TRANSPORTATION

FACILITIES

Team - 2

Purpose

To create a reusable knowledge graph that accurately represents the network of bus routes in Ulaanbaatar city. This involves connecting various bus stops throughout the city to establish a visualization of the public transportation system. By connecting these routes and stops, the knowledge graph will serve as a valuable resource for urban planners, public transportation planners for decision making, optimizing routes, and enhancing overall efficiency and accessibility withing Ulaanbaatar’s transportation system.

Domain of Interest:

Current (2024) bus transportation system of Ulaanbaatar city.

Phase 1 - Scenarios

In life, people often encounter situations where they need to choose a bus route that goes to their desired destination. When this happens, finding the most useful route becomes particularly significant. We have defined 6 scenarios

Scenario 1:

An urban planner needs to optimize bus routes in a specific district of Ulaanbaatar to improve efficiency. They use the knowledge graph to analyze current route data, identify areas with high passenger demand, and propose adjustments to the bus network accordingly.

Scenario 2:

A tourist visiting Ulaanbaatar explores the city using public transportation. They use the knowledge graph to plan a route from their hotel to major tourist attractions.

Scenario 3:

A commuter who regularly catches bus from their home to work needs to plan their daily bus route. They use the knowledge graph to find the most convenient bus stops and departure times for their morning and evening commute.

Scenario 4:

An accessibility advocate wants to assess the overall accessibility of Ulaanbaatar's public transportation system for individuals with disabilities. They use the knowledge graph to identify wheelchair-accessible routes and stops, as well as areas where accessibility improvements are needed.

Scenario 5:

A commuter working late hours needs to catch a bus home before the service ends on weekdays. They rely on the knowledge graph to plan their route and departure time, ensuring they can reach their destination before the last bus of the night.

Scenario 6: A group of friends planning a weekend outing must catch a bus before the service ends on weekends. Using the knowledge graph, they coordinate their itinerary to ensure they can return home via public transportation within

Phase 1 - Persona

The characters involved are these:
Public transport companies: They run the bus and they plan the routes including paths.
Public transport passengers: They take the public bus from starting point to their destination.
Persona 1:
Bayaraa is an urban planner working for the transportation company of Ulaanbaatar. He is tasked with optimizing bus routes in a specific district to improve efficiency.
Persona 2:
John is a tourist visiting Ulaanbaatar to explore its cultural landmarks and attractions. John prefers using public transportation to immerse themselves in the local experience.
Persona 3:
Tsetseg is a regular commuter who relies on public transportation to travel from her home to work and back. She has a busy schedule and needs to plan her daily bus route efficiently to minimize commute time.
Persona 4:
Bold is an accessibility advocate who is passionate about promoting inclusive transportation options for individuals with disabilities. He advocates for improvements to Ulaanbaatar's public transportation system to ensure accessibility for all residents.
Persona 5:
Khuyagaa is a late-night commuter who works unconventional hours and relies on public transportation to commute home after finishing work late at night. He needs to catch a bus before the service ends on weekdays.
Persona 6:
Bat and his group of friends enjoy exploring different parts of Ulaanbaatar together on weekends. They prefer using public transportation for its convenience and affordability. However, they need to coordinate their itinerary to ensure

Competency questions

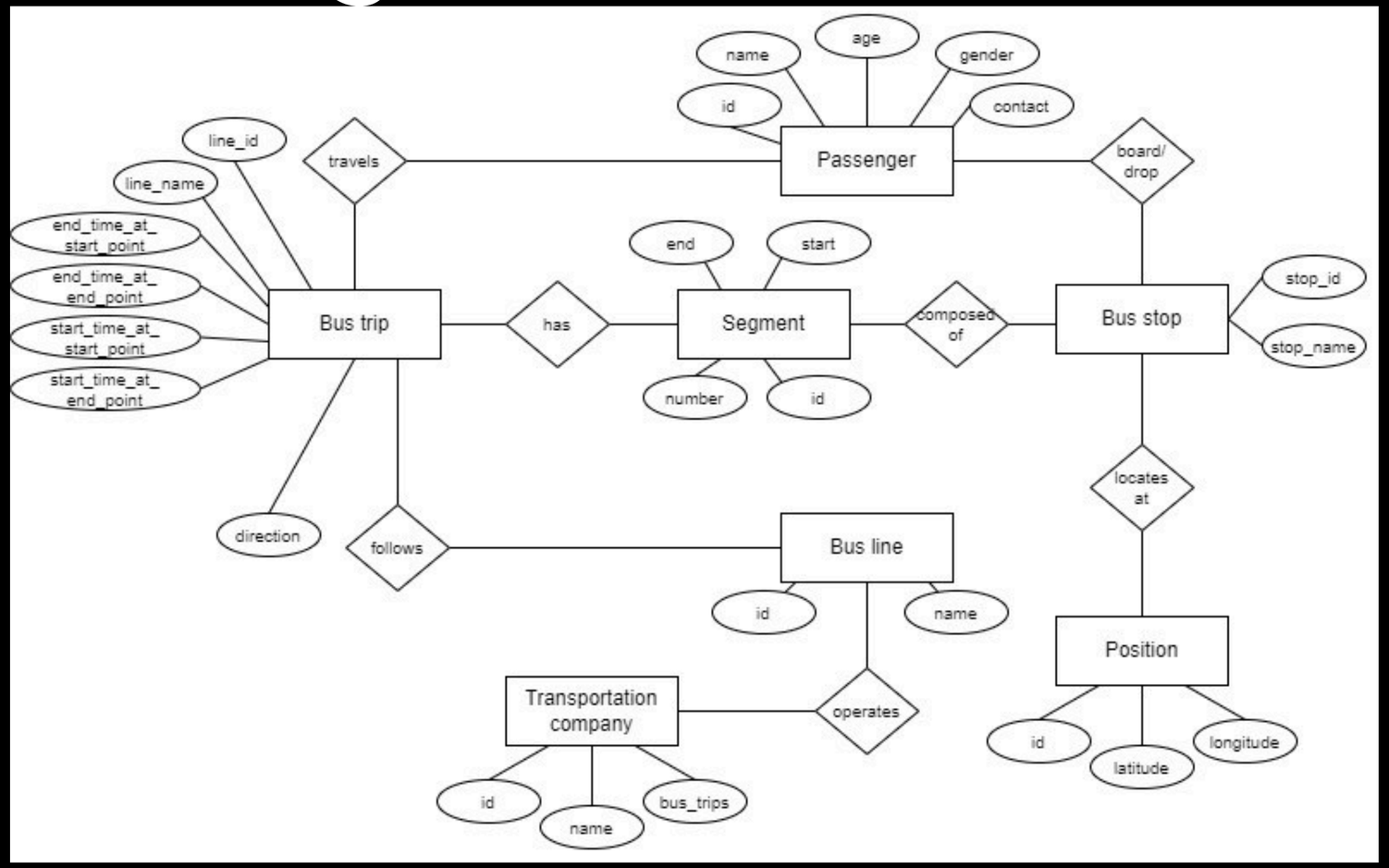
Competency Questions: a list of natural language questions. Each question defines a need (or query to the KG) that should be satisfied by the final KG. Each query refers to a Persona into a specific Scenario.

- How can the knowledge graph assist in identifying areas within a specific district of Ulaanbaatar with high passenger demand?
- Does the knowledge graph provide accurate information on bus routes and schedules relevant to John's exploration of Ulaanbaatar?
- Can Tsetseg easily access information on the most convenient bus stops and departure times for her morning and evening commute?
- What features does the knowledge graph offer to identify wheelchair-accessible routes and stops within Ulaanbaatar's public transportation system?
- Can Khuyagaa easily find information within the knowledge graph about bus routes and departure times before the service ends on weekdays?
- Can Khuyagaa determine whether he can reach home before the bus service stops?
- How can the knowledge graph assist in coordinating itineraries for groups like Bat and his friends to catch buses before the service ends on weekends?

Concepts identification

Scenarios	Personas	Competency Questions	Entities	Properties	Focus	Popularity
1,2,3,4,5,6	1,2,3,4,5,6,	1,2,3,4,5,6,7	Bus trip	id, start_point, end_point, distance, operating_hours, direction,accessibility,	Core	Contextual
2,3,4,5,6	2,3,4,5,6	3,4,7	Bus stop	id, name, address, nearby_buildings, served_routes	Core	Core
1,2,3,4,5,6	1,2,3,4,5,6	1,2,3,4,5,6,7	Bus line	id, name	Core	Core
2,3,4,5,6	2,3,4,5,6	3,4,7	Position	id, latitude, longitude	Common	Common
1,4	1,4	1,4	Transportation company	id, name, bus_lines	Common	Contextual
2,3,5,6	2,3,5,6	2,3,5,6,7	Passenger	id, name, age, gender, contact	Common	Contextual

ER diagram


name
age
gender
id
contact
Passenger
line_id
travels
board/
drop
line_name
end_time_at_
start_point
end_time_at_
end_point
start_time_at_
start_point
start_time_at_
end_point
Bus trip
end
start
has
Segment
composed
of
Bus stop
stop_id
stop_name
number
id
locates
at
direction
follows
Bus line
id
name
Position
Transportation
company
operates
id
latitude
longitude
id
name
bus_trips

Phase 2 - Information gathering

in this phase we have collected some datasets. And generate some datasets.

We run scripts to collect data from https://u-money.mn/

Collected datasets:
- bus lines of Ulaanbaatar
- bus trip (routes)

Generated datasets:
- passenger
- transportation company

From OSM:
- bus stop
- position (of bus stops)

We decided on CSV format for our datasets, so we performed some format alignment and created new CSV datasets from the JSON files we scraped on producer side.

Phase 3 - Language definition

1. **Concept selection**

We select purpose-specific concepts from the ER model and the Purpose Formalization Sheet.

2. Alignment

We check if a concept is available from the OSM page or the UKC. If the concept is there, we collect formal concept definitions. If the concept is not there, we define the new concepts formally.

Phase 3 - Language definition

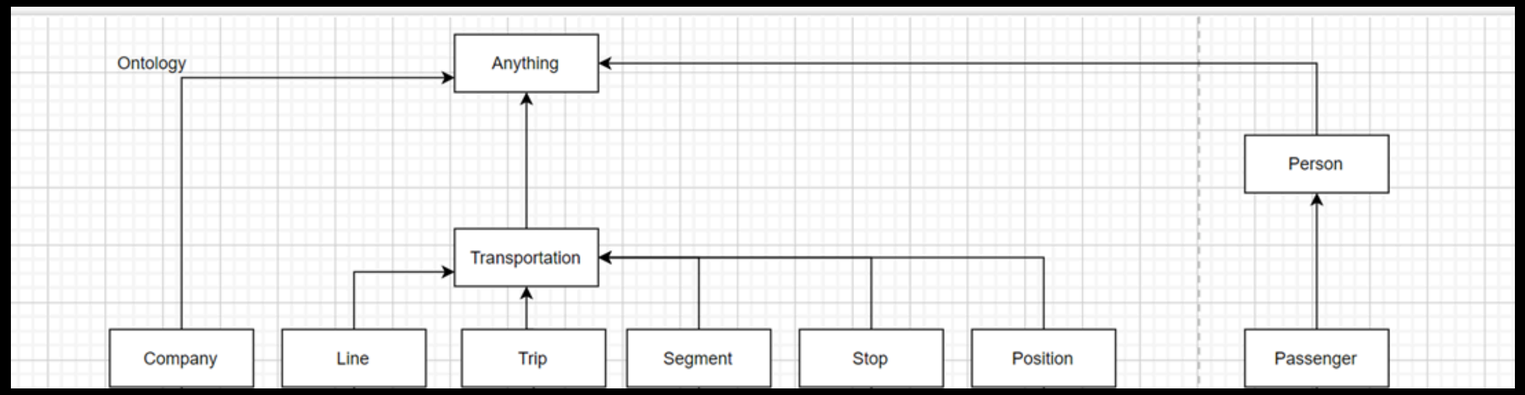
3. Language resources

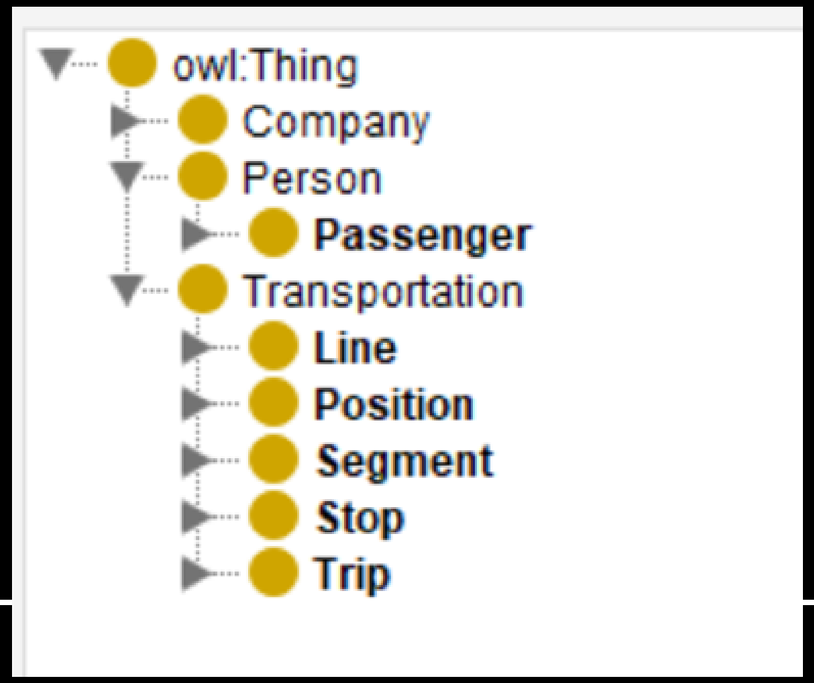
Word-mn	Gloss-mn
автобусны зогсоол	Жижиг автобусны буудал.
автобусны маршрут	Автобусны дагаж явах зам (төмөр зам биш) ба бусад тээврийн хэрэгсэлд тохиромжгүй.
зогсоолын байршил	Гудамжны нийтийн тээврийн хэрэгсэл зогсдог байрлал.
автобусны чиглэл	зорчигч тээврийн автобус тогтмол явдаг мартшрутын нэр
нийтийн тээврийн компани	Тодорхой газар зүйн бүсэд тогтоосон маршрутын дагуу нийтийн тээврийн хэрэгсүүдийг удирдан, зохицуулж зорчигчдыг тээвэрлэх үүрэг бүхий байгууллага.
зорчигч	Тээврийн (завь, автобус, машин, онгоц, галт тэрэг зэрэг) хэрэгслийг жолоодохгүйгээр сууж байгаа аялагч.

Word-en	Gloss-en	Word-mn
bus_stop	A small bus stop.	автобусны зогсоол
bus_guideway	A busway where the vehicle guided by the way (though not a railway) and is not suitable for other traffic.	автобусны маршрут
stop_position	The position on the street or rails where a public transport vehicle stops.	зогсоолын байршил
bus_route	the route regularly followed by a passenger bus	автобусны чиглэл
transportation company	An organization responsible for operating and managing buses, facilitating the transportation of passengers along designated routes within a specified geographical area.	нийтийн тээврийн компани
passenger	A traveler riding in a vehicle (a boat or bus or car or plane or train etc) who is not operating it	зорчигч

Phase 4 - Knowledge definition

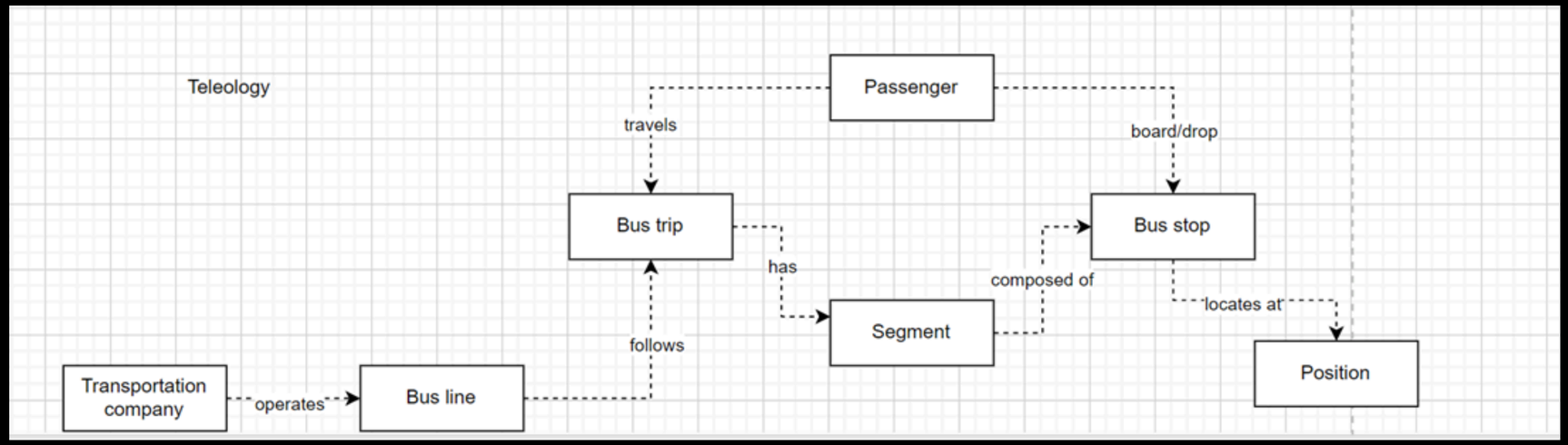
Lightweight ontology
In lightweight ontology modeling, we used top down method, by classifying datasets by hierarchical order, in which child node is a subset of parent superset. We also used some concepts from schema.org.





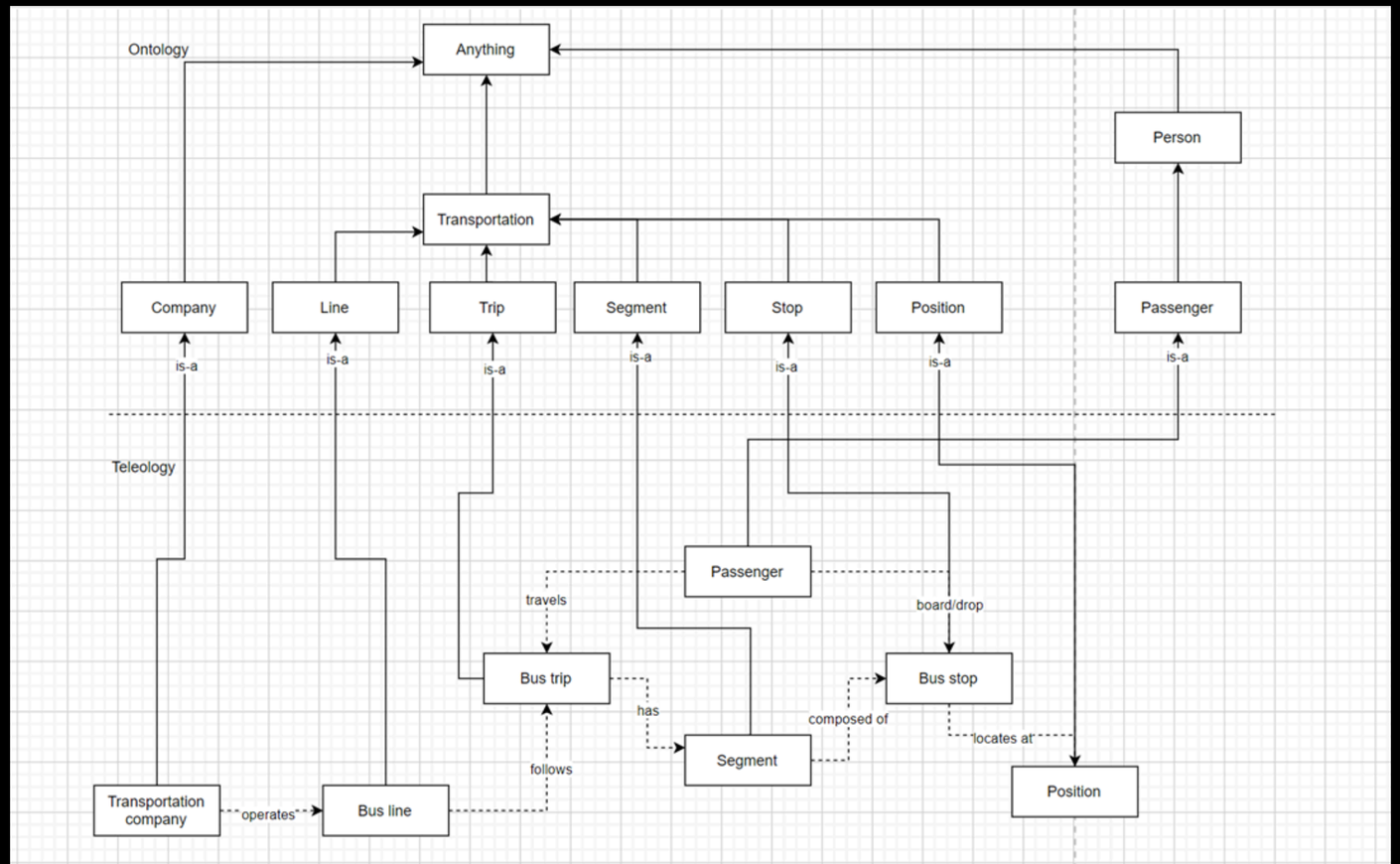
Phase 4 - Knowledge definition

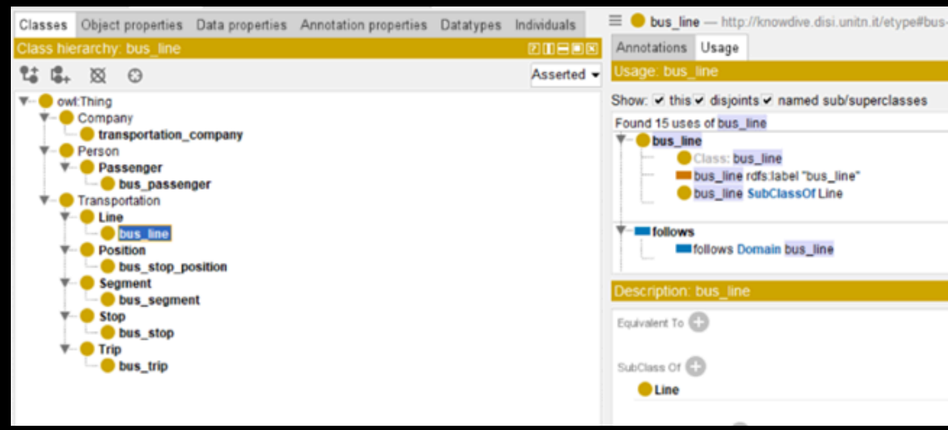
Teleology modeling
In teleology modeling, we used bottom-up method, starting with a tabulated list of Competency Questions (CQs) which encode the etypes and properties relevant to be modelled.



Phase 4 - Knowledge definition

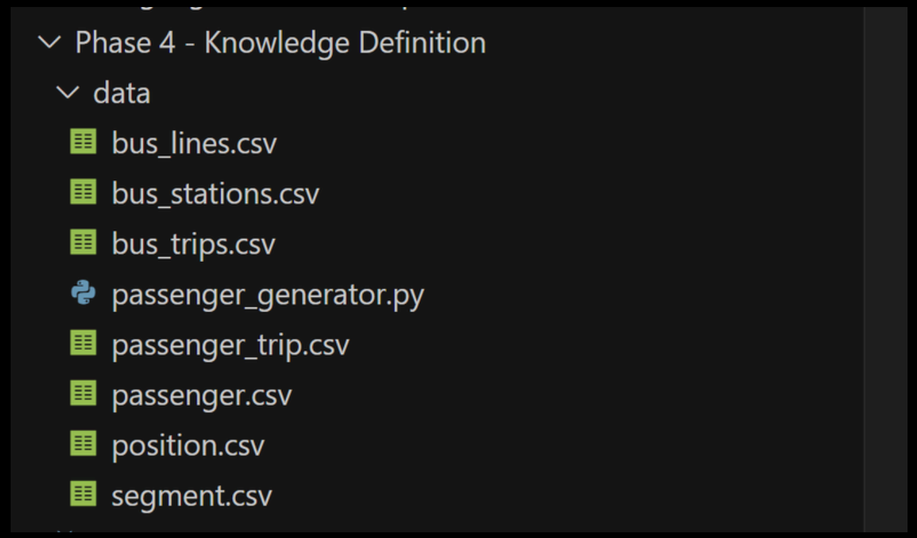
Teleontology modeling
In teleontology modelling, we used middle-out method, by semantically aligning teleology to the lightweight ontology. Each concept on the teleology is added as a child (via IS-A) to theair related general concept in the lightweight ontology.





Phase 4 - Knowledge definition

Dataset alignment
On the data layer, aligning the dataset previously collected, cleaned and formatted, with the modelling choices operated in the above parallel knowledge layer activity. We aligned datasets and data types regarding our teleontology.
Cleaned, formatted datas: https://github.com/belgeee/Transportation-Facilities-KGE-Project/tree/0446cdcd6884559d6a745ca49ad79c4d95271a7f/Phase%204%20-%20Knowledge%20Definition

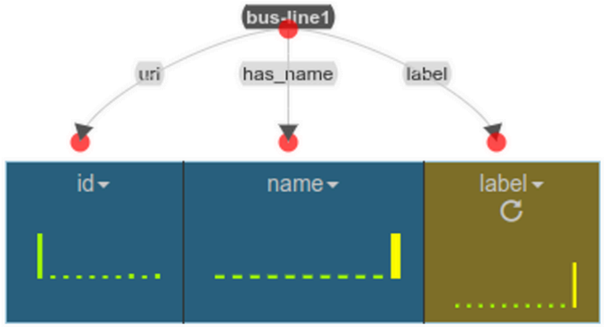


Phase 5 - Data Definition

- We use teleontology and cleaned, formatted dataset to generate knowledge graph.
- Import ontology and datasets into Karma tool. We perform entity mapping using Karma tool.
- Import generated ttl files into GraphDB knowledge graph visualizer tool to visualize our final knowledge graph.

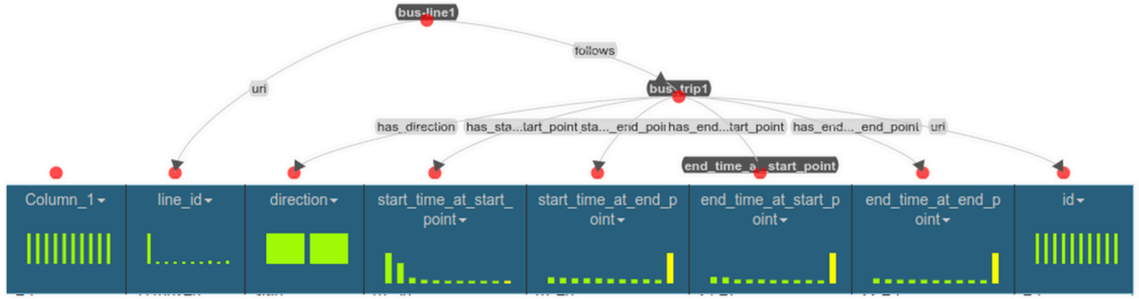
Phase 5 - Data Definition

8.1 Bus line dataset



Entities: bus_line

8.2 Bus trip dataset

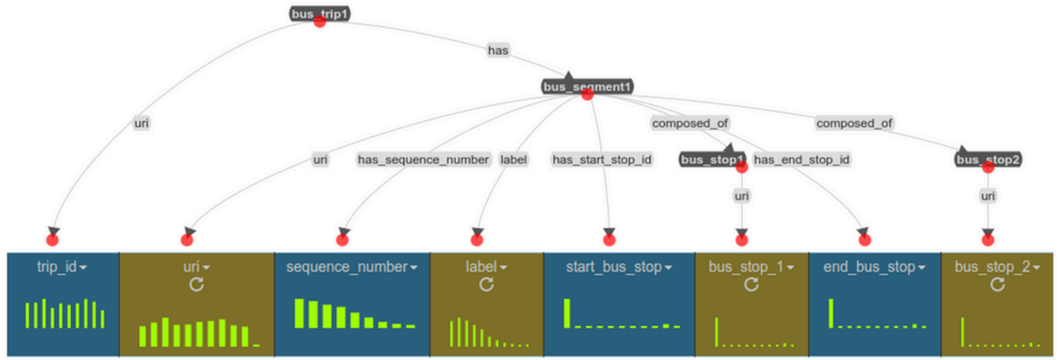


Entities: bus_line, bus_trip

Relationship: bus_line follows bus_trip

Phase 5 - Data Definition

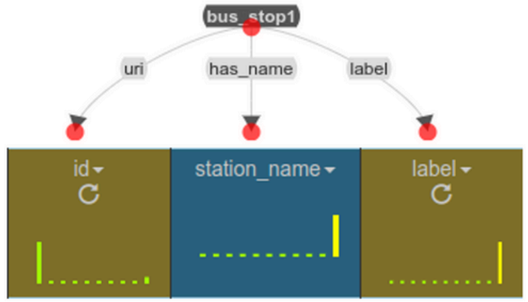
8.3 Segment dataset



Entities: bus_trip, bus_segment, bus-stop

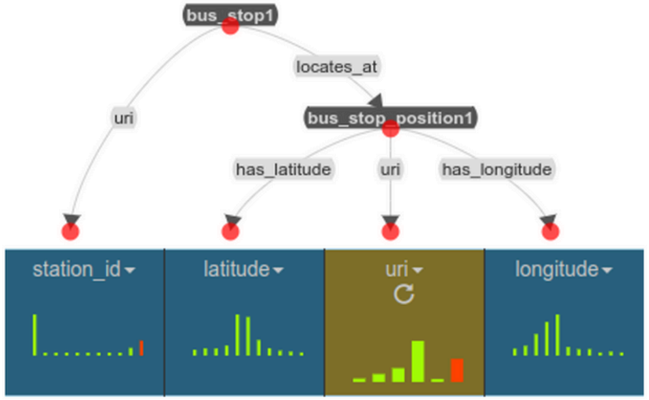
Relationship: bus_trip has bus_segment, bus_segment composed of bus stop

8.4 Bus stop dataset



Entities: bus_stop

8.5 Position dataset



Entities: bus_stop, bus_stop_position

Relationship: bus_stop locates at bus-stop_position

Evaluation

9.1.1 Teleontology

In the table below, there is a summary that takes into account the total number of etypes, object properties, and data properties, used for the calculation of the coverage

	Instances Count
Etypes	7
Object Properties	7
Data Properties	26

Evaluation

The coverage of the etypes, object properties, and data properties is calculated as follows. For example, for the etype, given a set of (CQE), the etype coverage (CovE) of the Teleontology (T) is:

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

where CQE is the number of etypes extracted from the CQs, and TE is the number of etypes of the Teleontology.

	Etypes CovE	Object Properties CovOP	Data Properties CovDP
Total identified from CQs	5	5	23
Total defined for the project	7 100%	7 100%	26 100%

Teleontology vs Competency Questions Coverage

- This table shows that for each criteria, the final Teleontology defines more etypes, data and object properties. This is due to the fact that during the initial phases of this project, the specific knowledge design choices and needs were not complete. They have been refined during the development of the project which lead to defining a better and complete knowledge structure to fulfill the purpose.

Evaluation

9.1.3 Teleontology vs Reference Ontologies

The coverage of the etypes (CovE), object properties (CovOP), and data properties (CovDP) is calculated as follows. For example, for the etype, given a set of (RO), the etype coverage (CovE) of the Teleontology (T) is:

$$Cov_E(RO_E) = \frac{|RO_E \cap T_E|}{|RO_E|}$$

where ROe is the number of etypes extracted from the ROs, and TE is the number of etypes of the Teleontology. Below, there is a table with the final evaluation, considering the etypes, object properties, and data properties coverage.

	Etypes CovE	Object Properties CovOP	Data Properties CovDP
Total in the ontology	7	7	23
Total reused in the project	7 100%	7 100%	23 100%

Data layer evaluation

The connectivity of a KG can be evaluated over two dimensions:

1. entity connectivity which evaluates the grades of connection between the different entities in the KG;
2. property connectivity which evaluates the grades of connection between each single KG's entity and its properties values.

	bus trip	segment	passenger	bus stop	position	bus line	transportation company
bus trip	0		0			0	
segment	0	0		0			
passenger			0				
bus stop		0	0	0	0		
position				0	0		
bus line	0					0	0
transportation company						0	0

The entity and property connectivity can be calculated by using the connectivity matrix, as represented in the table below.

Open issues

Incomplete Data Collection:

- Problem: Limited access to some different data sources to connect them.
- Solution: Allocate more time for data collection

Passenger Data Creation:

- Problem: Inconsistent data formats and missing information.
- Solution: We created passenger data based on personas ourselves.

Data Quality and Consistency:

- Problem: Variations in data quality across sources.
- Solution: Continuous data quality monitoring and validation.

THANK YOU

THAT'S MY PORTFOLIO SO FAR.