Design of Intelligence Gathering Model: A Semantic Web Based Approach

***[1]Jimoh, R.G., [1]Agbo, F.J., [1]Kissinger, S., [2]AbdulRaheem, M., [2]Tomori, R.A. and [3]Salimonu, I.R.**

[1]Department of Computer Science, University of Ilorin, Ilorin Nigeria
[2]COMSIT Directorate, University of Ilorin, Ilorin, Nigeria
[3]Department of Computer Science, Federal Polytechnic, Offa, Nigeria

Received: March 14, 2017; Revised: December 22, 2017; Accepted: February 8, 2018

Abstract

Intelligence information gathering, investigation and analysis are vital components of security management in any given society. The ability to know what, where and when events occur is a key element of investigation process especially in a large dataset. This is a thematic, spatial and temporal (TST) reporting issue. Semantic Web technologies such as Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) have been discovered to be a good approach to solving this problem. In this study, the researchers proposed a semantic web based intelligence gathering model for TST using OWL ontology, which has more expressive constructs than RDF. The study further demonstrated the use of Semantic Query-enhanced Web Rule Language (SQWRL), Semantic Web Rule Language (SWRL), Jess and OWL APIs for ontology reasoning and inferences. The model presents university ontology, which could be applied in implementing security intelligence gathering in university community. Thus, this paper established the possibility of gathering intelligence information that satisfied TST through semantic web based tools.

Keywords: Semantic web, Thematic, Spatial, Temporal, Intelligence Gathering

1.0 Introduction

One of the major issues facing many countries of the world today is the security challenge in different quotas. There is lack of consensus on how to deal with the problem leading to degeneration of the problem. The geographical spread of the problem in Nigeria was earlier revealed [1]. The advent of internet and the World Wide Web (WWW), with all its enormous benefits seemed to have broadened the scope of the security challenge. Tim Berners-Lee, the Director of World Wide Web Consortium (W3C) developed the World Wide Web or web technology for the public [2]. Little was known about it then and soon it became a great platform for commercialism, socialism, and networking all around the world. The web is now being categorized into Web 1.0, Web 2.0, Web 3.0 and 4.0. Some earlier categories of the web are characterized with static sites that do not permit interaction and are mostly proprietary in nature. Web 3.0's peculiarity lies with the use of Semantic web technology. The Semantic web is a project that intends to provide a universal platform for information exchange and by so doing provides a computer-processable meaning on the World Wide Web [3].

***Corresponding Author: Tel: +234(0)8168421369, E-mail: jimoh_rasheed@yahoo.com**

The concept of semantic web as developed by Berners-Lee was to have a web in which computers become capable of analyzing all data on the web i.e. the content, links and transactions between people and computers [4]. The semantic web as envisioned by its author provides technologies such as eXtensible Markup Language (XML), Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) that provide datasets of specific ontology that can be queried across the net by machines or intelligent agents, thereby providing relevant information to its assessors. XML was the first on line among all these technologies that enabled developers to organize data around tags that are well formed or well nested based on a rule written in Document Type Definitions (DTDs) or XML Schema. Then, it was followed by RDF which is often seen as a data model for representing object-attribute-value pattern called *statement*. RDF has been given XML syntax and it is domain independent, i.e. its applicability covers any real world domain. However, users of RDF may choose to define their own terminology by using a schema called RDF Schema (RDFS). The RDF/RDFS enable users to model particular domains such as the products/services. Another language for modeling concepts is the Web Ontology Language (OWL), which is a class of knowledge representation language to describe taxonomies and network classifications for various domains participating on the web.

The application of Thematic, Spatial and Temporal (TST) data model has been considered critical and an apt approach in analytical domain such as national security and criminal investigation [5]. TST model becomes more relevant in tackling issues involving complex datasets vis-à-vis **what**, **where** and **when** an event happens. This research will leverage and enhance on this model to create a framework for monitoring security issues and reporting to the appropriate personnel who is expected to take actions by way of responding to them. The idea here is that rule based solution that enables communication between the hosts or client and server will be developed and integrated in the proposed model.

The security agents can do little or nothing without adequate and prompt facts gathered through Intelligence Gathering (IG). It is cumbersome for the IG unit to ascertain precisely during investigation analysis process, the thematic, spatial and temporal elements of investigation [5]. These TST elements are critical to any investigation process for an informed decision to be made. For example, a bomb blast happened in Jos, Plateau State, and it took a while for the information to get to the appropriate quotas, each pieces of the information was gathered from different sources with different contents especially as regards to the TST facts. Semantic web technology is a powerful tool used to describe resources on the web that enables machines to read and infer certain meaning [4]. This is a good foundation for building a web based model for IG efficiency. Even though Semantic Web technologies has shown to be powerful in knowledge expressiveness, representation and description, very powerful tool known as Web Ontology Language (OWL) enhances on the RDF to give more expressiveness in building our ontology [3]. Furthermore, the process of studying and investigating large dataset especially in security domain in order to derive inference from them is usually cumbersome and time consuming. A model was presented for semantic analytics with properties of the thematic spatial and temporal model, making use of RDF whose query language may not generate result sets that have a robust relationship [5].

The present research seeks to close this identified gap by enhancing the work done by Perry, Sheth, Hakimpour and Jain [5], employing OWL as better ontologies with SQWRL as query language in order to realize an efficient inference system for investigating thematic, spatial and temporal dataset. Thus, the aim this research is to design a semantic web based intelligence gathering model with thematic, spatial, and temporal features. To achieve this goal the researchers defined two objectives. Firstly, all related works were reviewed to understand the concepts of Semantic Web technologies. Secondly, university ontology was developed with OWL as a knowledge base for future researches.

2.0 Review of Related Works

As early as the 1980s, significant researches appeared in information science literature about the development of expert systems for improving search results. Hundreds of universities, start-up companies, and major corporations have published researches and filed patents on various algorithmic techniques for machine-aided searching over three decades (and earlier when much of these works were classified as artificial intelligence) and in the late 1990s and early 2000s, these technologies began to be described as semantic search components [6].

The word semantic itself implies meaning or understanding. As such, the fundamental difference between Semantic Web technologies and other technologies related to data (such as relational databases or the World Wide Web) is that the Semantic Web is concerned with the meaning and not the structure of data [7]. One of the issues facing Semantic Web Computing is the creation and adoption of standardized ontologies in OWL for the various industry domains to precisely define the semantic meaning of the domain-specific concepts. The additional modeling effort incurred by ontologies must result in savings elsewhere [8]. Figure 1 depicts the different Semantic Web technologies stack and their relationships. Ontology refers to the means by which concepts or knowlegde in a specific domain can be specified and reprsented in form of vocabulary. It is noted that ontology is a specification about conceptualization. It is a way to model the relationship between entities in a domain [2].

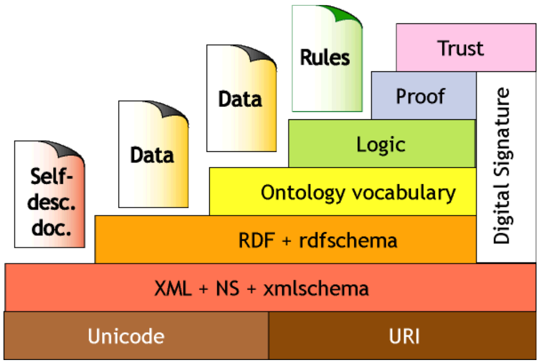

Figure 1: Semantic Web technologies stack [9]

In a research paper, Perry *et al*. [5] highlighted the importance of spatial and temporal data in analytical domains such as national security and criminal investigation. Often, the analytical process requires uncovering and analyzing complex thematic relationships between disparate people, places and events. A description of framework built around the RDF metadata model for analysis of thematic, spatial and temporal relationships between named entities was done while presenting a set of semantic query operators. A major achievement in their work was the modeling of spatial, temporal, and thematic data using ontologies and temporal RDF graphs. According to the decision model proposed by Zilberstein and Shlomo [10], a three layered architecture working concurrently will provide and improve the technology of information retrieval (IR) and information extraction (IE) from a large dataset. The three layers are:

a. User Interface layer (UI)
b. Decision Model Evaluation subsystem (DME)
c. Information Gathering subsystem (IG)

The model is believed to present a high filtering technique for information gathered from different sources in order to achieve reliability, efficiency, cost-effectiveness and decision making, which has been a huge challenge for man to carryout manually [10]. Sustainable knowledge management and more intelligent decision support are beneficial to collect, consolidate, store and share experiences in form of knowledge base or domain ontology in medical emergency management for mass gathering [11]. This paper describes the process of developing and evaluating a Domain Ontology for Mass Gatherings (DO4MG) with a focus on medical emergency management

3.0 Methodology

Foremost, OWL was adopted as Semantic Web technologies for this research because some particular properties such as transitive, symmetric, functional or inverse are not implementable with RDF/RDFS. Also RDF/RDFS lacks provision to specify disjointness classes. Besides, OWL has more expressivity tools than RDF/RDFS. The researchers designed the proposed model and university ontology using relevant tools, some of which include:

1. PROTÈGÈ 4.3 is a free, open source ontology editor with user friendly interface and knowledge acquisition system.

2. Android Software development Kit (SDK) and Android Developing Tool (ADT) are used to build android application and emulate Android device respectively.
3. Apache Server is used to deploy the core ontology and to also hold the information of the users for authentication and role based access control.

3.1 System Architecture

The diagram representing the architectural design of the proposed model is presented in this section to clearly show the decomposition of its major components. These components are subdivided into three levels of abstraction namely Users Level of Abstraction, Server Level of Abstraction and Security Analyst Level of Abstraction (Figure 2).

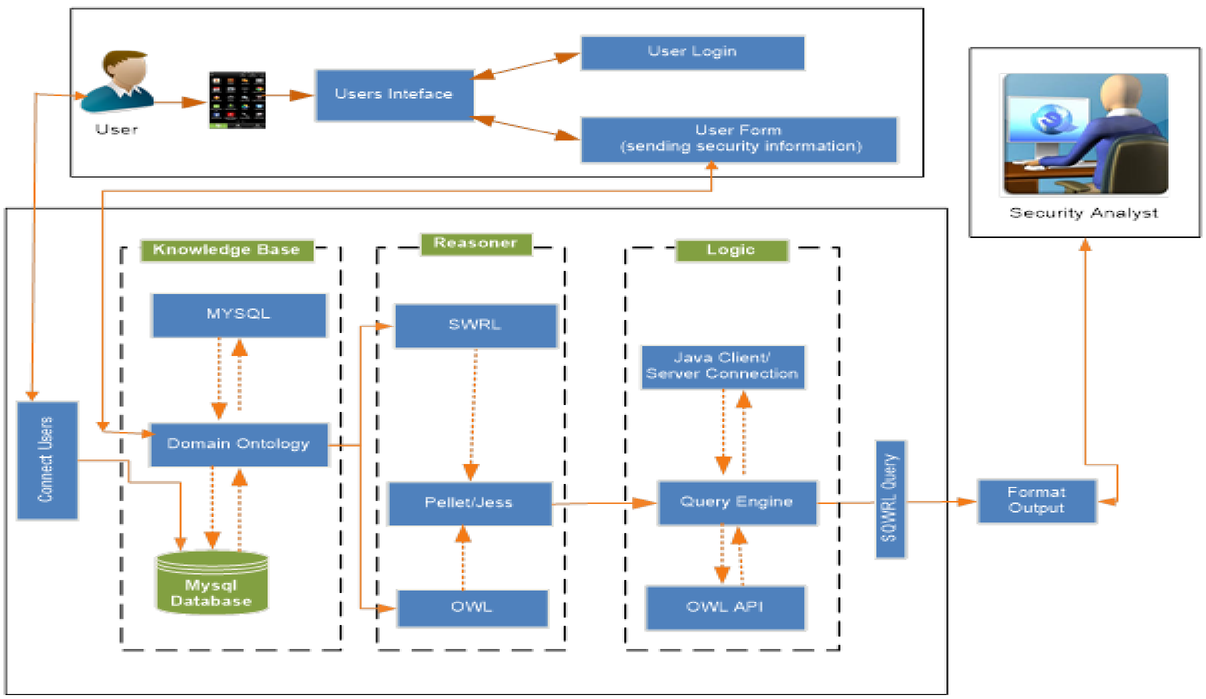

Figure 2: Model Architecture

3.2 System Components

The components of the system are depicted in the architectural diagram shown in Figure 2 and their detailed explanations are provided in subsections afterwards.

3.2.1 User Layer

User layer contains interface for user to login through the computer and or mobile device, so as to restrict access to only authenticated users. The form for security report is then forwarded to the Chief Security Officer (CSO) and a confirmatory message will be returned if submission is successful. As a prerequisite for sending a report to CSO, all users must be authenticated to ensure their valid identity.

3.2.2 Server Layer

Server layer contains three tiers namely; knowledge based (KB) tier, reasoner tier and logic tier. Each contains components that are interacting with one another or with another tier. The KB consists of MySQL and the domain ontology; both of them interact with the database server. The database holds all the information about the domain and the users. The reasoner consists of the SQWRL, the Jess or Pellet used for reasoning the ontology and the OWL, which is the defined domain ontology itself. The last component in the server layer is the logic tier. This part of the server ensures that the connection between the client and the server is established for communication between the two entities. The rule engine and the OWL API ensure that the implementation of SQWRL and SWRL rules written are established.

3.2.3 Security Analyst

The security analyst receives reported cases about the system and also run queries to ascertain the thematic, spatial and temporal element of the report before carrying out further investigation or action.

3.3 Ontology Design

Since the domain of this study is limited to security reporting and investiagtion model in university community, hence the design and description of the required ontologies that will be used restricted to the domain. Specifically, the developed ontology has three major entities which include students, staff and visitors to the university.

3.3.1 Steps in Designing Ontology

Some fundamental steps are required in designing ontology as highlighted by Mohan and Arumugam [12] who also suggested the following seven steps for developing and designing ontology:

i. Determine the domain and scope of the ontology.
ii. Consider the reuse of existing ontology.
iii. Enumerate important terms in the ontology.
iv. Define the classes and the class hierarchy.
v. Define the properties of the classes.
vi. Define the facets of the properties.
vii. Create instances.

3.3.2 Protégé

The seven highlighted steps were followed in creating the university ontologies using Protégé application package which is a free, open source ontology editor and a knowledge acquisition system. The package provides a graphic user interface to define ontologies by clicking tabs presented at the menu.

3.3.3 Design of Domain Ontology

In order to develop the ontology of university community with OWL for gathering intelligence information, this study determined the scope and domain of the ontology to be designed. Therefore, the domain comprised of students, staff and visitors, whereas the scope is the university. Thus, Figure 3 revealed the hierarchical view of the classes representing the domain.

3.3.4 Relationships between Classes and Individual of the Domain Ontology

Classes are the entities in the domain, while individuals are the instances of the classes. The relations (properties) are created to connect entities or classes and literal values. There are two types of properties (relations) that are used in OWL. These are DataProperty and ObjectProperty. Some of the DataProperties used here are hasAge, hasName, hasSex and livesIn while some of the ObjectProperties used are admitedTo, belongsTo, isStaffOf, isStudentOf, studiesIn, memberOf and teachesIn. Figures 4 and 5 show some of the ObjectProperties and DataProperties created and used in this ontology respectively.

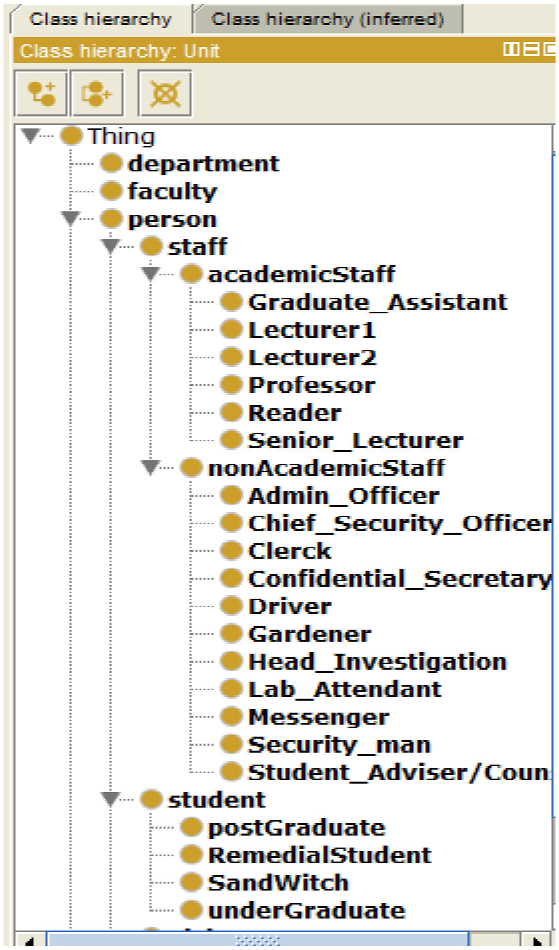

Figure 3: Domain classes and their hierarchy

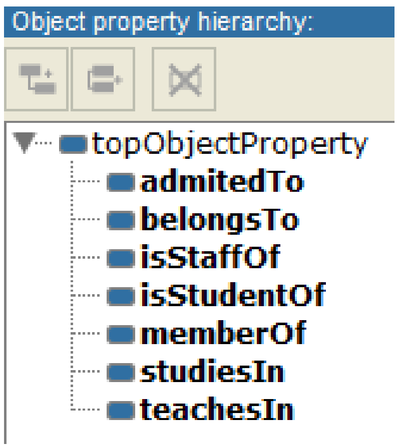

Figure 4: Object property hierarchy

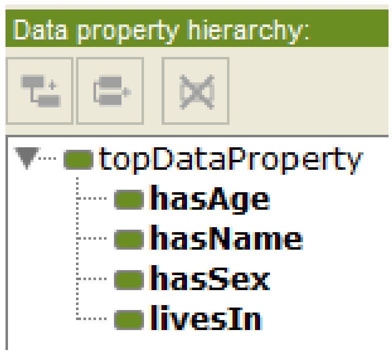

Figure 5: Data property hierarchy

3.3.5 OWL Visualization and Ontology Graph

One of the benefits of using protégé in designing ontology is the fact that it provides many easy to use and flexible plugins that simplify modeling activities. For example *OWLViz* plugin was used in this study to generate relationships between classes and subclasses for ontology as shown in Figures 6a and 6b. Similarly, *OntoGraph* plugin was used to build detailed graph for the classes and for the ontology as shown in Figure 7. In addition, a pictorial representation that shows all the classes, subclasses, individuals and their relationships is generated via a protégé plugin called *Jambalaya*. These graphs are presented as snapshots in Figure 8.

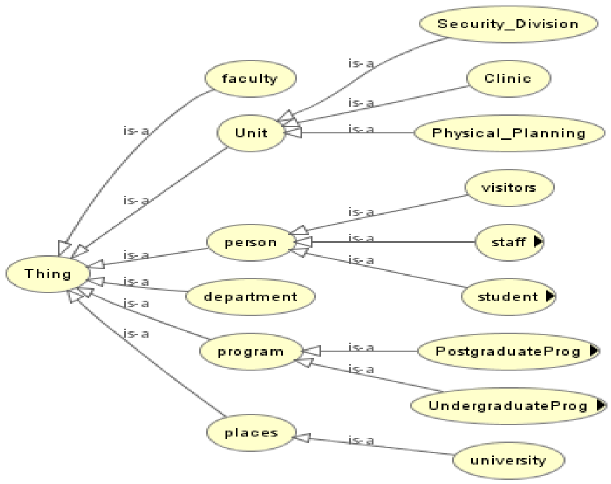

Figure 6a: Compact Classes and Sub-classes

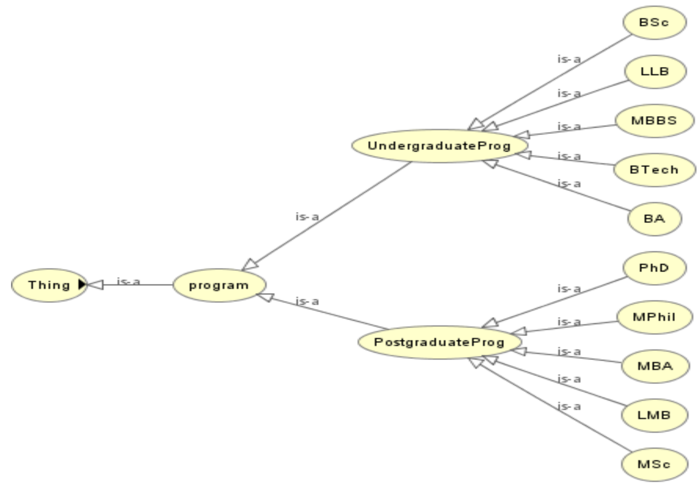

Figure 6b: Class Person and Sub-classes

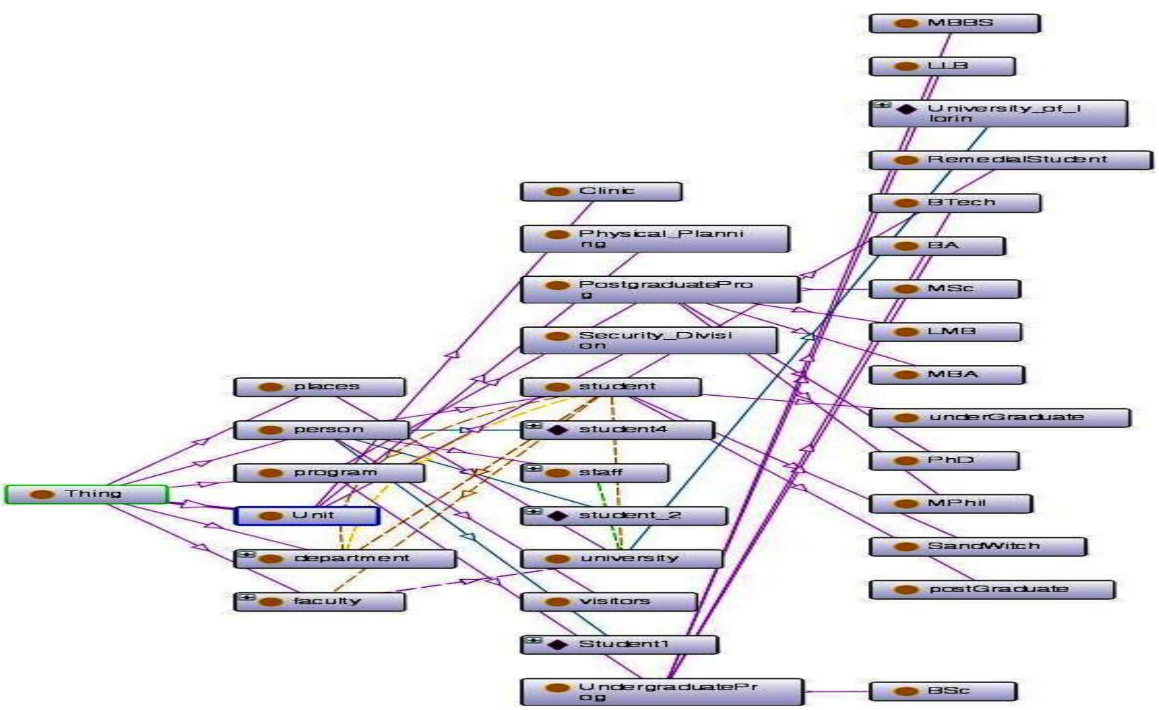

Figure 7: OntoGraph showing detailed classes and Individuals

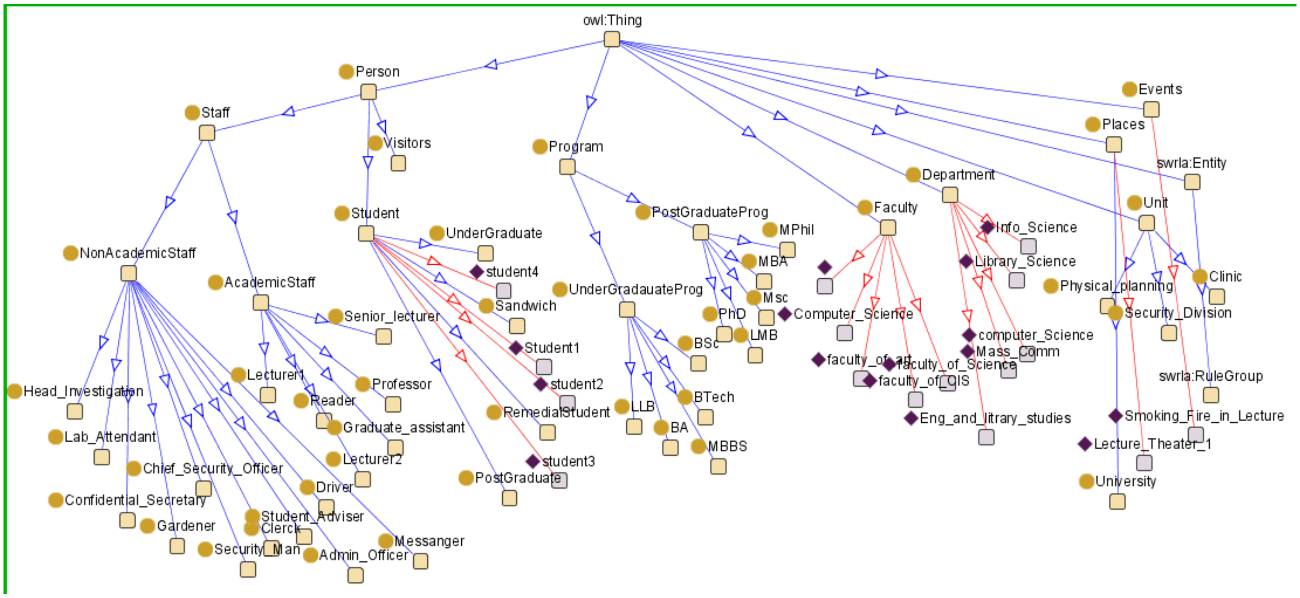

Figure 8: Jambalaya graphical representation of the ontology

4.0 Results and Discussion

The proposed model given in Figure 2 gives the building block for designing knowledge based system and particularly intelligence gathering system. As the basis for KB system, we have developed ontology in the domain of university, which can be adopted for other designs. Figures 3, 4 and 5 show the hierarchical construct of the classes, the data properties and object properties of those classes considered respectively. A better visualization of the designed ontology is shown in Figure 8, where the hierarchy starts from the root class **Thing** and flow down to other sub-classes. Figure 8 describes the full ontology based on the entities involved and their corresponding attributes for proper implementation.

5.0 Conclusion and Future Work

This paper presented a model for thematic, spatial and temporal intelligence gathering with a university ontology designed based on OWL technologies. This is a major contribution to knowledge as the ontology designed can be applied to solve any knowledge based problem in similar domain of discuss. Really, the domain of the ontology can be improved upon to extend to other public domains. Also, the implementation of the proposed design could be carried out to make intelligence gathering and investigation process more efficient. A rule based access control technique that enables communication between the client and server was developed and integrated in the model to allow users report security issues to a security agent or analyst.

References

[1] Albert, I.O. (2012). Explaining the Security Challenges in Contemporary Nigeria. Institute of African Studies, University of Ibadan.

[2] Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specifications. Journal of Knowledge Acquisition, Vol. 5, No. 2, pp. 199-220.

[3] Oyelade, O.N., Junaidu, S.B. and Obiniyi, A.A. (2014). Semantic Web Framework for E- Commerce Based on Web Ontology Language. International Journal of Computer Science, Vol. 11, Issue 3, No 2, pp: 145 - 152.

[4] Berners-Lee, T. (1994). Weaving the Web. MA, USA. The MIT Press, Cambridge.

[5] Perry, M., Sheth, A.P., Hakimpour, F. and Jain, P. (2007). Supporting Complex Thematic, Spatial and Temporal Queries over Semantic Web Data. 2nd International Conference on GeoSpatial Semantics. Mexico City, Mexico. Springer-Verlag Berlin.

[6] Wolfgang, G. (2010). Semantic Web. Retrieved June 25, 2015, from Semantic Web, Part of business world: http://www.semanticweb.rs/Article.aspx?iddoc=32&id=65&lang=2

[7] Cambridge, S. (2015). *Cambridge Semantics: The Smart Data Advantage*. Retrieved May 28, 2015, from Cambridge Semantics: http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web

[8] Dinger, U., Oberhauser, R. and Reichel, C. (2006). SWS-ASE: Leveraging Web Service-based Software Engineering. ICSEA '06 Proceedings of the International Conference on Software Engineering Advances. Washington, DC, USA

[9] Berners-Lee, T. and Hendler, J. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution possibilities. Scientific American.

[10] Zilberstein, S. and Lesser, V. (1996). Intelligent Information Gathering Using Decision Models. Computer Science Department, University of Massachusetts at Amherst. 96-35, Technical Report.

[11] Pari, D.H., Frada, B., Arkady, Z. and Paul, A. (2012). Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings. Elsevier.

[12] Mohan, R.A. and Arumugam, G. (2011). Constructing Railway Ontology Using Web Ontology Language and Semantic Web Rule Language.