KDI : Knowledge and Data Integration

Transportation

KDI Final Presentation

Contributors

- Data Scientist: Alberto Carbognin
- Data Scientist: Omid Jadidi
- Domain Expert: Fivos Kapidis
- Knowledge Engineer: Fivos Kapidis
- Knowledge Engineer: Antonio Luigi Stefani
- Project Manager: Antonio Luigi Stefani
- Tutor Data: Alessio Zamboni
- Tutor Knowledge: Mayukh Bagchi

Table of Contents

Table of Contents

Project description

Our ontology wants to explore the transportation domain in order to solve the daily problems of finding cheapest and shortest routes to reach a certain place.

Our goal is to provide all the more useful information regarding the road system, exploring also the railway system, the cycling routes and the mountain trails. In particular to reach our goal we have circumscribed the problem into two main issues:

- information regarding the road system itself: roads, modes of transport, prices, scheduling of the public transport.
- information regarding the facilities related to each road and public transport service.

Project description

To design our system first of all we put ourselves in the users shoes:

Name	Age	Interest	Usage
Claudio	17	Go to school, to his basketball training or matches all around the Trentino. Hang out with his friends	Check the public transportation schedules and the position of the facility he needs
Andrea	35	Go to the work by car or sometimes by public transports. Go hiking in the Trentino mountains	Check the fastest ways to get his several destination. Planning his hiking
Elisa	23	Go to the university, go cycling and camping with friends	Check cycling routes and facilities, check public transportation schedules and cheaper campsites
Maria	58	Go to the work by using her car or the car-sharing service, save as much time as possible. Visit the Trentino during the holidays	Check fastest and cheapest routes, check the schedule of the public transports, check the position of all the facilities she needs

Figure: Users Description

Table of Contents

Resources

Accordingly with the iTelos Methodology we have followed 5 different steps to achieve the final result. In particular we have identified 4 different stages in which we have used several tools in order to describe the different resources:

- Knowledge resources: firstly we used the yEd web application to design our model at an higher level (Informal Modeling), then we passed to Protégé to formalized the schema (Formal Modeling) and obtain the final ontology schema.
- Data resources: since the beginning we focused on finding datasets able to satisfy all the attributes of any EType designed in the EER Model also evaluating the *scraping* option that wasn't feasible in our case.
- Metadata resources: to have a better idea on the datasets and acquire knowledge on them we collected the metadata from time to time, filtering and formatting them in a more usable way, in particular we focused on finding a role to each attributes.

Resources: Knowledge

- EER Model (Informal Modeling): https://github.com/UNITN-KDI-2020/KDITransportation2020/blob/master/code/Schemas/InformalModeling/InformalModelingSchema.png
- Ontology (Formal Modeling): https://github.com/UNITN-KDI-2020/KDITransportation2020/blob/master/code/Schemas/FormalModeling/FullOntology.JPG

Resources: Data and Metadata

- Datasets (Final Results): https://github.com/UNITN-KDI-2020/KDITransportation2020/tree/master/dataset/final%20results
- Metadata (Final Results): https://github.com/UNITN-KDI-2020/KDITransportation2020/tree/master/dataset/final%20results/metadata

Table of Contents

Problems and Solutions

During the process to obtain the final Knowledge Graph we have encountered several issues on both knowledge side and data side:

- Knowledge: in the Formal Modeling Phase we faced several problems in suiting up our Ontology designed with Protégé to the KOS software. To solve these issues it required to change the version of Protégé several times and to do some minor changes in the Ontology.
- Data: in this case we faced several problems since the beginning, indeed we have many difficulties in collecting datasets regarding: public transportation datasets because of the privacy agreement of the service provider (Trenitalia); facilities prices datasets due to a leak of the data themselves; facilities ranking datasets since their provider required a payment to access them.

Table of Contents

Outcomes

Following the iTelos Methodology we obtained two main schemes:

- SKG. Comparing our one to those found in the net we can notice that is really small: 18 ETypes and 3 Enumerations versus 42 ETypes of the one chosen to compare our work.

Information	**Data**
Coverage	0.26
Flexibility	0.42
Extensiveness	0.11
Sparsity	0.63

As it is possible to see our strength is in the flexibility and in the sparsity, indeed the metrics show how our ontology is very good in supporting other ones working on the same domain.

Outcomes

- DKG. Considering our initial purposes we obtained a very good Knowledge Graph, indeed we have a coverage of all the queries of the 93%, we have not obtain the total coverage due to the problems faced during the datasets collecting phase.

At the end of the design of the DKG we also provided some queries solved to show how the KG works:

- KarmaLinker: https://github.com/UNITN-KDI-2020/KDITransportation2020/blob/master/code/Schemas/DataIntegration/GraphDB/Karmalinker.png
- GraphDB: https://github.com/UNITN-KDI-2020/KDITransportation2020/blob/master/code/Schemas/DataIntegration/GraphDB/VisualGraphDB_Facility1.png
- Queries: https://github.com/UNITN-KDI-2020/KDITransportation2020/blob/master/code/Schemas/DataIntegration/GraphDB/query11.JPG

Table of Contents

Open Issues & Future works

To improve our KG it is possible to proceed on both the knowledge side and data side:

- Knowledge: consider other type of routes and mode of transport (lake, river). Enlarge the schema in order to consider not only people transportation but also goods transportation.
- Data: find some datasets regarding facilities prices, fares and ranking. Enlarge our datasets by considering those provided by all the public transport agencies (Trenitalia). Develop a system able to provide traffic information (estimating them when missing).

KDI : Knowledge and Data Integration

Transportation

KDI Final Presentation