Project: SUMO On Protégé

Elayne Delgado

Clemmie Malley

Florida Atlantic University

Professor Dr. Ravi Shankar

COP5859 Semantic Web Programming

August 3rd, 2015

**Contents**

**Abstract 3**

**Project Overview 4**

**Project Methods 5**

**Project Results 6**

**Project Challenges 7**

**Conclusion 8**

**References 9**

**Appendix 10**

**Abstract**

The purpose of SUMO on Protégé project was to utilize the SUMO ontology presented by Adam Pease and successfully upload into the Protégé application with a traditional hierarchy. Initially when the SUMO.owl file was uploaded into Protégé, the hierarchy was flattened leaving subclasses at the same level as Entity which should be the highest branch. Through research and hard work we were able to utilize another SUMO.owl file, SUMO1.owl, and adjust the classes and instances to populate a SUMO.owl file that can be utilized in Protégé, SUMOFinal.owl. Although we did not resolve the original hierarchy issue by utilizing python to parse through and adjust the owl file, we were able use our knowledge and experience on Protégé and our research for this project to present a working ontology.

**Project Overview**

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**

**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. SUMO is the highest level ontology divided by physical and abstract classes. SUMO, MILO (Mid-Level Ontology) and other universal domain ontologies are combined to form the largest formal public ontology in existence.

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.

**Tools and Applications Required**

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 Research**

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.

**Project Methods**

We found that there have been many users that have attempted loading the file and had issues.

We first found a GitHub user, GoogleCodeExporter, who documented failure to load the ontology and the steps they took.

We decided to attempt replicating these steps:

1. Open Protégé 5.0 tool
2. File – open from URL – enter: <http://www.ontologyportal.org/translations/SUMO.owl>
3. This resulted in an error that the <http://www.adampease.org/translations/SUMO.owl> could not be found. We removed the /translations/ path and attempted load again.
4. The load was successful but still resulted in a flat hierarchy.

Continuing to research, we found a few questions and answers from Stanford.edu but these did not provide any useful files or changes. Additionally, we found that a user was attempting to create a tab plugin that would import SUMO into Protégé. We located the Protégé Plugin Library and search for SUMO but we were not able to locate a plugin at this time.

Then we found a website called ontohub.org where someone had submitted the SUMO owl ontology. Here we were able to download another SUMO.owl file to attempt loading. See SUMO1.owl file in project folder.

The following steps were taken to load the owl file:

1. Open Protégé 5.0 tool
2. File – open – find SUMO1.owl file (saved locally)
3. File took time but loaded.
4. The ontology now loads more hierarchal. Entity rolls up to Thing and Entity has subclasses Abstract and Physical. See Figure 3 for this hierarchy.

So far, this owl file appears to be the best option for using SUMO in protégé. The issues we found with the SUMO1.owl file are a few of the subclasses are missing. We believe that these may have been added at a later time then when the file was created. The missing subclasses are as follows:

* Entity - Physical – Process – NaturalProcess (Missing)
* Entity - Abstract – Graph (Missing)
* Entity - Abstract – GraphElement (Missing)

Although there are only three under Physical and Abstract branches missing at this level, there are also others missing in other sub-classes. We then evaluated the option of adding these branches and creating a new owl file from SUMO1.owl.

Our options for completing this project are:

1. Opening the URL (<http://www.adampease.org/OP/SUMO.owl>) in Protégé. This file loads in Protégé, see Figure 1, but the hierarchy is flat, see Figure 2. When attempting to parse through this code we found that we did not have enough technical knowledge to completely identify where the issue was. Initially, we planned to find the issue and use python to adjust the code but this did not work for us.
2. Utilize the SUMO1.owl file found from online research. When loaded this file appears as the correct hierarchy but is missing some of the subclasses. To fix this issue, we will compare the Sumo Hierarchy listed on Adam Pease’s website ( <http://virtual.cvut.cz/kifb/en/toc/229.html> ) to hierarchy that loads from the SUMO1 file. We will add subclasses and instance data that is missing and save to a new file.

**Project Results**

**Project Steps**

After evaluation of our options we decided that the SUMO1.owl file was our best option for being able to present a workable solution by the end of the semester.

Steps for completing this project are as follows:

1. Load SUMO1.owl file in Protégé 5.0 application
2. Access SUMO Hierarchy - <http://virtual.cvut.cz/kifb/en/toc/229.html>
3. Expand and evaluate each class, subclass, instance data and object property throughout the SUMO1 Hierarchy.
4. Where SUMO1 does not make the SUMO Hierarchy from Step 2 – adjust the SUMO1 file to match.
5. Repeat steps 3 and 4 until entire Hierarchy has been reviewed.
6. Save updated hierarchy as SUMOFinal.owl.

**Project Outcome**

After repeating steps from the Project Steps section above, we were able to create a SUMO hierarchy that reflected the Hierarchy listed on Adam Pease’s website, see Figure 5 and 6. Adding subclasses and instance data was a tedious effort as some classes had more than 20 instances attached that were missing from the SUMO1 file. Validating and adding these instances took a lot of time. Although the original SUMO file could have been resolved through utilizing code to parse through the file. As we are more functionally inclined then technical, we felt that option 2 allowed for us to utilize our knowledge of Protégé and Ontologies to compile a SUMO file that loads hierarchal into to Protégé.

**Project Challenges**

**Experience Challenge**

Inexperience with the tools and applications lead to project challenges because, although we have used each of these tools before, we were still in the learning phase. Our limited experience in programing was a challenge as we struggled with not being able to python to parse through the option 1 file. However, because we worked with Protégé from the beginning of the semester and because this tool is more functional in mindset we were able to use Protégé to produce a working product. Because we are both not technically inclined, our options for resolving this hierarchy issue were limited. Our experience was a major challenge that affected the outcome of our project.

**Time Challenge**

Time was a major challenge for successful completion of this project. The summer semester left a small portion of time to complete this project. Had we had additional time to complete this project, we could have evaluated other options for resolving the hierarchy issue. Although the project outcome was a sumo owl file that loads as a hierarchy instead of flat, we believe there are likely differences between our file and the file located on Adam Pease’s website.

**Conclusion**

Although our first option for using python to parse through the SUMO code did not pan out like initially planned we believe that we have used our knowledge of information and research to help come to a universal conclusion. The SUMO code on Adam Pease’s website was an estimated 4,000 pages when loaded into Microsoft Word and when uploaded into protégé the SUMO file presented a flat hierarchy. While doing research we found other users who had similar issues and provided their workarounds to resolving. We attempted all workarounds and ultimately we used a file that someone uploaded to the internet because the file, when loaded in Protégé, presented a hierarchal ontology versus flat.

However, this code was not complete and needed some adjustments. To make this a working SUMO.owl file we went over the classes and corresponding instances to validate they matched those that were in place by Adam Pease. This comparison took some time to complete, but in the end we were able to provide a more accurate look at how the hierarchy should be. SUMO is a very interesting ontology and everything we have researched has provided us with a clearer understanding of SUMO and its other domain ontologies.

**References**

Failed loading of SUMO ontology in OWL · Issue #18 · ronwalf/swoop. (n.d.). Retrieved July

24, 2015, from https://github.com/ronwalf/swoop/issues/18

SOCoP. (n.d.). Retrieved July 24, 2015, from <http://ontohub.org/socop/SUMO///formality_levels>

Subclass Hierarchy Tree. (n.d.). Retrieved July 27, 2015, from

<http://virtual.cvut.cz/kifb/en/toc/229.html>

The Suggested Upper Merged Ontology (SUMO) - Ontology Portal. (n.d.). Retrieved July 20,

2015, from <http://www.adampease.org/OP/>

**Appendix:**

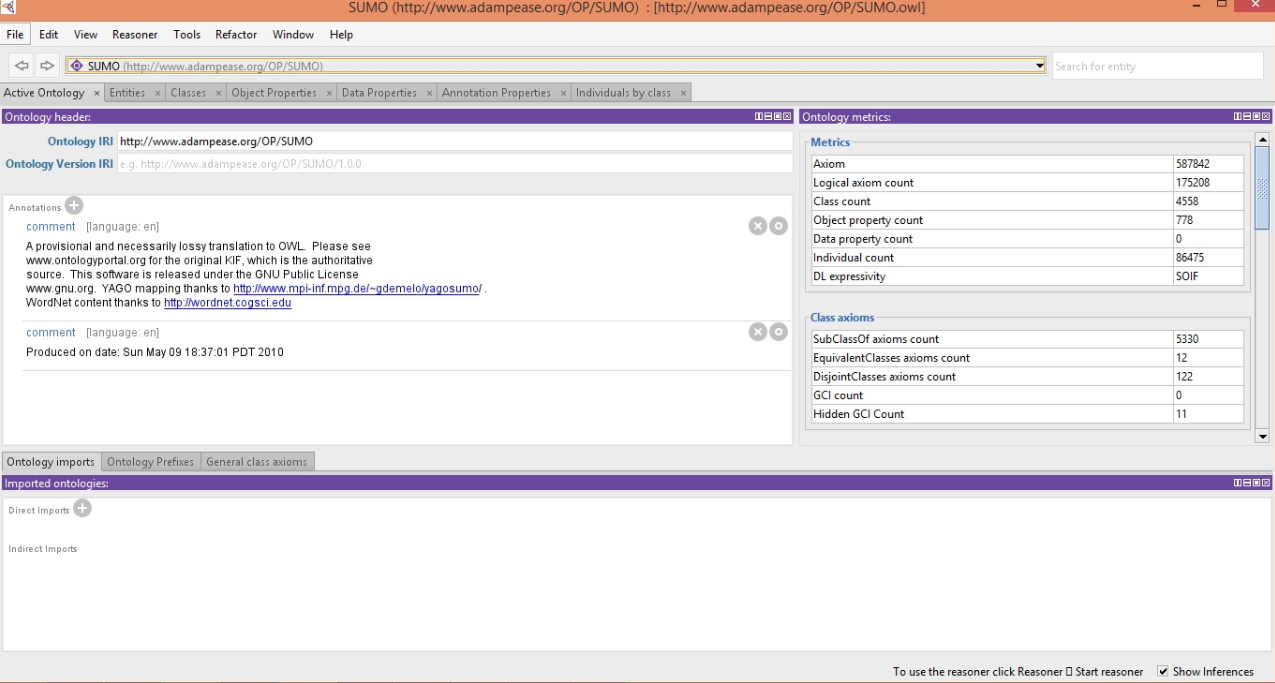


Figure 1: Active Ontology

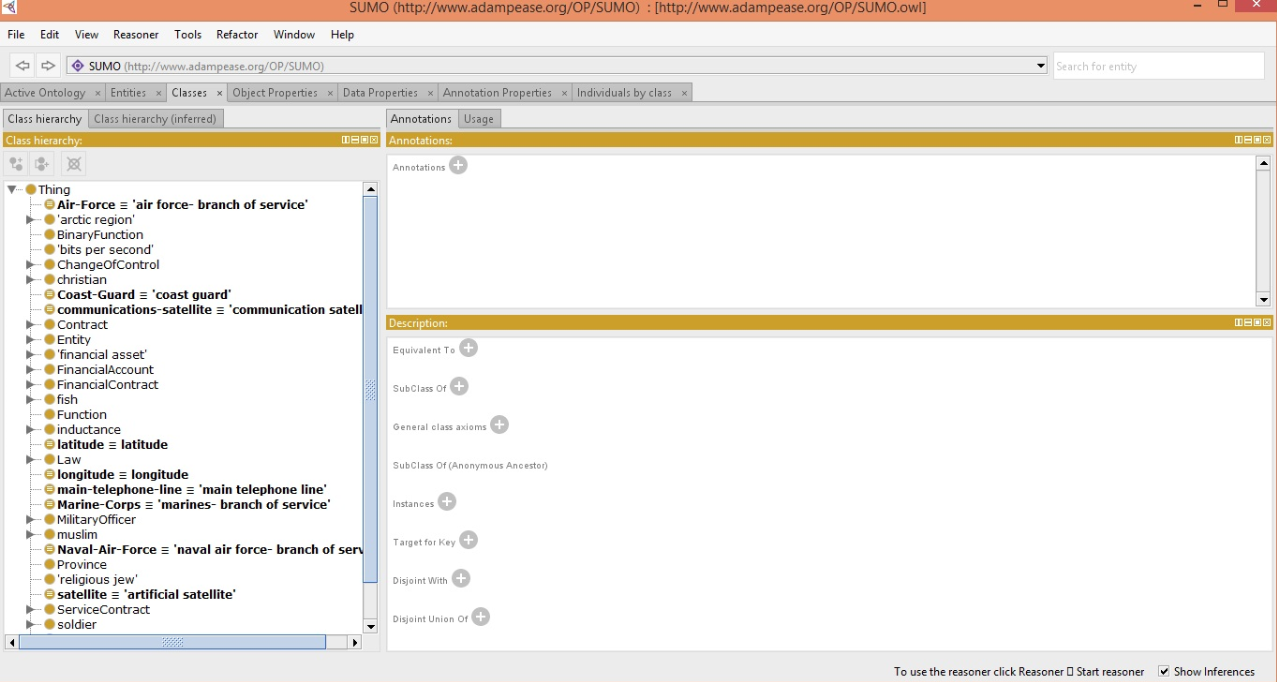


Figure 2: SUMO Class Hierarchy

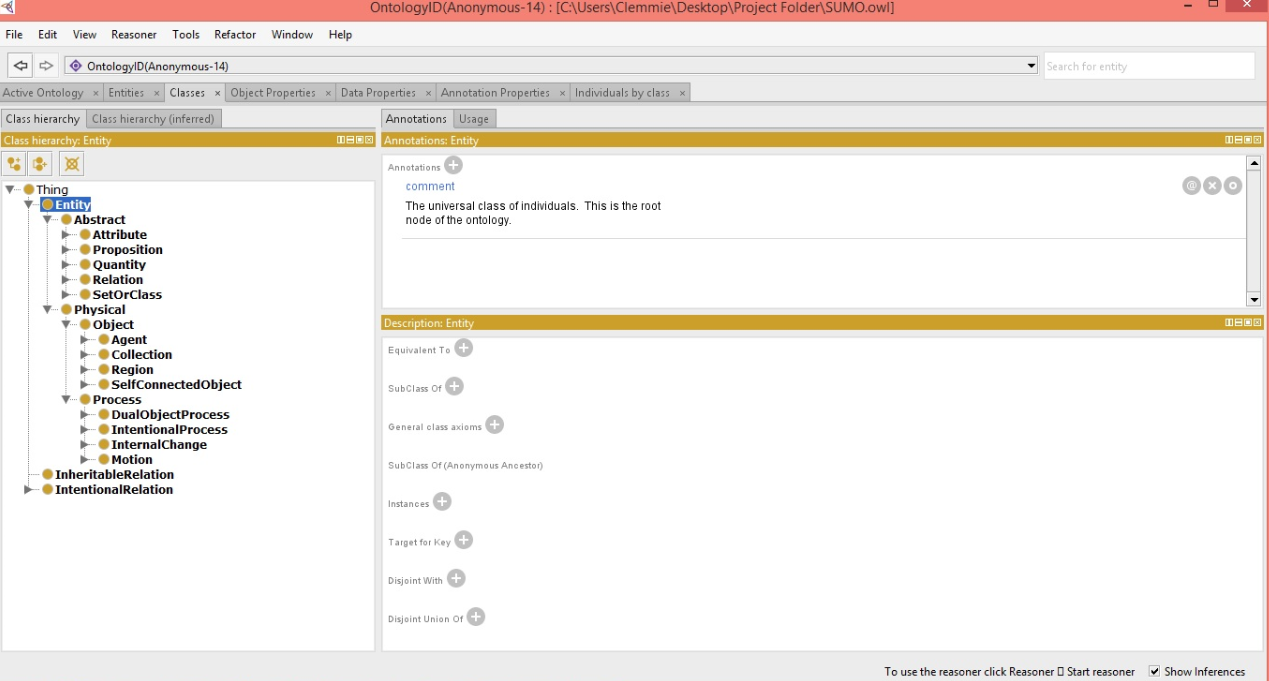


Figure 3: SUMO1 Class Hierarchy

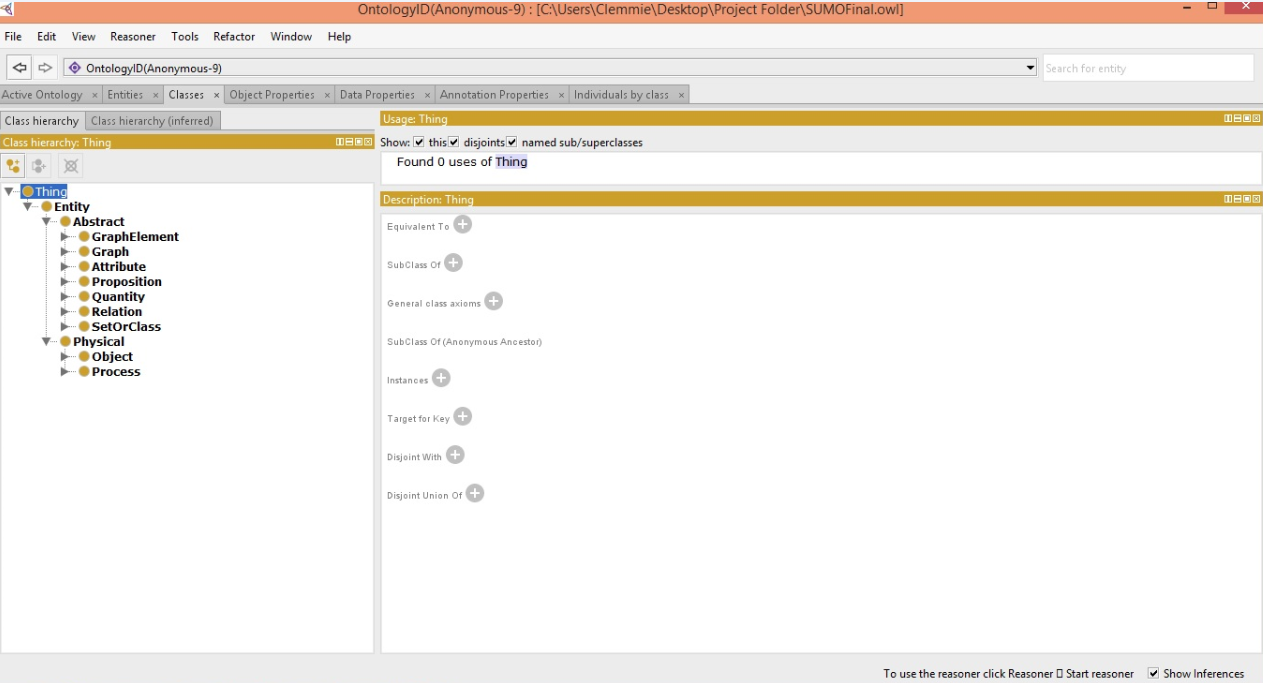


Figure 4: SUMO Final Class Hierarchy

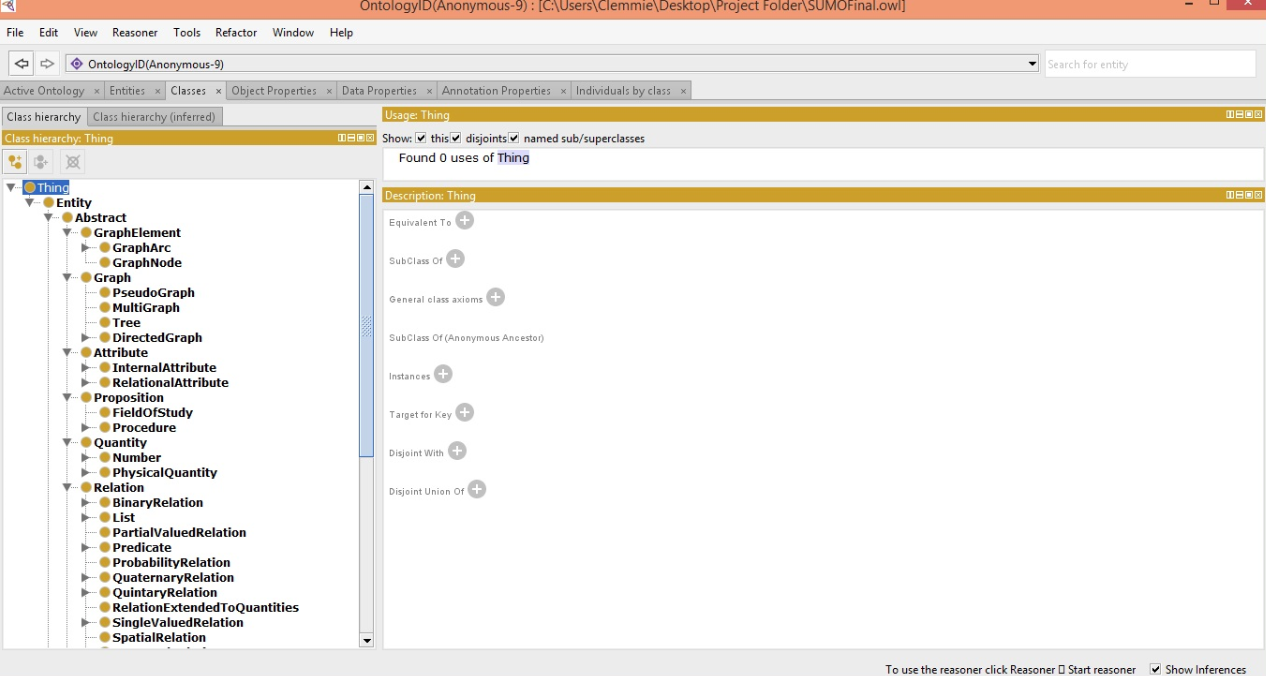


Figure 5: SUMO Final Hierarchy Expanded 1

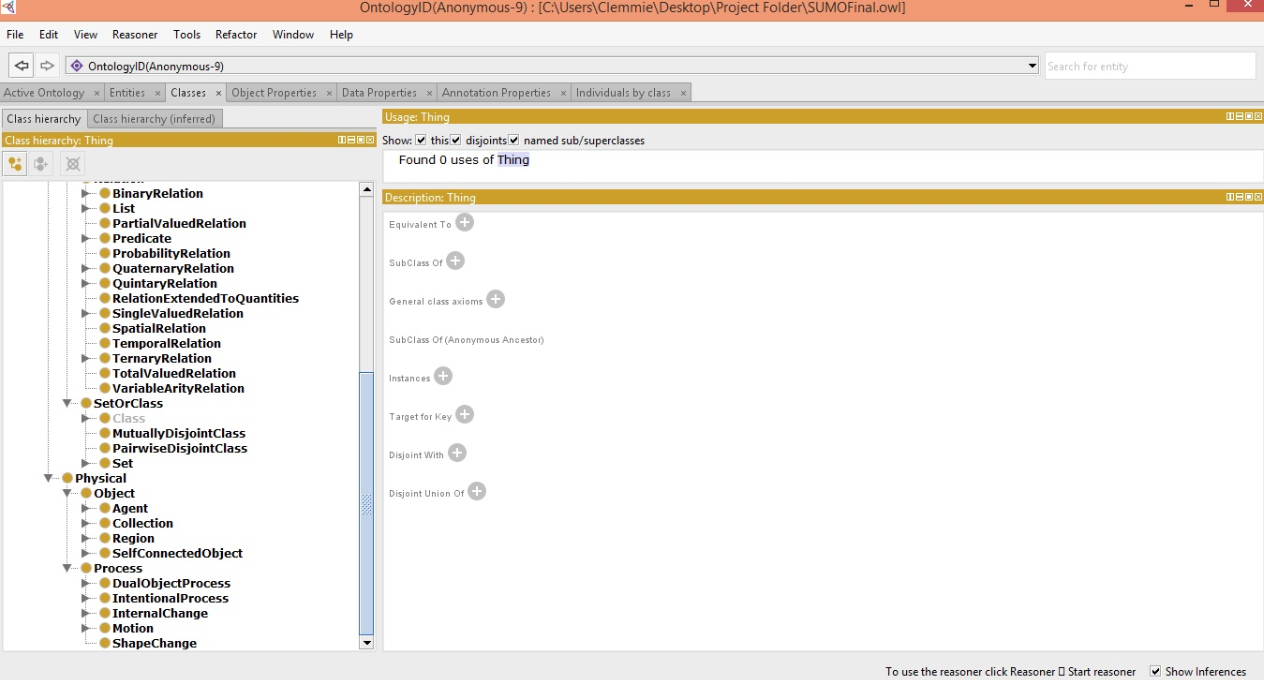


Figure 6: SUMO Final Hierarchy Expanded 2