**Capstone Project Proposal**

**Building Ontologies for the Semantic Web**

**Khuong D. Tong**

**Christopher R. Seemann**

A Special Project

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

In

Computer Information Technology

Department of Computer Science

Central Connecticut State University

New Britain, CT

**Spring 2016**

**Project Advisor: Dr. Neli Zlatareva**

**Introduction**

The semantic web is one of the important developing technologies that is reshaping the web by changing how information is represented. In earlier, and even recent implementations of web content are organized through links and the content is labeled by metatags that are accessed by web crawlers and search engines. Implementation of the semantic web hopes to improve the organization of content by replicating how information is represented in semantic memory where knowledge is represented relationally in a semantic net. In a semantic net, the link between objects provides additional information about the relation of content so that all content is related by information. When implementing a semantic web, the additional information of how content is related provides for more precise access to desired content in a query to produce inferences. However, the implementation of the semantic web has not been standardized across platforms. The following proposal examines and implements one of the means of organizing content in a semantic web, Protégé, and compares this to another means of organizing content in a semantic web, the triple-store.

**Background**

The standard of organization of content on the web at present utilizes links between content and is called a “Web of documents.” These links are organized through metatags that are accessed by web crawlers and search engines that attempt to determine related or relevant content. This limits what can be known about these links and how knowledge of these links can be used because the relationship is not specified. However, the more recently, and sparingly, implemented organization of information on the web is a semantic web refered to as “Web of data”. Semantic web differs from the traditional means of organizing web content since content is also organized by how content relates to other content (W3C).

The semantic web developed from research into semantic nets that attempted to replicate how knowledge is likely represented in the mind and has been investigated to represent semantic memory (Collins & Quillian, 1969). In semantic memory each object is represented by a set of nodes that define the objects features. These nodes are linked to form a semantic net. In semantic nets objects are not just related to each other, as is common on the internet, but the nature of these relationships is important by determining how objects are related. This distinction can be illustrated in an example. When trying to represent the concept of cat in a cat would have many features such as tail, legs, and mammal. However, in a semantic net the nature of these links are defined. For example, a cat is a mammal and has a tail and has legs and a cat is the combination of these features. This additional information allows for the use of relational logic to retrieve content from a semantic net leads to the ability to represent inheritance (Sowa, 1987) and spreading activation of information (Collins & Loftus, 1975).

The semantic web is an implementation of a semantic network to organize web data. To facilitate the implementation of a semantic web, the W3C has produced standards (W3C). The first of these standards characterizes linked data that like with metatags that classified content in the “Web of documents,” content in linked data is tagged by machine readable data that specifies both data relating to the content but also links that specify the relationship and is known as RDF. This allows for the automated agents to access the data. The next standard is that of ontologies. These are shared vocabularies that provide common descriptions of data and relationship and are known as OWL. This allows for linked data to be more easily combined with other linked data to increase the amount of accessible data. The next established standard is a means of querying the linked data for answers to questions. SPARQL provides a common query language that can be used to access multiple stores of linked data. The final standard established is for inference. Inference is used to manage different stores of linked data. Combining linked data stores provides the ability to increase the data that is linked to a piece of content by merging unique information from each of the linked data stores. When widely implemented these standards will help to increase personalization on the internet and develop human knowledge.

**Proposed Work**

The main intent of this project is to create an ontology in OWL using Protégé, make the ontology available online, and demonstrate the possible use of an ontology by an application. The ontology will convey meaningful knowledge from a selected domain. It will be utilized for building a web application, and will demonstrate ontology alignment and merging with another coexisting web application.

The second objective is to share our ontology with the general public. For this, we will host the ontology and make it available on a public domain. Sharing the ontology will allow fot its reuse by another application. We will demonstrate also the querying capabilities of SPARQL.

Our final objective is to create an application for the ontology. The application will make use of a triplestore coexisting on the web . The aim of the application is to demonstrate the viability of the semantic web as a consumer application. This means the application must demonstrate the benefits of the semantic web. This may include but is not limited to showing reuse of domain knowledge, make domain assumptions explicit, and share common understanding of structured information.

**Hardware & Software Requirements:**

* Windows/MacOS Compatible Computer
* Windows 7 (32 or 64 bit)
* MacOS 10.11
* Protégé
* SPARQL
* RDF, XML

**Initial Project Timeline**

|  |  |
| --- | --- |
| Week 1 | Introduction to semantic data |
| Week 2 | Introduction to Protégé |
| Week 3 | Classes, Properties, Individuals in Protege |
| Week 4 | Building ontologies with Protege |
| Week 5 | Inference capabilities: Fact++, Hermit |
| Week 6 | Querying capabilities with Protege |
| Week 7 | Building example ontology |
| Week 8 | Building example ontology |
| Week 9 | Exporting the ontology |
| Week 10 | Merging ontologies |
| Week 11 | Validation of merged ontologies |
| Week 12 | Documentation |
| Week 13 | Final project |
| Last two weeks | Final project and presentation preparation |

**References**

Berners-Lee, T., Hendler, J., & Lasilla, O. (2001). The semantic web. *Scientific American*, *284*

(5), 28-37.

Collins, A.M. & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal*

*Learning and Verbal Behavior, 8 (2),* 240-247.

Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic processing.

*Psychological Review, 82,* 407-428.

Sowa, J.F. (1987). Semantic networks. In Stuart C. Shapiro. *Encyclopedia of Artificial*

*Intelligence.* Retreived: http://www.jfsowa.com/pubs/semnet.htm

W3C. (2015). Semantic web. Retrieved from: https://www.w3.org/standards/semanticweb/