



UNIVERSITY
OF TRENTO - Italy



Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2022 - Project Report Point of Interest

Document Data:

January 10, 2023

Reference Persons:

Florian Dufaure, Simon Pichenot

© 2023 University of Trento
Trento, Italy

KnowDive (internal) reports are for internal only use within the KnowDive Group. They describe preliminary or instrumental work which should not be disclosed outside the group. KnowDive reports cannot be mentioned or cited by documents which are not KnowDive reports. KnowDive reports are the result of the collaborative work of members of the KnowDive group. The people whose names are in this page cannot be taken to be the authors of this report, but only the people who can better provide detailed information about its contents. Official, citable material produced by the KnowDive group may take any of the official Academic forms, for instance: Master and PhD theses, DISI technical reports, papers in conferences and journals, or books.



Index:

1	Introduction	1
2	Purpose and Domain of Interest (DoI)	1
3	Data Sources	1
4	Purpose Formalization	3
5	Inception	6
6	Informal Modeling	9
6.1	ER model	9
6.2	Teleology	11
7	Formal Modeling	13
7.1	Ontology	13
7.2	Teleontology	14
7.3	Final Teleontology	15
8	KGC	15
8.1	Data mapping	16
8.2	Entity matching	17
9	Outcome Exploitation	17
10	Conclusions & Open Issues	18

Revision History:

Revision	Date	Author	Description of Changes
0.1	20.04.2020	Fausto Giunchiglia	Document created

1 Introduction

Reusability is one of the main principles in the Knowledge Graph Engineering (KGE) process defined by iTelos. The KGE project documentation plays an important role in order to enhance the reusability of the resources handled and produced during the process. A clear description of the resources and the process developed, provides a clear understanding of the KGE project, thus serving such an information to external readers in order to exploit that in new projects.

The current document aims to provide a detailed report of the KGE project developed following the iTelos methodology. The report is structured, to describe:

- Section 2: The project's purpose and the domain of interest and the resources involved (both schema and data resources) in the integration process.
- Section 2: The input resources considered by the KGE project.
- Section 4, 5, 6, 7: The integration process along the different iTelos phases, respectively.
- Section 8: How the result of the KGE process (the KG) can be exploited.
- Section 9: Conclusions and open issues summary.

2 Purpose and Domain of Interest (DoI)

- Our application will be able to provide several points of Interest in the city of Trento. So users can find places they're interested in to go during their vacations. Our different point of interest will be separated in categories to help the user to find the information he wants. A point of interest is a specific location which defines something interesting. For examples restaurants, retail stores and grocery stores are points of interest
- The application will cover all the points of interest in the city of Trento, such as museums, restaurants, bars, art galleries, parks. Our application also contains information about buses and car rentals; in order to provide the people using it to move from different point of interest to another.

3 Data Sources

This section has to report and describe the input resources considered:

- **Knowledge sources:**
 - Commercial building
 - Transportation Area

- Transportation building
- Educational building

- **Data sources:**

For the points of interest we have found this data set : Link to the raw dataset
In the data set we kept only the following attributes for each entities.

- Location: coordinate (lat, long), address (country, street, city, postalCode, region)
- Description: english and italian
- Contact: phone, mail, web site url
- Category: name of the category
- Name: english, italian

The data from OpenData wasn't clear. First, all object hasn't key and the file wasn't a real json.

```

1  /* 10 */
2  {
3      "_id" : "it.trentour.domains.core.PointOfInterest.529ef58f3391
baee95b52c84",
4      "domainId" : "529ef58f3391baee95b52c84",
5      "type" : "it.trentour.domains.core.PointOfInterest",
6      ...
7  }
8

```

Before each object we have this comment `/* [a number] */`. We have used this script to modify the file and create a real json.

```

1 import re
2
3 file = open("poi.txt", encoding="utf8")
4 lines = file.readlines()
5
6 # counter for the number of object
7 cpt = 0
8 for i in range(0, len(lines)-1):
9     # use to match with the comment
10    if re.match(r" /* [0-9]* \*/\n", lines[i]) != None:
11        # replace the comment by a key for the object
12        lines[i] = re.sub(r"/\* [0-9]+ \*/", "", lines[i])
13        lines[i] += str(cpt) + ":" + lines[i]
14        cpt+=1
15
16 poi = open("poi-clean.json", "w", encoding="utf8")
17 poi.write("{")
18 for i in range(0, len(lines)-1):
19     poi.write(lines[i])
20
21 poi.write("}")
22 poi.close()
23 file.close()

```

Listing 1: Python cleaning script

After using the script the previous object be transform like that.

```
1  {
2      "0": {
3          "_id": "it.trentour.domains.core.PointOfInterest.52af11431512
4 e121e8dd32ee",
5          "domainId": "52af11431512e121e8dd32ee",
6          "type": "it.trentour.domains.core.PointOfInterest",
7          ...
8      },
9      ...
10 }
```

For data sources about Transportation (buses in our case), we used : Trentino Transporti Trentino Transporti It provide us data for bus, train, and cable car in the GTFS format. So for our application we only kept data about buses.

4 Purpose Formalization

The Purpose formalization section has to report:

- **Personas:**

- Arianna, 19 years old, student at university, she's new in town, she's in Trento as an Erasmus for the year. It's the first time for her that she's leaving alone. She wants to meet new people and enjoy her student life.
- Oscar, 34 years old, he's a tourist who wants to visit the most places he can. He's brother studies in the science university. He's going to a tourist Office to learn more about the city. He also wants to enjoy the local food.
- Simone, 27 years old, an employee who needs to prove himself in his company. He needs to organize a big seminar for his enterprise in the city to celebrate the 10th anniversary.
- Romi, 28 years old, she's a sport addict and she's with a group of friends for a weekend. They want to enjoy the best eyesight in Trento, especially on the sunset. Also do some hiking since the region is the best for it.

- **Scenarios:**

- Arianna is very stressed about her exams. She's searching for a library with a few people to study quietly for her KGE exam. After that she proposed to her friends to go take a coffee in the city center to relax.
- Oscar arrived in town, his hotel is available in 5 hours, during this time he wants to see his brother in the science university in Povo. He searched for a bus to come

from the train station to Povo. After the check-in he want to invite his brother to restaurant.

- Simone need to book a restaurant for 25 people for 20h. Also he need to come from the train station to the restaurant with bus. After, he wants to go at a big bar to drink some beers. To go to this bar they can take a bus.
- Romi and his friends love the mountains, before coming in the region they want there hikes. But none of use have car so they need to rent a car to go in the hotel in Bondone. To find some hikes in the region she will go to the tourist office of Bondone.

- **Competency Questions (CQs):** the list of CQs created considering the personas in the scenarios defined.

Personas	Scenario	Competency Question
Arianna	1	Where is the library in the city ?
Arianna	1	Is the library open regularly ?
Arianna	1	What is the best time to go there ?
Arianna	1	Where can I take a coffee ?
Oscar	2	What are the most visited places ?
Oscar	2	Where is a pizzeria ?
Oscar	2	Where can I buy bus tickets ?
Simone	3	Where are the places to eat for at least 25 people ?
Simon	3	Where is the nearest bus stop of the restaurant ?
Simone	3	Wich are the less crowded bars ?
Romi	4	Where are the finest spots in mountains ?
Romi	4	Where are the spots to hike ?
Romi	4	Where is the hotel ?
Romi	4	Where is the tourit office in Bondone ?
Romi	4	Where can I rent a car in Trento ?

- **Entities identified:**

- **Structures:**

- * Bar
- * Restaurant
- * University libraries
- * University
- * Tourist office
- * Car rental

- **Location:**

- * Ski station

- **Conveyance:**

- * Bus

* Car

In protege the entities look like that:

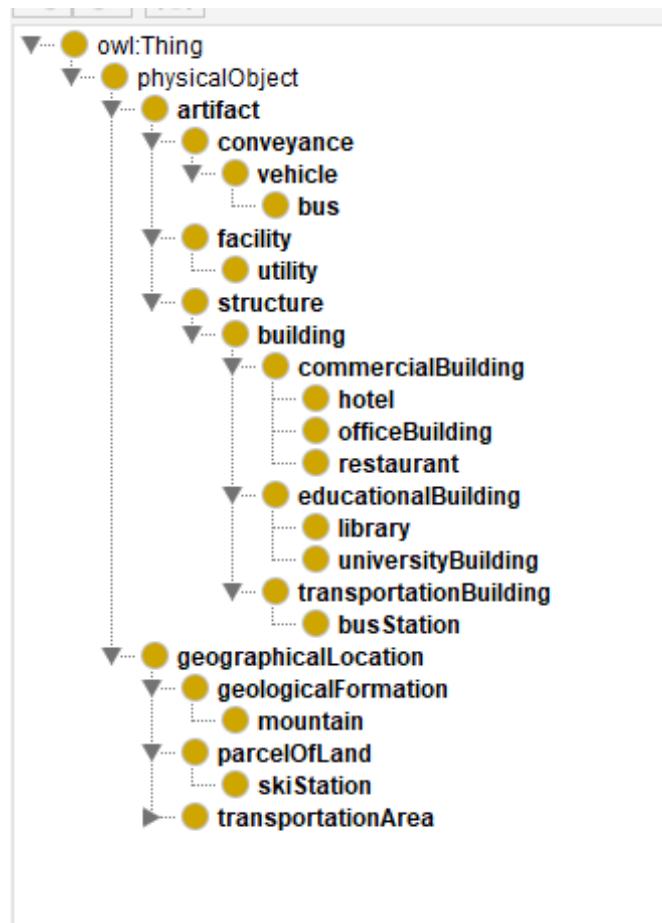


Figure 1: Protege entities

At the end of the project (during the data integration) we have some problem. To try solving this problem we have change entities in protege and we have the following tree.

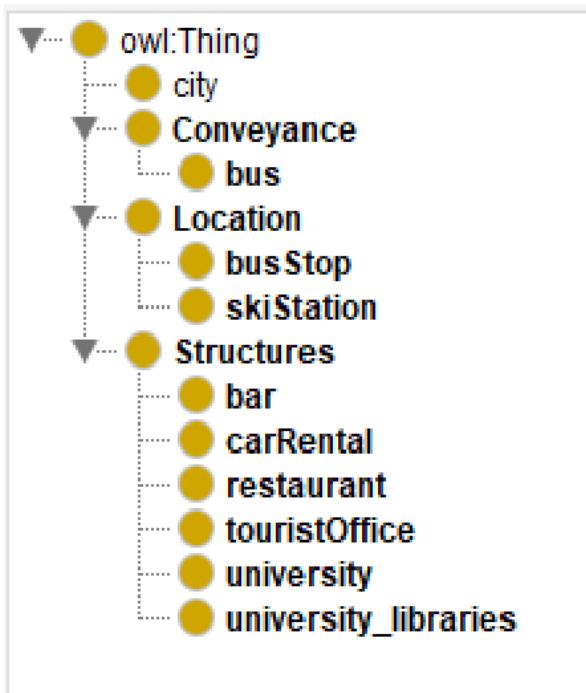


Figure 2: Protege entities to fix bug

We have maked a smaller entite tree to have an better view in what append in our data integration.

5 Inception

This section aims to report the KGE sub process performed during the inception phase, by describing each activities both in schema and data layer.

Inception sub activities:

- Resources collection/scraping

We went on OpenData to get the data of the city of Trento, we found a JSON on Point of Interest which was very useful.

We didn't had to scrap any data in this case. Because in the data set all the information were useful to us, so we kept everything.

About the data for the buses. We scrap in order to keep only information that we needed, which are : bus stop and bus route

- Resources filtering and classification over common, core and contextual

We get the restaurants and the bars, we filtering in order to keep all the data where we had information, such as : name, number, address, coordinates. We also find museums to respond the 2nd scenario.

We also had many data on hotels. Florian didn't want to keep them because they're not part of competency questions that we have. On the other hand Simon wanted to have them because they're still point of interest, especially for tourists, such as Oscar in our scenarios. So we didn't have them for the moment, but we still have them and integrate them very easily

About the data about hikes, we had some difficulties to them any of it. So for now our solution is to create our own data by scraping other data collection.

- Resources formatting (semi-formal transformation)
 - bar : {address: string, phone: number, mail: string, description: string, url?: string}
 - restaurant : {address: string, phone: number, mail: string, description: string, url?: string}

```
1  {
2      "content" : {
3          "poiData" : {
4              "location" : {
5                  "coordinate" : {
6                      "latitude" : 46.119,
7                      "longitude" : 11.6881,
8                  },
9                  "addresses" : {
10                     "IT" : {
11                         "country" : "ITA",
12                         "state" : null,
13                         "region" : "TN",
14                         "city" : "Castello Tesino",
15                         "postalCode" : "38053",
16                         "street" : "Loc. Passo Brocon"
17                     }
18                 }
19             },
20             "contact" : {
21                 "phone" : "+39 (0461) 594364",
22                 "mail" : "booking@brocon.it",
23                 "url" : "www.albergopassobrocon.it"
24             },
25             "timetable" : {}
26         },
27         "objData" : {
28             "name" : {
29                 "EN" : "",
30                 "IT" : "Passo Brocon \\"Da Sciopo\\"
31             },
32             "description" : {
33                 "EN" : "Typical Restaurant",
34                 "IT" : "Ristorante tipico"
35             },
36             "category" : "ristorante",
37         }
38     },
39 }
```

Link to the raw dataset

6 Informal Modeling

This section is dedicated to the description of the informal modeling phase. Like in the previous section, it aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced.

More in details, this section provides a description of the following activities:

- ER model description. Both the creation process (decisions taken) and the outcome have to be described.
- Teleology building.
- Datasets filtering and alignment with teleology.
- Phase open issues.

6.1 ER model

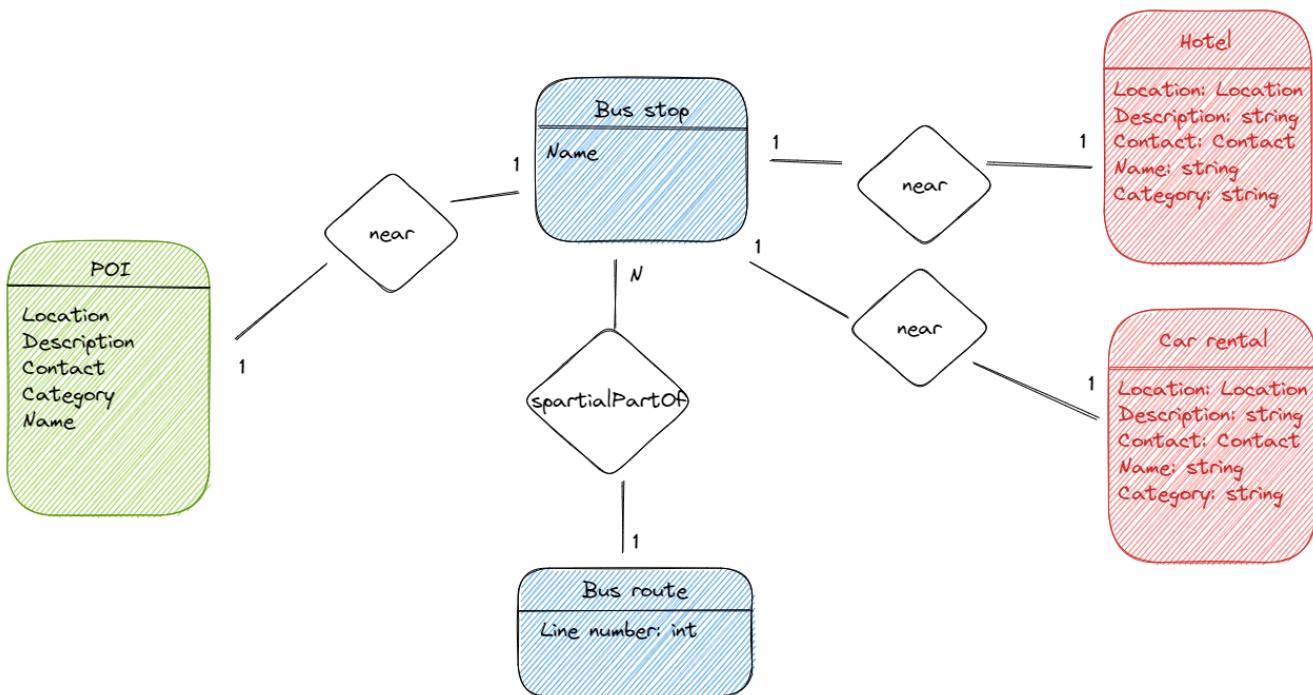


Figure 3: ER diagram

Red	contextual entities
Blue	common entities
Green	core entities

The POI object in the figure 3 is a group of all following entities: bar, restaurant, tourist office, university stuff. We made the diagram like that to simplify.

We decided to made an special entity of Hotel and Car rental, because for us they're not a Point of Interest but it's usefull for our use cases since our personnas can stay for more than one day in Trento or rent a car.

So our first ER Model was the one above and after reflection we saw that it contains many problems. We needed to had People in this diagram to represent the ones that will be use this service. And with that our diagram is more relevant for our use.

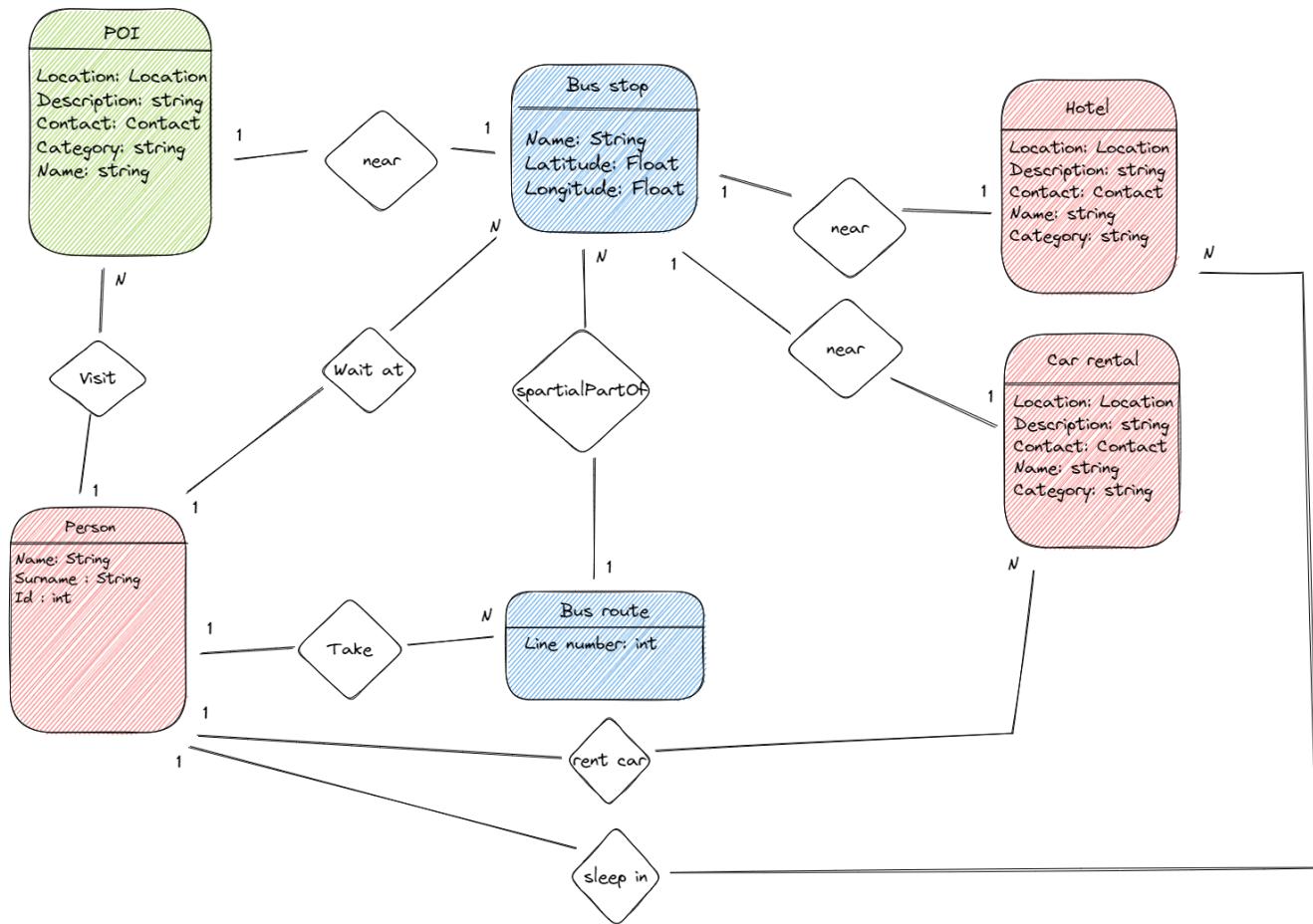


Figure 4: Final ER diagram

Red	contextual entities
Blue	common entities
Green	core entities

6.2 Teleology

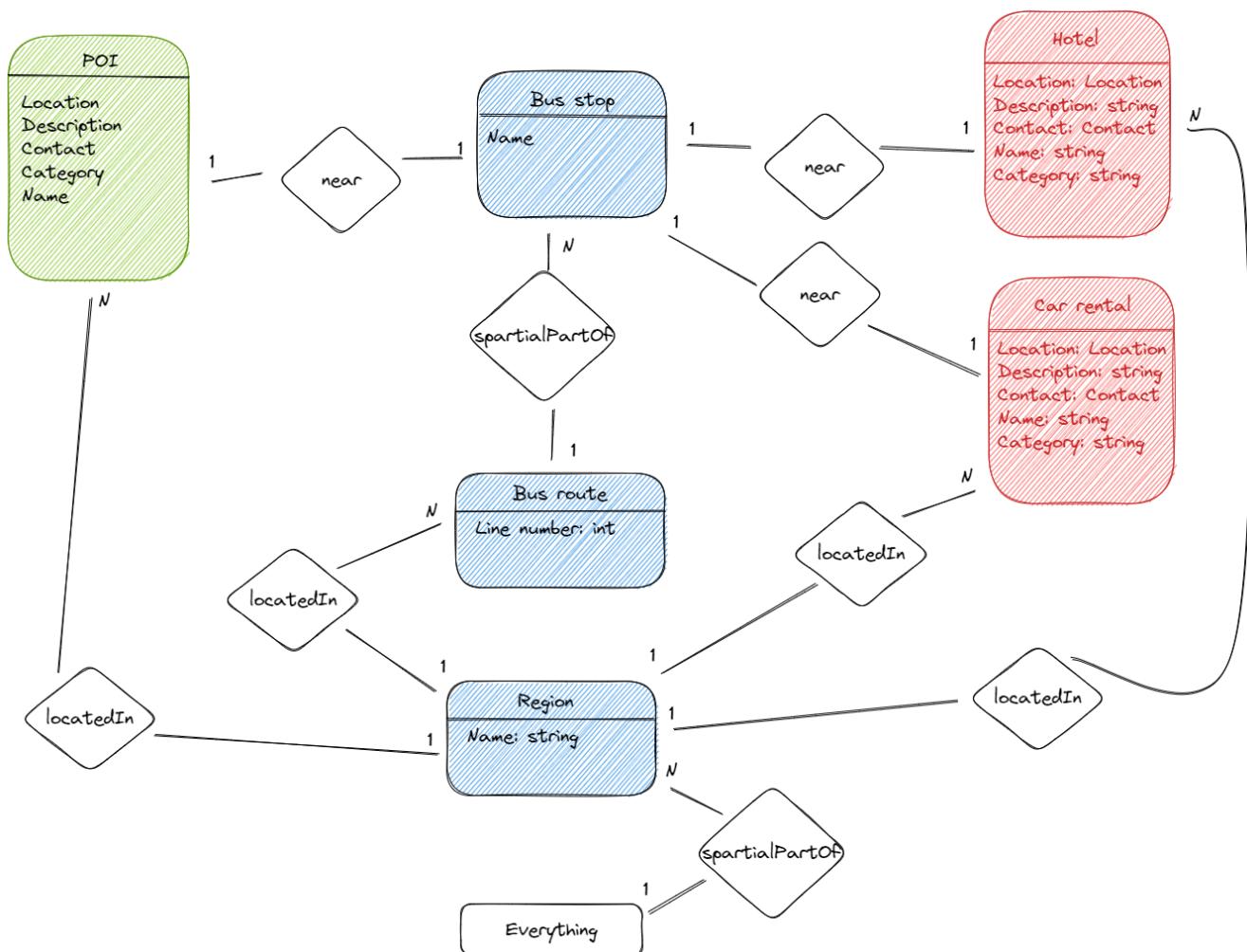


Figure 5: Teleology diagram

Red	contextual entities
Blue	common entities
Green	core entities

In the same way as before, we modified the teleology with our previous modification to have the correct one.

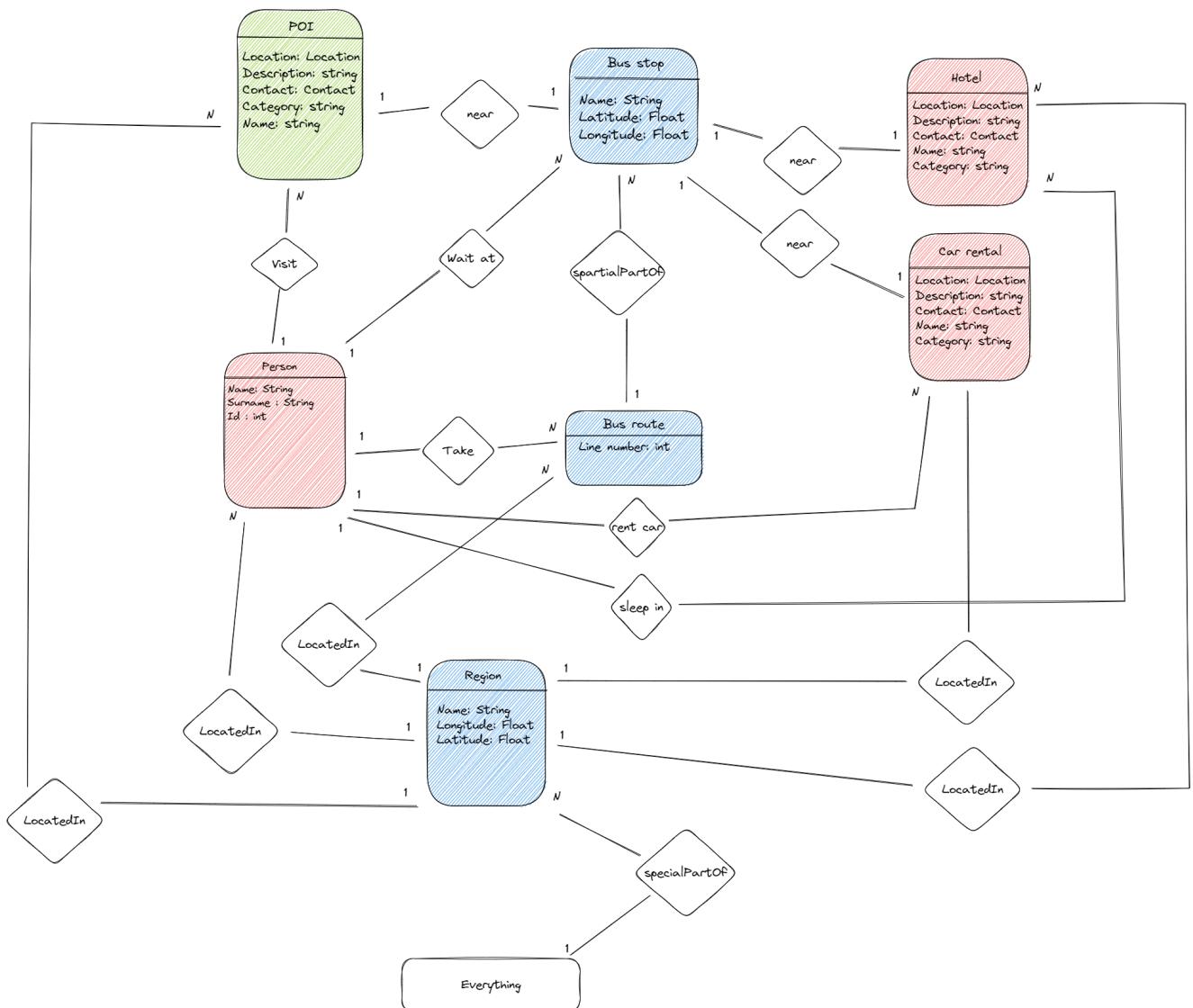


Figure 6: Final teleology diagram

Red	contextual entities
Blue	common entities
Green	core entities

7 Formal Modeling

This section is dedicated to the description of the formal modeling phase. Like in the previous section, the current one aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced.

More in details, this section provides a description of the following activities:

- ETG generation
- Language alignment
- Formal data creation, by aligning datasets with the ETG created.
- Phase open issues

Like the previous phase, also the current one has to report the decision made during the phase activities, with the weak and strong point associate to them. If difficulties and/or open issue have been encountered, they should be reported as well.

7.1 Ontology

So far our ontology we saw that we had 3 distinct important entities that's why we decided to make it this way.

We have of course all the Point of Interest that are represented in our data. After that, we have everything that is related to the buses, to reach all the different point of interest in the city. And to conclude we added the different people that are going to use our service, which are tourist or students.

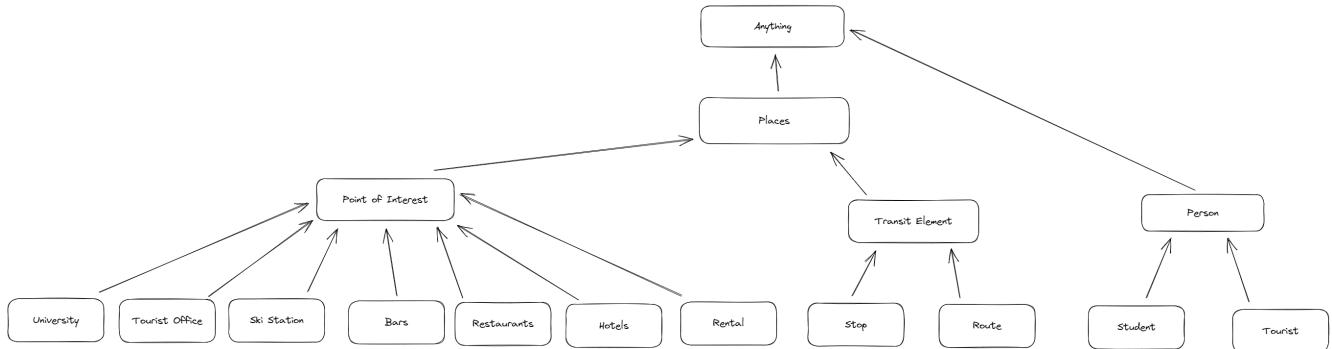


Figure 7: Ontology diagram

7.2 Teleontology

So we linked all our entities from our teleology to the matching one in the ontoly in order to create our Teleontology.

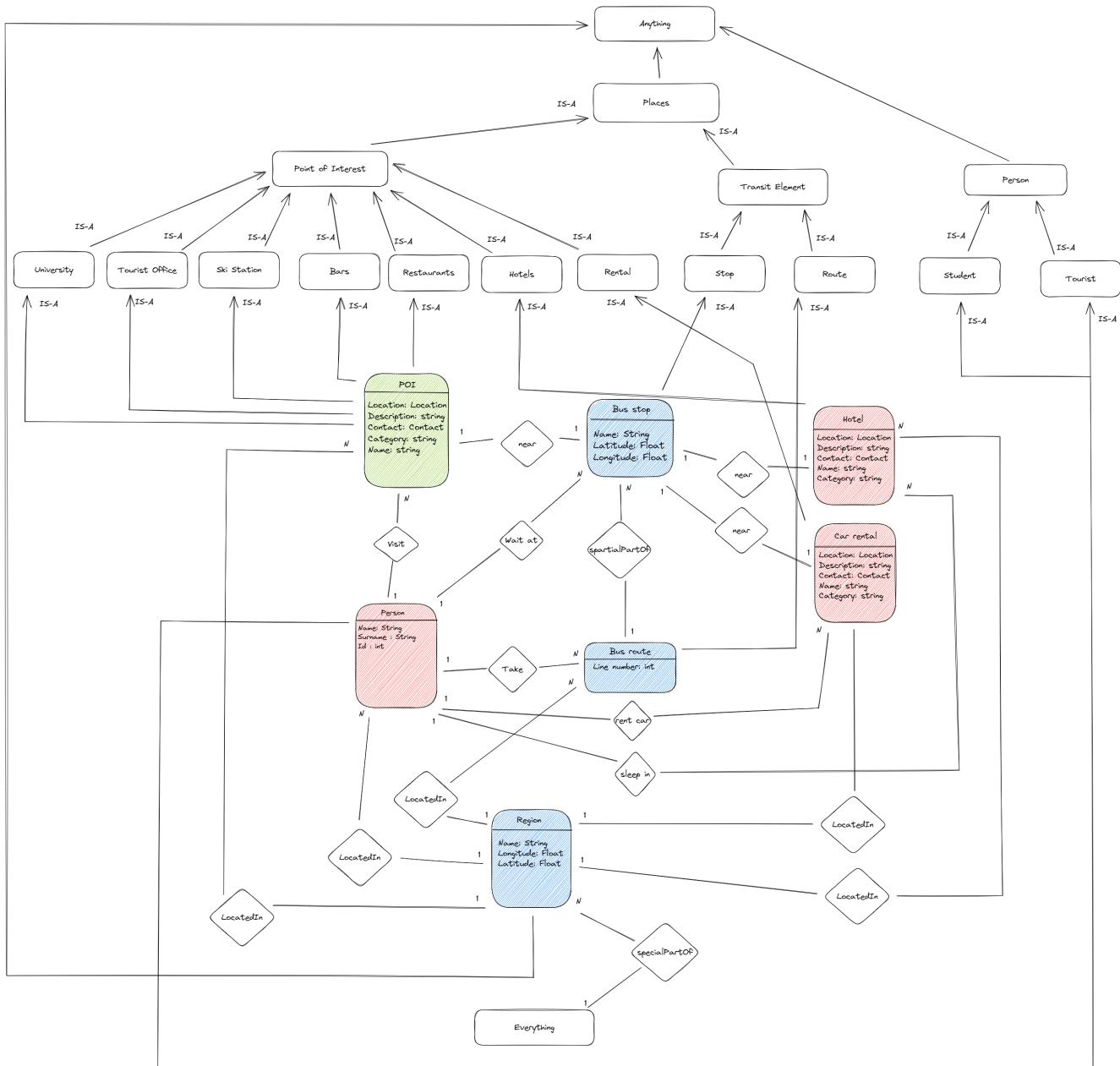


Figure 8: Teleontology diagram

7.3 Final Teleontology

In order to get our final teleontology, we just removed the entities : Everything and Anything to have our final result.

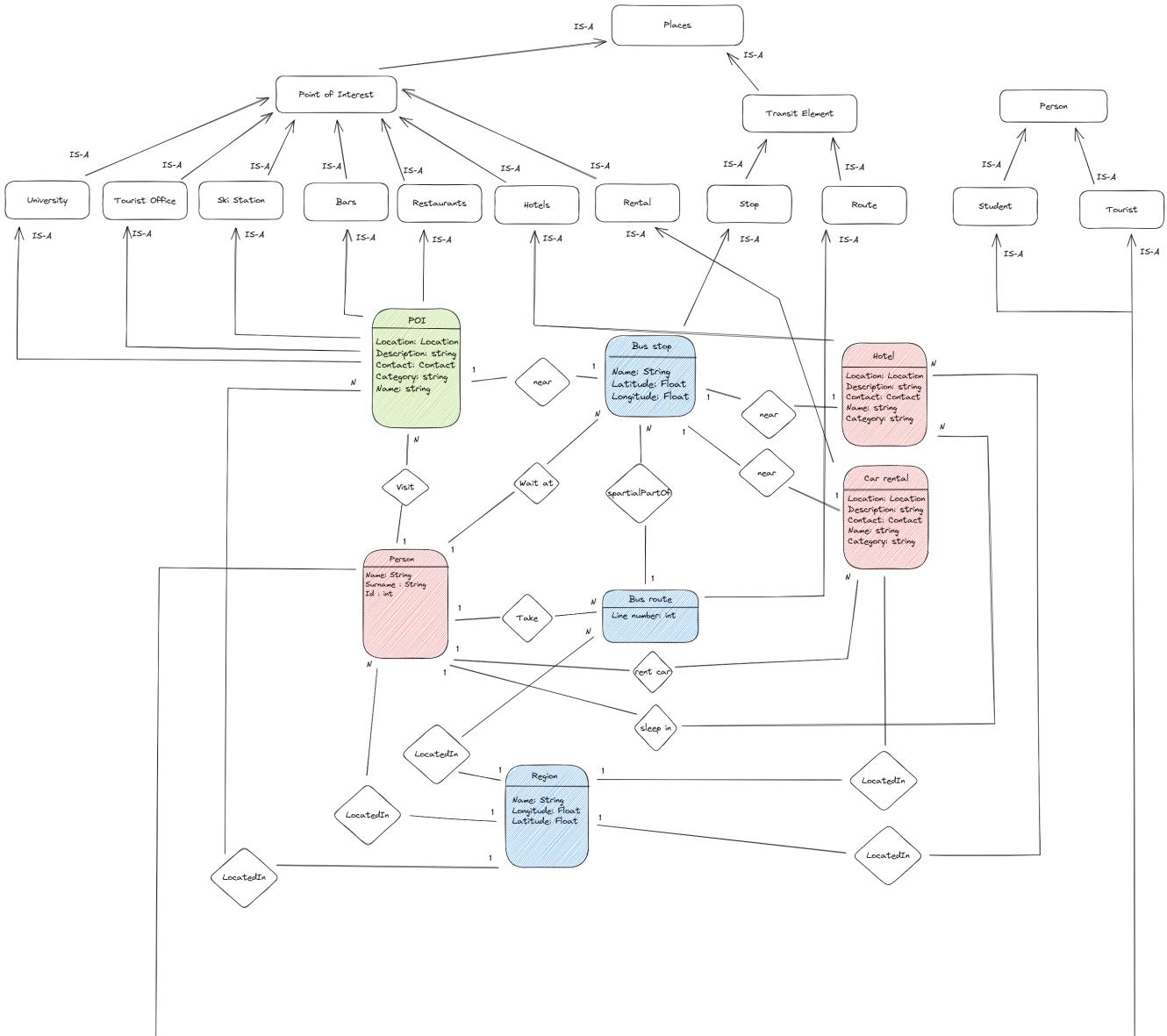


Figure 9: Final Teleontology diagram

8 KGC

This section is dedicated to the description of the KGC phase. Like in the previous section, the current one aims to describe the different sub activities performed by all the team



members, as well as the phase outcomes produced.

More in details, this section provides a description of the following activities:

- Data mapping. The mapping activities describe how the final KG is created. Provide a description of such activities for the datasets considered.
- Entity matching (semantic heterogeneity). Describe how different representation of the same real world objects have been handled.
- Phase open issues

8.1 Data mapping

When we have take the data from openData, we have clean it and reorganized it to have clear and good format for Karma.

After we have link the data and the teleology in Karma.

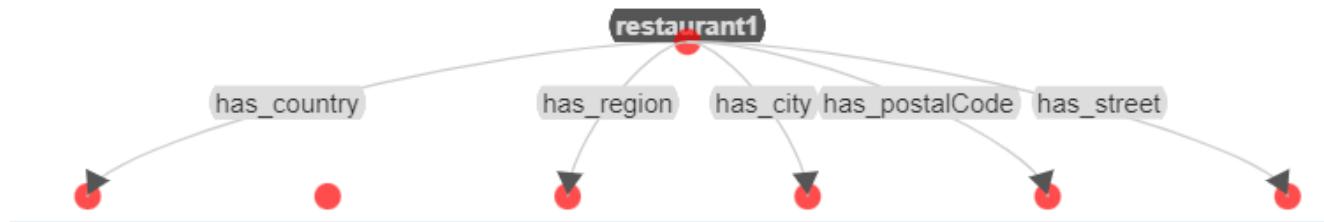


Figure 10: Karma diagram

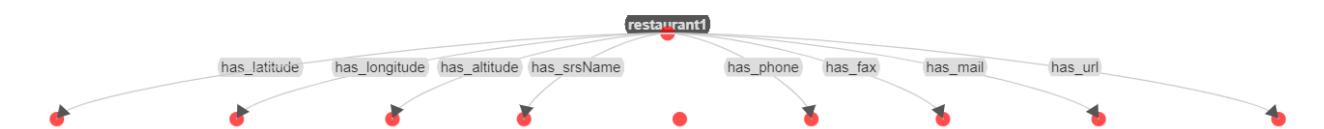


Figure 11: Karma diagram 2

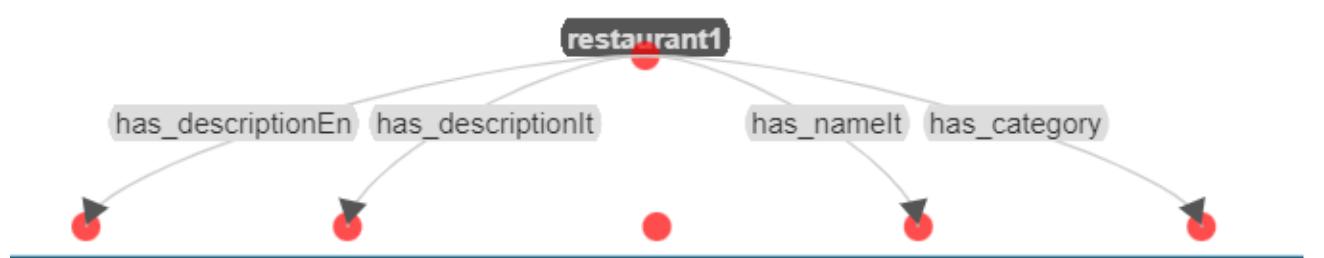


Figure 12: Karma diagram 3

The 4th schema are for the Point of Interest but we have cut it on 3 for more readability.



8.2 Entity matching

All Point of Interest have the same format in our data set. So that was not a difficult task to respect a semantic heterogeneity. Like you can see in the previous schema with the restaurant, it's the same for every POI.

Some attributes are missing like the state or the english name but it's missing for all POI.

9 Outcome Exploitation

This section aims to provide a description of the KGE process outcome. Here you have to report the final Knowledge Graph information statistics (like, number of etypes and properties, number of entities for each etype, and so on). Moreover this section has to provide a description for the KG possible exploitation, like examples of queries executed, execution time, and so on.

We finished our knowledge graph but we couldn't make it work in GraphDb. So we tried to make it work only using one point of interest (a lower set of data).

We also tried to change our ontology model with a litter one (with just restaurant entity) but it didn't worked out also.

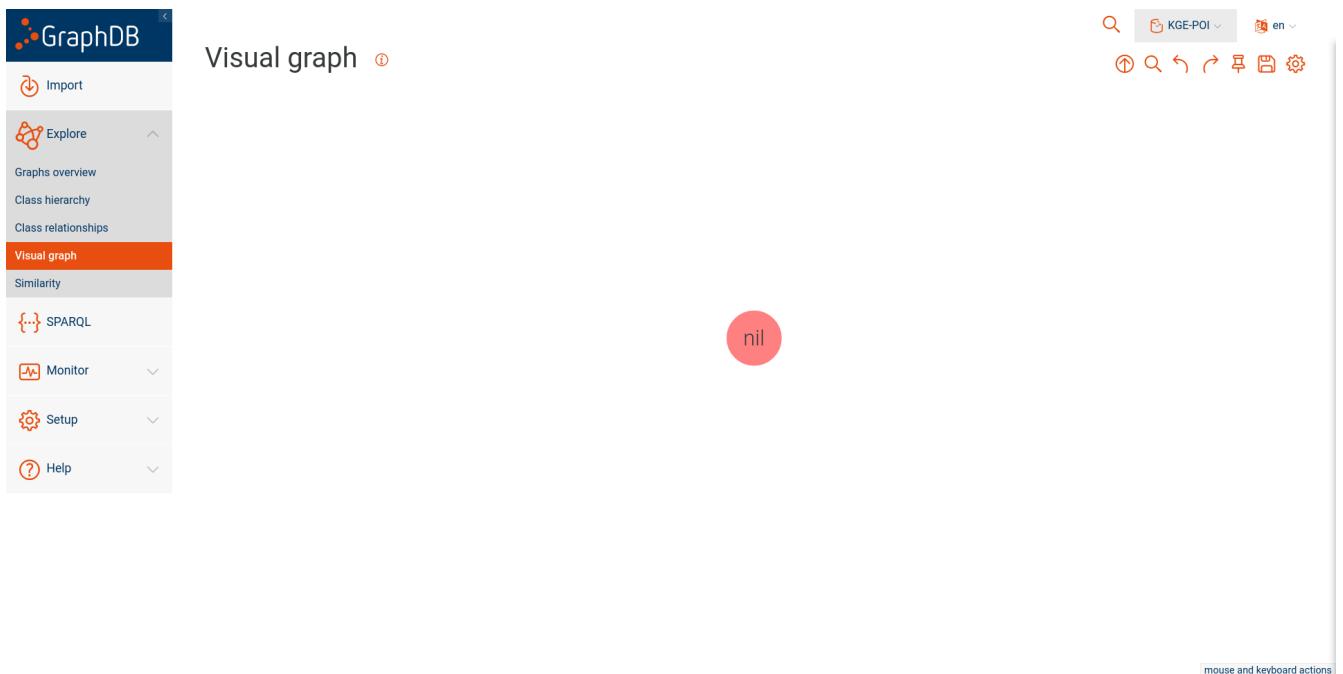


Figure 13: Screenshot from GraphDb

10 Conclusions & Open Issues

This section concludes the current document with final conclusions regarding the quality of the process and final outcome:

- Did the project respect the scheduling expected in the beginning ?
- Are the final results able to satisfy the initial Purpose ?
 - If no, or not entirely, why ? which parts of the Purpose have not been covered ?

Moreover, this section aims to summarize the most relevant issues/problems remained open along the KGE process. The description of open issues has to provide a clear explanation about the problems, the approaches adopted while trying to solve them and, eventually, any proposed solution that has not been applied.

At the end, the project final result are not able to satisfy the intial Purpose since we're not able to make some request out of our knowledge graph. But all the theory part is done and we think functional. Like we describe before we tried several solutions in order to solve this problem but we couldn't do it in time. Maybe with more time or more mentoring and help we could have make it work.

In order to have everything done we may should have organize ourself a bit more a the begining of the project in order to have more time at the end to deal with the outcome exploitation and the issues that comes with it.