DLSU RESEARCH CONGRESS 2023
MANILA, PHILIPPINES
JULY 5-7, 2023

Fostering a Humane and Green Future:
Pathways to Inclusive Societies and Sustainable Development

# Adding a Temporal Aspect to the Ontology Design of an Ontology Population System for Climate Change and Its Impact on Marine Life

Marvin Eljhune B. Lam, Anthony Philip F. Nocom Jr, Rinzai Dharma T. Terrado, and Nathalie Rose Lim-Cheng*

*De La Salle University*

**Corresponding Author: nathalie.lim@dlsu.edu.ph*

**Abstract:** Climate change refers to drastic shifts in temperature and weather. Different societies, habitats, and living organisms are forced to adapt to its irreversible effects, specifically the marine ecosystems. Various studies provide insight into the different effects of climate change and the processes they affect. Existing systems provide a structured collection of studies on the different processes through the use of ontologies. An existing study by Chua et al. (2022) provides insight into the population of ontologies about climate change and its impact on marine organisms. The ontology used by Chua et al. (2022) is extended by adding a temporal aspect. Adding a temporal aspect on the ontology design allows the viewing of data about climate change and its impact on marine life over the years easier. In this paper, we discuss the improved ontology design, extraction of temporal aspect data via bootstrapping, and results from the bootstrapping experiments. Eight (8) papers were used in the bootstrapping experiment and the results were compared with the same set of papers but manually annotated. Initial results show a low number of retrieved data as compared to referenced data. Improvements to the extraction via bootstrapping can be made by experimenting with more articles and/or standardizing date formats in the preprocessing stage.

**Key Words:** Climate Change; Ontology; Marine life; Bootstrapping; Temporal Aspect

## 1. INTRODUCTION

Climate change refers to drastic shifts in temperature and weather. To name a few effects of climate change, temperatures are hotter, glaciers are melting which results in rising sea levels, and storms are more severe. (United Nations, 2022) With these effects, different societies, habitats, and living organisms are forced to adapt to the situation because the impact of its effects are irreversible. One of the affected organisms are marine organisms.

Studies on climate change effects on marine organisms have been conducted through the years. Pankhurst and Munday (2011) conducted a study on the effects of climate change on fish reproduction and early life history stages. The study mentioned that seasonal changes in temperature have a significant effect on reproduction in fish. The study also tackled ocean acidification due to an increase in atmospheric carbon dioxide. Other studies in the same domain, such as Doney et al. (2012) indicate that rising atmospheric carbon dioxide has substantial effects on the physiology of an organism, population of an individual species, and community composition or biodiversity.

Correlating the findings and results of different studies about the effects of climate change to marine life would be difficult due to its volume. An effective way to correlate these studies is to use ontologies. An ontology defines a common vocabulary for researchers who need to share information in a domain. Ontologies also include machine-interpretable definitions of basic concepts in the domain and the relations among them (Noy & Mcguinness, 2001).

An existing ontology system design (Chua et al., 2022) correlates the effects of climate change to the habitat and the organisms living in it. The existing ontology system could potentially benefit scientists or other researchers that are studying the behavioral and physical changes of different marine species due to climate change.

An improvement to the existing ontology system design is done by adding a temporal aspect to its design. With a temporal aspect added, experts can view climate change data over the years and make predictions and analysis easier.

In Section 2, we discuss the design and the purpose of the existing ontology system made by Chua, Legaspi and Lim-Cheng (2022). In Section 3, the extended ontology design is discussed and how the temporal aspect is used. In Section 4, extracting the data using the bootstrapping method is discussed. In Section 5, the results of the experiments are discussed.

## 2. EXISTING ONTOLOGY POPULATION SYSTEM

The existing ontology design by Chua et. al (2022), seen in Figure 1, consists of 10 classes which are Habitat, Organism, Common Name, Location, Genus, Family, Effect, EffectValue, AffectedbyFactor, and FactorValue. The design also contains 9 relationships which are isLocatedIn, livesin, isAffectedby, belongsToGenus, hasCommonName, belongsToFamily, causedBy, effectValueis, and factorValueis. (Chua, et al., 2022)

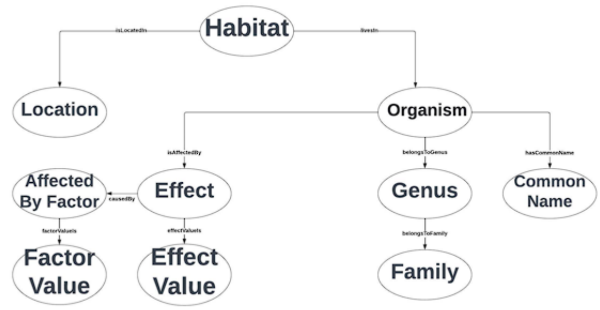



Fig. 1.Ontology Design by Chua, Legaspi, and Lim-Cheng (2022)

The purpose of the previous design is to correlate the effects of climate change to the habitat and organisms that exist within it, and how it affects the organism within the said habitat.

The ontology population system can extract important information from scientific publications using a bootstrapping algorithm. The bootstrapping algorithm will be used to scan through data, extract information, and find relationships between those pieces of information. The algorithm also finds possible patterns to how the extraction of information from the papers will occur, increasing the likelihood of extracting relevant information—in the subsequent runs—if the sentence patterns are similar.

One of the limitations of the existing design is that the correlation of effects of climate change on marine life does not take into account the changes in data over time. To alleviate this, a temporal aspect is added to the design. The temporal aspect is in the form of tagging the dates/durations in which effects of climate change were observed.

## 3. ONTOLOGY DESIGN

As seen in Figure 2, in the extended ontology design, 3 classes were added which are TimeStampObserved, StartTimeStampObserved, and EndTimeStampObserved. 3 relationships were also added. These relationships are hasTimeStampObserved, hasStartTimeStampObserved, and hasEndTimeStampObserved. The purpose of the extended design is to help the experts correlate the effects of climate change on marine organisms over the years. These classes and relationships were added so that the experts can view the data on climate change and its impacts over the years. An example of viewing data over the years is—for example—extracting information about ocean acidification in a particular location or habitat from a scientific publication from 1980 vs one from 2020. Marine experts will then be able to view this data over time and more easily analyze or predict trends in data on different marine organisms or habitats.

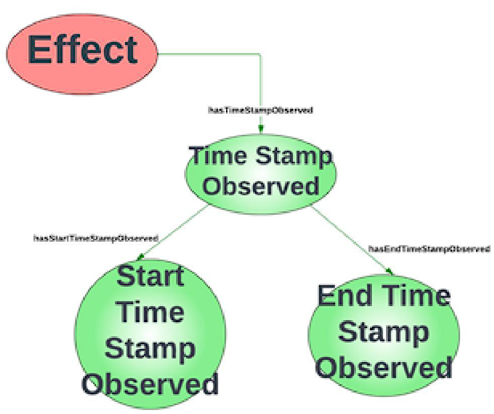



Fig. 2. Extended Ontology Design

TimestampObserved class represents the set of dates and/or duration of the observed effects of climate change. This class contains 2 subclasses namely StartTimeStampObserved and EndTimeStampObserved. The StartTimeStampObserved subclass represents the date of the beginning of the observation, while the EndTimeStampObserved subclass represents the date of the end of the observation. Data is retrieved from the ontology and displayed through the user interface for the user to view and analyze.

An example of hasTimeStampObserved is high atmospheric pressure from May to August. The seed tuple is formatted as follows: `hasTimeStampObserved (Effect : TimeStampObserved)`. In this case, the relation is `hasTimeStampObserved(high atmospheric pressure : May to August)`. Seed tuples are two entities that have a relationship between each other and they are used by the bootstrapping algorithm to extract seed patterns and other seed tuples.

To give a better understanding of the design, the entries to be stored for the marine organism "Mauve Stinger" can be seen in Figure 3.

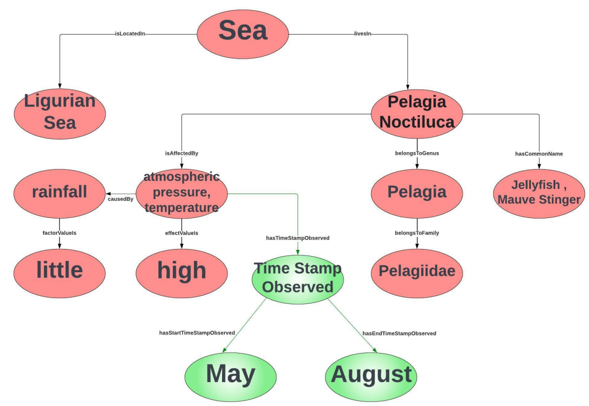



Fig. 3. Sample Entry of Marine Climate Change Ontology

## 4. EXTRACTING THROUGH BOOTSTRAP

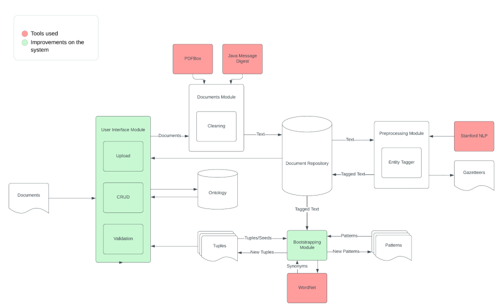


Fig. 4. Updated Ontology Population System Architecture

The implementation of information extraction through the use of bootstrapping is performed using a modified system architecture originally by Chua et. al (2022). The modified system architecture can be seen in Figure 4. The flow is as follows: the document is uploaded from the user interface module, then cleaned in the documents module. Afterward, it is processed and tagged within the preprocessing module before being sent to the bootstrapping module. In the bootstrapping module, entities are retrieved and validated by the user.

### *4.1 User Interface Module*

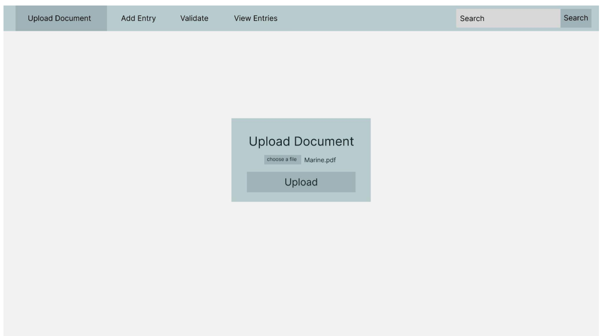
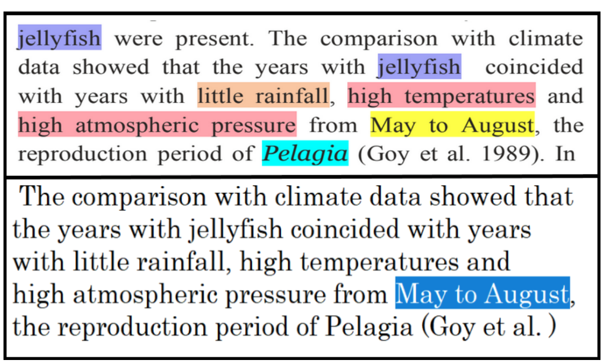



Fig. 5. User Interface for Document Uploading

The user interface module allows the user to interact with the system by providing a graphical interface with the use of a web application. Aside from handling the uploading of documents, this is also where users can validate the entries extracted before adding to the ontology. The interface for uploading documents can be seen in Figure 5.

### *4.2 Documents Module*



jellyfish were present. The comparison with climate data showed that the years with jellyfish coincided with years with little rainfall, high temperatures and high atmospheric pressure from May to August, the reproduction period of *Pelagia* (Goy et al. 1989). In

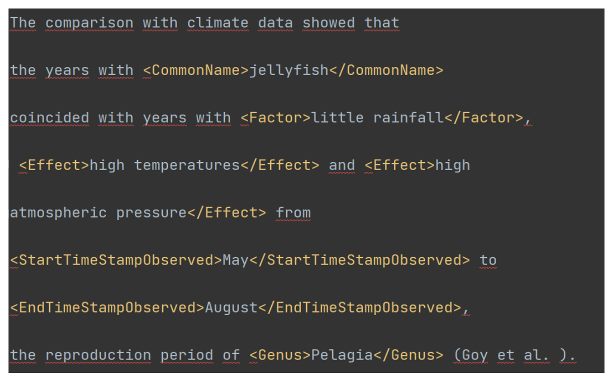
The comparison with climate data showed that the years with jellyfish coincided with years with little rainfall, high temperatures and high atmospheric pressure from May to August, the reproduction period of Pelagia (Goy et al. )

Fig. 6. Temporal Aspect in Sample Document after Text Cleaning and Conversion

The documents module handles the document cleaning and conversion. The sample document undergoes PDF-to-text conversion and text cleaning. PDF-to-text conversion is done through a Java library called Apache PDFBox, and text cleaning is done through regular expressions or RegEx. In text cleaning, unnecessary characters, white spaces, and new lines are removed.

Highlighted in the upper portion of Figure 6 are the entities that are manually tagged to be extracted through bootstrapping later on. The sample timestamp “May to August” is highlighted in yellow. The bottom portion of Figure 6 shows the sample document after it has been processed by the documents module.

### *4.3 Preprocessing Module*



```
The comparison with climate data showed that
the years with <CommonName>jellyfish</CommonName>
coincided with years with <Factor>little rainfall</Factor>,
 <Effect>high temperatures</Effect> and <Effect>high
atmospheric pressure</Effect> from
<StartTimeStampObserved>May</StartTimeStampObserved> to
<EndTimeStampObserved>August</EndTimeStampObserved>,
the reproduction period of <Genus>Pelagia</Genus> (Goy et al. ).
```

Fig. 7. Sample Document after Preprocessing and Tagging

The preprocessing module tags the entities inside the document using gazetteers (list of entities) and Stanford Named Entity Recognizer (Stanford NER).

The preprocessing module’s entity tagger (seen in Figure 4: System Architecture) is updated to accommodate tagging timestamp entities (TimeStampObserved). The other tags are based on the classes from the marine climate change ontology design. An excerpt of the sample document after going through the preprocessing module can be seen in Figure 7.

The tags help the bootstrapping process identify entities and retrieve their seed pattern/relation.

### *4.4 Bootstrapping Module*

```
▼<Seed>
   <Tag1>Effect</Tag1>
   <Tag2>TimeStampObserved</Tag2>
  ▼<Relation>
     <Pattern>beginning of</Pattern>
     <Pattern>between</Pattern>
     <Pattern>during</Pattern>
     <Pattern>end of</Pattern>
     <Pattern>from</Pattern>
     <Pattern>in the summers of</Pattern>
     <Pattern>in the winter of</Pattern>
     <Pattern>start of</Pattern>
     <Pattern>to</Pattern>
   </Relation>
 </Seed>
```

Fig. 8. Sample Seed Pattern Generated

The bootstrapping module generates seed patterns by extracting phrases between two tagged entities. It also uses synonyms of entities from WordNet so that syntactically different phrases and words that have the same meaning and other related words will not be missed by the system.

The patterns retrieved from two entities can be seen in Figure 8. These patterns are common phrases between two tagged entities (Effect and TimeStampObserved entities). Later on, the patterns will help retrieve entities to be validated by the user. The patterns are stored and used to generate new patterns for the next run.

## 5. RESULTS AND DISCUSSION

Eight (8) papers are used in the bootstrapping process to evaluate the system's temporal aspect. These papers are manually annotated by the proponents following the updated annotation guidelines based on the research by Chua et. al (2022). Annotation guidelines are a guide created for researchers to manually identify and annotate entities and compare them to the system's results. The guidelines were updated to accommodate the temporal aspect. The first three (3) papers are used to initialize the seed data. The bootstrapping experiment runs the algorithm once, then the results from the bootstrapping are compared to the set of manually annotated papers using computation metrics such as: accuracy, precision, recall, and F-measure.

True positives (TP), false positives (FP), and false negatives (FN) were used for computing the metric scores.

- True Positive (TP) - The total number of elements that actually belong to the positive class (correctly extracted information).
- False Positive (FP) - The total number of elements that does not belong to the positive class (information that should not be extracted).
- False Negative (FN) - The total number of elements that actually belong to the positive class but was not extracted.

Table. 1. Results per Article

| Articles | Referenced | Retrieved | TP | FP | FN | Accuracy | Precision | Recall | F-Measure |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 12 | 10 | 5 | 5 | 7 | 0.29 | 0.5 | 0.42 | 0.46 |
| 2 | 5 | 3 | 1 | 2 | 4 | 0.14 | 0.33 | 0.2 | 0.25 |
| 3 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 4 | 6 | 4 | 3 | 1 | 3 | 0.43 | 0.75 | 0.5 | 0.6 |
| 5 | 2 | 0 | 0 | 0 | 2 | 0 | NaN | 0 | NaN |
| 6 | 2 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | NaN |
| 7 | 2 | 1 | 1 | 0 | 1 | 0.5 | 1 | 0.5 | 0.67 |
| 8 | 2 | 2 | 2 | 0 | 0 | 1 | 1 | 1 | 1 |
| Total/Average | 32 | 22 | 13 | 9 | 19 | 0.42 | 0.65 | 0.45 | 0.66 |

The "retrieved data" in Table 1 pertains to the results of the bootstrapping experiment, while the "referenced data" pertains to the manual annotations by the proponents.

Extracted entity relations (pertaining to TimeStampObserved classes only) can be seen in Table 1. The number of correct data retrieved (TP) are 59.09%, while the number of incorrect data retrieved (FP) by the bootstrapping process are 40.91% of the 22 entries.

Overall, the system retrieves a low number of dates in non-standardized formats, and retrieves most dates in the common date formats such as: "from May to August" or "December 13 - 16". Non-standardized date formats include: "in the summers of '', or "end of spring".

Since the journals and authors do not have a particular standard for how the dates and durations should be written in their papers, this leads to fewer patterns being retrieved, and subsequently, fewer date entities extracted. Additionally, if phrases contain seasons (ex. summer of 1991), the year is extracted but it does not include the season.

Aside from only using eight (8) articles, there is also a low-to-medium amount of date entities that appear inside the articles themselves, totalling only 32 date entities for eight articles. These factors lead to a lower number of retrieved data and true positives.

## 6. CONCLUSION AND FUTURE WORK

An addition of a temporal aspect in an existing ontology population system has been made. The ontology population system can extract dates and durations of observed climate change effects from scientific publications. However, the low number of date entities and occurrences of non-standardized date formats lead to a lower number of true positives.

Improvements to the extraction via bootstrapping can be made by experimenting with more articles and/or standardizing date formats in the preprocessing stage. Future work can include experimenting with other clauses aside from verb phrases only. Additionally, adding the necessary front-end features in the future will allow experts to search for climate change effects observed during different time periods.

## 7. REFERENCES

Chua, J., Legaspi, J., and Lim-Cheng, N.R. (2022). Populating an Ontology for Climate Change and its Impact on Marine Life. De La Salle University.

Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., . . . Talley, L. D. (2012). Climate change impacts on marine ecosystems. Annual Review of Marine Science, 4 (1), 11-37.

Noy, N., & Mcguinness, D. (2001). Ontology development 101: A guide to creating your first ontology. Stanford University. Retrieved from https://protege.stanford.edu/publications/ontologydevelopment/ontology101-noy-mcguinness.html

Pankhurst, N. W., & Munday, P. L. (2011). Effects of climate change on fish reproduction and early life history stages. Marine and Freshwater Research, 62, 1015-1026.

United Nations. (2022). Causes and effects of climate change. United Nations. Retrieved from https://www.un.org/en/climatechange/science/causes-effects-climate-change