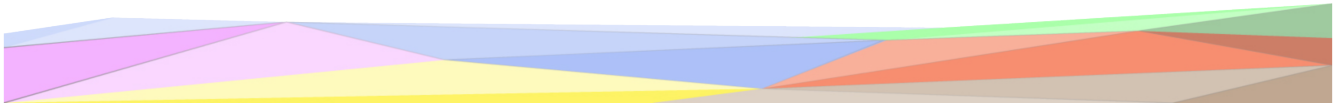
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 - Student life in Trento

Document Data:	Reference Persons:
October 30, 2024	Davide Cavicchini, Yesun-Erdene Jargalsaikhan



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 30, 2024	Davide Cavicchini, Yesun-Erdene Jargalsaikhan	Completed Phase1 - Purpose defiition

1 Introduction

In this project, we will develop a knowledge graph by integrating the following data source. The iLog app, which collects information from students studying at the University of Trento, Open Street Map, the open data resource for geographic data, and Trentino Trasporti for the public transportation data in Trento. The resulting knowledge graph will be utilized to facilitate informed decision-making.

2 Purpose Definition

This section introduces the purpose, domain of interest, scenarios and personas, competency questions, and concepts identification for the project.

2.1 Informal Purpose

The objective of this project is to build a knowledge graph that assists students in planning their trips from one location to another using public transportation in an efficient and comfortable manner. This tool aims to facilitate informed decision-making and enhance students' overall university experience. This will be achieved by integrating historical data on student commutes and activities, public transportation information and points of interest.

2.2 Domain of Interest

To focus our attention to what matters most and what are the distinctive features of the entities we will have to be able to handle.s In this section, we outline in which ways we ground our representations to the spatial-temporal domain in the world.

As per the spatial domain of the project, we are focusing our attention on the city of Trento, Italy. In particular, we are mostly interested in the commutes and daily activities of students around the city. For these reasons, we identify two main spatial domains of interest:

- Points of interest in Trento that students are interested in, such as bars, libraries
- Public transportation bus stop locations

While the main focus for the temporal domain is driven by the need to model the bus routes. For this reason, the distinction we have to make is between two classes:

- weekdays

- festive days

One interesting domain that we are interested in exploring is the state of crowdedness of places or routes and the emotional state of students. While this information can be grounded in the world by the location and the time at which it occurs, we also need to define their respective domains:

- The crowdedness state can take one of these four values: empty, low, high, full.
- The emotional state following the iLog data format is mapped to a scale from 0 to 5.

2.3 Personas & Scenarios definition

In this section, we introduce the personas and scenarios to ground the purpose on possible use-case of actual users of our knowledge graph.

2.3.1 Personas

To formalize the purpose of the project, we provide personas that covers various lifestyles among student which are useful to define diverse interactions with the knowledge graph.

Person 1 Alessia, a new international student, has recently started studying at the university.

Person 2 Paolo, a second-year master's student.

Person 3 Houda, an Erasmus student who wants to save up money

Person 4 Lucia, a student habit of dining in restaurants quite frequently

Person 5 Emanuele, a student who lives in San Bartolomeo student residence

2.3.2 Scenarios

For the persons we defined, we described some scenarios students could encounter during their university lifestyle in which our Knowledge Graph can assist for making decisions on planning.

1. **Social Interaction** - Alessia has recently moved to Italy and is excited to spend time with her new friends, exploring the city center of Trento, as she is eager to get to know the city.
2. **University Facilities** - Paolo is a second-year master's student at the University of Trento, currently working on his thesis. He wants to study in a quiet, uncrowded place, so he needs to choose one of the university's facilities.

3. **Daily life** - As an exchange student, Houda has started living in the city center and is planning to go grocery shopping. Since the atmosphere in supermarkets varies, he wants to choose the one that best suits his preferences.
4. **Dinner Place** - Regarding her dining habits, Lucia is looking for decent places to have dinner with her flatmate. While exploring restaurants she had both good and bad experiences with plates, so she doesn't want to choose a bad one.
5. **Personal Activity -** Emanuele is a professional athlete looking to have permanent training at the nearest sports facility to his student residence. A regular commute to the facility is an important part of his daily routine, so he needs to choose the one that will save him time.

2.4 Competency Questions

Following the paper on Big-Thick Data generation via reference and personal context unification, what we want to be able to answer are about personal-reference (PR) and reference-personal (RP) context questions. The following is a list of relevant questions for the scenarios and personas we defined which align with the purpose of our Knowledge Graph:

1. **P1-S1 PR**. Is public transport available to reach the destination?
2. **P1-S1**. Is the bus to the destination full, or does it have enough space for the group members?
3. **P1-S1**. How crowded are the social interaction locations in the city?
4. **P2-S2**. Which facility best fits the student's needs or has the least impact on their mood?
5. **P2-S2**. How crowded is BUC?
6. **P3-S3**. Which supermarket best meets the student's needs?
7. **P3-S3**. What was the student's mood when they were at the Coop supermarket?
8. **P3-S3**. What is the best route to the Coop supermarket?
9. **P3-S3**. How did I feel about the trip to the Coop supermarket when it was crowded?
10. **P4-S4**. Which restaurant served a meal that met the student's expectations?
11. **P5-S5**. Which sports facility is closest to the student's residence?
12. **P5-S5**. What is the best bus route to the sports facility?
13. **P5-S5**. What is the closest bus stop to reach the facility?

2.5 Concepts Identification

In this section, we try to come up with a mostly complete list and description of what type of concepts we are interested in modeling and which properties are fundamental for the Knowledge Graph. In the next section, we will use this information to guide the modeling of the ER diagram.

Scenarios	Personas	CQs	Entities	Properties	Focus
1-5	1-5	1-11	Student	student_id, name, current_position	Contextual
1, 5	1, 5	1, 13	Bus Stop	name, direction, time_table, location	Contextual
3, 5	2, 5	2, 8, 12	Bus route	number, festive, start_time	Contextual
4	4	10	Restaurant	civic_number, name, location	Common
2	2	3-5	University Facility	civic_number, name, location	Common
5	5	11	Sport Facility	civic_number, name, location	Common
1	1	3	Bar	civic_number, name, location	Common
5	5	11	Residence	intercom_name, civic_number, location	Contextual
3	3	6-9	Supermarket	civic_number, name, location	Common
2-4	2-4	4,7,9-10	Emotional state	time, duration, location, user, mood	Core
1-3	1-3	2-5,9	Crowdedness state	time, duration, location, density	Core

Table 1: Concepts Identification table

Table 1 reports the identified concepts and relates each to the specific competency question in which their use is required to correctly answer them. The most important concepts we identified for this part are the **Emotional State** and the **Crowdedness State** which will be used by our Knowledge Graph to address the majority of the reference-to-personal and personal-to-reference queries we aim to tackle.

2.6 ER modeling

Having defined the specific concept instances relevant to our identified competency questions, we now aim to formalize the newly acquired insights. To do so, we use the Entity-Relationship (ER) modeling to identify and define the entity types that we will need to manage and use. The resulting model is depicted in the image 1.

While modeling, we had to make several design choices.

One comes from the realization that, in all the questions we defined, we are only interested only in a subset of places that satisfy specific requirements, such as being suitable for hanging out with friends, having dinner, or working out. Therefore, it was unnecessary to model each place (e.g., restaurant, bar, sports facility) as a separate entity in the ER model. Instead, we collapsed them under a single **PoI** (Point of Interest) entity, differentiated by a *type* property.

Another important design choice is the need to separate the concepts of **Emotional state** and **Crowdedness state** used to describe a location from those used to describe a bus route. This is dictated by the No-Null and No-Repetition rules, which prevent us from collapsing these

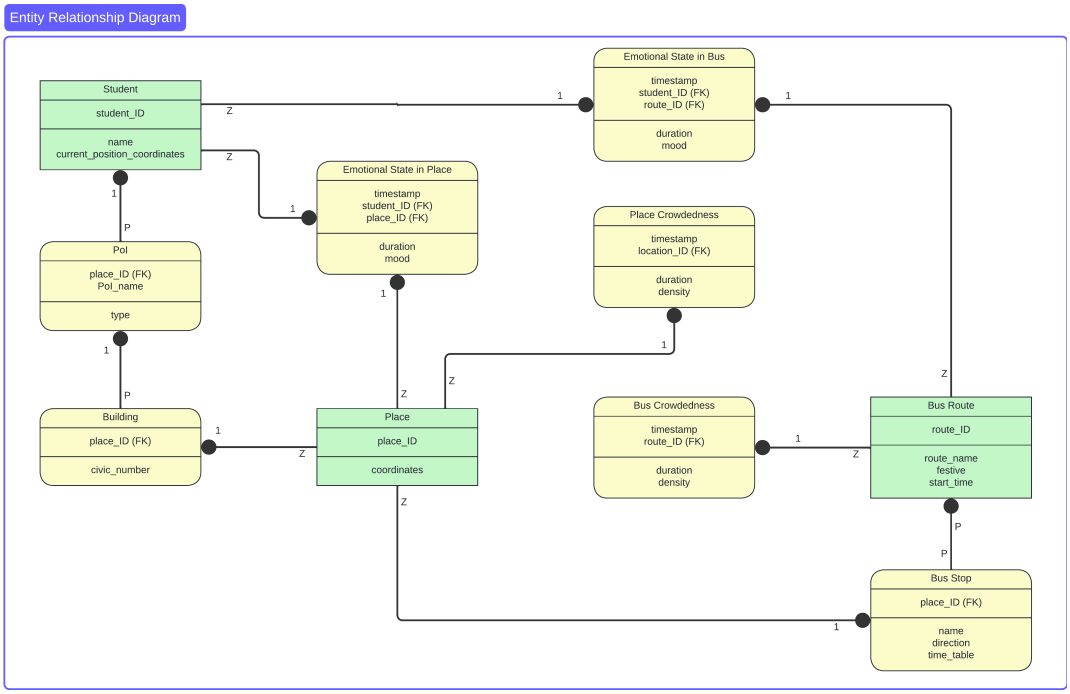


Figure 1: ER model

states into a single entity.