# Question Generation for the Validation of Mapping Adaptation

**Asma Ben Abacha, Julio Cesar Dos Reis, Yassine Mrabet**
Public Research Centre Henri Tudor
29, avenue J.F. Kennedy L-1855, Luxembourg
{asma.benabacha, julio.dosreis, yassine.mrabet}@tudor.lu

**Abstract**

Ontology mappings play a key role for information retrieval and integration in semantically-enabled systems. This requires guaranteeing the validity of correspondences over time, which forces knowledge engineers to adapt them according to ontology evolution. However, validating mapping adaptation remains an open research problem, due to the different types of modifications and to the factors that can trigger them. Existing mapping adaptation approaches rarely involve users. In this article, we propose an approach to support human experts in validating adapted mappings. The proposed 2-steps method starts by a first validation of mappings with Boolean questions. If the expert disapproves the adapted mapping, our approach exploits (i) mapping changes and (ii) the context of concepts in the ontology to propose new alternatives by means of Multiple Choice Questions (MCQ). The achieved results suggest the feasibility of our approach for the validation of mapping adaptation.

## 1 Introduction

Ontology mappings interconnect entities of domain-related ontologies through correspondences (Euzenat and Shvaiko, 2007). These correspondences play a key role in advanced interoperability tasks as they allow information systems to semantically interpret data annotated using different ontologies (Lambrix et al., 2009). Since ontologies continuously evolve, keeping these correspondences up-to-date has become a great challenge. Several techniques are proposed in the literature to semi-automatically adapt the existing mappings according to the changes affecting the ontology entities, and guarantee that they remain semantically valid and complete over time (Dos Reis et al., 2013).

Fully-automated methods are still error-prone due to the complexity of the task. The intervention of domain experts is therefore suited to validate the updated correspondences and to certify their correctness. However, adequate methods for expert validation require relevant techniques to facilitate the human intervention, and to assure a better quality of the resulting mappings. The validation demands a good understanding of the established correspondences as well as an easy access to alternative choices, to correct the automatically built adaptations.

Existing approaches for validation focus on revising and testing mappings issued from ontology alignment methods, based on automatic reasoning and case tests (Serpeloni et al., 2011) (Meilicke et al., 2008). Despite their effectiveness, we need further studies to deal with different versions of mappings in the validation process. Experts demand easier ways to understand correspondences without technical barriers. While ontology construction methods have explored question generation to validate ontological statements (Ben Abacha et al., 2013), generating questions for the validation of mapping adaptation remains an open issue.

Given an ontology evolution scenario where the ontologies evolve to new respective versions, the investigated problem consists in supporting human experts in validating a (semi-)automatic adaptation of the original correspondences. In this paper, we address this issue exploring Natural Language (NL) question generation.

## 2 Validation of Mapping Adaptation

We define a set of concepts of an ontology $O_x$ at time $j$, such that $j \in \mathbb{N}$ as $Concepts(O_x^j) = \{c_1^j, c_2^j, ..., c_n^j\}$. Each concept $c \in Concepts(O_x^j)$ has a unique identifier (it does not change over time) and is associated with a set of attributes $Attributes(c_1^j) = \{a_1, a_2, ..., a_p\}$ (*e.g.*, label, definition, synonym, *etc.*).

An ontology mapping $M^j_{O_S,O_T}$, established at time $j$, interrelates a set of concepts from two different ontologies $O_S$ and $O_T$, respectively, by so-called correspondences:

$$M^j_{O_S,O_T} = \{(c_s^j, c_t^j, semType^j_{st}) | c_s^j \in Concepts(O_S^j), c_t^j \in Concepts(O_T^j), semType^j_{st} \in \{\perp, \equiv, \leq, \geq, \approx\}\}$$

The $semType^j_{st}$ stands for the semantic relation connecting $c_s^j$ and $c_t^j$. We consider the following types of semantic relations: *unmappable* [$\perp$], *equivalent to* [$\equiv$], *more specific than* [$\leq$], *less specific than* [$\geq$] and *partially correspond to* [$\approx$]. For example, "*Nail-Patella Syndrome*" is more specific than [$\leq$] "*Congenital malformation syndromes predominantly involving limbs*".

### 2.1 The Question Generation Approach

Our question generation method uses question patterns, and a selection of candidate ontology entities to construct instances of *Boolean* – or *Multiple Choice* (MCQ) – questions. Then, our approach analyzes the answers provided by human experts to update the mapping adaptations. Figure 1 illustrates the approach overview.

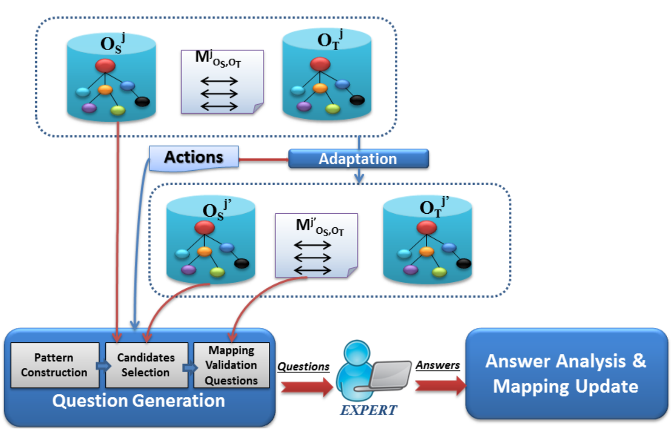


Figure 1: The proposed approach to validate mapping adaptation

### 2.2 STEP 1: Boolean Questions

In a first step, our method translates the proposed adapted correspondences into a NL question using textual patterns associated to each relation type. Let $X$ be the source concept label and $Y$ be the target concept label, the main patterns are as follows:

- (Is—Are) $X$ <equivalent to> $Y$?
- (Is—Are) $X$ <more specific meaning than> $Y$?
- (Is—Are) $X$ <less specific than> $Y$?
- Do(es) $X$ <partially correspond to> $Y$?
- $X$ <cannot be matched with> $Y$?

We present three examples borrowed from biomedical domain:

- *Are intestinal diseases equivalent to vascular disorder of intestines?*
- *Does the Trousseau sign partially correspond to ill-defined and unknown causes of morbidity and mortality?*
- *Is the Eisenmenger Complex more specific than other congenital malformations of cardiac septa?*

### 2.3 STEP 2: Multiple Choice Questions

In the second and main step, negative answers trigger MCQ that are submitted to the expert to detect alternative correct mappings between $c_s^{j'}$ (the source concept) and the concept $c_t^{j'}$ in target ontology $O_T^{j'}$. In our approach, we have three categories of questions:

**1. Revision of** $c_s$. This suggests revising the source concept by candidate proposals from the new source ontology $O_S^{j'}$. This category preserves the semantic mapping-relation between the new source concept $c_s^{j'}$ and the target concept $c_t^{j'}$, and proposes candidate concepts from $O_S^{j'}$ that are semantically close to the initial source concept $c_s^j$. In this MCQ category, we propose questions of the form:
*"What concept <semType> <target concept>?"* corresponds to the revision of the candidate source concept, where
*<semType>* refers to the type of mapping relation $semType^{j'}_{st}$ and

<target concept> consists in the label of the target concept $c_t^j$. For instance: *What concept is more specific than other restrictive cardiomyopathy?* The alternatives stand for the *top n* most semantically-close concepts to the initial source concept (*e.g.*, *Cardiomyopath, Restrictive*). Section 3 presents the selection of alternative candidate concepts.

**2. Revision of mapping relation** $(MR)$. This category of questions proposes revising the type of mapping relation. More precisely, in case of a negative answer in the previous MCQ option, our method preserves the initial concept candidate $c_s^{j'}$ and modifies the $semType$ of the adapted correspondence, selecting another alternative mapping relation (*cf.* Section 3). We propose questions of the form *"Choose the correct mapping relation alternative"*. The proposed alternatives are declarative sentences derived from the question patterns.

**3. Revision of both**. In case of a negative answer in the previous option, our method revises both candidate proposals and semantic relation types, aggregating both option 1 and 2 in a single multiple choice question. We formulate questions of the form *"Choose the correct source concept and relation type"* corresponding to the revision of both the candidate source concept and $semType$ of mapping. We present alternatives for the question generation in 3 columns format, where the first column consists of the list of selected source concept alternatives, the second column presents the list of new suggested types of semantic relations and the third column contains the target concept.

We present two examples in the following:

*1. Is Other spontaneous pneumothorax more specific than closed pneumothorax?*

- *Alternative source concepts:*
  - *Iatrogenic pneumothorax*
  - *Secondary spontaneous pneumothorax*

*2. Is Gastroparesis more specific than Diabetic Gastroparesis associated 2 diabetes mellitus?*

- *Alternative source concepts:*
  - *Acute dilatation of stomach*
  - *Dyspepsia and other specified disorders of function of stomach*

## 3 Selection of alternative concepts and mapping relations

Our approach based on MCQ (*cf.* Section 2.3) requires selecting candidate concepts and different $semType$ relations as suggested alternatives in question generation to support mapping adaptation validation. For this purpose, we propose an algorithm to select similar concepts to the original source concept.

We define the ***context of a concept*** $c_i$ in the ontology. This stands for the union of the sets of *super concepts*, *sub concepts* and *sibling concepts* of $c_i$:

$$CT(c_i^j) = sup(c_i^j) \cup sub(c_i^j) \cup sib(c_i^j) \tag{1}$$

where $c_k^j \in Concepts(O_x^j)$, such that

$$\begin{array}{lcl} sup(c_i^j) & = & \{c_i^j \sqsubset c_k^j \wedge c_i^j \neq c_k^j\} \\ sub(c_i^j) & = & \{c_k^j \sqsubset c_i^j \wedge c_i^j \neq c_k^j\} \\ sib(c_i^j) & = & \{sup(c_k^j) \cap sup(c_i^j) \neq \varnothing \wedge c_i^j \notin sup(c_k^j)\} \end{array} \tag{2}$$

where $c_i^j \sqsubset c_k^j$ means that "$c_i^j$ is a sub concept of $c_k^j$".

In the scope of source concept revision, we generate alternatives in MCQs by using candidate concepts from the context of the initial source concept in the ontology. We aim at combining the answers from these questions to propose re-adapting correspondences if necessary. For example, if a given correspondence between source concept $c_s^j$ and target concept $c_t^j$ is adapted, such that a concept $c_k^{j'} \in Concepts(O_S^{j'})$ replaces the original concept $c_s^j$, this generates an adapted correspondence at time $j'$ between $c_k^{j'}$ and $c_t^{j'} \in Concepts(O_T^{j'})$. Therefore, we retrieve from the ontological context $CT$ a set of other concepts which differs from $c_k^{j'}$, $Candidates = \{(c_{s_i}^{j'}, sim_i) i \in [1..n]\}$, where $c_i^{j'} \in CT(c_s^{j'})$.

Algorithm 1 presents the designed method to retrieve the candidate concepts from the context, given a source concept $c_s^j$ of a mapping. The algorithm sorts the best *top n* candidate concepts from $CT(c_s^{j'})$ using a similarity measure. We use the the *bigram* similarity measure following the observations of (Cheatham and Hitzler, 2013) on its suitability for ontology matching tasks. We compute the similarity between pairs of comparable attributes that are selected beforehand as a parameter (*e.g.*, the name and synonym attributes). We denote the similarity function as $simAtt(a_i^j.value, a_j^{j'}.value)$ between two attribute values $a_i^j.value$ and $a_j^{j'}.value$.

Given all attributes of the original source concept, the algorithm retrieves all concepts in context $CT$ at time $j'$. For all retrieved concepts different from the concept $c_k^{j'}$, to which the adapted mapping is associated, the algorithm selects their attributes and calculates the similarity between the attribute values (between attributes of the source concept and attributes of concepts in $CT$). For each candidate concept, the algorithm keeps the maximal

**Algorithm 1:** Find candidate concepts in context

**Require:** $c_s^j \in Concepts(O_S^j); CT(c_s^{j'}) \subset Concepts(O_S^{j'}); c_k^{j'} \in Concepts(O_S^{j'}); n \in \mathbb{N}$
$Cand \leftarrow \emptyset; maxSim \leftarrow 0;$
**for all** $a_p^j \in Attributes(c_s^j)$ **do**
  **for all** $c_i^{j'} \in CT(c_s^{j'})$ **do**
    **if** $c_i^{j'} \neq c_k^{j'}$ **then**
      **for all** $a_i^{j'} \in Attributes(c_i^{j'})$ **do**
        $s \leftarrow simAtt(a_p^j.value, a_i^{j'}.value);$
        **if** $maxSim < s$ **then**
          $maxSim \leftarrow s;$
        **end if**
      **end for**
      $Cand \leftarrow Cand \cup \{(c_i^{j'}, maxSim)\}; maxSim \leftarrow 0;$
    **end if**
  **end for**
**end for**
**return** $Cand \leftarrow sort(Cand, n);$

similarity value calculated among the attributes. Finally, the algorithm sorts the top $n$ retrieved candidate concepts according to the calculated similarity. We use these candidates as alternative answers in our MCQ approach, so they play a central role for the automatic generation of the questions.

Revising the semantic relation $semType$ demands retrieving alternatives for the second category of proposed MCQ (*cf.* Section 2.3). To this end, we recover a set of semantic relations $Rel_{alternatives} = \{(semType_i) i \in [1..n]\}$ where $semType_i \in \{\equiv, \leq, \geq, \approx\}$ such that $semType_i \neq semType_{st}^{t'}$. We use the $Rel_{alternatives}$ to formulate the question in the revision of *MR* category. The alternative relations are proposed from the most precise one to the more general one (*i.e.* $\equiv$, $\leq$, $\geq$, then $\approx$).

## 4 Experimental evaluation

We evaluate the natural-language quality of the questions. We use two biomedical ontologies SNOMED-CT[1] (SCT) and ICD-9-CM[2] (ICD9) and different versions of official mappings established between them. We aim to investigate to which degree it is possible to generate NL sentences that can adequately describe mappings. For this purpose, we evaluate the generated questions according to three standard measures in NL generation: grammaticality, fluency and meaning preservation. Since our approach aims to facilitate human intervention in mapping adaptation, we assume that it is relevant to assess the natural-language quality of the automatically-generated questions. We presented the generated questions to three different human assessors who were asked to associate a score value between 1 and 10 for each dimension and each question. Assessors were ontology experts and familiar with the biomedical domain. We evaluated the approach for the validation of 20 randomly-selected adapted mappings generated from the evolution of mappings between SCT and ICD9.

Table 1 presents the obtained results for the 20 Boolean questions that are generated for the 20 targeted mappings. We measure the Inter-Assessor Agreement (IAA) for grammaticality, fluency and meaning preservation. IAA corresponds to the average $\kappa$ measure defined in (Cohen, 1960). The $\kappa$ measure indicates how much the assessors' agreement is above the probability of an agreement by chance, and it is commonly used in computational linguistics. In order to have relevant measures, we define 3 score intervals for grammaticality, fluency and meaning preservation which are: [0..3], [4..6], [7..10]. We use these intervals as categories in the calculation of the $\kappa$ measure.

The overall assessment of the generated initial Boolean Questions (*cf.* Section 2.2) indicates good values for grammaticality and fluency because the attained average values are satisfactorily high regarding the used metric. The most important criterion for the mapping validation, which is meaning preservation, had the best score by the assessors. The $\kappa$ *Inter-Assessor Agreement* is also relatively high ($\kappa$ is not negative), which provides a positive test on the reliability of the assessors' ratings.

In the conducted experiments, the analysis of the quality-deficient of Boolean questions produced by the NL generation system led to two main causes: (i) the heterogeneity and length of the literal attributes that led to some inadequacies with the conceived patterns, *e.g.* "*Are other eye disorders more specific than family history degenerative disorder of macula?*" and (ii) errors in the concepts' attributes (mainly labels), *e.g.* " *Is other more specific than mechanical complication of suprapubic catheter?*".

## 5 Conclusion

The continuous evolution of ontologies and the complexity of the mapping adaptation task requires efficient methods to reduce the cost of mapping validation. Relying on domain experts to monitor the validation of automatically adapted correspondences can lead to various benefits, ensuring reliable communication between heterogeneous systems. In this article, we presented a question generation approach implementing automatic methods, to guide domain experts in the validation of adapted ontology mappings. The achieved results underscored the relevance of the approach in the completion of the mapping adaptation process.

[1] `www.nlm.nih.gov/research/umls/licensedcontent/snomedctarchive.html`

[2] `www.cdc.gov/nchs/icd/icd9cm.htm`

| | **Grammaticality** | | | **Fluency** | | | **Meaning** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Min. | Max. | Avg. | Min. | Max. | Avg. | Min. | Max. | Avg. |
| Assessor 1 | 0.4 | 1 | **0.775** | 0.4 | 1 | **0.81** | 0.4 | 1 | **0.915** |
| Assessor 2 | 0.4 | 1 | **0.745** | 0.4 | 1 | **0.77** | 0.4 | 1 | **0.86** |
| Assessor 3 | 0.6 | 0.9 | **0.735** | 0.6 | 0.9 | **0.745** | 0.6 | 0.9 | **0.77** |
| **average** $\kappa$ | 0.28 | | | 0.48 | | | 0.45 | | |

Table 1: Quality of the NL generated questions and average $\kappa$ Inter-Assessor Agreement

## References


[Ben Abacha et al.2013] Asma Ben Abacha, Marcos Da Silveira, and Cédric Pruski. 2013. Medical ontology validation through question answering. In Niels Peek, Roque Marín Morales, and Mor Peleg, editors, *Artificial Intelligence in Medicine*, volume 7885 of *Lecture Notes in Computer Science*, pages 196–205. Springer Berlin Heidelberg.

[Cheatham and Hitzler2013] Michelle Cheatham and Pascal Hitzler. 2013. String similarity metrics for ontology alignment. In Harith Alani, Lalana Kagal, and *et. al.*, editors, *ISWC 2013*, volume 8219 of *LNCS*, pages 294–309. Springer.

[Cohen1960] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

[Dos Reis et al.2013] J. C. Dos Reis, D. Dinh, C. Pruski, M. Da Silveira, and C. Reynaud-Delaitre. 2013. Mapping adaptation actions for the automatic reconciliation of dynamic ontologies. In *Proc. of the ACM International Conference on Information and Knowledge Management (CIKM 2013)*, pages 599–608.

[Euzenat and Shvaiko2007] Jérôme Euzenat and Pavel Shvaiko. 2007. *Ontology matching*. Springer.

[Lambrix et al.2009] Patrick Lambrix, Lena Strömbäch, and He Tan. 2009. Information Integration in Bioinformatics with Ontologies and Standards. In *Semantic Techniques for the Web*, pages 343–376.

[Meilicke et al.2008] Christian Meilicke, Heiner Stuckenschmidt, and Andrei Tamilin. 2008. Reasoning Support for Mapping Revision. *Journal of Logic and Computation*, 19(5).

[Serpeloni et al.2011] F. Serpeloni, R. Moraes, and R. Bonacin. 2011. Ontology mapping validation. *International Journal of Web Portals*, 3:1–11.