Automated Mapping for Semantic-based Conversion of Transportation Data Formats

Marjan Hosseini, Safia Kalwar, Matteo Rossi, and Mersedeh Sadeghi

Sem4Tra Karlsruhe
9th September 2019

Presented by: Safia Kalwar
Outline

• Problem Statement
• Objective
• Proposed Solution/Methodology
• Future Work
Problem Statement

**Objective:** "Mobility as Service"
User can build door to door trips through single entry point.

**Problem:**
- Divergence of transportation standards
- Heterogeneity of data representations, formats and models.
Research Objective

• Facilitate interoperability in transport domain

Targets and Goals

• Establish semantic interoperability rather than only syntactic

• Use semantic mapping to create communication between different systems

• Reduce the need of adopting unified data model

• Automate conversion process
Methodology

- Lifting: Transforming from source format to intermediate representation based on reference ontology.
- Lowering: Transforming from intermediate representation to target format

- **Objective:** Making annotation process automatic using machine-learning to increase performance and efficiency of system.
Method

**Source Format**

- PreBooking
  - Traveler name
  - Source
  - Destination
  - Date

**Reference Transform format**

- PreRegister
  - Passenger name
  - Trip
  - Time
  - Start Place
  - Stop Place
**Word2Vec**

**t:** 1. 2. 3. 4. 5. 6. 7.

**w:** Bus operates between central station to North

1. **Input Vector**

   - 1
   - 0
   - 0
   - 0
   - 0
   - 0
   - 0

2. **Pre-trained Model**

3. **Output Vector**

   - 14
   - 12
   - 10
   - 2
   - 11
   - 3
   - 15

**Sun**

**Shuttle**

**Bus**

---

Special Project Initiative (SPI)

Grant Agreement No: 826172

POLITECNICO MILANO 1863
Work Flow

Source Format

Concept Identification

<Route>

Trained Word2Vec Model

<Google News>

most similar words to 'Route' are:
1. routes
2. shortest_route
3. Path

Final Output

Path rank = 3/100
Method Assumptions

Assumption 1: The language is both sides of the mapping is English

Assumption 2: For each concept in source system there is at least one concept in target system

Assumption 3: One to one relationship between source and target data formats

Assumption 4: All concepts exist in word2vec model
Future work/Conclusion

- This model is based on Word2Vec Model.
- In a scenario for common terms this pre trained model is performing well.

Challenges
- Compound words might not exist in word2Vec model
  E.g: pre_booking, stop_place
- Due to absence of some terms in source/reference ontology mapping terms in target might face some difficulty.

Future work
- To validate the method possible approach is to prepare two datasets with different data formats containing equivalent instances.
Thank you for your Time

Any Questions?