

# Assessing Candidate Ontologies for Reuse

Reham Alharbi<sup>1</sup>

University of Liverpool, Liverpool L69 3BX, UK  
r.alharbi@liverpool.ac.uk

**Abstract.** Ontology reuse is a complex process that requires the support of methodologies and tools to minimise errors and keep the ontologies consistent. Although many efforts have investigated ontology reuse for different tasks and purposes, this body of work does not seem to translate to practice. The goal of this research is a comprehensive “Ontology Reusability Assessment”, that builds on and extends the current state of the art. In this paper, we describe this overall aim and two preliminary results: 1) a community questionnaire to gain an insight on the gap between theory and practice, and 2) a case study to see if it is possible to identify similar functional requirements across different domains.

## 1 Introduction

Ontologies are explicit specifications of conceptualised domains shared by a community of users, and play a significant role in conceptual data modelling and AI [4]. In the context of ontology engineering, reusing existing knowledge models is largely recognised as a key factor in the development of ontologies that are both cost-effective and of higher quality, by reusing components that have previously and independently been validated [14]. Despite this, ontology reuse does not seem a consolidated practice [8]: for instance, an analysis of 377 biomedical ontologies in BioPortal<sup>1</sup> concluded that reuse was very limited (<5%) [9].

From state of the art reviewed, it is clear that some ontologies are reused more than others, possibly because they documented when compared with others [13]. One of the aspects affecting the reusability of an ontology is the perception of its quality with respect to some evaluation criteria. However, although much work has addressed the evaluation of the quality of ontologies from the perspective of reuse, there are no definitive principles and practices, and especially, there is no practical mechanism for providing developers with a qualitative and quantitative assessment of reusability.

The PhD research presented in this paper, currently in its second year, aims to develop a comprehensive framework for assessing ontologies for reuse, capturing the requirements for reusing an ontology, identifying the modalities of reuse, the ontology features, and any additional information that ontology engineers need to be aware of when selecting a candidate ontology to reuse.

To achieve this objective, we analysed state of the art in a top-down and bottom-up manner: we reviewed the current literature on the theory of ontology

<sup>1</sup> <https://www.bioontology.org/>

reuse, whilst we compiled a questionnaire aimed at ontology practitioners to identify mismatches between theory and common practices.

In parallel with the questionnaire, we also started investigating how ontological requirements can help determine whether an ontology can be reused for a given purpose. Competency questions (CQs), the natural language questions outlining the scope of knowledge represented by an ontology and assisting in its development and maintenance [23], are often seen as a way of expressing functional ontological requirements that can be used to promote ontology reuse and that facilitate the identification of patterns that implement such requirements [6]. Our approach aims to contribute a qualitative estimator of the reusability of a given design pattern. We illustrate this through a case study that develops an ontology for the role game Dungeons and Dragons (DnD)<sup>2</sup>. In the case study we objectively assess the similarity between the DnD ontological requirements and those of candidate ontologies to reuse or of design patterns that could be reused.

This paper reports our preliminary results on these two directions of research. We start by defining the problem statement Section 2. In Section 3 we discuss state of the art on how ontology engineers/developers make use of known definitions and measures in reuse. We present the methodology Section 4, the questionnaire and its preliminary results in Section 5, and in Section 6 we present the case study assessing the similarity between ontological requirements. Finally, in Section 7 we conclude and discuss future research directions.

## 2 Problem statement

Ontology reuse is one of the fundamental phases in all ontology design methodologies, and it is deemed as crucial because it allows developing ontologies by saving effort and by reusing ontological fragments that have been independently validated [12]. However, a number of reasons can hamper the effective reuse of existing ontologies:

(i) possible deficiencies in the ontology’s documentation make it difficult to find all the ontologies suitable for reuse [11]; (ii) the lack of a standard way to verify the accuracy of CQs against an ontology typically results in misconceptions in determining the reusability of a candidate ontology [23]; (iii) the lack of standardisation in designing an ontology through reuse can introduce future errors, for example, by failing to keep track of changes in the reused ontologies [8]; (iv) insufficient information about the requirements that ontology engineers aim to satisfy when assessing candidate ontologies for reuse [11,8]. The current PhD research aims to answer the following main research question: “*RQ 1: To what extent and what methods can support ontology developers assess qualitatively or quantitatively the reusability of an ontology?*” This research question can be further decomposed in the following subsidiary questions:

RQ 1.1: Why ontology developers do not reuse existing ontologies?

RQ 1.2: When is an ontology good for reuse from the community point of view?

RQ 1.3: What methods can be used to assess the reusability of an ontology?

<sup>2</sup> A fantasy tabletop role-playing game, <https://dnd.wizards.com/>

RQ 1.4: To what extent identifying similar requirements could indicate the reusability of a target ontology?

RQ 1.5: What are the linguistic features in the requirements that can indicate ontology reusability?

These research questions lead to the definition of the following four objectives that this research aims to address:

- A 1: Identify the reasons that hamper reuse (RQ 1.1)
- A 2: Develop methods for reusing an ontology by identifying: the requirements for reusing an ontology (RQ 1.1), and the ontology features and additional information that ontology engineers need to know to make an informed decision (RQ 1.3, RQ 1.4).
- A 3: Develop mechanisms to assess the reusability of an ontology against the requirements identified and summarise them into an assessment metric (RQ 1.1, RQ 1.2).
- A 4: Evaluate the proposed methodology and metrics with the community to estimate the prospective level of uptake (RQ 1).

From reviewing the extensive state of the art on ontology reuse, we have established that reuse is based on the ability to determine the scope and the purpose for which an ontology is built. Therefore, we start our investigation from the analysis of the role that CQs play to express ontology requirements, and we investigate whether the semantic similarity between CQs can be used as an indicator of reusability of a candidate ontology, which could be used to determine an initial set of reusable ontologies to be selected by ontology developers.

### 3 Related work

In selecting the literature to review to investigate our research questions we also included ontology evaluation papers, as the quality of an ontology is an influential factor in determining whether to reuse an ontology. The literature review provides us with a set of widely used metrics and techniques that determine the ontology quality from different perspectives, even though these efforts are not always explicitly tailored to the task of ontology reuse. Furthermore, whilst several metrics have been proposed to support ontology reuse, few efforts have attempted to assess how ontology reuse theories affect development practices [8], which is the aim of this research.

The assessment method we aim to develop as part of this PhD combines these metrics to understand the thinking behind ontology selection and reuse, with a particular interest in the practical way of assessing the reusability of an ontology. Our assessment metric is based on the combination of four main aspects: schema metrics, instance metrics, community and social metrics, and documentation metrics:

**Schema metrics** [21,19,13]: pertain to the quality of the design of the ontology schema. This can be assessed by indicating the richness, width, depth, and inheritance of an ontology schema.

**Instance Metrics** [21,19,13]: evaluate the way an ontology is populated. The placement and distribution of instance data can indicate the effectiveness of the ontology design and the knowledge represented by the ontology.

**Metrics related to Community and Social aspects** [11,20]: capture the several social and community aspects that can affect the quality of an ontology in the evaluation process.

**Metrics Related to Documentation** [11,7,20]: determine whether an ontology includes some specification of the requirements, for instance the Ontology Requirement Specification Document or CQs.

In addition to identifying these metrics, we also used the literature review to guide us in preparing a questionnaire aimed at ontology practitioners that validates the importance of the identified measures in common practices.

The literature review also includes studies that investigate functional ontological requirements [2] and how these are represented by means of CQs, since these are fundamental in defining the scope of an ontology. For example, an analytical study presented by [16] included a list of 30 methodologies to build ontologies that start by defining CQs.

The purposes of many studies of ontological functional requirements can be categorised into: (i) **ontology evaluation**, by evaluating CQs over an ontology to check whether knowledge is entailed [1,2], or measuring the quality and accuracy of the knowledge encoded in the ontology [15,3]. (ii) **ontology verification**, by comparing an ontology against the ontological requirements to ensure that an ontology is built correctly [18]. (iii) **methodology definition**, by defining a methodology that could consistently drive the ontology engineers in developing an ontology from scratch. This by using a goal modelling approach (Tropos process) to capture, model and reason CQs [18]. (iv) **analysis of CQs**, by analysing the structure of CQs and proposing some popular linguistic patterns that can be reused to specify requirements [10,23,15,3]. (v) **CQs formalisation**, by automatically formulating CQs in a controlled natural language based on CQs patterns [10], and producing a glossary of terms for each domain [22].

All these studies can improve the requirement specification activity by identifying possible problems in the definition of requirements [6]. Our contribution proposes the assessment of the similarity between requirements to decide on the reusability of an ontology.

## 4 Methodology

This section introduces the different methodologies used for investigating the main research question, and its sub-questions.

RQ 1.1<sup>3</sup> has been addressed by a systematic review to determine the challenges to ontology reuse. We have then validated the identified challenges with a qualitative study through an online questionnaire aimed at ontology developers (Section 5), that assesses the extent to which these challenges hamper reuse, and

<sup>3</sup> Why ontology developers do not reuse existing ontologies?

therefore addresses both RQ 1.2 <sup>4</sup> and RQ 1.3 <sup>5</sup>. Based on the survey’s initial analysis, we recognised the importance of ontological requirements, and added two further sub-questions: RQ 1.4 <sup>6</sup> and RQ 1.5 <sup>7</sup>, whose aim investigates the feasibility of (semi-) automatic similarity assessment of requirements. We conduct a case study (Section 6) that develops a Dungeons and Dragons (DnD) ontology covering the different characters and the possible relationships between the resources of the game. The ontology development is divided in three phases, and is described in Section 6: 1) analysis of the DnD requirements by identifying concepts and relationships that are relevant in the domain; 2) implementation of the requirements using CLaRO [10], a template-based Controlled Natural Language resource to author CQs; and 3) identification of similar requirements in the CORAL curated data set [6] of ontological requirements. In order to assess the similarity between requirements and CQs we adopted the BERTSimilarity method [5], that allows us to compute the similarity between complete requirements, taking into account their grammatical structure. Wordnet was also used to account for entity and propriety synonyms.

The overall assessment mechanism(RQ 1) is meant as a summary of the measures for reusing an ontology. It is intended to be a visual graphical summary for all reasons that justify why an ontology can be reused, and thus a visual representation of the ontology reusability with respect to the requirements defined by an ontology developer.

## 5 The reuse questionnaire

The review of state of the art has highlighted the existence of a gap between theory and practice. However, it is not clear what are the barriers to a more extensive uptake of the practice: it is essential, therefore, to go back to the community, and elicit the information directly from ontology developers.

We designed an online questionnaire to understand how ontologists and knowledge engineers in different domains search for, evaluate, and select an ontology for reuse. The questionnaire was disseminated during the period March to November 2020 through personal contacts, and relevant mailing lists and conferences (e.g. the 19th International Semantic Web Conference). In total, 54 eligible respondents participated in this study, but we excluded respondents who classed themselves as having no ontology engineering experience.

The analysis of the answers to the questionnaire is intended to inform how each metric contributes to the assessment mechanism, based on how much the respondents consider a given metric when deciding whether to reuse an ontology.

---

<sup>4</sup> When is an ontology good for reuse from the community point of view?

<sup>5</sup> What methods can be used to assess the reusability of an ontology?

<sup>6</sup> To what extent identifying similar requirements could indicate the reusability of a target ontology?

<sup>7</sup> What are the linguistic features in the requirements that can indicate ontology reusability?

The initial findings can be accessed here <sup>8</sup>. The results are still being analysed; however, it is worth noting that a large proportion of respondents (95%) indicates that they are likely or very likely to consider any ontology documentation, including CQs, when deciding reuse. These results are the basis for the DnD case study, presented in Section 6.

## 6 Case Study: Developing Dungeon and Dragons through reuse

CQs are often used as a starting point in the ontology development process and capture the functional requirements of the ontology. It is, therefore plausible to consider first the reuse of ontologies that have similar functional requirements to the ontology being developed. In order to verify this hypothesis, we devised an experiment where we develop an ontology for the role game Dungeons and Dragons (DnD) <sup>9</sup> using the NeOn methodology [17] and reusing other ontologies from disparate domains do not bear an immediate similarity with the DnD one. The underlying idea is that there could be ontologies that cover different domains from the one of the ontology being designed, but that could still provide terms or patterns that can be reused because their semantics are similar to those in the domain of interest. For instance, the notion of a boss in an organisation can be considered similar to the one of a dungeon master in the game.

Therefore, we aim to assess whether the similarity between requirements can be used as an indicator of the reusability of an ontology. We verify this hypothesis by designing a case study, where we use the Corpus of Ontological Requirements Annotated with Lexico-syntactic patterns, CORAL [6], as a repository of requirements to be matched. CORAL is an openly available corpus of 834 ontological requirements extracted from ontologies modelling different domains, including video games, buildings and terrains and business organisation. We chose to model DnD because its rules and roles in the games are potentially similar to the requirements in CORAL.

We worked with a domain expert and we identified 45 CQs for DnD, which then were analysed through natural language processing to the main terms of interest, using the vocabulary agnostic patterns proposed by [23]. We then tested those terms against CORAL, and we identified five candidate ontologies to reuse (SAREF4ENV, OneM2M, SAREF, SAREF4BLD and OntoDT) [8] with 38 requirements matching the DnD CQs. The remaining 7 CQs from DnD remained unmatched. We then applied question relevance methods, investigated by the question answering community, to detect duplicate CQs and measure the similarity between requirements <sup>10</sup>.

<sup>8</sup> <https://ralharbi9.wixsite.com/reham-alharbi?uid=b3768c75-e033-45c8-88ae-8b6fb426f701>

<sup>9</sup> <https://dnd.wizards.com/articles/features/basicrules>

<sup>10</sup> <https://medium.com/@drcjudelhi/bert-fine-tuning-on-quora-question-pairs-b48277787285>

The underlying hypothesis in this case study is that the similarity scores between requirements can indicate the reusability of the part of the ontology modelling the requirement. For example, in DnD, we ask, "What does a campaign have?". The corresponding requirement in CLARO is "An organisation has users", and we notice that one is a generalised form of the other (that indicate an implicit relationship). We use WordNet (a large lexical database of English) and apply the Wu-Palmer Similarity<sup>11</sup> to assess how similar the word senses of the terms used in the requirements are and the relative position of Synsets in the hypernym tree. This shows that some terms in CLARO are more general than the ones in DnD, and therefore we could reuse the pattern used to define the corresponding CQs in CLARO in DnD.

The ability to start selecting and identifying the relevant ontologies by analysing requirements has a clear potential and is worth exploring further. Applying this methodology to ontology repositories like BioPortal could enable ontology developers to identify ontologies to reuse or to model some of their requirements.

## 7 Conclusion and Future work

We argued that despite the many community efforts, and reuse being a key step in ontology development, the practice is not yet widespread. Therefore this PhD research aims to identify the barriers to a more extensive uptake of the practice and proposes an "Ontology Reusability Assessment" to provide an objective assessment to guide the reuse of an ontology. As part of this effort we sought feedback from the community through an online questionnaire aimed at understanding how ontologists and knowledge engineers select an ontology to reuse.

From an initial analysis of the responses we identified that the ability to analyse and match CQs is a potential good indicator of the reusability of an ontology. Choosing the reuse of a candidate ontology by exploring the similarity between a set of given requirements and those documenting existing ontologies could facilitate and support their reuse. As a next step, we aim to identify linguistic features in the requirements that can be an indication of whether an ontology could be reused. The proposed mechanism for reuse will be then presented back to the community for evaluation. An evaluation study including expert ontologist and community members will ensure that the reuse assessment is both practical and relevant.

## References

1. Bezerra, C., Freitas, F.: Verifying description logic ontologies based on competency questions and unit testing. pp. 159–164 (2017)
2. Bezerra, C., Freitas, F., Santana, F.: Evaluating ontologies with competency questions. vol. 3, pp. 284–285 (2013)

<sup>11</sup> <https://wordnet.princeton.edu/>

3. Bezerra, C., Santana, F., Freitas, F.: Cqchecker: a tool to check ontologies in owl-dl using competency questions written in controlled natural language **12**(2), 4 (2014)
4. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice **31**, 273–318 (2008)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2019)
6. Fernández-Izquierdo, A., Poveda-Villalón, M., García-Castro, R.: Coral: a corpus of ontological requirements annotated with lexico-syntactic patterns. pp. 443–458 (2019)
7. Fernández-López, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: Methodological guidelines for reusing general ontologies **86**, 242–275 (2013)
8. Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A.: Why are ontologies not reused across the same domain? *Journal of Web Semantics* **57**, 100492 (2019)
9. Kamdar, M.R., Tudorache, T., Musen, M.A.: A systematic analysis of term reuse and term overlap across biomedical ontologies. *Semantic web* **8**(6), 853–871 (2017)
10. Keet, C.M., Mahlaza, Z., Antia, M.J.: Claro: a controlled language for authoring competency questions. pp. 3–15 (2019)
11. Matentzoglou, N., Malone, J., Mungall, C., Stevens, R.: Miro: guidelines for minimum information for the reporting of an ontology. *Journal of biomedical semantics* **9**(1), 6 (2018)
12. Noy, N.F., McGuinness, D.L., et al.: *Ontology development 101: A guide to creating your first ontology* (2001)
13. Park, J., Oh, S., Ahn, J.: Ontology selection ranking model for knowledge reuse **38**(5), 5133–5144 (2011)
14. Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A.: Reusing ontology design patterns in a context ontology network. pp. 35–52 (2010)
15. Ren, Y., Parvizi, A., Mellish, C., Pan, J.Z., Van Deemter, K., Stevens, R.: Towards competency question-driven ontology authoring. pp. 752–767 (2014)
16. Soares, A.: *Towards ontology-driven information systems: Guidelines to the creation of new methodologies to build ontologies* (2009)
17. Suárez-Figueroa, M.C.: *NeOn Methodology for building ontology networks: specification, scheduling and reuse*. Ph.D. thesis (2010)
18. Suarez-Figueroa, M.C., Gómez-Pérez, A.: First attempt towards a standard glossary of ontology engineering terminology. pp. 1–16 (2008)
19. Supekar, K., Patel, C., Lee, Y.: Characterizing quality of knowledge on semantic web. pp. 472–478 (2004)
20. Talebpour, M., Sykora, M.D., Jackson, T.: The role of community and social metrics in ontology evaluation: An interview study of ontology reuse. pp. 119–127 (2017)
21. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B.: *Ontoqa: Metric-based ontology quality analysis* (2005)
22. Wisniewski, D., Ławrynowicz, A.: A tagger for glossary of terms extraction from ontology competency questions. pp. 181–185 (2019)
23. Wiśniewski, D., Potoniec, J., Ławrynowicz, A., Keet, C.M.: Analysis of ontology competency questions and their formalizations in sparql-owl **59**, 100534 (2019)