Project Proposal: SUMO On Protégé

Elayne Delgado

Clemmie Malley

Florida Atlantic University

Professor Dr. Ravi Shankar

COP5859 Semantic Web Programming

July 15th, 2015

**Project Proposal**

The SUMO on Protégé project’s objective is to use the owl file provided by Adam Pease, the Protégé application and python code to repair the owl code. Currently when the SUMO.owl file is loaded in to Protégé the ontology is flattened. For example: The MISO entries are at the highest level with Entity. Therefore there is a need to update the owl file in order for it to be read correctly through any application moving forward, steps for how this will be achieved can be found in the Preliminary Exploration section of this paper.

**Project Requirements**

The outcome for this project is to compile a working SUMO.owl file that correctly loads all hierarchies, classes, instances, etc.

The following tools and applications will be needed in order to compile a working file:

1. SUMO.owl file found on Adam Pease’s website (http://www.adampease.org/OP/SUMO.owl)
2. Protégé 5.0 – application used for loading SUMO.owl to view the SUMO ontology.
3. WingIDE & iPython Notebook – applications used for coding python that will help identify the hierarchy issue and also assist with adjusting the errors once identified.
4. Internet – used for initial investigation on repairing SUMO ontology. Could provide answer for flat hierarchy issue.
5. Blackboard – documentation from Professor can be found on blackboard. Use this documentation to better understand SUMO.
6. Github – repository where final owl file will be submitted.

**Preliminary Exploration**

**Understanding SUMO**

An ontology is defined as a shared conceptualization of a domain and a set of definitions in a formal language for terms describing the world. The Suggested Upper Merged Ontology, SUMO, is a large, open source, formal ontology stated in first-order logic that is mapped to WordNet, a large multi-lingual lexicon and as of May 2002 roughly 50,000 WorldNet synsets have been mapped to the SUMO. The purpose of linking the SUMO to WordNet is to promote the use of SUMO in natural language understanding applications. The SUMO was created by merging existing upper level ontologies. Sumo is currently divided into 11 sections and each are unique ie: the Structural Ontology, Base Ontology, Set/Class theory, and Graph theory.

The SUMO ontology begins with an Entity and then is divided into subcategories, Physical which includes any entities that have a position in space and time, and Abstract which includes all other entities. Many additional subclasses are found under these two branches, which allow for further distinction of each of these branches of Entity. In ‘SUMO Overview’ documentation, each of these branches are discussed and further defined as disjoint where applicable. Understanding this hierarchy will provide insight as to where adjustments need to be made.

**Initial Actions**

We have opened the SUMO.owl file in protégé by following the actions below:

1. Obtain the SUMO.owl URL from Adam Pease’s website.
2. Open Protégé application. Load SUMO.owl file by following steps: File – Open from URL… - Enter <http://www.adampease.org/OP/SUMO.owl> - Continue - File will take a while to load.
3. See figure 1 for Active Ontology Tab. See figure 2 for Classes Tab showing the flattened hierarchy.

When the SUMO.owl file is loaded a hierarchy can be found along with object properties and data properties.

**Next Steps**

Moving forward, we plan to research online to determine if there is a simple solution to fixing the read error. If no solution is found, we plan to divide the owl file and validate each line is correctly transferred to Protégé. If a line is missing or written incorrectly we will make these adjustments in protégé and save to the owl file.

**Project Significance**

By adjusting the SUMO to owl code our hope is this will allow for universal use of the owl file. Currently when the SUMO owl file is read in Protégé the hierarchy is flattened and therefore showing incorrectly. The end result of this project will be an owl file uploaded to GitHub that is corrected and accessible for other people throughout the internet to use. Additionally, we will compile a technical paper documenting our actions and conclusions from the assignment. This paper will explain our actions and help those who access our file from Github to understand the changes we made.

**Project Challenges**

**Experience Risk**

Inexperience with the tools and applications will lead to project challenges because, although we have used each of these tools before, we are still in the learning phase. Our limited experience in programing can be seen as a risk as we may have issues creating working code in python. However, because we have been working with Protégé since the beginning of the semester and have been gaining knowledge about python little by little, we should be able to adjust the code so that everyone can be able to upload SUMO.owl file into Protégé correctly.

**Time Risk**

Additionally, time is a major risk for this project. As the SUMO file is so large it will take time to find the exact issue. Although we may be able to use Python to parse through the code faster, this may not be a viable solution once we begin our project. We hope that we will be able to achieve a completed working file by the end of the semester. We will keep the professor updated.

Our timeline for this project is to work on the paper as we develop a working file because there are pieces of the final paper that can be documented prior to complete. Our goal is to have a completed file by July 31st, allowing for the weekend to finalize the corresponding paper.

**Project Development**

**Initial Development**

In order to start this project we have to do further research on SUMO. We need to better understand the hierarchies and purposes for these. The extract provided by Professor Shankar from Adam Pease on ‘SUMO Overview’ will help to better understand this ontology and its purpose. Additionally, we will need to research and see if there is a known solution for solving the hierarchy issue with the owl file.

Once we become more familiar with SUMO and if we have not found another solution, we will start reading through the code to see why the hierarchy is loading flat.

We will likely need to use python to read portions of the owl file at a time to identify where the errors begin and end as the file is extremely large and could take weeks to read through completely if the error is not evident.

**Working Project Development**

After researching SUMO further and viewing the SUMO.owl file loaded in Protégé, the issue of the hierarchy is clear. The subclasses are showing at the same level as entity when entity should be the main branch. Although under entity the subclasses appear to be correct, the lower subclasses are where the issue is found. As the entities become more specifically defined it appears this is where the file is flattened. Our next step is trying to identify why the file is loading this way and why the hierarchy is not being reflected correctly.

Appendix:

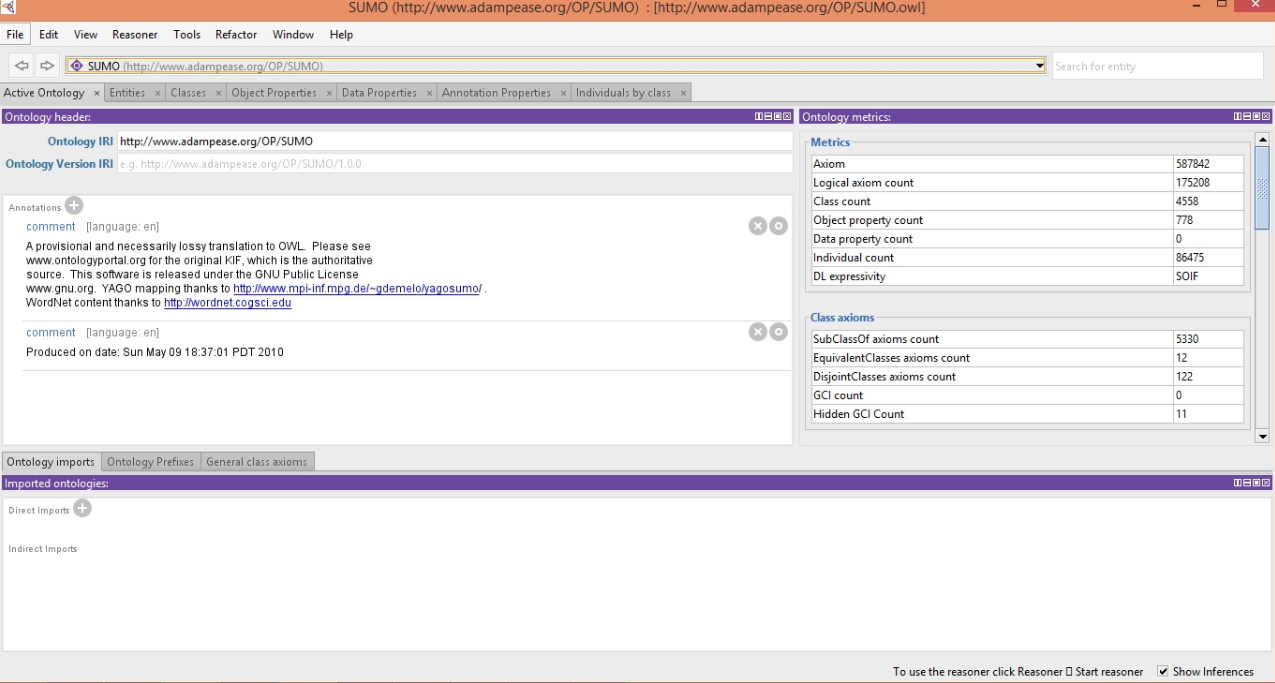


Figure 1: Active Ontology

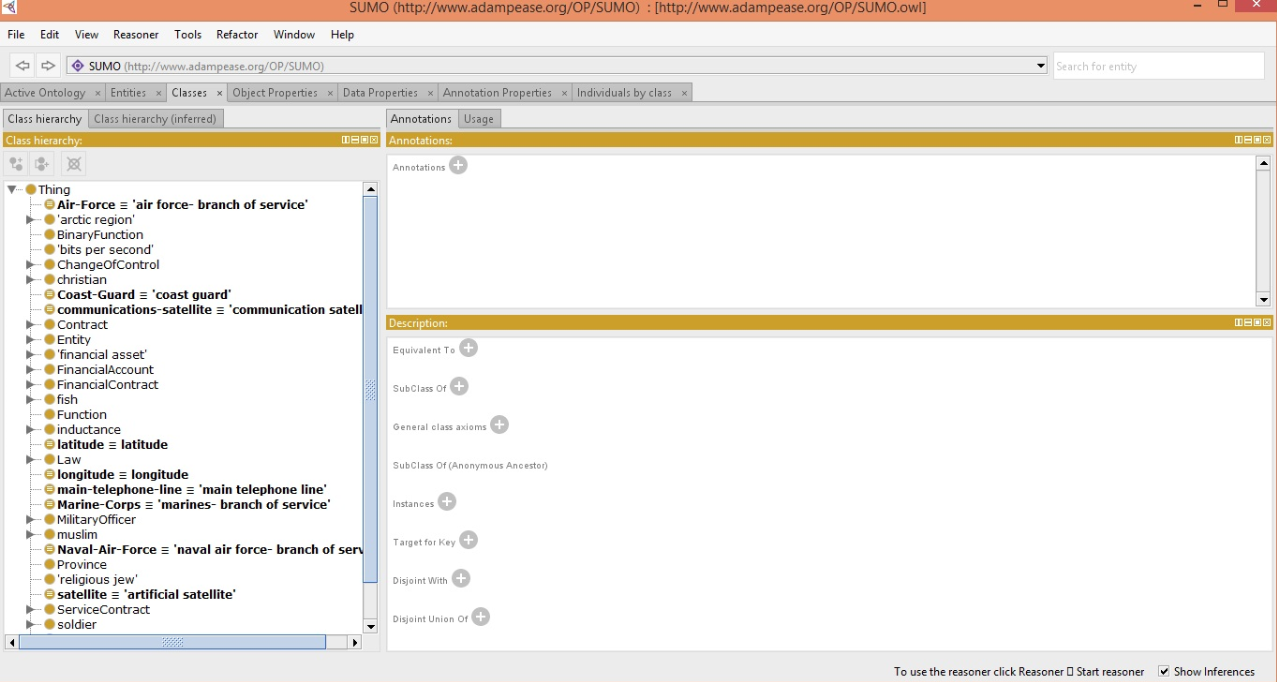


Figure 2: Class Hierarchy