



Building Ontology for Nigerian Tribes and Languages

Ebietomere E.P.^{1,*}, Aghaunor C.T.² and Ekuobase G.O.³

^{1,2,3} Department of Computer Science, University of Benin, Benin City, Edo State

*Corresponding author's e-mail: princewill.ebietomere@uniben.edu

Article Information

Keywords:

Ontology engineering, knowledge representation, semantic web, Nigerian languages

Received 25 June 2019

Revised 08 July 2019

Accepted 11 July 2019

Available online 12 August 2019

ISSN-2682-5821/© 2019 NIPES Pub.

All rights reserved

Abstract

The cultural heritage of the Nigerian people is gradually going extinct as the growing generation is being brought up to speak foreign languages at the neglect of their native or indigenous language. Native languages which remain the most viable cultural identity factor must therefore be preserved. Preservation of our culture definitely includes the preservation of its domain knowledge. Ontology is a veritable tool in computing for representing, and thus preserving, domain knowledge for both human and machine processing. This paper builds a foundational ontology for Nigerian languages and tribes (ONLT). ONLT was built using the Noy and McGuinness methodology and specifically created using the Protégé 3.5 OWL editor. The ONLT was built on, and evaluated by queries; using a set of competency questions. The ONLT was found to strongly align with its purpose of representing domain knowledge and thus provides required information on Nigerian tribes, and languages.

1. Introduction

Communication is a vital part of human life. Without communication, humans cannot interact with one another and for communication to be established, it must be done in a certain language understood by the parties involved. The author [1] in his “Philosophy of style” states that “Language is not the invention of yesterday; it is one of the most precious heirlooms bestowed by the divinity at the moment of creation, ‘IN PRINCIPIO ERAT VERBUM’ – in the beginning was the word”. Languages are important vehicles for the preservation of culture.

Nigeria is a country with rich and diverse ethnic ethos encapsulated as tribe usually with unique values, food and language such that a tribe is usually named after their unique native language. Native language is the most important factor of cultural identity, in distinguishing one tribe from another [2]. The gradual erosion of this ethnic ethos may result in moral turpitude and loss of ancestral identity. Sadly, as the growing generation in Nigeria is being brought up to speak the English language with little or no regard for their native languages [3–4], the extinction of Nigeria’s ethnic ethos is imminent.

Communicating via the renowned English language (as well as French and other foreign languages) is good as it enhances and promotes intra-and-inter-country communication for social, economic and political benefits amongst others. This notwithstanding, foreign languages should not be imbibed at the detriment of the Nation’s native languages. It is never too late to begin mother tongue learning and reading [5]. Irrespective of civilization, the Nigeria people should hold in high esteem their mother tongue like the Asian countries and the Germans; indigenous languages must be preserved. Having similar fate as the indigenous languages are Nigeria native food, cultural values and beliefs. Ontology is a veritable tool for achieving this preservation of

Nigeria's ethnic ethos. Ontology is a defacto way the content of the web can be marked up for automatic information processing [6–7].

This paper picked particular interest in languages, tribes and food because they are not in conflict with any religious or legal standing and as such, builds a foundational ontology for the Nigerian tribes, languages and food to aid the preservation and promotion of Nigeria's ethnic ethos towards incorporating the Nigeria's culture into the true vision of Semantic Web – comprehension by both man and machine.

1.1 Concept of Ontology

Ontology stores information (i.e. data and it's meaning within a domain) while database stores data without its meaning; and thus unlike database, ontology is both human and machine comprehensible. Ontology is particularly useful in this era of information overload and knowledge management for enhanced information processing, comprehension and application; as it enables cooperation between man and machine (computer) in the processing and application of information as knowledge.

Ontology as a concept has its root in Philosophy, where it is seen as “the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality” [8]; it is from here the term ontology was borrowed to be used in Computer Science and Artificial Intelligence (AI), in particular. The most prominent definition of the term in AI is that of [9] who defined ontology as an “explicit specification of a conceptualization”. A latter but more complete and most referred to definition of the term ontology is that given by [10], that “an ontology is a formal, explicit specification of a shared conceptualization”. The work of [11] made it clear that “ontology may take a variety of forms, but necessarily it will include a vocabulary of terms and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms”. This paper stuck to the definition of [12], where ontology is seen as “a formal explicit description of concepts in a domain of discourse (classes), properties of each concept describing various features and attributes of the concept (slots) and restrictions on the slots (facets)”; because the definition appears more explicit and realistic [13].

1.1.1 Classification of Ontology

Several ontologies exist in literature [14-17]. Generally, we can classify ontology based on purpose or commitment, formality or complexity, scope of object described by the ontology etc., but the most pronounced is the classification by purpose or commitment [13]. The work of [18] describes three types of ontological commitments – task commitment, method commitment and domain commitment.

1.1.2 Building Ontology

There are several reasons for building ontology. Some of which are as follows [12] and [19]:

- i. share common understanding of the structure of information among people or software agents.
- ii. enable reuse of domain knowledge.
- iii. make domain assumptions explicit.
- iv. separate domain knowledge from the operational knowledge.
- v. analyze domain knowledge.

Today, several standard methodologies exist for the development and evaluation of ontology. The authors [12] made it clear that there is no single correct ontology design methodology, hence, which to follow is a matter of choice of the ontology creator. Some of these methodologies include: Noy and McGuiness' methodology, Lenat and Guha's methodology, Gruninger and Fox's methodology, KATUS methodology and Methontology [12], [13], and [20–24]. In using any of

these methodologies, we can employ any of the three basic approaches for rendering ontology; top-down, middle-out and bottom-up.

This paper, employed the Noy and McGuinness' methodology and the top-down approach for simplicity and reliability sake [13], [25].

1.1.3 Ontology Languages

Ontologies are state-of-the-art knowledge base component of modern information systems [18], [26]. Consequently, a number of representational formats have been proposed to support and express them completely. Ontology languages can be classified into three [27], namely: (i) Traditional ontology – CycL, OCML, Loom, and Telos; (ii) web standards – XML and RDF; and (iii) web based ontology languages – OIL, DAML+OIL, SHOE, OWL, RDFS and XOL .

Of all the languages for rendering ontology, the Web Ontology Language (OWL) which is the standard ontology language recommended by the World Wide Web Consortium (W3C), was used in rendering this ontology for tribes and languages in Nigeria. Basically, there are three OWL sub-languages [28]; OWL Lite, OWL DL and OWL Full; with OWL Lite being the least expressive and OWL Full being the most expressive. Moreover, OWL DL is the only sublanguage of OWL that allows for consistency check, hence, OWL DL was employed in this paper.

Since this paper builds a foundational ontology for the Nigerian people and culture, it is expedient we give brief background information about the Nigerian tribes and languages highlighted in section 2.2.

1.2 Brief History of Nigerian Languages

Long before 1500, much of modern-Nigeria was divided into states identified with contemporary Ethnic groups. These early States included tribes like the Yoruba, Igbo, Hausa, Benin, and the Nupe. Additionally, numerous small States to the west and south of Lake Chad were absorbed or displaced. Other states probably existed but the absence of archaeological data prevents accurate dating. The three largest and most dominant ethnic groups are the Hausa, Yoruba and Igbo. Other smaller groups include the Ijaw, Ibibio, Tiv and Edo.

Contemporary Nigeria is divided into the north, south, east, and west. The North is popularly known as the Hausa speaking part of the Nation and there we have states like Kano, Katsina, Sokoto, FCT (Abuja), Adamawa, Bauchi, Kebbi, Plateau, Benue and some parts of Kogi state. They all have diverse tribes with unique languages but the official language in most of the northern state is Hausa. Tribes found in the North include Abaji, Bwari, Bele, Barke, Egede, Etolu, Idoma, Tiv, Mulgwa, and Babur.

Southern Nigeria comes with a mix up of tribes and languages. It is not actually dominated by any spoken language in particular and the states found here include: Edo, Delta, Akwa Ibom, Bayelsa, Cross river and Rivers. These states are commonly referred to as the Niger Delta states. Tribes found in the South include Ibibio, Eket, Biseni, Ekpeye, Oruma, Abayon, Adim, Ijaw, Anioma, Isoko, Benin, Ebirra, Abua, and Ebana. Historically the Yoruba Language has been the dominant language on the west bank of Nigeria. In the West we have states like Ekiti, Kwara, Ogun, Lagos, Ondo, Osun and Oyo. Tribes found include Ebirra, Egun, Awori, Ogori, and Yagba. The Eastern part of Nigeria is characterized by the Igbo speaking people. States found here include Abia, Anambra, Ebonyi, Enugu and Imo. Tribes are Igbo and Mbembe. Figure 1 depicts the distribution of principal Tribes and Languages in Nigeria.

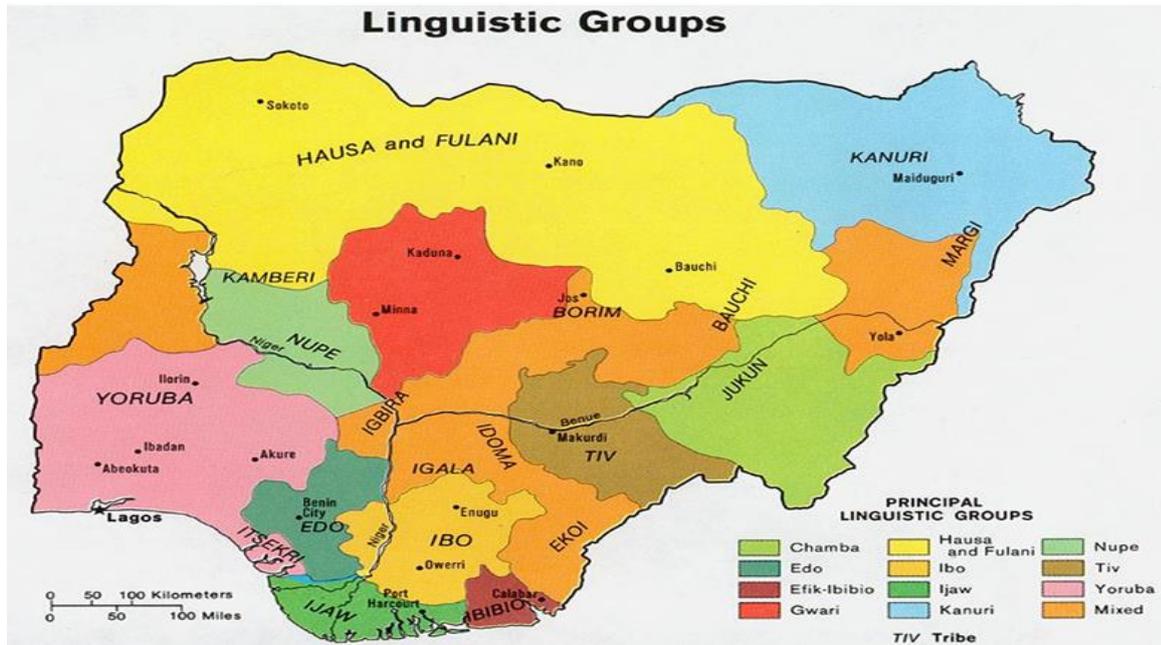


Figure 1: The Distribution of Principal Tribes and Languages in Nigeria

1.2.1 Meaning of Language and Tribe

Different definitions have been adduced to the term, language viz:

- i. A language consists of symbols that convey meaning, plus rules for combining those symbols, that can be used to generate an infinite variety of messages [29].
- ii. A language is a system of arbitrary vocal symbols by means of which a social group cooperates [30].
- iii. Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntary produced symbols [31].

Indigenous languages are important vehicles for the preservation of our culture. These languages with their various cultures and traditions originated from our forefathers and are spoken by different tribes today, each tribe with a unique language. With one's spoken native language, you can easily tell where a person is from in Nigeria.

Culture is the totality of learned, socially transmitted customs, knowledge, material objects and behavior. It includes the ideas, values, customs and artefacts of a group of people [32]. Culture can be transmitted or acquired through information or symbol. Cultural identities are those attributes, behavioral patterns, lifestyles, social structures and norms that distinguish a set of people from another [33].

Tribe is often seen as a group of people of the same race, and with the same customs, language, religion, values, etc., living in a particular area and often led by a chief. It could also be seen as the social division in a traditional society consisting of families or communities linked by social, economic, religious, or blood ties, with a common culture and language, typically having a recognized leader.

2. Methodology

There are several editors in existence for rendering ontology, some of which are; Protégé, ontoedit, Chimaera, Ontolingua, WebODE, OilED, Ontolingua, pOWL and SWOOP and OntoGen [34-36]. Protégé was chosen amongst these editors because it is [13]: (i) an open source ontology editor and knowledge base framework (ii) supported by a strong community of developers ranging from the academics and government officials to corporate users for knowledge solutions in areas like biomedicine, intelligence gathering and corporate modelling (iii) based on Java and provides a

plug-and-play environment that makes it a flexible base for rapid prototyping and application development (iv) platform independent and supports Resource description Framework (RDF), Web Ontology Language (OWL) and Extensible Mark-up Language (XML). Specifically, Protégé OWL was used for modeling the ontology for expressivity and consistency sake. Our codes are in OWL (RDF/XML) format because of its popularity in rendering ontologies [24].

The ontology was built using the Noy and McGuinness' methodology [12] and evaluated using a set of competency questions. Competency questions, like functional software requirements, define the ontology task and are usually thought out at the commencement of ontology construction. The relevant information used in building the ontology for tribes and languages in Nigeria were captured, from literatures and consultation with some language experts in our locality.

The steps involved in the Noy and McGuinness' methodology are as follows:

- Step I:** Determine the domain and scope of the ontology
- Step II:** Consider reusing existing ontologies
- Step III:** Enumerate important terms in the ontology
- Step IV:** Define the classes and the class hierarchy
- Step V:** Define the properties of the classes
- Step VI:** Define the value of the slots
- Step VII:** Create instances

The formulated competency questions enabled us to keep focus on the purpose of the ontology and as such, used both to ensure proper design and evaluation of the ontology. The formulated competency questions are:

- 1) What are the Geopolitical Zones in Nigeria?
- 2) What are the states in Nigeria?
- 3) What are the states in each geopolitical zone?
- 4) What are the foods in each geopolitical zone?
- 5) What are the tribes in each of the state?
- 6) What are the languages in each of the state?

Furthermore, the high level design of the ontology was modeled using a Directed Acyclic Graph (DAG) as shown in Figure 2.

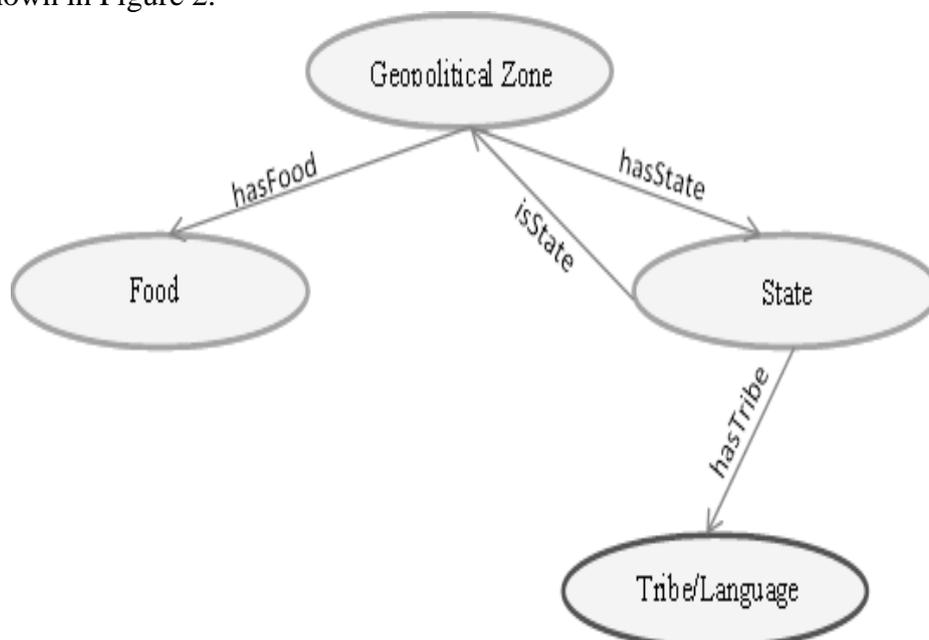


Figure 2: High Level Design of ONLT

From Figure 2, Geopolitical Zone, Food, and State are the major concepts of the ontology. Tribe/Language was defined as objects of the concept “State”. The concepts in the ontology are related by the hasFood, hasState, isState and the hasTribe relations.

The concepts and the properties of ONLT as implemented in Protégé 3.5 beta are as shown in Figure 3a and 3b respectively. Figure 4 shows the Domain/Range of the ONTL and the class/individual tree of the ONLT is as shown in Figure 5.

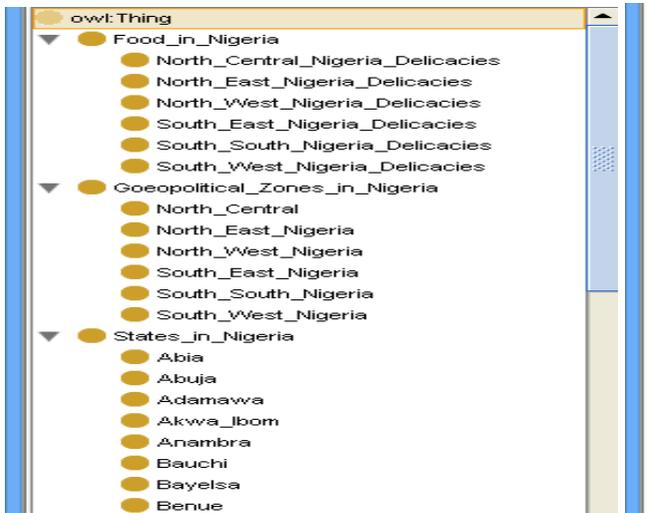


Figure 3a: ONLT concepts

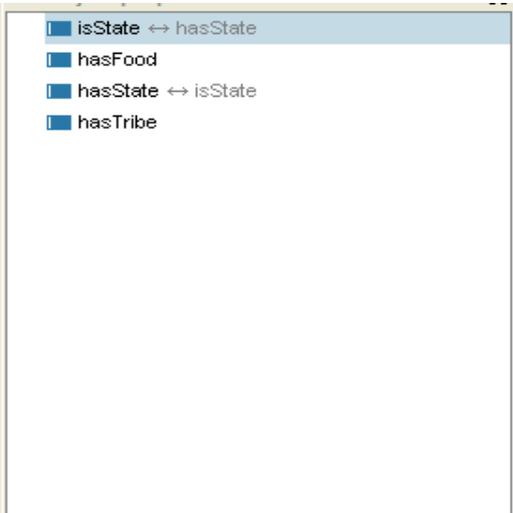


Figure 3b: ONLT properties

Every owl ontology in Protégé starts from a root usually denoted by “owl: Thing”. The concepts; Food_in_Nigeria, Geopolitical_Zones_in_Nigeria and States_in_Nigeria are subsumed by the superclass “owl:Thing” and each of these concepts in turn are with their subclasses as evident in Figure 3a. Figure 3b shows the properties relating the concepts of the ONLT ontology and it is obvious therefrom that the property “isState” is the inverse of hasState.

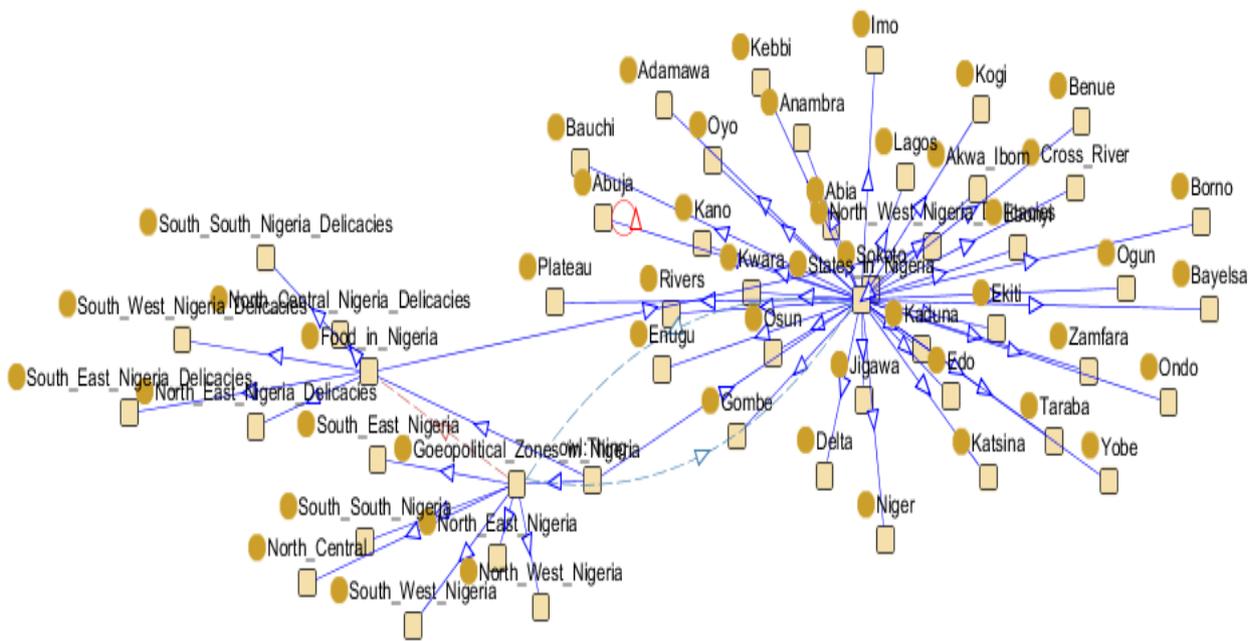


Figure 4: ONLT Domain/Range

The Domain/Range is usually used to depict which concept plays the role of a subject and which is the object at a given instance since ontologies are usually expressed as triples (i.e. subject, predicate, object). The domain is the subject, properties stands as the predicate and range stands as the object. From Figure 4, it is evident for example that, South_South_Nigeria_Delicacies (subject) is_a (predicate) Food_in_Nigeria (object).

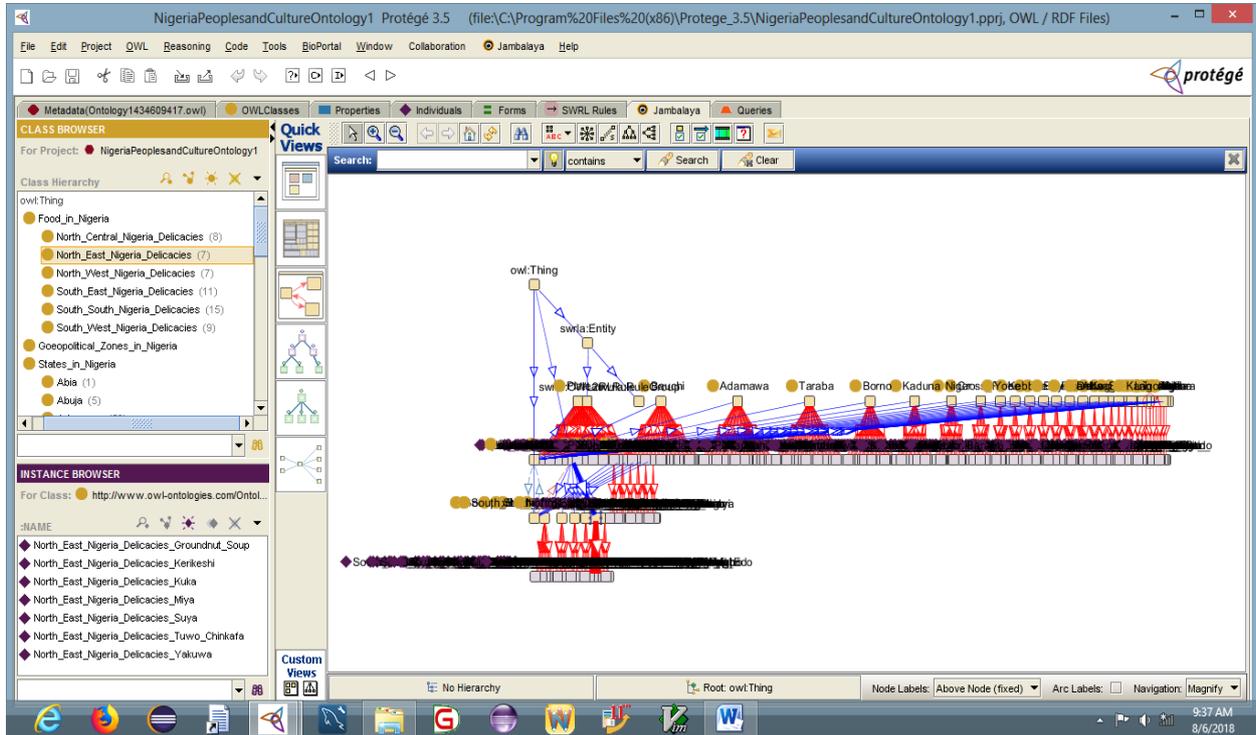


Figure 5: ONLT Class and Instance Tree

Figure 5 is a visualization of class and individual tree defined in the ONLT. This tree simply shows the hierarchical relationships among the concepts of the ontology and the instances of the concepts.

3. Result

Standard ways of presenting ontology results include exposing the intricacies (such as concepts, properties, domain, range, class and instance) of the ontology built [13, 14, 37]; as exposed in section 3. This approach is now often complemented by querying the ontology using a set of competency questions that captures the ontological commitment [12, 17, 37]. This new way of presenting ontology results is becoming popular because it clearly exposes the correctness of the ontology towards its purpose in addition to being lucid to human, particularly the non-experts.

As with similar researches [17, 37], we queried ONLT in line with the competency questions as shown in section 3 using Protégé tool query and export tab plug-in. A few of these queries and their results were captured in Figure 6a, Figure 6b and Figure 6c.

Query 1: what are the tribes in Akwa-Ibom state?

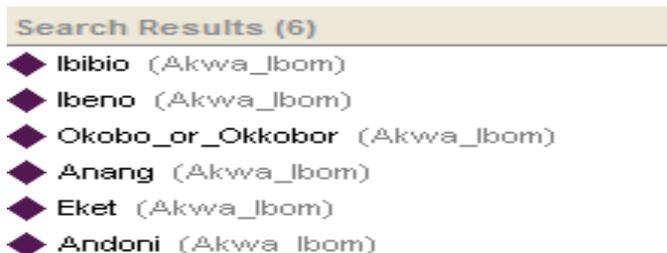


Figure 6a: Search results for query 1

Query 2: What are the native languages in Yobe state?



Figure 6b: Search results for query 2.

Query 3: what are the delicacies in South-South Nigeria?



Figure 6c: Search results for query 3.

The results obtained from the queries were found to be correct. For instance, as evident from Query 1 (Figure 6a), it is correct that the tribes in Akwa Ibom State include Ibibio, Ibeno, Okobo, Anang, Eket and Andoni. Thus, ONLT has been shown to be in strong alignment with its purpose of providing specific information about the Nigerian people and culture.

4. Conclusion

The problem of the gradual erosion of indigenous tribes and languages of the Nigerian peoples is severe. In preserving this ethnic ethos, ontology was identified as a veritable tool as it allows for machine and human process-ability of this ethnic ethos. Hence, ONLT for Nigerian indigenous tribes and languages, hitherto not existing was built using the Noy and McGuinness' methodology. The ONLT was modeled using Directed Acyclic Graph (DAG) and rendered using the Web Ontology Language (OWL) via protégé 3.5 beta editor. The ONLT was visualized using the Jambalaya plug-in of Protégé editor and its knowledge base queried using the query tab of Protégé editor in line with the competency questions formulated. The query results were apt, an indication that ONLT is in strong alignment with its purpose of providing specific information about the Nigerian people and culture.

References

- [1] Spencer, H. (1945). Philosophy of Style”, In G. Ehusani, “Language and the Nigerian Project. A paper presented at the 2nd National Conference of the School of Languages, Federal College of Education, Okene, May 3, pp. 1-7, 1945.
- [2] Ehusani, G. (2005). Language and the Nigerian Project. A paper presented at the 2nd National Conference of the School of Languages, Federal College of Education, Okene, May 3, pp. 1-7.
- [3] Akande, W. (2002). The Drawbacks of Cultural Globalization. *Yellow Times*, November 10, 2pp.
- [4] Ojo, G. (2010). Preserving our Local languages. Thisday Live. 5th July, 3pp.
- [5] Krashen, S. (2004). The Power of Reading: Insights from the Research. Heinemann, Portsmouth, Retrieved online at <http://esl.fis.edu> on 12/12/2015.
- [6] Setchi, R. M. and Tang, Q. (2007). Concept Indexing using Ontology and Supervised Machine Learning. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*. Vol. 1, No. 1, pp. 89-94.
- [7] Maedche, A. and Staab, S. (2002). Measuring similarity between ontologies. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, EKAW02*, Springer Verlag, LNAI. Berlin, 2473:pp. 251-263.
- [8] Smith, B. (2003). *Ontology*. Blackwell Guide to the Philosophy of computing and information, Oxford, Blackwell, pp 1-19.
- [9] Gruber, T. R. (1993). *A Translation Approach to Portable Ontology Specifications*. Stanford University, Stanford, California, pp. 1-23.
- [10] Studer, R., Benjamins, V. R. and Fensel, D. (1998). *Knowledge Engineering: Principles and Methods*. *Data Knowledge Engineering*, Vol. 25, No. 1, pp. 161-197.
- [11] Uschold, M. and Jasper, R. (1999). A Framework for Understanding and Classifying Ontology Applications. *The Proceedings of the IJCAI-99 Workshop On Ontologies and Problem-Solving Methods (KRR5)*, Stockholm, Sweden, 12pp.
- [12] Noy, N. F. and McGuinness, D. L. (2001). *Ontology Development 101: A guide to Creating Your First Ontology*. Stanford University, Stanford, CA, USA, pp. 1-25.
- [13] Ekuobase, G. O. and Ebietomere, E. P.(2013). *Ontology for Nigeria Case Laws*. *African Journal of Computing and ICTs*. Vol.6, No.2, pp.177-194.
- [14] Breuker, J., Elhag, A., Petkov, E. and Winkels, R.(2002). *Ontologies for Legal Information Serving and Knowledge Management*. *Legal Knowledge and Information Systems, Jurix '02: The fifteenth Annual Conference*, Amsterdam, IOS Press, pp. 73-82.
- [15] Breuker, J., Hoekstra, R., Boer, A., Berg, K., Sartor, G., Rubino, R., Wyner, A. and Bench-Capon, T. (2007). *OWL Ontology of Basic Legal Concepts (LKIF-Core)*. *ESTRALLA*, 138pp.
- [16] Guarino, N. (1998). *Formal Ontology and Information System*. IOS Press, Amsterdam, 13pp.
- [17] Ekuobase, G. O. and Ebietomere, E. P. (2016). *Ontology for Alleviating Poverty among Farmers in Nigeria*. *ACM International Conference on Informatics and Systems (INFOS'16)*, Cairo University, Egypt, ACM, pp. 28-34.
- [18] Visser, P. R. and Bench-Capon, T. J. (1998). *A Comparison of Four Ontologies for the Design of Legal Knowledge Systems*. *Artificial Intelligence and Law*, Kluwer Academic Publisher, pp.27-57.
- [19] Bench-Capon, T. J. and Visser, P. R. (1997). *Ontologies in Legal Information Systems: The Need for Explicit Specification of Domain Conceptualisations*. *ICAIL*, Melbourne, pp. 132-141.
- [20] Gruninger, M. and Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*. *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, 9pp.
- [21] Berneras, A., Laresgoiti, I. and Corera, J. (1996). *Building and Reusing Ontologies for Electrical Network Applications*. In C. N. Caralt, “*Modelling Legal Knowledge through Ontologies*. *OPKJ: the Ontology of Professional Judicial Knowledge*”, Ph.D. Thesis, Departament de Ciencia Política i Dret Public, Universitat Autònoma De Barcelona, 527pp.
- [22] Gavrilova, T., Farzan, R. and Brusilovsky, P. (2005). *One Practical Algorithm of Creating Teaching Ontologies*. University of Pittsburgh, USA, 9pp.
- [23] Chaware, S. and Rao, S. (2010). *Integrated Approach to Ontology Development Methodology with Case Study*. *International Journal of Database Management Systems*, Vol.2, No.3, India, pp. 13-19.
- [24] Sanchez, M. F. (2009). *Semantically Enhanced Information Retrieval: An Ontology-based Approach*. Ph.D. Thesis, Escuela Politécnica Superior, Universidad Autónoma De Madrid, 254pp.
- [25] Venturi, G. and Montemagni, S. (2012). *Ontology Learning in Legal Domain: An Introduction*. Retrieved online at: http://summerschoollex.cirsfid.unibo.it/wp-content/uploads/2012/09/Ontology_Learnin_in_the_legal_domain_an_introduction.pdf, 17/07/2014, 80pp.
- [26] Caralt, C. N. (2008). *Modelling Legal Knowledge through Ontologies*. *OPKJ: the Ontology of Professional Judicial Knowledge*. Ph.D. Thesis, Departament de Ciencia Política i Dret Public, Universitat Autònoma De Barcelona, 527pp.
- [27] Su, X. and Iiebrekke, L. (2002). *A Comparative Study of Ontology Languages and Tools*. Springer, Verlag, pp. 3-12.
- [28] Horridge, M., Jupp, S., Moulton, G., Rector, A., Stevens, R. and Wroe, C. (2007). *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.1*. University of Manchester, UK, 103pp.
- [29] Weiten, W. (2007). *Psychology: Themes and Variations*. Wadsworth Publishing, 7th ed., 880pp.
- [30] Bloch, B. and Trager, G. (1942). *Outline of Linguistic Analysis*. Waverly Press, 84pp.
- [31] Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Harcourt Bruce and Company, pp. 1-22.
- [32] Schaefer, R.T. (2002). *Sociology: A brief Introduction*. 4th ed. Boston: McGraw Hill.
- [33] Omekwu, C. O. (2003). *Information technology revolution, libraries and cultural values: Issues, impacts and inevitable challenges for Africa*. 69th IFLA General Conference and council, Berlin, August 1-9.

- [34] Fortuna, B., Grobelnik, M. and Mladenic, D. (2007). OntoGen: Semi-automatic Ontology Editor. In: M. J. Smith and G. Salvendy (Eds.), *Human Interface*, LNCS 4558, Springer-Verlag, Berlin, Heidelberg, pp. 309-318.
- [35] Escorcio, L. and Cardoso, J. (2007). *Editing Tools for Ontology Construction*. In: *Semantic Web Services: Theory, Tools and Applications*, Idea Group, 27pp.
- [36] Alatrish, E. (2013). Comparison Some of Ontology Editors. *Management Information systems*, Vol. 8, No. 2, pp. 18-24.
- [37] Wyner, A. and Hoekstra, R. (2010). A Legal Case OWL Ontology with an Instantiation of Popov v. Hayashi. *The Knowledge Engineering Review*, Vol. 14, No. 2, pp. 1-24.