Towards an Improved Semantic Model: An Ontological Based Framework for Misinformation Detection.

Hussaini Abubakar Zubairu[1], Enesi Femi Aminu[2],
[1.]Department of Information Technology, [2.]Department of Computer Science, [1,2]Federal University of Technology, Minna Nigeria.

Abstract

The scourge of false information across different real life scenarios in this digital age is alarming and worrisome. While open social networks and artificial intelligence have changed the narration of computing by making data available and accessible, on the contrary, they are aiding and abating misinformation. State of the art techniques such as machine learning models, natural language processing, and recently ontology semantic technology have been geared towards addressing this challenge. However, the promising strength of ontology semantic technology in terms of contextual knowledge and handling concept semantic conflict have not been adequately addressed. In view of this development, this research aims to design an ontological semantic framework for the detection of false information considering the domains of politics, science and health. An existing fake news taxonomy data was reused along with the newly harvested domain based dataset. The ontology, dubbed as *iFaIDet* is developed based on Noy-McGuiness methodology, and the results are promising. The individual OWL component of the ontology far outweigh the results of three other existing related ontologies. Similarly, the results of seven out of eight metrics of its structural based evaluation are promising. For instance, average population and class utilization metrics show 3.45 and 0.88 against the existing 1,413 OWL ontologies' average values of 1.34 and 0.54 respectively. However, the ontology is still work in progress as standard rule based misinformation database will be fused into it, along with design collaborative algorithm for top level semantic lexical databases.

Keywords: iFaIDet, Semantic Model, Contextual Knowledge, False Information Ontology, Rule Based Misinformation Database

1. Introduction

Advancement in technology in this present age has caused a lot of rapid development in virtually all real life scenarios. With the high volume of data traffic, while more data are being generated which has a lot of usefulness in computing; on the other hand, some of those data are equally misleading the uninformed populace thereby causing false information to be consumed. More worrisome is the use of social media platforms (Almaliki, 2019) to propagate the false information, which may appear as news items or unsuspected fact (Wang, et al., 2019). The rise of false information spread weather intentional or unintentional is alarming because it has led to so many social and political unrest (Kaliyar, et al., 2020). During the outbreak of the global pandemic known as CoVid-19, several attempted and suggested solutions were resisted owing to the fast spread of misleading information (Van Der Linden, 2022). The terms misinformation, disinformation, and fake news are often used interchangeably; they might portray the same semantic however, they are differed in driven purpose (Broda, & Strömbäck, 2024; Dame Adjin-Tettey, 2022; Petratos, 2021). While misinformation is a false and misleading information but without the intention to harm unsuspected information consumers, disinformation is a deliberate act of passing wrong information. Fake news, the frequently used term among the three terms, and the same time arguably, the most problematic in terms of definition (Hassan, 2023; Wang, et al., 2019). Fake news may roughly defined as the process of misrepresentation of true information or a deliberate act of manipulating false information which may appears real. It is still a very complex task to identify fake news (Aïmeur, et al., 2023).

Regardless of the variance in the driving purpose and intention of the terms: misinformation, disinformation, and fake news; they still have a point of convergence where they all still point out to false or misleading information. In fact, fake news may be regarded as the product or resultant effect of misinformation or disinformation. Besides human beings, which are well known carriers of misinformation, is also the social bots. The bots pretend as normal human users on the social media platforms to spread false information across large followers (Himelein-Wachowiak, et al., 2021). Therefore, no amount of effort is too much to address this menace. In view of this development, there have been several attempts to employed artificial intelligence and machine learning models to address the problem (Capuano, et al., 2023; Hunt, et al., 2022; Zhao and Yan, 2021); this has come with some level of progress however, not without room to advance and exploit other technologies. Alenezi and Alqenaei (2021) proposed machine learning based detection model for COVID-19 misinformation on a twitter platform by exploring long short-term memory (LSTM), k-nearest neighbors (KNN), and multichannel convolutional neural networks (MCCNN). Similarly, Chen, et al., (2023) deployed series of deep learning models such as bidirectional long and short-term memory (BiLSTM) for detecting false information about COVID-19.

While considerable measures of progress have been recorded using these technologies, there is still need for an improvement owing to the complex nature of data in context as a result of ambiguities of natural languages (Jiang, 2020). The semantic of terms based on various domains in terms of synonyms and polysemy have to be considered for a more accurate detection of misinformation when hybrid approach of content and context based is a focus. When the complex data are situated in context in relation to their domains and similar concepts, such data can easily and accurately classify by machine learning models. Consequently, Semantic based technologies such as, ontology knowledge-representation approach have demonstrated capacity for this objective (Bani-Hani, et al., 2022; Mazepa et al., 2021). Ontology, according to Gruber can be defined as a formal and explicit specification of shared conceptualization (Aminu, et al., 2021). Ontology as semantic knowledge representation strategy can be developed based on adopted or adapted methodology such as Noy-McGuiness, Fox-Gruninger, Uschold and King, and methontology (Aminu, et al., 2020). Furthermore, the harvested concepts from a domain is represented for machine understanding and interpretation using some of the popular ontology languages such as resource description framework (RDF), and web ontology language (OWL).

Recently, researchers have started to exploit ontology approach to design a semantic based database for onward and further utilization by other cutting-edge technologies (Sharma & Kumar 2022; 2023). Bani-Hani, et al., 2022 designed a semantic based model christened Fandet ontology to provide contextual information for the detection of fake news on social media. However, the contextual information of Fandet primarily lies between the main concepts and their properties as no other standard contextual based model is integrated into the ontology to take care of real-life concepts mismatch and for robust performance. In view of this development, this research proposed an ontological based framework, which is intended to leverages on the existing Fandet towards an improved semantic model for detection of misinformation on health, politics, and science domains. The remaining part of this work include the review of related studies as section 2. Section 3 presents the proposed methodology and conceptual framework for the enhanced semantic model. The preliminary results and its evaluation for the ontology is presented by section 4, while section 5 accounts for the conclusion.

2. Related Studies

Today, different digital contents across various real life scenario suffer from the menace of false information. Interested researchers in this area have advanced series of approaches and methodologies to address the menace. However, the research attention given to the different domains are not equitable. The research of Petratos, (2021) clearly pointed that the attention of misinformation, disinformation, and fake news for political domain is wider than any other domains for example, business and science. The position of this literature under review clearly shown that some areas or domains have little or no attention given to them. In order to proffer robust solutions to these challenges, several efforts have been attempted. Most of these dominating efforts include the application of machine learning techniques. However, in a recent time, effort to employed semantic and ontology technologies evolved. Jabardi and Hadi, 2020 proposed a novel approach to detect fake twitter account using ontology and its rule language.

In an effort to deal with misinformation menace in biomedical domain, Amith and Tao, (2018) developed vaccine misinformation ontology in order to detect false information about vaccines. The developed ontology is an extension of the existing misinformation ontology (MO) that connects the model of nanopublications resource description framework. The authors re-encoded some main concepts of RDF based MO into OWL using protégé editor. It contains 116 classes with 20 and 6 object and data properties respectively. The ontology has no instances as concepts.

The semantic salient and unique property of ontology design to deal with the issue of data complexity and natural languages ambiguities cannot be overemphasized (Brasoveanu and Andonie, 2020). The research of Bajpai and Chaturvedi, (2024) exploited this property to address the challenge of detecting the initiator of false information in an open social network, which is against the previous attempts on only addressing the structures of the networks. An RDF lower level three-layer architecture based ontology was designed to mapped with the concepts of upper level ontology. Twitter network is considered as the domain for the harvested concepts to carried out the design and experiment. SPARQL is used to query the ontology developed. However, for expressivity gain of the ontology, OWL and Semantic Query enabled Web Rule Language (SQWRL) would have been more appropriate for the ontology design and query process respectively. Similarly, the research of Jabardi and Hadi, (2020) in an attempt to detect false information in a twitter network designed an OWL based ontology with combined strength of semantic web rule language (SWRL) to infer fake accounted created within the network. Based on the evaluation, the rule-based ontology is promising with the results obtained and reported; however, to take care of concept ambiguity, which is common in natural language, there would have been need to map the concepts with any semantic based top-level ontology. In that way, a more robust result would have been reported.

There have been argument on which methods to detect false information within a given sample of data. While some literature described content-based method for machine learning based techniques, the context-based approach has been adjudged to be well suited for semantic technology based techniques. This is the motivation behind the OWL based semantic model developed by the work of Bani-Hani, et al., (2020) to detect false information from news article. However, the proposed model is the representation of the taxonomy concepts of news items in that the model lacks the capability to infer knowledge from the representation. Based on the analysis and position of this literature, the hybrid approach of false information detection would be more robust especially when there are handshake between the ontology knowledge representation and machine learning models.

Furthermore, the literature of Du et al., (2021) designed a machine learning based model to detect misinformation regarding the vaccine for human papillomavirus. The authors employed five models each with its own classification result. The results of some of the models were not satisfactory as most labelled data of misinformation or none-misinformation were reported as misinformation. Besides, it can be argued that binary forms of results is not sufficient as the contextual or hidden knowledge would not be taken into account. The authors concluded and suggested further feature engineering process to be carried out to improve the quality of results.

The threat of misinformation in this digital age is alarming taken the advantages of open social networks, and artificial intelligence based systems such as bots. The dimensions of its consequences if not properly addressed might be unimaginable. Consequently, several interested researchers have taken the bull by the horns to advance a robust research in order to mitigate the menace of misinformation. While existing approaches such as content based, and context based are being advanced to detect false information, some other literature come up with a hybridized approach such as knowledge and linguistic based approaches to solve the problem (Seddari, et al., 2022). The authors reported that hybrid strategies are promising in terms of semantic consideration, inference of hidden knowledge, and overall, better results accuracy.

On this age of digital news, the spread of fake news or misinformation in general is on the high rise. This development propelled the research of Ahmed and Ahmed (2023) to create an integrated approach of data collection, analysis, and classification to determine fake news in India solely employed machine-learning model. However, their dataset is limited to a particular location, which makes the results of the fake news detection model to lack diversity. The researchers affirmed that ontology techniques have the capacity to reports accurate information while dispelling the false claim made in an online news.

Bani-Hani, et al., 2022 in an attempt to solve the issue of fake news designed an ontology-based model primarily for semantic annotation of microblogging news items and news datasets. This process making approach is achieved in three strata; firstly, the classification of news entities as taxonomy, followed by semantic model (ontology) to developed the taxonomy and lastly, the sketchy use case scenario to demonstrates the implementation and evaluation based on some dummy news items. However, the development of Fandet ontology is not premised on any known or new analytical methodology. Besides, the contextual capability of Fandet is limited as the ontology is not communicating or mapping with any other synonym-based models or sources for robust results. More so, intersecting machine-learning techniques with the ontology model is most likely to churn out more accurate fake news detection's results.

The role of semantic techniques (such as ontology) in bridging the vocabulary gap when user and query search systems use different terms to represent the same intention is heartwarming in any given information retrieval system (IRS). In order to demonstrate the efficacy of this technology, the research of Sharma and Kumar (2023) proposed hybrid semantic data indexing system based on domain ontology and machine learning model. The computer science domain based ontology is used to determine the concepts for annotating unstructured documents. Similarly, multiple features ranking algorithm based on word-embedding technique was proposed to assign relevance weight to the annotations of unstructured document. However, an improvement can be considered by expanding semantic contextual possibility of concepts by integrating a lexical data source.

3. Proposed Methodology for the Improved Semantic Model

Every system or model developed must leverages on particular methodology(ies). Thus, the proposed ontological model for detection of misinformation in this research is premised on ontology engineering principle called Noy McGuiness methodology. The rationale behind the choice of this methodology is not farfetched. The methodology is identified as a viable methodology when it comes to ontology reuse. The methodology consist of seven steps as shown by Figure 1.

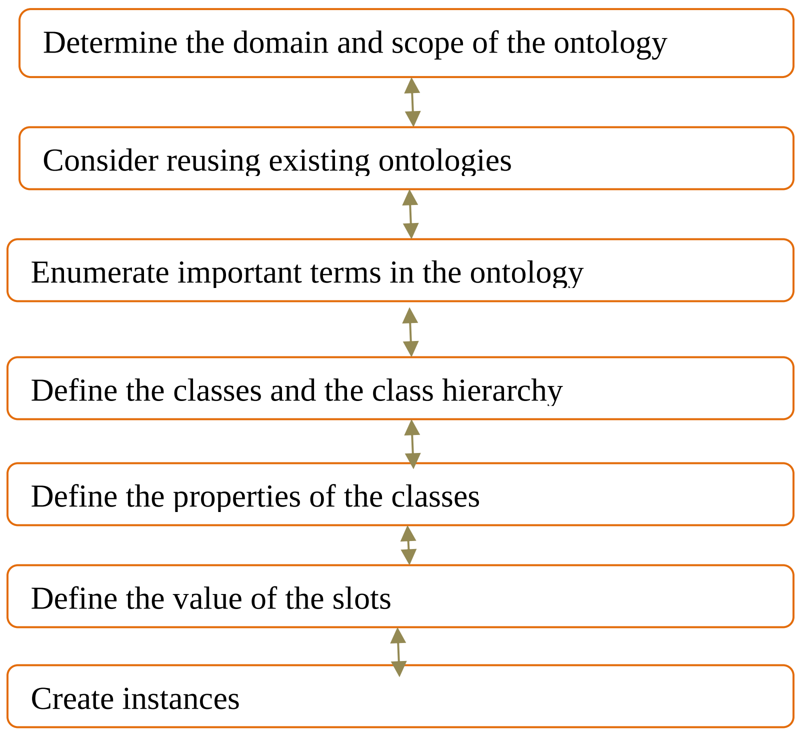

Figure 1: Ontology Development Stack of Noy and McGuiness Methodology, (Noy and McGuiness, 2001)

i. The first step has to with the determination of the domain or coverage area of the ontology by engineer(s). The domains for the proposed ontology in this research include health, politics, and science. The motivation for these choices is not farfetched because the digital online space has suffered a lot of false contents in the areas of health, politics, and science.

ii. The second step assist developer to ascertain the need to develop a new ontology or rather to reuse an existing ontology. This research work consider the reuse of the Fandet ontology taxonomical concepts and improve it to be more robust semantic model for detection of false information.

iii. The third step is to identify and list the needed concepts to model the ontology. Regardless of the ontology representation language, three categories of terms or concepts are always identified. In the case of OWL, they are classes, property, and individuals. Middle out strategy of top down manual design approach in collaboration with domain experts is considered in this research.

iv. The fourth step is to clearly defines the concepts that are classes, both superclass (hypernyms) and subclasses (holonyms) in hierarchical order. By default, the super and the root class of OWL is owl:Thing. For instance, both *Disease* and *COVID-19* are classes however, Disease can be regarded as superclass, while Disease COVID-19 is subclass.
v. The fifth step is to define other concepts, which is called the property of the class concepts. Every class has some properties or characteristics that define it. In the case of OWL, property is described as object property. It indicates the relation that exist between two or more classes for example, "*COVID-19* ***was caused by*** *5G technology"* the concept "***wasCausedBy***" is regarded as object property.
vi. The sixth step describes the features of class for example, the class "*5G_Technology*" has speed therefore, hasSpeed becomes it data property.
vii. The seventh step pointed out to another concept of a domain dataset, which usually spans out as values from the nonterminal classes. The values become the terminal concepts called individuals or instances in OWL representations.

Therefore, for this ontology to be developed based on the methodology earlier presented, and for this research work to achieve its objective of an improved semantic model, Figure 2 presents the overall conceptual framework.

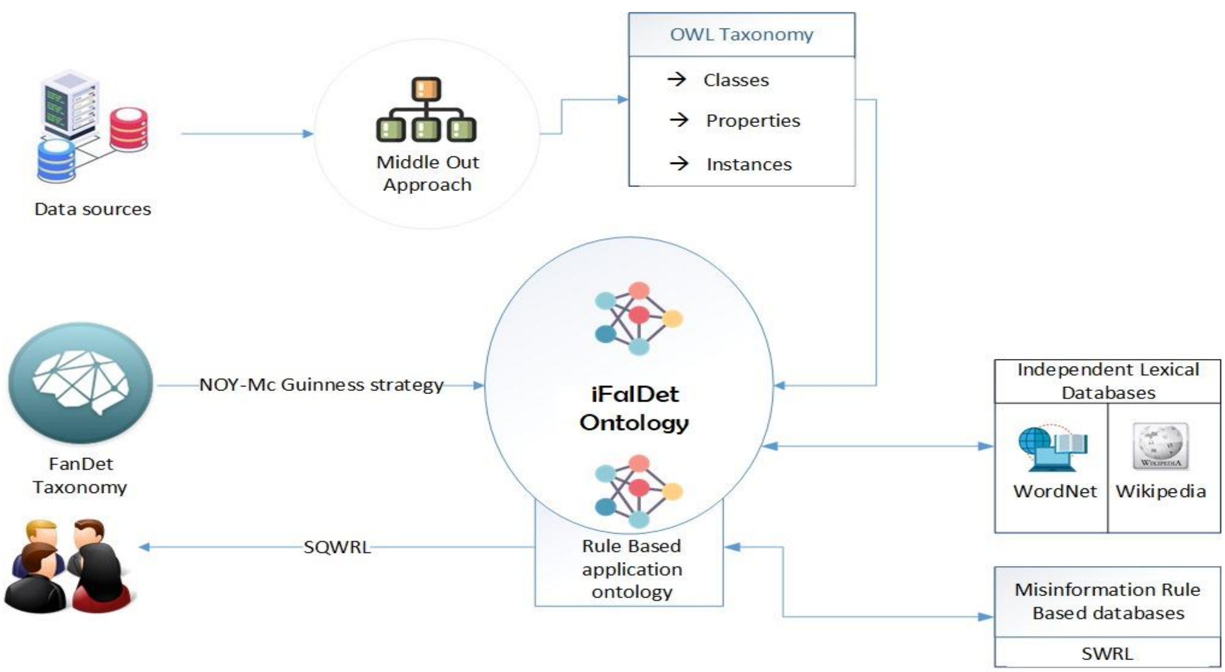

Figure 2: Conceptual Framework for the Improved Semantic Model

From the top left hand side of Figure 2, the required data are harvested from various sources, which are defined into three major components of OWL through a middle out strategy of specifying concepts. That is, the concepts are specified into classes, properties, and instances to form what can be analogously described as OWL taxonomy. The taxonomy as well as the concepts of the existing FanDet taxonomy are encoded based on the OWL knowledge representation language in accordance with the Noy-McGuiness methodology to model what is dubbed in this research as iFaIDet ontology. The acronym stands for improved False Information Detection ontology (iFaIDet).

3.1 Integration of Misinformation Rule Database into iFaIDet

In order to have a robust rule based application ontology, IFaIDet is expected to be enriched with a special database called misinformation rule based database encoded with semantic web rule language (SWRL). The database would be modeled based on first-order-logic (FOL) in order to unearth hidden knowledge imbued in any information thereby enriched the semantic strength of the ontology. Table 1 shows examples of information that need to be verified.

Table 1: Examples of Probable Misinformation

S/N	Ordinary Information	Verification Information
1.	COVID-19 was caused by 5G technology	• Information source: Does the website have a "Contact Us" or "About Us" page? Does it have a trusted domain, like ".edu" or ".gov?" What are the author's credentials • Understand what COVID-19 entails • Understand what 5G Technology entails • Check if there exist a scientific link between them
2.	COVID-19 Vaccines cause infertility	• Information source: Does the website have a "Contact Us" or "About Us" page? Does it have a trusted domain, like ".edu" or ".gov?" What are the author's credentials • What are the components of the vaccines and their effects? • What are the causes of infertility? • Does any relationship exist between infertility and components of vaccines?
3.	The U.S. Capitol police gave the protesters an "okay" to enter the Capitol.	• Information source: Does the website have a "Contact Us" or "About Us" page? Does it have a trusted domain, like ".edu" or ".gov?" What are the author's credentials • What are the expected roles of police during protest? • Scientific evidence from police to protesters.

Table 1 presents three likely examples of popular misinformation with the first two on science and health domain, and the third example on politics. The first column contains what is called ordinary information in this research because it has not been verified and validated to ensure its authenticity. In order to achieve this goal, column two provides additional related information as hidden knowledge that can unearth and detect the authenticity of the information. The additional information is modeled in FOL, and encoded in SWRL, to create the misinformation rule based database. Table 2 presents the overall representation of Table 1 in FOL.

Table 2: Modeled Examples of Probable Misinformation

Information (Informal)	FOL Representation (Formal)
COVID-19 was caused by 5G technology	$\forall x \forall y \forall z{:}\ \mathit{COVID\text{-}19}(x) \wedge \mathit{5GTech}(y) \wedge \mathit{InformationSource}(z)\ \exists \mathit{wasCausedBy}(x,y) \wedge \mathit{mayBeSource}(\mathit{Website}, z) \wedge \neg \mathit{has}(z, \mathit{ContactUs} \vee \mathit{AboutUs}) \wedge \neg \mathit{has}(z, \mathit{TrustedDomain}) \Rightarrow \neg \mathit{wasCausedBy}(x, y) \Leftrightarrow \mathrm{isCausedBy}(x, \mathrm{SARS\text{-}CoV\text{-}2Virus}) \wedge \mathit{isPartOfGerm}(\mathit{Virus}, \mathit{Germ}) \wedge \mathit{isACellular}(y, \mathit{CellularNetwork}) \wedge \neg \mathit{creates}\ (\mathit{CellularNetwork}, \mathit{Germs}) \Rightarrow \neg \mathit{wasCausedBy}(x, y)$

COVID-19 Vaccines cause infertility	∀x∀y∀z: COVID-19Vaccines(x) ∧ Infertility(y) ∧InformationSource(z) ∃causesInfertility(x,y) ∧mayBeSource(Website, z) ∧¬ has(z, ContactUs ∨ AboutUs)∧ ¬ has(z, TrustedDomain) ⇒ ¬caused(x, y) ⇔ isCommonlyCausedBy(y, OvulationProblems∨PoorQualitySemen) ∧areMajorIngredientsOf(MessengerRibonucliecAcid∧Lipids∧Cholesterol, x)∧ ¬caused(MessengerRibonucliecAcid∧Lipids∧Cholesterol, OvulationProblems∨PoorQualitySemen) ⇒ ¬caused(x, y)
The U.S. Capitol police gave the protesters an "okay" to enter the Capitol.	∀s∀t∀q: Police(s) ∧∃s U.S.CapitolPolice(s) ∧ Protesters(t)∧InformationSource(q) ∧ ∃allowProtestIn(s,Capitol)∧mayBeSource(Website, z) ∧¬ has(z, ContactUs ∨ AboutUs)∧ ¬ has(z, TrustedDomain) ⇒ ¬allowed(s, t) ⇔ isLocatedIn(US Capitol Building, Washington) ∧ isTheGovernmentHouseOf(US Capitol Building, USA) ∧requiresMaximumProtectionFrom(US Capitol Building, t) ∧ areWellTrainedToEnforceCivil(s,CivilRule)⇒ ¬allowed(s, t) ∧ breakTheRuleToEnterThe(t, Capitol)

In order to get a well detailed results of the information to be detected, which is presented in natural languages, it entails further, some contextual meaning to unveil the hidden knowledge in it. The entailments add contextual details to the information, which are formalized with the aid a logic-modeling tool (FOL) as clearly shown in the second column of Table 2. For example, the interpretation of the first information is as follow:

For all x, y, and z, where x isA COVID-19, y isA 5GTechnology, and z isAn Information Source, if there exist an information that: COVID-19 wasCausedBy 5GTechnology, and the information source may be website for example. If the website has no ContactUs or AboutUs page, and as well has no trusted domain such as .edu, .gov, it implies that COVID-19 was not caused by 5G Technology. Furthermore, based on knowledge, COVID-19 is caused by SARS-CoV-2 Virus, and the virus is part of germs. Conversely, 5G technology is a cellular network therefore, cellular network cannot creates germs. The proposed ontological model can concludes that COVID-19 isNotCausedBy 5GTechnology and affirm it as false information.

3.2 Collaboration of Independent Lexical Databases into iFaIDet

Considering the upper part of the bottom right hand side of Figure 2, two independent lexical databases: WordNet and Wikipedia are proposed for iFaIDet to have a handshake with the goal of enriching the ontology with semantic contents. While WordNet is anticipated to handle the synonyms of single concept or term, Wikipedia will take care of compound concept or concepts in phrases. For example, the concept U.S. is synonymous with United States or United States of America (USA). Similarly, the ontology should understand that coronavirus and COVID-19 are synonymous. If a relevant concept is inputted into the ontology as input string term and the term is not found in the ontology, such term(s) will be fact checked by the wordnet or Wikipedia in order to augment the robustness of the ontology and its operability.

4. Discussion

Following the outline methodology and the conceptual framework in the previous section, the proposed ontological based semantic model is being designed using protégé editor version 5.6.3 encoded in OWL. The proposed improved false information detection (iFaIDet) ontology is still working in progress; however, the growth considering considering the metrics shown by Figure 3 is promising.

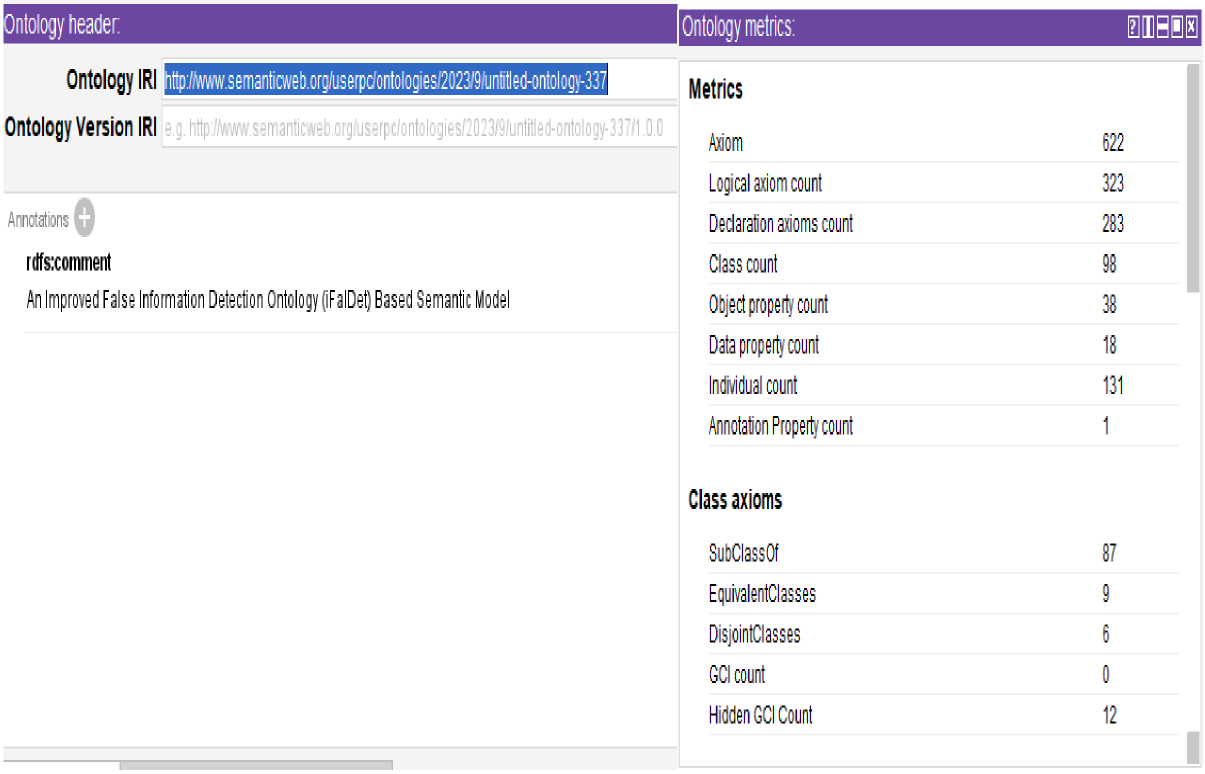

Figure 3: Metrics of the Proposed iFaIDet

As at the moment of writing this manuscript, a total axiom of 622 is built, while total number of classes, object properties, data properties, and individuals stands at 98, 38, 18, and 131 respectively for the proposed iFaIDet. This progress can be favorably compared with the statistics of other related works shown by Table 3.

Table 3: Statistics of OWL Ontologies Building Blocks

The Ontology	Number of Classes	Number of Object Properties	Number of Data Properties	Number of Instances
Amith and Tao, (2018)	116	20	6	-
Fandet (Bani-Hani, et al., 2022)	47	25	35	-
Bajpai and Chaturvedi, 2024	62	59	7	-
Proposed Ontology (iFaIDet)	98	38	18	131

As clearly shown by Table 3, while some existing ontologies does not explicitly account for instances or individual concepts of the OWL ontologies, some do not totally consider those concepts are within the component of instances. Instances are OWL componenets that can adequately represent terminal concepts for example, *US_Capitol_Police* may be regarded as terminal concept of another nonterminal concept such as *Police_Force*. From Table 3, the three ontologies does not explicitly reported any concepts that are instances; conversely, the proposed ontology has reported 131 number of individuals and still counting as the ontology's development progress. Figure 4 represents Table 3 in a more graphical structure for easier comparative analysis.

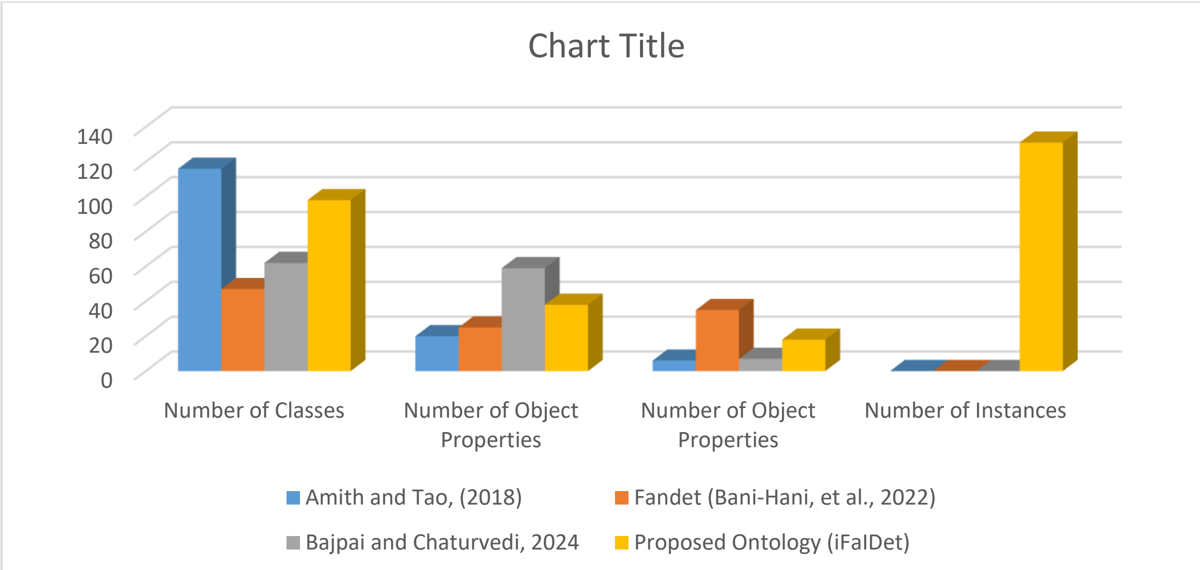

Figure 4: Graphical Representation of Proposed ontology's Building Blocks

Aside from the ontology designed by Amith and Tao (2018), iFaIDet class building block is promising and highly optimistic to surpass it. However, none of the reported ontology in the Table encoded instance concepts building block as iFaIDet, which is clearly shown by the first yellow bar from the right side of Figure 4. For emphasis and evidence sake, Figure 5 presents the fragments of number of classes and object properties have been encoded so far.

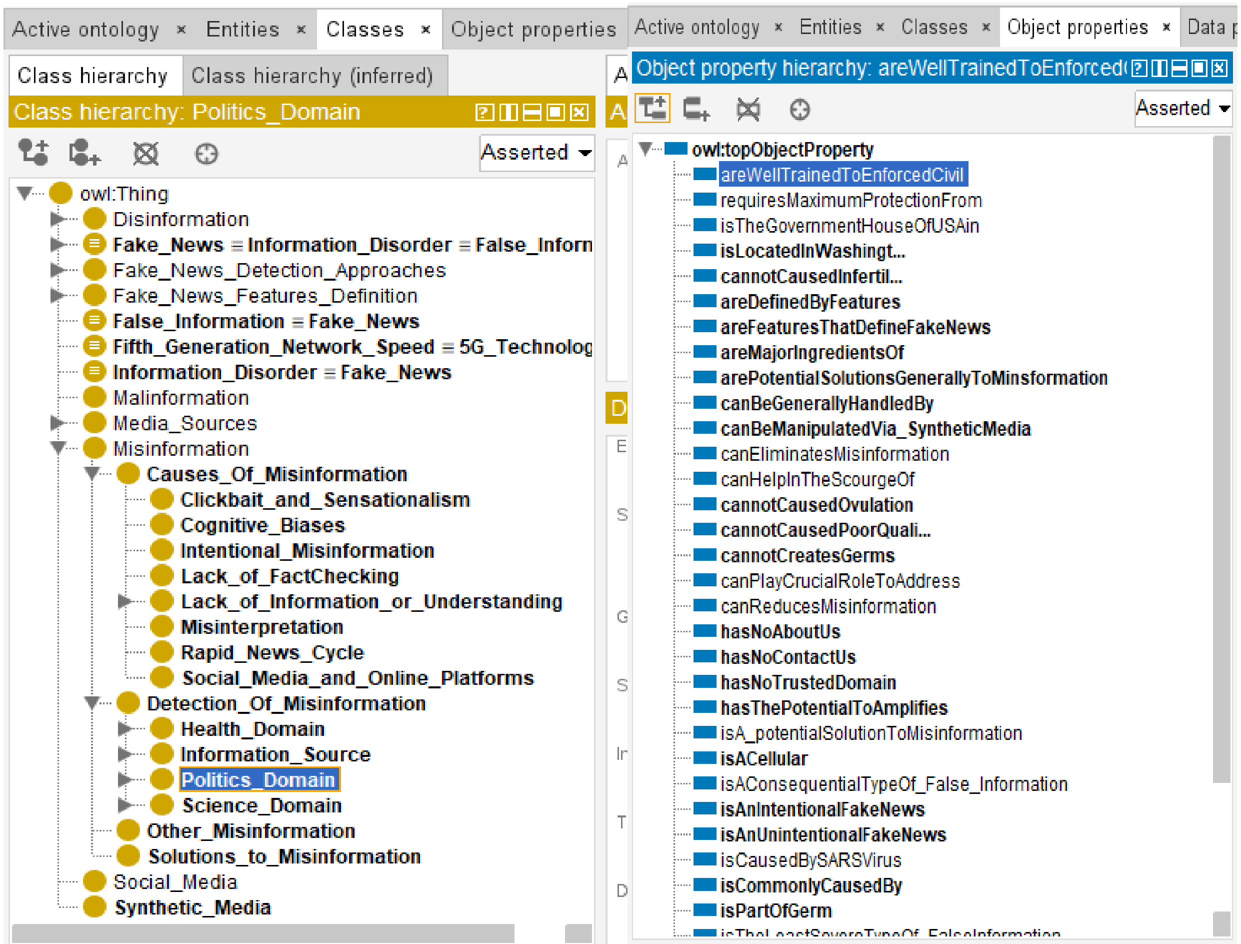

Figure 5: Samples of iFaIDet's Classes and Object Properties

The yellow color from the left hand side of Figure 5 shows some samples out of the 98 classes components that constitutes the ontology, while the terminologies appeared in blue from the right hand side of the Figure represents some samples of the 38 object properties. Similarly, the graphical objects of Figure 6 shows a fragment of some super classes (that are nonterminal) include the root class of every OWL class.

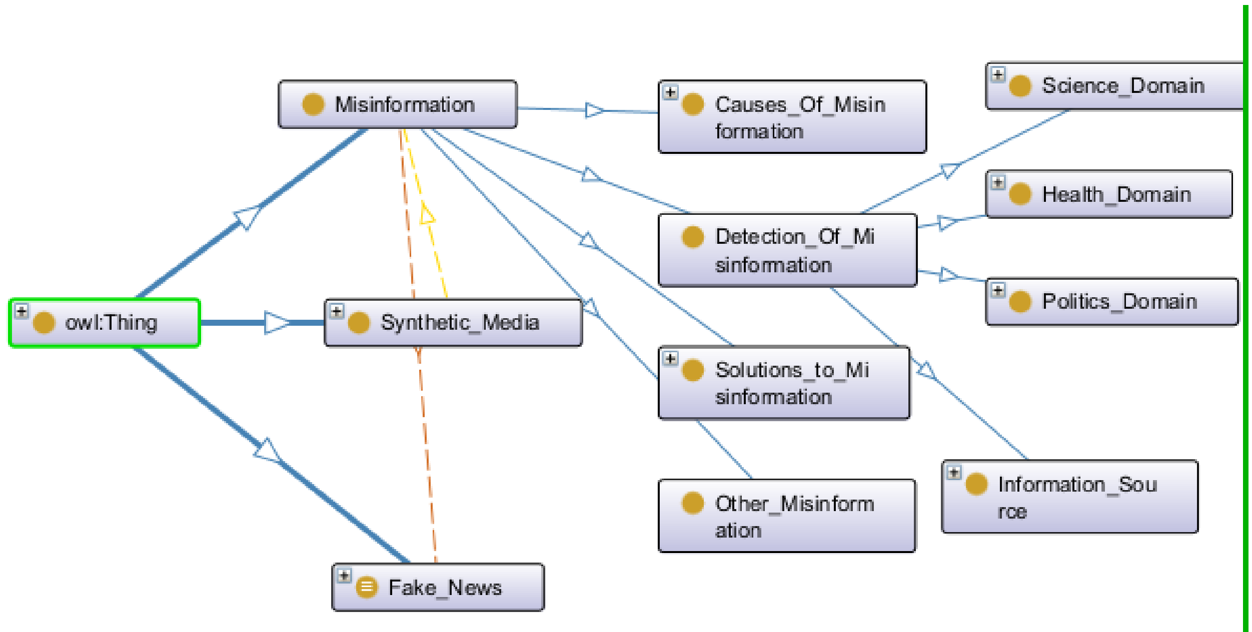

Figure 6: Graphical View of Misinformation Detection Concept

As shown by Figure 6, every OWL ontology has a default class called owl:Thing from where other user defined classes span from. As in the case of this research, Misinformation class is a superclass that spans from the default class. Other super classes equally span from it for example, classes science, health, and politics domains. Those rectangular shaped classes with plus (+) sign indicate that the classes are hypernyms, meaning they contain some other subclasses. Furthermore, Fake_News and Synthetic_Media classes were part of the existing Fandet ontology that was reused and being improved upon with the other classes of false information such as the misinformation concepts.

4.1 iFaIDet Ontology Structural based Evaluation

Furthermore, the results of iFaIDet's evaluation based on structure is presented by Table 4 considering the eight popular metrics against the average and median values of 1413 OWL ontologies (Aminu, et al., 2022; Sicilia, et al., 2012; Wang and Wang, 2018). The eight metrics are classes' numbers (*cn*), individuals' numbers (*in*), object properties (*op*), root classes' numbers (*rcn*), average population (*ap*), utilization of class (*uc*), schema deepness (*sd*), and diversity of relationship (*dr*).

Table 4: iFaIDet Structural Evaluations

Ontology Structural Metrics	iFaIDet	1413 OWL Ontologies Average	1413 OWL Ontologies Median
classes' numbers (cn)	98	36.11	6
individuals' numbers (in)	131	28.13	6
binary relation properties (op)	38	24	0
root classes' numbers (rcn)	23	6.69	5
average population (ap)	3.45	1.34	1
utilization of class (uc)	0.88	0.54	0.72
schema deepness (sd)	0.89	0.34	0
diversity of relationship (dr)	0.30	2.78	1

Based on the metrics, the results of iFaIDet structural evaluations are presented by Table 4 against the existing ontologies. The results of the first four metrics (that is, *cn, in, op and rcn*) for iFaIDet as earlier shown by Table 3 and Figure 4 are far higher than the average and median values of the existing 1,413 OWL ontologies obtained from the repository of swoogle. Thus, the results validate the proposed ontology is promising. Metric average population is obtained by dividing the absolute values of individuals' sizes by classes' sizes; that is, $ap = |in| / |cn|$ which is equal to 131/38. Similarly, utilization of classes is obtained by dividing the absolute values of the number of classes that have a minimum of one individual(s) by the size of classes; that is, $uc = |c| / |cn|$ which is equal to (98-12)/98 = 86/98. $|c|$ is calculated by subtracting numbers of classes without individuals, which is 12 from the total number of classes. With 0.88 for *uc*, it indicates that the ontology's individuals adequately utilized the classes. More so, *sd* is calculated by dividing the number of subclasses by the total number of classes; that is, $sd = |nsc| / |cn|$ which is 87/98. Subtracting the number of root classes from number of the classes derives the number of subclasses; that is, $|nsc| = |cn| - |rcn|$. The value of *sd* clearly determine the robustness of ontology in terms of how the classes are spread to form the ontology inheritance graph. A higher value obtained for *sd* compared to the benchmark values in the Table 4 shows that the ontology is deep. Else, if the obtained value of *sd* is less than the benchmark values, such ontology is described to be flat.

Finally, the last metric considered (*dr*) is obtained by dividing the properties number by the summation of number of subclasses and properties number. That is, $dr = |op| / |nsc+op|$ (38/(87+38)) = 38/125. This last metric intends to signify the multiplicity of ontology's properties (relations) which is based on theory that ontology consist many relations other than meronyms and hypernyms which are class inheritance relations. Therefore, the value of *dr* in this research is less than the average and median values. Since the total numbers of object properties (38) are the user defined relations and the ontology is still work in progress, it can be concluded that the value for diversity of relationship will be better. In conclusion, the structure of iFaIDet is evidently better than the average and median values of the existing 1,413 OWL based ontologies.

5. Conclusion

In this research work, an improved semantic model is proposed by developing an OWL based ontology capable to detect false information (misinformation) considering the domains of health, science, and politics. The proposed ontology, which is christened iFaIDet in this work, leverages on the standard Noy-McGuiness ontology development methodology. The choice of the methodology is owing to its completeness of predevelopment, development, and postdevelopment stages. Besides, the iFaIDet is an improvement on the existing Fandet ontology's taxonomy consequently, concepts have to be reused, as such, and the methodology is best suited for ontology reuse, which is a cardinal principle of ontology development. The ontology is developed using the protégé 5.6.3 editor by following the conceptual framework rigorously. The consistency of the ontology is validated by employing the use of HermiT 1.4.3. 456 reasoner against others such as EKL and Pellet owing to its robustness in checking the semantic context of the ontology's concepts. The iFaIDet is still a work in progress ontology however; it contains 622 axioms and 98 classes. The ontology has been subjected to structured based evaluation, which uses eight metrics to make comparison with existing 1,413 OWL based ontologies. Except, the last metric, which is diversity of relationship that turn out a low performance, iFaIDet is promising because it shows superiority considering the other seven metrics over the existing ontologies.

The work equally modeled the misinformation database of the ontology based on FOL in order to unveil hidden knowledge that would strengthen the ontology for resound inferences. Presently, the FOL models are being encoded in SWRL and to be queried with SQWRL. In addition, as part of the future work, an algorithm will be designed that would enable the domain ontology (iFaIDet) to collaborate with other top level semantic based lexical databases such as WordNet and Wikipedia. While the former assist the ontology to check the semantic relation of user string concept to the domain under consideration, the latter would assist to handle user string concepts that are compound terms or phrases. At the moment, the design framework is what is presented in this research. The domain ontology itself is still under development; at the end, the axioms, and the OWL components will be exponential grow. At the successful completion of the research, it is anticipated to have what is described as a robust rule based application ontology that would detect misinformation by taking into account contextual knowledge of user's query. The designed framework can as well be adapted to other domains aside politics, science, and health.

References

Ahmed, J., & Ahmed, M. (2023). Classification, detection and sentiment analysis using machine learning over next generation communication platforms. *Microprocessors and Microsystems, 98*, 104795.

Alenezi, M. N., & Alqenaei, Z. M. (2021). Machine learning in detecting covid-19 misinformation on twitter. *Future Internet*, *13*(10), 244.

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, *13*(1), 30.

Almaliki, M. (2019, April). Online misinformation spread: A systematic literature map. In *Proceedings of the 2019 3rd international conference on information system and data mining* (pp. 171-178).

Amith, M., & Tao, C. (2018). Representing vaccine misinformation using ontologies. *Journal of biomedical semantics*, *9*, 1-13.

Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2020). A Review on Ontology Development Methodologies for Developing Ontological Knowledge Representation Systems for various Domains. *International Journal of Information Engineering and Electronic Business*, *12*(2), 28–39. https://doi.org/10.5815/ijieeb.2020.02.05

Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2021) Modeling Competency Questions Based Ontology for the Domain of Maize Crop: SIMcOnto. ICACA 2021. Applied informatics and Information System, Springer Book Series

Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2022) "MaCOnto: A Robust Maize Crop Ontology Based on Soils, Fertiliz-ers and Irrigation Knowledge". *Intelligence Systems with Applications*, 200125. DOI:

https://doi.org/10.1016/j.iswa.2022.200125

Bajpai, S., & Chaturvedi, A. (2024). Toward Detecting Rumor Initiator in Online Social Networks Using Ontology-Driven Model. *Arabian Journal for Science and Engineering*, 1-17.

Bani-Hani, A., Adedugbe, O., Benkhelifa, E., & Majdalawieh, M. (2022). Fandet semantic model: an OWL ontology for context-based fake news detection on social media. *Combating Fake News with Computational Intelligence Techniques*, 91-125.

Bani-Hani, A., Adedugbe, O., Benkhelifa, E., Majdalawieh, M., & Al-Obeidat, F. (2020, November). A semantic model for context-based fake news detection on social media. In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1-7). IEEE.

Braşoveanu, A. M., & Andonie, R. (2021). Integrating machine learning techniques in semantic fake news detection. *Neural Processing Letters*, *53*(5), 3055-3072.

Broda, E., & Strömbäck, J. (2024). Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review. *Annals of the International Communication Association*, 1-28.

Capuano, N., Fenza, G., Loia, V., & Nota, F. D. (2023). Content-based fake news detection with machine and deep learning: A systematic review. *Neurocomputing*, *530*, 91-103.

Chen, M. Y., Lai, Y. W., & Lian, J. W. (2023). Using deep learning models to detect fake news about COVID-19. *ACM Transactions on Internet Technology*, *23*(2), 1-23.

Dame Adjin-Tettey, T. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent arts & humanities*, *9*(1), 2037229.

Du, J., Preston, S., Sun, H., Shegog, R., Cunningham, R., Boom, J., ... & Tao, C. (2021). Using machine learning–based approaches for the detection and classification of human papillomavirus vaccine misinformation: Infodemiology study of reddit discussions. *Journal of Medical Internet Research*, *23*(8), e26478.

Hassan, I. (2023). Dissemination of disinformation on political and electoral processes in Nigeria: An exploratory study. *Cogent Arts & Humanities*, *10*(1), 2216983.

Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., ... & Curtis, B. (2021). Bots and misinformation spread on social media: Implications for COVID-19. *Journal of medical Internet research*, *23*(5), e26933.

Jabardi, M. H., & Hadi, A. S. (2021). Ontology Meter for Twitter Fake Accounts Detection. *International Journal of Intelligent Engineering and Systems*, *14*(1), 410-9.

Jiang, Y. (2020). Semantically-enhanced information retrieval using multiple knowledge sources. *Cluster Computing*, *23*(4), 2925-2944.

Hunt, K., Agarwal, P., & Zhuang, J. (2022). Monitoring misinformation on Twitter during crisis events: a machine learning approach. *Risk analysis*, *42*(8), 1728-1748.

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, *61*, 32-44.

Petratos, P. N. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons*, *64*(6), 763-774.

Mazepa, S., Banakh, S., Melnyk, A., Pugach, S., Yavorska, O., & Golota, N. (2021, September). An ontological approach to detecting fake news in online media. In *2021 11th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 531-535). IEEE.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-Muhtadi, J., & Bouras, A. (2022). A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. *IEEE Access*, *10*, 62097-62109.

Sharma, A., & Kumar, S. (2022). Shallow neural network and ontology-based novel semantic document indexing for information retrieval. *Intelligent Automation & Soft Computing*, *34*(3), 1989-2005.

Sharma, A., & Kumar, S. (2023). Machine learning and ontology-based novel semantic document indexing for information retrieval. *Computers & Industrial Engineering*, *176*, 108940.

Sicilia, M. A., Rodríguez, D., García-Barriocanal, E., & Sánchez-Alonso, S. (2012). Empirical findings on ontology metrics. *Expert Systems with Applications*, *39*(8), 6706–6711. https://doi.org/10.1016/j.eswa.2011.11.094

Van Der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature medicine*, *28*(3), 460-467.

Wang, Yi, & Wang, Y. (2018). Citrus ontology development based on the eight-point charter of agriculture. *Computers and Electronics in Agriculture*, *155*(October), 359–370. https://doi.org/10.1016/j.compag.2018.10.034

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social science & medicine*, *240*, 112552.

Zhao, Y., Da, J., & Yan, J. (2021). Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management*, *58*(1), 102390.