A Comparative Study on Representation Formalism Adoption for Semantic Knowledge Retrieval in Agriculture using Open Research Knowledge Graph

Patience U. Usip[1,2,*,†], Ekemini Johnson[1,*,†], Fernando Ortiz-Rodriguez[3,†], Funebi F. Ijebu[1,†], Babatunde O. Akinkunmi4[4,†], Ifiok J. Udo[1,†], Usoro I. Usip[1,†], Temitope Fakiyesi[1,†], Suzanlindsy Asuquo[1,†] and Etim Ekpenyong Attah[1,†]

[1]Department of Computer Science, University of Uyo, Uyo, Nigeria
[2]TETFund Center of Excellence in Computational Intelligence Research, University of Uyo, Uyo, Nigeria
[3]Universidad Autonoma de Tamaulipas : Reynosa, Tamaulipas, Mexico
[4]Department of Computer Science, University of Ibadan, Ibadan, Nigeria

Abstract

Today's focus on machine understandable designs that capture real world or domain information is required for solving complex problems in agriculture. Question answering in this domain require a robust semantic knowledge retrieval system that is vast in that domain and adequately represented knowledge for accurate and efficient semantic reasoning. This work is aimed at carrying out a critical review on the various representation methods, implementation approaches, and evaluation tools adopted in conducting agriculture-based researches. The reviewed articles are collected in the domain of agriculture. Open Research Knowledge Graph (ORKG) is adopted for creating comparisons used in this critical review including that of representation methods for multilingual machine translations. Visualizations from these comparisons are used in answering some competency questions surrounding this work and communicating the various results from this research. The results show 63 % analytical implementation work of most researches conducted in agriculture domain and 23 % automated. Knowledge graphs is mostly adopted in locations other than India and Nigeria. The report shows a high level of usage of general evaluation metrices such as accuracy, precision and recall for knowledge graph and ontology representations, pointing knowledge engineers to more researches on specific evaluation tools which are only being considered by very few of these researches. This will enhance the semantic knowledge retrieval procedures in agricultural domains as well as knowledge representation and reasoning for Semantic Web.

Keywords

Knowledge representation and reasoning, representation formalism, semantic knowledge retrieval, crop yield improvement, multilingualism,

MSW-2024: Fourth International Workshop on Multilingual Semantic Web. December 10, 2024, co-located with 6th Knowledge Graph and Semantic Conference (KGSWC-2024), December 11-13, 2024, Paris, France.
*Corresponding author.
†These authors contributed equally.
patienceusip@uniuyo.edu.ng (P. U. Usip); eke5461@gmail.com (E. Johnson); fernandortizrodriguez5@gmail.com (F. Ortiz-Rodriguez); ijebufrancis@uniuyo.edu.ng (F. F. Ijebu); ope34648@yahoo.com (B. O. Akinkunmi4); udojames@uniuyo.edu.ng (I. J. Udo); usorousip@uniuyo.edu.ng (U. I. Usip); temitopefakiyesi@uniuyo.edu.ng (T. Fakiyesi); sulindsy@gmail.com (S. Asuquo); attahdavid1@gmail.com (E. E. Attah)
0000-0002-6516-5194 (P. U. Usip); 0009-0000-5266-742X (E. Johnson)

1. Introduction

Knowledge representation and reasoning is a widely adopted sub-field of artificial intelligence with focus on machine understandable designs that capture real world or domain information required for solving complex problems. In line with the sustainable development goals (SDGs) number 1 and 3 which is centered on no poverty and good health and well-being by 2030, ORKG African Research Observatory aimed at studying researches done in agriculture and health with emphasis on comparing the various representation formalisms or methods used in the design for understanding and reuse by both humans and machines in providing solutions. Most of the proposed solutions with adequate representations of domain knowledge have made use of ontologies, knowledge graphs, logics, natural language processing models, transformer models, etc. The solutions have their justifications made either from analytical approaches or through automation of their frameworks followed by evaluations using standard tools. Other properties under considerations included the study location, the dataset used and the research problem. The reviewed articles in the comparisons were grouped based on common problem statements. The comparisons that are considered in this review include Representations and Reasoning for Semantic Knowledge Retrieval in Agriculture, Multilingual machine translation, Crop Yield Improvement and Irrigation and Erosion Control in Farming. The remaining parts of this work are organized as follows: Section 2 gives the review of contributions covered in the comparisons. Section 3 shows the representation and reasoning comparison results obtained from ORKG platform for selected agriculture related problem statements. Section 4 gives the visualizations of results and discussion of the comparisons in the previous section. Section 5 gives the conclusion of the reviews.

2. Related Literature

Choudhary et al. [1] used a smart farm ontology that comprises concepts and attributes relevant to the agricultural domain in addition to analyzing crop data gathered from an agriculture site in Rajasthan, India, that includes both Rabi and Kharif cropping patterns. The study creates a knowledge graph by connecting the gathered data to the smart farm ontology. Using SPARQL queries, the created knowledge graph was used to aggregate data and offer structural information. To the advantage of farmers and other stakeholders, the combined data is additionally utilized in machine learning algorithms to forecast crop yield. The outcomes for the various machine learning models that were employed were examined and contrasted by the researchers. For the Rabi crops data set, the Gradient Boosting Regressor model performed best, whereas the XGBoost Regressor model worked best for the Kharif-Crops data set. Inadequate data was the study's limitation. By combining deep learning with semantic web technologies of ontologies and knowledge graphs, Aminu et al. [2] present disease classification, addressing the challenges of (i) a lack of contextual information (e.g., soil or plant information) and domain-expert knowledge in deep learning-based systems; (ii) a lack of explainability of deep learning-based systems (e.g., disease information), especially to non-experts like farmers. The results demonstrate that the proposed method is superior to the state-of-the-art in terms of performance and explainability and is also appropriate for real-world scenarios. They include 0.905 (90.5 %)

prediction accuracy on large noisy dataset, the ability to generate user-level explanations about disease and incorporate contextual and domain knowledge, an average prediction latency of 3.8514s on 5268 samples, and 95 % of users finding the proposed method's explanation useful. Additionally, 85 % of users found the generated explanations to be easily understandable. Anand et al. [3] used natural language processing and deep learning to create a system for estimating crop disease severity and detection. Various convolution neural networks are developed to diagnose the severity of diseases that are discovered. Both the models that are fine-tuned by transfer learning and the trained networks that are learned from scratch are evaluated for performance. With a 99 % accuracy rate on a validation set, GoogleNet trained by transfer learning is the industry's leading model. Farmers can receive information regarding crop disease and its severity using natural language processing. With a 95 % accuracy rate for the Marathi language, the NLP model was trained on an English to Marathi dataset. The suggested approach might be able to reduce illness and boost agricultural productivity. To methodically extract information from expert-written text and create a knowledge graph, Wang et al. [4] devised a technique that uses relation extraction. Chinese texts were employed mostly in the study. First, after carefully selecting information authored by experts on soybean diseases and pests, the researchers identified the entities and categorized their thematic relationships into five groups. Next, utilizing cutting-edge deep learning architectures, they developed and trained three relation extraction models, then assessed each model's performance on the given task. Lastly, a sample knowledge network was constructed using the retrieved data to show how they could be used for automatic reasoning and the recommendation of solutions for the control of diseases and pests. The study sampled 1569 relation instances and 1038 entities in total. The top model, the Bert-Based Model, was able to identify relations from text with an F1 score of 98.49 %, according to the experimental data. The usefulness of the sample knowledge graph was further demonstrated by the experimental results. A fuzzy logic-based crop recommendation system is proposed by Banerjee et al. [5] to support farmers in rural areas. The goal of the project is to use fuzzy logic to create an effective and reliable crop recommendation system for the Indian state of West Bengal that considers topographical pattern, rainfall, and soil factors. Eleven soil properties, land elevation, mean annual rainfall, and associated cultivation index were included in the dataset as input parameters and output parameters, respectively. Eight key crops grown in the state of West Bengal are covered by the suggested model. For each crop, distinct fuzzy rule bases were developed to facilitate faster parallel processing. A wide range of datasets have been used to validate the model's performance, which has produced an accuracy of roughly 92 %. The ontology knowledge base for the durian pests and illnesses retrieval system is presented by Visutsak [6]. The system's primary contributions are: i) the information base on durian illnesses and pests that has been stored; and ii) the diagnosis of durian diseases and recommendations for treatment. Eight primary classes comprise the ontological knowledge: illnesses, pests, cultivars, bunch symptoms, leaf area symptoms, branch and trunk symptoms, fruit symptoms, and root and growth symptoms. 100 % precision, 88.33 % recall, and 93.8 % overall performance were obtained from the experiment. The goal of Aydin & Aydin [7] is to discuss how to relate site-specific parameters to sensor measurement outputs using crop-specific trait ontologies. To achieve the following goals, a data-integration approach for syntactic and semantic interoperability is proposed: (i) gather domain-specific data about specific agricultural products using ontology-based data acquisition forms created by domain stakeholders using crop-specific trait ontologies; (ii) gather and visualize stream data about site-specific parameters of specific agricultural products through WSNs; (iii) create domain-specific linked open data using mapping rules created by any domain stakeholder using crop-specific trait ontologies; (iv) store semantically annotated agricultural data across various databases and files, including relational databases, graph databases, XML files, and RDF files, etc.; (v) to provide syntactical interoperability using web services and APIs, which allow stakeholders to share data for a particular agricultural product between different kinds of software applications; (vi) to publish well-defined, well-structured and semantically annotated data concerning a particular agricultural product using open standard in appropriate formats. An open-data platform is developed, and its usability is evaluated to justify the viability of the proposed approach. Furthermore, this research shows how to use web services and APIs to carry out the syntactic interoperability of sensor data in agriculture domain. A plan for creating an ontology specifically for the agriculture domain is presented by Kaushik and Chatterjee [8]. The suggested strategy operates in two stages. First, it automatically extracts agricultural-related words using natural language processing techniques and domain-dependent regular expressions. Semantic connections between the retrieved terms and sentences are found in the second stage. For the task, the rule-based reasoning algorithm RelExOnt has been suggested. Precision and recall obtained from human review of the word extraction output are 75.7 % and 60 %, respectively. With an average precision of 86.89 %, the relation extraction algorithm RelExOnt performs admirably. A review and trend study of knowledge graphs for crop diseases and pests was conducted by Xuesong et al. in 2019 [9]. The study finds that while knowledge graphs offer a novel and more flexible approach to knowledge administration, current knowledge management systems share several fundamental flaws in terms of efficiency, scalability, and application. This study examined and categorized the most important knowledge graph technology approaches and procedures in the field of agricultural diseases and pests in recent years, considering the unique characteristics of crop illnesses and pest data. The definition and meaning of crop disease and pest knowledge were presented, and a detailed analysis of the current creation process was conducted from four perspectives: knowledge representation, extraction, fusion, and reasoning. Additionally, the expert system, search engine, and knowledge question-answering system were all given a thorough introduction to the use of crop diseases and pest knowledge graphs. In conclusion, the study provided an overview of the key issues and concerns related to diseases and pest knowledge graphs, as well as an outlook for the field based on the salient features and challenges of current knowledge graph research.

3. ORKG Comparisons of Representations and Reasoning in Agriculture

There are several researches with various representations and reasoning formalisms used in agriculture-based problem statements to include ontologies, knowledge graphs, logics, natural language processing techniques, etc. Some of the problem statements in agriculture under consideration include crop yield improvement, semantic knowledge retrieval, multilingual machine translation, pest management, etc. But this critical review is centered on the first three with the number of contributions or articles in each comparison checked against each problem statement as shown in ??.

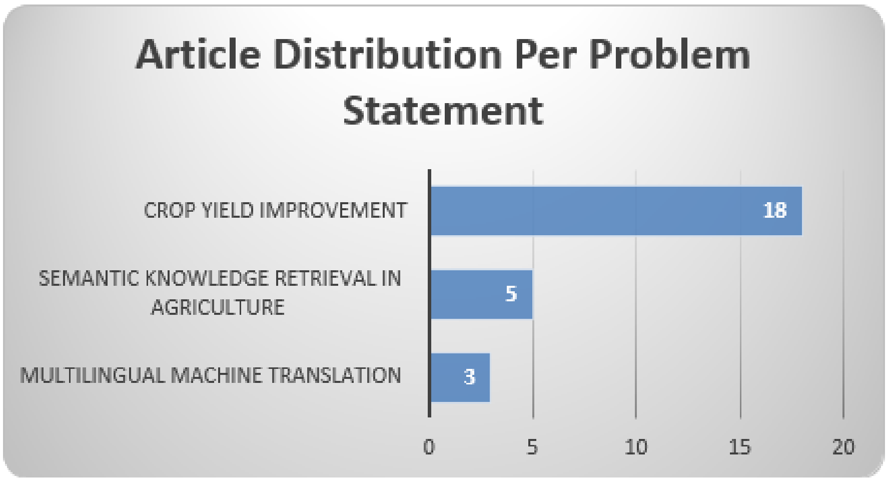


Figure 1: Bar Chart showing Article Distribution by Problem Statement

From ??, more numbering up to 20 articles were reviewed around crop yield improvement knowledge representation while less than 5 articles were reviewed around machine language translation. The comparisons for the problem statements under consideration with their properties are described in the following sub sections.

3.1. Comparison on Representation and Reasoning for Semantic Knowledge Retrieval in Agriculture

Various contributions reviewed on representations and reasoning tools adopted for the semantic knowledge retrieval of knowledge on crop pest and disease management based on featured properties including problem statement, problem domain, method, implementation approach, evaluation tool, study area and control parameters as considered in the ORKG comparison published in the link: https://orkg.org/comparison/R801641https://orkg.org/comparison/R599390/ as shown in ?? [10]. The comparison shows that ontology has been widely adopted, and implementation approaches being 80 % automated and 20 % analytical. Despite the adoption of ontology and automation, the evaluation tool in used is the F1 score and none of the reviewed articles adopted the ontology evaluation tools such OOPs! and TDDOnto.

3.2. Comparison on Representations for Multilingual machine translation

Reviews made on representations and reasoning for multilingual machine translations in several domains including agriculture and other related domains have properties considered in the ORKG comparison published in the link: https://orkg.org/comparison/R801643 as shown in ?? [11] The comparison featured the use of logic, knowledge graphs and natural language processing as representation and reasoning formalisms and several language pairs including English–Malayalam, English–Hindi, English–Tamil, and English–Punjabi. Other languages

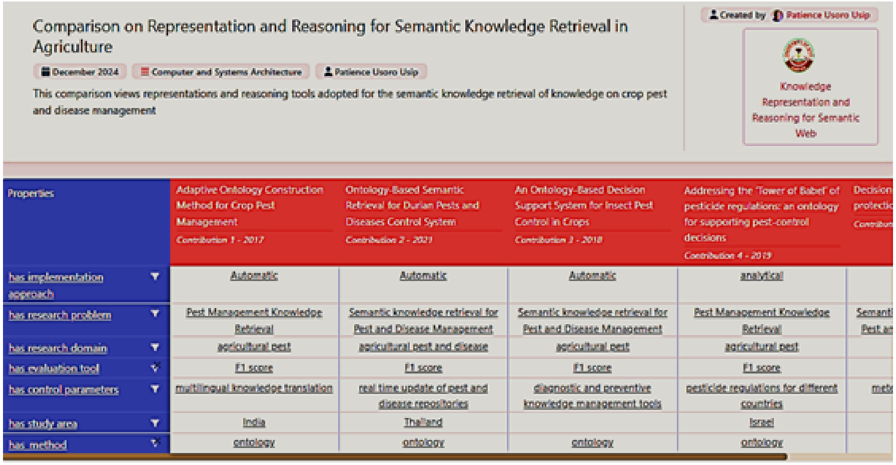


Figure 2: Representation and Reasoning for Semantic Knowledge Retrieval in Agriculture

considered in the reviews includes English, Chinese, Japanese, German Korean, Swahili and other Western translation. Although not much have been done on under resourced languages.

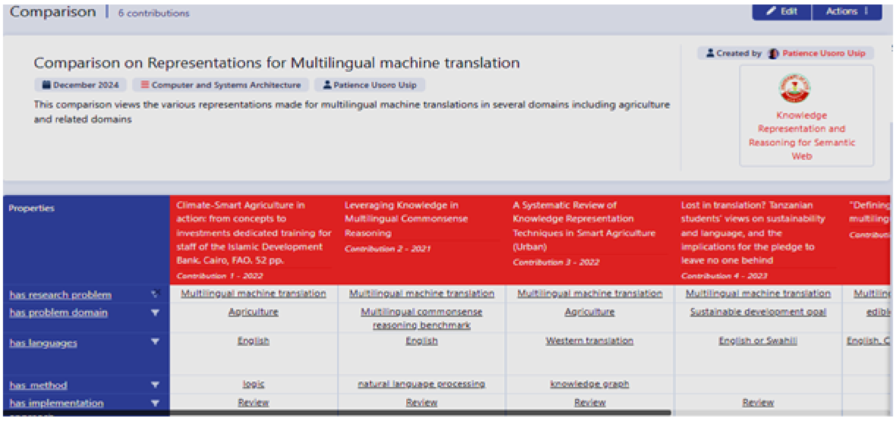


Figure 3: Representation for Multilingual machine translation

3.3. Comparison for representation and reasoning for crop yield improvement

This comparison considered several contributions on crop yield improvement based on featured properties considered in the ORKG comparison and published in the link: https://orkg.org/comparison/R659189 as shown in **??** [12]. It covers a wide range of representation and reasoning formalisms including ontology, knowledge graphs, logics, natural language processing with about 73.68 % of implementation approach being analytical as shown in **??** (c). **??**(a) shows statistical tools (such as F1 score, accuracy, R-square, MAE, RMSE, precision, recall), were widely used for evaluations across these articles with no mention made on specific evaluation tools for ontologies, knowledge graphs, etc. This observation is seen across various study locations including China, USA, Nigeria, India, Australia, Sri Lanka, Russia, Ireland and Canada as shown in ?? (b).

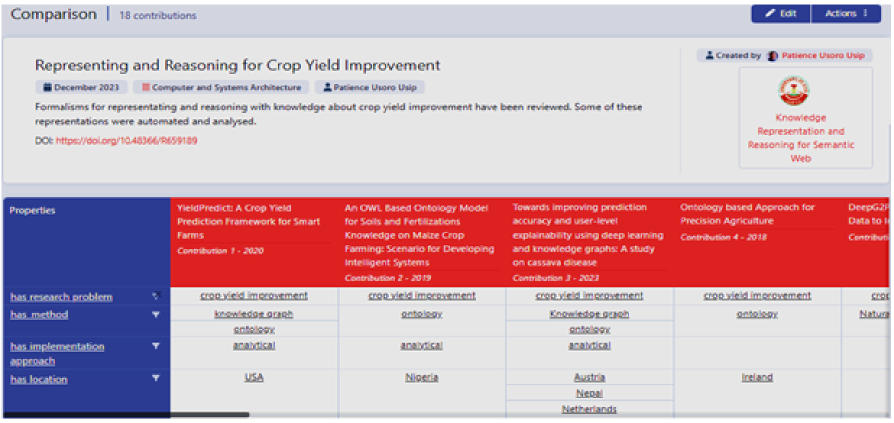
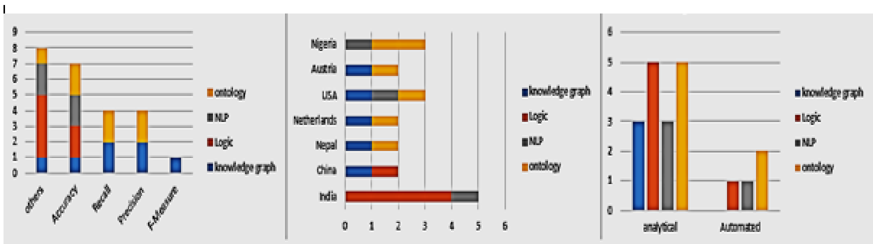


Figure 4: Comparison for Representation and Reasoning Crop Yield improvement



(a) (b) (c)

Figure 5: Representing and Reasoning for Crop Yield Improvement

Reviewed articles called contributions in ORKG comparisons are graphically represented with all the key concepts, attributes and the relationships clearly represented. For example, an article titled “A Machine Learning-Based Mobile Chatbot for Crop Farmers” is graphically represented in ORKG and graphically viewed as shown in ?? with graph depth 1, 2 and 3 viewed as shown in ??a, b and c respectively [13]. Authors can deploy their translators [14] or find a match using ORKG comparisons as they embark on further study. The ORKG graph view uses different shapes and colours to represent concepts, sub-concepts, individuals (data items) while the relationships clearly shown on the labelled directed vertices on the graph.

4. Results and Discussion

In this review, the sample comparisons cutting across three problem statements clearly show the initial concerns of the reviews made in each case or has answered some competency questions such as: Do these contributions use representation and reasoning formalisms? Was the representation formalism used analyzed or automated? Did the contribution adopt any

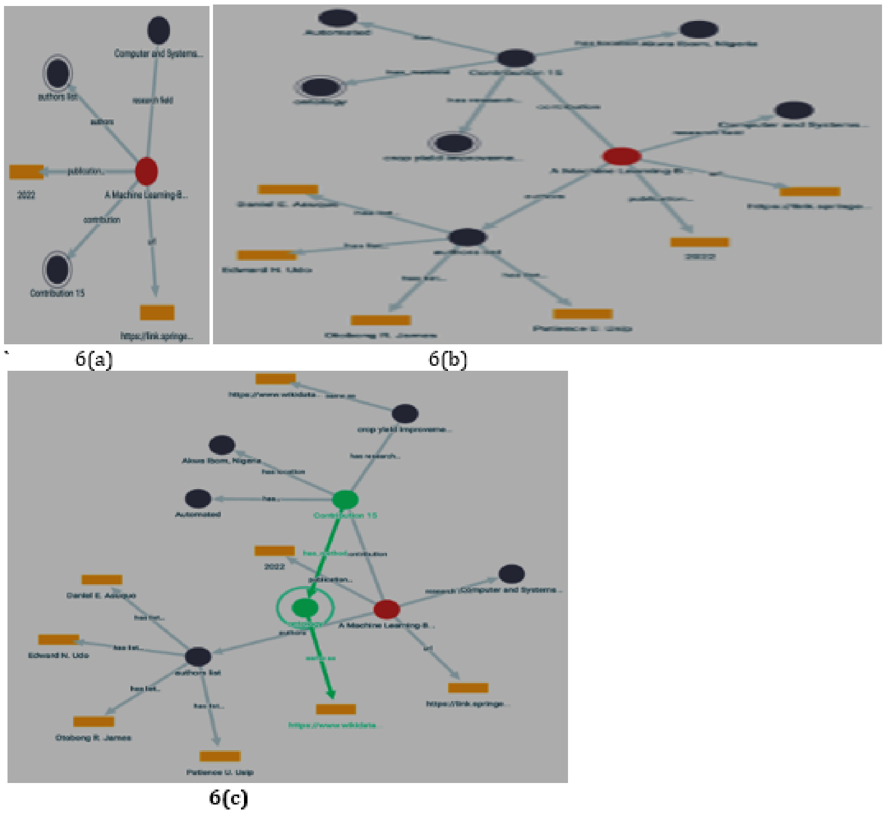


Figure 6: An ORKG knowledge graph view of a Sample Contribution

evaluation tool? Are the tools standard tools for the adopted formalism? From the total number of contributions considered in agriculture domain, the overall count of implementation approach in **??** shows that about 63 % were implemented analytically, only 26 % were automated while 11 % either adopted both or not specified. The study areas of the overall reviewed articles given in **??** shows India as the lead followed by works carried out in Nigeria, China, USA and others.

?? shows the overall count of representation methods from the reviewed literature, with ontology as the most widely used, followed by logic and then knowledge graph. The chart in **??** gives the evaluation tool by representation method with F-measure mostly used for evaluations in ontology. Other evaluation tools such as accuracy, precision and recall are also widely applicable in ontologies as well as precision and recall in knowledge graphs. For the evaluation of logical systems, other evaluation tools were widely used outside accuracy, f-measure, precision and recall.

?? gives the chart showing the implementation approach by representation method to describe that most of the ontologies adopted in most of the literature were analyzed and automated. Logic and NLP are seldomly automated. Knowledge graphs and logics for reasoning sparing use other implementation approaches other than the identified ones in this review. From the reviewed literature, logic is widely adopted for representations and reasoning in agriculture

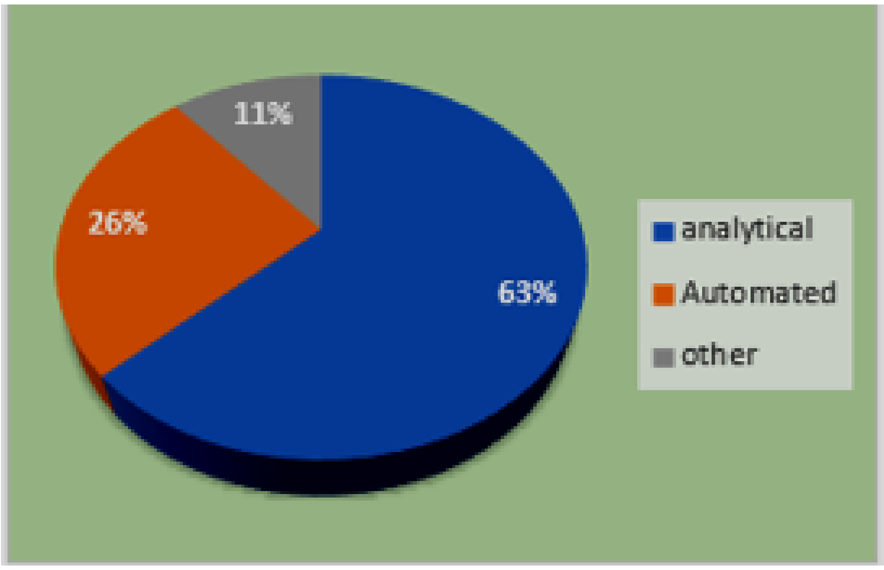


Figure 7: Overall Count of Implementation Approachn

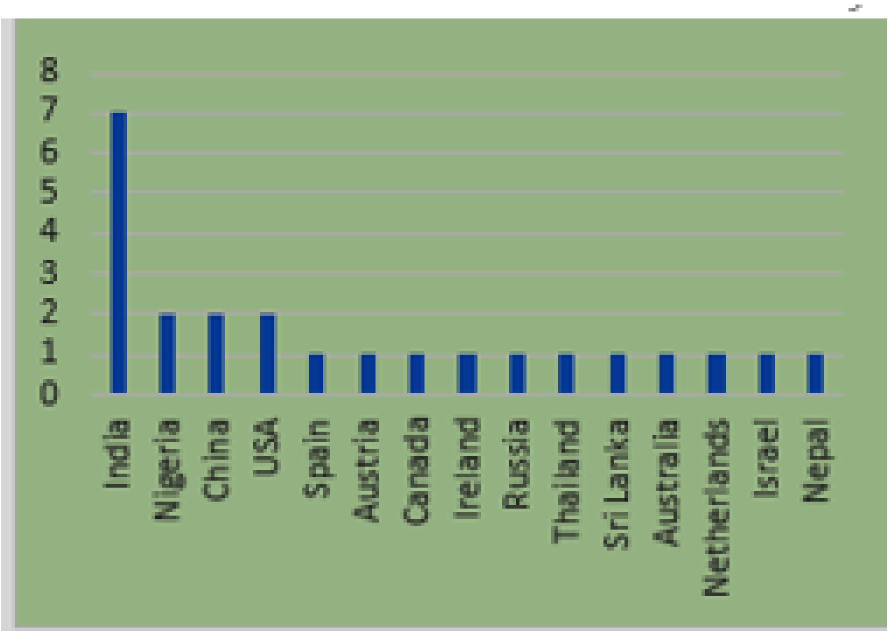


Figure 8: Locational Distribution of Research Articles

related research in India, sparingly being used in China, and not fully embraced in Nigeria, USA, Netherlands, Austria and Nepal. However, research in agriculture embracing knowledge graphs were conducted in several locations except India and Nigeria as shown in in ??. On the other hand, India, Nigeria and USA have adopted NLP which were not used in other locations

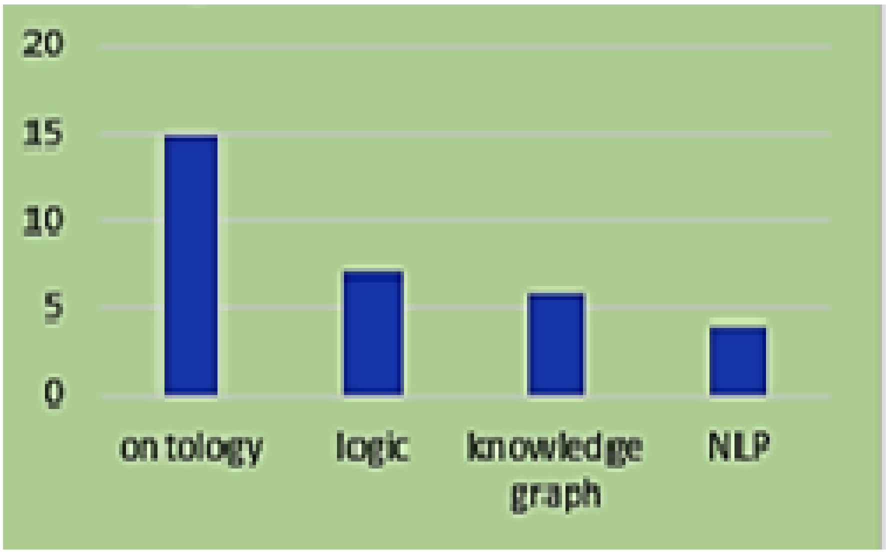


Figure 9: Overall Count of Representation Method

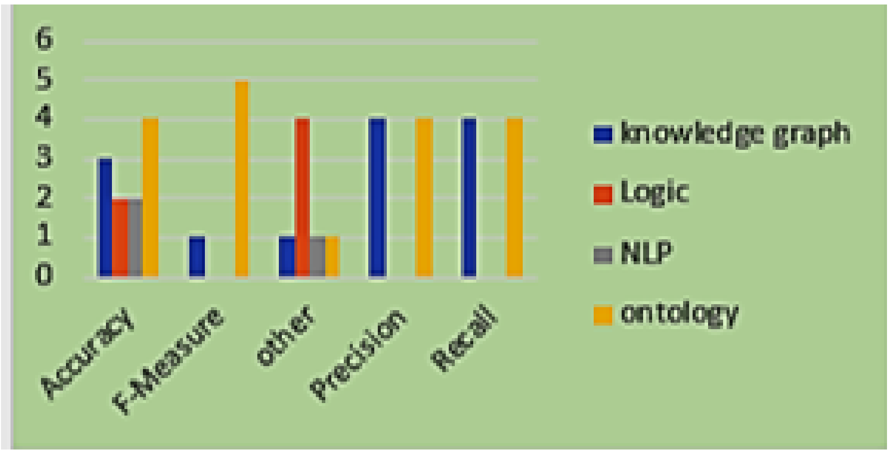


Figure 10: Evaluation Tool by Representation Method

including Netherlands, China, Austria and Nepal.

5. Conclusion

The automation of most researches conducted in agriculture domain have not yet been fully done as a higher percentage of 63 reviewed articles still used analytical implemented approach. This opens up more doors for researchers in India and Nigeria to embrace the use of knowledge graphs in most of their researches as seen and adopted in other locations. The report shows a high level of usage of general evaluation metrices such as accuracy, precision and recall

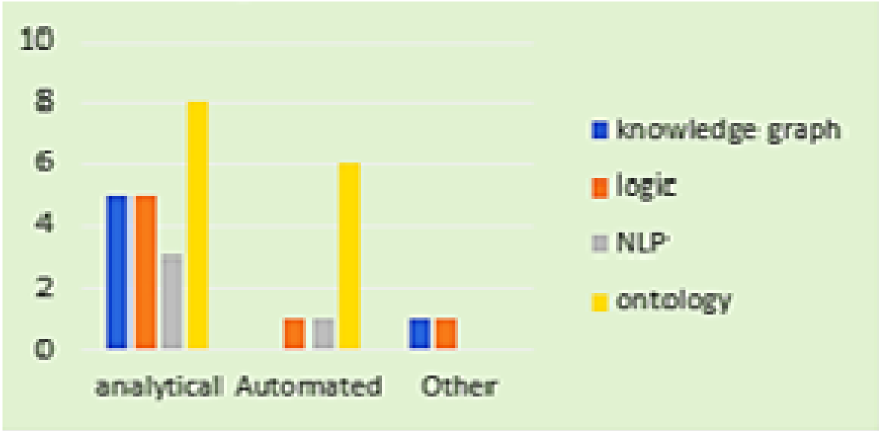


Figure 11: Chart showing Implementation Approach by Representation Method

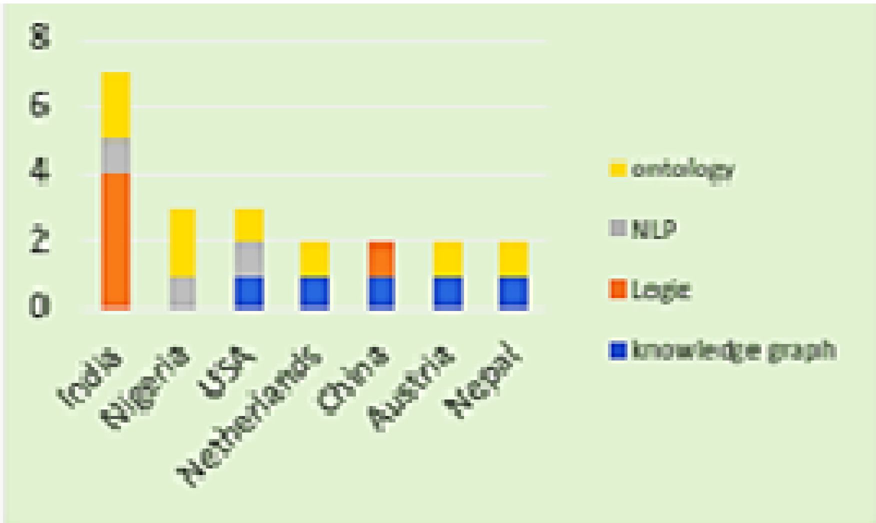


Figure 12: Chart showing Locational Distribution Representation Method

for knowledge graph and ontology representations, pointing knowledge engineers to more researches on specific evaluation tools such as TDDOnto and OOPs! which are only being considered by very few of these researches. This will enhance the semantic knowledge retrieval procedures in agricultural domains as well as knowledge representation and reasoning for Semantic Web.

6. Acknowledgements

The authors of this work appreciate the entire TIB Research Lab, Hannover, Germany and the Team led by Souren Auer, for sponsoring this work through their 2023 Curation grant. We also acknowledge research partners, Shodhguru Labs, India and the members of the FoRRK Research Lab team, Computer Science Department, University of Uyo and the interns/mentees for all their efforts. We also acknowledge TETFund Centre of Excellence, University of Uyo for funding this research publication.

References

[1] N. K. Choudhary, S. S. L. Chukkapalli, S. Mittal, M. Gupta, M. Abdelsalam, A. Joshi, Yieldpredict: A crop yield prediction framework for smart farms, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 2340–2349.

[2] E. F. Aminu, I. O. Oyefolahan, M. B. Abdullahi, M. T. Salaudeen, An owl based ontology model for soils and fertilizations knowledge on maize crop farming: Scenario for developing intelligent systems, in: 2019 15th International Conference on Electronics, Computer and Computation (ICECCO), IEEE, 2019, pp. 1–8.

[3] M. Anand, K. B. Sahay, M. A. Ahmed, D. Sultan, R. R. Chandan, B. Singh, Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques, Theoretical Computer Science 943 (2023) 203–218.

[4] P. Wang, C. Zhang, D. Wang, S. Zhang, J. Wang, X. Wang, L. Huang, Relation extraction for knowledge graph generation in the agriculture domain: A case study on soybean pests and disease, Applied Engineering in Agriculture 39 (2023) 215–224.

[5] G. Banerjee, U. Sarkar, I. Ghosh, A fuzzy logic-based crop recommendation system, in: Proceedings of International Conference on Frontiers in Computing and Systems: COMSYS 2020, Springer, 2021, pp. 57–69.

[6] P. Visutsak, Ontology-based semantic retrieval for durian pests and diseases control system, International Journal of Machine Learning and Computing 11 (2021) 92–97.

[7] S. Aydin, M. N. Aydin, Semantic and syntactic interoperability for agricultural open-data platforms in the context of iot using crop-specific trait ontologies, Applied Sciences 10 (2020) 4460.

[8] N. Kaushik, N. Chatterjee, Automatic relationship extraction from agricultural text for ontology construction, Information processing in agriculture 5 (2018) 60–73.

[9] L. Xiaoxue, B. Xuesong, W. Longhe, R. Bingyuan, L. Shuhan, L. Lin, Review and trend analysis of knowledge graphs for crop pest and diseases, IEEE Access 7 (2019) 62251–62264.

[10] P. U. Usip, Comparison on representation and reasoning for semantic knowledge retrieval in agriculture, 2023. URL: https://orkg.org/comparison/R599390/. doi:10.48366/R599390.

[11] Patience Usoro Usip, Comparison on representations for multilingual machine translation, 2024. URL: https://orkg.org/comparison/R763758. doi:10.48366/R763758.

[12] Patience Usoro Usip, Representing and reasoning for crop yield improvement, 2023. URL: https://orkg.org/comparison/R659189/. doi:10.48366/R659189.

[13] P. U. Usip, E. N. Udo, D. E. Asuquo, O. R. James, A machine learning-based mobile chatbot for crop farmers, in: International Conference on Electronic Governance with Emerging Technologies, Springer, 2022, pp. 192–211.

[14] A. Oboho-Etuk, P. U. Usip, O. S. Adeoye, I. Ollawa, A machine-learning based text

classification and machine translator for selected under-resourced languages., in: SWIoT+ MSW@ KGSWC, 2023, pp. 98–105.