UNIVERSITY OF TRENTO

KGE 2024 - Trentino Territory & Transportation

Academic year 2024-2025

248490 - Mores Nicola, nicola.mores@studenti.unitn.it

248491 - Roccon Marco, marco.roccon@studenti.unitn.it

Purpose Definition

- **Informal purpose:** “A person wants to move in an easy and efficient way through the Trentino region using public transports and other transport services available”

- **Domain of Interest**: Transportation services
- **Spatial domain**: Trentino region and in particular the city of Trento
- **Timespan**: From September 2024 to the end of June 2025

Activities of this phase:

- Definition of **Scenarios**, **Personas** and **CQs**
- **Concept Identification** and creation of **PFSheet**
- Formalized the purpose through an **ER Model (IDEF1X Notation)**

Output of this phase:

- The **ER model**, representing the **first implementation of the final Knowledge Layer**

Information Gathering

Activities of this phase:

- **Source Identification:**
  - **Data Values Source**: OPENdata Trentino, OpenStreetMap, Trentino Trasporti Website (extremely variable)
  - **Knowledge Source**: SCHEMA.org and GTFS.org (high quality)
  - **Language Source**: Universal Knowledge Core (UKC), SCHEMA.org, and OpenStreetMap Wiki (high quality)
- **Dataset Collection & Cleaning:**
  - **Removal** of dataset, entities and properties not aligned to our formal purpose
  - **Modification** and **translation** of dataset whenever needed
  - **Conversion** of data value dataset in high-quality, **standardized format** such as CSV.

Output of this phase:

- A **set of dataset at various level**, useful for future phases
- A collection of **cleaned data value dataset** and organized in standardized formats

Language Definition

Activities of this phase:

- **Formal language sources**:
  - **Universal Knowledge Core**
  - Otherwise, **OpenStreetMap Wiki**
  - Otherwise, **handcrafted** ones
- **Formal language definition**:
  - **Formalization of every concept**, specifying for each of them a **Concept ID,** a **label,** and a **gloss**
  - Identification of eventual concepts' **synonyms**

Output of this phase:

- **Language resource file**, containing the formalization of every concept in English and Italian

Knowledge Definition

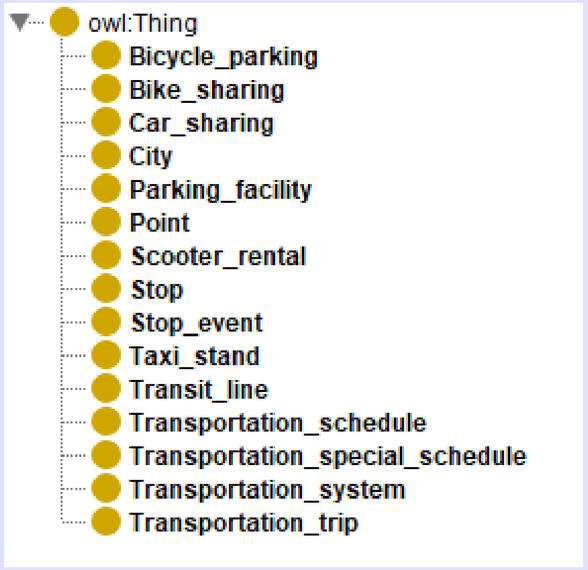
Activities of this phase:

1. **Formalization** of the **ER** in **Protégé**, creating a **Teleontology**
2. **Integration** of **external ontologies**, aligning our concepts to standard ones and introducing a hierarchical structure through IS-A relationships
3. **Simplified schema alignment**, transforming the Teleontology into a **Teleology**
    a. Identification of leaf nodes for which data availability was assured
    b. Removal of higher hierarchical ETypes, transferring their object and data properties to lower entity types.

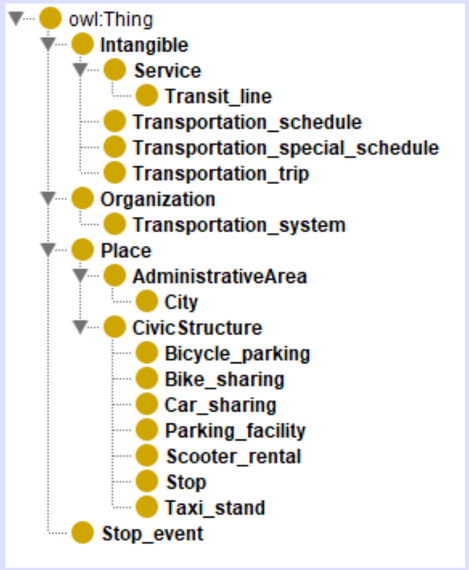
Output of this phase:

- A refined version of SCHEMA.org ontology, aligned to our domain
- A Teleontology, serving as foundation for next phases
- A structured Teleology, providing our knowledge representation

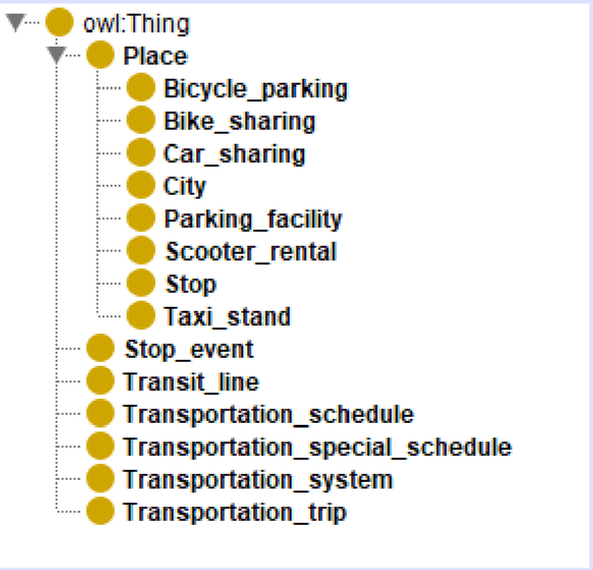
Knowledge Definition



Teleontology:
Obtained from the **ER Model**



Extended Teleontology:
Obtained from
Teleontology + Reference Ontology



Teleology:
Obtained from the
Schema Allignment applied on
the Extended Teleontology

Entity Definition

Activities of this phase:

- **Entity Matching: Merge of dataset** regarding the same ETypes
  - Urban and Extra-urban transportation datasets: Modified conflicting IDs and assure the absence of duplicate entities
  - Bike_sharing.csv and Centro_in_bici.csv: Removed eventual duplicate entities using location data as discriminant
- **Entity Identification:** Introduction of **identifiers** in datasets were this wasn't provided
- **Entity Mapping: Mapping** of dataset **to** the **Teleontology** through **KarmaLinker**

Output of this phase:

- A **Mapping Model**, linking Data and Knowledge layer
- The final **Knowledge Graph**

Knowledge Layer Evaluation

In this phase we **evaluated the result obtained** through the iTelos process, verifying the quality of our final Knowledge Graph, following the criteria defined by the iTelos methodology:

Our primary objective is to evaluate **how well the Teleontology covers** the entities and properties extracted from the initial **Competency Questions.** To this purpose, we compute:

- **EType Coverage:** Measures how well the extracted **ETypes from** a set of Competency Questions (**CQ**) **match** those defined in a **Teleontology**

$$Cov_E(CQ_E) = \frac{|CQ_E \cap T_E|}{CQ_E} = \frac{15}{15} = 1$$

Obtaining CovE = 1, **all extracted ETypes are present in the Teleontology.**

- **Property Coverage:** Evaluates how well the **properties** extracted **from Competency Questions align** with those defined in the **Teleontology**.

$$Cov_P(CQ_P) = \frac{|CQ_P \cap T_P|}{CQ_P} = \frac{19}{20} = 0.95$$

Only **95%** of the extracted properties are **covered** by the Teleontology, since we are **missing Opening Hours for Parking Facilities.**

Data Layer Evaluation

- **Connectivity Matrix:** a matrix whose (X,Y) cells assume these values:
  - If **X = Y** , the cell **(X,Y)** is equal to the number of **non-null data properties values** for all the entities mapped on the **EType X**.
  - if **X != Y** , the cell **(X,Y)** is equal to the number of **non-null object properties values** for all the object properties having the **EType X as domain and Y as range**.

	City	Place	Bicycle_parking	Bike_sharing	Car_sharing	Parking_facility	Scooter_rental	Stop	Taxi_stand	Stop_event
City	8	0	451	47	8	572	39	0	9	0
Place	0	10188	0	0	0	0	0	0	0	0
Bicycle_parking	0	0	2706	0	0	0	0	0	0	0
Bike_sharing	0	0	0	282	0	0	0	0	0	0
Car_sharing	0	0	0	0	48	0	0	0	0	0
Parking_facility	0	0	0	0	0	4576	0	0	0	0

- **Entity connectivity**: Evaluates **how much** the **different entities** in the Knowledge Graph **are interconnected**

$$EC(X) = \frac{\sum_{Y=1}^{N}(X,Y)}{OP(X)}$$

$$EC(KG) = \sum_{X=1}^{N} EC(X) = 339042.14$$

- **Property connectivity:** Measures **how well each entity** in the Knowledge Graph is **connected with its property values**

$$PC(X) = \frac{(X,X)}{DP(X)}$$

$$PC(KG) = \sum_{X=1}^{N} PC(X) = 185278$$

Exploitation

In order to demonstrate the quality of the obtained result we also concentrated on **graph interrogation**, utilizing **GraphDB** where we can upload all the .ttl files representing our Knowledge Graph

We **successfully translated all CQs** into functional SPARQL queries, with a single **exception**:

- **CQ 7.2** - *What are the opening hours of the “Piazza di Fiera” parking lot?*

We initially assumed that information on parking facility opening hours would be easily accessible but we later found that this data was missing for most entities in the dataset we gathered. Consequently, we decided to exclude this query.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

# 4.3 (Marco): On line 7, how many stops are there from “"Gorizia Adamello” to “Gocciadoro Arcate”?

select distinct ?line_name ?departure_name ?arrival_name
((xsd:integer(?arrival_sequence) - xsd:integer(?departure_sequence)) as ?n_stops) where {

    # Define the departure and arrival event for every trip and filter by direction
    ?trip etype:Contains ?departure_event, ?arrival_event .
    ?trip etype:trip_direction ?trip_direction .
    FILTER(STR(?trip_direction) = "0")

    # Select the departure stop by name
    ?departure_stop etype:name ?departure_name ;
    FILTER(STR(?departure_name) = "Gorizia Adamello")

    # Select the arrival stop by name
    ?arrival_stop etype:name ?arrival_name .
    FILTER(STR(?arrival_name) = "Gocciadoro \"Arcate\"")

    # Connect the previously defined stops with the corresponding stop events
    ?departure_stop etype:is_location_for ?departure_event .
    ?departure_event etype:sequence_number ?departure_sequence ;
                     etype:arrival_time ?departure_time .

    ?arrival_stop etype:is_location_for ?arrival_event .
    ?arrival_event etype:sequence_number ?arrival_sequence ;
                   etype:arrival_time ?arrival_time .

    # Connect the lines with trips
    ?line etype:composed_of ?trip .
    ?line etype:line_short_name ?line_name .

    # We want only the arrival stops that appen after the departure event
    FILTER(str(?arrival_sequence) > str(?departure_sequence))

}
```

Metadata Definition

To effectively share our results we will use a **dataScientia** Catalog, organizing resources based on their layers. These catalogs contains only the metadata and not the data itself, but have a crucial role for what concern the reusability, organization and discovery of resources

The **metadata produced** for this project has been organized in three different categories:

- **People** Metadata
- **Project** Metadata
- **Dataset** Metadata

In **addition to the standard metadata**, we introduced an **additional attribute** called **DatLatestModificationTimestamp**

DatURL	DatKeyword	DatPublisher	DatCreator	DatOwner
https://dati.trentino.it/dataset/trasporti-pubblici-del-trentino-formato-	Transportation System, Trento	OPENdata Trentino	Servizio Trasporti Pubblici	Provincia Autonoma di Trento
https://dati.trentino.it/dataset/taxi-open-data	Taxi, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/car-sharing-open-data	Car, Sharing Service, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/bike-sharing-open-data	Bike, Sharing Service, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/c-entro-in-bici-open-data	Bike, Sharing Service, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/rastrelliere-per-biciclette-open-data	Bike Rack, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/parcheggio-protetto-per-biciclette-ope	Bike Rack, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://dati.trentino.it/dataset/punti-sosta-monopattini-condivisi-a-ta	Scooter, Sharing Service, Trento	OPENdata Trentino	Comune di Trento	Comune di Trento
https://github.com/NikoMrs/KGE-Trentino-Territory-Transportation/b	Parking, Trento	Open Street Map	Mores Nicola, Roccon Marco	Open Street Map

Open Issues

- **Inability to fully answer** some **C**ompetency **Q**uestions:
  - Ones related to **parking lot opening hours**. Despite extensive research, we did **not find datasets** providing comprehensive information on the accessibility to parking lot during the day.
  - **Research/Acquisition of new data** on this regard could allow us to fully answer to every CQs

- Integration of **German** into the **Language Resource file**:
  - Not included so far due to **limited proficiency** in the **language**

- Other valuable **enhancements**:
  - Incorporation of data about **actual departure and arrival times** considering **delays** and **anticipations** for public transportation services.
  - Addition of data regarding **average occupancy** at different times of the day for specific routes.