Survey on Defining Practices in Ontologies

– Report Summary –

This document reports the results of a survey on defining practices in ontologies conducted in preparation of the International Workshop on Definitions in Ontologies (DO 2013) held on July 7, 2013, in Montreal, in conjunction with the Fourth International Conference on Biomedical Ontologies 2013 (ICBO2013), itself part of the Semantic Trilogy '13 event.

1. Background

Ontologies built using OBO Foundry principles are advised to include both formal (logical) definitions and natural language definitions. Depending on the effort, one or the other can be underrepresented. Possible explanations to this bottleneck are the high cost of producing well-written definitions; an insufficient understanding of the nature of natural language definitions or of logic; the lack of an operational theory of definitions; the lack of studies that evaluate usability and effectiveness of definitions in ontologies; a paucity of tools to help with definition authoring and checking.

Producing natural language definitions is time-consuming, costly and prone to all kinds of inconsistencies. Producing logical definitions that are effective, correct, and communicative is also difficult. It is therefore worth exploring different ways of assisting, with automation, creation and quality control of definitions.

Accordingly, we thought it would be useful to gather interested researchers and developers to reflect upon general themes as the selection and modeling of defining information; the relation between definitions in specific domains as opposed to domain-independent definitions; the theoretical underpinnings of definitions; tools that can facilitate relating logical and natural language definitions. In addition, we wanted to encourage participation by different communities using definitions so that their needs can be exposed.

To address these issues, we organized a half-day workshop aimed at discussing questions, ideas and existing projects regarding definitions in ontologies. The expected outcomes of the workshop were to get an overall view of the needs of the users so as to best orient research on the definition authoring side, as well as to get a diagnosis of the difficulties faced by the latter in order to guide groundwork on definitions and their production.

We present here the results of the survey on defining practices that was conducted in preparation of the general discussion at the workshop.

2. Objectives

The objective was to gather information on the practices and needs of the ontology community with respect to definitions – logical and textual – in order to guide the discussion session aimed at creating a prioritized list of needs and best practices in definitions. We invited the ontology community to give us feedback on their experience by filling in a questionnaire published on the Internet. The web-based survey was sent to several ontology lists; 14 people responded to the questionnaire. The small number of participants does not allow us to draw statistically

significant conclusions; their answers are nevertheless indicative of the practices and needs related to definitions in ontologies.

3. Methodology

The 15 questions – some of which include sub-questions – of the questionnaire can be grouped into three larger categories:

- **User-oriented questions:** types of users; their role in the ontology project on which they are working; their use of logical and/or textual definitions; their training in logical and/or textual definition authoring; the kind of assistance needed with respect to definitions in ontologies. [Q: 1, 2, 8a, 9a-d, 10a-b, 11a-b, 14]

- **Ontology-oriented questions:** inclusion of logical and/or textual definitions in the ontologies. [Q: 3, 4a-c, 5a-b]

- **Definition-oriented questions:** usefulness of logical and textual definitions; major problems in definitions; desired enhancements to textual definitions. [Q: 6, 7, 8b, 12, 13]

Several types of questions were asked: closed yes/no questions; multiple choice questions with single or multiple answers, and open-ended textual (qualitative) questions.

4. Summary of the results

4.1. Users and their needs

4.1.1. Respondents' profile (Q1-2)

Most of the respondents work as ontologists regardless of their primary profession. They are thus more likely to be involved in definition authoring and to express needs related to these activities, which is confirmed by the results to the other questions in this section.

4.1.2. Use of definitions (Q8a-b)

The majority of the respondents report using – consulting or writing – definitions 'often', which is indicative of the fact that definitions are central to the ontology work.

Two types of uses seem to emerge: mostly internal uses related to the activity of ontology development, and, to a lesser extent, external uses related to the application of ontologies.

The answers suggest that respondents are primarily concerned with logical definitions. The lesser use of textual definitions may be due to their lacking quality. These results suggest, in turn, that the roles of the term, the logical definition and the textual one in ontologies could be more precisely defined.

4.1.3. Definition consultation (Q9a-d)

All of the respondents report using logical and/or textual definitions to get a clear understanding of the terms in the ontologies; moreover, the majority of them report using definitions 'often' rather than 'sometimes'.

The use of logical definitions is quantitatively closer to 'very often' than to 'rarely' (7/12 vs. 5/12 respondents). However, the use of textual definitions is even more frequent than that of logical definitions (9/11 respondents).

The frequent use of both logical and textual definitions seems to indicate that they play an important role in the proper understanding of what is represented in the ontologies.

4.1.4. Definition writing (Q10a-b)

The majority of the respondents report engaging in definition authoring activities.

The defining activity is not only limited to definition creation, generally, from texts and consultation of experts; it also includes definition revision and 'translation' of textual definitions to/from logical ones.

As for the defining form, the classical definition structure – genus + differentia – is the preferred one.

4.1.5. Training in definition writing (Q11a-b)

Half of the respondents have had no training in definition writing. Among the other half of the respondents, most have had training in both logical and textual definition writing, and one only in logical definition writing. In only a few cases the training in definition writing was ontology-oriented. It would thus be interesting to create this kind of specific training.

4.1.6. Users' needs (Q14)

The ontology community would mostly welcome general principles for definition writing. Half of the respondents were also interested in ontology-specific training for writing logical definitions. The results also suggest that training and tools related to textual definitions tend to be considered as nice-to-have but not as important as assistance with logical definitions.

4.2. Ontologies and Definitions

4.2.1. Kinds of ontologies (Q3)

Most of the respondents work on ontologies related to the biomedical domain; two work on an upper level ontology, the Basic Formal Ontology. The other ontologies cover varied areas.

4.2.2. Importance of definitions in ontologies

4.2.2.1. Importance of logical definitions (Q4a-c)

The majority of the ontologies on which the respondents answered these questions include logical definitions. However, in most of the ontologies, less than half of the entities are logically defined; only in one are 75-100% of the entities defined.

These results suggest to us that more logical definitions will be added in the future, in particular if the ontology developers want to comply with the OBO Foundry principles. Hence, authoring tools that allow for the semi-automatic creation of logical definitions would probably be helpful.

4.2.2.2.   Importance of textual definitions (Q5a-b)

All the ontologies on which the respondents answered these questions except one have textual definitions. Moreover, by contrast with logical definitions, the textual definitions are well represented: in 10/12 ontologies, more than half of the entities are defined with a textual definition; the coverage rate in 2/3 of all the ontologies is even comprised between 75% and 100% of the entities.

These results tend to indicate that the needs related to textual definitions may be less pronounced than those related to logical definitions.

4.3. Definitions in Ontologies

4.3.1.   Usefulness of definitions (Q6-7)

Both logical and textual definitions are subjectively considered by the respondents as extremely important in ontologies.

4.3.2.   Problems with definitions (Q12)

Four large types of problems were mentioned by the respondents. These are related to:
1.   the information content of textual definitions
    - insufficiently informative
    - too informative/too complex
    - outdated
    - absence of standard defining patterns
2.   logical issues
    - vague
    - circular
    - self-contradictory
3.   the writing and style of the definitions
    - poorly written
    - inconsistent in style
4.   coverage
    - multiple definitions
    - absence of definitions

4.3.3.   Desired enhancements in textual definitions (Q13)

The most frequently mentioned desired enhancements to textual definitions relate (i) to their authoring methods – the creation of definition templates –, and (ii) to their content and form – an increase in the readability of the definitions. The latter enhancement includes not only stylistic matters, but also adaptability of the defining vocabulary to different types of users, which is also related to the adaptability of the defining content. Current user-oriented trends of research in terminology and lexicography could be helpful in this respect.

Among the other mentioned enhancements, we note the development of tools or methods to convert textual definitions to/from logical ones, issues that have started to be explored in the ontology community. Finally, the inclusion of examples is also mentioned, although this enhancement is not as such related to definitions; it may however be indicative of definitions

that are not explicit and content-wise not rich enough to be useful to the users – although in some (or maybe many) cases it might not be related to the lacking of definitions at all, only to the fact that examples tend to fulfill a different cognitive need.

4.3.4. Further comments and suggestions (Q15)

Tools should be developed to help ontology developers implement general principles on definitions.

5. Conclusion

In conclusion, this survey on defining practices in ontologies suggests that definitions are central to ontologies, not only for computational reasons, but also for their proper development and use by humans.

Concerning users' needs, the survey results indicate that it would be valuable to establish ontology-oriented defining principles and manuals, backed up with tools to support ontology developers in implementing the recommendations. Moreover, specific ontology-oriented definition writing training courses or tutorials would also be among the priorities, in particular for logical definitions.

Finally, the current rather low definition coverage rate in ontologies suggests to us that, in light of the standards of good practice in ontology development, more logical, but also textual, definitions will (or, at least, should) be added in the future. Therefore, research efforts could be geared towards developing tools that allow for the (semi-)automatic creation of definitions, for example by generating textual definitions to/from logical ones.

A detailed analysis of the results of the questionnaire is available on the DO 2013 workshop's website: http://definitionsinontologies.weebly.com.