Foundations of Semantic Knowledge Graphs

Prof. Dr. Stefan Linus Zander

Data Graphs

What you will learn in this Lecture

We first discuss a **selection of graph-structured data models** that are commonly used in practice to represent data graphs.

We then discuss the primitives that form the basis of **graph query languages** used to interrogate such data graphs.

Didactical Model

- At the core, we start with a mathematical definition and representation of graphs
- we then use extend these definition towards a graph-based model in CS
- We introduce data graphs and discuss different types and models
- We then add schema, identity, and context to transform data graphs into knowledge graphs
- We add semantics and ontologies to build semantic knowledge graphs and introduce a representation framework for SKGs

Didactical Model

This lecture uses a continuum of increasing "intelligence" as a didactical model to explain the the additions needed for the transition of simple data graphs towards semantic knowledge graphs.

A Graph-based Data Representation provides a number of Advantages over other, traditional Data Models

Why there is a need for data graphs

- Requirements and peculiarities of domains of interest are often not fully known at the beginning
- Additional features might needed
  - multi-language support
  - multiple names and descriptions
  - multiple types
  - incomplete or missing information might harm schema compliance
- Schemas might evolve over time as the complexity of the domain of interest increases
- Schema adaptations are expensive and require lots of testing and validation
  - remodelling, reloading, and reindexing of data needed
- Diversity

Example

Initial schema for an event database (table with 5 columns)

$$Event(\underline{name}, venue, type, \underline{start}, end)$$

A more flexible and elaborated schema solution

$$EventName(id, name), EventStart(id, start), EventEnd(id, end), EventVenue(id, venue), EventType(id, type)$$

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi$$

Remodelling often results in binary relations between entities --> close to modelling a graph

Using a data-graph model removes the necessity of an upfront schema and facilitates refinements

Overview of the different Types of Data Graphs

1. Directed Edge-Labelled Graphs
2. Heterogeneous Graphs
3. Property Graphs
4. Graph Datasets
5. Other Graph Data Models
6. Graph Stores

Directed Edge-Labelled Graphs

Definition A directed edge-labelled graph is a tuple $G = (V, E, L)$, where $V \subseteq \mathbf{Con}$ is a set of nodes, $L \subseteq \mathbf{Con}$ is a set of edge labels, and $E = V \times L \times V$ is a set of edges.

Source: https://kgbook.org/

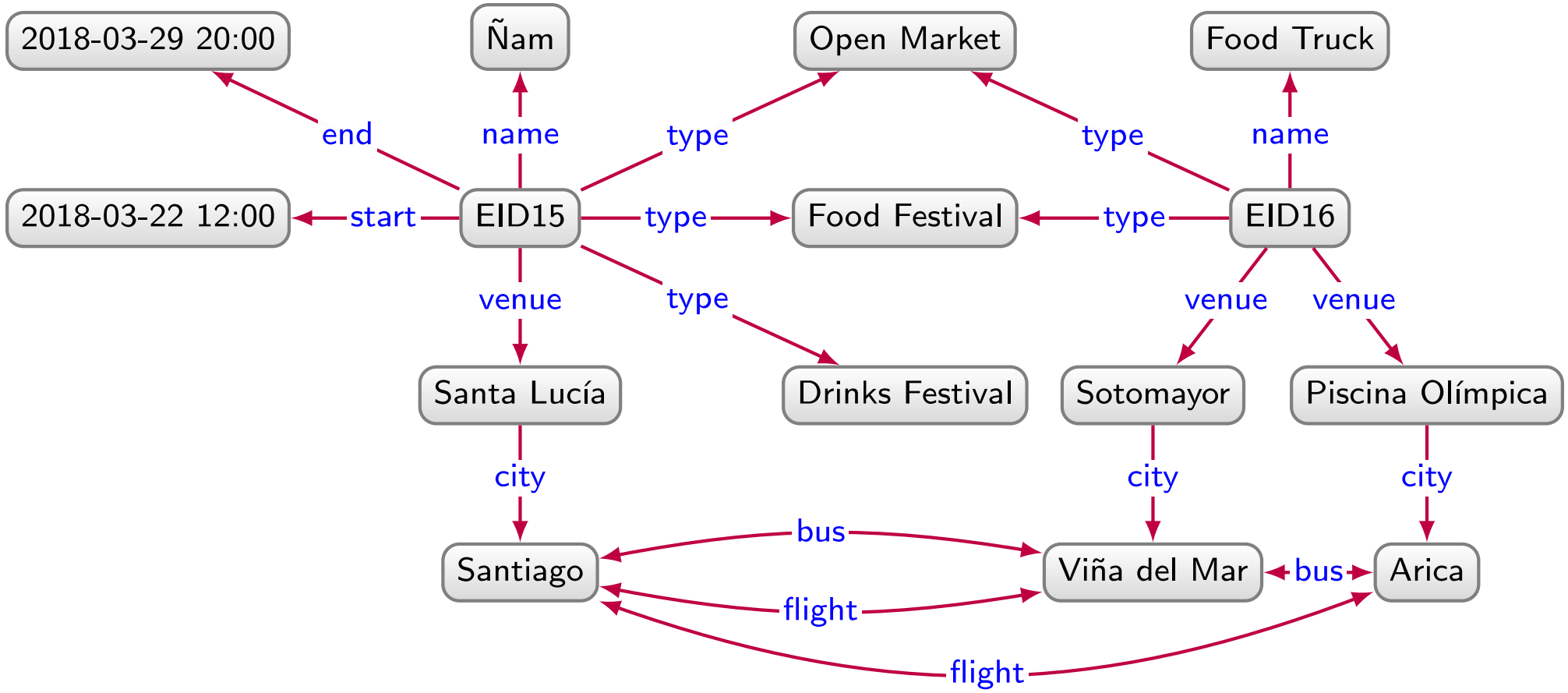
Characteristics

A directed edge-labelled graph (sometimes known as a multi-relational graph) is defined as a set of nodes – like Santiago, Arica, EID16, 2018-03-22 12:00 (with types?)– and a set of directed labelled edges between those nodes, like Santa Lucía–city➛Santiago.

Features

...

Example


2018-03-29 20:00
Ñam
Open Market
Food Truck
end
name
type
type
name
2018-03-22 12:00
start
EID15
type
Food Festival
type
EID16
venue
type
venue
venue
Santa Lucía
Drinks Festival
Sotomayor
Piscina Olímpica
city
city
city
bus
Santiago
Viña del Mar
bus
Arica
flight
flight

The foundation of any knowledge graphs rests on the principle of first applying a graph abstraction to data

...