GOSPL: Grounding Ontologies with Social Processes and Natural Language

Christophe Debruyne, Quentin Reul, Robert Meersman
STARLab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Introduction

The GOSPL (Grounding Ontologies with Social Processes and Natural Language) wiki aims to enable communities to develop and maintain a representation of their world. This is essential for facilitating the uptake of the Linked Open Data (LOD) [1] initiative, which aims to annotate and expose dataset on the World Wide Web. Thus, the community as a whole needs to reach an agreement on the meaning of annotations associated with the (legacy) data. However, the LOD initiative relies on RDF(S) [2] and URI mechanisms to represent these annotations, which is not IT laymen friendly.

As a result, the GOSPL application is based on the DOGMA framework [3], which is an ontology engineering approach grounded in natural language. Furthermore, GOSPL relies on Social Web technologies to allow every member of a community to express their own knowledge, and to support provenance by tracking changes to the common agreement on this knowledge.

Grounding in Natural Language

RDF(S) provides constructs for naming and accessing resources on the Web by both machines and humans. These constructs use URIs to achieve this, but URIs are not always the easiest thing for humans to understand even though constructs such as **rdfs:label** and **rdfs:comment** are available, but are not imposed. IT laymen may be reluctant to learn a new paradigm and feel more comfortable expressing knowledge in natural language.

DOGMA has its groundings in natural language by storing its conceptualizations as lexons [3]. Lexons are formally described as a 5-tuple ⟨*γ, head term, role, co-role, tail term*⟩, where γ is an abstract context identifier used to group lexons that are logically related to each other, *head term* plays *role* on *tail term* and *tail term* plays *co-role* on *head term*. GOSPL uses the same lexons as the elementary building blocks for constructing ontologies.

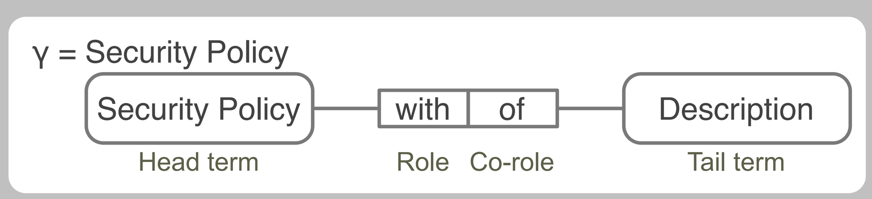

Figure 1: Example lexon

Visualization

Compared to existing wiki-based ontology engineering tools, GOSPL provides a graphical modeling language to represent concepts and the relations between them in Natural Language using NORM-Trees [4], which are undirected trees constructed by concatenating lexons. We make a distinction between the formal model and natural language description of the model for both viewing and editing the model.

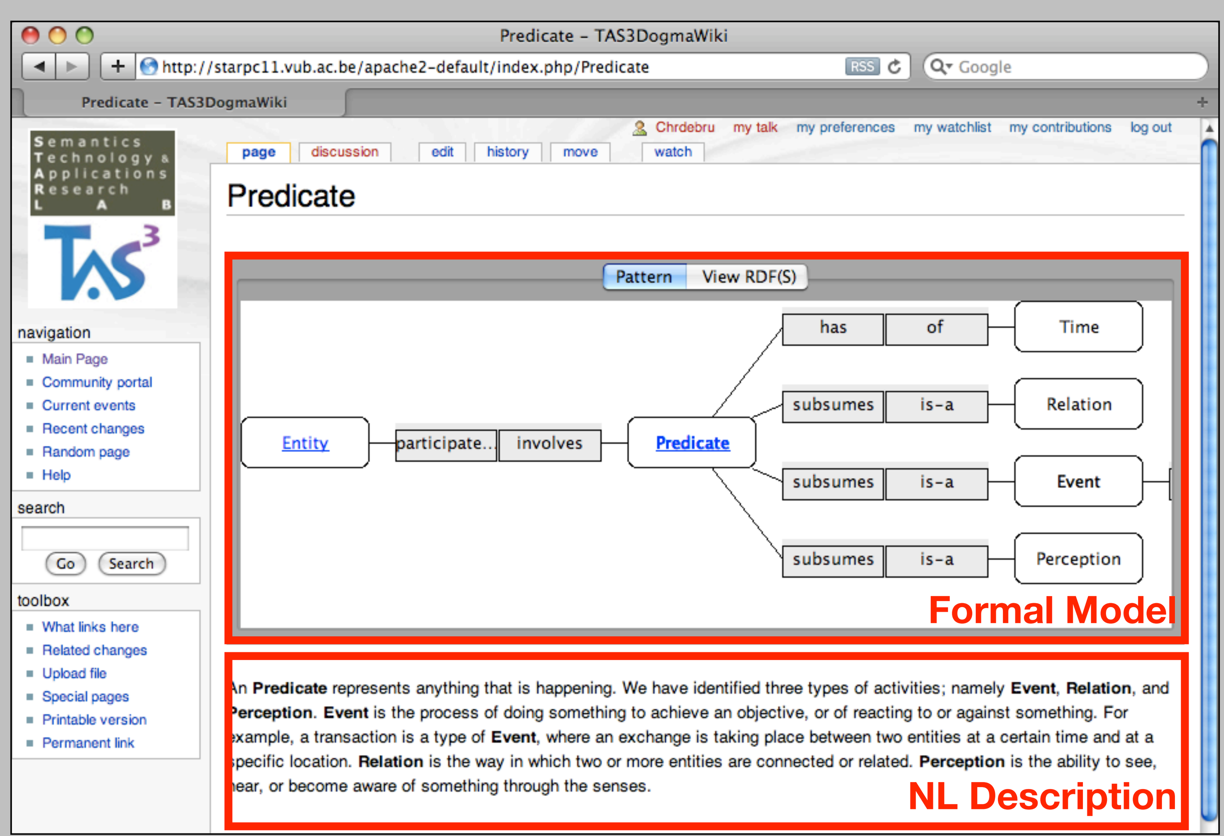

Figure 2: Viewing a concept and its lexons

Each page represents a concept and contains its relations to other concepts. These concepts can, in turn, have their own page and can be easily accessed via a hyperlink. The lexons are stored on the DOGMA server [3] and manipulated through the wiki. These lexons can then be consulted through other tools such as DOGMA Studio Workbench [5].

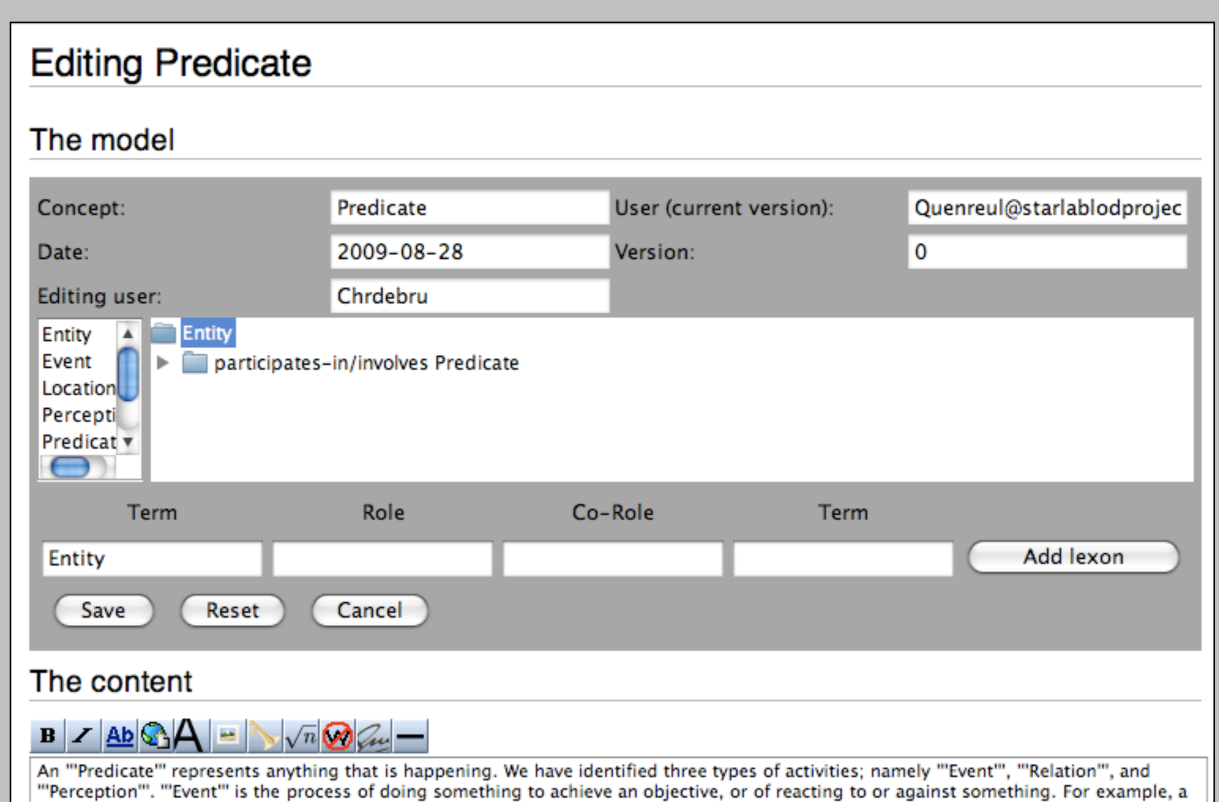

Figure 3: Editing a concept

Provenance

In GOSPL, we track changes on two levels: one for the model and a second for the textual content. Separating the evolution of the formal and the non-formal content has as an advantage that one can apply heuristics to the history of the formal model (e.g., set-operations on the facts of two different versions). We use the time stamp to analyze the relationship between the two evolutionary processes of a semantic object.

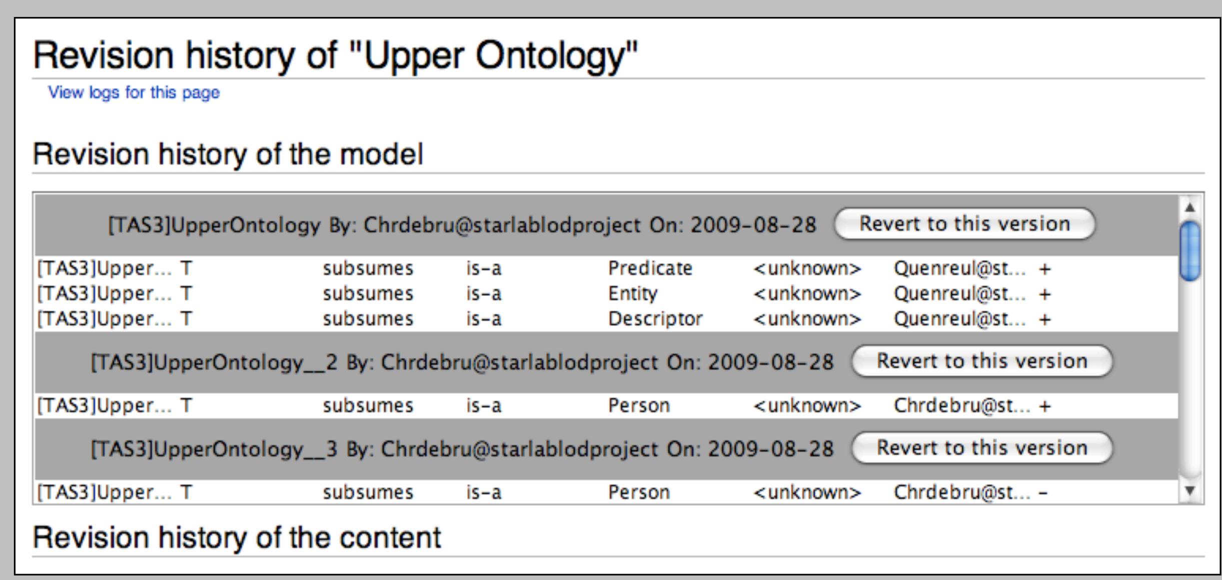

Figure 4: Revision history of a concept and its lexons

Transforming Lexons into RDF(S)

Lexons are easily transformed into RDF(S) by generating classes, properties and their respective hierarchies from the lexons. The advantage of the tool is that the agreement is reached based on natural language and that this shared agreement is then transformed into RDF(S) to annotate datasets, enabling a community to publish their data on the Web.

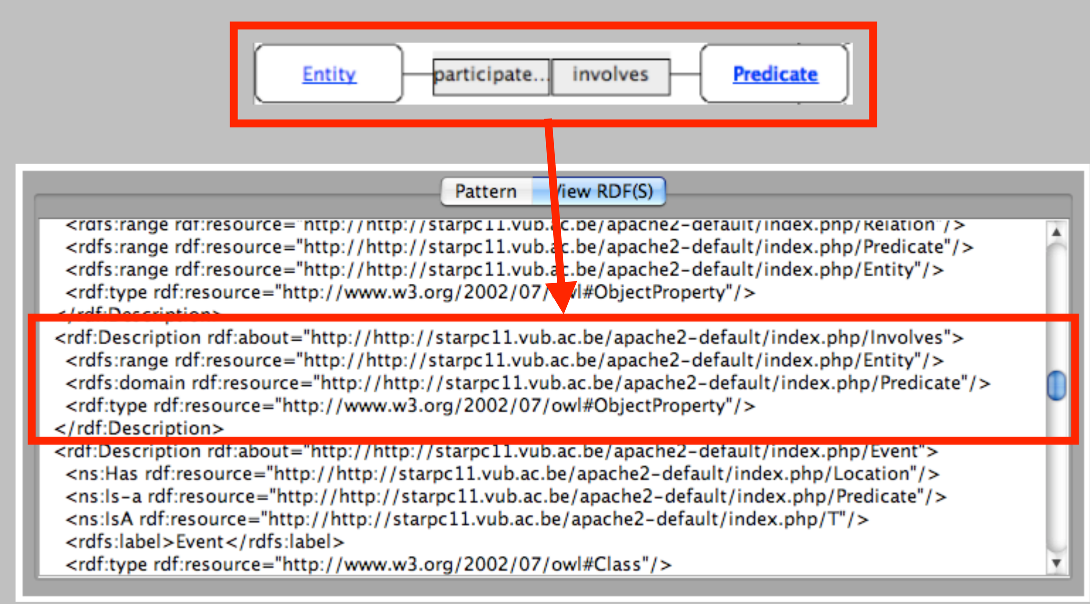

Figure 5: Exporting a concept and its lexons to RDF(S)

Applications

We have used GOSPL in the TAS³ [6] project to allow end users (e.g. security and privacy experts) to easily develop conceptual models on the domain of security and privacy. The resulting ontologies are then used to facilitate access control policy interoperability within the TAS³ project

References

1. http://www.linkeddata.org
2. Brickley, Guha and McBride. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 2004. http://www. w3.org/TR/rdf- schema/
3. Meersman. Semantics ontology tools in information system design. In ISMIS 1999, volume 1609, Springer, 1999.
4. Trog, Vereecken, Christiaens, De Leenheer, and Meersman. T-Lex: A role-based ontology engineering tool. In OTM 2006 (ORM 2006), pages 1191–1200. Springer, 2006.
5. De Leenheer and Debruyne. DOGMA-MESS: A tool for fact oriented collaborative ontology evolution. In OTM 2008 (ORM 2008), pages 797-806, Springer 2008
6. Reul, Zhao, and Meersman. Ontology-based Access Control Policy Interoperability. In Proceedings of the 1st International Conference on Mobility, Individualization, Socialization, and Connectivity (MISC 2010).

Acknowledgements

The research leading to these results has received funding from the EC's FP7 program under grant agreement number 216287 (TAS³ - Trusted Architecture for Securely Shared Services).