

Document-centric workflow: a framework for ontology development

Aisha Blfgeh^{1,2} and Phillip Lord¹

¹ School of Computing Science, Newcastle University

² Faculty of Computing and Information Technology, King Abdulaziz University

Ontologies are widely used in the field of health-care and life sciences, because of their capability to adapt with such a complex and heterogeneous information. Therefore, the process of ontology construction is important especially with such complex and potentially changeable knowledge; also cooperation between domain specialists and ontology developers is essential, which will improve the reliability and robustness of the ontology overtime.

However, tools used by ontology developers are less understood by the domain specialists who use different formalisms and tools to represent and manipulate their data. As developing ontologies is a collaborative process between different specialities with a variety of tools, it is challenging the integration of the different mindsets in dealing with knowledge in order to develop robust ontologies. Therefore, incorporating those tools into one framework to develop ontologies is a promising mechanism to interact directly with specialists through their preferred tools. In this research, we are proposing a document-centric workflow centred around the use of English and standard office tooling which should ease the interaction between a domain specialist and an ontologist. There are two popular tools that most biologists tend to use in managing their data; Excel spreadsheet and Word document. Hence, the proposed work which has been done so far is facilitating the use of these two applications. This workflow using Word documents is illustrated in Figure 1.

First, the framework is started by the biologists creating an Excel spreadsheet and formatting their sheet as they preferred using all Excel facilities. This spreadsheet is then digested into an ontology by a developer using a programming textual environment of Tawny-OWL [1]. Any modifications to the spreadsheet is possible as it remains as a part of the ontology's source code. The second framework is started when ontologists generate a literate ontology using Lentic package for documentation³ which can be transformed into Word documents, which biologists can then read and edit using Microsoft Word facilities. The biologists are not required to understand any ontological statements or scripts as the source code is embedded within the text and we have the choice to show or hide it. Also, the ontology developers can incorporate all changes to the text and ontological statements as needed. In this case, an arbitrary and frequently transformation of documents will be performed using Tawny-OWL, as well as

³ <http://homepages.cs.ncl.ac.uk/phillip.lord/lentic/lenticular.html>

maintaining the original files as part of the ontology source code. Similar approaches such as CLCE (Common Logic Controlled English) ⁴ and Intermediate representation [2] also facilitate using English language to increase readability of the source code. Nonetheless, we produce more readable and also editable document which can be commented and updated in a normal and more familiar environment for the biologists and users but still need ontologists to recast any changes in the source code according to the updated document.

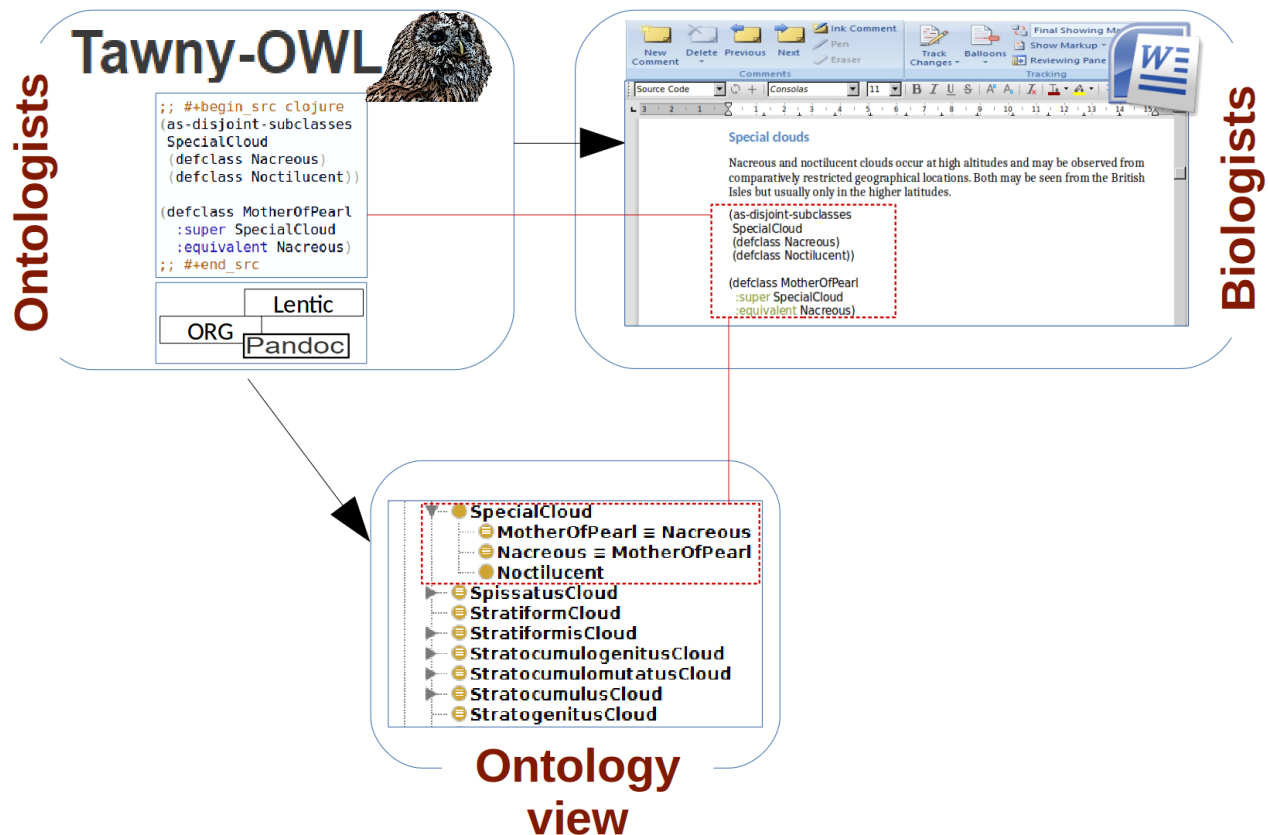


Fig. 1. Workflow for building ontologies using Word document

In this study, we designed a framework for developing ontologies which is centred around the use of Excel spreadsheet and Word document to cooperate with non-ontological specialists who are satisfied to interact with text environment rather

⁴ <http://www.jfsowa.com/clce/specs.htm>

than a complex ontology development environment. We can successfully generate Word versions of a literate ontology.

We are now moving toward the testing phase; one aim of this poster is to approach life sciences community to help in appraising our methodology of constructing ontologies. In this early stage of the study, we would like to encourage interested SWAT4LS' participants (biologists and developers) to test and evaluate our proposed framework. This engagement will be beneficial to improve the interaction activities between different specialities' users.

References

1. Lord, P.: The Semantic Web takes Wing: Programming Ontologies with Tawny-OWL. OWLED 2013 (March 2013)
2. Rector, A.L., Zanstra, P.E., Solomon, W.D., Rogers, J.E., Baud, R., Ceusters, W., Claassen, W., Kirby, J., Rodrigues, J.M., Mori, A.R., der Haring, E.J.V., Wagner, J.: Reconciling users' needs and formal requirements: issues in developing a reusable ontology for medicine. *IEEE Transactions on Information Technology in Biomedicine* **2**(4) (Dec 1998) 229–242