Ontology ABox Comparison

Jan Martin Keil[(✉)]

Heinz Nixdorf Chair for Distributed Information Systems,
Institute for Computer Science, Friedrich Schiller University Jena, Jena, Germany
`jan-martin.keil@uni-jena.de`

Abstract. Correctness and completeness of facts are relevant criteria for the evaluation and selection of ontologies. They can be determined through the comparison with a gold standard. But a gold standard is usually not available. Hence, we propose to determine these criteria through the comparison of ontology ABoxes. However, the comparison of ontology ABoxes is only rarely addressed in the literature. We aim to identify appropriate methods and criteria for the semi-automatic comparison of the ABox of ontologies. They will be integrated into an ontology comparison framework. This framework will provide novel opportunities to assess and improve the correctness and completeness of ontologies.

Keywords: Ontology ABox · Ontology comparison
Ontology evaluation · Ontology quality · Ontology selection

1 Introduction

Ontologies can provide domain knowledge in various applications. The accuracy of the provided knowledge is important for their faultless operation. Ideally, an ontology has been carefully evaluated before its employment. The correctness and completeness of facts are relevant criteria for the evaluation and selection [1,2]. For example, before using an ontology of measurement units in a data transformation application, the correctness of conversion factors and the wide coverage of units should have been evaluated by the ontology developer or the user. The further rise of knowledge available in ontologies [3] increasingly complicates the determination of correctness and completeness. The manual review of large ontologies is possible only at great expense. Hence, automation is required.

Earlier research provided methods to verify the inferential consistency [4] or the compliance with given constraints [1]. However, an inferential consistent and constraint compliant ontology is not necessarily in line with the real world. Therefore, an additional method is required to complement existing methods for the evaluation of ABox rich ontologies. To verify correctness and completeness, modeled facts must be compared to actual facts. As it is impossible to directly compare the facts with the real world, a trusted data source representing the real world is required. An ontology could be automatically compared with this

---

data source to determine the correctness and completeness of facts. However, the aim of building an ontology is to create such a trusted data source. Therefore, it seems unlikely to having another trusted data source available.

Alternatively, multiple ontologies of the same domain could be compared to each other. Even if they have only a partial overlap, a comparison can provide valuable information about the correctness and completeness. As the number of available ontologies is continuously increasing, the existence of several ontologies of the same domain is becoming more likely. For example, several ontologies provide knowledge about measurement units [5]. To select the most appropriate ontology for the application at hand, all of them have to be analyzed anyway. Thus, the additional effort during the ontology selection is low. However, methods for comparing the facts in ontologies have not previously been investigated.

To fill this gap, we aim to identify appropriate methods and criteria for the semi-automatic comparison of the ABox of ontologies. Moreover, we will provide a framework implementing these methods. In the remainder of this article we will summarize the state of the art in Sect. 2, formulate the problem in Sect. 3, present our approach in Sect. 4, present preliminary results in Sect. 5, and describe the evaluation plan in Sect. 6.

2 State of the Art

The term *comparison* is ambiguously used in the context of ontologies and semantic web technologies. We will use the term to describe (a) the comparison of entire ontologies regarding certain aspects to evaluate or select ontologies. This is in contrast to other notions of the term to describe (b) the comparison of different versions of one ontology to highlight changes [6], (c) the comparison of single entities or sets of entities to calculate recommendations of entities [7–9], or (d) the calculation of the similarity of single or a few entities from different ontologies to match or merge these ontologies [10–12].

Visser et al. [13] compared four law ontologies regarding the dimensions *epistemological adequacy*, *operationality*, and *reusability*. The criteria for adequacy included the *epistemological completeness*. However, they concluded that the assessment of the completeness is difficult due to the lack of a gold standard.

Botke et al. [14] compared two ontologies regarding their expressive power. To achieve this, they tested if the ontologies can be expressed using the other ontology. An ontology is more powerful if it can express the other ontology, but not vice versa.

Xue et al. [10] proposed an algorithm to calculate the similarity of trees, which could be used to compare ontologies. They focused on the application of their algorithm for ontology integration. Beside that, they claimed that it can be used to evaluate ontologies: An ontology is considered more trustworthy, if it is more similar to other domain ontologies.

Blázquez et al. [15] presented an approach to compare three hydrographical ontologies. The criteria mainly focus on the ontology development process and used data sources that could justify trust into an ontology.

Kretz et al. [16] described a system to select, compare and align ontologies. The comparison is based on observational, structural, functional, processing time, usability, and domain relevance criteria. They named some example criteria without further explanation. Most of these criteria have been described in [17]. Further, the system generates additional candidate ontologies by combining two candidate ontologies. The selection of combination pairs is based on the similarity of the ontologies which is calculated using the confidence of matches between contained entities.

Steinberg et al. [18] compared the schemata of seven unit ontologies based on a set of 16 use cases. A metric was developed to measure the suitability of an ontology with respect to those use cases.

Katsumi et al. [19] designed a procedure to compare candidate ontologies for reuse. The comparison is based on competency questions that specify the intended models of an ontology. They state, that it is easier to address errors of omitted models than superfluous models. Therefore, an ontology is preferred for reuse if it omits the fewest intended models and has no superfluous models.

Färber et al. [20] provided an extensive comparison of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. They applied 34 metrics covering eleven data quality dimensions like accuracy, consistency, completeness, and accessibility. This included metrics for *semantic validity of triples*, *population completeness*, and *column completeness*. *Semantic validity of triples* is the proportion of correct triples compared to a gold standard. Due to the lack of a comprehensive gold standard the evaluation of the semantic validity of triples remains "very challenging" and it is hard to reveal meaningful differences. Therefore they proposed to use test cases [1] to check the semantic validity of triples, although this would not detect all wrong values. *Population completeness* is the coverage of a basic population, which is defined by a gold standard. *Column completeness* is the proportion of the individuals of a class that have a property that is dedicated to this class. Due to properties that do not apply for each individual of the class (e.g., `:hasChild`), they only used a subset of the properties.

The metric for *column completeness* in [20] is similar to the methods presented in [21–23]. Galárraga et al. [21] proposed to use mined rules to assess the completeness of entities. Ahmeti et al. [22] presented a tool to asses the completeness of entities in Wikidata, based on a comparison with similar entities. Similarly, Hitz-Gamper [23] proposed to compare entities of the same class for the detection of missing facts in Wikidata. More general, Razniewski et al. [2] discussed the challenges regarding the completeness of knowledge bases. They proposed to use *mark and recapture* techniques from the field of ecology to estimate the size of a population.

The ABox comparison of ontologies is also related to the field of *data integration*. Naumann et al. [24] described data integration as a process of the three main steps (1) schema matching, (2) duplicate detection, and (3) data fusion. Their work is focused on the integration of databases. Therefore *schema matching* is the process of detecting corresponding tables and attributes. *Duplicate detection* refers to the detection of rows from multiple sources describing the

same real-world entity. *Data fusion* is the step of integrating corresponding rows and resolving contradictions. Methods from this field could be adapted for the usage in the ABox comparison of ontologies.

3 Problem Statement and Contributions

The purpose of our work is to explore appropriate methods and criteria for the semi-automatic comparison of the ABox of ontologies to aid ontology evaluation and selection. The lack of a gold standard to evaluate ontologies has already been identified as a problem in early work on ontology comparison [13]. Still 20 years later this remains an open issue [20]. We aim to contribute on this issue by employing the comparison of the ABox of ontologies.

The ABox of an ontology consists of membership, property, and equality assertions about individuals. The TBox describes available vocabulary by defining relations and constraints of classes and properties. ABox rich ontologies contain many ABox axioms compared to the number of TBox axioms. As the vocabulary used in the ABox is defined in the TBox, the comparison of ABoxes must take place in context of the according TBoxes: The comparison of ABoxes requires a matching of the according TBoxes to identify corresponding assertion axioms. Constraints defined in the TBoxes might imply particular comparison rules. Further, the TBoxes can be used to restrict the ABox comparison on a reasonable set of assertions. It is unlikely to find corresponding property assertions without a corresponding property. Neither can it be expected to find corresponding individuals without a corresponding class.

We assume that the ontologies to compare are given by the user. If the ontologies have been independently developed, it is unlikely that the same accidental error occurred multiple times at the same fact. Therefore we expect that the replacement of a gold standard by competing ontologies will uncover a relevant number of issues. Even if not all wrong facts will be discovered, this will enable ontology authors to correct their ontologies. The results of our preliminary work in the domain of measurement units show promise [5]. Further, the number of errors can be used to assess the quality of an ontology. Ontology users can involve this assessment in their choice of an ontology.

Equally, it is unlikely that independent ontologies contain the same subset of individuals and facts of an unknown population. Therefore we expect that the replacement of a gold standard by competing ontologies will uncover a relevant number of missing individuals and facts. In extreme case, two ontologies do not share any corresponding individuals and facts and therefore, all of them will be treated as missing in the other ontology. Even if not all missing entities and facts will be discovered, this will enable ontology authors to further enrich their ontologies. Further, the number of missing entities and facts can be used to assess the completeness of an ontology. Ontology users can involve this assessment in their choice of an ontology.

Hence we believe that the semi-automatic comparison of ontology ABoxes helps improve these ontologies or to select a suitable ontology:

Hypothesis 1. *Given two different ontologies that are overlapping, ABox rich, and **flawed**: The encountered number of **errors** in these ontologies after limited time using semi-automatic comparison of the ABoxes is greater than the encountered number of **errors** in these ontologies and after the same time using manual review.*

Hypothesis 2. *Given two different ontologies that are overlapping, ABox rich, and **incomplete**: The encountered number of **missing facts** in these ontologies after limited time using semi-automatic comparison of the ABoxes is greater than the encountered number of **missing facts** in these ontologies and after the same time using manual review.*

Hypothesis 3. *Given two different ontologies that are overlapping, ABox rich, and **incomplete**: The encountered number of **missing individuals** in these ontologies after limited time using semi-automatic comparison of the ABoxes is greater than the encountered number of **missing individuals** in these ontologies and after the same time using manual review.*

The comparison of the ABox requires to identify the corresponding facts of the ontologies. However, different ontologies of the same domain might use different approaches to model certain aspects of the domain. For example, there might be (a) properties corresponding to a chain of properties, (b) anonymous individuals corresponding to named individuals, (c) data properties corresponding to annotation properties, or (d) classes corresponding to individuals. This leads to the question:

Research Question 1. *How can different modeling approaches used in the ontologies be (semi-)automatically handled during the comparison of the ABox of ontologies?*

OWL 2 [25] allows the logical description of properties, such as the definition as functional or inverse functional properties. These might imply particular rules on the comparison of the ABox of ontologies. This motivates the question:

Research Question 2. *How can OWL 2 Object Property Axioms, Data Property Axioms, and Key Axioms be (semi-)automatically utilized in the comparison of the ABox of ontologies?*

The comparison of the facts about entities requires to match the corresponding entities. However, the matching relies on analogous facts about the entities. This might reflect the classic 'chicken and egg' problem. Hence, one question is:

Research Question 3. *Are general ontology matching methods sufficient for the comparison of ontology ABoxes to match incorrect entities or are specialized entity resolution methods required?*

The involvement of reasoning in the comparison process might, on the one hand, provide the implicit facts in ontologies for comparison too. On the other hand, high computational effort or inconsistencies could intercept the comparison process. However, the extend of reasoner usage is adjustable. Reasoning could be used for (a) at least each single ontology, to produce the implicit facts, (b) each single ontology and the according mappings to other ontologies, to also produce a schema translation, (c) each pair of ontologies and the according mappings, to also produce implicit facts implied by axioms from both ontologies, or (d) all ontologies and mappings at once, to also produce implicit facts implied by axioms from at least three ontologies. Therefore, a question is:

Research Question 4. *To which extent should reasoning be incorporated into the comparison of the ABox of ontologies?*

The comparison of the ABox of ontologies is supposed to complement other comparison methods. A general interpretation requires the result integration of all methods. Therefore the results of the ontology ABox comparison have to be embedded into an general ontology quality model. Thus, one question is:

Research Question 5. *How can the comparison of the ABox of ontologies be embedded into a general ontology quality model?*

4   Research Methodology and Approach

We propose a framework for the ABox comparison of ontologies. This framework will be implemented to verify our hypotheses. The implementation requires to answer the research questions. Our framework consist of five components, as shown in Fig. 1.
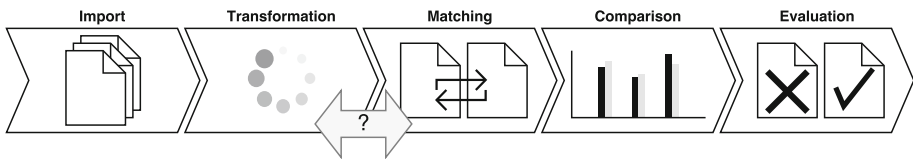
Fig. 1. Schematic of the ABox comparison framework. The order of the transformation component and the matching component is open regarding to Research Question 4.

The *import component* imports ontologies from the web or a local file. Further data sources like SPARQL endpoints, CSV files, or databases are also conceivable, but must be converted into RDF graphs during the import. One data source might have several versions, which in turn might consist of several files. The versions used in the further process can be selected.

The *transformation component* allows to generate additional axioms, to enable the user to employ domain knowledge that is not formalized in the ontology to deduce additional facts. The user could, e.g., provide SPARQL construct

queries to generate axioms. This provides high flexibility, which is important for the applicability in a wide range of domains. Further, reasoning could be employed to deduce facts that are not explicitly formalized in the ontology. However, it is open for evaluation if reasoning should better be involved after the matching of the ontologies. Therefore, the final cooperation and order of these components depends on the answer to Research Question 4.

The *matching component* matches the entities of all data sources. We will employ well known matching libraries using the Alignment API [26]. However, the suitability of general ontology matching methods is open for evaluation regarding to Research Question 3. It might be necessary to implement other, fault-tolerant entity resolution methods. This would employ adapted duplicate detection methods known from the field of data integration. Regardless of the method used, the automated matching could be complemented by user defined mappings and mapping exclusions. The matching of properties is a special issue at this component. The knowledge of the relation of the properties is essential for the further process. However, this might be hampered by different modeling approaches whose mapping goes beyond simple relations. Therefore, a special approach for the property mapping is required, as outlined in Research Question 1.

The *comparison component* provides comparative statistics of the ontologies. It automatically selects relevant classes and properties to compare. This might benefit from an analysis of property definitions, as outlined in Research Question 2. Additionally, classes and properties can be manually (de-)selected. It will compare the number of property assignments and individuals. *Mark and recapture* techniques could be used to estimate the total completeness of each ontology [2]. Finally, it will generate a report containing the comparison results. Regarding to Research Question 5, it remains open how to transfer the statistical values into quality metrics. This is required to embed the comparison into a quality model.

The *evaluation component* highlights deviations between property values of mapped individuals. It will use the classes and properties selected in the comparison component. Methods from the field of data integration could be adapted for this, depending on the aims of the user. If users wants to debug the ontologies, they must explicitly flag conflicting facts as wrong or correct. These flags can be reused for new versions. Therefore, facts do not need to be assessed twice. If users need a quality overview, a completely automated strategy, like a majority voting or a preferred ontology [24], might be sufficient. Wrong facts that have been generated in the transformation component are not a failure of the ontology, but they might hint to wrong source facts. Therefore, the provenance of these facts needs to be provided for assessment. Finally, the component will generate a report with missing, deviating, and wrong property values and missing or duplicated individuals. Regarding to Research Question 5, the evaluation results must be aggregated into meaningful metrics, too.

5   Preliminary Results

We performed a semi-automatized comparison of nine ontologies and knowledge bases in the domain of measurement units [5]. We used a collection of scripts to automatize a majority of the work. While we did not utilize the presented framework, our experiences from that work influenced our framework design. The comparison uncovered a surprisingly low overlap of the ontologies. Further more, we discovered several issues in all analyzed ontologies. The issues have been reported to the ontology authors. For some ontologies, this triggered new releases. The results of this work support the importance of ontology ABox comparison.

Challenges and Lessons learned. Due to the ongoing development of some of the analyzed ontologies, it was necessary on several occasions to reevaluate all ontologies considering a newly published version of one ontology. Even if a majority of the work was automatized, we had to reevaluate possible issues, because our scripts did not take earlier decisions into account. Hence the implementation of our framework should record decisions about potential issues to avoid repeated manual effort.

To improve the comparison of conversion values, we calculated their transitive closure inside each ontology. The localization of the actual causes of errors required to keep the provenance of the calculated values. Hence the implementation of the presented comparison framework should, on the one hand, allow to calculate further facts that have not been formalized in the ontologies, but, on the other hand, also record the provenance of these generated facts.

6   Evaluation Plan

In a first step, we plan to perform a further iteration of the comparison and evaluation of ontologies for measurement units. This is to prove the suitability of the framework in an already investigated context. Several ontologies in this domain are under active development. Therefore, this also provides an opportunity to evaluate the consideration of different versions in the framework. Subsequently, we will perform further ontology comparisons in other domains to prove the general applicability of the framework. Candidate domains are, e.g., species, chemical substances, publications, famous persons, locations, or diseases.

In addition, we intend to test our hypotheses by a series of experiments. Several test persons will inspect two ontologies without using our methods and two other ontologies using our methods. All ontologies will be based on the same set of artificial facts, but with different schemata, omissions and errors. To test the hypotheses we will measure the number of (1) errors, (2) missing facts, (3) missing individuals that a test person detects in a specific time. It is important that it is easy for the test person to validate the facts, but also that it is difficult to automatize the validation. However, the detailed design of this experiments remains open for future work.

7   Conclusions

The presented framework will provide novel opportunities to strengthen the trustworthiness of ontologies. Ontology developers will be able to keep track on the correctness and completeness of facts in their work. The framework will allow them to regularly compare their ontology with their competitors' ontologies and further data sources. Users of ontologies will be empowered to easily compare available ontologies. They will not longer have to blindly trust the represented facts or to perform a tedious manual review of each axiom. The framework will highlight questionable facts. That way, users will be able to take a more educated decision on the selection of ontologies for their projects.

References

1. Kontokostas, D., Westphal, P., Auer, S., et al.: Test-driven evaluation of linked data quality. In: Chung, C., Broder, A.Z., Shim, K., Suel, T. (eds.) 23rd International World Wide Web Conference, WWW 2014, pp. 747–758. ACM (2014). https://doi.org/10.1145/2566486.2568002
2. Razniewski, S., Suchanek, F.M., Nutt, W.: But what do we actually know? In: Pujara, J., Rocktäschel, T., Chen, D., Singh, S. (eds.) Proceedings of the 5th Workshop on Automated Knowledge Base Construction, AKBC@NAACL-HLT 2016, pp. 40–44. The Association for Computer Linguistics (2016)
3. Abele, A., McCrae, J.P., Buitelaar, P., et al.: The Linking Open Data Cloud Diagram (2017)
4. Sirin, E., Parsia, B., Grau, B.C., et al.: Pellet: A practical OWL-DL reasoner. J. Web Sem. **5**(2), 51–53 (2007). https://doi.org/10.1016/j.websem.2007.03.004
5. Keil, J.M., Schindler, S.: Comparison and evaluation of ontologies for units of measurement. In: Semantic Web (2018, in press)
6. Noy, N.F., Musen, M.A.: Using prompt ontology-comparison tools in the EON ontology alignment contest. In: Sure, Y., Corcho, Ó., Euzenat, J., Hughes, T. (eds.) Evaluation of Ontology-based Tools EON 2004 at ISWC 2004, vol. 128. CEUR Workshop Proceedings. CEUR-WS.org (2004)
7. Weinstein, P.C., Birmingham, W.P.: Comparing concepts in differentiated ontologies. In: Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management (KAW 1999) (1999)
8. Giménez-Lugo, G.A., Amandi, A., Simão Sichman, J., Godoy, D.: Enriching information agents' knowledge by ontology comparison: a case study. In: Garijo, F.J., Riquelme, J.C., Toro, M. (eds.) IBERAMIA 2002. LNCS (LNAI), vol. 2527, pp. 546–555. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36131-6_56
9. Bieliková, M., Andrejko, A.: Comparing instances of ontological concepts for personalized recommendation in large information spaces. Comput. Inf. **28**(4), 429–452 (2009)

10. Xue, Y., Wang, C., Ghenniwa, H.H., Shen, W.: A tree similarity measuring method and its application to ontology comparison. J. UCS **15**(9), 1766–1781 (2009). https://doi.org/10.3217/jucs-015-09-1766

11. Maedche, A., Staab, S.: Comparing Ontologies — Similarity Measures and a Comparison Study. Internal Report. 408. Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany, March 2001

12. Wang, J.Z., Ali, F.: An efficient ontology comparison tool for semantic web applications. In: Skowron, A., Agrawal, R., Luck, M., et al. (eds.) IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005) IEEE Computer Society 2005, pp. 372–378. https://doi.org/10.1109/WI.2005.28

13. Visser, P.R.S., Bench-Capon, T.J.M.: A comparison of four ontologies for the design of legal knowledge systems. Artif. Intell. Law **6**(1), 27–57 (1998). https://doi.org/10.1023/A:1008251913710

14. Botke, M., Hoede, C.: A comparison of two ontologies of regions. Memorandum 1511. Department of Applied Mathematics, University of Twente (2000)

15. Vilches-Blázquez, L.M., Ramos, J.A., López-Pellicer, F.J., Corcho, O., Nogueras-Iso, J.: An approach to comparing different ontologies in the context of hydrographical information. In: Popovich, V.V., Claramunt, C., Schrenk, M., Korolenko, K.V. (eds.) Information Fusion and Geographic Information Systems. Lecture Notes in Geoinformation and Cartography, pp. 193–207. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00304-2_13

16. Kretz, D.R., Phillips, W.D., Peoples, B.E., Toennies, J.W.: Method and system for ontology candidate selection, comparison, and alignment. US Patent 8,655,882. 18 February 2014 (2014)

17. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Ontology evaluation and validation. An integrated formal model for the quality diagnostic task. Technical report Laboratory for Applied Ontology, Institute of Cognitive Sciences and Technologies, Italian National Research Council, 6 September 2005 (2005)

18. Steinberg, M.D., Schindler, S., Keil, J.M.: Use cases and suitability metrics for unit ontologies. In: Dragoni, M., Poveda-Villalón, M., Jimenez-Ruiz, E. (eds.) OWLED/ORE -2016. LNCS, vol. 10161, pp. 40–54. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54627-8_4

19. Katsumi, M., Grüninger, M.: Choosing ontologies for reuse. In: Applied Ontology, November 2016, pp. 1–27 (2016). https://doi.org/10.3233/AO-160171

20. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web **9**(1), 77–129 (2018). https://doi.org/10.3233/SW-170275

21. Galárraga, L., Razniewski, S., Amarilli, A., Suchanek, F.M.: Predicting completeness in knowledge bases. In: Rijke, M. de Shokouhi, M., Tomkins, A., Zhang, M. (eds.) Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM 2017), pp. 375–383. ACM (2017). https://doi.org/10.1145/3018661.3018739

22. Ahmeti, A., Razniewski, S., Polleres, A.: Assessing the completeness of entities in knowledge bases. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) ESWC 2017. LNCS, vol. 10577, pp. 7–11. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70407-4_2

23. Hitz-Gamper, B.: What do others say about similar things - predicate comparing for a linked data quality boost. In: Aroyo, L., Gandon, F. (eds.) Proceedings of the Doctoral Consortium at the 16th International Semantic Web Conference (ISWC 2017), vol. 1962. CEUR Workshop Proceedings (2017). CEUR-WS.org

24. Naumann, F., Bilke, A., Bleiholder, J., Weis, M.: Data fusion in three steps: resolving schema, tuple, and value inconsistencies. IEEE Data Eng. Bull. **29**(2), 21–31 (2006)
25. Motik, B., Patel-Schneider, P.F., Parsia, B., eds.: OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax, 2nd edn. 11 December 2012. http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/
26. David, J., Euzenat, J., Scharffe, F., dos Santos, C.T.: The Alignment API 4.0. Semantic Web **2**(1), 3–10 (2011). https://doi.org/10.3233/SW-2011-0028