

KnowHax - Climate Data Challenge

Participant - Sayli Shivalkar



Creating a Climate Data Literature Ontology

1 Parse Input Files

The first step involves thoroughly parsing all input files to extract and organize the relevant data. This lays the foundation for establishing the relationships between different attributes and entities.

1

2

Establish Relationships

With the data parsed, the next step is to carefully examine the connections and dependencies between the various attributes, such as institutions, sources, experiments, activities, citations, and licenses. This relational understanding is crucial for building a comprehensive ontology.

3

Build the Ontology

The final step is to construct the climate data literature ontology, leveraging the insights gained from the previous steps. This ontology will serve as a structured, interconnected knowledge base, enabling efficient storage, retrieval, and analysis of the climate data landscape.

Identifying Used Data in Scientific Text

1 Understand Entity Labeling

The first step is to gain a deep understanding of the entity labeling mechanism, which is essential for accurately identifying the used data within the scientific text.

2 Build an Entity Recognition Model

Leveraging the SciBERT model, a specialized language model for scientific text, the next step is to build an entity recognition model that can effectively identify and label the scientific entities within the text.

3 Optimize and Evaluate Performance

To ensure the model's effectiveness, it is crucial to train and fine-tune the model over multiple hyperparameters, evaluating its performance using the SeqEval metric to achieve the best possible results.

Obtaining Publication Metadata

Understand Website Structure

The first step in this process is to thoroughly understand the structure and layout of the website hosting the publication data. This knowledge is crucial for effectively navigating and extracting the relevant information.

Parse Publication Data

With the website structure understood, the next step is to implement a robust data parsing mechanism to extract the publication data from the website. This involves identifying and extracting the various metadata fields for each publication.

Structure the Metadata

The final step is to organize the extracted metadata in a structured and standardized format, ensuring that the information is easily accessible and can be effectively utilized for further analysis and applications.