

Natural Language Parsing with GOLD and SWRL

Graham Wilcock

University of Helsinki
graham.wilcock@helsinki.fi

We show that natural language parsing can be done with SWRL. Using GOLD, the General Ontology for Linguistic Description (<http://www.linguistics-ontology.org/>) as a general framework for grammatical classes in OWL, we add a lexicon of specific English words. We write context-free grammar rules in SWRL that refer to the word classes and create instances of syntactic constructions. Using SWRLTab in Protégé, we can then parse test strings to classify a range of simple phrases and sentences.

The OWL ontology in Fig. 1 has five instances of *LexicalItem* (*cat*, *mat*, *on*, *sat*, *the*) but no instances of *SyntacticConstruction* (neither *Clause* nor *Phrase*). In Fig. 2 the SWRL grammar rules are executed in SWRLTab by Jess. Fig. 3 shows the OWL ontology after parsing. There are now two instances of *NounPhrase* (*the cat*, *the mat*), one *PrepositionalPhrase* (*on the mat*), and two instances of *VerbPhrase* (*sat*, *sat on the mat*). The test strings *the cat sat* and *the cat sat on the mat* have been correctly classified as instances of *MainClause* (sentence).

The demo uses SWRL builtins for strings (in Protégé 3.2 beta, build 324 or later).

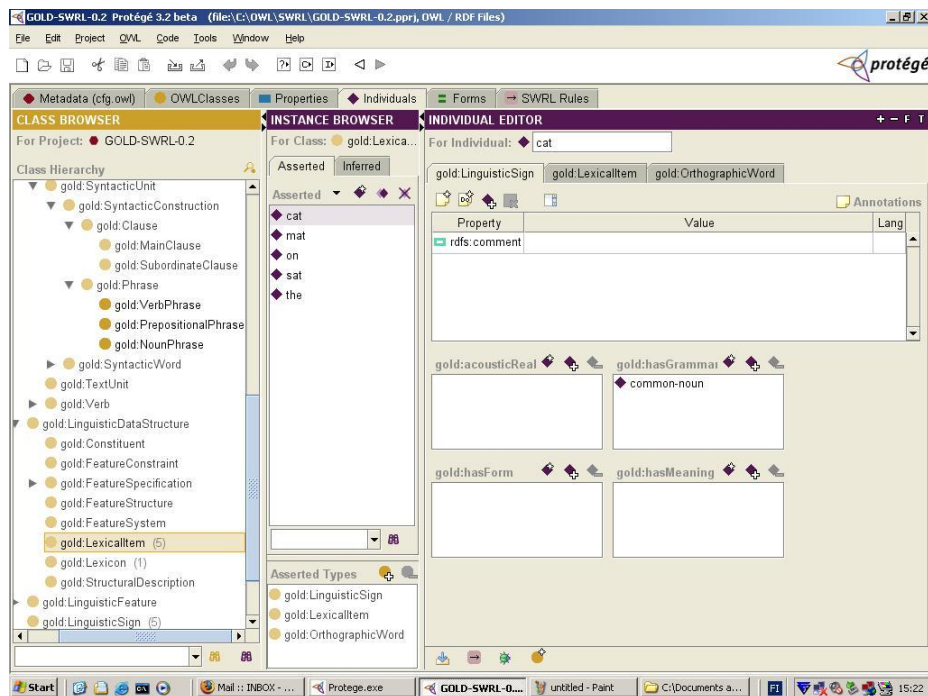


Fig. 1. The OWL ontology before parsing.

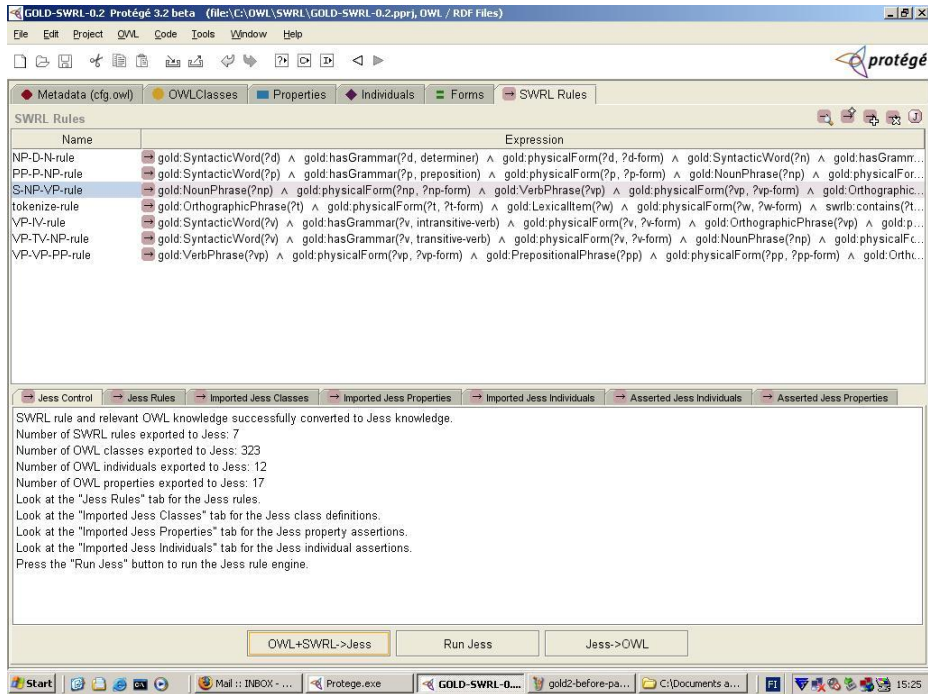


Fig. 2. Executing the SWRL grammar rules using SWRLTab.

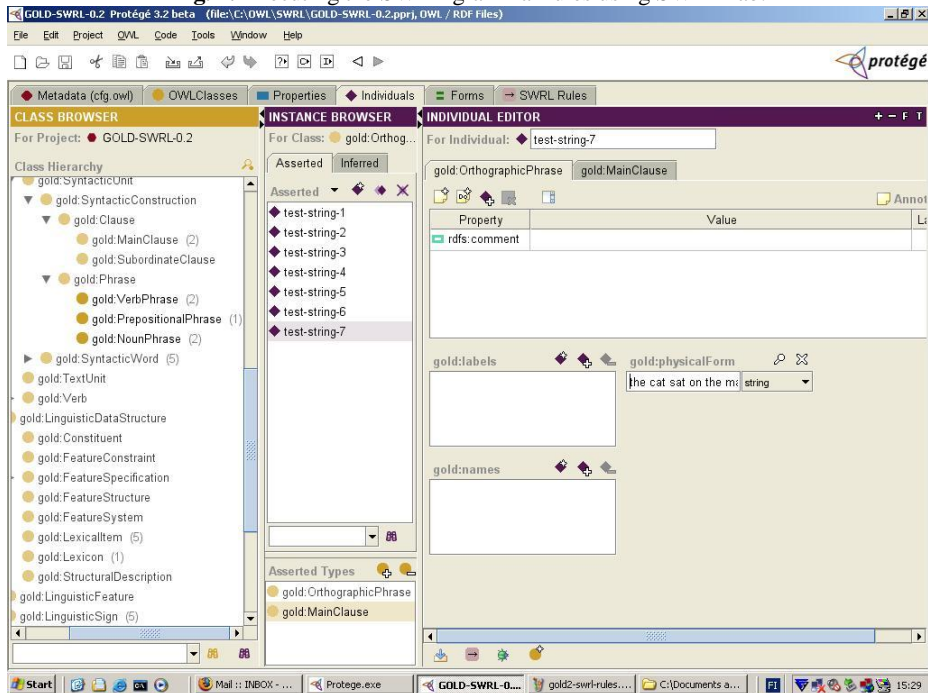


Fig. 3. The OWL ontology after parsing.