Trentino Education

Samuele Bortolotti, Erich Robbi

UNIVERSITÀ
DI TRENTO

Knowledge and Graph Engineering 2022

UNIVERSITÀ
DI TRENTO

Contents

Introduction

UNIVERSITÀ DI TRENTO

Material

KNOW DIVE

The entire project material is available publicly and freely:

- GitHub repository:

  https://github.com/samuelebortolotti/Education-Trentino

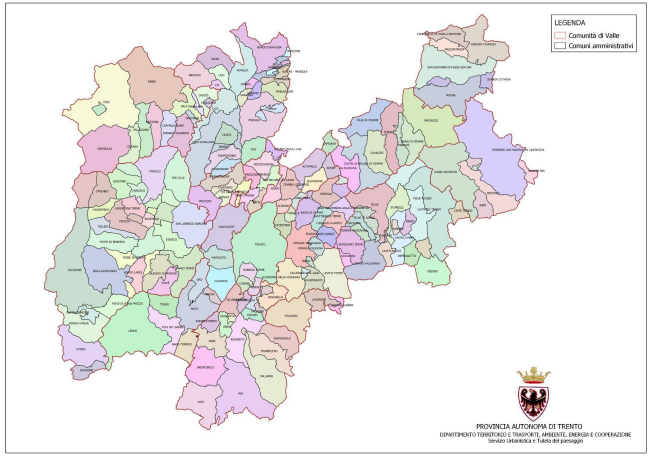
- Google Drive:

  https://drive.google.com/drive/folders/1X7iKpIhawEzTfQhQWdEJWfbMpVyZ6xl_?usp=sharing

Domain of Interest

This research seeks to provide a complete analysis of the educational facilities in Trentino which comprehend geospatial and temporal domains.

Trentino has a wide range of educational institutions, including public and private kindergartens, elementary schools, secondary schools, universities, and other vocational training facilities. The temporal domain refers to the last few years (2020 - 2022).



UNIVERSITÀ DI TRENTO

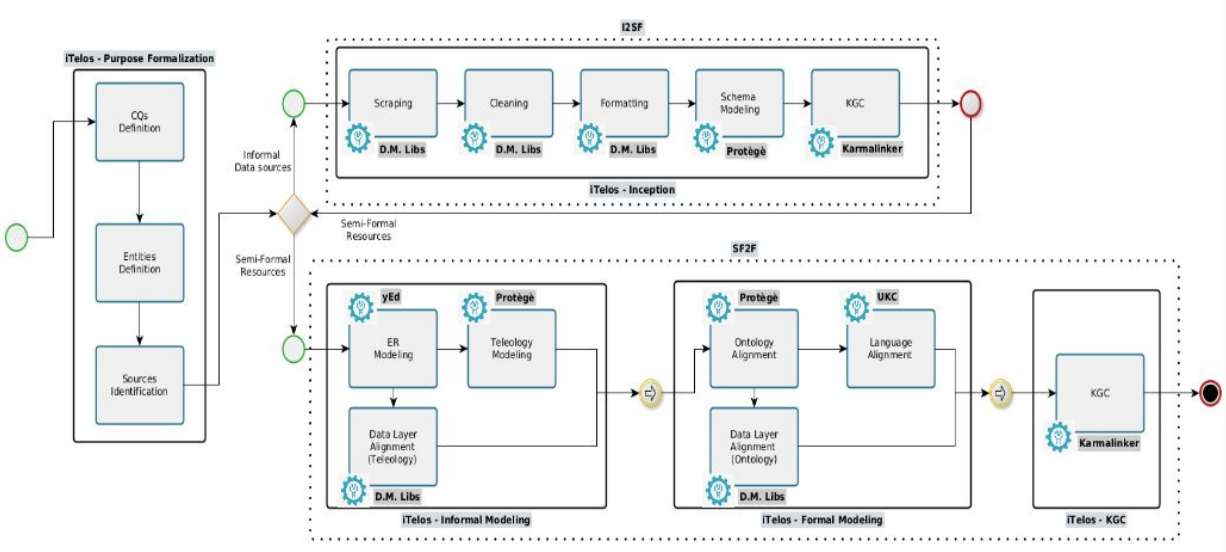
Initial Purpose

“A service that will facilitate the finding of schools, including details about the school and courses offered, in the region of Trentino based on city, commune, school type, course duration, and teaching activities schedules.”

iTelos

The process comprises the following phases:

- Purpose Formalization
- Inception
- Informal Modeling
- Formal Modeling
- KGC



Personas

11 Personas

5 Needs

- Autonomy
- Bridging
- Competence
- Curiosity
- Security

3 Main Roles

- Students
- Workers
- Parents

Personas - Example

Hermann Schmidt

Differentiator:
Hermann is a 55-year-old man from Austria and he works as a senior researcher at Stanford University, United States. His research area is mostly focused on the importance of education and how to improve it during each stage of life. His latest work is carried out in collaboration with the European Schoolnet, which is an organization headquartered in Brussels, with the aim of innovating both education and learning in Europe.

Purpose:
Hermann wants to perform a statistical analysis of all the schools located in the Italian territory.

Personas - Example

Vincente Trevisan

Differentiator:
Vincente is a 14 years old boy who lives in Margone, a small village in Trentino. He has just finished middle school and is interested in attending a high school that allows him to become a software engineer. Despite being a native of Trentino, he does not know which nearby high schools provide such a kind of education

Purpose:
Vincente wants to find a school where he can learn to be a software developer.

UNIVERSITÀ DI TRENTO

Scenarios

A scenario depicts an environment (where a person may attempt to execute a task) in narrative form.

10 Scenarios were defined.

Scenarios

We made scenarios with the intention of describing personas in different events:

- Common Events
- Accidents
- Ideal Scenarios

Scenarios - Example

Work and Study

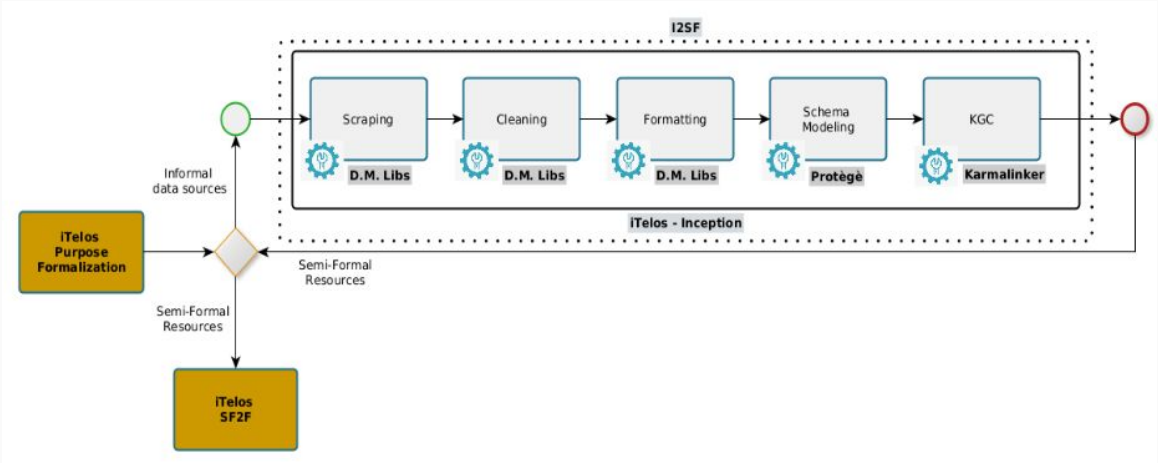
Description:
I am a student and I am looking for a school that would allow me not to quit my job. I am currently employed part-time and I cannot afford to quit my job to attend school full-time. I am looking for a school that has an online or evening program that would allow me to continue working while attending school.

Inception Phase

UNIVERSITÀ
DI TRENTO

iTelos - Inception

In the Inception phase, the goal is to take in input the user's purpose and formalize it into a set of CQs and a set of input data sets and input teleologies.



Competency Questions

With 11 Personas and 10 Scenarios, we were able to define 40 Competency Questions.

UNIVERSITÀ DI TRENTO

Competency Questions - Example

Persona

Wants to find a school where he can learn to be a software developer.

Scenario

What school provides what I need

CQ

Give me how many schools have courses on Software Engineering

Entity Types Classification

Common Entities:

- School
- Institute
- Contact Information
- Educational Institute

Entity Types Classification

Core Entities:

- Subregional Academic Division
- Institute Contact Information
- Professor

Entity Types Classification

Contextual Entities:

- Study Course
- Education Quality
- School Statistics
- Review
- Invalsi Score

Resource Collection - Schools

Concerning Schools in Trentino, the main sources of information initially were:

- Open Data Trentino
- Data from the last year’s KDI project

Resource Collection - Schools

Problems:

- Incomplete Data
- Some Resources were not accessible from Open Data (bad URL)

Resource Collection - Schools

Solution:

- To have as much data as possible, we scraped from the vivoscuola portal
- For the missing resources, we managed to retrieve the correct URL (from ISPAT) and download them

Resource Collection - UniTN

For data about the University of Trento, we relied mainly on data provided by the Digital University portal, obtaining information about:

- Professors
- University Courses

Resource Collection - Education Quality

Reviews and average scores were used to define quality of education, which was obtained by scraping the google reviews box of every school

UNIVERSITÀ DI TRENTO

Resource Collection

- 700 schools
- 270 institutes
- Study Courses (Schools and University)
- Education quality of collected Schools
- Statistics of collected Schools
- Invalsi Scores of a Municipality with Schools

UNIVERSITÀ DI TRENTO

Formalized Purpose

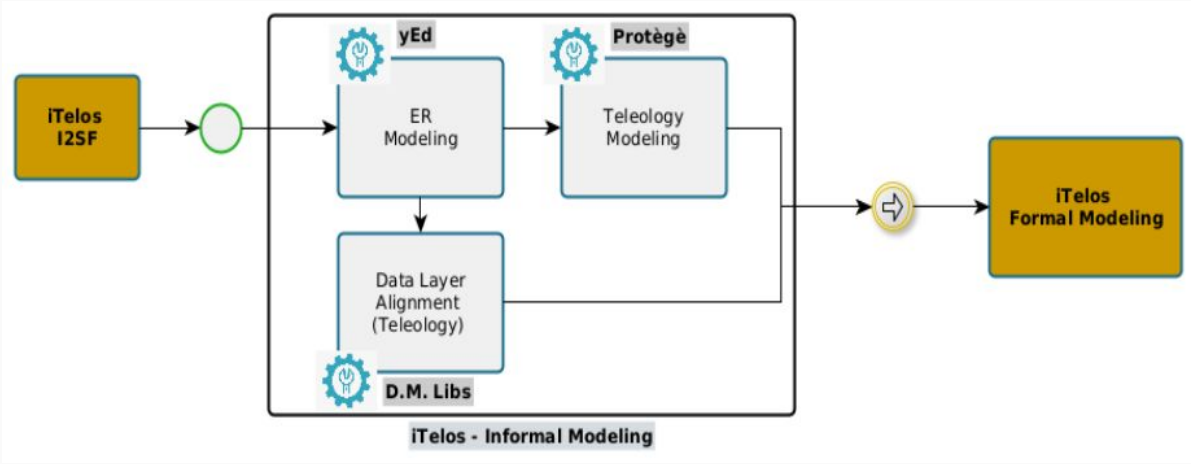
“A service that will help parents and students to find schools, including details about schools and courses offered, in the region of Trentino based on city, municipality, school type, course duration, teaching activities schedules, contact information, and education statistics.”

Informal Modeling Phase

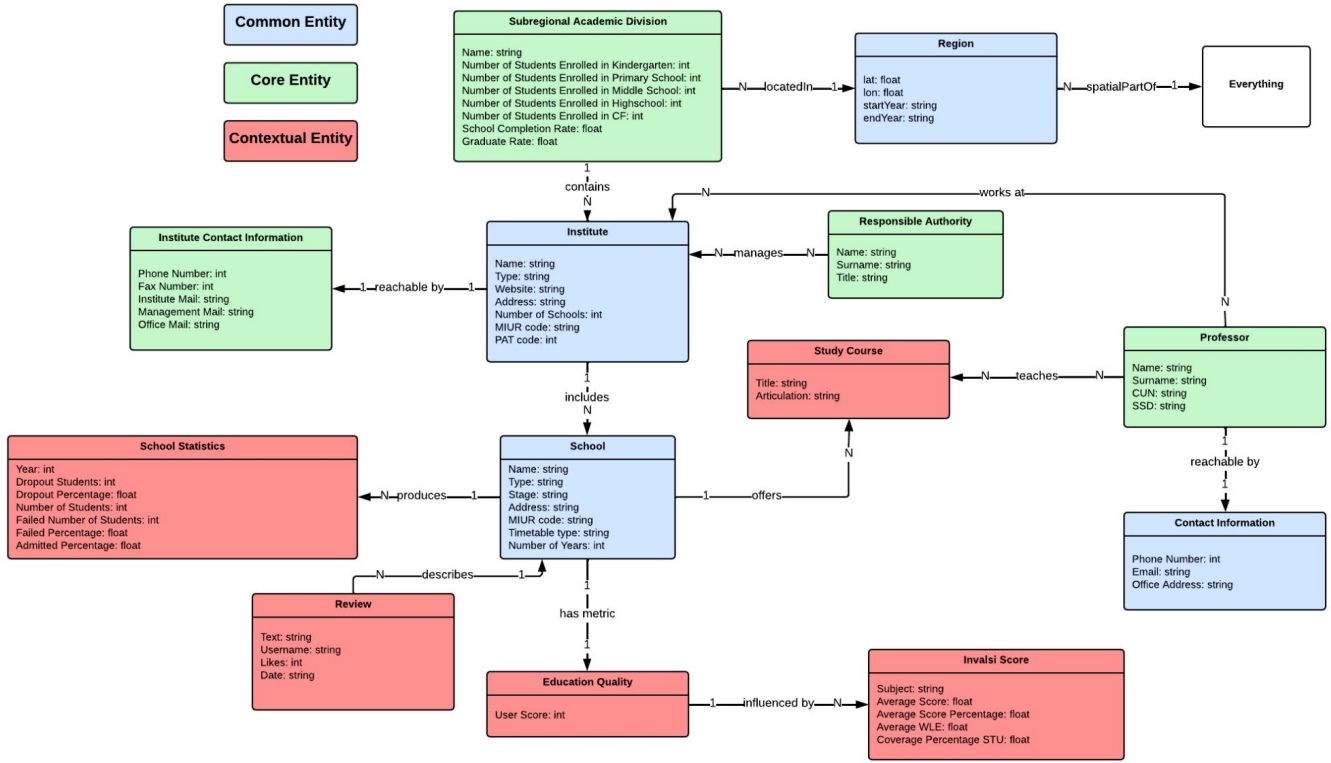
UNIVERSITÀ DI TRENTO

iTelos - Informal Modeling

In the Informal modeling phase, the goal is to take in input the semi-formal resources and the formalized purpose, and produce an ER model.



ER Model



UNIVERSITÀ DI TRENTO

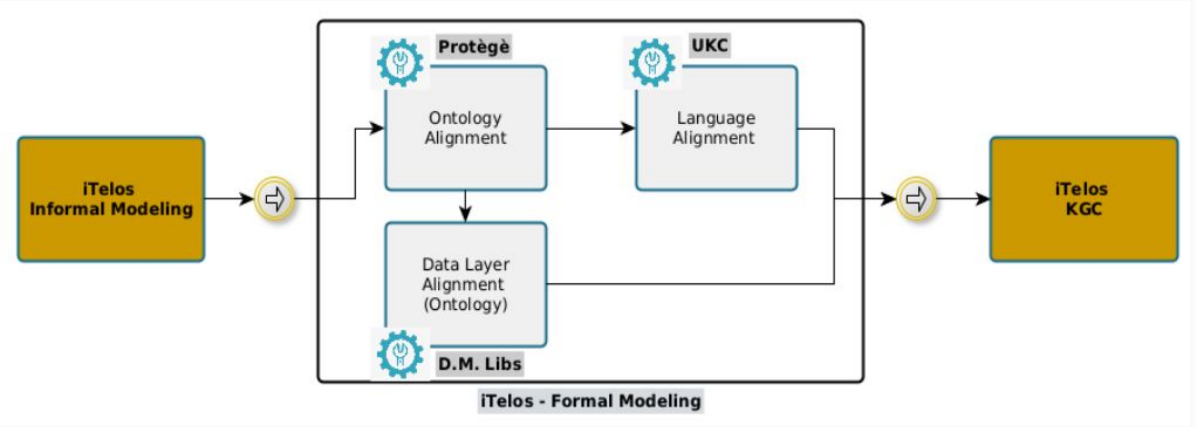
KNOW DIVE

Formal Modeling Phase

UNIVERSITÀ
DI TRENTO

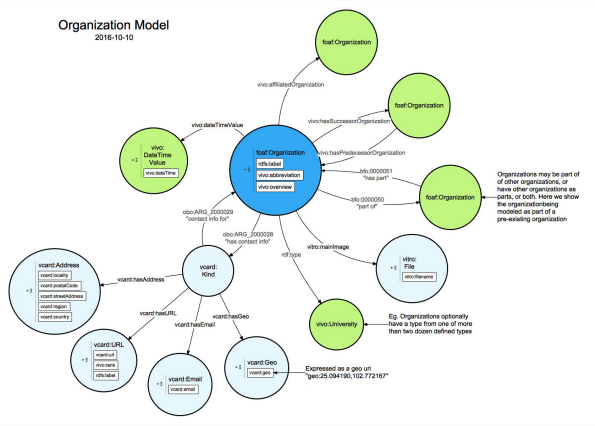
iTelos - Formal Modeling

In the formal modeling phase, the goal is to take in input the teleontology plus the semi-formal datasets and produce the formal version of the dataset aligned to the purpose plus the ETG.



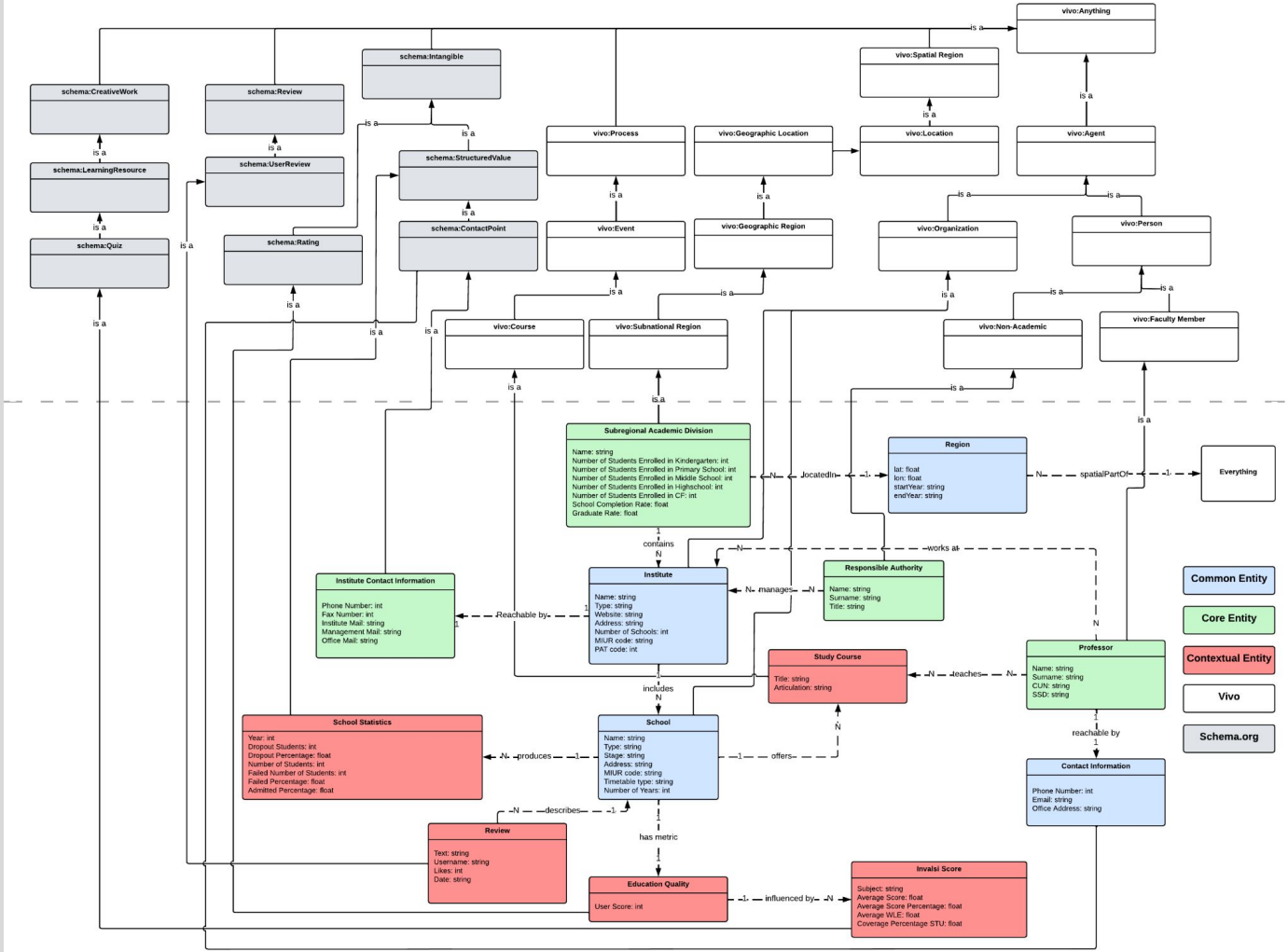
Ontology Selection

We opted to explore the **VIVO** portal in order to find a first reference teleontology, expecting to discover any specific contextual entity as we are modeling entities related to education, academia, and school and university courses. However, we still had to use **schema.org** to find additional specific entities that were previously not found



schema.org

Teleontolgy Model



UNIVERSITÀ DI TRENTO

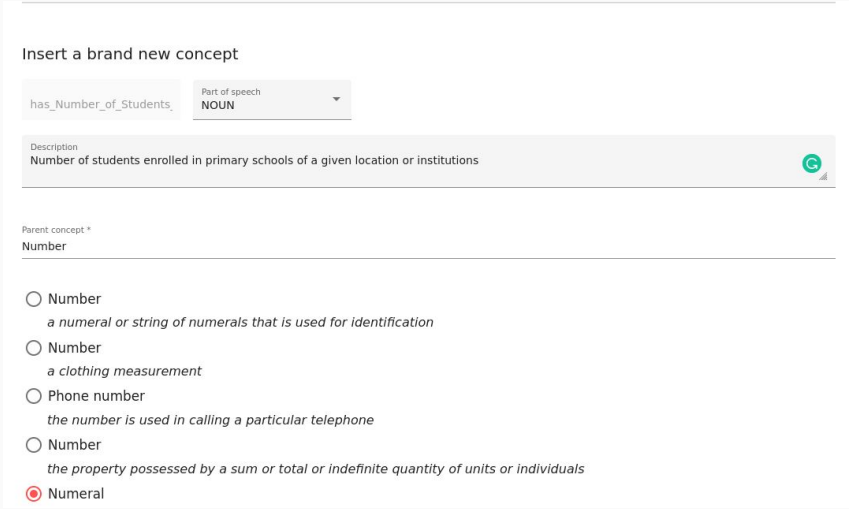
KNOW DIVE

Data Integration

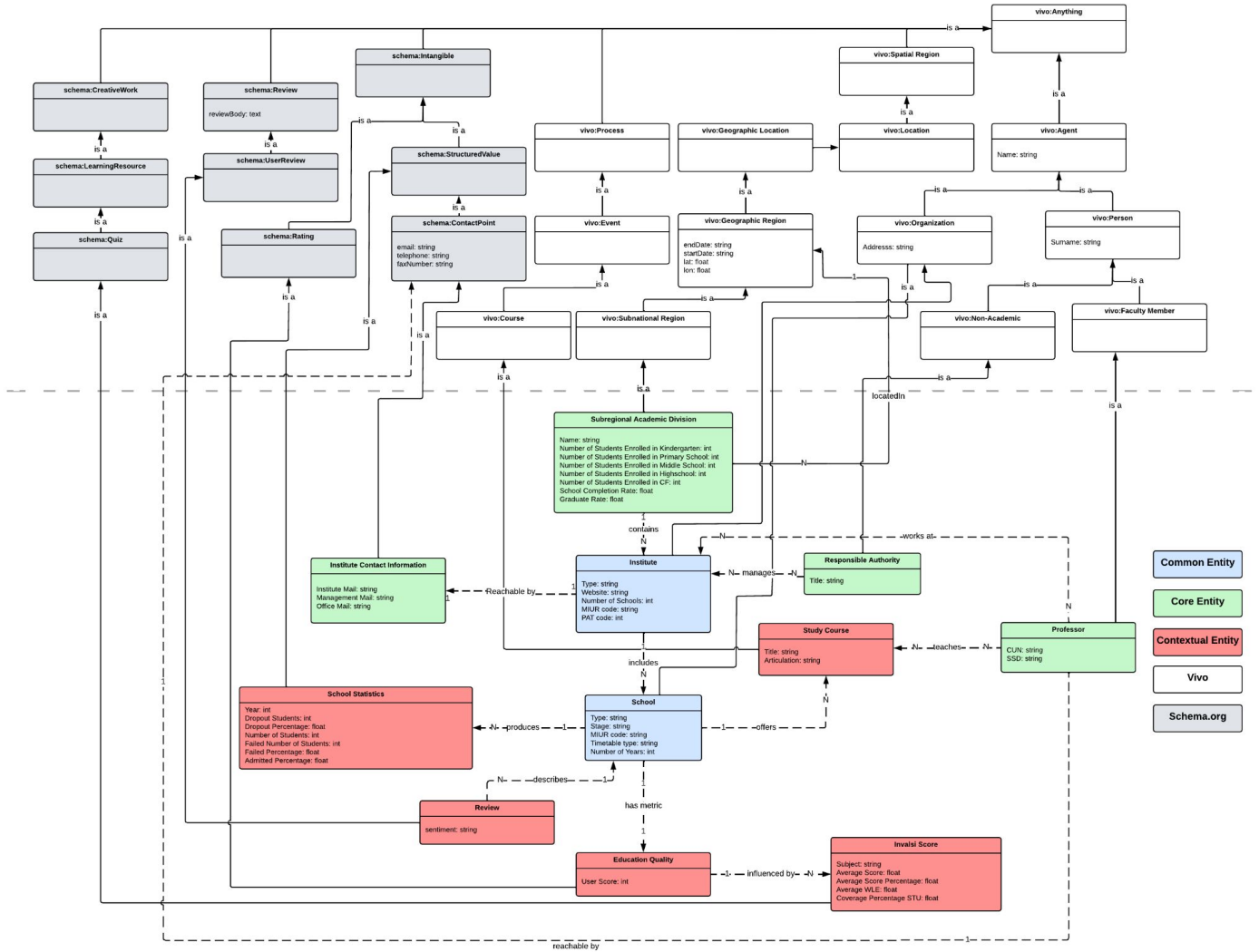
UNIVERSITÀ
DI TRENTO

Language Alignment

Since the ER model still has the problem of **conceptual diversity** (L1) and **language diversity** (L2) it can be seen as an intermediate version of ETG. To align the language we have employed **UKC**, via the **KOS2** web application, which is a specialized tool to address problems related to conceptual and language diversity.



Final ETG

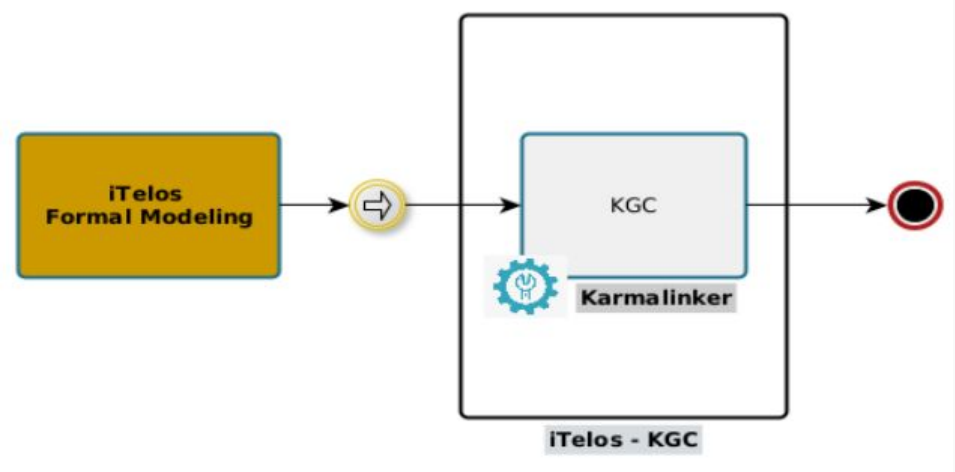


UNIVERSITÀ DI TRENTO

KNOW DIVE

iTelos - KGC

In the KGC phase, the goal is to take in input the final KG's ETG plus the set of formal data resources and produce the final KG.



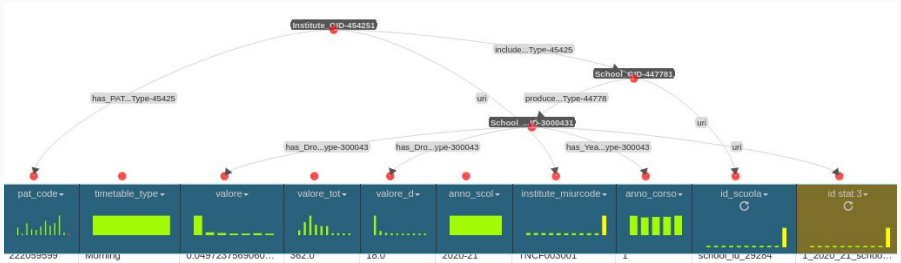
UNIVERSITÀ DI TRENTO

KNOW DIVE

Entity Matching

There are various ways to represent the same real-world objects; this phenomenon is known as **semantic heterogeneity**.

The key point of entity matching is to define each entity **Identifying Set**. The idea is that identifiers are not always included in datasets, thus, we should find those properties which allow us to uniquely identify an entity.



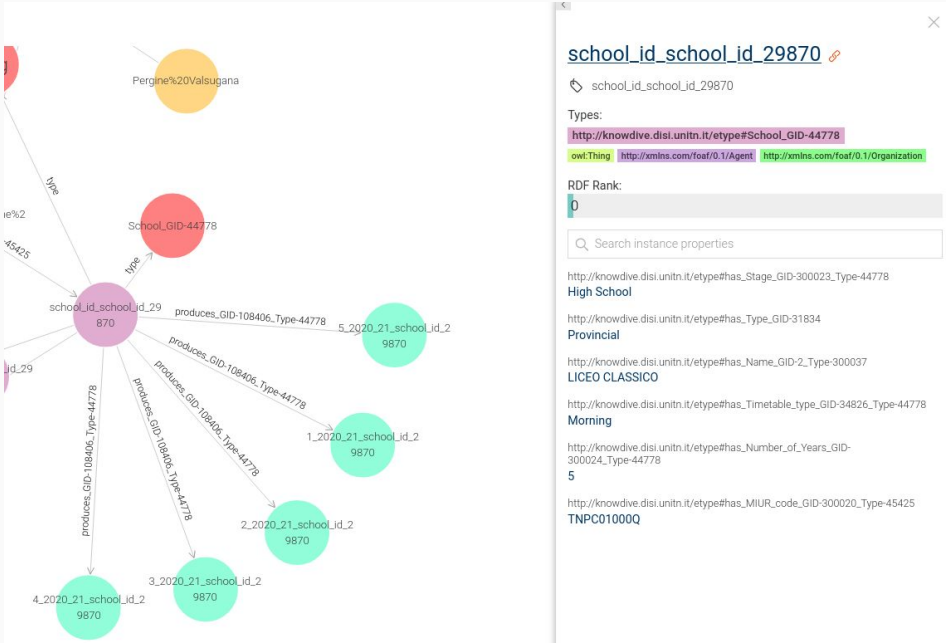
pat_code	timetable_type	valore	valore_tot	valore_d	anno_scol	institute_miurcode	anno_corso	id_scuola	id stat 3
222059599	Morning	0.0497237569060…	362.0	18.0	2020-21	TNCF003001	2	school_id_29284	2_2020_21_schoo…
222059599	Morning	0.0497237569060…	362.0	18.0	2020-21	TNCF003001	3	school_id_29284	3_2020_21_schoo…
222059599	Morning	0.0497237569060…	362.0	18.0	2020-21	TNCF003001	4	school_id_29284	4_2020_21_schoo…
222059599	Morning	0.0497237569060…	362.0	18.0	2020-21	TNCF003001	5	school_id_29284	5_2020_21_schoo…
221179501	Morning	0.0012048192771…	830.0	1.0	2020-21	TNIS00300A	1	school_id_29803	1_2020_21_schoo…

Conclusion

UNIVERSITÀ
DI TRENTO

Knowledge Graph Overview

Etype	Instances
Subregional Academic Division	183
Institute	273
School	724
Institute Contact Information	273
Responsible Authority	1239
Study Course	3326
Professor	1804
School Statistics	3846
Review	944
Education Quality	344
Invalsi score	36



Knowledge Graph Application (1)

Get the number of institutes and the number of students per school type grouped by Subregional Academic Division

SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?municipality_name ?n_k ?n_ps ?n_ms ?n_hs ?n_cf ?total_number_of_schools where {
    ?municipality rdf:type ds:Subregional_Academic_Division_GID-300037 .

    ?municipality ds:has_School_Completion_Rate_GID-300011_Type-300037
    ↪  ?school_completion_rate .
    ?municipality ds:has_Name_GID-2_Type-300037 ?municipality_name .
    ?municipality ds:has_Number_of_Students_Enrolled_in_Kindergaten_GID-300025_Type-300037
    ↪  ?n_k .
    ?municipality
    ↪  ds:has_Number_of_Students_Enrolled_in_Primary_School_GID-300022_Type-300037 ?n_ps .
    ?municipality
    ↪  ds:has_Number_of_Students_Enrolled_in_Middle_School_GID-300032_Type-300037 ?n_ms .
    ?municipality ds:has_Number_of_Students_Enrolled_in_Highschool_GID-300018_Type-300037
    ↪  ?n_hs .
    ?municipality ds:has_Number_of_Students_Enrolled_in_CF_GID-300027_Type-300037 ?n_cf .

    {
        select ?local_municipality_name (sum(xsd:integer(?institute_number_of_school)) as
        ↪  ?total_number_of_schools) where {
                ?local_municipality rdf:type
                ↪  ds:Subregional_Academic_Division_GID-300037 .
            ?institute rdf:type ds:Institute_GID-45425 .

            ?local_municipality ds:contains_GID-113311_Type-300037 ?institute .

            ?institute ds:has_Number_of_Schools_GID-300019_Type-45425
            ↪  ?institute_number_of_school .
                ?local_municipality ds:has_Name_GID-2_Type-300037
                ↪  ?local_municipality_name .
        } group by(?local_municipality_name)
    }
    FILTER(?municipality_name = ?local_municipality_name)
}
```

Result

Subregional_Academic_Division... | School_GID-44778

```
?total_number_of_schools) where {
                ?local_municipality rdf:type ds:Subregional_Academic_Division_GID-300037
.
            ?institute rdf:type ds:Institute_GID-45425 .

            ?local_municipality ds:contains_GID-113311_Type-300037 ?institute .

            ?institute ds:has_Number_of_Schools_GID-300019_Type-45425
?institute_number_of_school .
                ?local_municipality ds:has_Name_GID-2_Type-300037
?local_municipality_name .
        } group by(?local_municipality_name)
    }
    FILTER(?municipality_name = ?local_municipality_name)
}
```

Run | Press Alt+Enter to keyboard shortcut

Table | Raw Response | Pivot Table | Google Chart | Download as

Filter query results | Showing results from 1 to 95 of 95. Query took 0.1s, moments ago.

	municipality_name	n_k	n_ps	n_ms	n_hs	n_cf	total_number_of_chools
1	"Trento"	"2915"	"5529"	"3882"	"8631"	"1923"	"133"^^xsd:integer
2	"Roncegno Terme"	"66"	"142"	"146"	"0"	"0"	"11"^^xsd:integer
3	"Cavalese"	"70"	"180"	"307"	"366"	"0"	"20"^^xsd:integer
4	"Brentonico"	"99"	"177"	"125"	"0"	"0"	"4"^^xsd:integer
5	"Rovereto"	"1041"	"2124"	"1812"	"4582"	"1413"	"69"^^xsd:integer
6	"Arco"	"408"	"972"	"621"	"173"	"352"	"15"^^xsd:integer
7	"Pergine Valsugana"	"559"	"1058"	"801"	"671"	"0"	"39"^^xsd:integer
8	"Borgo Valsugana"	"153"	"277"	"197"	"891"	"188"	"27"^^xsd:integer

Knowledge Graph Application (2)

Schools with the highest number of positive reviews

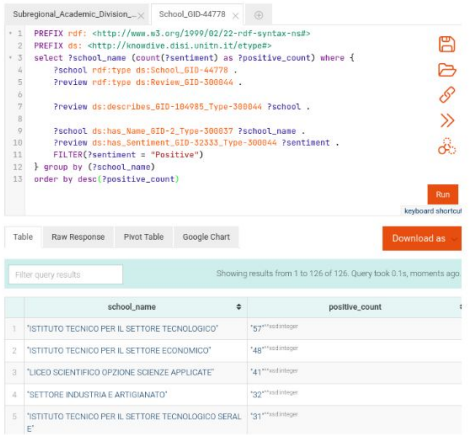
SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02
 ↪  /22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn
 ↪  .it/etype#>
select ?school_name (count(?sentiment)
 ↪  as ?positive_count) where {
    ?school rdf:type
     ↪  ds:School_GID-44778 .
    ?review rdf:type
     ↪  ds:Review_GID-300044 .

    ?review ds:describes_GID-104985_Ty
     ↪  pe-300044 ?school
     ↪  .

    ?school
     ↪  ds:has_Name_GID-2_Type-300037
     ↪  ?school_name .
    ?review ds:has_Sentiment_GID-32333
     ↪  _Type-300044 ?sentiment
     ↪  .
    FILTER(?sentiment = "Positive")
} group by (?school_name)
order by desc(?positive_count)
```

Result



Knowledge Graph Application (3)

Give me schools that have courses in the multidisciplinary area

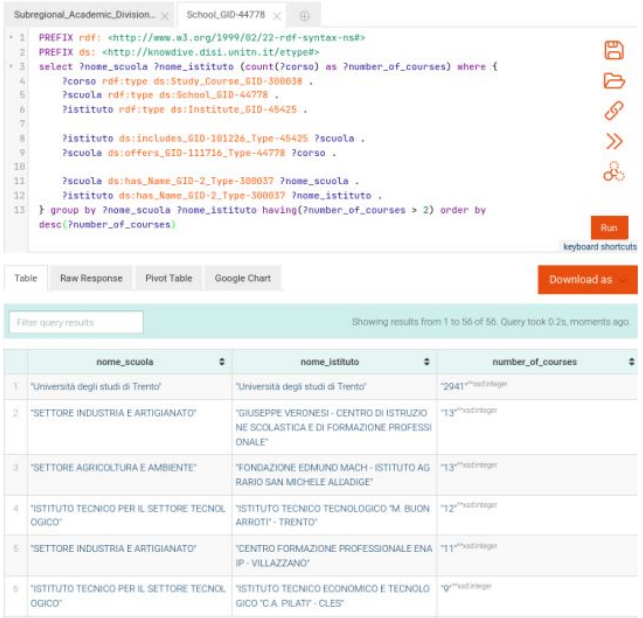
SPARQL Query

```
PREFIX rdf:
↪  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
select ?nome_scuola ?nome_istituto (count(?corso)
↪  as ?number_of_courses) where {
    ?corso rdf:type ds:Study_Course_GID-300038 .
    ?scuola rdf:type ds:School_GID-44778 .
    ?istituto rdf:type ds:Institute_GID-45425 .

    ?istituto ds:includes_GID-101226_Type-45425
    ↪  ?scuola .
    ?scuola ds:offers_GID-111716_Type-44778 ?corso
    ↪  .

    ?scuola ds:has_Name_GID-2_Type-300037
    ↪  ?nome_scuola .
    ?istituto ds:has_Name_GID-2_Type-300037
    ↪  ?nome_istituto .
} group by ?nome_scuola ?nome_istituto
↪  having(?number_of_courses > 2) order by
↪  desc(?number_of_courses)
```

Result



UNIVERSITÀ DI TRENTO

Knowledge Graph Evaluation

Coverage

$$Cov_e(CQ_e) = \frac{|CQ_e \cap ETG_e|}{CQ_e} = \frac{11}{12} \approx 0.91$$

Knowledge Graph Evaluation

Coverage

$$Cov_p(CQ_p) = \frac{|CQ_p \cap ETG_p|}{CQ_p} \approx \frac{39}{40} \approx 0.975$$

UNIVERSITÀ DI TRENTO

Knowledge Graph Evaluation

Coverage

$$Cov_e(RO_e) = \frac{|RO_e \cap ETG_e|}{RO_e} = \frac{5}{22} \approx 0.227$$

UNIVERSITÀ DI TRENTO

Knowledge Graph Evaluation

KNOW DIVE

Coverage

$$Cov_p(RO_p) = \frac{|RO_p \cap ETG_p|}{RO_p} = \frac{6}{24} = 0.25$$

UNIVERSITÀ DI TRENTO

Knowledge Graph Evaluation

Dense and Connected KG Analysis

$$\sum_{k=1}^{N} E(T_k) = 12992$$

$$\sum_{k=1}^{N} Op(T_k) = 7968$$

$$\sum_{k=1}^{N} Dp(T_k) = 116332$$

UNIVERSITÀ DI TRENTO

Conclusion and Future Work

We think that the objective of creating a Knowledge Graph that contains details and data on educational institutions in Trentino has been attained. Our project's outcomes are practical and usable in real-world situations.

Although the final Knowledge Graph is straightforwardly applicable as is, there is still room for improvement. As a future work, we propose to:

- integrate data related to other regions;
- retrieve data for defining the popularity of a school;
- integrates high/middle school professors;
- gather more precise information for the timetables of the courses.