Automated Climate Analyses Using Knowledge Graph

Jiantao Wu\textsuperscript{1,2} Huan Chen\textsuperscript{2} Fabrizio Orlandi\textsuperscript{1,3} Yee Hui Lee\textsuperscript{4} Declan O’Sullivan\textsuperscript{1,3} Soumyabrata Dev\textsuperscript{1,3}

\textsuperscript{1}The ADAPT SFI Research Centre, Ireland
\textsuperscript{2}School of Computer Science, University College Dublin
\textsuperscript{3}School of Computer Science and Statistics, Trinity College Dublin, Ireland
\textsuperscript{4}School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore

November 8, 2021, AP-S/URSI
1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4. Future Work
Outline

1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4. Future Work
Motivation

- The FAIR (Findable, Accessible, Interoperable, Reusable) data principles are fundamental for the current digital ecosystem. However,
  - extensive climate data are available on the web or other medium as separate data pieces in various formats\(^1\), e.g. CSV, JSON, RDB...

- While modern knowledge graphs (KG) have effectively contributed to the establishment of FAIR data in a variety of domains, including Wikipedia and Facebook’s social networks, they are seldom built for the climate domain.

---

Objective and approaches of the study

- Facilitating the creation of “FAIR”-oriented data in the climate domain
- Creating a climate knowledge graph that
  - provides semantic context for climate data by ontology modeling
  - presents climate data in RDF data framework which is more interoperable and becomes Linked Data as others link to it
  - allows users to use semantic query language to fetch the climate data and perform climate analysis
Outline

1 Objective & Approaches

2 Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3 Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4 Future Work
Table of Contents

1 Objective & Approaches

2 Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3 Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4 Future Work
All begin with the definitions

CA ontology[^1] defines Classes, Properties, Relations, etc. for climate observations using RDF data structure (i.e. the triples).

We build the climate knowledge graph model on NOAA climate data, and reuse vocabularies from other ontologies. Here are some examples:

- `<ca: TemperatureObservation > < rdf : type > < rdfs : Class >`
- `<ca: TemperatureObservation > < rdfs : subClassOf > < sosa : Observation >`
- `<ca: isLocatedIn > < rdf : type > < rdf : Property >`
- `...`

# Table of Contents

1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4. Future Work
Apache Jena Fuseki

- Our knowledge graph is stored using Apache Fuseki.
- Fuseki expose our data as open data on the web. The data is retrievable and resolvable via URIs for more facts over HTTP.
- The query language to the triplestore is the standard SPARQL.

Figure: A screenshot of our configured fuseki server
Outline

1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4. Future Work
Table of Contents

1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
     - Case 1: Variation of monthly temperature
     - Case 2: Distribution of weather type

4. Future Work
SAPRQL examples retrieving KG’s climate data

Listing 1: SPARQL query to obtain the average temperature in Dublin from 1980 to 2019

```sparql
BASE <http://jresearch.ucd.ie/climate-kg/>
PREFIX ca_prop:<http://jresearch.ucd.ie/climate-kg/ca/property/>
PREFIX sosa:<http://www.w3.org/ns/sosa/>
PREFIX qudt:<http://qudt.org/1.1/schema/qudt#>
SELECT ?tavg ?date
WHERE {
  ?obs ca_prop:sourceStation <resource/station/GHCND:EI000003969> ;
  sosa:hasResult ?res ;
  sosa:resultTime ?date .
  ?res ca_prop:withDataType <resource/datatype/TAVG> ;
  qudt:numERICValue ?tavg .
  filter(year(?date)>1979 && year(?date)<2020)
}
ORDER BY DESC ( ?date )
```

Data can be queried dynamically on demand in instead of fixed data dumps, for example, a big CSV file.

- PREFIX specifies the URI’s name space and is often followed by a semantic word. For instance, in the listing, “sourceStation”.

Listing 2: SPARQL query to gain the weather type data in Sculthorpe England from 1951 to 1963

```sparql
BASE <http://jresearch.ucd.ie/climate-kg/>
PREFIX ca_prop:<http://jresearch.ucd.ie/climate-kg/ca/property/>
PREFIX sosa:<http://www.w3.org/ns/sosa/>
PREFIX qudt:<http://qudt.org/1.1/schema/qudt#>
SELECT ?fog ?date
WHERE{
  ?obs ca_prop:sourceStation <resource/station/GHCND:UKW00035036> ;
  sosa:hasResult ?res ;
  sosa:resultTime ?date .
  ?res ca_prop:withDataType <resource/datatype/WT01> ;
  qudt:numERICValue ?fog .
  filter(year(?date)>1950 && year(?date)<1964).
}
```
Table of Contents

1 Objective & Approaches

2 Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3 Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4 Future Work
Case Study  
Case 1: Variation of monthly temperature

Monthly temperature variation in Dublin (IE) and Manston (UK)

- Each packet contains monthly statistics for 40 years daily temperature records.
- The temperatures in Dublin and Manston reached their highest levels in July and are close in the winter from November to January but in other seasons temperature in Manston is generally a bit higher than Dublin.

Figure: Monthly temperature in Dublin and Manston from 1980 to 2019
Table of Contents

1. Objective & Approaches
2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool
3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type
4. Future Work
Case 2: Distribution of weather type

Weather type distribution in Sculthorpe, UK

- weather type data (binary data) is visualized using strip plot
- clear to see raining is very frequent in Sculthorpe

**Figure:** Distribution of days by weather type in Sculthorpe, UK
Outline

1. Objective & Approaches

2. Methodology
   - KG Data Schema—“Climate Analysis (CA)” Ontology
   - Triple Store—The KG Publishing Tool

3. Case Study
   - Formulating the SPARQL queries
   - Case 1: Variation of monthly temperature
   - Case 2: Distribution of weather type

4. Future Work
Future Work

- A continuous extension to the types of the classes and properties in CA ontology, catering for different datasets
- After having a rich set of vocabularies, define some climate domain-specific rules to bring the CA ontology an inference ability (i.e. exploring the knowledge graph applications)
- Leveraging the power of Linked Data to a certain degree by integrating CA ontology with other published ontologies in order to enable users to have some knowledge on how climate influences other domains.
Thank You

Please feel free to reach me at email: jiantao.wu@ucdconnect.ie