https://dx.doi.org/10.4314/dujopas.v9i2a.35
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 9 No. 2a June 2023*

Nija-Onto: An Ontology of The Nigerian Languages Version 1

Al-amin Dahiru[1], Abba Lawan[1].
[1]Department of Computer Science,
Federal University Dutse.
Jigawa,
Nigeria.

Email: abbalawan@fud.edu.ng

Abstract
Ontologies are becoming increasingly important as a knowledge-based resource for man and machines. 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. Each language is annotated with its location information showing geo-political zone, states, and places where the Language originates. The ontology of Nigerian languages was built using the Methontology approach and specifically created using the Protégé 5.5.0 OWL editor. The Ontology was built on and evaluated using queries and a set of competency questions. The Nija-Onto was found to strongly align with its purpose of representing domain knowledge and thus provides visualized and highly tractable information on Nigerian languages.

Keywords: Ontology, Nigerian Languages, Information retrieval, semantic web.

INTRODUCTION

Natural Language Ontology (NLO) provides a formal specification of the linguistic semantics inherent in a given natural language.

Communication is extremely important in everyday activities of humans and people can communicate with one another based on the language understood by both parties. Various languages are being spoken in Nigeria in which their actual number cannot be specified. However, three major languages spoken in Nigeria are: Hausa, Yoruba, and Igbo and they provide respective means for communication between the Nigerian populace (Ayeomoni & Omoniyi, 2012).

The semantic web was introduced by the W3C consortium and in its approach, data is given a well-expressed format that models the meaning of information that can be found on the web, as well as applications and services (Zou, Finin, et al., 2004). The semantic web allows the conceptual structuring of the web in an explicit machine-readable way (T. & M., 1999). While the Semantic web is but an extension of our current web, it however, provides information in a more meaningful and understandable way for humans and machines alike. It enables the expression of contents and services in machine-readable form and enables some of the services to be automated.

**Author for Correspondence*

The Internet and the growth of the semantic web have made it possible to publish structured legal data online. The bundle of resources that make up the Semantic web of technologies that expand the existing web by permitting direct data connections between online documents, allowing for simple understanding and manipulation by both machines (computers) and humans. This technique provides online content in a way that both machines and people can understand. Making online material machine-processable enables the use of intelligent strategies to benefit from these representations. A given domain is explicitly conceptualized in an ontology, which essentially defines the domain model with potential limitations. A knowledgebase is created for each collection of individual items or classes described by an ontology. Instead than describing the precise state of things inside the domain, the conceptualization simply describes the knowledge that can be derived from the domain. This implies that conceptualization does not, or only occasionally does, change. So, this conceptualization's ontology is a specification (Thompson & Biralatei, 2022).

The backbone of the semantic web is an ontology, which provides a formal definition of a given domain thus allowing the conversion of the current web from machine-readable to machine understandable (Senthil, et al., 2022). While the web provides a source of reference for information on several topics on the other hand, semantic web is developed to build a web of meaning (Mustafa, 2010). Ontology as a component of database allows data to be stored as well as the metadata and it can be viewed by both humans and machines, usually in a graphical form (tree-like structure) which makes it unstructured and easier to view.

In other words, Ontology in computing is used for presenting and enumerating domain knowledge for human and machine processing (Ines, Ben, et al 2021). Ontology keeps information for future use i.e., data entities and their meaning within a domain. Unlike a database, which simply stores or keeps data without its meaning (E.P, C.T, & G.O, 2019). In this era of information overload and unprecedented technology, ontology played a vital role in information processing, comprehension, and application. This is because it provides interaction and cooperation between humans and computers, which help in comprehending, processing, and utilization of information.

IDEA AND SCOPE

The recent era of information and communication technology made information processing and retrieval an emerging technique. But most information saved or stored in a database for future use contains only the object or entity, excluding its meaning (which means the retrieval method is based on string matching and lacks semantic information for processing and cannot understand the users' needs well) while the advent of ontology allows information to be stored as well as its meaning within a domain for easy processing and retrieval (Kalaivani, Anandaraj, et al., 2011). The overall idea to conduct this project is to provide a well-developed ontology for Nigerian languages domain showing their geographical location and demographic data, so that information retrieval related to the northwest languages of Nigeria can be easier for knowledge or software development.

LITERATURE REVIEW

Ontology languages begin to emerge at the beginning of the 1990s as the evolution of existing knowledge representation languages. The advent of the internet led to the development of ontology languages for exploring the characteristics of the web. Such languages are referred to as

web-based ontology or ontology markup languages, their syntax is based on existing markup languages we have such as HTML and XML whose purpose is not ontology creation but data presentation and data exchange (Corcho, Fernandez-Lopez, et al 2007).

Ontology in computer science can be developed using an application that is developed using Java called Protégé which provides an environment with a large community of users that are active in the field (Kalaivani, Anandaraj, et al 2011). Protégé is not the only ontology editor but it provides a good environment for developing ontology as it was developed using Java programming language.

Ontology is a shared formal conceptualization of knowledge in a domain, which signifies that an ontology can be used as a standard vocabulary that will be used to share knowledge, and allows the reuse of information in a domain (Afolabi, Daramola, et al 2014). Ontology could be a shared formal conceptualization of a domain that gives a veritable stage to handle some of the existing threats to the Nigerian languages. Typically, since ontology will serve as a standard vocabulary, which will cultivate a shared understanding of the Nigerian languages' domain concepts, it will also allow knowledge reuse for the domain in such a way that it will validate certifiable data on languages such that it gets to be valuable for man and frameworks and make a national knowledgebase of dependable occasions that's capable of supporting sundry learning and proficient purposes (Afolabi, Daramola, et al 2014). Ontology plays important role in solving the problem of interoperability between applications across different areas by providing a shared understanding of knowledge within a domain.

The ontology idea was first applied by researchers in artificial intelligence who used computers to generate simulations using automated reasoning. From an IT standpoint, an ontology describes a collection of depictive primitives in a specific domain of knowledge. Classes, attributes, and relationships, as well as their outcomes and boundaries, are examples of frequently used primitives. A variety of programming languages and frameworks, including description logic (DL), first-order logic, relational models, and UML, can be used to express ontologies. Web ontology language is the most widely used ontology language in the semantic web's information stack (OWL) (Q, et al., 2022).

Ontologies have gained popularity as semantic-based representations crucial significance is one of the methods used to define and share knowledge most frequently in various domains including knowledge engineering, business modeling, software engineering/development, and e-learning systems. The method that teaching and learning are conducted may alter as a result of the ongoing development of new technology. Since they offer "a shared and common understanding of a topic that can be conveyed across individuals and application systems," ontologies are particularly popular. In order to create the scope and variety needed for the building of educational frameworks, ontology can be built. Ontologies are used in e-learning for a variety of purposes, from representing domain knowledge modules to automatically creating and evaluating individualized learning materials. A helpful tool that integrates linked resources, communicates knowledge, and gets rid of extraneous data is the idea of ontology. A fundamental account of the information in the world is provided by an ontology. A collection of concepts and the relationships between them are referred to as an ontology in computing. Ontology is used to logically reason and validate concepts in the semantic knowledge model within the context of a chosen discipline. A formal, clear specification of a shared

conceptualization is what an ontology is, according to theory. It provides a common language that may be used to build the domain knowledge model, which includes things like concepts, attributes, and relationships (Ahmed, Jawad, & László, 2022).

Ontologies are being used more frequently in numerous fields to specify a representational language that is used across a variety of topic areas, improving search and integration as well as resolving conceptual difficulties. Ontologies are verbal or visual representations of a specific domain's conceptualization. Since they offer the building blocks for modeling legal concepts, it offers many solutions to the semantic issues with understanding the legal subdomain. An ontology that is clearly defined helps to resolve the ambiguity problem, where a concept may imply various things to distinct fields. Several researchers have used ontological methodologies in a variety of disciplines to address issues with reasoning, ontology querying in natural language, and information retrieval. (Thompson & Biralatei, 2022).

The success of machine represented web known as the semantic web largely relies on ontologies. Ontology is a data modeling technique for structured data repositories premised on a collection of concepts with their semantic relationships and constraints on the domain. Ontology in computing is used for presenting and enumerating domain knowledge for Human and machine processing (E.P, C.T, et al 2019). Ontology keeps information for future use (i.e data entities and their meaning within a domain) unlike databases it stores or keeps data without its meaning, which makes ontology well-defined and understood for both human and machine processing.

From AI perspective, ontologies are formal representations of a common understanding within a domain. As a result, ontologies are regarded as a vital tool for Knowledge Management since they strive to bring consensus in the way a particular field of knowledge is characterized. This is because Knowledge Management has its roots in Artificial Intelligence. This consensus extends not only to terminology but also to the way concepts and objects may be organized and structured within the domain due to the formality of ontologies they also have the advantage to be modeled through information systems. Any subject of study can benefit from the use of ontologies to more effectively organize the data, knowledge, and ideas that comprise the field. Even though ontology may take on different forms, it will always have a term vocabulary and some related term meaning definition. Ontologies can therefore be utilized in the domain (Păvăloaia, 2012). Unlike in database data stored in ontological form can be viewed graphically rather than in table form which makes it easier for human and machine processing.

It plays a vital role in the semantic web as it is a component of a database, with the advent of ontology, people and software agents have access to information on the World Wide Web which makes users access information easily (Lakel & Bendella, 2015). Ontology improves information or knowledge processing by specifying or defining the major semantic relationship within the ontology, ontology is used in solving challenges of keyword search systems.

RELEVANT WORKS

An ontology of Nigerian history was developed by (Afolabi, Daramola, et al 2014). where the main objective of developing the ontology was to provide genuine support for various needs that require historical knowledge. Where he extracts terms from various textual documented sources, modeling and implementing the gathered data on the Protégé framework. The result of his ontology has the potential to competently support knowledge-based software endeavors.

Similarly, (Theresa, Ibukun, et al 2019). developed an ontology for the Yoruba language where he gathered data and terms from online books, blogs, websites, and the Yoruba dictionary he developed his ontology contribute to the development of digital resources for Yoruba language where its primary use is in AI applications, providing support in the advancement of natural language understanding for human beings, processing and generation. The ontology also makes the Yoruba language accessible for digital processing. His ontology was developed using the Protégé framework.

There are also a few domain ontologies that have been developed for educational purposes, specifically for automated grading, it was an ontology developed for higher education in the economics domain. This ontology helped to solve the problem of taxonomic shortcomings in economics as an academic disciple. Their approach to ontology creation is based on (Mesaric & Dukic, 2007).

(Boyce and Pahl 2007), developed a domain ontology for database course content. This followed a process called, Knowledge Engineering in Educational Technology (KEET). The KEET approach includes the following steps: 1) enumerate important terms 2) define the classes (concepts) 3) define the properties of classes and 4) finally create instances of the classes.

So also, an ontology was developed for the semantic search approach for geospatial information systems by (Panchanathan, Lawan, et al 2015). to help resolve the issue of semantic heterogeneity in metadata catalogs, using the Malaysian geospatial data infrastructure service portal. And the ontology was successfully merged with the developed Java software called MyGeo-Explorer (a web application coupled with a semantic search engine for exploring geospatial information).
Likewise, (Ishkewy, Harb, et al 2014). developed a software module called Azhary, which is a lexical ontology for the Arabic language, the ontology was evaluated using the gold standard-based approach and Arabic Word Net was used as the gold standard.

An ontology was developed by (E.P, C.T, & G.O, 2019) about Nigerian Languages and tribes using Noy and McGuinness methodology and was successfully developed on protégé 3.5 OWL editor. The ontology was also evaluated by queries; using competency questions and the ontology strongly align its purpose of representing domain knowledge.

METHODOLOGY

The methodology adopted majorly follows the knowledge engineering principles. Figure 1 shows the Nigerian languages ontology development life cycle.

Figure 1 Nigerian languages ontology life cycle

i. Specification the goal of the specification phase is to produce either an informal, semi-formal or formal ontology specification document written in natural language, aim and scope of the ontology, using a set of intermediate representations or using competency questions, respectively.
ii. Knowledge acquisition the knowledge elicitation about the Languages in the Northwest region is procured from (Roger, 2020), there are about 450 languages in the paper, from which the refined data about the languages in the Northwest region are found. The data is well formalized and organized so that no data is missing about the scope of this project.
iii. Conceptualization the terms used in this ontology are selected based on the competence questions the domain will answer, and the class names and object properties used in the ontology are not too difficult to avoid confusion, where there is no repetition in the terms.
iv. Integration there are different relationships between the classes, and the formalization is done with Description logic (DL's) notation, which is a formal language for knowledge representation and reasoning.

Implementation this ontology is implemented on Protégé editor 5.5.0 and represented using web ontology language (OWL).

DOMAIN CONCEPT

Some of the domain concepts are Nigeria, geopolitical zones, states, people, languages, northwest, northeast, north-central, southeast, south-south, and southwest.

Figure 2 showing domain concepts.

IMPLEMENTATION

In this section, the developed Nigerian languages ontology is engineered in the Protégée ontology editor (Protege, n.d.) . Ontograph is used to visualize the developed ontology. OntoGraf is a plugin in protégée that enables visualizing ontology stored on a protégée in graphical representation. Below are the methodology stages followed to achieve the objective of the ontology development.

Specification Stage: in this stage we develop competency questions that will be addressed by an Ontology of Nigerian languages. The scope of the project is also specified where it was limited to those languages in the Northwest region of Nigerian. Some of the formulated competency questions that NIJA-ONTO is designed to answer include: What are the Geopolitical Zones in Nigeria? What are the states in Nigeria? What are the states in the Northwest geopolitical zone? (Scope of the project) What are the languages spoken in each of the states in the Northwest region? Combinations of States or places that have a certain language in common? List of Northwest States and their local Governments, etc.

Knowledge Acquisition Stage: data about the Languages in the Northwest region were procured in (Roger, 2020) there are about 450 Nigerian languages identified in the paper, from which the refined data about the languages in the Northwest region are found. The data is thus formalized to cover the scope of this project. Whereas the data about States and Local government areas of Nigeria and Geopolitical zones were mostly collected from the internet.
In developing the ontology, optimal reuse of existing resources is considered. Although ontologies in the specific domain of the Nigerian language does not exist, some concept about other ontologies were found and added into the design of NIJA-ONTO.

Conceptualization Stage: in this stage, the classes and their subclasses, object properties were identified and asserted. This is shown in figure 3 & 4 below: where classes were created in hierarchies (figure 3) and their object property assertions are shown (figure 4).

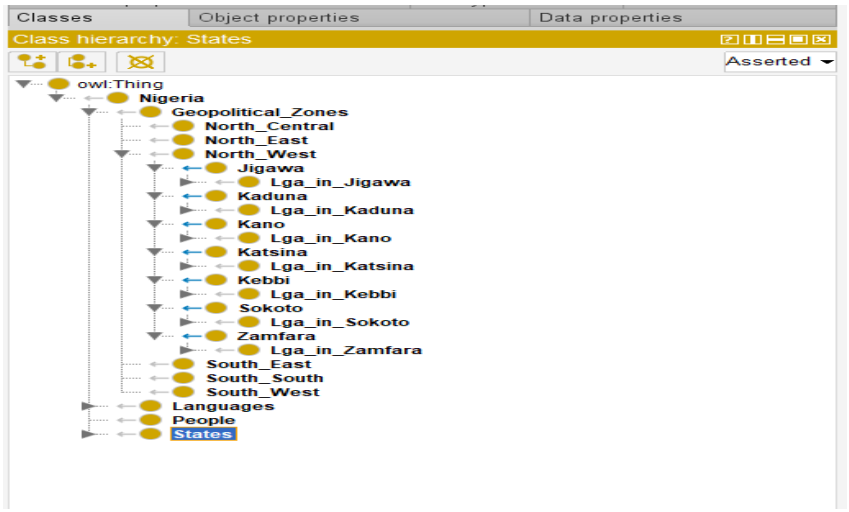

Figure 3 showing Classes hierarchy

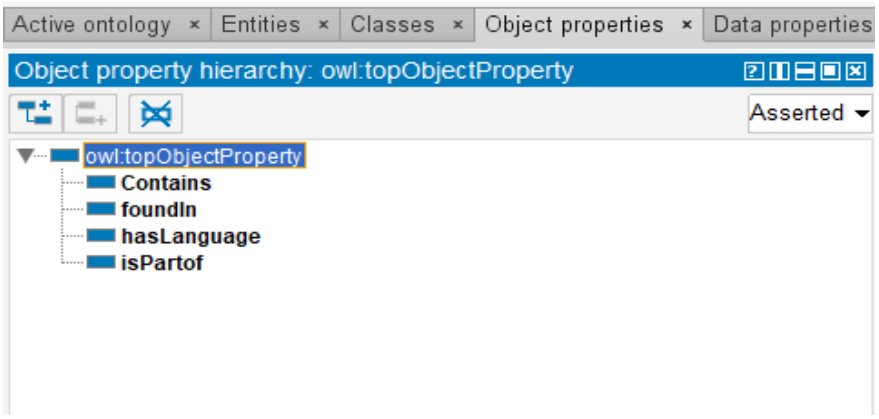

Figure 4 showing Object property.

Integration Stage: In the integration stage, classes are related to one another through object properties and data properties, where necessary. Visualization results of the Integration is shown in figure 5 below.

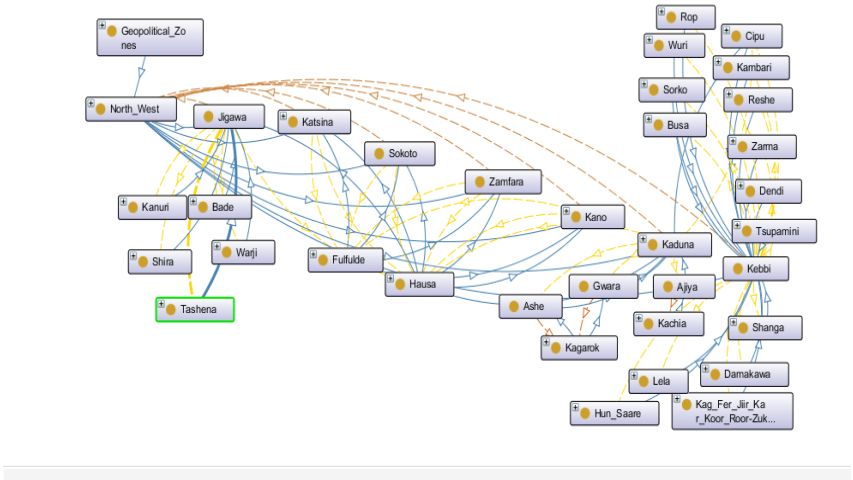

Figure 5 presents Visualization results of the integration

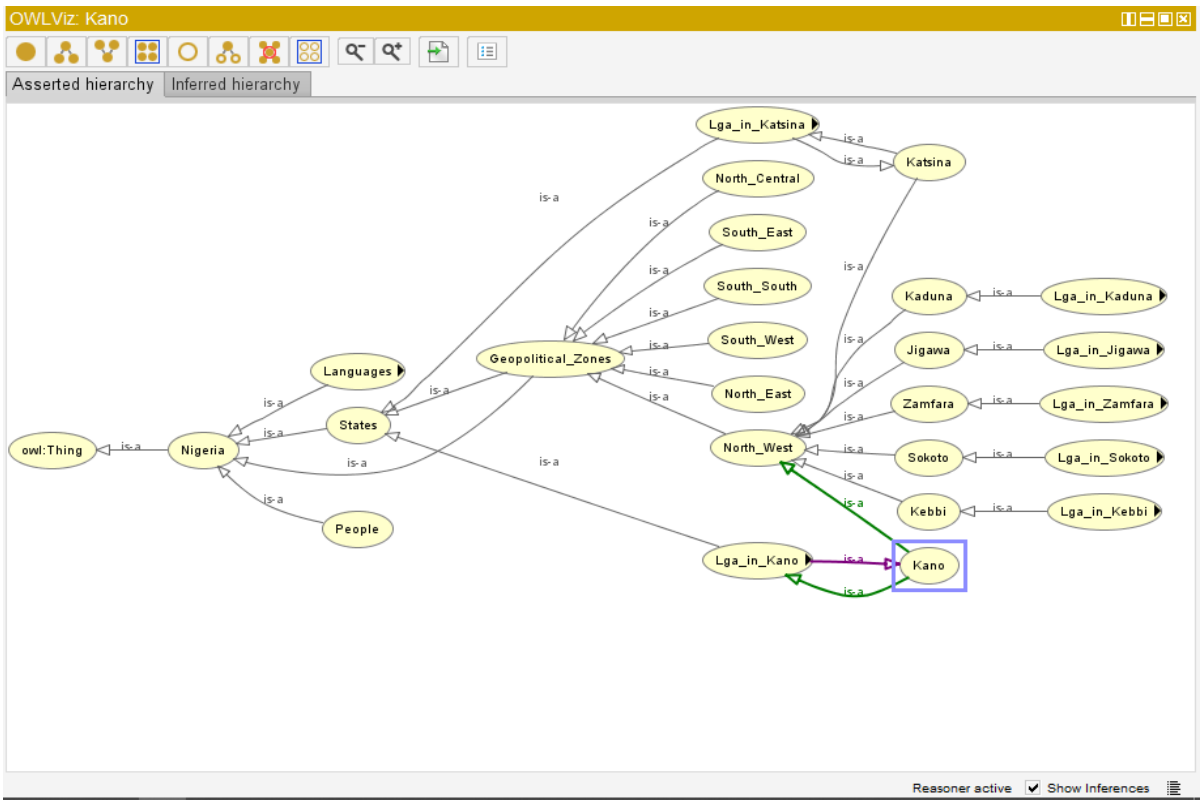

Figure 6 presents visualization of the developed Nigerian languages ontology.

Implementation Stage

The ontology is implemented using the Protégé Ontology editor 5.5.0 and specified using Web Ontology Language (OWL). Reasoner is enabled to show automatic inferences – called inferred hierarchy from manual assertions – known as the Asserted hierarchy. An OWLVIZ is used to view these assertions in a pictorial view as shown in figure 6 above.

DISCUSSION AND CONCLUSION

In order to check the validity of the developed ontology some of the competence questions are answered, as they are the key to evaluating how applicable the ontology is in the domain. Competency questions were questions that the ontology must answer, these questions were formulated or written before the ontology development started. They served as a requirements specification for ontology development.

The NIJA-ONTO is evaluated based on selected competency questions such as the following: (1) "What are the languages in Kebbi state?" the result of this question is answered using DL query as shown below in figure 7 (top) and a pictorial ontograph view is also shown in figure 6 (bottom).

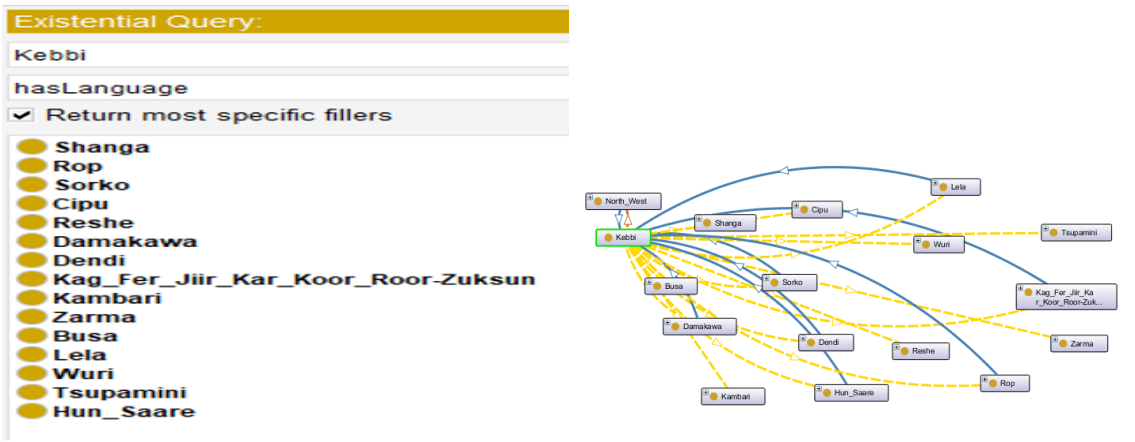

Figure 7 showing result of DL Query for languages in Kebbi state

(2) Another Competency Question is where the language is used as the object of attention and its information are sought such as in figure 7 below, which also shows the result of the query: "What are the states that speaks the Cipu language?". This is formulated as: *hasLanguage some Cipu* in the DL query editor.

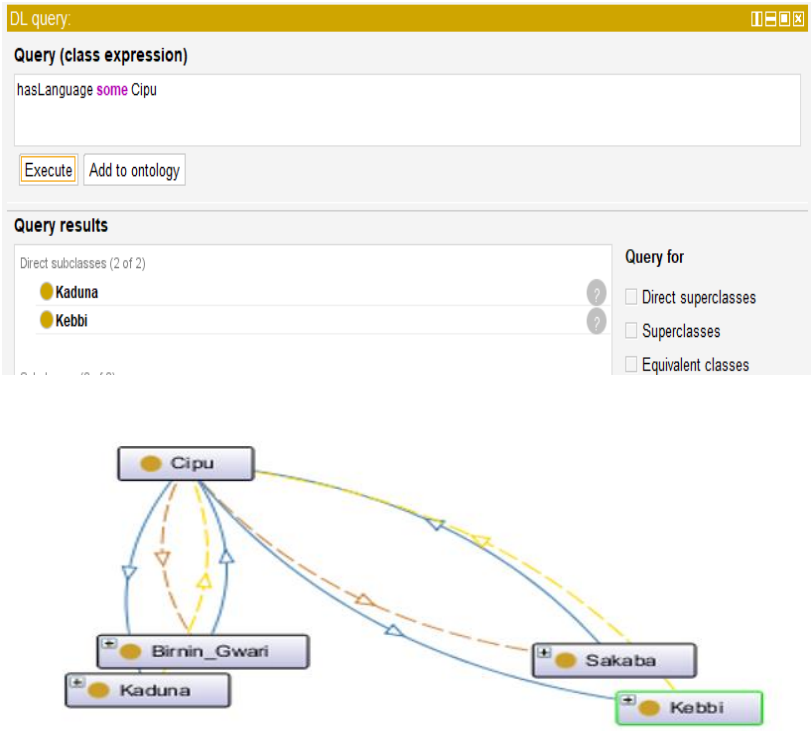

Figure 8 - showing result of States that speaks the Cipu Language. The Query (top) and pictorial view (bottom).

(3) So also, another Competency Question is answered where the language is used as the object of attention and its information are sought such as in figure 9 below, which also shows the result of the query: "What are the languages in some of Kaduna local government?". This is shown in the pictorial ontograph view: where a certain language Jju can be found in both Kachia and

Jema'a local government which is clearly showing the relationship between the local governments in Kaduna state.

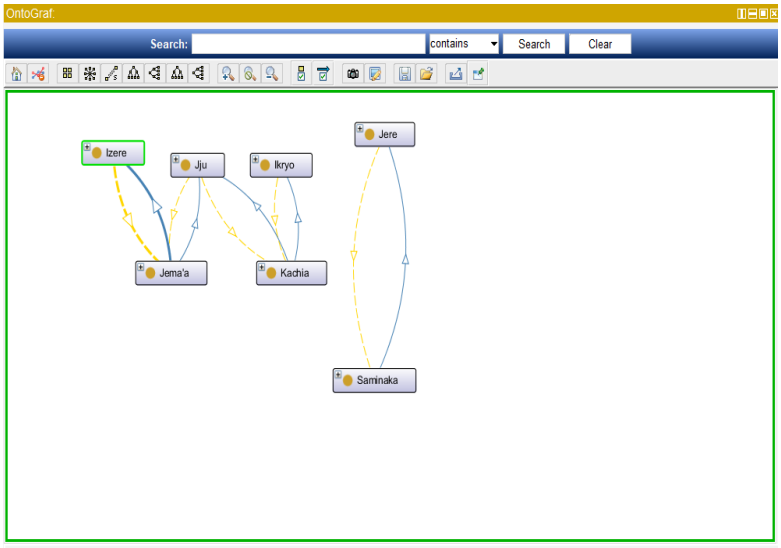

Figure 9 - showing result of languages in some of Kaduna local government

The answer to some of the competency question shows that the developed ontology will offer significant support for many automations needs that pertains to Nigerian languages domain, so also, the problem that arises when information is retrieved from the web, (where some certain or irrelevant result is displayed, and maybe the intended result is not within the results), the ontology will help to solve that problem so that irrelevant results are not returned.

The ontology has been applied in many different ways. The core meaning within computer science is that it is a model for describing the world that consists of a set of types, properties, and relationship types. This project demonstrated that ontology plays a critical role in knowledge acquisition and knowledge management processes. It helps make the knowledge storage and retrieval process significantly more intelligent. In future work, we shall expand the scope of the ontology from the North-west region to the whole geopolitical region of Nigeria. and it is very reasonable to encourage the construction of many more ontologies of different domains in Nigeria and Africa as a whole.

REFERENCES

Afolabi, I., Daramola, O., & Adio, T. (April 2014). Developing Domain Ontology for Nigerian History. *Australian Journal of Basic and Applied Sciences, 8*(6), 30-39. Retrieved from www.ajbasweb.com

Ahmed, H. G., Jawad, A., & László, K. (2022). Development of Ontology-based Domain Knowledge Model for IT Domain in e-Tutor Systems. *) International Journal of Advanced Computer Science and Applications, 13*(5). Retrieved from www.ijacsa.thesai.org

Ayeomoni, & Omoniyi, M. (2012). The Languages in Nigerian Socio-political Domains: Features and functions. *English Language Teaching, 5*(10), 12-19. doi:10.5539/elt.v5n10p12

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2007). What Are Ontologies and How Can We Build Them? *Ontological Engineering*.

E.P, E., C.T, A., & G.O, E. (2019). Building Ontology for Nigerian Tribes and Languages. *Journal of Science and Technology Research*, 14-23.

Ines, Osman, S., Ben, Yahia, G., & Diallo. (2021). Ontology Integration Approaches and challenging issues. *Halal Open Science*.

Ishkewy, H., Harb, H., & Farahat, H. (2014). In International. *Azhary: An Arabic Lexical Ontology. journal of Web & Semantic Technolog, 5*. doi:10.5121/ijwest.2014.5405

Kalaivani, P., Anandaraj, A., & Raja, K. (2011). AN ONTOLOGY CONSTRUCTION APPROACH FOR THE DOMAIN OF POULTRY SCIENCE USING PROTÉGÉ. *International Journal of Information Technology and Management Sciences, 1*(2), 134-140.

Lakel, K., & Bendella, F. (2015). Procedia Computer Science. *Dynamic Evaluation of ontologies, 73*, 16-23. Retrieved from https://doi.org/https://doi.org/10.1016/j.procs.2015.

Mesaric, J., & Dukic, B. (2007). An approach to Creating Domain Ontology for Higher Education. *Proceedings of the ITI 2007 29th In Conf.om Information Technology Interfaces*.

Mustafa, M. T. (2010). Understanding Semantic Web and Ontologies: theory and applications. *Journal of Computing, 2*(6). Retrieved from WWW.JOURNALOFCOMPUTING.ORG, HTTPS://SITES.GOOGLE.COM/SITE/JOURNALOFCOMPUTING/

ontology languages for semantic web. (2002). *Intelligent Systems, IEEE*. doi:10.1109/5254.988453

Panchanathan, S., Lawan, A., & Rakib, A. (2015, December). MyGeo-EXPLORER: A SEMANTIC SEARCH TOOL FOR QUERYING. *ARPN Journal of Engineering and Applied Sciences, 10*(23), 18012-18020. Retrieved from ww.arpnjournals.com

Păvăloaia, V.-D. (2012). EXPERIMENTS AND RESULTS ON THE USE OF ONTOLOGIES IN THE ARTIFICIAL INTELLIGENCE DOMAIN. *ECONSTOR*.

Protege. (n.d.). *Protege*. Retrieved from Stanford University Protege: http://protege.stanford.edu/community.php

Q, F., M, R., Ahsan M, R., A., N., Basem, A., Z., F., . . . Ahmed, S. (2022). Apllications of Ontology in the Internet of Things: A Systematic Analysis. *Electronics*.

Roger, B. (2020). An Atlas of Nigerian Languages. *McDonald Institute for Archaeological Research University of Cambridge*, 1-198. Retrieved from http://www.rogerblench.info/RBOP.htm

Senthil, Kumar, Narayanasamy, K., Srinivisan, Y., Chung, Hu, S., . . . Huang. (2022). A Contemporary review on utilizing semantic web technologies in healthcare, virtual communities, and ontology-based information processing systems. *Electronics*.

T., B.-L., & M., F. (1999). The Original Design and Ultimate Destiny of the World Wide Web by its Inventor, Harper, San Francisco, 1999. *Weaving the web*.

Theresa, O., Ibukun, A., Olamma, I., & Sunday O., O. (2019). Building Ontology for Yorùbá Language.

Thompson, I. P., & Biralatei, F. (2022). An Adaptable Ontology for Easy and Efficient University Data Management in Niger Delta University. *European Journal of Computer Science and Information Technology, 10*(2), 49-57. Retrieved from https://www.eajournals.org/

Zou, Y., Finin, T., Chen, & H. (2004). F- OWL: An inference engine for the semantic web. *Proceedings of the Third International workshop*, (pp. 16-18).