On the Roles of Competency Questions in Ontology Engineering

C. Maria Keet[1][0000−0002−8281−0853] and Zubeida Casmod Khan[2][0000−0002−1081−9322]

[1] Department of Computer Science, University of Cape Town, South Africa
mkeet@cs.uct.ac.za
[2] Council for Scientific and Industrial Research, Pretoria, South Africa
zdawood@csir.co.za

Abstract. Competency Questions (CQs) are not merely intended for scoping the prospective content of an ontology and as information-seeking queries posed over an ontology, but serve manifold purposes in the ontology engineering processes. This position paper argues that CQs should be viewed as complex acts with underlying motivations beyond just eliciting facts. We explore the concept of questions in general, drawing on philosophical and logical perspectives to highlight the complexity of questions and how they function beyond information seeking. This understanding is applied to CQs for ontologies, revealing how they can serve various purposes in ontology development, such as knowledge acquisition, knowledge organisation, and validation. The paper also introduces the notion of types of CQs for ontology engineering and a first taxonomy of CQ types. Having identified different types of CQs, it may assist research into devising more specific methods and tools to support the development of CQs—be it manual authoring or automating it—and their use at various stages and tasks in ontology engineering, as well as contribute to a notion of quality of a CQ.

Keywords: Competency Question · Ontology Engineering · Ontology Development · Foundational Ontology · Ontology methodology

1 Introduction

Ontology engineering concerns the development and structuring of knowledge in various domains and applications. One of the key tools in the processes is the use of Competency Questions (CQs). These questions are not just queries for obtaining information but serve deeper purposes, playing a role also in defining the scope, requirements, and validation of ontologies, as demonstrated in, among others, [1, 4, 6, 7, 9, 15–17]. A systematic mapping of which CQs are to be used where for which task in which ontology development methodology is still outstanding.

This position paper argues that CQs should be viewed as complex acts with underlying motivations beyond merely obtaining information. This requires a focused analysis on the technical functionality and limitations of CQs within ontologies. By exploring the concept of questions through philosophical and logical perspectives, we aim to highlight the complexity of questions and how they function beyond mere information seeking and scoping of a prospective ontology's content. This understanding is then applied

to CQs, revealing how they can serve various purposes in ontology development, such as in knowledge acquisition, organisation, and validation.

There is an argument that philosophical and logical perspectives on questions might not be directly applicable to the specific context of CQs in ontology engineering. Notably, Sowa states that "the problem of matching language to logic is unsolvable if the two are considered totally different, irreconcilable systems" [14], which might also apply to CQs. This opposing viewpoint also aligns with the work of Bezerra and Santana, who present CQs as a method for evaluating ontologies [7]. In their approach, CQs function primarily as information-seeking queries within the domain. However, we argue that such philosophical and logical perspectives can provide context for the technical functionality and limitations of CQs. Analysing CQs based on these perspectives help to understand how to improve CQ formulation and interpretation, as well as their use, authoring, research into it, and development of tools for CQs. This is demonstrated by the development of a first taxonomy of CQ types, which provides a more informed understanding of their roles and functions.

The remainder of the paper is structured as follows. In Section 2 we examine CQs in ontologies, including a discussion of faulty CQs. Section 3 introduces pertinent philosophical and logical perspectives on questions, which inform clarifying types of CQs for ontologies. The taxonomy and a library of CQs as a direction towards a solution are presented in Section 4. Section 5 discusses and considers future research directions, and Section 6 concludes the paper.

2 Challenges with Determining the Quality of CQs

While the largest published dataset with CQs [11] may assist in understanding CQs, it was also found that not all questions were of exemplar quality [11, 2], which is also an issue when trying to automatically generate them [3, 1]. Therefore it is of use to further the understanding of CQs for ontologies to consider what a 'faulty' or 'bad' CQ is, and whether that would be in the absolute sense or relative to something.

Upfront, and applicable to all types of CQs, first, there are problems that all types of CQs may exhibit: syntax issues. The sentence may be a grammatically incomprehensible or an ambiguous question, or not be a question but a statement appended with a question mark. Second, there may be questions that no ontology will ever be able to answer, like "how do I apply for promotion to full professor?". Accordingly, we critically assessed the dataset of 234 CQs of [11]. Both authors evaluated each entry on grammar and semantic issues (including vagueness, ambiguity, answerability) to earn a Yes or a No, and then discussed to harmonise any differences. The outcome of the first round was 40 and 49 as problematic CQs of which 28 were initially judged differently. Discussion resolved each, mostly resulting in No (n=22), bringing the overall number of problematic questions to 53 (i.e., 23%; see the new ROCQS dataset in Section 4). Analysing those, 17 of the 53 are easily solvable grammar issues (e.g., 'what' versus 'which'), 9 were about 'can I do x'/'how to do x' rather than about content of the ontology, which are thus strictly unanswerable, and the rest had a range of issues, such as asking for the "fastest" software, imprecision with, e.g., "possibly problematic" behaviour, asking "where" to find something, and others.

There also exist questions that Wiśniewski calls "semantically faulty" [19]. He provides the example of "Which natural number is smaller than 0?". The answer is 'none', but that is presumably not intended because 'none' is not a natural number that was asked for in the question. Applying this notion to a CQ for an ontology, then it means that a CQ should not have always the empty set as answer, as a minimum, and maybe also that at least the category of the entity is among the intended one. For instance, if a physical object is expected as answer, to not have abstract objects in the answer. How this may be managed when querying an ontology is yet to be determined, especially regarding specifying upfront the category or upper-level entity or top-domain or domain-level class in the ontology, which may not be known upfront if the ontology is not aligned to a top-level ontology, let alone at the scoping stage of ontology development.

Then there are subtle issues that interfere with the goodness of a CQ, or its usability at least. We identify and discuss three. First, and especially for CQs used for validating the ontology: some questions cannot be converted into SPARQL or SPARQL-OWL (or a similar query language for another ontology language) to query the ontology due to the lack of expressiveness of the query language. One then either has to reformulate the question or check it manually. For instance, negation is not fully supported in SPARQL, and ranking is difficult. Example CQs that are at least ambiguous if not impossible to answer due to query language restrictions are, e.g.: "To what extent does [the software] support appropriate open standards?" due to the gradation inherent in 'to what extent' and the imprecision of what counts as 'appropriate'; "Which is the fastest software to read [this data]?" due to the comparator 'fastest'; and "Is there an animal that does not drink water?" due to the negation's intent. Of course, at least some of them can be tweaked to become answerable, such as "Which animal represented in the ontology is known not to drink water?" to limit it to declared knowledge in the ontology. Observe that this also shows why the set of CQs for scoping the ontology at the start of ontology development may differ from those CQs used to validate the ontology once built: a desired scope may be represented in more or less detail in the ontology eventually, be it due to scope creep or language limitations or rewording of an imprecise question.

Conversely, the query language may be expressive enough, but the ontology language it is used with is not. That is, second, there are CQs that cannot be answered by a particular ontology due to the restriction on the language that the ontology is represented in. For instance, a CQ "Does a narcissist love himself?" concerns the relational property of reflexivity, but if the ontology language prohibits it, it cannot possibly be answered as intended; likewise for a CQ inquiring about reflexivity of love. This does not make it a bad CQ of itself, however, just not a good one for the target ontology; or the CQ is good, but the ontology does not meet the user's requirements for their intended use. Thus, a CQ may be deemed good or bad within a specific context of use.

Third, there are questions that cannot be answered by a particular ontology because it lacks the coverage with respect to the content. For instance, with the "Does a narcissist love himself?" and where the ontology language allows reflexivity, but now it lacks the vocabulary, i.e., there is no Narcissist and/or no love in the ontology. The question is then not answerable for that version of the ontology. It could still be a good CQ. It would contextually turn into a 'not good' one only if the scope had been changed.

3 Philosophical and Logical Perspectives

As the previous section illustrated, all questions are not alike, and, in fact, already drew from the theories of questions, being that of Wiśniewski's notion of semantically faulty questions. Therefore, to arrive at a proposed solution direction, we first take a step back to consider theoretical works that focus on questions in general, which are more varied than CQs specifically for ontologies. Part of those insights are relevant, which we highlight here, where key aspects are italicised; their use will be demonstrated in the second paragraph of Section 4 and first paragraph of Section 5.

Cohen [8] reignited the 'question question' in research for the past century. He argues that questions are more than requests for information, and rather that they are complex entities that play a crucial role in shaping human thought and discourse. *Question complexity* concerns not only the number of variables involved, but also, according to Cohen, that questions may ask more than one thing at a time, that it is impossible to avoid ambiguous phrasing, and that question contain implicit assumptions. Most recently, Watson [18] highlights the limited attention given to questions from the philosophical point of view and argues that questions are an integral part of human life. She conducted a survey on questions and created a living question collection with as aim to construct a definition of questions. She found that defining a question as an interrogative sentence is insufficient and instead one must focus on its *function*, being an "information-seeking act". Ram [12] already refined the information-seeking aspect into underlying *knowledge acquisition* and *knowledge organisation* goals. Watson considers also *motivations* behind asking questions; e.g., information-seeking questions are asked with the motivation to "expose" a colleague for non-performance of duties [18].

Logics and linguistics-based approaches to questions are reviewed by Wiśniewski [19], who covers the history of research on questions, the different approaches and methods used by logicians and linguists, and the formal systems developed for representing questions. It covers various theories that aim to model natural language questions, including: questions as *sets of declaratives* (similarly advocated in [10]), as epistemic imperatives, as interrogative speech acts, as sentential functions, as *inquisitive semantics* etc. Lastly, types of questions (normal, regular, self-rhetorical, and proper) and answers (complete, partial, eliminative and corrective) are presented. Wiśniewski also noted that there is no universally accepted theory of questions yet.

Neither considers types of questions extensively, nor inventarises purposes, or goals for asking the questions. A practical example of a possible distinction between questions is demonstrated by Bertolazzi et al. [5]. Their work, they argue, provides a basis for understanding how questions evolve as cognitive abilities develop. They explore how ChatGPT builds and refines its hypothesis space through asking questions in the Twenty Questions Game. They use so-called *hypothesis-scanning* questions, which explicitly mention one of the candidate items (e.g., "is it a melon?"), and *constraint-seeking* questions, which do not (e.g., "can you eat it?"). Ruggeri and Lambrozo [13] refer to hypothesis-scanning questions are those that narrow down the search space by testing specific hypotheses or potential solutions, whereas constraint-seeking questions eliminate unlikely options by identifying features that many solutions have. From an ontological perspective, the former is to ascertain whether the fact holds and the latter to ascertain whether an entity has the property mentioned in the question. For instance, a

hypothesis-scanning CQ from [20] may be: "'Which software tool created [this data]?" and a constraint-seeking one "Does [this software] provide XML editing?", but extant CQs for ontologies have not yet been examined this way, to which we turn next.

4 Formulating CQs in Ontologies

Despite a recent increase in popularity in adoption of CQs, what they exactly are, or should be, or what 'exemplary' or 'good' CQs should look like, remains unclear, let alone the idea that there may be different types of CQs. The insights gleaned from philosophical and logical perspectives on questions can assist with this. By considering the complexity, function, motivation, and logical structure of questions, we can design CQs that not only seek information but also enhance the overall quality and usage of the ontology. We briefly describe a small taxonomy of types of CQs and a library of CQs in this section.

We first draw from the existing philosophical and logical works regarding questions. Applying the notions of hypothesis-scanning and constraint-seeking questions [13], we noted that these both may apply to ontology CQs; e.g., for a software ontology, the CQ "Who is the subject in the process of programming?", it is constraint-seeking, while for the CQ "Do I need a password to use the software?", it is hypothesis-scanning. Cohen's claim of complex and multi-faceted questions [8] holds for CQs whereby different terminologies are used (scoping vs. validation), each phrase chunk of a CQ may match vocabulary in the ontology, or a CQ may describe a property or characteristic of an entity, such as the 'unfolding in time' for perdurants. Cohen's claim on underlying assumptions ties in with Watson's motivations behind asking questions [18], which apply to CQs in that there are specific motivations, such as elicitating requirements, validating the content of an ontology, and obtaining a foundational ontology alignment. The 'questions as sets of declaratives' alludes to a tight relation between CQs and their formalisation in an ontology, whereas 'questions as inquisitive semantics' points to CQs' use to help with trying to determine the ontological nature of the entity that needs to be represented in the ontology.

Building on our analysis of what motivates asking questions and the components that make them up, we have identified five main types of CQs used in ontology development: scoping competency questions (SCQ), validating competency questions (VCQ), foundational competency questions (FCQ), metaproperty competency questions (MpCQ), and relationship competency questions (RCQ), and we describe them briefly here, alongside a small hierarchy, as shown in Figure 1. For structuring purposes, they are divided into ontological CQs (OCQs) and domain CQs (DCQs), where the former focus on the ontological nature of the entity that is being interrogated and the latter on the entities with respect to the subject domain of the prospective ontology. The other ones are summarised as follows.

- Scoping CQ (SCQ): A question that mentions a domain entity and helps define the scope of an ontology for a specific subject domain. SCQs help establish what the ontology will be about.

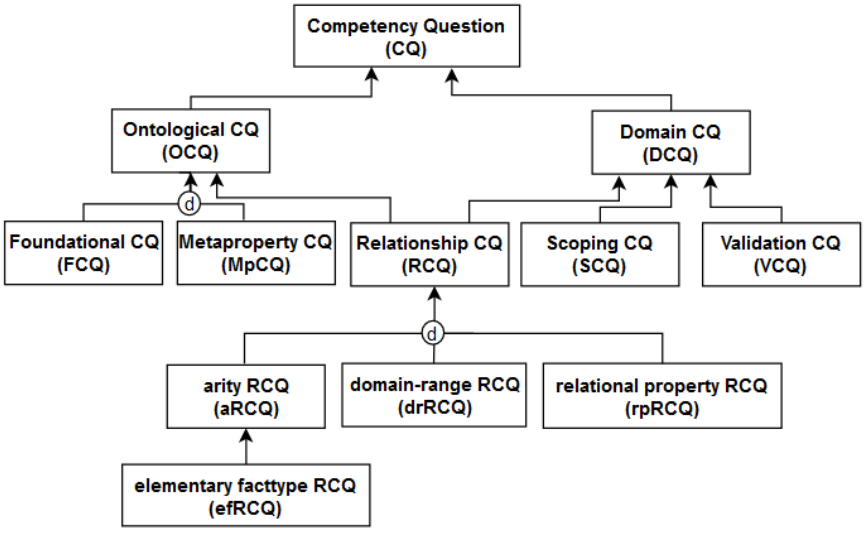

Fig. 1. Overview of the main types of CQs for ontologies, in EER diagram notation.

- Validation CQ (VCQ): A question used to validate the content of an ontology by checking if the ontology adheres to its intended meaning and knowledge representation. VCQs ensure the ontology accurately reflects the domain it represents.
- Foundational CQ (FCQ): A question used to align a domain entity to an entity within a foundational ontology (a more general ontology). FCQs help ensure consistency between the domain ontology and a higher-level ontology.
- Relationship CQ (RCQ): A question that explores various characteristics of relationships within an ontology. There are four sub-types of RCQs:
  - Arity CQ (aRCQ): Determines the number of participants in a relationship.
  - Elementary Fact Type CQ (efRCQ): Asks whether a relationship can be further decomposed without losing information if split up (e.g., a ternary recast as two binary relationships).
  - Domain-Range CQ (drRCQ): Identifies the entities that participate in a relationship (domain and range); these may be either OCQs or DCQs.
  - Relational Property CQ (rpRCQ): Investigates specific properties of a relationship, such as transitivity.
- Metaproperty CQ (MpCQ): A question that classifies an entity according to a predefined set of metaproperties. Metaproperties are general characteristics that hold true across ontologies, such as being a sortal or whether it is telic (has a goal).

To foster further analysis of CQS, as well as their use and reuse at different stages of ontology development, we created a basic Repository of Ontology CQs, ROCQS[3]. It contains 38 FCQs, 33 VCQs, 323 SCQ, 27 RCQs, and 17 MpCQs, 48 of which were newly created by the authors. They are annotated with, including among others and

[3] ROCQS is accessible from http://www.meteck.org/files/ROCQS/ROCQS.htm.

Type of OCQ	Purpose	Examples
Scoping CQ (SCQ)	Define the domain and scope of the ontology	Which predators eat rabbits? What information is clinically relevant for social interaction assessment?
Validation CQ (VCQ)	Verify the accuracy of the content	Is ruby a type of chocolate? What is the Base of ThinAndCrispyPizza?
Foundational CQ (FCQ)	Align entities with a foundational ontology	Is water bottle classified as a Material Entity in the BFO foundational ontology? Is coffee something that cannot be counted, or only in specific quantities?
Relationship CQ (RCQ)	Investigate the characteristics of relationships	What is the domain and range of the eating relationship? If a body contains a heart and a heart contains a cell, does the body contain the cell?
Metaproperty CQ (MpCQ)	Classify entities based on metaproperties	Is each instance of a coffee bean necessarily (at all times of its existence) an instance of a coffee bean? Does a thesis defense have a definite endpoint?

Properties recorded for each type of CQ:
SCQ: ID, SCQ, Ontology, Template (Y/N), Source, Link
VCQ: ID, VCQ,Ontology, Template (Y/N),Source, Link
FCQ: ID, FCQ, Foundational Ontology, Entity/ies described, Template (Y/N), Instantiates Template, Source, Link
RCQ: ID, RCQ,Entity/ies, Template (Y/N), Instantiates template, Source, Link, Sentence pattern
MpCQ: ID, MpCQ, Entity/ies, Template (Y/N), Instantiates template, Source, Link

Fig. 2. Illustrative table in ROCQS and the initial list of properties recorded for each type of CQ (screenshot from the ROCQS repository).

depending on the type of CQ: source, whether they are templates or concrete CQs and if concrete which template they instantiate, which element of the question is the principal one (be it of the FO or the meta- or relational property), which ontology they were created for (if any), and it contains an illustrative table with examples, as illustrated in Fig. 2. The aforementioned SCQs evaluated as problematic (discussed in Section 2) are included in a separate tab and appended with a brief reason why.

5 Discussion

Our analysis of the philosophical and logical works reveals insights into the role of questions, particularly CQs, in ontology development. The insights from the philosophical and logical domains served as a foundation for analysing, identifying, and devising CQs. Recognising questions as complex acts for eliciting information, as in [18, 12], informed by the logical and philosophical theories of questions [19], shed light on how to develop better CQs. For instance, FCQs fit the notion of Ram's [12] "knowledge organisation" goal—where precisely to link it to that ontology—whereas MpCQs and RCQs are information-seeking in the sense of knowledge acquisition, and VCQs align with Knuth's formalisation of Cox's definition of a question as "a system of assertions that answers that question" [10].

Also, by recognising ambiguity and various motivations behind CQs, we were able to elucidate types of CQs and catalog them in a repository for CQs, ROCQS. Both may motivate further research on multiple aspects, such as guiding automated CQ generation, which sentences structures work better than others, demonstrable effects of using questions of each type of CQ, quality metrics of CQs, and methods and techniques for more effective use of CQs in the various ontology development tasks.

Previous work regarding the generation of CQs using LLMs [1] did not include the distinction between hypothesis-scanning and constraint-seeking questions, but it is relevant given the examples provided in Section 3 and the LLM's demonstrated ability to deal with them [5]. Further investigation into such LLM-associated questions may benefit from the separation into different types of CQs and structure the training or output evaluation or filtering accordingly. One also might be able to take existing CQs, convert them to statements, use that to process a corpus in the domain of the prospective ontology, and from found matching sentences, generate new CQs.

We considered creating an ontology of CQs, but since the topic is not sufficiently stabilised yet and rather suggests more research avenues, we deemed it too premature to already develop the artefact. This is, perhaps, also due to, there being no universally accepted theory of questions, as already observed in [19], and possibly how to best structure the types of CQs we identified. Further, especially the collection of FCQs, RCQs, and MpCQs for ontologies would be helpful. To have guidelines and tools for that would assist the endeavour, yet, such research and development will also be assisted by having enough sample CQs to base it off.

Avenues for practical future work include, among others, integration of ROCQS with existing platforms, such as Protégé, to allow users to switch between formulating CQs and examining the relevant parts of the ontology. Another avenue is to conduct user studies with ontology developers to evaluate ROCQS to design it for both usability and CQ collection for analysis.

6 Conclusion

By recognising CQs as complex acts with diverse purposes that go further than information seeking, we can use them more effectively throughout ontology development. This paper explored questions through philosophical and logical perspective which revealed that CQs are used throughout the ontology engineering lifecycle—from scoping and validation to alignment and metaproperty analysis. This was captured in a taxonomy of CQ types that goes beyond previous work, providing a more comprehensive framework for understanding and applying CQs in ontology development. This, in turn, generated various new research and innovation directions to improve CQ quality and use.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alharbi, R., Tamma, V., Grasso, F., Payne, T.R.: An experiment in retrofitting competency questions for existing ontologies. ArXiv **abs/2311.05662** (2023)
2. Antia, M.J., Keet, C.M.: Assessing and enhancing bottom-up CNL design for competency questions for ontologies. In: Proceedings of the Seventh International Workshop on Controlled Natural Language (CNL 2020/21). ACL (2021), https://aclanthology.org/2021.cnl-1.11

3. Antia, M.J., Keet, C.M.: Automating the generation of competency questions for ontologies with agocqs. In: Ortiz-Rodriguez, F., Villazón-Terrazas, B., Tiwari, S., Bobed, C. (eds.) Knowledge Graphs and Semantic Web. pp. 213–227. Springer, Cham (2023)
4. Bernabé, C.H., Keet, C.M., Khan, Z.C., Mahlaza, Z.: A method to improve alignments between domain and foundational ontologies. In: 14th International Conference on Formal Ontology in Information Systems 2023 (FOIS'23). FAIA, vol. 377, pp. 125–139. IOS Press (2023)
5. Bertolazzi, L., Mazzaccara, D., Merlo, F., Bernardi, R.: ChatGPT's Information Seeking Strategy: Insights from the 20-Questions Game. In: Keet, C.M., Lee, H., Zarrieß, S. (eds.) Proceedings of the 16th International Natural Language Generation Conference, INLG 2023, Prague, Czechia, September 11 - 15, 2023. pp. 153–162. Association for Computational Linguistics (2023), https://aclanthology.org/2023.inlg-main.11
6. Bezerra, C., Freitas, F.: Verifying description logic ontologies based on competency questions and unit testing. In: ONTOBRAS'17. pp. 159–164 (2017)
7. Bezerra, C., Freitas, F., Santana, F.: Evaluating ontologies with competency questions. In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 03. pp. 284–285. WI-IAT '13, IEEE Computer Society (2013). https://doi.org/10.1109/WI-IAT.2013.199
8. Cohen, F.S.: What is a question? The monist pp. 350–364 (1929)
9. Keet, C.M., Lawrynowicz, A.: Test-driven development of ontologies. In: Sack, H., et al. (eds.) Proceedings of the 13th Extended Semantic Web Conference (ESWC'16). LNCS, vol. 9678, pp. 642–657. Springer, Berlin (2016), 29 May - 2 June, 2016, Crete, Greece
10. Knuth, K.H.: What is a question? In: AIP Conference Proceedings. vol. 659, pp. 227–242. AIP (2003). https://doi.org/10.1063/1.1570546
11. Potoniec, J., Wisniewski, D., Lawrynowicz, A., Keet, C.M.: Dataset of ontology competency questions to SPARQL-OWL queries translations. Data in Brief **29**, 105098 (2020). https://doi.org/10.1016/j.dib.2019.105098
12. Ram, A.: A theory of questions and question asking. Journal of Learning Sciences **1**(3/4), 273–318 (1991)
13. Ruggeri, A., Lombrozo, T.: Learning by asking: How children ask questions to achieve efficient search. In: Proceedings of the 36th Annual Meeting of the Cognitive Science Society. pp. 1335–1340 (2014), https://escholarship.org/uc/item/32c734r9
14. Sowa, J.F.: The Role of Logic and Ontology in Language and Reasoning, pp. 231–263. Springer Netherlands, Dordrecht (2010)
15. Suarez-Figueroa, M.C., de Cea, G.A., Buil, C., Dellschaft, K., Fernandez-Lopez, M., Garcia, A., Gómez-Pérez, A., Herrero, G., Montiel-Ponsoda, E., Sabou, M., Villazon-Terrazas, B., Yufei, Z.: NeOn methodology for building contextualized ontology networks. NeOn Deliverable D5.4.1, NeOn Project (2008)
16. Thiéblin, E., Haemmerlé, O., Trojahn, C.: Complex matching based on competency questions for alignment: a first sketch. In: 13th International Workshop on Ontology Matching (OM@ISWC 2018). pp. 66–70. CEUR-WS, Monterey, US (2018)
17. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. The Knowledge Engineering Review **11**(2), 93–136 (1996). https://doi.org/10.1017/S0269888900007797
18. Watson, L.: What is a question. Royal Institute of Philosophy Supplement **89**, 273–297 (2021). https://doi.org/10.1017/s1358246121000114
19. Wiśniewski, A.: Semantics of questions. The handbook of contemporary semantic theory pp. 271–313 (2015)
20. Wiśniewski, D., Potoniec, J., Ławrynowicz, A., Keet, C.M.: Analysis of ontology competency questions and their formalizations in SPARQL-OWL. Journal of Web Semantics **59**, 100534 (2019). https://doi.org/10.1016/j.websem.2019.100534