An Ontology Model for Climatic Data Analysis

Jiantao Wu\textsuperscript{1,2} Fabrizio Orlandi\textsuperscript{1,3} 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

June 15, 2021, IGARSS
Outline

1. Objective & Approaches

2. Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3. Future Work
Objective & Approaches

2 Methodology
   • Data Collection
   • Ontology Modeling
   • A Case Study

3 Future Work
Many researchers tend to analyze data solely in single domain.

- A neglect of cross-domain impacts such as climate effects on urbanization

Remote sensing data are presented in different formats on the web such as CSV, JSON, RDB.

- A need to integrate them uniformly for the data access ease\(^1\).

Ontology is a popular data model with superiority in integrating heterogeneous data.

Objective of the study

- Facilitating the integration of various climate data sources
- Proposing a Knowledge Graph (KG) model that is able to fuse heterogeneous climate remote sensing observations
- Create a platform for climate remote sensing data such that the data can be easily interlinked to other datasets
Outline

1. Objective & Approaches

2. Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3. Future Work
Table of Contents

1. Objective & Approaches

2. Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3. Future Work
Choice of climate remote sensing data

Today’s remote sensing data contains various climate features including precipitation, temperature, sea level pressure, wind speeds, etc. gather records from climate stations throughout the world.

I. Issues
A. Though data are available across the Web, datasets are hosted separately by data vendors such as KNMI Climate Explorer\textsuperscript{2}, NOAA \textsuperscript{3}
B. Less attention is paid on how to effectively integrate these datasets.

II. Proposed solution
A. Defining an ontology called “Climate Analysis” converting climate data into RDF triples and republish them as online accessible Linked Data.
B. We choose NOAA climate data as a driving use-case.

\textsuperscript{2} \url{http://climexp.knmi.nl/}
\textsuperscript{3} \url{https://www.ncdc.noaa.gov/data-access}
Table of Contents

1 Objective & Approaches

2 Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3 Future Work
Core techniques used in this workflow

- **Ontology**: human-readable vocabularies to classify and describe the relations between data, *e.g.* “Dublin”, “capitalOf”, “Ireland”, every vocabulary corresponds to a real data entity.

- **RDF Triples**: a data model to construct triples (*e.g.* “Dublin”–“capitalOf”–“Ireland”) that can be stored in computers.

- **Linked Data**: It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers.

- **Triple Store**: a database to store and publish Linked Data across the Web; the compatible query language is SPARQL.
Our proposed ontology model for climate data

We model this ontology on NOAA data.
Table of Contents

1. Objective & Approaches

2. Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3. Future Work
Using SPARQL queries to analyse weather

We perform automated analysis on Shanghai’s and Dublin’s weather.

- Apart from winter, Shanghai has much higher temperature than Dublin.
- The temperature variance against the months is much greater in Shanghai.

**Figure:** Monthly temperature summarized from recent 70 years of Dublin and Shanghai
Outline

1. Objective & Approaches

2. Methodology
   - Data Collection
   - Ontology Modeling
   - A Case Study

3. 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