



# Quasi-Equivalent Concept Trade-Off in Ontology Design: Initial Considerations and Analyses

Vojtěch Svátek<sup>(✉)</sup> , Anna Nesterova, and Viet Bach Nguyen 

Department of Information and Knowledge Engineering,  
Prague University of Economics and Business, Prague, Czech Republic  
{svatek,qnesa01,viet.nguyen}@vse.cz

**Abstract.** The problem of concept equivalence is often addressed within ontology alignment. A similar problem is however encountered in ontology design: the decision whether to express multiple semantically close informal concepts as one or more formal classes, for which we coin the term *concept quasi-equivalence trade-off*. We outline its formal framework as well as an initial set of decision-making criteria. We also tried to collect traces of the trade-off from two sources: the LOV vocabulary catalog and ontology design experts addressed through a questionnaire. Finally, we discuss possible modalities of a software support.

**Keywords:** Ontology design · Concept equivalence · Ontology merging

## 1 Introduction

Concept *merging* is, in semantic web realms, associated with *ontology alignment* [2], which aims to find equivalence or subsumption links between classes from *pre-existing* ontologies. Ontology alignment techniques are also usually executed for the whole ontologies *in bulk*, whether *automatically* or (less often) *interactively*, relying on the matching of entity name strings, structural patterns and instance pools. The main purpose is to achieve interoperability of data (or document) sets described by independently developed ontologies. The existence of such data sets mandates the *soft merging* of classes, whose instance bases become bi- or unidirectionally subsumed but the classes themselves are kept.

A less investigated concept merging scenario can be however identified in the process of designing a *new ontology*. On several occasions, its designer/s may consider pairs (or, generally, n-tuples) of concepts whose semantics is very close, and decide whether to merge them or keep as separate; ‘quasi-equivalent’

---

Supported by the IGA VŠE project № 56/2021, and by the COST Action NexusLinguarum – “European network for Web-centered linguistic data science” (CA18209), supported by COST (European Cooperation in Science and Technology), [www.cost.eu](http://www.cost.eu). We are also grateful to the providers of example cases in Section 4: Jorge Gracia, Fahad Khan, Chris Mungall and Kateřina Haniková.

© The Author(s) 2022

O. Corcho et al. (Eds.): EKAW 2022, LNAI 13514, pp. 209–216, 2022.

[https://doi.org/10.1007/978-3-031-17105-5\\_16](https://doi.org/10.1007/978-3-031-17105-5_16)

concepts may for example be identified by cross-checking verbally expressed *competency questions*. While one or more of these informal concepts may already be expressed by a class in a pre-existing ontology, the goal is not to align existing ontologies but to reach a fine-grained modeling decision for the new ontology. The result of the decision can be not only soft merging (resulting in set-theoretically linked classes), but also a hard merging (a single class, possibly reused from an external ontology), or, on the other hand, the preservation of concepts in the form of separate classes (but, most likely, linked by some non-set-theoretical property). There may be arguments both for the merging and for the separation of the concepts. From now on, we will call this situation as *quasi-equivalent concept (QuEC) trade-off*. We hypothesize that abstracting elements of the rationale used in this trade-off, expressing them as guidelines, and, eventually, transforming to software support, could possibly make the life easier for OE novices.

The short paper aims to serve as an initial exploration of the quasi-equivalent concept trade-off. In Sect. 2 we formulate and exemplify the QuEC trade-off, outline an initial set of criteria that may support its resolution, and hypothesize about the visible signs of such a process in existing ontologies. In Sect. 3 we consequently analyze a collection of ontologies with respect to the presence of links considered as such signs. In Sect. 4 we provide real examples of the QuEC trade-off as provided by ontology design experts through a questionnaire. Finally, in Sect. 5 we discuss possible modalities of a software support for such decision making, and in Sect. 6 we wrap up the paper. More details about the research carried out can be found in a thesis [4].

## 2 Quasi-Equivalent Concept Problem Input/Outcome

The problem can be characterized as follows, in terms of input and outcome:

- Input: informal conceptualization (i.e., the designer’s mental model) of the domain, containing, among other, two<sup>1</sup> input concepts,  $\mathcal{C}_1$  and  $\mathcal{C}_2$ .
- There are two ‘canonical’ variants (with sub-variants) of the modeling process outcome, in terms of the content of the output formal (OWL) ontology  $O$ :
  - (Merging outcome:)  $O$  contains in its signature either
    - \* (Hard merging:) a class  $c$  representing both  $\mathcal{C}_1$  and  $\mathcal{C}_2$
    - \* (Soft merging with equivalence/subsumption:) classes  $c_1, c_2$  such that  $c_1$  represents  $\mathcal{C}_1$ ,  $c_2$  represents  $\mathcal{C}_2$ , and either  $c_1 \equiv c_2$ ,  $c_1 \sqsubseteq c_2$  or  $c_2 \sqsubseteq c_1$  holds in the deductive closure of the ontology
    - \* (Soft merging with overlap:) classes  $c_1, c_2, c$  such that  $c_1$  represents  $\mathcal{C}_1$ ,  $c_2$  represents  $\mathcal{C}_2$ , and both  $c_1 \sqsubseteq c$  and  $c_2 \sqsubseteq c$  hold in the deductive closure of the ontology, whilst  $c_1 \sqcap c_2 \sqsubseteq \emptyset$  does not.
  - (Separation outcome:)  $O$  contains in its signature classes  $c_1, c_2$  such that  $c_1$  represents  $\mathcal{C}_1$ ,  $c_2$  represents  $\mathcal{C}_2$ , and  $c_1 \sqcap c_2 \sqsubseteq \emptyset$  holds in the deductive closure of the ontology; furthermore, there is a (logical or annotation) axiom  $(c_1, p, c_2) \in O$  such that  $p$  is some predicate expressing the ‘relatedness’ of two concepts in other than set-theoretical terms.

<sup>1</sup> Variants for more than two concepts could be derived in a combinatorial manner.

Notably, real-world cases need not fully correspond to such ‘canonical’ structures, for example, in the separation outcome, the disjointness axiom  $c_1 \sqcap c_2 \sqsubseteq \emptyset$  may not be present explicitly. The model also does not explicitly handle the setting with  $C_1$  and/or  $C_2$  already mapped on class/es from existing ontologies. Presumably, such classes would then be reused in the new ontology.

As an example, consider the design of an ontology of academic positions and grades.  $C_1$  could then be the concept of *Professor* as a role associated with a particular position at a university (among other, implying being a head of a group), and  $C_2$  the concept of *Professor* as being a grade recognized nation-wide and entitling, as such, to executing some responsibility by the law, at whatever academic institution. Both concepts however correspond to a person role requiring university education, implying the right to supervise PhD students, etc. A (soft) merging outcome could be, for example, the setting with three classes: *ProfessorByPosition*, *ProfessorByGrade*, and their common superclass *Professor*. A separation outcome, in turn, would be that of the first two classes being merely interconnected by a ‘relatedness’ predicate, for example:

*:ProfessorByPosition skos:closeMatch :ProfessorByGrade*

Various factors may influence the decision of the ontology designers. Among other, *merging* may be supported by the following arguments:

- *M1*: The ontology has to be kept small, for manageability/comprehensibility concerns (this only supports the hard merging).
- *M2*: Merging the concepts allows to keep all respective data instances under the same type, making the management of data easier.

On the other hand, *separation* may be supported by the following arguments:

- *S1*: Few or no plausible axioms could be formulated for the merged concept, while the separate concepts could be axiomatized more richly.
- *S2*: There are stakeholders behind each of the concepts who prefer to see it as separate (this is consistent with soft merging but not with hard merging).

In practical terms, how would the process of resolving the QuEC trade-off be manifested in an ontology – considering we can only access the content of  $O$ , and not the informal concepts  $C_1, C_2$  (which were just in the heads of the ontology engineers) or discussions with stakeholders? Consequently to the above discussion, we can expect that the *merging outcome* would result in: (1) *equivalence* or *subclass* axioms in the ontology; (2) class definitions *poor in axioms*. Since the subclass axioms would most often truly correspond to subordination rather than to quasi-equivalence of the pre-cursor informal concepts, and the scarcity of axioms can also have numerous other reasons, the only sensible sign of *merging* seems to be the presence of *equivalence axioms*. The *separation outcome*, in turn, would result in pairs of classes being declared as *disjoint* but connected by some *linking property* expressing their relatedness.

In all, the possible (but, surely, not fully discriminative) manifestation of the quasi-equivalence tradeoff in the design of an ontology seems to be the presence of a pair of classes directly interconnected by a certain kind of axiom: equivalence, disjointness, or the assertion of a linking property.

### 3 LOV Link Analysis

Referring to the above considerations, we set out on analyzing, quantitatively and qualitatively, the structure of the ontologies indexed by the *Linked Open Vocabularies* (LOV) catalog,<sup>2</sup> starting from the presence of the three kinds of axioms (equivalence, disjointness, linking property). This analysis is still ongoing; some initial results (merely for equivalence and linking properties) follow.

Via a literature review we identified 21 candidate linking properties, of which we shortlisted four well-known ones (their approximate count in LOV ontologies, as of November 2021, is in parentheses): *rdfs:seeAlso* (7000), *owl:sameAs* (5000), *skos:exactMatch* (700) and *skos:closeMatch* (300). *owl:equivalentClass* axioms (among named classes) were even more frequent (14000).

Examples of possible (separation) results of the QuEC tradeoff are:

- *dbo:Annotation*<sup>3</sup> *owl:equivalentClass* *bibo>Note*<sup>4</sup>
- *cwmo:Idea*<sup>5</sup> *rdfs:seeAlso* *skos:Concept*<sup>6</sup>;
- *swrc:PersonalName*<sup>7</sup> *owl:sameAs* *foaf:name*<sup>8</sup>
- *ldr:Agent*<sup>9</sup> *skos:exactMatch* *odrl:Party*<sup>10</sup>

All these correspond to concepts that are declared, at lexical level, as synonyms by respected (e.g., Oxford’s) dictionaries. At the same time, their textual descriptions in the ontologies indicate subtle differences in their features.

### 4 Real-World Cases

We compiled a questionnaire on the QuEC trade-off that we advertised, throughout 2021, via direct mailing (to approx. 50 experts) and a few mailing lists, to the ontology engineering community,<sup>11</sup> yielding three fillings.<sup>12</sup> Additionally, we introduced a fourth case, which arose in an ongoing project related to a SARS-CoV-2 antigen testing knowledge graph, at our institute.

<sup>2</sup> <https://lov.linkeddata.es/dataset/lov/>.

<sup>3</sup> <http://dbpedia.org/ontology/Annotation>.

<sup>4</sup> <http://purl.org/ontology/bibo/Note>.

<sup>5</sup> <http://purl.org/cwmo/#Idea>.

<sup>6</sup> <http://www.w3.org/2004/02/skos/core#Concept>.

<sup>7</sup> <http://sparql.cwrc.ca/ontologies/cwrc#PersonalName>.

<sup>8</sup> <http://xmlns.com/foaf/0.1/name>.

<sup>9</sup> <http://purl.oclc.org/NET/ldr/ns#Agent>.

<sup>10</sup> <http://www.w3.org/ns/odrl/2/Party>.

<sup>11</sup> The questionnaire is still ready for input, at <https://forms.gle/ZBXyFzXwmBC8ymob9>.

<sup>12</sup> The reason for this low response may be the unfamiliarity of the topic under the given framing, in combination with the Covid-19 pandemics.

#### 4.1 Case 1: *Entry* vs. *LexicalEntry* in *OntoLex*

The concept *LexicalEntry*<sup>13</sup> pre-existed in the core module of the *OntoLex* ontology. When the new *lexicog* (for ‘lexicography’) module was being developed, a concept called *Entry*<sup>14</sup> was proposed for it, which considered the position of the entry in a dictionary rather than merely its linguistic features. Although the semantics of the concepts was similar, both were retained (after consultation with experts), in order to provide the ‘lexicographic view’ of the entry for the respective stakeholders while at the same time allowing to only use the core module when the lexicographic view is not essential. The module-internal *describes*<sup>15</sup> property was proposed to express the link from *Entry* to *LexicalEntry*.

#### 4.2 Case 2: *Attestation* in *lemonBib* vs. *Citation* in *CiTO*

In the *lemonBib*<sup>16</sup> ontology it was deemed useful to model the notion of *Attestation*, similar to the notion of *Citation*<sup>17</sup> in the existing CiTO ontology. The two concepts were however identified as pertaining to different levels of description [3]. In lexicography, *attesting* some property of a word means referencing an external text in which this property is manifested by a word occurrence. According to CiTO, a citation is “a conceptual directional link from a citing entity to a cited entity, created by a human performative act of making a citation”. This definition ignores the purpose of citing, which was, however, crucial for *lemonBib*; for example, a citation may refer to a word occurrence in order to attest a particular one of its senses, or its rhetorical role, which each correspond to a different attestation target (while the citation target remains the same). Therefore, the entities were kept as separate. To capture their interrelationship, a custom linking property *attestationCitation*<sup>18</sup> was used to connect their instances.

#### 4.3 Case 3: *Fanconi Anemia* in *Mondo Disease Ontology*

Mondo Disease Ontology has been semi-automatically merged from multiple disease resources. One of the merged concepts is that of *Fanconi anemia*,<sup>19</sup> a hereditary DNA repair disorder. It had been a sub-concept of numerous concepts in the source models; these concepts mostly address a specific organ/tissue whose development is affected by the disorder, e.g., ‘genetic skin disease’ or ‘congenital limb malformation’. The quasi-equivalence was concluded to be a true equivalence (the same disorder), while the positioning of the merged concept in 11 different branches of the ontology reflects its diverse perceived manifestations.

<sup>13</sup> <http://www.w3.org/ns/lemon/ontolex#LexicalEntry>.

<sup>14</sup> <http://www.w3.org/ns/lemon/lexicog#Entry>.

<sup>15</sup> <http://www.w3.org/ns/lemon/lexicog#describes>.

<sup>16</sup> <http://lari-datasets.ilc.cnr.it/lemonBib>.

<sup>17</sup> <http://purl.org/spar/cito/Citation>.

<sup>18</sup> <http://lari-datasets.ilc.cnr.it/lemonBib#attestationCitation>.

<sup>19</sup> [http://purl.obolibrary.org/obo/MONDO\\_0019391](http://purl.obolibrary.org/obo/MONDO_0019391).

#### 4.4 Case 4: Notions of ‘Evaluation’ in the Antigen Test Ontology

In the context of developing<sup>20</sup> a knowledge graph on various kinds of SARS-CoV-2 antigen tests, a number of concepts are being considered for the ontological schema, some of which have the character of ‘evaluation’ of a test. Some ‘evaluations’ are, essentially, *claims* (on test sensitivity) made by manufacturers based on their proprietary sources. Some ‘evaluations’, in turn, are statements made by independent organizations or bodies, already having the character of *certification*. Furthermore, some of these independent evaluations are accompanied with quantitative results from either in vitro or clinical studies (again, as sensitivity figures), while some other are mere verdicts (passed/failed). Finally, the tests are also ‘evaluated’ with respect to their *listing* within national or EU-level lists. The publishers of the lists however do not perform any study; they merely verify the fulfillment of common criteria through existing studies. For example, a test listed in the EU Common List should reach at least a 90% sensitivity and a 97% specificity,<sup>21</sup> and must have been validated by at least one Member State based on a study providing details on the methodology.

The plethora of trade-offs remains yet unresolved, but the separation of ‘claims’ from ‘certifications’ appears more likely than their unification. On the other hand, the independent evaluations by authorities may deserve a common over-arching class, whether quantitative evidence is present or not. Finally, the notion of ‘list’ should be modeled separately from that of ‘evaluation’, but their instances should be connected via a domain property.

#### 4.5 Comparison of the Cases

Two of the cases (3 and 4) are from the biomedical domain; this is unsurprising given the prominent role of this domain in knowledge/ontology engineering research. The reason why there are also two cases from linguistics/lexicography can be explained by an initial bias in choosing the direct mailing subjects.<sup>22</sup> As regards the criteria used to merging/splitting the quasi-equivalent concepts, apparently, in Cases 2 and 3 it was primarily their semantic ‘essence’ of the concepts; the same will probably hold for the ultimate decision in Case 4. In Case 1 it seems that the semantic difference might have been accommodated within one concept (*lexicog*’s *Entry*), the positioning information only being optional; the assumption of two different stakeholders groups (one requiring the richer version of the concept in *lexicog*, and one being fine with the core *Ontolex*), however lead to separation.

### 5 Software Support Considerations

Starting from the premise that a criterion in the QuEC trade-off is the proportion of axioms in/valid for both quasi-equivalent concepts, the interplay between

<sup>20</sup> Starting from a database source behind the <https://covidtesty.vse.cz/english/> portal.

<sup>21</sup> <https://data.consilium.europa.eu/doc/document/ST-5451-2021-INIT/en/pdf>, p. 8.

<sup>22</sup> Namely, the fact that the research is partially aligned with the Nexus Linguarum COST Action, <https://nexuslinguarum.eu/>.

concepts and their axioms in the ontology will be of interest. This leads us to seeking inspiration from *knowledge elicitation* techniques, such as the *personal construct theory* made popular in the 1980s through the ETS system [1]. During the process of incrementally eliciting entities and their features from the expert, the tool repeatedly asks either about features that are common or discriminate between given entities, or about new entities that differ from given entities (in some feature). With respect to our QuEC challenge, the approach might have to be extended from the level of entities to a two-level system of *concepts and their instances*, and the role of features would be played by *structured axioms* (namely, Tbox and Abox ones) instead of propositional features. The system would elicit axioms common for or distinguishing between the quasi-equivalent concepts, as well as between the instances of those concepts (potentially leading to further concept splitting). A criterion for the merging/separation would be the number/proportion of axioms that could be asserted for the chosen constellation of concepts. The process would have a dual effect: aside the conclusion on merging/separation, the axioms would be elicited.

While in the 1980s the experts were the dominant source of knowledge, in the semantic web era we pay attention to the *reuse of structured knowledge*. In the simplest scenario, this would mean that not all the axioms brought into the analysis would have to be elicited from the user but would rather be picked up from existing ontologies or even inductively learned from knowledge graphs.

Finally, *textual resources* should be consulted. A focused version of concept description learning [5], where the axioms would be specifically sought for the chosen quasi-equivalent concepts (with the user serving as oracle, assigning them to either one or the other), might be applied.

## 6 Conclusions

We have presented the assumption that ontology engineers (frequently, or at least occasionally) encounter the quasi-equivalent concept trade-off, and outlined the principles that may govern the decision making in such cases. The empirical evidence collected from both existing (LOV) ontologies and experts addressed via a questionnaire is so far rather limited. While we also provide initial considerations on what kind of software support could alleviate the described challenge, further empirical research would probably be needed first in order to ascertain the cost/benefit ratio of developing such a support.

## References

1. Boose, J.H.: Personal construct theory and the transfer of human expertise. In: IAAA 1984, pp. 27–33 (1984)
2. Euzenat, J., Shvaiko, P.: *Ontology Matching*, 2nd edn. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-38721-0>
3. Khan, A.F., Boschetti, F.: Towards a representation of citations in linked data lexical resources. In: Euralex 2018 (2018)

4. Nesterova, A.: Management of Quasi-equivalent concepts in ontologies. MSc. thesis. Prague University of Economics and Business (2021). [https://insis.vse.cz/zp/index.pl?podrobnosti\\_zp=75522](https://insis.vse.cz/zp/index.pl?podrobnosti_zp=75522)
5. Petrucci, G.: Learning to learn concept descriptions. Ph.D. thesis, University of Trento, Italy (2018)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

