A Methodology for Developing Onion Ontology in Nigerian Domain

Aliyu Rufai Yauri, Mustapha Abdullahi
Faculty of Engineering
Kebbi State University of Science and Technology, Aliero.

Email: rufaialeey@yahoo.com, almustapha.abdullahi@ssu.edu.ng

Abstract-*With the recent Nigerian government's effort for the diversification of its economy, efforts are put in place to revitalize Nigerian Agricultural sector. Part of those efforts is to incorporate technology in to the process of revitalizing of Nigerian agriculture. This paper proposed a methodology for developing ontology for Nigerian agriculture domain. Although few works attempted to develop Ontology of Nigerian agriculture domain, due to the complex nature of natural language the area is still subject to research. This paper proposed a methodology for developing effective Nigerian agriculture ontology development methodology. The proposed methodology will be tested using protogee ontology editor and OntoGraf is used for visualizing the developed ontology.*

Keywords: Ontology, Agriculture, Information Retrieval, Semantic

1. INTRODUCTION

Huge amount of heterogeneous data being deposited on the Web has increased the quest for effective information retrieval systems. Information retrieval systems such as Google, Yahoo has played a major role on the access of data on the Web. However these retrieval systems are based on traditional keyword search which lacks semantics and as a result, allot of irrelevant results are returned. To overcome the shortcomings of the traditional keyword search systems, the concept of the Semantic Web was introduced by the W3C consortium. In semantic web approach, data is given a well-defined format that models the meaning of information on the Web, as well as applications and services, so as to discover, annotate, process and publish data that is encoded in them (Zou, Finin, & Chen, 2004.). Data in semantic web is represented in ontology triple representation (RDF) format.

In simple terms ontology can be seen as objects that may exist in a particular domain and the relationships that may exist between those objects. Data represented ontology representation (RDF triple) format is stored in the knowledgebase in order to facilitate manipulation and querying semantically (Ciccarese, Ocana, Garcia, Sudeshna & Clerk , 2011). Therefore in order to build a knowledgebase for a particular domain, ontology for that such domain needs to be built.

Although several ontology development methodologies have been proposed by researchers, most of these methodologies are domain specific which cope with the challenges of individual domains. As a result different working group employs their own methodology (Fernandez Lopex, M, 1999). In this research, we intend to propose a new ontology development methodology on Nigerian Agriculture domain. Having standard ontology for Nigeria Agriculture will enables the development of various agriculture based applications that will contribute to the Nigerian agricultural system such as semantic agricultural information retrieval systems.

Although ontology has proven to be significant towards solving the challenges of traditional keyword search systems, most of the current ontology Methodology approaches are based on Manual approach using domain expert to identify ontology concepts and link the concept via various relationships. Manual ontology development requires domain expert to do the development of ontology manually which is tedious and time consuming. Additionally in term of ontology development, ontology development in Nigerian Agriculture domain is still at infant stage especially in agriculture domain. Work in (Syed Malek F. D Syed Mustapha and Emmanuel Ukpe, 2013) proposed ontology development methodology for Nigeria agriculture, but the paper has not clearly described their methodology and no proper evaluation of their methodology has been done to test the effectiveness of the methodology.

In the work, we proposed an ontology development methodology for Nigeria Agriculture domain. The proposed methodology will be based on Semi-automatic ontology development methodology for Nigerian Agriculture domain.

2.0 LITERATURE REVIEW

The main building block of the semantic web is ontology, which transforms web content into a machine-readable format that can be manipulated (Ahmed & Gerhard, 2007). Ontology

is the main building block of the semantic web which transforms web content into a machine-readable and format that can be manipulated (Ahmed & Gerhard, 2007). Ontology, in other words Web Ontology Language (OWL), is commonly defined as formal, and explicit specifications of shared conceptualization. Formal signifies ontology as a machine-readable format. Whereas, the concepts or entities used are explicitly described, shared, and displayed, ontology is concept that captures knowledge in a widely acceptable standard, and its conceptualization reflects ontology as a notion that identifies entities in the real world (Hu, 2004). In other words ontology can simply be seen as the study of entities that exist in the real world, and the things they have in common (Lawson, 2004). Ontology facilitates standards for integrating and sharing data in a conceptual schema. Objects, entities or concepts are identified and annotated with the relationships that exist between them.

In the concept of ontology, an entity or object is referred to as the same thing. This research will be using 'concept' to denote an entity or object, while 'relationship' is seen as the things concepts have in common, known as properties. Properties can be classified into object properties and data properties. Object properties represent the semantic relationship between concepts, while data properties define the relationship between a concept and its literals. Annotation of concepts enables better descriptions of the concepts in the form of metadata, facilitating greater meaning for human and machines to easily process and share.

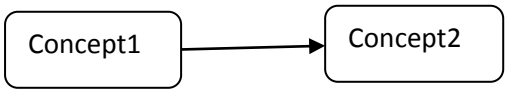

Figure 1: Example of object property

Figure 1 is a graphical representation of an object property, where semantic mapping between two concepts (Concept1 and Concept2) is provided. The semantic mapping gives a better description of the concepts.

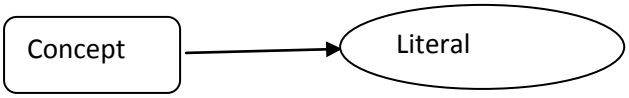

Figure 2: Example of data property

Figure 2 is a graphical representation of data property, where a concept is mapped with its literal. A literal is a mechanism for describing a concept itself. It gives an additional description of a concept. Figure 3 denotes an example of ontology representation for students.

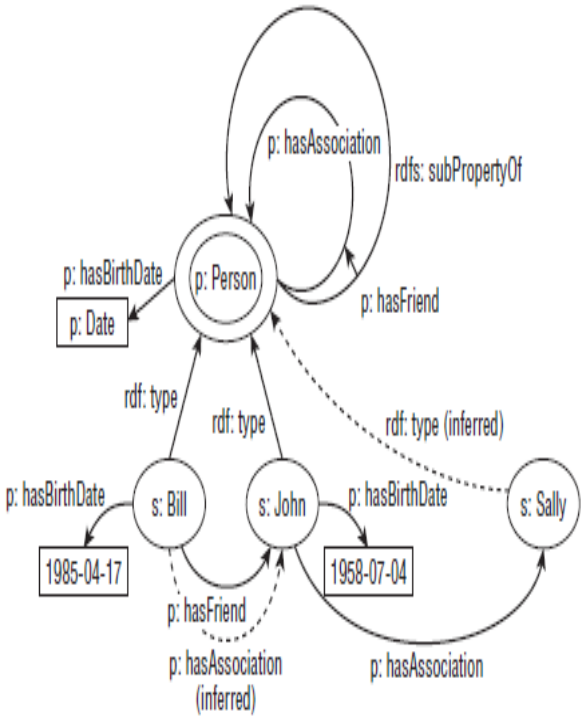

Figure 3: Example of student ontology representation

Ontology can be created automatically, semi-automatically, or manually (Erdmann, Maedche, Schnurr, & Staab, 2000). Automatic creation of an ontology involves using an automated tool to automatically generate the ontology from a domain (Balakrishna & Srikanth, 2008). Semi-automatic ontology creation involves a combination of human effort and automated tools (Balakrishna, Moldovan, Tatu, & Olteanu, 2013). Manual ontology is usually complex and time-consuming especially when dealing with a great deal of data (Ahmed & Gerhard, 2010). Manual ontology creation involves the design and creation of an ontology completely by a human expert (Tao, Embley, & Liddle, 2009).

Quite a number of researches have been presented by various researchers on ontology development methodology (Jean Vincent, Fonou-Dombeu, & Magda Huisman, 2011). An ontology development methodology is the process that describes various steps by step procedures that must be performed when building ontology. A domain expert analyses the domain and adds relationships by considering things such as what the ontology will cover, what the ontology will be used for, and the type of questions the ontology will serve. Relationships are added to enrich the ontology knowledge base that serves the semantic search systems. A domain expert therefore needs to look at the objectives of the project, i.e. the expectations of the functionalities

using a particular domain. The expert should also consider the domain boundaries by looking at the scope of the domain and considering what should be included and what should not be included in the descriptions added to the concepts. Potential sources of information should also be considered when adding description to the ontology concepts. The domain expert should consider the source of information that will be used to add relationships to the ontology concepts. The sources of information for ontology development can be online sources, textbooks, and articles among others, that are related to the domain. Comprehensive studies of these previously reported ontology development methodologies are provided in (Fernandez Lopex, M, 1999) ,(H. Beck and H.S Pinto,2003) , (C. Calero, F. Ruiz and M. Piattini,2006). However, it is important to know that each domain come with different challenges during the development of such domain ontology. Although different ontology development methodologies have been reported previously by different researchers, most of those methodologies are manually based which requires domain expert to analyze the date, identify domain ontology concepts, identify the relationship and link the concepts with relationship. This process is hectic and more venerable to human error. Additionally, in terms of Nigeria agriculture domain, to the best of our knowledge, we identify one research that focuses on ontology in Nigeria agriculture which a work in (Syed Malek F. D Syed Mustapha and Emmanuel Ukpe, 2013). In their work they reported ontology development methodology in Nigeria agriculture domain but, their work fails to clearly describe the methodology and shows prove that the ontology has been developed and evaluated. Therefore the effectiveness of their methodology cannot be guaranteed.

The primary objective of this study is to propose semi-automatic ontology development methodology in Nigeria agriculture domain which will combine both automatic and manual ontology development process.

3.0 THE PROPOSED RESEARCH METHODOLOGY

This section will provide a step by step procedure for ontology development in Nigerian agriculture domain. This research is based upon a case of onion crop in Nigeria. The proposed approach is semi-automatic whereby some part of the framework is carried out manually while other parts are done automatically by the system as presented figure 3.1.

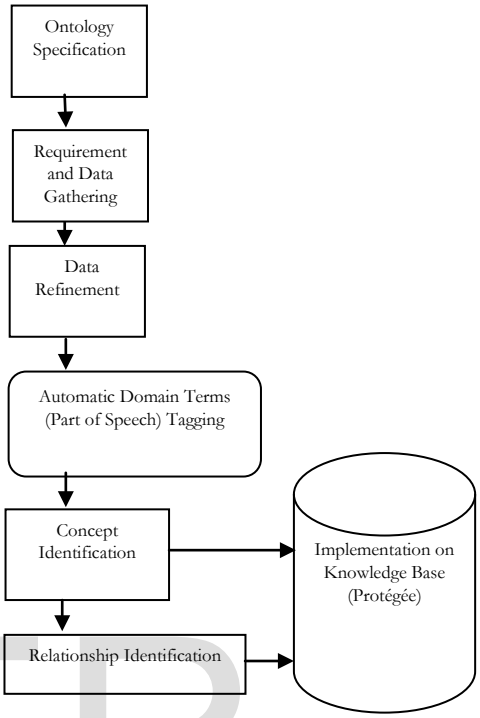

Figure 3.1: Framework of Onion Crop Production Ontology Development Methodology

Figure 3.1 shows the step by step procedures that will be involved in the implementation of this research. The research task for the Development of the onion crop production ontology is divided into the following stages;

- Ontology specification
- Data and requirement gathering
- Data Refinement
- Automatic Domain Terms (Part of Speech) Tagging
- Concept Identification
- Relationship Identification
- Implementation on Knowledgebase (Protégée)

The first step of the proposed methodology is Ontology specification. In this stage domain and scope of the ontology are clearly defined. The onion crop production ontology was started by defining its domain and scope. The ontology will cover onion crop production from land preparation to harvesting.

The second step is Data and Requirement Gathering (Acquiring Domain Knowledge). Here, a deep insight into and a thorough knowledge of the respective domain is prerequisite to construction of any domain ontology. Secondary data was used to generate data about the onion domain.

The next step is Data Normalization and Refinement. This stage also involved the data formalization and organization such that all redundancy is removed and avoided for better presentation. Also it involves making sure that all critical data is not missing.

After normalization, the next step is Automatic Domain Terms (Part of Speech) Tagging. A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

From the domain tagged token an automatic Concept Identification is carried out to ascertain terms that are concepts in the onion domain .Concepts that represent the domain are identified and hence implemented by means of classes. This is the stage in which the conceptual model of the ontology is built.

Haven identified concepts of the onion domain, the next thing it identify relationship in order to get triple representation (subject, predicate, object). **Relationships Identification** identifies relationship between concepts in the onion domain

Lastly, after all concepts in onion domain are identified with their corresponding relationship, we now have onion ontology from a row onion domain data. The developed onion ontology is now stored in Knowledgebase. (Protégé). A knowledge base is a repository that contains data derived from machine-readable format such as RDF triple. A knowledgebase could be a database where knowledge is stored. An example of a knowledgebase is DBpedia. In this research, protégée ontology editor is used to store the onion ontology

4.0 IMPLEMENTATION

In this section, the developed onion ontology is stored in the protégée ontology editor. Ontograph is used to visualize the developed ontology. OntoGraf is plugin in protégée that enables visualizing ontology stored on a protégée in graphical representation. Figure 4.1 presents visualization of the developed onion ontology.

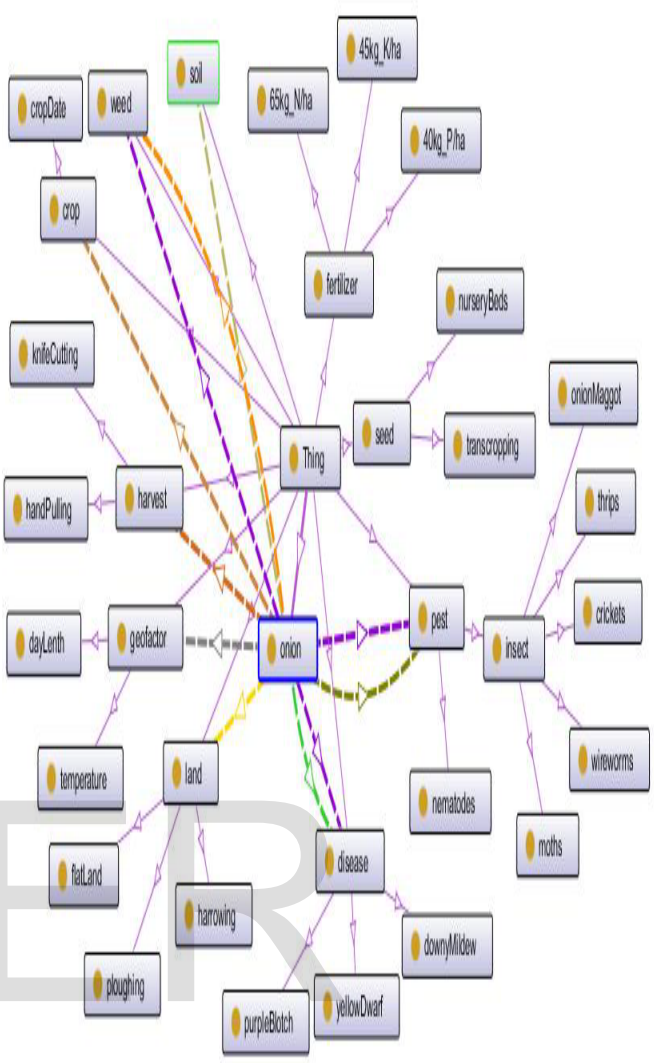

Figure 4.1: Visualization view of the Ontology Using OntoGraf Plugin

5.0 CONCLUSION

This research has presented a methodology for developing onion ontology in Nigeria. The research has demonstrated that ontology plays a critical role in knowledge acquisition and knowledge management processes. It helps make knowledge storage and retrieval process significantly more intelligent. So it is very reasonable to encourage the construction of many more ontology of different Agriculture domain in Nigeria. In the next phase of this research, a retrieval mechanism for the ontology will be focussed on.

REFERENCE

1. Zou, Y., Finin, T., & Chen H. (2004). F-OWL: An inference engine for the semantic web, Proceedings of the Third International Workshop (FAABS), April 16–18, 2004, USA.
2. Ciccarese, P., Ocana, M., Garcia Castro, L. J., Das, S., & Clark, T. (2011). An open annotation ontology for science on web 3.0, Journal of Biomedical Semantics, 2041-1480-2-S2-S4.
3. Ahmed, Z., & Gerhard, D. (2007). Web to Semantic Web & Role of Ontology, In the proceedings of National Conference on Information and Communication Technologies, (NCICT-2007), Pakistan, 9th May 2007.
4. Lawson, T. (2004) A Conception of Ontology, University of Cambridge, Faculty of Economics, Cambridge.
5. M. Erdmann, A. Maedche, H. Schnurr, and S. Staab. (2000). From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In, Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content, August 2000.
6. Balakrishna, M., & Srikanth, M. (2008). Automatic Ontology Creation from Text for National Intelligence Priorities Framework (NIPF).In Proceeding of 3rd International conference for Ontology for the Intelligent Community (OIC), Fairfax, VA, USA, Dec 3-4.
7. Tao, C., Embley, D. W., & Liddle, S. W. (2009). FOCIH : Form-based Ontology Creation and Information Harvesting. In Proceedings of the 28th International Conference on Conceptual Modeling , Gramado, Brazil.
8. Syed Malek F. D Syed Mustapha and Emmanuel Ukpe, (2013). Agriculture Ontology for Sustainable Development in Nigeria. Journal of Advances in Computing 2013, 3(3): 57-59 DOI: 10.5923/j.ac.20130303.04
9. Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman and Masrah Azrifah Azmi Murad(2013) Quranic Verse Extraction base on Concepts using OWL-DL Ontology. Research Journal of Applied Sciences, Engineering and Technology 6(23): 4492-4498, 2013
10. Jean Vincent, Fonou-Dombeu, & Magda Huisman, (2011). Combining Ontology Development Methodologies and Semantic Web Platforms for E-government Domain Ontology Development. International Journal of Web & Semantic Technology (IJWesT) Vol.2, No.2, April 2011