deti universidade de aveiro
departamento de electrónica,
telecomunicações e informática

Knowledge Representation

Semantic Expression

Knowledge Graphs (KG)

Formalization

The Triple (i)

- Previously, it was found that a sufficiently generic data model to accommodate both data and its description (*metadata*) is the use of the key-value model associated with identifiers.
  - Identifiers represent the various entities present in the data.
  - Keys and their values characterize these entities.

- This data model consists of three fields and is called a Triple:
  - (ID – Property – Value)
  - It's the most elementary and fundamental building block to implement semantic representations.

The Triple (ii)

- The Triple data model is formally defined by the following components, that codifies a semantic statement:
  - (Subject – Predicate – Object)
- Subject
  - It's an entity, identified by an ID
  - Something for what we have a conceptual class
    - Exs: persons, locations, concrete objects and some less concrete things, like time periods and simple ideas

The Triple (iii)

- Predicate
  - Represents an entity's property or a relationship between two entities
  - Properties
    - Exs: name, birthdate, email, etc.
  - Relationships
    - Exs: father, mother, brother, sister, friend, owns, etc.

The Triple (iv)

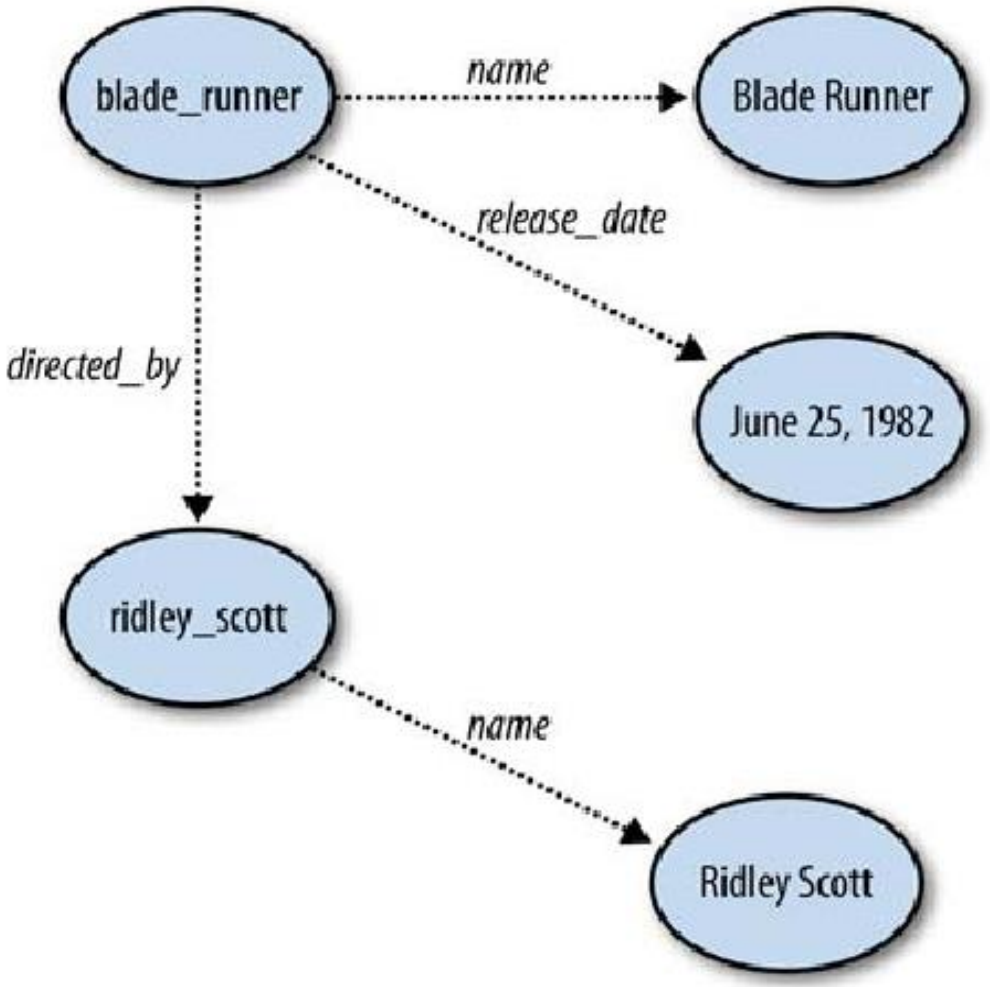
- Object
  - Represents a property's value or another entity.
  - Properties' values
    - Exs: text, numbers
  - Entities
    - Exs: entity, like the subject, identified by an ID

Knowledge Graphs (KG)

- Using the same entities, as subjects or objects, in multiple and different triples, leads to:
  - Create a Web of Triples, representing the relationships between different entities
  - In sum, create a Knowledge Graph (KG), in the form of an oriented graph

Example: Movies (i)

- Use of a knowledge graph to represent data about the movie “Blade Runner”



Example: Movies (ii)

- Writing the KG in triples:

  (blade_runner name “Blade Runner”)

  (blade_runner release_date “1982/06/25”)

  (blade_runner directed_by ridley_scott)

  (ridley_scott name “Ridley Scott”)

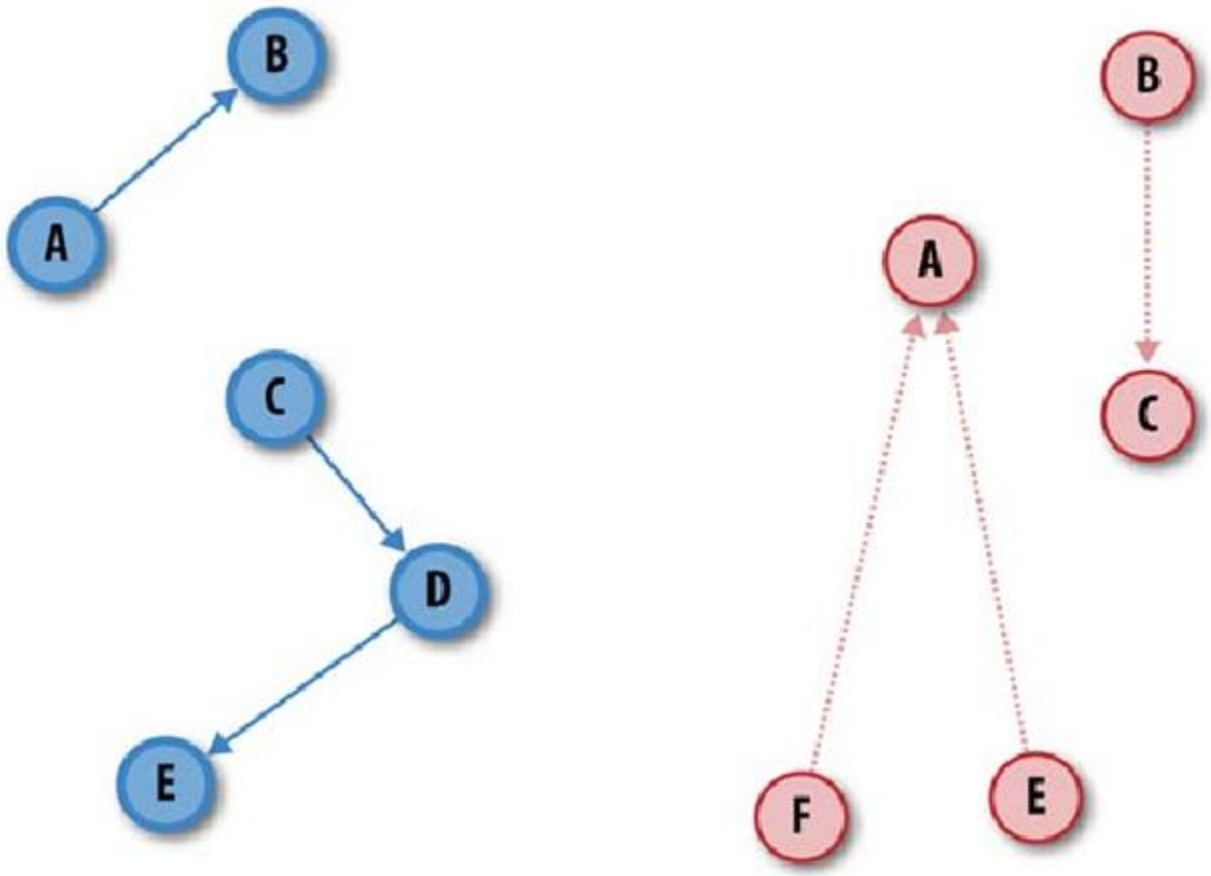
- Meaning:
  - blade_runner is the movie’s ID and ridley_scoot is the director’s ID
  - directed_by states a relationship between the 2 entities
  - name and release_date are entities’ properties – property name is reused by both

Expanding Knowledge

- Using triples (graphs) to convey knowledge allows for its expansion by simply adding new triples or merging existing graphs.
- It only needs that subjects and objects have a consistent system of identifiers.
  - The same entities are always identified by the same identifier.
- And if a triple exists in multiple graphs, it must be merged into a single one.

Expanding Knowledge

- Example of separated graphs:



Expanding Knowledge

- Example of merged graphs: