Project Proposal – “SUMO on Protégé”

Hafeez U. Khan

College of Business - Florida Atlantic University

Prof. Dr. Ravi Shankar

Semantic Web Programming

COP 5859

July 17, 2015

Abstract

This is a research paper proposal for “SUMO Protégé Project”, required for COP 5859 – Semantic Web Programming. This project proposal will elaborate my objectives during this project and class about the SUMO Protégé Ontology.

SUMO On Protégé

**Introduction**

**Project Requirement**

The project “SUMO on Protégé” code is located at the link to SUMO in .owl format at <http://www.adampease.org/OP/> , listed under "Documentation and other resources." When SUMO.owl is imported to Protégé, the SUMO ontology is flattened. To run the SUMO.owl successfully using Protégé, find the errors by:

* Going online and determine if there is a solution; if not,
* Review the SUMO.owl file to see where the error was made and get it right, or,
* As a last alternative, import part of the SUMO.owl file and add incrementally the remainder.

This process may require the use of Python to systematically achieve the result.

**Project Scope**

The goal of SUMO on Protégé project is to create a SUMO ontology using Protégé. This project encourages the student to learn the SUMO-The Suggested Upper Merged Ontology in detail through the process of (1) running the SUMO ontology file in Protégé, (2) analyze the reason of flat ontology, (3) find the trend of correct ontology, (4) replace the non-operational trend with the operational trend, and (5) finally, run the correct SUMO ontology in Protégé. Additionally, the project encourages the student to expand the knowledge in Python. By using Python Regular expression operations, ontology trend can be found and replaced by the correct ontology code.

**Introduction to SUMO**

SUMO is a Large Ontology for the Semantic Web and its Applications. The Suggested Upper Merged Ontology (SUMO) was created by merging a number of existing upper-level ontologies. The SUMO is created as foundation ontology for a variety of information technology processing systems. It was originally developed by the Teknowledge Corporation and now is maintained by Articulate Software. The SUMO is an open source. Initially, SUMO was released in December 2000. SUMO defines a hierarchy classes and related rules and relationship. It is available free and owned by the IEEE. Today, the SUMO- Suggested Upper Merged Ontology and its domain ontologies form the largest formal public ontology. SUMO is being used in research, search applications, reasoning and linguistics. SUMO is the only formal ontology that has been mapped to the entire WordNet lexicon. SUMO is written in the SUO-KIF language, the ontologies that extend SUMO are available under GNU General Public License. Mr. Adam Pease is the Technical Editor of the SUMO.

**Use of Semantic Translation and Mapping During the Project**

This project involves the Semantic Translation. The Semantic Translation is the process of using semantic information to assist in data translation from one data model or representation to another or data model or representation. By using the concept semantic, the semantic translation takes the benefit to associate the meaning of one dictionary data elements to create same meaning in another system. XML conversion is an example of semantic translation. Using XML conversion, one data model can be converted to another data model with the help of formal ontologies called OWL-Web Ontology Language, for each system. During this project, the semantic mapping will be used. This semantic mapping will be different from data mapping tool. Because the data mapping just do the simple one-to-one translation of data from one system to another system. Data mapping does not associate meaning of data elements from one to another. While, in semantic mapping it’s totally opposite and will be followed during this project to associate the actual associated meaning.

The semantic translation of this project requires that data elements in the source and destination systems have "semantic mappings" to a central registry or registries of data elements. The simplest mapping is where we will find the equivalence. During this project I will look for these three types of Semantic equivalence:

**Class Equivalence** – Class Equivalence means, the class or concepts, both are equivalent. For example: "ProcessSteps" is the same as "Process Flow"

**Property Equivalence** – Property Equivalence means, two properties are equivalent. For example: "AgentGivenName" is the same as "FirstName"

**Instance Equivalence** – Instance Equivalence means that two individual instances of objects are equivalent. For example: "OrgProcess" is the same person as "OrganizationalProcess"

**Participants**

The “SUMO on Protégé” is a single student project. I, Hafeez Khan will be the only participant of this project.

**Project Deadline**

The deadline for this project is August 3rd, 2015.

**Project Requestor**

Prof. Dr. Ravi Shankar is the requestor for this project. Prof. Dr. Shankar’s contact information is given as under:

Center of System Integration Director

FAU-Florida Atlantic University

College of Engineering and Computer Science

777 Glades Road – 513 EE (Building 96) Boca Raton, Florida 33431

Tel- 561- 297-3470

Fax (561) 297-2800

Email: [shankar@fau.edu](mailto:shankar@fau.edu)

**Data Source**

Project file name is SUMO.owl, located at “<http://www.adampease.org/OP/SUMO.owl>”. This file is also located at COP 5859- Semantic Web Programming Course webpage at Black Board under Code Resources and Semantic Web Resources tab.

**Method**

**Data Provision**

First, I will try to download the data from <http://www.adampease.org/OP/SUMO.owl> , SUMO Ontology site. Since, the code had 819237 lines; the copy paste option might freeze the browser. So, I will download the SUMO.owl file from the COP 5859- Semantic Web Programming Course’s webpage at Black Board by using the right click and safe as option.

**Tools and Materials**

I will use following tools during the project:

**Python:** I will use the Python’s regular expression operations to find the errors and replace the correct code.

**Protégé 4.3:** I will use the Protégé version 4.3 to run the SUMO.owl file to see the ontology of my project

**Search Engines**: I will use the Google search engine to search for the project solution during the option No. 1 of the project.

**Reference Sites**: To get the information about the SUMO and project related material I will use the Suggested Upper Merged Ontology (SUMO) and <https://www.python.org/> websites.

**LinkedIn**: I will use LinkedIn to communicate with Prof. Dr. Shankar to discuss any related problem during my project cycle.

**Skype**: I will use Skype meeting to discuss the project issues and to get the project guidance from Prof. Dr. Shankar.

**GitHub**: I will use the GitHub site to submit the project related files under “Semantic- Web-Hafeez-K-SUMO on Protégé Project” folder.

**Project Communication**

LinkedIn and FAU email will be use for communication related to this project. Skype will be used for online meeting and discussion purpose.

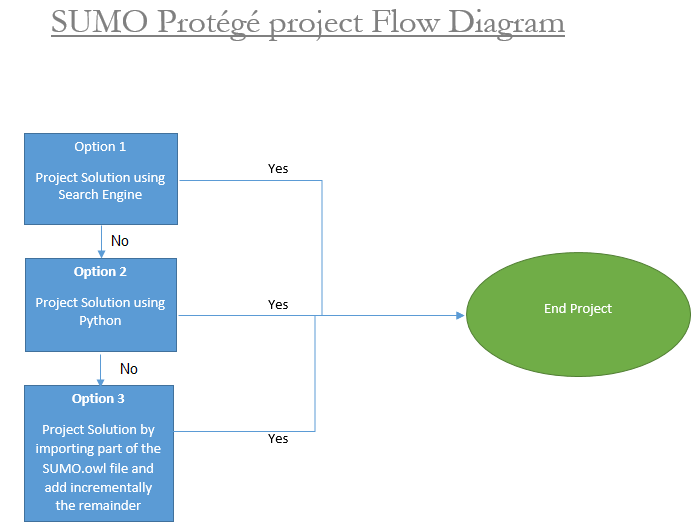
**Project Design**

To meet the project requirement, I will adopt three options.

**Option 1:** I willsearch the project solution by going online through the Google search engine. I will try to search the SUMO ontology blogs, project reviews and online books. If I will not be able to find the solution online, then I will go for option 2.

**Option 2:** In option 2, I will review the SUMO.owl. I will try to find the error in the file. If I find the errors and trend in SUMO.owl file then I will try to replace the error code with the correct ontology code using Python. I will use the Python’s regular expression operations during this option to find the errors and replace the correct code. If I will not be able to make Python code to find and replace the code, I will adopt the option 3.

**Option 3:** In option 3, I will simply import part of the SUMO.owl file and add incrementally the remainder. First, I will transfer main classes, then sub-classes then finally, sub-sub-classes and so on.

****

**Project Timeline**

Since, the deadline to submit the project is August 3rd, 2015; I will commence the project on 18th July 2015 and complete the project on 1st August 2015, within two weeks.

**Week One** (18th July 2015 to 24th July 2015)**:** On 18th July I will search for the project solution as mentioned in option No. 1. If I will not find the solution, I will utilize 6 days of first week to solve the project b using option No. 2. If, I find the solution as per option No. 2, I will further use Python code to facilitate the Python to make project more ease.

**Week Two** (25th July 2015 to 1st August 2015)**:** I will not be able to fine the Python code for the solution then I will use second week to complete the project as per option No. 3.

**Unforeseen Situation**: On 2nd August 2015, I will handle any unforeseen situation of the project.

**Project Expenses**

There is nofinancial expense involved to complete the SUMO on Protégé project. Time wise, I will utilize 100 hour to complete this project.

**Risks Analysis**

Time factor is the only risk associated with this project because the deadline to complete the project is August 3, 2015. While, option No. 2 is time utilizing situation. If I fail to find the solution through option 2, I will only have one week to finish the project through option No. 3.

**Project Completion and Submission**

After completion of the project, I will submit the screenshots and related files to GitHub, under SUMO on Protégé Project folder. Also I will upload all files and screenshots to Blackboard with a link to LinkedIn.

References

Suggested Upper Merged Ontology (SUMO). (2015, July 7). Retrieved from http://www.adampease.org/OP/index.html

Pease, A. (2002). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. Retrieved from https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf