A Hybrid Question Answering Model with Ontological Integration for Environmental Information*

Tianda Sun[1,*], Jamie Carr[1] and Dimitar Kazakov[1,*]

[1]University of York, Heslington, York YO10 5DD, United Kingdom

Abstract

This paper presents a novel approach to constructing a Question Answering model for analysing Nationally Determined Contributions (NDC) reports within the environmental sector. The approach is based on Large Language Models (LLMs) equipped with Retrieval Augmented Generation (RAG) and enhanced by ontology integration. Acknowledging the challenges inherent in directly applying RAG, our approach begins with the development of a specialised ontology framework for NDC reports. This framework supports the construction of a knowledge graph that provides essential, verifiable information for a Question Answering (QA) model. In the next step, the model combines RAG embeddings with ontology-based queries, aiming to enhance the reliability of answers across various NDC reports. We evaluate the performance of our hybrid model through testing with a set of questions and human/AI evaluation across different LLMs. While the results indicate improvements in the efficiency of climate change-related QA models, they also underscore the complexity of achieving significant enhancements in this domain. Our findings contribute to ongoing discussions about the potential and limitations of integrating ontological methods with LLM for environmental information retrieval.

Keywords

Relation Extraction, Knowledge Graph Construction, Retrieval Augmentation Generation,

1. Introduction

Ontologies and the knowledge graphs derived from them are valuable knowledge representation approaches in many application areas. Designing such ontologies is a laborious process requiring domain expertise and knowledge engineering skills. The process can be assisted by tapping into existing sources of knowledge containing a structured component. One example of this is using the labelled information shown in the little grey rectangle at the top of each Wikipedia page to produce a vast ontology known as DBpedia [1]. Domain-specific work in the biomedical, finance[2, 3] or clinical. Most textual information is available in a non-structured form though, and extracting concepts and relations is a valuable but challenging task [4].

Retrieval Augmented Generation (RAG) offers a transformative approach to enhancing Question Answering systems by dynamically retrieving and incorporating external knowledge during the generation process. In the context of ontologies, RAG can be particularly powerful. By integrating RAG with a robust ontology, the system can access a structured repository of domain-specific knowledge, such as those concerning environmental policies and practices outlined in Nationally Determined Contributions (NDC) reports. This integration allows the QA system to fetch relevant information and contextually adapt its responses based on ontological relationships and entities. The methodology leverages the precision of ontology-based data retrieval and the flexibility of generative models to produce more accurate and contextually relevant answers. Studies such as those by Li et al. (2024) have demonstrated the efficacy of RAG in various domains by showing significant improvements in the accuracy and relevance of generated answers[5]. Applying these principles, our model aims to tackle environmental data's complex terminology and inter-relation characteristics, thereby enhancing the decision-making process in climate change mitigation efforts.

DAO-XAI 2024: Workshop on Data meets Applied Ontologies in Explainable AI, October 19, 2024, Santiago de Compostela, Spain
*Corresponding author.
tianda.sun@york.ac.uk (T. Sun); jamie.carr@york.ac.uk (J. Carr); dimitar.kazakov@york.ac.uk (D. Kazakov)
https://www-users.york.ac.uk/~dlk2/ (D. Kazakov)
0000-0003-1998-5228 (T. Sun); 0000-0003-2219-5374 (J. Carr); 0000-0002-6476-0098 (D. Kazakov)

In this paper, we start with these ideas to investigate how LLM can construct domain-specific knowledge graphs with limited participation of domain experts. Then, we involve our constructed ontology into a retrieval augmentation generation (RAG) system for the question and answering(QA) system to answer a set of NDC-related questions and evaluate the answer. Our contributions are as follows:

- We implement a novel environmental knowledge graph framework manually and apply it to extract information from environmental reports on climate change. This approach aims to capture important information and organise it into an ontology.
- Based on the ontology we constructed, we developed a hybrid RAG system to retrieve information from both our ontology and NDC report. Our evaluation shows that our method performs better than the baseline method.

2. Background

2.1. Ontologies and Knowledge Graphs

Large knowledge graph such as Freebase [6] and DBpedia [1] have shown a remarkable ability to provide a well-structured data source in a wide range of NLP tasks, including question-answering [7], information retrieval [8], chatbots [9], recommendation systems [10] and machine translation [11]. With structured vocabulary and predefined relationships, knowledge graphs and ontologies facilitate automated associations between data and knowledge, enabling nuanced analyses and hypothesis generation in various scientific domains such as finance [3] and health care [12].

However, constructing domain-specific ontology is a multifaceted, challenging task requiring the deep participation of domain experts. Researchers usually divided knowledge construction into separate NLP tasks, including named entity recognition(NER) and relation extraction(RE). Since models and NLP tools based on the general domain cannot satisfy the form of scientific corpus or data, all steps require independent fine-tuning or adjustment for different scientific requirements, as well as human annotation with domain experts [4]. Some joint relation extraction methods, such as NovelTagging [13] or TPLinker [14], demonstrate the concept of unifying NER and RE steps, which can mitigate the error propagation between different models and reduce the dataset annotation cost and also improve the performance for tackling overlapping or complex relation scenarios. These also push the work on scientific knowledge graph construction.

2.2. LLM-based Relation Extraction and knowledge construction

With the rapid growth of LLMs, the few/zero-shot learning abilities via prompt learning on LLMs are widely investigated by NLP researchers. Recent progress in large language models (LLMs) like GPT series(GPT-3 [15], ChatGPT, GPT-4 [16]), LLama-series(LLama [17], Alpaca [18], Vicuna [19]), PaLM [20] and GLM [21] has showcased remarkable performance across a spectrum of natural language processing (NLP) tasks. The ongoing expansion in the number of model parameters and the size of training datasets has endowed LLMs with emergent capabilities, facilitating them to partake in in-context learning (ICL). In ICL, these models can derive insights from a limited set of demonstrative examples presented within the input context and exceed the previous baselines in multiple NLP tasks under few-shot or zero-shot settings, as well as relation extraction and knowledge construction [22].

Based on the ICL paradigm, Wadhwa et al. [23] propose comparing few-shot LLM-based relation extraction and fine-tuned baselines. At the same time, the result indicates that Few-shot prompting with GPT-3 achieves near SOTA performance and can exceed the previous supervised SOTA performance when using the Chain-of-Thought (COT)[24] prompt for enhancement. In addition, in the domain-specific area, Agrawal et al. [25] point out that LLMs perform well at zero- and few-shot clinical relation extraction despite not being explicitly trained in clinical records. Rajpoot et al. [26] developed the GPT-FinRE model using ICL and achieved a 0.718 F1 score on their financial dataset, REFinD. Moreover, some works such as LLM2KB [27] attempt to connect LLM to the knowledge construction task, which aims to construct a knowledge base from input relation triples, and link to the external knowledge base such as Wikidata to capture the related information. They achieved an average F1 score of 0.6185 across 21 relations in the LM-KBC challenge held at the ISWC 2023 conference.

2.3. Retrieval Augmentation Generation(RAG)

Retrieval-augmented generation (RAG) is a method that integrates traditional large language models (LLMs) with external knowledge sources to enhance their output quality and relevance. The primary goal of RAG systems is to overcome the limitations inherent in LLMs, such as content hallucination, outdated information, and lack of traceable reasoning processes by dynamically incorporating up-to-date information from external databases or other structured knowledge forms during the generation process [28]. This methodology significantly improves the models' accuracy and reliability by ensuring that the generated content is contextually relevant and factually accurate.

The operational framework of an RAG system typically involves querying an external database using a retrieval mechanism that can pull relevant information based on the input query or context. This retrieved content is then fed into the LLM, enabling the model to produce outputs enhanced by the externally sourced data. The integration of retrieval capabilities allows RAG systems to remain dynamic, adapting to new data and evolving user requirements.

Using ontology in RAG systems introduces a structured way of organizing information that can significantly enhance the retrieval process. Ontologies provide a framework for representing knowledge as a set of concepts within a domain and the relationships between those concepts. This structured approach allows RAG systems to perform more precise and context-aware retrievals, essential for generating accurate and relevant outputs.

A notable example is presented in the study by Sabrina Toro et al., where they explore the "Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI)" system. [29] This system utilizes ontologies to dynamically generate and update knowledge, illustrating a profound enhancement in managing and utilizing structured domain knowledge effectively in real-time applications. This methodology is particularly beneficial in fields requiring precise and up-to-date information, such as biomedical and environmental sciences, where it helps mitigate the substantial collaborative efforts typically required from domain experts.

Further, Julien Delile et al. introduce a graph-based retriever that captures the long-tailed data of biomedical knowledge. Their work emphasizes the utility of integrating graph-based knowledge representations with RAG systems to handle rare or less frequently seen information better.[30] By incorporating these knowledge graphs, the system achieves a more comprehensive retrieval coverage, significantly enhancing the generation capabilities of LLMs in the biomedical research domain. This approach enhances the retrieval of relevant data. It improves the overall accuracy and utility of the generated content by ensuring it reflects rare yet critical information that might otherwise be overlooked.

These studies demonstrate how ontologies and knowledge graphs can significantly enhance RAG systems by providing a more structured, nuanced, and comprehensive approach to information retrieval and generation. The structured nature of ontological frameworks allows for a deeper semantic understanding of queries and the relationships between different pieces of information, leading to more accurate and contextually appropriate outputs. However, there is still a lack of research which directly links to the ontology-enhanced RAG system for extracting information in the climate change area, which is the motivation for our research.

3. Case Study: Processing Climate Change Policies

To demonstrate our framework, we considered the Nationally Determined Contributions (NDCs) to the United Nations Framework Convention on Climate Change (UNFCCC). These documents, which are updated every five years, outline countries' national strategies to mitigate and adapt to climate change. Specifically, we focused on elements of NDCs pertaining to Nature-based Solutions (NBS), which, broadly speaking, is the sustainable use or use of natural features and processes to tackle socio-environmental issues (in this case, climate change).

As the urgency to mitigate and adapt to climate change becomes increasingly pressing, consideration of climate strategies and policies by analysts is likely to become a more common process. As such, the development of ontologies on relevant issues is likely to be of great utility to those involved, as they will essentially allow rapid and automated comparisons to be made between countries, institutions and other entities. Moreover, given that climate change issues typically cut across multiple themes and sectors (e.g. energy, transport, or, in the case of NBS, biodiversity conservation and natural resource management), efforts to assess (and ultimately maximise) levels of cross-sectoral alignment between policies and intended activities may also be expected to increase, and can also be facilitated through the use of topic-specific ontologies. In this regard, we hope that our efforts to develop an ontology specific to NBS will have further utility beyond this demonstration.

4. Method Design

4.1. Dataset

We collected the environmental reports we needed from the Convention on Biological Diversity(CBD) and the United Nations Climate Change(UNFCCC) website. We chose 10 reports from all publications by different countries, including Albania, Angola, Antigua and Barbuda, Armenia, Australia, Azerbaijan, Bahrain, Bangladesh, Nigeria, and Papua New Guinea (PNG).

4.2. Relation Triple Extraction Process

As figure 1 shows, our model mainly includes two main processes: Relation Triple Extraction aims to extract important features from the given NDC reports and output them as a fixed JSON format, while Relation Triple Alignment receive the structural result combining with a pre-designed ontology framework to assign entities to the specified class.

In the Relation Triple Extraction process, we follow the general RAG structure to load different NDC reports and split them into pages. Through one LLM as a paging encoder to generate different embeddings and store them in a vector database. Then, we design a structural prompt to let LLM extract featured information with JSON-style output by computing the similarity between the page's vectors. Our prompt and example result for the structural extraction are shown in Table 1.

4.3. NDC Ontology Framework Design

Then, to mostly capture the relation between different initiatives from different countries, using Stanford's WebProtégé [31], we create our knowledge base framework with one domain expert. The main structure(components) of the ontology and some related examples are shown as follows:

- **Object properties**: Object properties link pairs of instances (individuals). They describe the relationship between two individuals, e.g. *country A* benefits_from *project B*.
- **Data properties**: Data properties link instances (individuals) to data values. They describe the attributes of instances by assigning them specific data values, such as numbers or strings. E.g. Data property: *Project A* has_start_date *01/01/99*.
- **Class/Subclass**: A class is a category or type representing a set of entities with common properties. Classes can inherit properties from other classes, which means they can be arranged in a hierarchical structure. E.g. Class: *Country*
- **Named Individuals**: Named Individuals are specific instances of classes. They are concrete examples or entities that belong to one or more classes. In WebProtégé, individuals are created

Table 1
Prompt for Relation Triple Extraction

Prompt:
NDC, or Nationally Determined Contribution, is a climate action plan to cut emissions and adapt to climate impacts. I am interested in Country NDCs. In particular, I want to know some of the nature-based solutions proposed in those NDCs, who the actors are involved in, and what the challenges these solutions address are. I also want to know what specific initiatives are implementing these solutions, over what period, where, and with what budget. Please structure your reply as a list of labelled items, where each item is one of the following: nature-based solution, actor, challenge, start date, end date, location, and budget. If any of the information is not available, please leave it N/A. Please list as many initiatives as you can find in the document. Please output it as a JSON object as follows:

{{ "Initiative 1": {{
"Nature-Based Solution": "",
"Actor": "",
"Challenge": "",
"Start Date": "",
"End Date": "",
"Location": "",
"Budget": ""
}},

and start with Initiative 1, Initiative 2, etc. Please gives out as many as initiatives that mentioned in the NDC document as you can find.

Example result:
Here are some initiatives related to nature-based solutions in Papua New Guinea's NDC, based on the provided document:
{ "Initiative 1": {
"Nature-Based Solution": "Reducing Emissions from Deforestation and Forest Degradation and Enhancement of Carbon Stocks (REDD+)",
"Actor": "Papua New Guinea government, with support from development partners",
"Challenge": "Forest degradation and deforestation, particularly due to logging and land use changes for agriculture",
"Start Date": "2017",
"End Date": "2027",
"Location": "Nationwide",
"Budget": "N/A" }
...

to represent specific objects or entities in the domain of interest. E.g. *Indonesian* is a named individual and belongs to the class *Country*.

Figure 2 shows the most general classes, as well as some of the class hierarchies in our ontology (known as Aris). We defined seven main class categories that conclude different kinds of national policy attributes in 'Action', 'Actor', 'Budget', 'Challenge', 'Document', 'Location' and 'Outcome'. Similarly, in figure 3, we show the basic design for object and data properties, which are defined to reflect the relationship between classes and individuals. Figure 4 indicates an example ontology individual called 'FoodSecurityInNigeria' that belongs to the 'FoodSecurity' subclass, which links through several different subclass under the main class 'Challenge' and connected to the other three different main classes: 'Outcome', 'Actions' and 'Actor'.

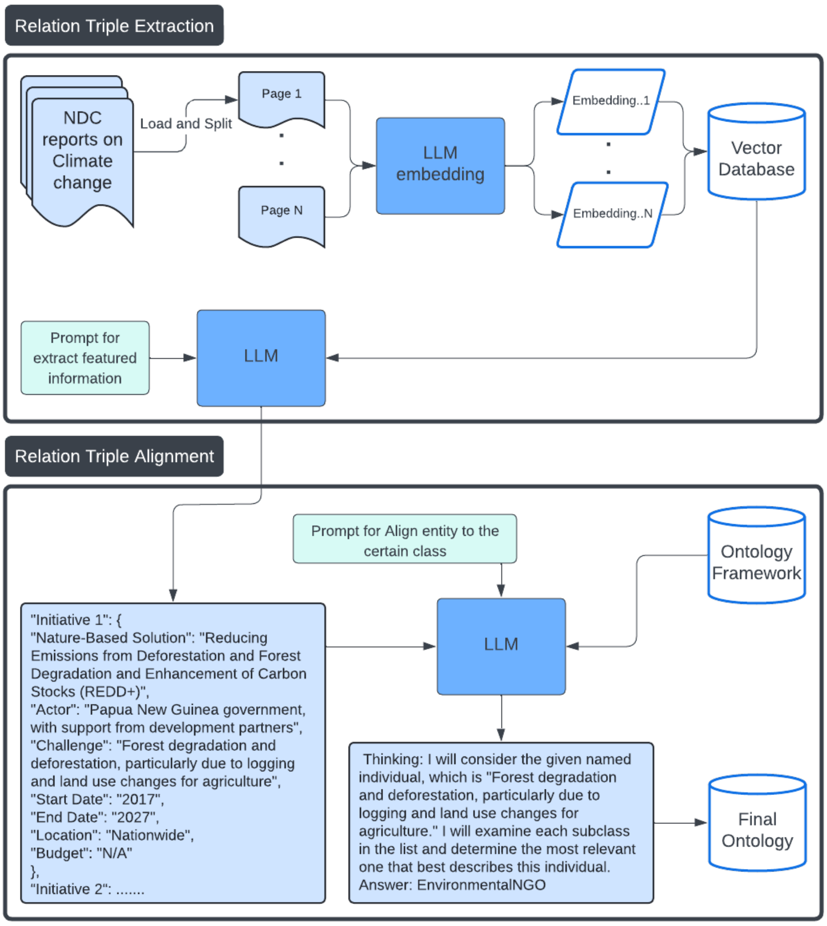

Figure 1: Model structure on LLM-based for knowledge graph completion

4.4. Relation Triple Alignment Process

For the relation triple alignment process, based on the output of the last step, we develop another prompt combined with the hierarchy structure of our Aris ontology framework to let LLM assign an entity to a leaf subclass such as the 'FoodShortages' in figure 2. Specifically, given a set of tree-structure class-subclass names as {JSON}, the LLM will classify the entity to a certain subclass and give the thinking process. Our prompts and example result are shown in Table 2, with content in parentheses used as a placeholder:

Following the progress above, we constructed the fulfilled Aris ontology with 10 countries' NDC reports. The complete version of Aris ontology contains 5923 relation triples with 401 classes and subclasses and 19 different relations. We extracted 155 environmental initiatives across 10 countries that contain nature-based solutions, challenges, actors and time period information.

4.5. RAG-based Question&Answer model Construction

In the next step, to verify the performance of our ontology straightforwardly, we implement an RAG-based Q&A model that combines the retrieval information from both vector database and structure ontology. As the figure 5 shows, the model will retrieve a set of select page content ranking by the similarity score with cosine similarity function, as well as generate related SPARQL query to achieve information from the Aris ontology to generate answers. Specifically, our method will extract entities related to the country in the question and then generate SPARQL query to extract related information

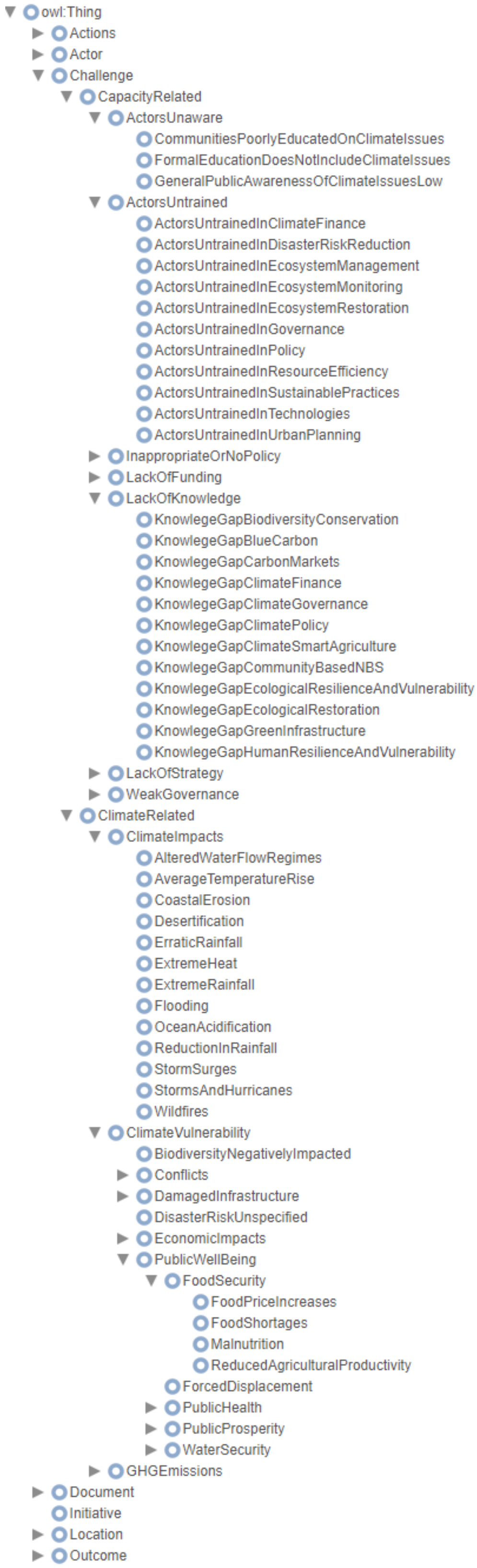

Figure 2: The main structure of Aris ontology, here we are showing the design of the Actions classes and subclasses.

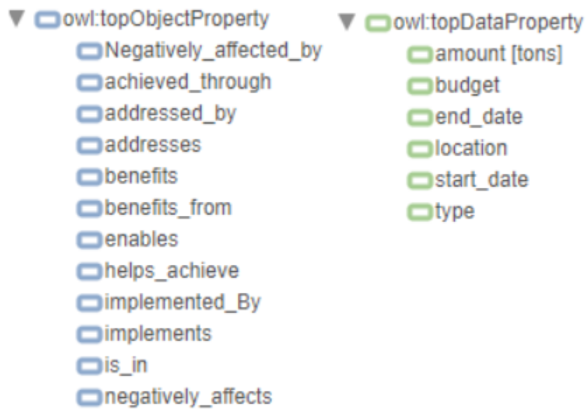

Figure 3: Defined Object and Data relation in Aris ontology.

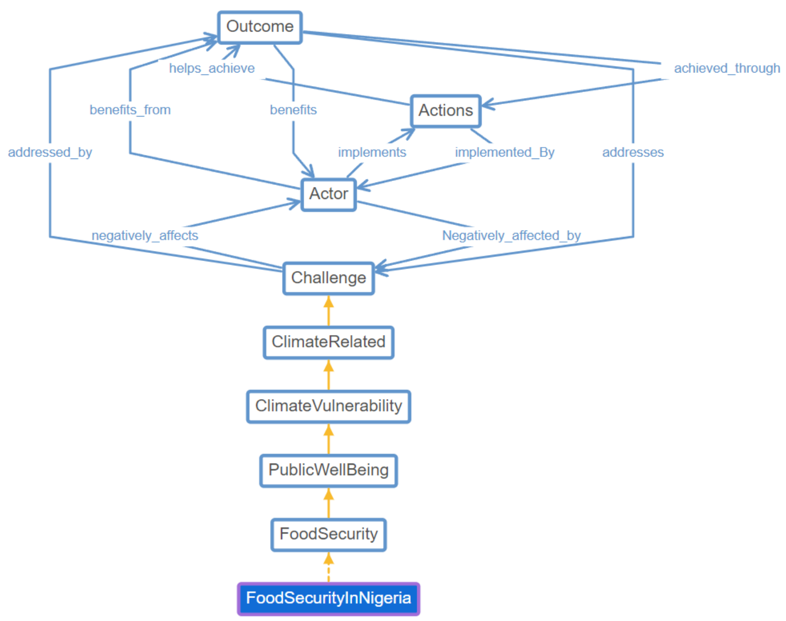

Figure 4: An example ontology individual defined in Aris ontology. It belongs to the subclass 'FoodSecurity' and is linked with the other top-level classes ('PublicWellBeing', 'ClimateVulnerability','ClimateRelated'), which belongs to the main class 'Challenge'

from ARIS ontology to generate the answer. To compare our result, we set the baseline as the general RAG framework with the retrieval vector from the page-split database.

5. Result Discussion

5.1. Question Set and Evaluation Setting

To evaluate our result, we propose a set of questions across the climate-change-based information we are interested in. To perform a clear structure for evaluating the result, we design the prompt prefix for the questions and let the LLM answer it by listing its answer, reason and reference. Some examples are shown in Table 3.

Based on these questions, we formulate our dataset across 10 countries and their combinations for single and double-country queries, then generate 40 questions for a single query and 45 questions for a double-country query. The RAG-based model will answer the questions based on the prompt to list

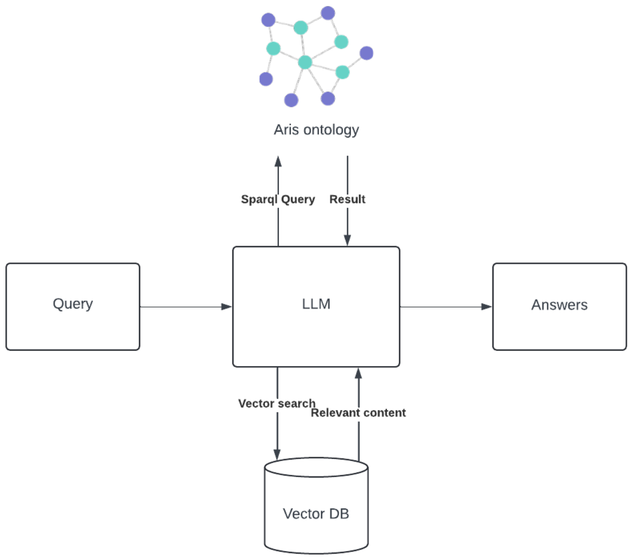

Figure 5: Structure of the hybrid RAG that mixed the knowledge graph data and retrieval content from the NDC report

Table 2
Prompt and example result for Relation Triple Alignment

Prompt:
Here is a list string that represents some classes in an ontology: {JSON}. Now, I want to give you a named individual:{Ind}, which should belong to one of these subclasses.
If we consider the JSON I give to you as a tree structure, please find the most relevant class that the individual belongs to by the meaning.
If you think no leaf subclass is relevant please answer "No match found".
Please only give one answer for the suitable subclass, but not two or more.
Please answer this question in this format precisely without any additional explanation or information:
Thinking: {{The steps you are thinking}}
Answer: {{The subclass name}}

Example Answer from LLM:
Thinking: The individual "Papua New Guinea government, with support from development partners" is a political entity at the national level. I will search for a matching subclass in the provided ontology that represents this concept.
Answer: NationalGovernment

points in enumerates. We manually verified our result based on the facts in the original NDC report against the following criteria:

Precision: How many answers match the facts in the NDC report?
Recall: How many answers from the model cover the points in the NDC reports?

We choose Mixtral 8*7B [32] as the core LLM for both embedding and ontology/RAG model. Compared to the other open-source model, the Mixtral 8*7B model has demonstrated superior performance across a variety of benchmarks, including code generation, math problem-solving, and general language understanding.

It matches or exceeds the capabilities of more extensive models like Llama 2 70B and GPT-3.5 with

Table 3
Prompt on RAG-based QA System and 2 example questions

Prompt:
Here is the retrieval context from the NDC report, please answer the question based only on the following context: {context}
Question: {question} Please think of the question step by step and provide a detailed answer, It also gives the reason and references that describe which page comes from the context of this format:
Answer:

Point 1:
Reason:

Point 2:
Reason:

Point 3:
...
Please provide as many answers as you can find.
Example Single Country Query Question:
"What specific climate-related challenges is {country} hoping to address through the use of nature-based solutions?"
Example Double Country Query Question:
Are {country X} and {country Y} sharing any same challenges or natural-based solutions in the climate change aspect?

the efficient memory requirement and longer context length, up to 32000 tokens. We locally run the Mixtral 8*7B model with 1 H100 GPU for all the tasks we proposed above.

5.2. Result Analysis

The result in table 4 shows the performance of our model. We separately test our model on the baseline model and the Hybrid RAG model with human evaluation to check the result. Compared to the baseline model, our model achieved 0.836 precision and 0.588 recall in the single-country query questions and 0.676 precision and 0.582 recall in the cross-country query questions, which indicates the success of the enhancement based on the ontology data during the RAG progress.

Table 4
Performance of methods across different strategies.

Method	Question type	Precision	Recall
Baseline (Raw RAG)	Single-country	0.686	0.482
	Cross-country	0.522	0.474
Hybrid	Single-country	0.836	0.588
	Cross-country	0.676	0.582

Below we show an example from both our ontology-enhanced RAG model and the baseline:

Question: "With regards to nature-based solutions, is country Albania setting measurable, time-bound targets? If so, what is being measured, and over what time period? "

Baseline: "No, Angola is not setting measurable, time-bound targets specifically for nature-based solutions in the provided context. The document outlines Angola's commitment to reduce greenhouse gas emissions and adapt to climate change but does not explicitly mention numerical targets for nature-based solutions."

Hybrid: ”Yes, Angola is setting measurable and time-bound targets with regards to nature-based solutions. The nature-based solutions mentioned in the ontology data are linked to specific challenges and have a start date of 2020 and an end date of 2030. This indicates that the country has committed to implementing these solutions over a 10-year period. The solutions being measured include Coastal Protection and Mangrove Restoration, Community-based Natural Resource Management and Conservation, Conservation and Restoration of Wetlands, Peatlands, and Other Blue Carbon Ecosystems, Disaster Risk Reduction and Early Warning Systems for Climate-related Hazards, Integrated Water Resources Management and Watershed Protection...”

Ontology data: {’solution’: ’IntegratedWaterResourcesManagementAndWatershedProtection’, ’challenge’: ’WaterScarcity_Flooding_AndClimateChangeImpactsOnWaterResources’, ’Actor’: ’MinistryOfEnergyAndWater’, ’budget’: ’N/A’, ’start_date’: ’2020’, ’end_date’: ’2030’}, ...

Here we can see that our model successfully answer the question with the ontology data even if the baseline hasn’t retrieve the related information from the text. Our model suffer from the manual evaluation efficiency which led to a limited question dataset to expand the question type to other type of environmental problem. However, our result and example demonstrate that to integrate with the ontology we design, our model can exceed the raw RAG system and sufficient in answer the NDC-based environment questions.

While our study introduces a novel integration of ontological frameworks with retrieval-augmented generation models, it also encounters specific limitations that pave the way for future research opportunities. Firstly, the current ontology is restricted to extracting relational triples from Nationally Determined Contributions (NDC) reports across ten countries. This limitation confines the breadth of our ontological database and potentially impacts the model’s applicability to global environmental data. Future efforts will aim to expand our method to include a wider range of countries, thereby enriching the ontology’s diversity and representativeness.

Secondly, the scope of our evaluation is constrained by the manual effort required, limiting us to a smaller dataset of questions and answers. This has potentially restricted our ability to thoroughly assess the model’s performance across more complex environmental queries and diverse data scenarios. Moving forward, we plan to develop more robust evaluation methodologies that can handle larger and more intricate datasets of environmental-related questions. This will not only improve the accuracy and reliability of our model but also enhance its utility for more comprehensive environmental information retrieval tasks.

By addressing these limitations, future research can significantly enhance the model’s functionality and applicability, providing more detailed and extensive support for environmental decision-making processes.

6. Conclusion

This study presents a novel integration of Retrieval Augmented Generation (RAG) with a specialized ontology tailored for environmental policy documents, particularly focusing on Nationally Determined Contributions (NDC) reports. The development of this hybrid model showcases an innovative approach to improving the accuracy and relevance of question-answering systems in the environmental domain. Our primary contribution lies in the formulation of a domain-specific knowledge graph that significantly augments the question-answering capabilities of large language models (LLMs) through ontology-based queries. The combination of RAG embeddings with these queries allows our model to deliver responses that are not only accurate but also deeply contextualized, drawing on the structured knowledge embedded in environmental policies. Testing across different LLMs has confirmed that our model offers a clear improvement over traditional RAG systems in handling complex environmental data. The ability to integrate and manipulate domain-specific ontologies enables the model to address intricate queries about climate change mitigation strategies and policies with a higher degree of precision. Our future work will focus on expanding the ontology’s scope, refining the integration mechanisms, and exploring the application of this model to other domains requiring high accuracy and context-specific responses. This endeavour not only advances the field of environmental informatics but also contributes to the broader discussion on the potential and limitations of applying advanced AI techniques in highly specialized and impactful areas.

References

[1] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, C. Bizer, DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia, Semantic Web 6 (2015) 167–195. URL: https://content.iospress.com/articles/semantic-web/sw134. doi:10.3233/SW-140134, publisher: IOS Press.

[2] C. Erten, N. Chotai, D. L. Kazakov, Pair Trading with an Ontology of SEC Financial Reports, in: The 2020 IEEE Symposium Series on Computational Intelligence:IEEE Symposium on Computational Intelligence for Financial Engineeting and Economics (CIFEr 2020), 2020. URL: https://eprints.whiterose.ac.uk/169191/.

[3] C. Erten, D. L. Kazakov, Ontology Graph Embeddings and ILP for Financial Forecasting, in: Proceedings of the 30th International Conference on Inductive Logic Programming, LNAI, Springer, GRC, 2021. URL: https://eprints.whiterose.ac.uk/178510/.

[4] M. Hofer, D. Obraczka, A. Saeedi, H. Köpcke, E. Rahm, Construction of Knowledge Graphs: State and Challenges, 2023. URL: https://arxiv.org/abs/2302.11509v1.

[5] J. Li, Y. Yuan, Z. Zhang, Enhancing llm factual accuracy with rag to counter hallucinations: A case study on domain-specific queries in private knowledge-bases, arXiv preprint arXiv:2403.10446 (2024).

[6] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, SIGMOD '08, Association for Computing Machinery, New York, NY, USA, 2008, pp. 1247–1250. URL: https://dl.acm.org/doi/10.1145/1376616.1376746. doi:10.1145/1376616.1376746.

[7] A. Saha, V. Pahuja, M. Khapra, K. Sankaranarayanan, S. Chandar, Complex Sequential Question Answering: Towards Learning to Converse Over Linked Question Answer Pairs with a Knowledge Graph, Proceedings of the AAAI Conference on Artificial Intelligence 32 (2018). URL: https://ojs.aaai.org/index.php/AAAI/article/view/11332. doi:10.1609/aaai.v32i1.11332, number: 1.

[8] Z. Liu, C. Xiong, M. Sun, Z. Liu, Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 2395–2405. URL: https://aclanthology.org/P18-1223. doi:10.18653/v1/P18-1223.

[9] L. Xu, Q. Zhou, K. Gong, X. Liang, J. Tang, L. Lin, End-to-End Knowledge-Routed Relational Dialogue System for Automatic Diagnosis, Proceedings of the AAAI Conference on Artificial Intelligence 33 (2019) 7346–7353. URL: https://ojs.aaai.org/index.php/AAAI/article/view/4722. doi:10.1609/aaai.v33i01.33017346, number: 01.

[10] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, M. Guo, RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 417–426. URL: https://doi.org/10.1145/3269206.3271739. doi:10.1145/3269206.3271739.

[11] M. Schmitt, S. Sharifzadeh, V. Tresp, H. Schütze, An Unsupervised Joint System for Text Generation from Knowledge Graphs and Semantic Parsing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 7117–7130. URL: https://aclanthology.org/2020.emnlp-main.577. doi:10.18653/v1/2020.emnlp-main.577.

[12] T. J. Callahan, A. L. Stefanski, J. M. Wyrwa, C. Zeng, A. Ostropolets, J. M. Banda, W. A. Baumgartner,

R. D. Boyce, E. Casiraghi, B. D. Coleman, J. H. Collins, S. J. Deakyne Davies, J. A. Feinstein, A. Y. Lin, B. Martin, N. A. Matentzoglu, D. Meeker, J. Reese, J. Sinclair, S. B. Taneja, K. E. Trinkley, N. A. Vasilevsky, A. E. Williams, X. A. Zhang, J. C. Denny, P. B. Ryan, G. Hripcsak, T. D. Bennett, M. A. Haendel, P. N. Robinson, L. E. Hunter, M. G. Kahn, Ontologizing health systems data at scale: making translational discovery a reality, npj Digital Medicine 6 (2023) 1–18. URL: https://www.nature.com/articles/s41746-023-00830-x. doi:10.1038/s41746-023-00830-x, number: 1 Publisher: Nature Publishing Group.

[13] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, B. Xu, Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1227–1236. URL: https://aclanthology.org/P17-1113. doi:10.18653/v1/P17-1113.

[14] Y. Wang, B. Yu, Y. Zhang, T. Liu, H. Zhu, L. Sun, TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 1572–1582. URL: https://aclanthology.org/2020.coling-main.138. doi:10.18653/v1/2020.coling-main.138.

[15] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language Models are Few-Shot Learners, in: Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

[16] OpenAI, GPT-4 Technical Report, 2023. URL: http://arxiv.org/abs/2303.08774. doi:10.48550/arXiv.2303.08774, arXiv:2303.08774 [cs].

[17] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, LLaMA: Open and Efficient Foundation Language Models, 2023. URL: http://arxiv.org/abs/2302.13971. doi:10.48550/arXiv.2302.13971, arXiv:2302.13971 [cs].

[18] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, T. B. Hashimoto, Stanford Alpaca: An Instruction-following LLaMA model, 2023. URL: https://github.com/tatsu-lab/stanford_alpaca, publication Title: GitHub repository.

[19] W.-L. Chiang, Z. Li, Z. Lin, Y. Sheng, Z. Wu, H. Zhang, L. Zheng, S. Zhuang, Y. Zhuang, J. E. Gonzalez, I. Stoica, E. P. Xing, Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality, 2023. URL: https://lmsys.org/blog/2023-03-30-vicuna/.

[20] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski, X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ippolito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi, D. Dohan, S. Agrawal, M. Omernick, A. M. Dai, T. S. Pillai, M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov, K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta, J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, N. Fiedel, PaLM: Scaling Language Modeling with Pathways, 2022. URL: http://arxiv.org/abs/2204.02311. doi:10.48550/arXiv.2204.02311, arXiv:2204.02311 [cs].

[21] Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, Z. Yang, J. Tang, GLM: General Language Model Pretraining with Autoregressive Blank Infilling, 2022. URL: http://arxiv.org/abs/2103.10360, arXiv:2103.10360 [cs].

[22] X. Zhao, Y. Deng, M. Yang, L. Wang, R. Zhang, H. Cheng, W. Lam, Y. Shen, R. Xu, A Comprehensive Survey on Deep Learning for Relation Extraction: Recent Advances and New Frontiers, 2023. URL: http://arxiv.org/abs/2306.02051, arXiv:2306.02051 [cs].

[23] S. Wadhwa, S. Amir, B. Wallace, Revisiting Relation Extraction in the era of Large Language Models, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 15566–15589. URL: https://aclanthology.org/2023.acl-long.868. doi:10.18653/v1/2023.acl-long.868.

[24] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, D. Zhou, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, 2023. URL: http://arxiv.org/abs/2201.11903, arXiv:2201.11903 [cs].

[25] M. Agrawal, S. Hegselmann, H. Lang, Y. Kim, D. Sontag, Large language models are few-shot clinical information extractors, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 1998–2022. URL: https://aclanthology.org/2022.emnlp-main.130. doi:10.18653/v1/2022.emnlp-main.130.

[26] P. K. Rajpoot, A. Parikh, GPT-FinRE: In-context Learning for Financial Relation Extraction using Large Language Models, 2023. URL: http://arxiv.org/abs/2306.17519. doi:10.48550/arXiv.2306.17519, arXiv:2306.17519 [cs].

[27] A. Nayak, H. P. Timmapathini, LLM2KB: Constructing Knowledge Bases using instruction tuned context aware Large Language Models, 2023. URL: http://arxiv.org/abs/2308.13207, arXiv:2308.13207 [cs].

[28] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, Q. Guo, M. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, ArXiv abs/2312.10997 (2023). URL: https://api.semanticscholar.org/CorpusID:266359151.

[29] S. Toro, A. V. Anagnostopoulos, S. Bello, K. Blumberg, R. Cameron, L. Carmody, A. D. Diehl, D. M. Dooley, W. Duncan, P. Fey, P. Gaudet, N. L. Harris, marcin p. joachimiak, L. Kiani, T. Lubiana, M. C. Munoz-Torres, S. T. O'Neil, D. Osumi-Sutherland, A. Puig, J. P. Reese, L. Reiser, S. M. C. Robb, T. Ruemping, J. Seager, E. Sid, R. Stefancsik, M. Weber, V. Wood, M. A. Haendel, C. J. Mungall, Dynamic retrieval augmented generation of ontologies using artificial intelligence (dragon-ai), ArXiv abs/2312.10904 (2023). URL: https://api.semanticscholar.org/CorpusID:266359619.

[30] J. Delile, S. Mukherjee, A. V. Pamel, L. Zhukov, Graph-based retriever captures the long tail of biomedical knowledge, ArXiv abs/2402.12352 (2024). URL: https://api.semanticscholar.org/CorpusID:267751251.

[31] T. Tudorache, C. Nyulas, N. F. Noy, M. A. Musen, WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the Web, Semantic Web 4 (2013) 89–99.

[32] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. de Las Casas, E. B. Hanna, F. Bressand, G. Lengyel, G. Bour, G. Lample, L. R. Lavaud, L. Saulnier, M.-A. Lachaux, P. Stock, S. Subramanian, S. Yang, S. Antoniak, T. L. Scao, T. Gervet, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mixtral of experts, ArXiv abs/2401.04088 (2024). URL: https://api.semanticscholar.org/CorpusID:266844877.