
2	LogMapLite	151	0.987	0.661	0.755	0.637	0.392
2	POMap	161	0.663	0.709	0.685	1.165	0.908
2	XMap	153	0.987	0.619	0.761	0.638	0.415
3	ALOD2Vec	1727	0.941	0.213	0.347	0.182	0.061
3	AML	6988	0.923	0.762	0.835	0.803	0.662
3	DOME	1530	0.988	0.198	0.330	0.160	0.052
3	FCAMapX	7582	0.955	0.815	0.879	0.853	0.657
3	KEPLER	4005	0.822	0.424	0.559	0.450	0.251
3	LogMap	6282	0.947	0.690	0.798	0.718	0.321
3	LogMapBio	6319	0.947	0.693	0.800	0.722	0.332
3	LogMapLite	1642	0.968	0.208	0.342	0.173	0.058
3	POMAP++	2163	0.906	0.260	0.404	0.227	0.081
3	XMap	5815	0.962	0.647	0.774	0.669	0.402
4	ALOD2Vec	2528	0.972	0.839	0.901	0.845	0.748
4	AML	2723	0.958	0.910	0.933	0.912	0.763
4	DOME	2248	0.985	0.764	0.861	0.756	0.503
4	FCAMapX	2828	0.948	0.911	0.929	0.940	0.865
4	KEPLER	2506	0.960	0.831	0.891	0.841	0.881
4	LogMap	2747	0.944	0.897	0.920	0.920	0.747
4	LogMapBio	2776	0.941	0.902	0.921	0.929	0.914
4	LogMapLite	2480	0.967	0.819	0.887	0.829	0.676
4	POMAP++	2414	0.979	0.814	0.889	0.808	0.637
4	XMap	2315	0.977	0.783	0.869	0.778	0.514

TABLE 3. $ICC(3, 1)$ results.

	precision	recall	F-measure
$\hat{\lambda}_D$	-0.179	0.934	0.838
$\hat{\lambda}_C$	-0.159	0.837	0.697

decrease in the value of $ICC(3, 1)$ by only about 0.1 compared to the recall.

C. RESULTS INTERPRETATION

Both measures proved to be consistent with recall and, to a lesser extent, with F-measure. Lack of compliance with precision is not unfounded. As the alignment size increases, the probability of occurrence of incorrect correspondence increases, and thus its precision decreases.

As the proposed methods evaluate each individual correspondence without knowing its correctness, these values will increase with the size of the alignment regardless of whether it is correct. High values of λ_D and λ_C mean that this alignment contains a lot of mappings of classes located deep in the taxonomic hierarchy of integrated ontologies, which are also in close relationships with each other. The ontology integration process is semi-automatic, which means that the generated mappings have to be verified and corrected by specialists anyway. Despite the low precision, this alignment can be more useful because repairing incorrect connections could be more effective and require less effort than manually replenishing missing counterparts.

λ_D and λ_C measures can be used in real-world comparison problems of competing alignment systems when there is no ‘gold standard’. They will be particularly helpful in situations where the compared alignments have a similar size.

For example, alignments of AML and LogMap systems in task 2 have respectively 233 and 235 mappings, while their λ_D ratings are equal to 71.059 and 70.478. This fact entails that despite LogMap creates “bigger” alignments, their quality is lower than alignments designated by AML. A comparison of these values can aid in choosing alignment that will be selected for the next stages: verification, correction, and implementation.

VII. FUTURE WORKS AND SUMMARY

The main contribution of this work is a definition of new methods that allow comparing ontology alignments without the need for reference mapping. It consists of four main Research Goals formulated in Section I. Reaching these goals resulted in definitions of innovative functions for assessing the quality of mappings adapted for two levels of ontology - classes, and instances. The proposed methods are build using a criterion based on the depth of the mapped classes, and a criterion based on the continuity of mapped classes. To formulate them, two types of inaccuracies based on the taxonomic properties of ontologies were also defined: circular inheritance and disjoint class inheritance.

The methods for assessing ontology alignments of classes presented in Section V-A have been implemented using Java programming language. To conduct an experiment, data provided by the Ontology Alignment Evaluation Initiative were used. Obtained results showed compliance of both aforementioned measures with reference to widely used assessments of mapping correctness - completeness and measure F. Thanks to that, it was concluded that the proposed criteria can be used to reliably evaluate the quality of mappings on a concept level without any kind of pre-prepared reference alignment. Verification of methods concerning alignments of instances can be performed in an analogous way.

Presented methods offer a fresh perspective on the assessment of ontology alignments. At this point, it should be emphasized that their purpose is not to assess the correctness, but quality with a focus on the taxonomic architecture of integrated ontologies. Admittedly, defined inconsistencies are taken into account in scoring, but still, many incorrect connections are not detected in this way due to correctness in the context of the ontology structure. However, compliance with recall and F-measure is their additional significant asset. Therefore, it is possible to define useful methods for comparing ontology mapping states that do not require a reference mappings. This can become invaluable in modern applications related to, for example, smart spaces ([17]) or the Internet of Things ([14]). The multiplicity of communication requirements makes it impossible to predict which ontologies will be aligned and, in consequence, to prepare the needed mappings.

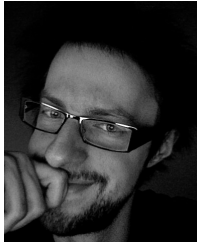
In the nearest future, we plan to formulate inconsistencies resulting from non-compliance with restrictions imposed on the relationships between elements of the ontology. Such an extension of the methods proposed in the following article can have a positive influence on their effectiveness.

Currently, such situations are ignored, which means that the appearance of inconsistent mappings is put on a par with its absence.

REFERENCES

- [1] S. Bowman, “Coordinating SNOMED-CT and ICD-10: Getting the most out of electronic health record systems,” *J. AHIMA*, vol. 76, no. 7, pp. 60–61, Jul./Aug. 2005.
- [2] I. F. Cruz, F. P. Antonelli, C. Stroe, U. C. Keles, and A. Maduko, “Using AgreementMaker to align ontologies for OAEI 2009: Overview, results, and outlook,” in *Proc. 4th Int. Workshop Ontology Matching OM Collocated 8th Int. Semantic Web Conf. ISWC*, 2009, pp. 135–146.
- [3] M. Ehrig and J. Euzenat, “Relaxed precision and recall for ontology matching,” in *Proc. Integrating Ontologies K-CAP Workshop Integrating Ontologies*, vol. 2, Oct. 2005, pp. 25–32.
- [4] J. David and J. Euzenat, “On fixing semantic alignment evaluation measures,” in *Proc. 3rd Int. Workshop Ontology Matching OM Collocated 7th Int. Semantic Web Conf. (ISWC)*, Karlsruhe, Germany, Oct. 2008, pp. 1–13.
- [5] J. Euzenat, “Semantic precision and recall for ontology alignment evaluation,” in *Proc. 20th Int. Joint Conf. Artif. Intell. IJCAI*, Hyderabad, India, Jan. 2007, pp. 348–353.
- [6] J. Euzenat, M. Ch, H. Stuckenschmidt, P. Shvaiko, and C. Trojahn, “Ontology alignment evaluation initiative: Six years of experience,” *J. Data Semantics*, vol. 15, pp. 158–192, Aug. 2011.
- [7] D. Fleischhacker and H. Stuckenschmidt, “Implementing semantic precision and recall,” in *Proc. 4th Int. Workshop Ontology Matching OM Collocated 8th Int. Semantic Web Conf., ISWC*, Oct. 2009, pp. 254–255.
- [8] L. Hollink, M. van Assem, S. Wang, A. Isaac, and C. Schreiber, “Two variations on ontology alignment evaluation: Methodological issues,” in *The Semantic Web: Research and Applications. ESWC (Lecture Notes in Computer Science)*, vol. 5021, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds. Berlin, Germany: Springer, 2008, doi: [10.1007/978-3-540-68234-9_30](https://doi.org/10.1007/978-3-540-68234-9_30).
- [9] M. Koutlaki, N. Preda, and D. Vodislav, “Online relation alignment for linked datasets,” in *Proc. Eur. Semantic Web Conf.*, Cham, Switzerland: Springer, 2017, pp. 152–168.
- [10] A. Kozierekiewicz and M. Pietranik, “Triggering ontology alignment revaluation based on the degree of change significance on the ontology concept level,” in *Proc. Int. Conf. Bus. Inf. Syst.*, Cham, Switzerland: Springer, 2019, pp. 137–148.
- [11] H. Li, Z. Dragisic, D. Faria, V. Ivanova, E. Jiménez-Ruiz, P. Lambrix, and C. Pesquita, “User validation in ontology alignment: Functional assessment and impact,” *Knowl. Eng. Rev.*, vol. 34, p. E15, 2019, doi: [10.1017/S0269888919000080](https://doi.org/10.1017/S0269888919000080).
- [12] D. Manning Ch, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [13] C. Meilicke and H. Stuckenschmidt, “Incoherence as a basis for measuring the quality of ontology mappings,” in *Proc. 3rd Int. Workshop Ontology Matching (OM) Collocated, Iswc*, 2008, pp. 1–12.
- [14] S. D. Nagawah, H. B. Sta, and B. A. Gobin-Rahimbux, “Towards achieving semantic interoperability in an IoT-enabled smart campus,” in *Proc. IEEE Int. Smart Cities Conf. (ISC)*, Oct. 2019, pp. 593–598, doi: [10.1109/ISC246665.2019.9071694](https://doi.org/10.1109/ISC246665.2019.9071694).
- [15] J. C. D. Reis, C. Pruski, and C. Reynaud-Delaître, “State-of-the-art on mapping maintenance and challenges towards a fully automatic approach,” *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1465–1478, Feb. 2015.
- [16] J. M. Rodrigues, S. Schulz, A. Rector, K. Spackman, B. Üstün, C. G. Chute, V. D. Mea, J. Millar, and K. B. Persson, “Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability,” in *Proc. 14th World Congr. Med. Health Inform. (MEDINFO)* (Studies in Health Technology and Informatics), vol. 192, no. 1–2, 1st–2nd ed. Amsterdam, The Netherlands: IOS Press, 2013, pp. 343–346, doi: [10.3233/978-1-61499-289-9-343](https://doi.org/10.3233/978-1-61499-289-9-343).
- [17] A. Smirnov, A. Kashevnik, N. Shilov, S. Balandin, and I. S. O. Boldyrev, “On-the-fly ontology matching in smart spaces: A multi-model approach,” in *Smart Spaces and Next Generation Wired/Wireless Networking* (Lecture Notes in Computer Science), vol. 6294, S. Balandin, R. Dunaytsev, and Y. Koucheryavy, eds. Berlin, Germany: Springer, 2010, pp. 72–83, doi: [10.1007/978-3-642-14891-0_7](https://doi.org/10.1007/978-3-642-14891-0_7).

- [18] A. Solimando, E. Jiménez-Ruiz, and G. Guerrini, "Detecting and correcting conservativity principle violations in ontology-to-ontology mappings," in *Proc. Int. Semantic Web Conf.*, vol. 2, 2014, pp. 1–16.
- [19] A. Solimando, E. Jiménez-Ruiz, and G. Guerrini, "A multi-strategy approach for detecting and correcting conservativity principle violations in ontology alignments," in *Proc. OWL, Exper. Directions Workshop OWLED*, 2014, pp. 13–24.
- [20] *Ontology Alignment Evaluation Initiative*. Accessed: Mar. 9, 2020. [Online]. Available: <http://oaei.ontologymatching.org>



MARCIN PIETRANIK received the M.Sc. and Ph.D. degrees in computer science from the Wrocław University of Science and Technology, in 2008 and 2014, respectively. Since 2016, he has been an Assistant Professor with the Wrocław University of Science and Technology. His scientific interests include knowledge integration and topics related to ontology management (especially ontology evolution and ontology alignment). He has authored or coauthored over 28 articles and co-organized several conferences. He also has practical background and a strong experience in modern web technologies used within medical projects.



ADRIANNA KOZIERKIEWICZ received the M.Sc. degree in computer science, in 2007, and the Ph.D. degree, in 2011. She is currently an Assistant Professor with Wrocław University of Science and Technology. Her research interests include ontologies, knowledge management, recommendation systems, consensus theory, and other artificial intelligence applications. She is an author or coauthor of over 40 publications in prestigious journals and proceedings of international conferences. She is also an Editor of a book titled "*Modern approaches for intelligent information and database systems*." She has also been the chair of several conferences, a member of program committee of many conferences, and a member of review boards of some journals. She has been awarded with a few scholarships and grants, for her research projects.



MATEUSZ WESOŁOWSKI received the M.Sc. degree in computer science, in 2019. His day job focuses on distributed systems used within public transportation oriented projects. His scientific interests include data integration and knowledge management.

...