Ontology Synthesis Using Semi-Automatic Semantic AI Using Deep Learning and Reinforcement Learning

Anubrat Bora, Gerard Deepak
Department of Computer Science and Engineering
Manipal Institute of Technology, Bengaluru
gerard.deepak.cse.nitt@gmail.com

Abstract

The need for automatic modeling of ontologies specifically in the era of Web 3.0 as ontologies form the structural metadata of the Web 3.0 and specifically for domains, although which may seem common like environmental studies has been neglected and formal ontologies are not present. Such domains attempt a model for generating ontologies using auxiliary knowledge from the Web 3.0 by seed term pool generation from the data set generation and its classification using a Bi-LSTM model. Subsequently, the AdaBoost classifier classifies the data set Pearson’s correlation coefficient helps in extracting features through federation between two classifiers provisioning sedated learning. Google's Knowledge Graph Search API also contributes to intensification of incorporating additional information into the model, thereby providing a bold knowledge infrastructure into the model. Morista's overlap index helps in semantics-oriented relevance computations and thereby helps in reasoning out through semantic similarity measures. The Imperialist Competitive Algorithm serves as a metaheuristic optimization strategy in the proposed model. The Yago and the Google Knowledge Graph API helps facilitate in adding to the density of the auxiliary knowledge of the model an overall precision of dash and estimation of F-measure with the lowest value of FDR. A proposed framework which is a best-in-class model for ontology automatic ontology generation that targets environmental journalism as the preferred domain.

Keywords: Bi-LSTM, AdaBoost, Knowledge Graph, Semantics AI, Deep Learning, Knowledge Centric Approach, Ontological Model

1. Introduction

A key component in forming the basis of Web 3.0 (the Semantic Web) is an ontology. Ontology is the structural metadata that supports the information and knowledge that is available online in the context of Web 3.0. It tries to make content machine-readable and interpretable, going beyond the traditional HTML-based web, which primarily focuses on providing information in a human-readable fashion. By offering an organized framework for expressing concepts, their characteristics, connections, and logical principles, ontology promotes a deeper and more uniform comprehension of data on the internet. The smooth integration and sharing of data and information across numerous applications and disciplines is made possible by this organized knowledge representation. Ontology-driven data will enable intelligent systems as Web 3.0 develops, enabling sophisticated features like context-aware search, automated reasoning, and improved data interoperability. Ontology is essentially the foundation of the next generation of the web, where data is no longer just content but rather structured, connected knowledge that can be processed by robots. This will revolutionize how we interact with data and the internet in general. When combined with agents, ontologies can be employed in a variety of fields since they can bridge communication gaps by supplying a common language and act as a user's personal assistant by classifying and recommending material. Semantic retrieval, the semantic web, data enrichment and mining, detection of topic and tracking, natural language processing (NLP), semantic search over diverse networks, knowledge engineering along with management, digital trade, sentiment evaluation, scientific communication, bioinformatics, biomedical, and human-computer interaction are some of the research fields that can benefit from using the proposed ontology, which is based on modeling of topic. Building an ontology is also advantageous for ontology-based recommendation systems, information retrieval, link prediction, classification, clustering and Association Rule Mining (ARM). Building strong AI systems requires the capacity to transfer knowledge into and out of various representation languages. This is because it makes the process of creating new knowledge bases easier and faster, and it also makes it easier for current ones to share knowledge. In order to facilitate automatic knowledge exchange, we examine in this study the challenge of giving declarative languages semantics through formal ontologies. One suggestion for handling the intrinsic heterogeneity found in knowledge derived from many sources is to use formal ontologies. The concept of a formal ontology is defined differently by several methods; they include taxonomic hierarchies of classes, vocabularies of concepts defined by text that can be read by humans, and sets of formal restricting axioms. The degree of commitment made by the communicating agents to the shared ontology is another difference. This can range from having all agents adhere to the same common ontology (the standardization approach) to having a network of mediators and facilitators that allow the various ontologies of the agents to be translated. We take the logical theory approach of an ontology for our purposes, and the defining of our semantics will be much aided by the constraining axioms. A common ontology that is expressive enough to interpret the ideas in each agent's ontology must exist, even while we permit the communicating agents to have their own declarative languages and ontologies. The term "ontological engineering" describes a collection of activities related to the creation of ontologies, their life cycle, their building techniques, and the toolkits and languages that underpin them. Ontologies and Ontological Engineering have garnered more and more interest in recent years. These days, ontologies are being utilized extensively in computer science, knowledge engineering, and artificial intelligence (AI) applications related to database integration, e-commerce, knowledge management, biological informatics, natural language processing, intelligent information integration, information retrieval, and even emerging fields such as the Semantic Web. A formal, clear specification of a common understanding is called an ontology. which, after identifying the pertinent concepts of a phenomenon, refers to an abstract representation of that occurrence in the real world. Explicit refers to the kinds of ideas that are employed as well as the limitations placed on their application. The term "formal" describes the requirement for machine-readable ontology. The idea that an ontology obtains consensus knowledge—that is, knowledge that is accepted by a group rather than the private life of an individual—is reflected in the concept of shared. Developers can reuse and exchange application domain knowledge by using ontologies, which give diverse software platforms a common vocabulary. This eventually frees them up to focus on the task at hand and the domain's structure, rather than being distracted by implementation details. One significant turning point in the development of ontologies has been the creation of the Semantic Web. Tim Berners-Lee claims that the Semantic Web is an expansion of the current Web, giving data a clear meaning and improving the ability of machines and humans to collaborate. Since common knowledge-components can be used to accomplish this cooperation, ontologies have emerged as crucial tools in the Semantic Web's development. The emphasis has switched from closed, comparatively applications with poor data to systems and searching applications, integrating, utilizing massive volumes of information which are already available as a result of the expanding availability of information. Since ontologies offer the semantic foundation for intelligent data access, integration, exchange, and utilization, this technology has truly grown to be strategic. Ontologies offer a critical technological component that facilitates web interoperability and allows for the semantic integration of processes and data. In fact, ontologies have become so popular in recent years that even "traditional" businesses, for instance, IBM have launched their own ontology management systems. Additionally, taxonomies and ontologies are ranked third in the top 10 technologies of the past few years by market research firms like Gartner. Thus, right now we are moving into a new stage in which ontologies are being created at a higher rate and with more complexity than before. As a result, we now face both fresh chances and fresh difficulties. Today, we can create a fresh wave of intricate systems that can fully utilize the never-before-seen availability of massive amounts of data as well as enormous, reusable semantic resources, as ontologies grow in quantity, complexity, and size. These technologies are expected to offer novel features in the developing semantic web, business-to-business relationship automation, and corporate intranets. Simultaneously, we have a difficulty: the existing approaches and technologies, which date back to the era of shut systems with inadequate data which were merely insufficient to facilitate the entire application development process for this novel category of application of semantics. The ability of knowledge engineers to capture, represent, evaluate, and use knowledge is necessary for a knowledge-based system to succeed. Methodological development of an ontology engineering project requires defining and standardizing the life cycle models to be used in the project, as well as the development and maintenance processes (from requirements specification to upkeep of the finished product). Software engineering is the foundation upon which methodologies for creating knowledge-based systems (KBSs) have been presented. The whole life cycle of the ontology creation process, including best practices for each step of the process, is provided by these methods, which take into consideration the unique peculiarities of this kind of system. The ontology development process is different from the ontology lifecycle model in that it dictates the order in which the actions performed in the former must be completed. Among the approaches mentioned above, methodology is the one we concentrate on in this section in order to explain and provide examples of these ideas. For journalistic and other news-related objectives, large and open data sources can be made more easily accessible through the use of semantic knowledge graphs and other semantic technologies. They provide a common framework and auxiliary materials for exchanging, handling, and preserving factual information at the syntactic and semantic levels. Thus, these knowledge graphs provide a means of improving the integration and understanding of large, open, and other data sources. They enable the integration of the incredibly diverse material found on the Internet and increase its accessibility for news-related and journalistic objectives. Semantic knowledge graphs have a lot in common with the web of semantics. The goals are to give a broad overview of the field, explore its definition, and identify areas that require more study and investigation. Semantic knowledge graphs and news are both understood in a broad way. We cover semantic knowledge graphs as well as the applications of enabling semantic technologies for semantically linked (open) data and the semantic web, including RDF, OWL, and SPARQL. For journalistic and other news-related objectives, large and open data sources can be made more easily accessible through the use of semantic knowledge graphs and other semantic technologies. They provide a common framework and auxiliary materials for exchanging, handling, and preserving factual information at the syntactic and semantic levels. Thus, these knowledge graphs provide a means of improving the integration and understanding of large, open, and other data sources. They enable the integration of the incredibly diverse material found on the Internet and increase its accessibility for news-related and journalistic objectives. Semantic knowledge graphs have a lot in common with the Semantic Web.

Motivation: Building a structured knowledge representation framework for the Semantic Web is a complex procedure, which is what is involved in the production and synthesis of ontologies utilizing semi-automatic models to populate Web 3.0 structurally at its metadata. Web 3.0's objective is to make web content machine- and human readable. This is accomplished by developing ontologies that specify the ideas, connections, and characteristics found domains. To create these ontologies, semi-automatic models make use of both automated procedures and human knowledge. These models collect and combine instances, or data points or examples, from different Web 3.0 knowledge resources. Databases, websites, connected data sources, and other organized data repositories are a few examples of these information resources. The semi-automatic nature of ontology generation implies that although certain portions can be automated, human intervention is frequently required to guarantee precision and accurate capture of domain-specific subtleties. Within the ontology, instances from various sources are examined, categorized, and arranged into ideas and relationships. In this procedure, properties and their attributes are defined, hierarchies and similarities are found, and logical rules governing the domain-specific knowledge are established. The final ontology is a structured metadata framework that improves Semantic Web information interpretation, automated reasoning, and data interoperability. It expands the capabilities of Web 3.0 by enabling robots to process, reason with, and integrate data from many sources more effectively. Thus, a crucial first step toward achieving the goal of a more intelligent and linked web—one in which information and knowledge are easily reachable and understandable to both humans and machinery—is the synthesis of ontologies using semi-automatic models.

Contributions: The primary contributions of the proposed framework are - Metadata generation from the seed term pool which is obtained from the dataset itself and the metadata categorization using the Bi-LSTM classifier is quite novel, AdaBoost classifier and its federated co learning of the entities from the classified metadata which cover a Bi-LSTM classifier where feature selection takes place from the classified instances using the Pearson's Correlation Coefficient to cooperate with AdaBoost classifier to classify the dataset is quite novel to the model, the Tversky's Index and Morisita's Overlap Index facilitate semantic similarity computation. The Google knowledge graph API Yago serves as knowledge intensification repositories for knowledge addition and attenuation. Imperial Competitive Algorithm is the best-in-class optimization algorithm for this aspect.

Organization: The remainder of the paper is structured. Section 2 is devoted to the Literature Review. In Section 3, the Proposed Methodology is covered. The implementation is thoroughly explained in Section 4. Section 5 shows the results and performance evaluation. In Section 6, the paper finally ended.

2. Related Works

2.1 Knowledge-Based System

Usmani et al. [1] have stated application development for smart cities and infrastructure development are heavily reliant on data supplied as a kind of geographical context and comprehensive constructing knowledge. The adoption of open-formats in information engineering and data capture speeds up the development of geospatial technologies for knowledge-based systems and sustainable urban environments. In this field, BIM and GIS technologies are well-known to be leaders. Nora Yahia et al. [2] have stated how XML, or Extensible Markup Language, is a data transmission standard that may be applied to several fields. It facilitates data interchange between parties by establishing a shared understanding of the fundamental ideas in the field. XML provides syntactic coverage but does not provide reasoning assistance. A semantic representation of domain knowledge, supported by ontology, can facilitate effective reasoning and expressive capability. The Web Ontology Language (OWL) is one of the popular ontology languages. Classes, attributes, axioms, and instances can be used to describe domain knowledge for use in a distributed context like the World Wide Web. Davies, J., Studer, R., & Warren, P. (Eds.). [3] have described the contents of Web documents, the Semantic Web integrates the data-centric, configurable XML with the descriptive languages RDF. By writing more intelligent software systems, it has made it possible to automate the analysis and utilization of web-based data thanks to these machine-interpretable descriptions. It has provided an extensive summary of important semantic knowledge technologies and research is given by Semantic Web Technologies and an explanation of information extraction, ontology administration, and semi-automatic ontology development.

2.2 Fuzzy Logic

Wei Chen ˈQing Yang et al., [4] stated that because traditional ontology is unable to capture the fuzziness and ambiguity of domain knowledge, it has not been able to support modeling in the real world. Determining if a concept belongs in a particular sector might be difficult. Then, one solution to knowledge uncertainty is to use fuzzy logic into an ontology. It is possible to handle fuzzy knowledge with fuzzy ontologies. However, using the pre-established concept hierarchy to manually produce fuzzy ontology is quite challenging and sometimes requires a significant amount of expert interpretation based on some implicit subject knowledge. Thus, research on automatically generating technology would be prudent. In the process of automatically creating a fuzzy ontology, the idea hierarchy construction should be considered. Ontology learning can be aided by various strategies, such as statistical models and association rules. A formal method for analyzing data and presenting information is formal concept analysis. Because it offers an ontology class—an explicit hierarchy relation—and a distinct hierarchy structure among ideas, the resultant idea lattice is seen as a useful tool for machine learning. However, the concept lattice's reflection captures the precise relationship between an entity and its attribute; thus, it is not suitable for building a fuzzy ontology model. The goal of this study is to enable automatic or semiautomatic modeling for fuzzy ontology generation by utilizing the fuzzy concept lattice from the fuzzy formal context. The next portions of the study deal with the formal concept analysis of fuzzy, which includes fuzzy idea lattices, fuzzy formal contexts, and fuzzy clustering-based fuzzy concept hierarchy construction for fuzzy ontology classes. At last, a fuzzy ontology is produced. Ramaprasad, A., & Syn, T. [5] have developed an ontological meta-analysis and synthesis as a technique to map, review, and visualize the body of research in an area in a systematic, logical, cumulative, and systemic manner. The approach brought to light the domain's significantly emphasized bright spots, mildly stressed light spots, unemphasized blank spots, and unnoticed blind spots.

2.3 Deep Learning & Machine Learning

Grubic, T., & Fan, I. S [6] have provided a new method of ontological reasoning based on deep learning instead of formal reasoning based on logic, a novel statistical relational learning model based on deep recursive neural networks along with the experimental proof that it can readily match or surpass current logic-based reasoning challenge in ontology. More specifically, we ran many large standard benchmark dataset comparisons between our implemented system and RDFox, one of the strongest logic-based ontology reasoners available today and discovered that although our system was up to two orders of magnitude faster, it still achieved high reasoning quality. Lytvyn et al. [7] have proposed a method for creating a computer system that builds an ontology foundation automatically. With the aid of the CROCUS (Cognition Relations or Concepts Using Semantics) software model, an architecture of synthesis for the ontology system is developed. The primary system modules are explained along with their roles. SDK's choice to realize the system is warranted. The suggested system's application can automatically fill an ontology of subject matter. Bourguin et al. [8] have examined the benefits and viability of a method intended to use domain ontologies to produce ontologically explicable classifiers. The method is demonstrated using the Pizzas ontology, which is employed to build a synthetic picture classifier that can offer visual explanations for a range of ontological characteristics. To put the strategy into practice, ontological tensors produced from the ontology represented in Description Logic are used to finish a DL model. Kulmanov et al. [9] have proposed a summary of the techniques that use ontologies to calculate similarity and integrate them into machine learning techniques; specifically, how ontologies provide limits that enhance machine learning models and how metrics of semantic similarity, embeddings in ontology can take advantage using prior understanding of biomedical ontologies are described. Sulaeman et al. [10] have stated six supply chain ontology models and the three comparison framework criteria were found. Following application, a framework has been found, which ought to shed light on prospective directions for further investigation. Overall, the conclusion is that ontologies are not yet a viable solution for information system interoperability issues. Their research has demonstrated that comprehending supply chain reality is given less weight than the structure and organization of supply chain human knowledge. Smith et al. [11] have stated the factors that have led to the development of ontology in the field of information science and provide an initial assessment of the present usage of the term. In closing, some reasons for hope about the future cooperation between information scientists and philosophical otologists are offered. Mike et al. [12] have outlined the synopsis of ontology research and one of its most important fields of application, which is knowledge management (KM), a business discipline. Their primary contribution to the synthesis of knowledge management and ontology technology is a visual classification framework known as the Semantic Web Matrix. It provides a clear, business-oriented explanation of application scenarios for both distributed and centralized ontology-based knowledge management. As they wrap up our poll, they talk about the biggest obstacle to using ontologies, which is a lack of awareness about the social character of knowledge and the effects this has on initiatives aimed at distributed knowledge management. Janowicz et al. [13] have stated regarding the Semantic Web. They contend that semantic rails are necessary for the data train in the following. They make the point that, rather than fighting over interpretation authority, inductive and deductive approaches operate best when combined to make sense of facts and provide fresh insights. Al-Aswadi et al. [14] have developed several methods, frameworks, and difficulties in automatically creating ontologies from textual data. Furthermore, future directions for improving the ontology creation process by introducing methods that bridge the gap between shallow and deep learning (DL) are explored. Croce et al. [15] have suggested many techniques for automatically annotating 2D and 3D data that maximize the use of the collaborative, web-based annotation platform Aïoli as well as machine or deep learning. The suggested methods are intended to help heritage specialists with the hybrid annotation of architectural items. Through their efforts, public and private partners tasked with rehabilitation and conservation efforts will be able to: i) keep data safe while it transitions to new representation types; ii) disseminate relevant information through digital media and iii) exchange information about heritage on the web using an open-source approach. Hassanzadeh et al. [16] have addressed several challenges facing the Semantic Annotation, including concerns with scalability, diversity, and consistency in the content of various web sites, as well as multilingualism. One of the major issues in this sector is automating the annotation process because Semantic Annotation systems need to operate on a wide range of domains and dynamic contexts. This problem has been resolved using a variety of machine learning approaches, including supervised, unsupervised learning, and more recently, semi-supervised learning and active learning. In this work, we examine the key concerns in the field of Semantic Annotation and propose an inclusive layered classification of the obstacles involved. We also examine and assess machine learning applications that address issues with semantic annotation.

2.5 Natural Language Processing

Gacitua et al. [17] have talked about the initiatives that use a combination of machine learning techniques and a certain degree of natural language processing to identify concepts and relationships. Quantitatively assessing the value or correctness of the methods and combinations of strategies when applied to an ontology model can be a difficult problem, though. Their goal is to create a system that can be used to extract an ontology from a vast amount of text domains. Regarding assessing various NLP and ML methods in the context of ontology learning, this approach helps. Their preliminary trial validates our hypotheses regarding the practicality of our methodology. Konys et al. [18] have examined how to combine tools for learning ontology from texts in the knowledge base with used techniques and ontology learning algorithm outputs to create a single, intricate, multifunctional solution. The suggested knowledge base includes a reasoning system based on competency questions that allows users to define their own profiles of ontology learning tools and covers the various applicabilities of current methods for learning ontologies from text. Moreover, applied reasoning is offered during the validation stage. Van Rooijen et al. [19] have talked about their goal of automating the requirement specification. The goal is to discriminate between inexperienced and experienced users while taking advantage of subject expertise gathered from earlier system runs. As they permitted trained users to submit instances of descriptions of behavior, they allowed unskilled people to provide unstructured natural language descriptions. Their objective in both situations was to create formal requirements models that resembled statecharts. Behavioral ontologies were learned from requirements specification procedures involving trained users, and they are then applied to support requirements specification processes involving unskilled users. Adelkhah et al. [20] have outlined the suggested strategy for creating an ontology for NLP (natural language processing). Their approach was semi-automatic, combining machine learning and rule-based methods to build and assemble an ontology with bilingual (English Persian) concept labels (lexicon) and then manually assess it. Using this methodology, an entire ontology of 887 concepts, 88 relations, and 71 characteristics is produced for the natural language processing domain.

2.6 AI in Digital & Environmental Journalism

Malgieri et al. [21] have talked about empowering professionals worldwide to make better decisions and build deeper connections by placing sophisticated, intelligent solutions and medical gadgets that are aided in several ways by artificial intelligence (AI) in their hands. The study looked at the effects of incorporating artificial intelligence (AI) into the certification process for software used in medical equipment. The certification of software that is subject to change based on the data it analyzes has prompted discussions with certifying bodies and raised the prospect of using software to treat certain diseases. They discussed the three pillars of AI: reasoning, data management, and picture management. They also explained how ontologies enable the synthesis and management of an incredibly potent and rich map of all available data along with a focus on Deep Learning methods. Iannone et al. [22] have investigated methods for semi-automatically inducing idea descriptions. Specifically, they also demonstrated an algorithm that can use instances provided by domain experts to infer definitions in the ALC Description Logic, an OWL-DL sub-language. An empirical evaluation of the method's efficacy in comparison to previous algorithms is conducted through an experiment conducted in the document image interpretation area. Konys, A. [23] have shown the reliance of the semantic web on ontologies for power has made ontology learning from many data sources an extremely promising area of study. It attempted to construct ontologies either automatically or semi-automatically from provided data sources with a minimal amount of human labor. The subject of ontology learning has required knowledge systematization due to the vast array of accessible methodologies and their notable variations. The article has presented the author's ontological elaboration of ontology learning techniques and their aspects, offering formal, technological, and practical assistance to knowledge management-based approaches to ontology learning methods. Tosi et al. [24] have talked about the systematic literature review to provide answers to a set of research questions and sum up the current state-of-the-art to support the semantic annotation of web services currently. A predetermined process that includes automatically searching reputable digital libraries was adhered to during the review. Consequently, 35 primary papers in all were found to be pertinent. Nine more primary studies that were not included in the digital libraries' automatic search were found through a manual search. These 44 papers' pertinent data was taken out and compared to the chosen study topics before being published. Broussard, et al. [25] have stated that the implications of AI for journalism must be understood in the broader digital context of the media and public affairs. This shift to applications, algorithms, social media, and the like has fundamentally altered the nature of journalism as an institution by disrupting work routines, upending business models, and unleashing a plethora of information alternatives to news, among other things. Therefore, whatever the short-, medium-, or long-term impact of AI technologies, they could be viewed as a part of a larger narrative about the reconfiguration of journalism in relation to computation.

2.7 Other Novel Related Works

Mizoguchi et al. [26] have defined an ontology at the outset of the study. Subsequently, they examined the eight levels of ontology use in depth and then talked about the specific benefits ontology may offer when solving problems in the real world. Ontology classification was the next issue of discussion. The extent of ontology engineering was also presented. Lastly, they summarized their findings to provide an example of ontology engineering. Bhatti el al. [27] have stated about a new service ecosystem trend in which service providers might use business-related service delivery features like delivery and distribution to enhance their core offerings. One bottleneck in the service ecosystem will be the semantic service definition of services for the business service provision. By fusing machine learning and interactive visualization approaches, the Visual Semantic Analysis approach is proposed to facilitate modeling of semantic service descriptions semi-automatically. Additionally, two application scenarios are provided as an assessment of the Visual Semantic Analysis approach from the German Federal Ministry of Economics and Technology-funded THESEUS-TEXO project. Pavlick et al. [28] have stated new models of lexical and compositional meaning that are involved in current deep learning techniques. Novel strategies adopt a top-down approach, treating representation of sentences as primary and word and syntax representations as emergent, in contrast to conventional models of distributional semantics that take a bottom-up approach, defining sentence meaning by the application of explicit composition functions to word meanings. This article summarizes our current understanding of how well such representations capture lexical semantics, composition, and world knowledge. The intention is to promote more cooperation in examining the consequences of general-purpose semantic models and other similar representations. Yoon et al. [29] have suggested processing an ontology population using a method that involves calculating the overlap between concepts and examples. Although this method is simple, it performs well. A vital area of contemporary journalism that explores the intricate interactions between human activity and the environment is environmental journalism. It includes a wide range of reporting, analysis, and narrative with the goal of identifying, comprehending, and disseminating information on problems pertaining to climate change, biodiversity loss, environmental degradation, conservation initiatives, and sustainable development. This type of journalism is essential for educating the public, decision-makers, and interested parties about urgent environmental issues, exposing their root causes and effects, and promoting solutions that advance the health of ecosystems and society.

The Related works for Ontology Synthesis have been categorized as using Knowledge-Based System and Knowledge wikis, works on Fuzzy Logic, Deep and Machine Learning Paradigms, Natural Language Processing, AI in Digital and Environmental Journalism and other novel related works.

3. Proposed Methodology

Three distinct components make up the architecture of the suggested system: ontology building, classification, and data preparation. This approach bridges the gap between Web 2.0 and Web 3.0 for classification of environmental journalism and analysis based on information from numerous social media platforms. It is quite concrete and unique because it uses a Dynamic Ontology Modeling scheme based on the real-world merging of popular domain-relevant tweets, hashtags, and auxiliary knowledge from news articles. It is fascinating how different knowledge from a variety of sources has been compiled for the domain of environmental journalism.

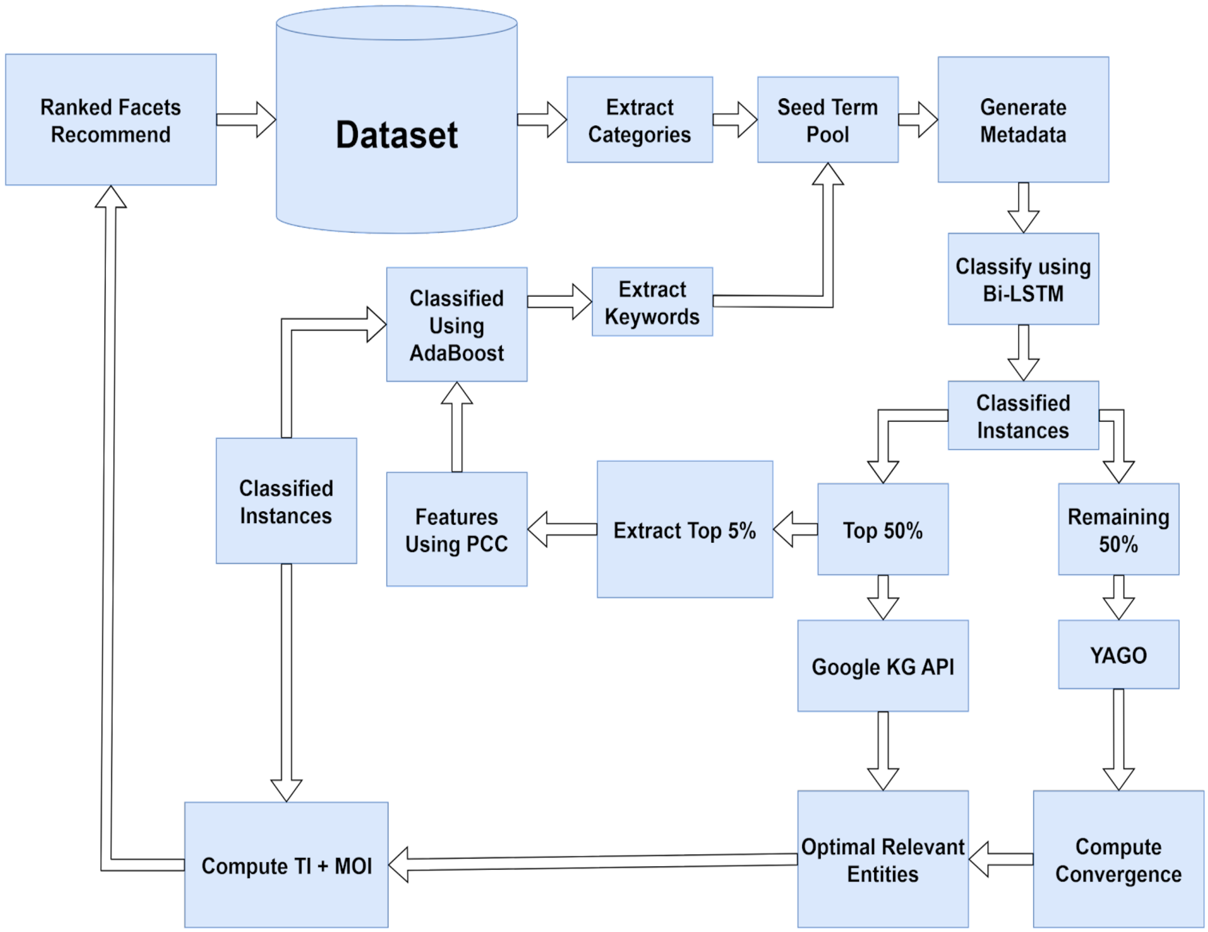


Figure 1: Phase 1 of the proposed architecture

Figure 1 depicts phase 1 of the proposed architecture of semantics-oriented ontology synthesis and generation for environmental policy as the domain of choice. The dataset of environmental policies is subject to extraction of keywords and categories, since the dataset is subject to categories, it is subjected to category extraction, and both these are integrated together to form an initial seed term pool. Seed term pool is subject to generation of metadata using a tool named Open Calais. The generation is achieved using Open Calais as the tool of choice. Open Calais is used as a web reference corpus and generates a large number of metadata into the model. Metadata being extensively large in volume, it cannot be handled as it is. So, it is subject to classification by employing a Bi-LSTM classifier, which is a bi-directional LSTM classifier. Bi-LSTM was chosen because it is a powerful classifier, a deep learning classifier that operates on the principle of auto handling selections and works wonders on diverse and exponentially large metadata. The classification itself from the Bi-LSTM is divided into two classification instances - Top 50% is first chosen and fed into the Google KG API to load strong and relevant knowledge graphs and subgraphs and the rest 50% is sent into YAGO and entities from it are obtained. The reason for choosing two distinct models is to improve the diversity. Convergence computation is achieved between the entities that come out of the Google KG API and YAGO using Tversky's Index that is set to a threshold of 0.60 and knowing its strength and this is done using Imperialist Competitive Algorithm where Tversky's Index is set as an objective of criteria function with the same step deviants measure. ICA is a multivariable algorithm which optimizes and yells irrelevant optimal entities which come out of the convergence from Google KG API and YAGO. ICA is a meta imperialistic algorithm and henceforth it is encompassed. Subsequently the top 5% of the top 50% of the classified instances which comes of the Bi-LSTM classifier i.e. classified metadata is subjected to feature extraction using Pearson's Correlation Coefficient with a step deviant of 0.10 as there are more number of penalties and most features have to be selected for this and that is the reason we're using Pearson's Correlation Coefficient and these features are included in AdaBoost classifier to classify the dataset. AdaBoost classifier is preferred because it uses machine learning, which is feature controlled and lightweight compared to a deep learning classifier, and it preserves domain deviance. Henceforth a machine learning classifier which is lightweight is preferred as it also helps in preserving the computational cycle. The entities that come out of the AdaBoost classifier and the optimal relevant entities computed through the ICA is further subjected to computation Tversky's Index (set to a threshold of 0.60) and Morisita's Overlap Index (adjusted to a 0.60 step deviation). As already limited optimal entities are computed so henceforth the thresholds are not made very stringent. The stringency is maintained. This is the end of Phase 1.

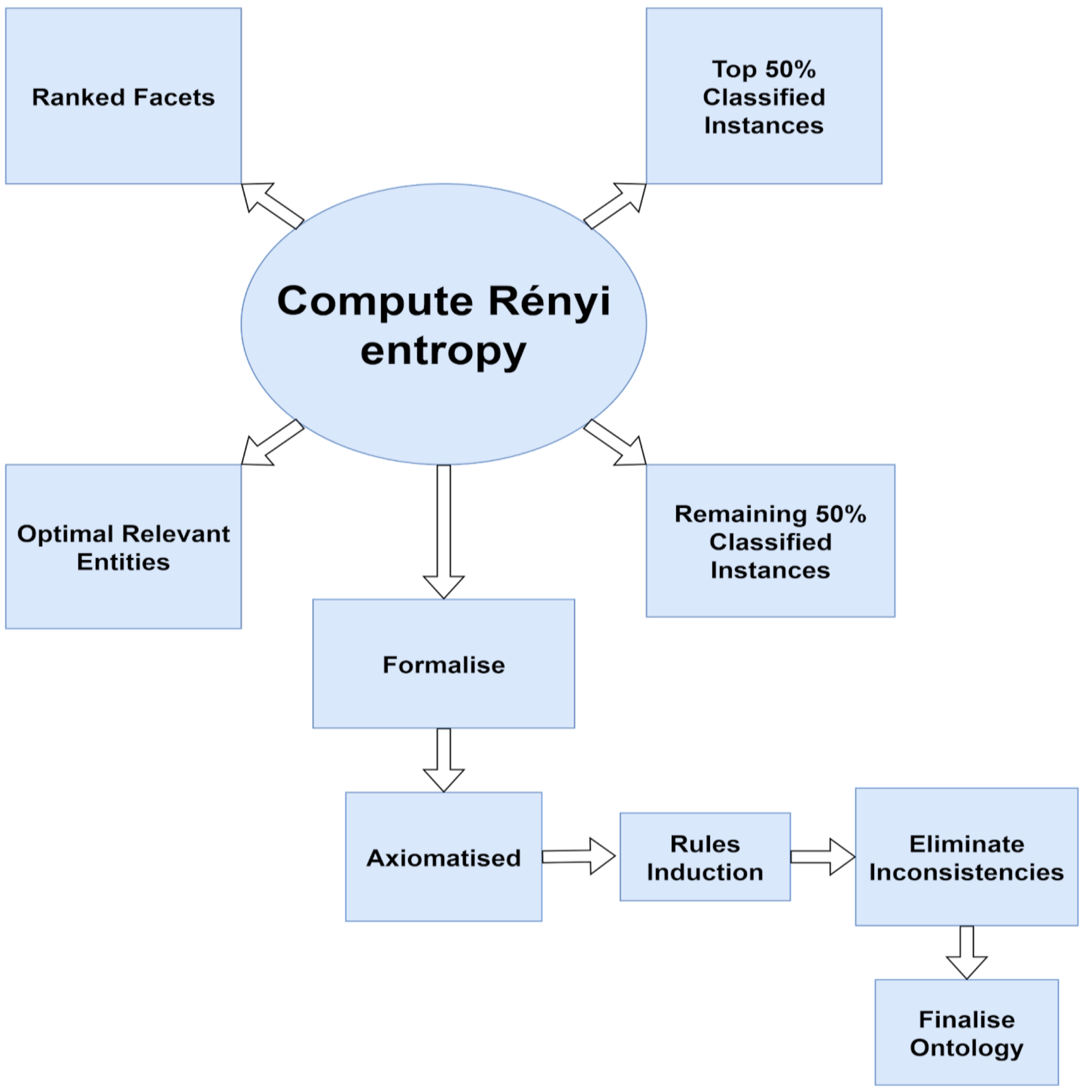


Figure 2: Phase 2 of the proposed architecture

Figure 2 depicts Phase 2, the ranked facets, 50% of the classified instances, and the remaining 50% of the classified instances that exit the Bi-LSTM classifier i.e., classified metadata and the relevant entities are all made use of from phase 1. All this is a subjective computation of Rényi entropy with a step deviance of 0.15. It is made stringent because many entities must be harnessed as the most optimal relevant entities are already computed. This is formalized by creating an edge based on the Rényi entropies information measure equivalents and is axiomatized by injecting four rules - IsaPartOf, HasaPartOf, IsaSuperClassOf and IsaSubClassOf. It is further subjected to rule induction based on auxiliary relationships and hierarchies and finally, the resulting ontology is subjected to elimination of inconsistencies to reasoning using the HermiT reasoner. And so, when all the inconsistencies are eliminated of the HermiT reasoner, it is subjected to finalization as an Ontology and sent for review to domain experts which is further finalized after review.

3.1 Tversky Index

A mathematical metric used to compare or contrast two sets of things is called the Tversky index, also known as the Tversky coefficient or Tversky similarity coefficient. In human cognitive psychology, notably in investigations of judgements of human similarity, it was first proposed by psychologists Amos Tversky and Daniel Kahneman. However, it has been used in many areas, such as bioinformatics, data mining, and information retrieval. When comparing two sets in terms of how many components they share and how many are unique to each set, the Tversky index is extremely helpful. The following formula defines it:

$$S(X, Y) = \frac{X \cap Y}{X \cap Y + \alpha\,(X \mid Y) + \beta\,(Y \mid X)} \; \ldots (1)$$

Formula 1 depicts the Tversky Index, a mathematical metric used to assess how similar or dissimilar two sets are and includes several essential elements. First, $X \cap Y$ denotes the total number of elements that belong to both sets X and Y. The overlap or commonality between the sets, or the number of items they share, is measured in this component of the calculation. Next, $X \mid Y$ denotes the quantity of elements in set X but absent from set Y. This section quantifies set X's exclusivity by listing the pieces that make up the set. The symbol $Y \mid X$ denotes the distinct items unique to set Y by denoting the number of elements found in set Y but not in set X. The formula additionally has two non-negative parameters, α and β, which are used to regulate the relative weights given to shared and unique components in the similarity computation. With the use of these settings, you may change how much focus is put on the similar elements compared to the different elements to suit your own requirements. You can fine-tune the Tversky index to give one aspect priority over another by changing the values of and. In conclusion, the Tversky index provides a flexible method for determining the degree of similarity or dissimilarity between sets by considering both shared and unique features, with the flexibility to adjust the weighting of each aspect using the parameters and. Due to its adaptability, it is a useful tool for comparing and analyzing data in many different sectors.

3.2 Pearson’s Correlation Coefficient

The Pearson's Correlation Coefficient is defined as the covariance of two variables divided by the sum of their standard deviations, or "r," displayed in formula 2. It evaluates a linear relationship between two continuous variables in its direction and importance. The definition is given as a "product moment," which is the statistical mean of the first moment of the origin multiplied by the random variables adjusted for mean. This is where the word "product-moment" originates.

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}} \ldots (2)$$

Perfect positive linear relationships are represented by a value of 1, perfect negative linear relationships are represented by a value of -1, and no linear relationships are represented by a value of 0. on a scale from -1 to 1. A robust measure, "r" considers the covariance between the variables and their individual variances in its formula. It is crucial in disciplines like statistics, the social sciences, and data analysis since it presupposes linearity, normal distribution, and sensitivity to outliers. It is used by researchers in data analysis for feature selection and understanding variable interactions, and to quantify linkages like height and weight correlations.

3.3 Morisita's Overlap Index

Formula 3 depicts the Morisita's Overlap Index, which is a statistical metric used to express how much two ecological communities or collections of objects overlap or are similar. Within these communities, it considers both the quantity and distribution of the various species or components. Higher values of the index which goes from 0 to 1. It specifically considers the product of the total abundances in each community divided by the sum of the products of the abundances of shared species or objects. With 0 indicating no overlap and 1 showing total overlap, Morisita's overlap index provides a mechanism to evaluate how similar or dissimilar two communities are in terms of their species composition or element distribution.

$$C_H = \frac{2 \sum_{i=1}^{S} x_i \, y_i}{\left(\frac{\sum_{i=1}^{S} x_i^2}{X^2} + \frac{\sum_{i=1}^{S} y_i^2}{Y^2}\right) X \, Y} \ldots (3)$$

3.4 Bi-LSTM Classifier

An adaptation of the neural network architecture known as Long Short-Term Memory (LSTM), which is also popularly employed for sequential data analysis and tasks involving natural language processing, is the Bidirectional Long Short-Term Memory (Bi-LSTM) classifier. The primary innovation of a Bi-LSTM is its ability to store contextual information from both the past and the future. This makes it particularly useful for comprehending sequences when bi-directional relationships are significant. In a Bi-LSTM classifier, the network is made up of two parallel LSTM layers, one of which goes through the input sequence forward (from start to finish), and the other of which goes backward (from end to start). As opposed to conventional LSTMs, which solely consider previous data, this bidirectional processing allows the model to take information from the entire sequence into account. To produce a thorough representation of each time step, the outputs from these two LSTM layers are often merged in some fashion, such as concatenation or summation. Bi-LSTM classifiers' capacity to recognize distant relationships in sequential data is one of its key features. This is especially useful when doing tasks like natural language understanding because words' meanings frequently depend on the words that come before and after them. Numerous NLP applications, such as sentiment analysis, part-of-speech tagging, and named entity recognition, benefit from the use of bi-LSTMs. To train a Bi-LSTM classifier, labeled data must be provided to the network. Through gradient descent and backpropagation, the network then learns to extract pertinent features and associations from the input sequences. By learning both past and future context, the model is better able to generate accurate predictions or classifications. However, when working with tiny datasets, Bi-LSTMs can be computationally taxing and potentially prone to overfitting. Due to its parallelization and attention methods, more complex architectures like Transformers have occasionally acquired prominence in NLP applications, but Bi-LSTMs continue to be an excellent option when working with sequential input that has bidirectional relationships.

3.5 Imperialist Competitive Algorithm

The Imperialist Competitive Algorithm (ICA) is a nature-inspired optimization algorithm that takes cues from social and political phenomena, especially the struggle for dominance and access to resources across various nations and regions. Esmat Rashedi, Hossein Nezamabadi-pour, and Saeid Saryazdi introduced it in 2009. ICA is a type of computational approach used to tackle several types of optimization problems in computer science. ICA does not require the function's gradient during its optimization process, unlike most methods in computational evolution. In a certain sense, ICA can be compared to genetic algorithms (GAs) as the social analog. While GAs is based on a species' biological evolution, a computer simulation and mathematical model of human social evolution. The population of potential solutions that ICA works with is often called colonies or countries. Each colony stands for a potential answer to the optimization issue. Each colony's fitness, which measures how well it performs in relation to the goal function, is first assessed by the algorithm. Higher fitness colonies are given control over weaker colonies since they are seen as imperialistic. The imperialistic colonies incorporate the weaker ones throughout the colonization phase. This procedure entails exchanging knowledge, integrating approaches, or using other tools to raise the standard of the colonies. By allowing the stronger colonies to affect the weaker ones, the goal is to improve the population's performance. ICA includes a revolution phase to maintain diversity and investigate various areas of the solution space. The opportunity for weaker colonies to overthrow their imperialist overlords and obtain independence exists. This makes sure the algorithm keeps exploring new options and avoids getting caught in local optima. Numerous optimization issues, such as function optimization, scheduling, and machine learning parameter tweaking, have been addressed by ICA. Researchers continue to investigate its potential and create versions to efficiently address a variety of complicated optimization challenges. Its unique simulation of political and social dynamics offers a new perspective on optimization. But like many optimization methods, ICA's effectiveness might vary depending on parameter choices and problem-specific factors, necessitating careful adjustment for the best outcomes under various conditions.

Algorithm: Imperialist Competitive Algorithm

Inputs:
- Objective function f(x)
- Number of variables n
- Number of imperialists N_imp
- Number of colonies N_col
- Maximum number of iterations max_iter
- Lower bound lb
- Upper bound ub

Output:
- Best solution x*

Step 1: Initialization:
- Generate N_imp random imperialist solutions x_imp[i], i = 1 to N_imp
- Generate N_col random colony solutions x_col[j], j = 1 to N_col

Step 2: Repeat for max_iter iterations:
a. Competition:

- Evaluate the fitness f(x_imp[i]) for each imperialist solution x_imp[i]
- Evaluate the fitness f(x_col[j]) for each colony solution x_col[j]
- Sort imperialists and colonies based on fitness values in ascending order

b. Replace:
- For each colony j, if f(x_col[j]) < f(x_imp[j]):
  - Replace the corresponding imperialist with the colony: x_imp[j] = x_col[j]

c. Assimilation:
- Merge remaining colonies into the imperialist population:
  - Sort imperialists based on fitness values in ascending order
  - Replace the least fit imperialists with remaining colonies

Step 3: Return x_imp[1] as the best solution x*

Algorithm 1: Imperialist Competitive Algorithm

Algorithm 1 explains the order to identify the best answers to optimization issues, the Imperialist Competitive Algorithm (ICA) simulates the dynamics of imperialist nations and their colonies. First, a population of practical solutions is produced at random and represented as colonies. A fitness value that represents each colony's quality in relation to the optimization problem is assigned to it. After that, these colonies are arranged into empires, with one serving as the "imperialist" colony and the others serving as "colonies." Imperialists strive for supremacy by absorbing colonies with lower fitness values, which raises their own fitness, during each algorithmic iteration. Transferring colonies from weaker to stronger empires is a part of this assimilation process, which reflects the sociopolitical idea of imperialism. Furthermore, if colonies reach a higher fitness level, they can rebel against their imperialists, which could result in the relocation of colonies throughout empires. Until a termination condition, such as the achievement of a satisfactory solution or a maximum number of iterations is reached, this iterative process of competition, assimilation, and potential revolution continues. The method efficiently explores the search space and converges towards optimal or nearly optimal solutions for a wide variety of optimization problems through this dynamic competition and assimilation mechanism.

3.6 Ontology Modeling

An ontology is a formal knowledge representation in computer science that describes ideas, attributes, connections, and axioms within a certain domain. It acts as a well-structured framework for knowledge organization and comprehension, promoting both human and automated comprehension. Artificial intelligence, the semantic web, natural language processing, data integration, and several more applications all require ontologies. In addition to improving data interoperability, reasoning, and decision-making in a variety of computational contexts, such as the semantic web, expert systems, data integration, and health informatics, they encompass concepts, properties, and relationships and enable the representation of domain-specific knowledge. OWL, RDF, and Protégé are examples of common ontology languages and tools. The process of developing a structured and formal representation of knowledge within a particular domain or subject area is known as ontology modeling. It entails defining the ideas, traits, connections, and logical principles that define that field. Typically, the initial step in the modeling process is to determine the essential ideas which are important to the domain and then to characterize the relationships between those concepts. These ideas' properties and attributes are specified to include extra information or features, and relationships are established between concepts to show how they interact or are arranged hierarchically. Making axioms, or rules, to guide the ontology's behavior and constraints, is another aspect of ontology modeling. Information retrieval, reasoning, and data integration are made easier in a variety of computer science applications, from artificial intelligence to the semantic web and beyond, thanks to this modeling approach's ability to capture, organize, and share complex knowledge in a structured and machine-readable format. To solve complicated problems and increase the effectiveness of computational systems, it enables a greater grasp of the topic. Ontologies are conducted on a single large dataset in five integrated datasets namely Jamie Matthews [30] proposed environmental communication [Dataset], Singh et al. [31] stated environmental issues in MPAND approach of media [Dataset], Paige B Jarreau et al. [32] best practices in environmental communication [Dataset], GlobalData Plc CECO Environmental CorpMedia [33] proposed Ads & Top Trends [Dataset] and Secretariat of the Pacific Regional Environment Programme [34] suggested media resources for the Pacific Islands' State of Environment and Conservation [Dataset]. These datasets were strategically integrated by using a customized algorithm to annotate all datasets. Once these datasets were annotated, these were reprioritized. On annotation and reprioritization, a customized web crawler is utilized to extract documents from the structural data from the web. Much more documents comprising the categories and annotations generated along with keywords of these documents where we used to crawl for the documents, document space was maintained and integrating the documents on the web and rearranged into categories, one huge dataset is formed onto which we extract categories.

4. Implementation

Jamie Matthews [30] has stated that Environmental communication is a field within the communication discipline, advertently cutting across various other disciplines such as cultural theory, media theory and social movement theory and has stated that his overall experience of the environment and the actual field of environmental communication is embedded with symbolism and metaphors to describe the natural world. It is in this context that nature is surrounding them from outside of their minds and experiences. Yet through language and metaphors, we feel closer to the natural world when we deploy terms like Mother Nature. So, it is language and symbolic action that we use to create our perceptions and values of the environment. Jai dev Singh et al. [31] have discussed the primary reasons for the current environmental problems and the main obstacles that the media and media professionals encounter when covering environmental topics in the state of Madhya Pradesh, India, and the possible immediate cause correction. Paige B Jarreau el al. [32] have proposed the best practices in environmental communication. GlobalData Plc CECO Environmental CorpMedia [33] have stated the top trends in environment journalism. Secretariat of the Pacific Regional Environment Programme [34] have stated the sources of media for the State of Environment and Conservation in the Pacific Islands: 2020 Regional Report.

Algorithm: Environmental Policy Datasets to Environmental Journalism Ontology
Inputs: D = Environmental Policy Dataset **Outputs:** O = Environmental Journalism Ontology **begin** **1. Preprocessing:**

D: Environmental policy dataset (set of documents)
K(D): Function mapping D to unique keyword set
C(D): Function mapping D to unique category set
Seed Term Pool (S):

$S = K(D) \cup C(D)$

2. Metadata Generation:
OC = OpenCalais() employed to generate metadata (M) from the seed term pool (S)

$M = OC(S)$

3. Metadata Classification:
B = Bi-directional Long Short-Term Memory (Bi-LSTM) classifier
M_c = classified metadata
M = metadata

$M_c = B(M)$

4. Knowledge Graph Integration:
M_c = classified metadata (divided into two parts)
$M_{c(top)}$ = top 50% of the classified metadata
$M_{c(bottom)}$ = bottom 50% of the classified metadata
$KG_{\{API\}}$ = Google Knowledge Graph API
YK = YAGO knowledge Graph
$KG_{\{G,S\}}$ = extracted Graphs and Subgraphs form $KG_{\{API\}}$
E_y= extracted entities from YAGO

$KG_{\{G,S\}} = KG_{\{API\}}(Mc_{(top)})$

$E_y = YK(M_{c(bottom)})$

5. Entity Convergence:
T = Tversky's Index
τ = predefined threshold
ICA = Imperialist Competitive Algorithm
E_{opt} = set of optimal relevant entities
employed to measure the similarity between entities retrieved from Google KG API (E_G) and YAGO (E_Y)

$Sim(E_G, E_Y) = T(E_G, E_Y) > \tau$
$E_{opt} = ICA(E_G, E_Y)$

6. Feature Extraction and Classification:
The top 5% features (F_{top}) are extracted from the top 50% classified metadata ($M_{c(top)}$)) using Pearson's Correlation Coefficient (PCC) with a step deviation (σ). An AdaBoost classifier (A) is then trained on these features to further refine the classification of entities and stored in E_A.

$Ftop = PCC(M_{c(top)}, \sigma)$
$E_A = A(F_{top})$

7. Entity Refinement:

Entities from AdaBoost (E_A) are merged with optimal entities (E_{opt}). Tversky's Index (threshold: T , step deviation: σ_T) and MOI (Morisita's Overlap Index) (threshold: M, step deviation: σ_M) are used to refine the combined set (E).

$E = E_A \cup E_{opt}$
$Sim(E_i, E_j) = T(E_i, E_j) > T \cap MOI(E_i, E_j) >: M (\forall E_i, E_j \in E)$

8. Ranked Facet Generation:
Rényi entropy (H_R) with a step deviation (σ_R) is computed on three elements:
a) Ranked facets (F_R)
b) Remaining 50% classified metadata from Bi-LSTM ($M_{c(bottom)}$))
c) Relevant entities (E)
This step aims to incorporate many entities into the ontology construction.

$H_R(F_R, M_{c(bottom)}, E) > \sigma_R$

9. Ontology Construction
Based on the Rényi entropy (H_R) information measures, edges are created in the ontology (O) to represent relationships between entities and concepts. Four foundational axioms are injected into the ontology:
a) IsaPartOf (is-a-part-of)
b) HasaPartOf (has-a-part-of)
c) IsaSuperClassOf (is-a-superclass-of)
d) IsaSubClassOf (is-a-subclass-

end

Algorithm 2: Ontology Generation

Algorithm 2 explains the way to create a seed term pool. This algorithm first preprocesses the Environmental Policy Dataset (D) by removing distinct keywords and categories. This pool is used to generate metadata using OpenCalais. The metadata is subsequently categorized by a Bi-directional Long Short-Term Memory (Bi-LSTM) classifier. Knowledge graph integration is then carried out by selecting pertinent entities and comparing them between the YAGO knowledge graph and the Google Knowledge Graph API. The Imperialist Competitive Algorithm and Tversky's Index converge on ideal entities. Pearson's Correlation Coefficient is used to extract features, and an AdaBoost classifier is used to fine-tune entity classification. Next, using similarity measures, the entity collection is further refined. In addition to ranking aspects and residually categorized metadata, entities are included in the ontology development process by Rényi entropy computation. Ultimately, using foundational axioms, an ontology based on Rényi entropy information measures is built, facilitating the establishment of links between entities and concepts in environmental policy. By building an Environmental Journalism Ontology (O) from the provided dataset, this all-inclusive procedure hopes to improve environmental journalism efforts.

5. Results and Performance Evaluation

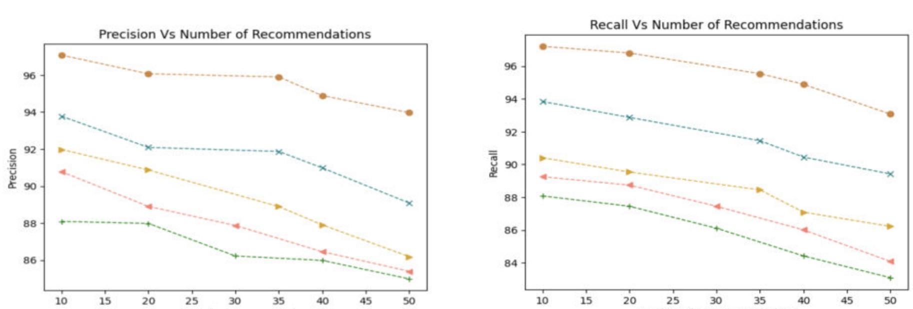
The performance of the proposed OSSS framework, a semantics-oriented framework for knowledge-centric ontology synthesis for environmental journalism, is a semi-automatic approach. This is evaluated using Precision, Recall, Accuracy, F-measure percentages and False Discovery Rate, FDR (as potential metrics). The reason for choosing precision, recall, accuracy, F-measure percentages as potential metrics since it measures the sets' relevance. False positive rate inherently depicts the error rate in the model as it measures the number of false positives and therefore, they are chosen as potential metrics.

Table 1: Performance Comparison of the proposed OSSS

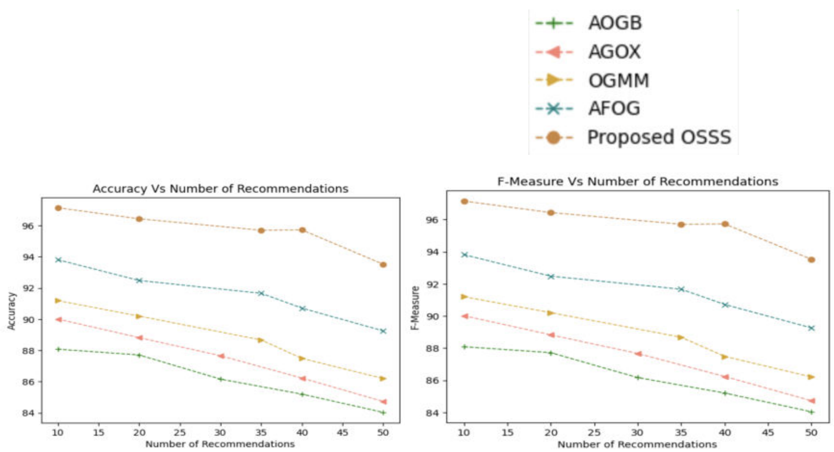
Model	**Average Precision %**	**Average Recall %**	**Average Accuracy %** (P+R)/2	**Average F-Measure %** 2*P*R/(P+R)	**FDR**
AOGB[1]	86.62	87.09	86.85	86.85	0.14
AGOX [2]	87.08	88.12	87.60	87.60	0.13
OGMM [3]	88.35	90.09	89.22	89.21	0.12
AFOG [4]	91.08	93.45	92.27	92.24	0.08
Proposed OSSS	95.15	96.08	95.62	95.60	0.05

From *Table 1*, it is indicated that the proposed OSSS has the highest precision accuracy, F-measure percentages of 95.15, 96.08 respectively. The lowest value of FDR of 0.05. To evaluate the performance of the OSSS is based with four distinct models namely the AOGB, AGOX, OGMM and AFOG respectively. The AOGB has yielded 86.62% average precision, 87.09 % average recall, 86.85 % average accuracy, 86.85 % F-measure, FDR of 0.14. The AGOX has yielded 87.08% average precision, 88.12 % average recall, 87.60 % average accuracy, 87.60 % accuracy, 87.60 % F-measure with an FDR of 0.13. OGMM yielded 88.35 % average precision, 90.09 % average recall, 89.22 % average accuracy, 89.21 % F-measure, FDR of 0.12. The AFOG framework has yielded 91.08% average precision, 93.45 % average recall, 92.27 % average accuracy, 92.24 % accuracy F-measure with an FDR 0.08. The fact that the proposed OSSS has yielded the highest precision, accuracy, F-measure percentages, and the lowest value of FDR is because of the reason that metadata generation is facilitated from the keywords and categories harvested from the data set and metadata raises the auxiliary knowledge's density. The metadata generated is classified using the Bi-LSTM classifier, that is a strong classification model which classifies the metadata under the presence of AdaBoost and the Bi-LSTM is a very strong learning infrastructure, one being a machine learning classifier, which is strong, lightweight, which is feature controlled, namely the AdaBoost, preserves domain deviants and Bi-LSTM classifies the metadata much accurately increasing the overall intermediate accuracy rate. When it comes to semantic relatedness computation and reasoning using semantic similarity measures, the Pearson's Correlation Coefficient with the Tverskey's Index and Morisita's Overlap Index with differential thresholds and an empirically determined step deviance measure ensures a strong relevance computation mechanism. The Yago knowledge store further adds to the amount of auxiliary knowledge into the framework. The Pearsons' Correlation Coefficient helps in harvesting the features, thereby improving the overall classification accuracy of the AdaBoost model as well. The Imperialist Competitive Algorithm is an intermediate metaheuristic-based optimization model which optimizes the intermediate solution set to yield the most optimal solution set and Google KG also adds to the amount of auxiliary knowledge, thereby system rating convergence to a better rate of optimality in the model. Thus, the suggested framework performs better than any other baseline models. and shows that it is the best-in-class model for Ontology Synthesis. The reason the baseline models, namely the AOGB performs differently than anticipated in comparison to the suggested framework, it is because although it is an automatic ontology generation using BIM and GIS data, this framework uses a formal method of ontology generation and is synthesized by using XSD documents and transformation patterns. It is only an XSD schema with transformation, which is a form of XML Schema, is used to transform it directly into the ontology, which is a transfiguration from one data format to another data format alone. There is no harvesting, no synthesis of auxiliary knowledge, there's only static knowledge which is transformed into an ontology. There is no new knowledge synthesis, and there are no strong classification models, and there are no learning frameworks.

There are no semantic reasoning models, and henceforth the AOGB model lags compared to the proposed framework. The explanation for why the alternative framework i.e. The AGOX framework does not function as intended when compared to the framework because, although the AGOX is an automatic ontology generation from XML source, the XML is transformed to web ontology language where static ontology along with domain knowledge is represented as an ontology. There are no auxiliary knowledge addition schemes, no knowledge schemes in the framework and no presence of learning models. There is no possibility of harvesting auxiliary knowledge into the model and henceforth the AGOX model does not perform as expected, due to the absence of learning infrastructure into the model, and due to the absence of the semantic computation mechanism, the AGOX model also lags drastically than the proposed framework. The reason the OGMM model also does not perform as expected, although it is OGMM is an ontology-based generation of medical multi terms. Here, to implement this, the concepts for generation of multiterm MCQs was used to generate the ontology for the exact same dataset, which was incorporated in the proposed framework, which involved generating case-based questions providing explanation for incorrectness. And furthermore, implementation of procedure using medical oncology, medical ontology was synthesized from the natural text and a medical knowledge base which is synthesized but the exact same technique was used. But however, this framework also did not perform as expected even though it was a tweak for ontology modeling and generation. There was too much manly intervention in the OGMM, auxiliary knowledge was not fed into the framework but however knowledge was fed through questioning questions but reasoning for the auxiliary knowledge and transforming to an atomic permeable state was not done and it lags a strong running infrastructure. Henceforth the OGMM model also lags than the proposed framework. Finally, the AFOG framework also lags in contrast to the suggested framework because although it is a creation of fuzzy ontologies, using fuzzy concepts, the fuzzy ontology is produced through expansion of rudimentary ontologies. The fuzzy idea originated in a particular formal context, so fuzzy is not accurate because its rules are approximate because there is some uncertainty in fuzzy sets. Exactly because of this, the AFOG model also lagged and most importantly no learning infrastructure was quite weak or absent and there was no auxiliary knowledge or any semantic reasoning mechanism in the AFOG model. And hence it lagged compared to the proposed ontology. Since the proposed framework is a semantics oriented framework which functions on the perspective of incremental knowledge derivation and generation, specifically the metadata generation from the structural World Wide Web and since it has a very strong learning infrastructure in the form of a highly efficient deep learning classifier, the Bi-LSTM, and a powerful classifier which uses machine learning, the AdaBoost, which targets the metadata and the data set respectively. And, since it has a strong relevance computation mechanism through semantics-oriented reasoning using Tversky's Index, Morisita's Overlap Index, Pearson's Correlation Coefficient, and a strong metaheuristic algorithm for optimization, namely the Imperialist Competitive Algorithm. The suggested framework works better than all other baseline models and serves as the top of the class model for ontology synthesis and generation.



(a) **(b)**



(c) **(d)**

Figure 3: **Performance Comparison (a) Precision (b) Recall (c) Accuracy (d) F-Measure**

The distribution curve is depicted as *figure 3* wherein the suggested OSSS holds the top position in the hierarchy, bottom position within the hierarchy is occupied by AOGB, and the last position in hierarchy is occupied by AGOX. AFOG holds the second from the top in the hierarchy. OGMM holds the central position within the hierarchy. The best precision, accuracy, F-measure percentages, and lowest false discovery rate (FDR) of the proposed OSSS model demonstrate its superior performance, which can be attributable to several important aspects. The effective production of metadata, which is accomplished by removing keywords and categories from the dataset, is one essential component. The depth of the auxiliary knowledge within the model is increased by this information injection. Furthermore, a powerful Bi-LSTM classifier is utilized for metadata classification. This classifier, which makes use of AdaBoost, a feature-controlled classifier that uses machine learning, is a potent tool for categorizing metadata. The performance of the model is improved using AdaBoost by maintaining domain-specific distinctions. Bi-LSTM and AdaBoost work together to produce accurate metadata categorization, which raises the model's total accuracy. To calculate semantic relatedness and aid in reasoning, the model makes use of innovative metrics like Pearson's Correlation Coefficient in conjunction with Tversky's Index, Morisita's Overlap Index, and differential thresholds. These semantic similarity measurements aid in the efficient computation of relevance. The model gains additional auxiliary knowledge thanks to the inclusion of the Yago knowledge repository. The AdaBoost model's classification accuracy is improved through feature extraction, which benefits from the Pearsons' Correlation Coefficient. The Imperialist Competitive Algorithm, an intermediate metaheuristic-based optimization algorithm, is utilized for enhancement of the model's performance and execution. The most ideal solution set is produced by this procedure, which refines the intermediate solution set. Finally, the incorporation of Google Knowledge Graph (KG) increases the model's ability to converge towards highest optimality by introducing extra auxiliary knowledge. Baseline models' inferior performance, especially that of the AOGB, as compared to the suggested framework can be due to significant variations in their ontology generating strategies. The proposed framework uses a formal technique utilizing XSD documents and transformation patterns as opposed to the AOGB, which depends on automatic ontology development from BIM and GIS data. XSD schema with transformations, a method that directly translates data formats into ontology, is what the AOGB principally makes use of. It is incapable of dynamic categorization models and learning frameworks, knowledge harvesting, the synthesis of auxiliary information, or any of these. The suggested framework, in contrast, integrates these components and differs from the AOGB by offering robust classification models, learning capabilities, and semantic reasoning. Due to its reliance on automatic ontology synthesis from XML sources, the AGOX framework is similarly inadequate when compared to the proposed framework. With the help of AGOX, static ontology and domain-specific knowledge are represented alongside each other in a web ontology language. AGOX, however, lacks integration methods for complementary knowledge, knowledge structures, or learning models. AGOX's performance is hampered by the lack of a knowledge harvesting procedure and a semantic computation mechanism, which causes a considerable performance gap between it and the suggested framework. In a similar vein, the ontology based OGMM paradigm, which focuses on creating medical multi-term ontologies, has performance challenges. From the same dataset as the suggested framework, it creates ontology concepts for multiple-choice questions (MCQs). The OGMM model, however, heavily relies on manual intervention and does not adequately account for auxiliary knowledge. It relies on question-based knowledge input and omits the crucial steps of inferring auxiliary knowledge and transforming it into a flexible, atomic state. This restriction leads to a weak infrastructure, which lowers the effectiveness of the OGMM model in contrast to the suggested framework. The AFOG framework's fuzzy ontology generation method, which is based on fuzzy ideas, also makes it inferior to the proposed framework. Fuzzy ontologies build upon basic ontologies and rely on fuzzy formal contexts, which by virtue of the approximate nature of fuzzy rules, involve inherent ambiguity. The AFOG model performs poorly when compared to the suggested ontology because it lacks a strong learning infrastructure, auxiliary knowledge integration, and a semantic reasoning mechanism This is one of the first frameworks for ontology generation and synthesis for environmental journalism as a domain of choice, which generates a strong formal ontology for strong environmental journalism.

6. Conclusions

This paper puts forth a strategic framework for generating ontology using semantic intelligence approaches and federated learning paradigms. Federated cooperative learning paradigms for Web 3.0 focusing on environmental studies as the prospective domain of choice. The framework encompasses metadata generation from seed term pool generation by obtaining the entities directly from the data set itself. A strong running infrastructure is provided through the AdaBoost classifier and the Bi-LSTM classifier for the dataset classification and the metadata classification, respectively. The classified cases that come out of Bi-LSTM classifiers are sent as federated features to the AdaBoost classifier through Perason's Correlation Coefficient as a feature selector mechanism. The Tverski's Index along with the Morisita's Overlap Index serve as strong indexes for relevance computation and providing semantic similarity-based reasoning. Tversky's Index under competitive Imperialist Algorithm also helps in transforming the initial or intermediate solution state to a much more optimal solution state. Google search graph API Yago further contributes to the intensification of the density of the supporting knowledge which has been included into the structure, thereby providing fact-based verification of knowledge and an overall precision, recall, accuracy, F-measure which generates ontologies automatically focusing on environmental journalism as a prospective domain of choice.

7. References

1. Usmani, A. U., Jadidi, M., and Sohn, G.: AUTOMATIC ONTOLOGY GENERATION OF BIM AND GIS DATA, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B4-2020, 77–80, https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-77-2020, 2020.

2. Nora Yahia, Sahar A. Mokhtar, AbdelWahab Ahmed, Cornell University

3. Davies, J., Studer, R., & Warren, P. (Eds.). (2006). *Semantic Web technologies: trends and research in ontology-based systems*. John Wiley & Sons.

4. Wei Chen˙Qing Yang, Li Zhu, Bin Wen Department of Computer Science˙HuaZhong Normal University˙Wuhan , China.

5. Ramaprasad, A., & Syn, T. (2013, August). Ontological meta-analysis and synthesis. In *Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois*.

6. Grubic, T., & Fan, I. S. (2010). Supply chain ontology: Review, analysis and synthesis. *Computers in Industry*, *61*(8), 776-786.Hohenecker, P., & Lukasiewicz, T. (2017). Deep learning for ontology reasoning. *arXiv preprint arXiv:1705.10342*.

7. Lytvyn, V., Vysotska, V., Wojcik, W., & Dosyn, D. (2017). A method of construction of automated basic ontology. *Computational linguistics andintelligent systems (COLINS 2017)*.

8. Bourguin, G., Lewandowski, A., Bouneffa, M., & Ahmad, A. (2021). Towards ontologically explainable classifiers. In *Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30* (pp. 472-484). Springer International Publishing.

9. Kulmanov, M., Smaili, F. Z., Gao, X., & Hoehndorf, R. (2020). Machine learning with biomedical ontologies. *biorxiv*, 2020-05.

10. Sulaeman, M. M., & Harsono, M. (2021). Supply chain ontology: model overview and synthesis. *Jurnal Mantik*, *5*(2), 790-799.

11. Smith, B., & Welty, C. A. (2001, October). FOIS introduction: Ontology-towards a new synthesis. In *FOIS* (Vol. 1, pp. 3-9).

12. Mika, P., & Akkermans, H. (2004). Towards a new synthesis of ontology technology and knowledge management. *The Knowledge Engineering Review*, *19*(4), 317-345.

13. Janowicz, K., Van Harmelen, F., Hendler, J. A., & Hitzler, P. (2015). Why the data train needs semantic rails. *AI Magazine*, *36*(1), 5-14.

14. Al-Aswadi, F. N., Chan, H. Y., & Gan, K. H. (2020). Automatic ontology construction from text: a review from shallow to deep learning trend. *Artificial Intelligence Review*, *53*, 3901-3928.

15. Croce, V., Manuel, A., Caroti, G., Piemonte, A., De Luca, L., & Véron, P. (2023). Semi-automatic classification of digital heritage on the Aïoli open source 2D/3D annotation platform via machine learning and deep learning. *Journal of Cultural Heritage*, *62*, 187-197.

16. Hassanzadeh, H., & Keyvanpour, M. (2011). A machine learning based analytical framework for semantic annotation requirements. *arXiv preprint arXiv:1104.4950*.

17. Gacitua, R., Sawyer, P., Piao, S., & Rayson, P. (2007). Ontology acquisition process: A framework for experimenting with different nlp techniques. In *Proceedings of the UK e-Science All Hands Conference 2007* (pp. 561-567).

18. Konys, A. (2019). Knowledge repository of ontology learning tools from text. *Procedia Computer Science*, *159*, 1614-1628.

19. Van Rooijen, L., Bäumer, F. S., Platenius, M. C., Geierhos, M., Hamann, H., & Engels, G. (2017, September). From user demand to software service: using machine learning to automate the requirements specification process. In *2017 IEEE 25Th international requirements engineering conference workshops (REW)* (pp. 379-385). IEEE.

20. Adelkhah, R., Shamsfard, M., & Naderian, N. (2019, April). The ontology of natural language processing. In *2019 5th International Conference on Web Research (ICWR)* (pp. 128-133). IEEE.

21. Malgieri, L. E. (2023). Ontologies, Machine Learning and Deep Learning in Obstetrics. In *Practical Guide to Simulation in Delivery Room Emergencies* (pp. 29-64). Cham: Springer International Publishing.

22. Iannone, L., Palmisano, I., & Fanizzi, N. (2007). An algorithm based on counterfactuals for concept learning in the semantic web. *Applied Intelligence*, *26*(2), 139-159.

23. Konys, A. (2018). Knowledge systematization for ontology learning methods. *Procedia computer science*, *126*, 2194-2207.

24. Tosi, D., & Morasca, S. (2015). Supporting the semi-automatic semantic annotation of web services: A systematic literature review. *Information and Software Technology*, *61*, 16-32.

25. Broussard, Meredith, et al. "Artificial intelligence and journalism." *Journalism & mass communication quarterly* 96.3 (2019): 673-695.

26. Mizoguchi, Riichiro, and Mitsuru Ikeda. "Towards ontology engineering." *Journal-Japanese Society for Artificial Intelligence* 13 (1998): 9-10.

27. Bhatti, N., & Fellner, D. W. (2012). Visual semantic analysis to support semi-automatic modeling of semantic service descriptions. In *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 986-1033). IGI Global.

28. Pavlick, Ellie. "Semantic structure in deep learning." *Annual Review of Linguistics* 8 (2022): 447-471.

29. Yoon, H. G., Han, Y. J., Park, S. B., & Park, S. Y. (2007, August). Ontology population from unstructured and semi-structured texts. In *Sixth international conference on advanced language processing and Web Information Technology (ALPIT 2007)* (pp. 135-139). IEEE.

30. Jamie Matthews (2016). Environmental Communication [Dataset].

31. Jai dev Singh; Ram Awtar Yadav; Shrikant Singh (2023). ENVIRONMENTAL ISSUES IN MPAND APPROACH OF MEDIA [Dataset].

32. Paige B Jarreau; Zeynep Altinay (2023). Best Practices in Environmental Communication - Appendices [Dataset].

33. GlobalData Plc (2013). CECO Environmental CorpMedia, Ads & Top Trends [Dataset].

34. Secretariat of the Pacific Regional Environment Programme (2022). Media resources for the State of Environment and Conservation in the Pacific Islands: 2020 Regional Report [Dataset].