KnowUREnvironment: An Automated Knowledge Graph for Climate Change and Environmental Issues

Md Saiful Islam [1], Adiba Proma [1], Yilin Zhou [1],
Syeda Nahida Akter [2], Caleb Wohn [1], Ehsan Hoque [1]

[1] University of Rochester, United States
[2] Carnegie Mellon University, United States

Abstract

Despite climate change being one of the greatest threats to humanity, many people are still in denial or lack motivation for appropriate action. A structured source of knowledge can help increase public awareness while also helping crucial natural language understanding tasks such as information retrieval, question answering, and recommendation systems. We introduce *KnowUREnvironment* – a knowledge graph for climate change and related environmental issues, extracted from the scientific literature. We automatically identify $210,230$ domain-specific entities/concepts and encode how these concepts are interrelated with $411,860$ RDF triples backed up with evidence from the literature, without using any supervision or human intervention. Human evaluation shows our extracted triples are syntactically and factually correct (81.69% syntactic correctness and 75.85% precision). The proposed framework can be easily extended to any domain that can benefit from such a knowledge graph.

1 Introduction

Both misinformation and lack of information can be harmful when it comes to climate change. Misinformation can create confusion about crucial socio-political issues and makes it difficult for people to comprehend the true scenario. At the same time, a lack of knowledge about climate change and related environmental issues makes people underestimate the problem. Many people lack motivation for appropriate action (Joshi and Rahman 2015); many even deny that climate change is happening (Dunlap, McCright et al. 2011). The scenario can be improved by equipping the general public with accessible information. A relevant, appropriate knowledge graph can help people easily comprehend facts and connect different concepts, which eventually helps generate new forms of knowledge. Also, knowledge graphs have been frequently used to improve the performance and explainability of several downstream AI applications. One important application is automated fact-checking, where a statement is converted to a semantic graph and later compared against an existing knowledge graph to check for logical consistency (Gallagher 2006; Vedula and Parthasarathy 2021). Knowledge graphs are also extensively used for information retrieval, recommender systems, and question answering. Google uses a knowledge graph to improve understanding of queries and documents, thereby improving the relevance and diversity of search results (Zou 2020). Amazon introduced a product graph (Dong 2018) that improves product search and recommendation while consumers shop online. IBM Watson (Gliozzo et al. 2013) which defeated human experts in question answering, uses several knowledge bases as its source of information. Structured representation of knowledge also plays a vital role in chatbots and virtual assistants (e.g., Siri). Unlike deep-learning-based models lacking interpretability, knowledge graph-based models are explainable.

Several domains have benefitted from having a domain-specific knowledge graph. KnowEdu (Chen et al. 2018) is a knowledge graph for education that captures mathematical concepts and visualizes them as a graph, facilitating learning and scientific discovery. Knowledge graph for biomedical engineering is a hot topic, and has been applied for drug discovery (Sang et al. 2018; Zeng et al. 2022) and predicting disease co-morbidity (Biswas, Mitra, and Rao 2019). A knowledge graph for climate change can reduce misinformation on social media by using automated fact-checking. Such a knowledge graph can raise public awareness by automatically answering questions related to climate change, and help knowledge discovery and causal inference by finding relevant concepts in the knowledge graph. It can also enable recommendation systems to suggest more sustainable content (e.g., showing more eco-friendly products on e-commerce websites based on consumer preference (Islam et al. 2022)). Unfortunately, to the best of our knowledge, a knowledge graph focusing on climate change and related issues is not publicly available.

In this paper, we present **KnowUREnvironment** – a knowledge graph focusing on climate change and environmental issues, constructed automatically in an unsupervised manner. We use Resource Descriptor Framework (RDF) triples to represent our knowledge graph, where each triple consists of a subject, a predicate, and an object. Subjects and objects are nodes of the knowledge graph, often referred to as entities/concepts. Predicates are directed edges capturing the relationship between the subject and the object. As shown in Figure 1, “automobile” and “emission” are two nodes of the knowledge graph, where the “source” predicate denote their relationship that “automobile may

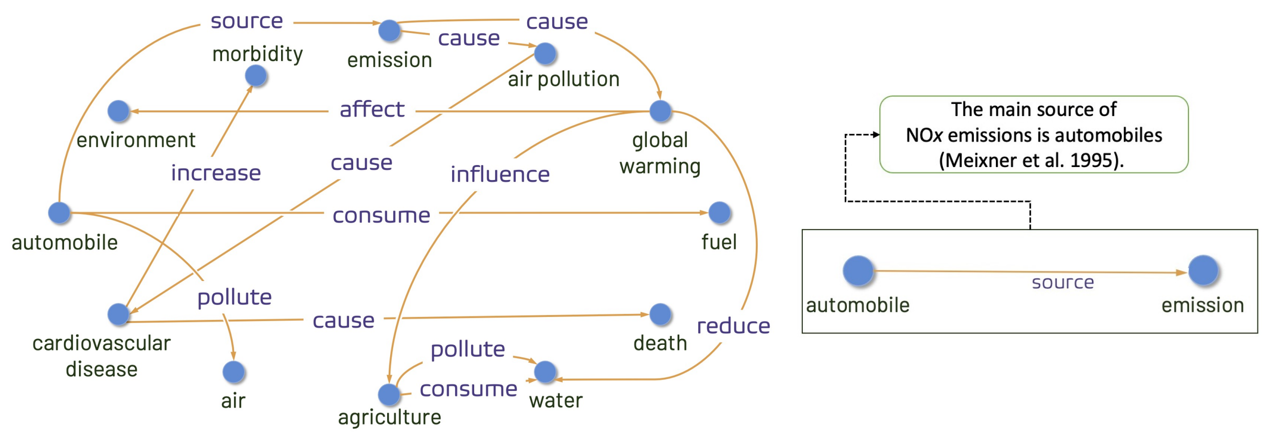

Figure 1: A snapshot of KnowUREnvironment depicts how 'automobile' could be related to climate change and related issues. Within a few hops, the graph can connect concepts from diverse yet relevant fields like environment, agriculture, and public health demonstrating how powerful yet so compact a knowledge graph can be. Additionally, each of the links in the graph is backed up by a scientific evidence.

source emission". We enrich each RDF triple with evidence sentences from scientific articles supporting the triple, thus making the knowledge graph explainable and more trustworthy. KnowUREnvironment connects diverse yet related concepts regarding climate change and makes it possible to infer complex relations. For example, it is possible to verify that "automobiles may impact the environment and public health" by analyzing the relations presented in Figure 1.

Automatic construction of a knowledge graph is a hard problem, and the most popular knowledge graphs are often constructed manually (e.g., Wikidata (Vrandečić and Krötzsch 2014), Wordnet (Miller 1995)). Despite having higher precision, the manual construction of a knowledge graph is very costly. Researchers also tried to combine named entity recognition (NER) and relation extraction (RE) systems to generate RDF triples (Maynard, Bontcheva, and Augenstein 2016; Mishra and Mittal 2021). Such systems need models trained on the domain-specific corpus to detect entities, requiring a significant amount of supervised data. Another disadvantage of relying on NER systems is that they are primarily trained to detect named entities like the date of an event, the name of a person, an organization, or a city. So, these systems struggle when identifying concepts (e.g., natural disasters, global warming, etc.) RE systems often rely on relations defined by an ontology (UzZaman and Allen 2010). Due to the unavailability of NER datasets and an ontology for climate change, we use semantic role labeling (SRL), an open information extraction system, to extract the RDF triples. From a given sentence, SRL can identify the key verb as an event and find associated roles related to the event. SRL is general purpose and believed to be more domain-independent compared to NER. Thus, researchers have adopted SRL for information extraction, specifically for domains where appropriate NER models are not available (Roopak and Deepak 2021).

We rely on the abstracts of $152,595$ relevant, methodologically collected scientific articles published in several conferences or journals as the unstructured source of knowledge. Although news articles, social media, blogs, or online websites can be used as alternative sources of knowledge, these are often not trusted (Kiousis 2001). We prefer abstracts to the entire paper because abstracts are usually publicly available and capture the essence of the whole article. However, working with scientific literature is challenging, as the abstracts often use complex sentences and passive voice. Due to these reasons, SRL performs poorly on scientific texts and generates erroneous triples (Groth et al. 2018). To improve the precision of the SRL extracted triples, we further impose syntax verification to throw away spurious triples. We also keep the source sentence(s) from where a triple is extracted as evidence for the triple. RDF triples passing syntax verification are "candidate triples". If an RDF triple is extracted from multiple sentences, it is said to have multiple evidence and is promoted to a "trusted triple". KnowUREnvironment consists of $24,263$ trusted triples that capture the interaction among $10,321$ unique concepts. Despite not being a direct part of the knowledge graph, $387,597$ candidate triples also capture potential facts that could be trusted upon further evidence from new scientific articles. As opposed to being limited to extracting a certain set of pre-defined relations, KnowUREnvironment consists of $4,323$ unique relations picked up automatically. Despite the complexity of the scientific text, the knowledge graph has 75.85% precision and 81.69% syntactic correctness, as estimated by human evaluation. This suggests the syntax verification and evidence count steps help improve the accuracy of the extracted triples.

In summary, we make the following contributions in this paper:

- We introduce KnowUREnvironment, a knowledge graph for climate change and environmental issues, that can help knowledge understanding and reasoning, as well as crucial NLP applications for climate change. The proposed method can be easily adapted for any academic research domain.

- We integrate evidence with each triple that can increase the explainability and trustworthiness of the knowledge graph.
- We show that syntax verification and counting evidence on top of semantic role labeling can help extract precise RDF triples.

The domain-specific article abstracts and the knowledge graph are made publicly available[1], hoping for further progress in natural language processing tasks for climate change.

2 Method

This section describes the text corpus and data models of KnowUREnvironment, followed by our proposed framework for generating the knowledge graph from the scientific literature on climate change and environmental science as depicted in Figure 2. Note that, although we are building a domain-specific knowledge graph, all the components of our pipeline can be replicated for any other domain.

2.1 Text Corpus

We depend on a large corpus of unstructured text to build our knowledge graph. Although many online sources offer a huge amount of information related to climate change, their credibility is questioned (Metzger, Flanagin, and Medders 2010; Metzger and Flanagin 2013; Abdulla et al. 2002). On the other hand, scientific articles published in different journals and conferences are mostly peer-reviewed and more trustworthy. We collect 1.04M sentences from 152,595 abstracts of academic papers highly relevant to climate change and associated environmental issues in two phases.

Phase 1: We use S2ORC (Lo et al. 2020) that provides 8.1M open-access papers with abstract and rich metadata from multiple academic disciplines. Then we use a complex search that tries to match the author-provided keywords available as the metadata with multiple keyword strings most relevant to climate change and associated environmental issues to find 228,860 relevant papers. Each keyword (both author-provided keywords and the search keywords) is normalized following four simple steps: (i) tokenization[2] using NLTK (Bird 2006) word_tokenize, (ii) removing punctuation and stop words, (iii) lemmatizing[3] each token with NLTK WordNet lemmatizer, and (iv) concatenate all the lemmatized tokens. The search keywords are hand-picked based on the popular issues associated with climate change: 'climate change', 'sustainability', 'pollution', 'global warming', 'sea-level rise', 'climate', 'water stress', 'coastal flooding'. We only consider the English abstract of papers where the abstract is publicly available, and the author-provided keywords match any of the search keywords in their normalized form.

Phase 2: We analyze the author-provided keywords of the entire S2ORC corpus and our extracted domain-specific papers to find 4,650 keywords that are most domain-frequent. The domain frequency for a particular keyword x, $DF_x = \text{COUNT-DOMAIN}(x)/\text{COUNT-ALL}(x)$ where COUNT-DOMAIN(x) and COUNT-ALL(x) are the number of times x appears as an author-provided keyword in the domain-specific papers and in all the papers of S2ORC corpus respectively. These $4,650$ keywords (out of $394,580$ distinct author-provided keywords) covered 80% of the keyword appearances in the domain-specific corpus. For each of the extracted domain-specific papers in phase 1, we only keep it in our corpus if its author-provided keywords contain any of the 4650 domain-frequent keywords. This phase further ensures our corpus is closely relevant to the issues we are investigating and eventually helps to build a knowledge graph highly relevant to the domain.

2.2 Data Model

We use RDF triples (Candan, Liu, and Suvarna 2001) data model to represent the facts extracted from the scientific literature. Each triple (t_s, t_p, t_o) consists of a subject (t_s), a predicate (t_p), and an object (t_o) in normalized (lemmatized) form, where t_s and t_o are domain-specific entities or concepts and t_p describes the relationship between the subject t_s and the object t_o. For example, the fact "Cardiovascular disease causes death." is represented as ("cardiovascular disease", "cause", "death").

To generate the knowledge graph, we convert the RDF triples into a labeled directed multigraph. Formally, a knowledge graph $KG = (V, E, s, t, l_E)$ is a graph where V is the set of nodes, E is the set of edges, $s : E \to V$ and $t : E \to V$ maps each edge to their source and target nodes respectively, and $l_E : E \to \Sigma^*$ maps edges to their corresponding labels in natural language (Σ is the alphabet of the language). In our knowledge graph, a node represents a subject/object of the RDF triples and the label of an edge represents a predicate.

Figure 1 presents a visual representation of a part of the knowledge graph. Since there may be multiple relationships between a subject-object pair (e.g., as in Figure 1, agriculture may consume water or agriculture may pollute water), we choose a multi-edged graph instead of a simple graph model. A subject or an object is also referred to as an entity or a concept throughout the paper in an interchangeable manner.

2.3 Knowledge Graph Construction

Our framework can be divided into three major steps: *(i) Triple Extraction* step uses methods from Open Information Extraction to generate RDF triples, which are further validated considering syntactic perspective in the *(ii) Syntax Verification* step, and finally, the triples are converted to a directed multigraph in *(iii) Evidence Counting and Graph Construction* step.

(i) Triple Extraction. An RDF triple typically encodes a fact consisting of a subject, an object, and a predicate that describes the relationship between the subject and object.

[1] https://github.com/saiful1105020/KnowUREnvironment

[2] Tokenization is the process of converting a piece of text (e.g., sentence) into smaller tokens (e.g., words).

[3] Lemmatization is the process of converting a word to its root form (e.g., better → good).

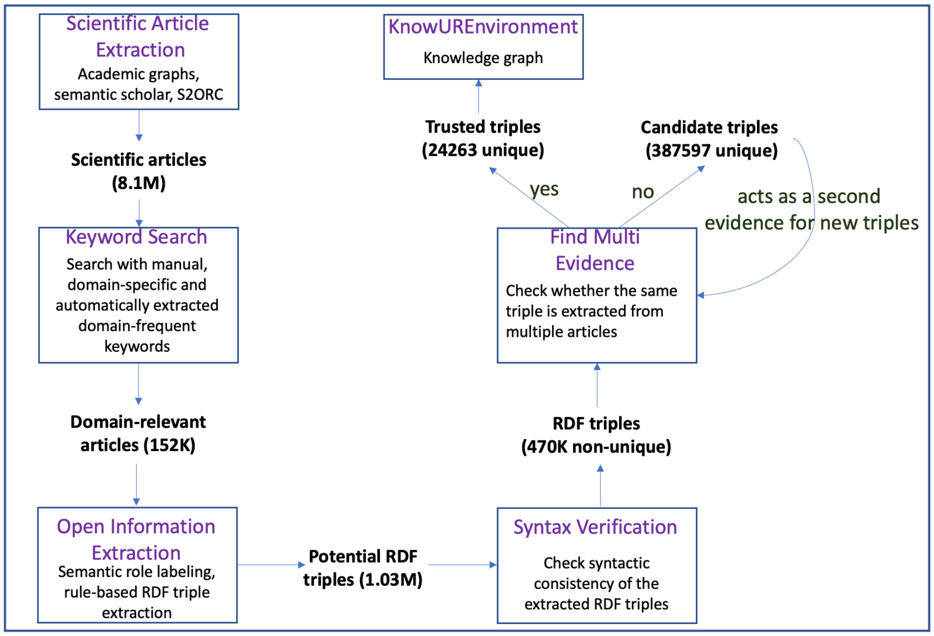

Figure 2: A brief overview of our methodology. Key methods are enclosed by a rectangle, while intermediate data (and their counts) are boldfaced. The trusted triples immediately get included in the knowledge graph. The candidate triples can act like evidence for future triples – if any future triple matches a candidate triple, it is promoted to be a trusted triple.

It is a common practice to use Named Entity Recognition (NER) (Nadeau and Sekine 2007; Lample et al. 2016) to identify subjects and objects from the text. However, using NER for climate change requires domain-specific training with a huge amount of human-annotated text, which is not available. For this reason, we choose to use Open Information Extraction (Etzioni et al. 2008; Stanovsky et al. 2018) techniques that are domain independent (Etzioni et al. 2008). Specifically, we use Abstract Meaning Representation (AMR) (Banarescu et al. 2013) for its availability and easy integration.

AMR is a semantic representation language that identifies the root verb of a sentence indicating an event and assigns semantic roles[4] for words/phrases in the sentence. In AMR, the frame arguments are the most important semantic roles for extracting RDF triples – more specifically, ARG0 and ARG1 representing the agent and the affected role of an event (following OntoNotes (Pradhan et al. 2007) convention) can be treated as the subject and object respectively. The root verb is represented following PropBank (Kingsbury and Palmer 2002) convention. For example, given a sentence "Automobiles emit CO2.", AMR converts it into a semantic graph:

```
(e / emit-01
        :ARG0 (a / automobile)
        :ARG1 (s / small-molecule
                :name (n / name
                        :op1 "CO2")))
```

[4]https://www.isi.edu/ ulf/amr/lib/roles.html

The AMR semantic graph captures that "emit" is the key event (predicate for the triple), 'automobile" is the subject, and "CO2" is the object. It also provides additional information that "CO2" is a small molecule. We use Penman (Goodman 2020) to parse the AMR graph and use rule-based methods to extract the arguments accurately. The RDF triple formed from this sentence is ("automobile", "emit", "CO2").

(ii) Syntax Verification. Although AMR provides domain-independent semantic representation, OpenIE systems in general, suffer when dealing with scientific literature (Groth et al. 2018) due to the complex nature of sentences presented in academic papers. For this reason, AMR extracts many invalid RDF triples and may yield poor alignment between the AMR graph and the text tokens, leading to faulty triple generation. We use rule-based systems to verify the syntactic consistency of the extracted triplets in Step (i). Before applying the consistency check, we apply some simple pre-processing techniques:

First person resolution: in academic literature, the first person (e.g., we) typically refers to the authors/researchers, and "this paper/article" refers to research. We create a manual map to convert these types of entities to their appropriate form (e.g., "we" $\rightarrow$ "researchers", "this paper" $\rightarrow$ "research", etc.)

Removing parenthesis-enclosed text: it is common to use an abbreviated form or an explanation right after the original phrase using a parenthesis. We make sure the parenthesis enclosed text is removed from the entities of RDF triples to generate a cleaner form of the entities.

For example, “united states (u.s.)” is converted to “united states”.

The syntax verification process consists of checking for three types of errors in the entities as mentioned below. If an RDF triple fails any of the error checks, it is flagged as an inconsistent triple and removed from our knowledge graph. Although this type of aggressive filtering may cause loss of recall, the procedures help us extract cleaner and more precise triples. We do not check the syntax of the predicates since they follow a standard convention (PropBank).

(a) Shared words between the subject and the object: AMR graphs can be erroneous for complex academic sentences. In many cases, AMR labels a substring of ARG0 as ARG1 and vice versa. For example, from the sentence “The use of total lipopolysaccharide (LPS) as a rapid biomarker for bacterial pollution was investigated at a bathing and surfing beach during the UK bathing season.” (Sattar, Jackson, and Bradley 2014), AMR labels “biomarker” as the root verb (predicate), “a rapid biomarker of bacterial pollution” as the subject, and“pollution” as the object. One way to identify this type of error is to check whether the subject and the object share common words (except for common English stop-words). We convert the subject and object in their lowercase form, remove punctuation and stop-words, and then check whether they have any common word left between them (indicating a possible error).

(b) Repeated words in an entity or too long entity: for complex arguments in AMR graph, we use AMR graph to text generation model (`gtos.generate()`) to generate the entities. However, AMR-to-text is still an active research topic with many limitations (Manning, Wein, and Schneider 2020). A common type of error we observed is repeated word generation. Moreover, the text-generation model may add meaningless/irrelevant text, thereby producing a subject or object that is very long. For each subject/object in RDF triples, we first check whether its length exceeds 50 characters. If so, we flag the triple as a possible error. Otherwise, we convert the subject/object to lowercase, remove punctuation and stopwords, and then check whether any word is repeated to further flag the triple as a possible error.

(c) Entity is not a noun phrase: in an RDF triple, the entities (subject/object) are supposed to be noun phrases while the predicate is in verb form. We use NLTK (Bird 2006) POS tagging accompanied with syntactic rules to flag the entities that are not noun phrases and eventually remove inconsistent triples. The syntactic rules are:

```
NNx → NN | NNP | NNPS | NNS | CD
NP → NNx
NP → DT NP
NP → JJ NP
NP → NNx NP
NP → NP IN NP
NP → NP CC NP
NP → NP , NP
```

Here, `NNx, CD, DT, JJ, IN, and CC` are NLTK POS tags representing different types of nouns, numerical entity, article/determiner, adjective, preposition, and coordinating conjunction respectively.

(iii) Evidence Counting and Graph Construction. We keep the sentence from where an RDF triple is extracted as an evidence. If the same triple is extracted from multiple sentences, it will have multiple evidence. We classify the triples verified in Step (ii) into two groups: (a) candidate triples having a single evidence and (b) trusted triples having multiple evidence. A candidate triple could be either erroneous or less studied. As scientific literature is ever growing, the knowledge graph construction needs to be done periodically with the new articles published. Whenever another evidence is found for a candidate triple in the new articles, the triple could be promoted to a trusted triple. This type of framework where the system is learning continuously and growing its knowledge base is itself an interesting research area (Carlson et al. 2010).

After being vetted by multiple evidence, all the trusted triples are combined to construct KnowUREnvironment, a directed multi-graph $G(V, E, s, t, l_E)$ that represents all the knowledge in a structured and compact graph. We first generate the list of all subjects and objects present in the trusted RDF triples, which becomes the vertex set of the graph (V). For each of the trusted triple (t_s, t_p, t_o), a unique id $e \in E$ is generated that represents an edge of the graph, and then $s[e] = t_s$ and $t[e] = t_o$ maps the source and target node of edge e. Finally, e is labeled with predicate t_p (i.e., $l_E[e] = t_p$). We use Python library Networkx (Hagberg, Swart, and S Chult 2008) for implementation, visualization, and analysis of KnowUREnvironment.

3 Evaluation and Analysis

In this section, we present a qualitative and quantitative evaluation of KnowUREnvironment. We also provide a preliminary analysis of the knowledge graph.

3.1 Human Evaluation

Automated evaluation of a knowledge graph is a hard problem, especially when a knowledge graph is domain-specific. A common approach is comparing the newly constructed knowledge graph with an existing one. However, to the best of our knowledge, no prior knowledge graph exists that focuses on climate change and related issues. Since there is no accessible and suitable ground truth, we had to rely on human judgment. We recruited three human annotators who voluntarily agreed to evaluate KnowUREnvironment. All the annotators were either undergraduate students or completed Bachelor’s degrees from a reputed Engineering university. We use A, B, and C to identify the annotators while respecting their privacy.

We are mainly interested in the precision and relevance of the knowledge graph since it is impossible to assess the recall without ground truth. We randomly sampled 650 triples from the knowledge graph that were evaluated by the human annotators. Each annotator evaluated 250 triples: 50 triples being the same for all and 200 triples being unique to each annotator. Annotators labeled accuracy, syntactic correctness, domain relevance, and ambiguity of each triple with binary judgments. Since the annotators may lack domain knowledge, we provided one evidence sentence for each triple from where the triple was extracted. As further context, the entire abstract containing the evidence sentence was also provided. When evaluating each triple, we requested the annotators to use their common sense and prior knowledge, in addition to considering the provided evidence sentence or abstract as needed.

We use the common 50 triples to assess inter-rater agreement among the annotators in terms of Fleiss' kappa score. We take the majority-agreed label as the ground truth when a triple has labels from multiple annotators. All 650 triples were used to estimate the performance of our knowledge graph.

Precision. Precision reflects the average accuracy of the facts presented in the knowledge graph. If the subject and object are well-formed (an ill-formed subject can be a nonsense word, or a phrase without meaning, typically generated by the NLP components due to their incorrect understanding of the language) and the predicate captures a correct relationship between the subject and object, then the triple is factually accurate and labeled with 1. Otherwise, the triple is labeled with 0. Table 1 shows the precision of our knowledge graph as evaluated by the annotators. We obtain 75.85% overall precision when reported on all the 650 triples evaluated.

Annotator	Accuracy of common triples	Accuracy of unique triples	Overall
A	64%	66.5%	66%
B	74%	71.5%	72%
C	82%	88.5%	87.2%
Majority	80%	–	**75.85%**

Table 1: The precision of KnowUREnvironment. Fleiss' kappa score of 0.455 shows moderate agreement among the annotators.

However, the performance estimate may be affected by the subjective bias of human annotators. The inter-rater agreement among the annotators is 0.455, showing moderate agreement. Out of the 50 common triples, three annotators fully agreed on 34 labels, while showing disagreement on 16 labels. Annotator A caused the most disagreement – labeling 8 triples as inaccurate while the other two labeled them as accurate. When evaluated on the common triples where the majority agreed label is considered to be the gold label, the precision is 80%.

Syntactic Correctness. In addition to precision, we also evaluate the syntactic correctness of the triples of KnowUREnvironment. If the subject and the object represent an entity or concept (typically in a noun phrase), and the predicate represents a relationship as a verb phrase, then the triple is syntactically correct and labeled as 1. For example, ("bird", "eat", "fish") is syntactically correct, but ("active", "impact", "climate change") is syntactically incorrect because "active" is not a concept/noun phrase ("activity" would have been correct).

Annotator	Syntactic correctness of common triples	Syntactic correctness of unique triples	Overall
A	78%	75%	75.6%
B	80%	81%	80.4%
C	82%	90%	88.4%
Majority	78%	–	**81.69%**

Table 2: Syntactic correctness of KnowUREnvironment. Fleiss' kappa score of 0.708 shows substantial agreement among the annotators.

Figure 3: Word cloud visualization of all the concepts captured in KnowUREnvironment.

As presented in Table 2, the average syntactic correctness of the triples is 81.69%. Unlike precision estimation, the inter-rater agreement is substantial for syntactic correctness evaluation, as indicated by a Fleiss' kappa score of 0.708.

Domain Relevance. In addition to evaluating factual and syntactic correctness, we also evaluate whether the concepts captured by KnowUREnvironment are relevant to the domain. We asked the human annotators to label a subject/object with 0 if they thought it was irrelevant to climate change and environmental issues (otherwise with 1). On average, 66.15% of the subjects and 60% of the objects were reported as domain-relevant. However, measuring domain relevance requires significant domain-specific expertise which our annotators lacked. This resulted in poor ($\kappa = -0.07$) and slight ($\kappa = 0.11$) inter-rater agreement while measuring the relevance of the subjects and objects respectively.

We also provide a word cloud visualization of all the concepts in our knowledge graph in Figure 3. The concepts that appear more frequently in the knowledge graph are presented in a larger font in a word cloud. Some of the most frequent concepts like "researchers", "pollution", "environment", "water", and "drought" can be easily related to climate change and academic research.

Ambiguity. We asked the annotators to rate each triple as understandable or ambiguous. A triple is understandable if the annotator can interpret a single, unique meaning from the triple without making any assumptions. Since a knowledge graph can lose important information due to its nature of aggressive summarization of text content, it can suffer from a lack of clarity.

Annotator	Understandability of common triples	Understandability of unique triples	Overall
A	18%	41.5%	36.8%
B	50%	34%	37.2%
C	64%	65.5%	65.2%
Majority	44%	–	**46.77%**

Table 3: Understandability of the triples presented in KnowUREnvironment. Annotators agreed fairly as indicated by a Fleiss' kappa score of 0.296.

Property	Value
# unique subjects	5240
# unique objects	6660
# unique predicates	4323
# triples	24263
# nodes	10321
average degree	4.70
diameter	9

Table 4: Size and basic properties of KnowUREnvironment.

46.77% of the triples in KnowUREnvironment were understandable without ambiguity, as evaluated by our human annotators. However, we also need to consider the high subjectivity of this evaluation as the inter-rater agreement is "fair" ($\kappa = 0.296$).

3.2 Knowledge Graph Analysis

We present further analysis of KnowUREnvironment in terms of the size of the knowledge graph and the diversity of the predicates. We also demonstrate how having multiple pieces of evidence affects the reliability of the triples.

Size. Table 4 presents some basic properties of the knowledge graph. KnowUREnvironment consists of $24,263$ trusted triples, $5,240$ unique subjects and $6,660$ unique objects. The network structure of the knowledge graph is sparse, having an average degree (in-degree + out-degree) of 4.70. This indicates a potential weakness of the knowledge graph in terms of missing out on existing relations (loss of recall). However, in this study, we mainly focused on obtaining accurate triples since further semi-supervised algorithms can be developed that exploit the patterns of extracted triples to find new triples. This strategy is often known as distant supervision (Mintz et al. 2009).

Since scientific literature is ever-evolving, we propose a structure that can integrate new information in the light of previously extracted information (see Figure 2) – a core criterion for never-ending learning (Mitchell et al. 2018). Although the knowledge graph consists of the trusted triples, candidate triples possess the potential to get promoted to trusted triples if further evidence is found in the scientific literature. We were able to extract $387,597$ candidate triples capturing $102,728$ unique subjects, $131,057$ unique objects, and $4,287$ unique predicates that have the potential to be included in the knowledge graph upon further evidence from the growing scientific literature.

Figure 4: Word cloud visualization of diverse predicates in our knowledge graph.

Diversity of Predicates. A significant strength of our knowledge graph is being able to capture diverse sets of predicates, while many studies focusing on automated knowledge graph construction are only able to extract a few pre-defined predicates (Fabian et al. 2007; Hoffart et al. 2013). As visualized in Figure 4, KnowUREnvironment contains $4,323$ unique predicates, focusing on causal interactions among entities relevant to climate change.

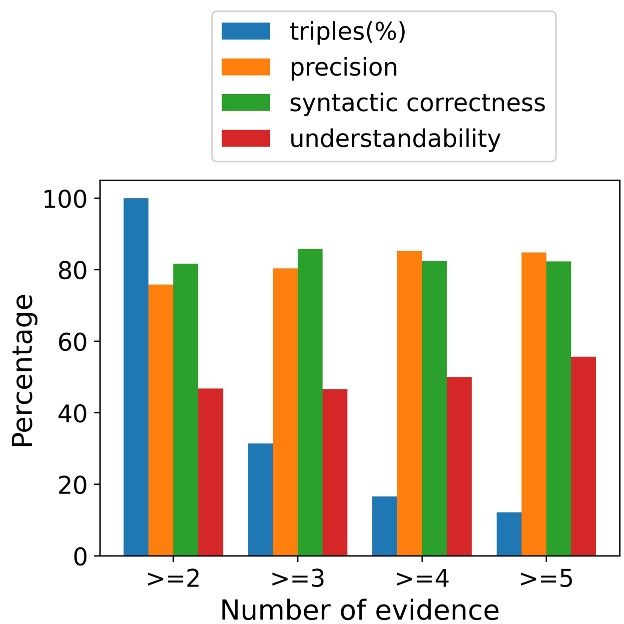

Figure 5: Triples that have more evidence are more accurate and understandable, although the number of triples decreases significantly. The triple count is presented as a percentage (100% = 650) to match the scale of other metrics.

Impact of Evidence Count. An evidence sentence is a sentence of a scientific article's abstract from where the triple is extracted. All the triples in KnowUREnvironment are backed up by multiple such evidence sentences. The core idea is that a triple with single evidence might be spurious or less studied. If a triple has multiple evidence sentences, we expect it to be more accurate and less susceptible to systematic errors. Based on the 650 triples evaluated by the human annotators, we analyze the effect of evidence count on the quality of the triples. As shown in Figure 5, triples with more evidence are more precise and understandable. As opposed to 75.85% precision and 46.77% understandability when triples have at least two evidence sentences, the precision and understandability reach 84.81% and 55.70% respectively when triples have at least five supporting evidence sentences. Although more evidence could indicate higher precision, the syntactic correctness remains mostly unaffected, perhaps due to our imposed syntax verification while constructing the knowledge graph. However, the downside of requiring more evidence to trust a triple is a significant reduction in the number of such triples.

4 Discussion

Although a large amount of unstructured information related to climate change is available, most automated systems can utilize them only when presented in a structured manner. A knowledge graph captures critical concepts and their relations in a structured and concise way, eventually helping further knowledge synthesis. However, automatically constructing such a knowledge graph is an open research problem. KnowUREnvironment is our first step toward a climate change knowledge graph and will require further iteration to reach its maturity. In this section, we identify key challenges and limitations of the work and briefly discuss some interesting future directions.

Limitations of the knowledge graph. Despite our best attempts, KnowUREnvironment still contains some erroneous RDF triples. The open information extraction pipeline often yields words out of English vocabulary, and the semantic roles obtained by parsing AMR graphs are often meaningless phrases. AMR also faces difficulty when a sentence is in a passive voice or contains negation, thus possibly extracting inaccurate triples.

Task specific benchmark datasets. We could not explore how KnowUREnvironment can be used in downstream NLP tasks like question answering, information retrieval, and fact-checking due to the unavailability of task-specific datasets. Introducing benchmark datasets for these tasks may open the door to climate change research using NLP.

Reducing complexity of scientific text. Scientific articles are one of the most trusted source of knowledge, since most published papers are peer-reviewed. However, when presented as evidence, these sentences are often complex and difficult to understand for a general audience. Also, NLP systems such as semantic role labeling performs poorly on complex scientific text. In the future, we plan to reduce the complexity of the sentences as a pre-processing step.

Resolving references and abbreviations. Coreference resolution is the task of resolving repeated reference to existing objects in the text (Sukthanker et al. 2020). Additionally, scientific articles include abbreviations of concepts and use them later to refer to the original concepts. Resolving references and abbreviations as a pre-processing step could help us extract more triples in the future.

Experimenting other SRL techniques. We used AMR for semantic role labeling from a given sentence. Although it is domain independent and easy to integrate with our pipeline, AMR suffers from several limitations. AMR provides the roles as a semantic graph and aligning AMR graph with corresponding text tokens is still an active research area. As alignment becomes more difficult with the increased complexity of scientific text, lots of erroneous triples were generated. We had to discard $562K$ out of $1.03M$ triples during syntax verification step – $61,982$ for having common words between the subject and the object, $116,728$ triples where the subject or the object was too long or had repeated words, and $439,009$ triples where the subject or the object were not noun phrases. In the future, we will experiment with other SRL techniques to improve this scenario.

Expert evaluation. For human evaluation, we employed undergraduate students with good reading skills. However, they are computer science majors and not experts in the area. Although decent inter-rater agreement scores indicate the reliability of our evaluation, being able to consult domain experts might enable us to reach better conclusions where specific expertise is necessary (e.g., evaluating the domain relevance of the concepts).

Standardized ontology. A knowledge graph becomes more valuable to the community when it is standardized. Having a standard concept ontology on climate change would enable researchers to share a common understanding of structured knowledge, reuse, and analyze domain knowledge (Noy, McGuinness et al. 2001). Building an ontology for climate change and linking KnowUREnvironment to the ontology could be rewarding for future research.

Bootstrap learning. Bootstrap learning uses already learned information to find further information with similar patterns (Mintz et al. 2009; Mitchell et al. 2018). In this paper, we find that RDF triples with multiple evidence sentences have higher precision and thus could be trusted more. It would be interesting to explore whether bootstrap learning succeeds to extract more triples using the "highly trusted" RDF triples.

5 Conclusion

In this paper, we introduce KnowUREnvironment, a knowledge graph specific to climate change and associated environmental issues. Although we had to use open information extraction methods in the absence of domain-specific NLP tools and datasets, our first attempt shows promising results – 75.85% precision and 81.69% syntactic correctness, as estimated by human evaluation. We plan to improve the knowledge graph and standardize it in the future, hoping to facilitate further NLP research on climate change.

Acknowledgement

This research was supported by funding from the Goergen Institute for Data Science. We acknowledge Aaron Steven White and Kate Seeman for their input in the earlier stage of this work. We would also like to sincerely thank Samuel Potter and Stela Ciko for voluntarily agreeing to evaluate the knowledge graph.

References

Abdulla, R. A.; Garrison, B.; Salwen, M.; Driscoll, P.; and Casey, D. 2002. The credibility of newspapers, television news, and online news. In *Education in Journalism Annual Convention, Florida USA*. Citeseer.

Banarescu, L.; Bonial, C.; Cai, S.; Georgescu, M.; Griffitt, K.; Hermjakob, U.; Knight, K.; Koehn, P.; Palmer, M.; and Schneider, N. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, 178–186.

Bird, S. 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, 69–72.

Biswas, S.; Mitra, P.; and Rao, K. S. 2019. Relation prediction of co-morbid diseases using knowledge graph completion. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2): 708–717.

Candan, K. S.; Liu, H.; and Suvarna, R. 2001. Resource description framework: metadata and its applications. *ACM SIGKDD Explorations Newsletter*, 3(1): 6–19.

Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka, E. R.; and Mitchell, T. M. 2010. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI conference on artificial intelligence*.

Chen, P.; Lu, Y.; Zheng, V. W.; Chen, X.; and Yang, B. 2018. Knowedu: A system to construct knowledge graph for education. *Ieee Access*, 6: 31553–31563.

Dong, X. L. 2018. Challenges and innovations in building a product knowledge graph. In *Proceedings of the 24th ACM SIGKDD International conference on knowledge discovery & data mining*, 2869–2869.

Dunlap, R. E.; McCright, A. M.; et al. 2011. Organized climate change denial. *The Oxford handbook of climate change and society*, 1: 144–160.

Etzioni, O.; Banko, M.; Soderland, S.; and Weld, D. S. 2008. Open information extraction from the web. *Communications of the ACM*, 51(12): 68–74.

Fabian, M.; Gjergji, K.; Gerhard, W.; et al. 2007. Yago: A core of semantic knowledge unifying wordnet and wikipedia. In *16th International world wide web conference, WWW*, 697–706.

Gallagher, B. 2006. Matching Structure and Semantics: A Survey on Graph-Based Pattern Matching. In *AAAI Fall Symposium: Capturing and Using Patterns for Evidence Detection*, volume 45.

Gliozzo, A.; Biran, O.; Patwardhan, S.; and McKeown, K. 2013. Semantic technologies in IBM Watson. In *Proceedings of the Fourth Workshop on Teaching NLP and CL*, 85–92.

Goodman, M. W. 2020. Penman: An Open-Source Library and Tool for AMR Graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 312–319. Online: Association for Computational Linguistics.

Groth, P.; Lauruhn, M.; Scerri, A.; and Daniel Jr, R. 2018. Open information extraction on scientific text: An evaluation. *arXiv preprint arXiv:1802.05574*.

Hagberg, A.; Swart, P.; and S Chult, D. 2008. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Hoffart, J.; Suchanek, F. M.; Berberich, K.; and Weikum, G. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial intelligence*, 194: 28–61.

Islam, M. S.; Mahbub, A.; Wohn, C.; Berger, K.; Uong, S.; Kumar, V.; Korfmacher, K. S.; and Hoque, E. 2022. SEER: Sustainable E-commerce with Environmental-impact Rating. *arXiv preprint arXiv:2209.06156*.

Joshi, Y.; and Rahman, Z. 2015. Factors affecting green purchase behaviour and future research directions. *International Strategic management review*, 3(1-2): 128–143.

Kingsbury, P. R.; and Palmer, M. 2002. From TreeBank to PropBank. In *LREC*, 1989–1993. Citeseer.

Kiousis, S. 2001. Public trust or mistrust? Perceptions of media credibility in the information age. *Mass communication & society*, 4(4): 381–403.

Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; and Dyer, C. 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 260–270. San Diego, California: Association for Computational Linguistics.

Lo, K.; Wang, L. L.; Neumann, M.; Kinney, R.; and Weld, D. 2020. S2ORC: The Semantic Scholar Open Research Corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4969–4983. Online: Association for Computational Linguistics.

Manning, E.; Wein, S.; and Schneider, N. 2020. A human evaluation of AMR-to-English generation systems. In *Proceedings of the 28th International Conference on Computational Linguistics*, 4773–4786.

Maynard, D.; Bontcheva, K.; and Augenstein, I. 2016. Natural language processing for the semantic web. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 6(2): 1–194.

Metzger, M. J.; and Flanagin, A. J. 2013. Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of pragmatics*, 59: 210–220.

Metzger, M. J.; Flanagin, A. J.; and Medders, R. B. 2010. Social and heuristic approaches to credibility evaluation online. *Journal of communication*, 60(3): 413–439.

Miller, G. A. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11): 39–41.

Mintz, M.; Bills, S.; Snow, R.; and Jurafsky, D. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 1003–1011.

Mishra, P.; and Mittal, R. 2021. NeuralNERE: Neural Named Entity Relationship Extraction for End-to-End Climate Change Knowledge Graph Construction. In *Tackling Climate Change with Machine Learning Workshop at ICML*.

Mitchell, T.; Cohen, W.; Hruschka, E.; Talukdar, P.; Yang, B.; Betteridge, J.; Carlson, A.; Dalvi, B.; Gardner, M.; Kisiel, B.; et al. 2018. Never-ending learning. *Communications of the ACM*, 61(5): 103–115.

Nadeau, D.; and Sekine, S. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1): 3–26.

Noy, N. F.; McGuinness, D. L.; et al. 2001. Ontology development 101: A guide to creating your first ontology.

Pradhan, S. S.; Hovy, E.; Marcus, M.; Palmer, M.; Ramshaw, L.; and Weischedel, R. 2007. Ontonotes: A unified relational semantic representation. In *International Conference on Semantic Computing (ICSC 2007)*, 517–526. IEEE.

Roopak, N.; and Deepak, G. 2021. KnowGen: a knowledge generation approach for tag recommendation using ontology and honey bee algorithm. In *European, Asian, Middle Eastern, North African Conference on Management & Information Systems*, 345–357. Springer.

Sang, S.; Yang, Z.; Liu, X.; Wang, L.; Lin, H.; Wang, J.; and Dumontier, M. 2018. GrEDeL: A knowledge graph embedding based method for drug discovery from biomedical literatures. *Ieee Access*, 7: 8404–8415.

Sattar, A. A.; Jackson, S. K.; and Bradley, G. 2014. The potential of lipopolysaccharide as a real-time biomarker of bacterial contamination in marine bathing water. *Journal of water and health*, 12(1): 105–112.

Stanovsky, G.; Michael, J.; Zettlemoyer, L.; and Dagan, I. 2018. Supervised open information extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 885–895.

Sukthanker, R.; Poria, S.; Cambria, E.; and Thirunavukarasu, R. 2020. Anaphora and coreference resolution: A review. *Information Fusion*, 59: 139–162.

UzZaman, N.; and Allen, J. 2010. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, 276–283.

Vedula, N.; and Parthasarathy, S. 2021. Face-keg: Fact checking explained using knowledge graphs. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 526–534.

Vrandečić, D.; and Krötzsch, M. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10): 78–85.

Zeng, X.; Tu, X.; Liu, Y.; Fu, X.; and Su, Y. 2022. Toward better drug discovery with knowledge graph. *Current opinion in structural biology*, 72: 114–126.

Zou, X. 2020. A survey on application of knowledge graph. In *Journal of Physics: Conference Series*, volume 1487, 012016. IOP Publishing.