Automated Climate Analyses Using Knowledge Graph

Jiantao Wu†, Huan Chen‡, Fabrizio Orlandi†, Yee Hui Lee§, Declan O’Sullivan‡, and Soumyabrata Dev‡†

† ADAPT SFI Research Centre, Dublin, Ireland
‡ School of Computer Science, University College Dublin, Ireland
§ School of Computer Science and Statistics, Trinity College Dublin, Ireland

Abstract—The FAIR (Findable, Accessible, Interoperable, Reusable) data principles are fundamental for climate researchers and all stakeholders in the current digital ecosystem. In this paper, we demonstrate how relational climate data can be "FAIR" and modeled using RDF, in line with Semantic Web technologies and our Climate Analysis ontology. Thus, heterogeneous climate data can be stored in graph databases and offered as Linked Data on the Web. As a result, climate researchers will be able to use the standard SPARQL query language to query these sources directly on the Web. In this paper, we demonstrate the usefulness of our SPARQL endpoint for automated climate analytics. We illustrate two sample use cases that establish the advantage of representing climate data as knowledge graphs.

I. INTRODUCTION

Knowledge Graphs (KG) have successfully contributed to a variety of applications, including information retrieval, chatbots [1], smart energy systems [2]. In this work, we explore how to perform weather analysis by querying our climate knowledge graph1. The data in our climate knowledge graph is modeled in RDF (Resource Description Framework) following our Climate Analysis (CA) ontology [3] and is stored in a SPARQL endpoint. The creation of climate KG provides an interoperable data structure for climate data, which makes the data easily integrated with other domain data modeled by other ontologies across federated endpoints. We use the data from the National Oceanic and Atmospheric Administration (NOAA), a reputable climate data center that provides raw datasets. There are two major classes in the CA ontology: (i) Station and (ii) Observation. A station is a site that generates climate observations. Geographical factors (longitude and latitude) or literal names can be used to identify a station. An observation is used to describe a climate feature of interest that has been detected that generated by station, the observed value, and the datatype that defines the feature of interest are typically associated with it.

III. ANALYZING CLIMATE DATA USING KG

This section showcases how to write SPARQL queries to gather climate data for study2. The first sub-section lists example SPARQL queries that are executed for producing data sources used by the climate analysis experiments in the second sub-section.

A. Data Collection

Listing 1 is the SPARQL query to gather data on the average temperature in Dublin from 1980 to 2019 and Listing 2 is to gain the weather types data in Sculthorpe, England, from 1951 to 19633.

2In the spirit of reproducible research, all the source code is available at https://github.com/futaoo/sparql-weatheranalysis.

3For this older time period, weather types (e.g. fog, haze) are available for analysis. These are categorical data for which the detailed processing procedure is given in Section III-B.
B. Use Case I: Variation of monthly temperature

To analyze the temperature in Dublin (IE) and Manston (UK) from 1980 to 2019, we construct a station-grouped box-plots based on the results (see Fig. 1). Each packet contains monthly statistics for 40 years daily temperature records. The figure shows that 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.

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 qdt: <http://purl.org/lii/schema/qdt#>

select ?year 
where { 
  ?obs ca_prop:sourceStation <resource/station/GHCND: ;
  sosa:hasResult ?res ;
  sosa:resultTime ?date .
  ?res ca_prop:withDataType <resource/datatype/TAVG ;
  qdt:numericValue ?temp .
  filter(year(?date)>1979 & & year(?date)<2020).}
ORDER BY DESC(?date)
```

C. Use Case II: Distribution of weather type

Finally, we present the strip plot (Fig. 2) to present the distribution of weather types (rain, fog, drizzle, snow, sun) on every day of the year 1951, 1955, 1959 and 1963. The weather in Sculthorpe was bad most of the time during 1950s as a result of frequent drizzle, fog, and haze, as seen in Fig. 2.

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 qdt: <http://purl.org/lii/schema/qdt#>

select ?date ?weather 
where { 
  ?obs ca_prop:sourceStation <resource/station/GHCND:UKX00035036 ;
  sosa:hasResult ?res ;
  sosa:resultTime ?date .
  ?res ca_prop:withDataType <resource/datatype/WT01 ;
  qdt:numericValue ?temp .
  filter(year(?date)>1950 & & year(?date)<1964).}
```

IV. Conclusion & Future Work

In this paper, we demonstrated how to use standards like RDF and SPARQL to obtain knowledge graph data modeled by the CA ontology for climate studies. According to the data type variety, this paper provides two different weather analyses of which the one is based on the statistics on the quantity data about temperature and the other is based on the categorical meteorological records. Both analyses provide basic understanding to the weather conditions in the corresponding cities (i.e. Dublin, Manston and Sculthorpe). In the future, we will leverage the power of semantic approaches to enhance the usability of our climate knowledge graph for climate researchers. For instance, integrating additional data sources for more advanced climate studies.

REFERENCES

[1] Y. Liang, F. Xu, S.-H. Zhang, Y.-K. Lai, and T. Mu, “Knowledge graph construction with structure and parameter learning for indoor scene design,” Computational Visual Media, vol. 4, no. 2, pp. 123–137, 2018.
[2] F. Orlandi, A. Meehan, M. Hossari, S. Dev, D. O’Sullivan, and T. AlSkaif, “Interlinking heterogeneous data for smart energy systems,” in Smart Energy Systems and Technologies (SEST). IEEE, 2019.
[3] J. Wu, F. Orlandi, D. O’Sullivan, and S. Dev, “An ontology model for climatic data analysis,” in Proc. IEEE IGARSS conference, 2021.
[4] S. Manandhar, S. Dev, Y. H. Lee, and Y. S. Meng, “On the importance of PWV in detecting precipitation,” in Proc. IEEE AP-S Symposium and USNC-URSI Radio Science Meeting. IEEE, 2018, pp. 89–90.
[5] S. W. P. Center, “National oceanic and atmospheric administration,” National Weather Service (NOAA/NWS), cited, pp. 1950–2011, 2012.
[6] W. D. Graham, J. Spruce, K. W. Ross, J. Gasser, and N. Grulke, “Growing degree vegetation production index (GDVPI): A novel and data-driven approach to delimit season cycles,” in AGU Abstracts, 2014.
[7] D. DiFranzo, A. Graves, J. S. Erickson, L. Ding, J. Michaelis, T. Lebo, E. Patton, G. T. Williams, X. Li, J. G. Zheng et al., “The web is my backend: Creating mashups with linked open government data,” in Linking government data. Springer, 2011, pp. 205–219.