Mining Logical Rules

Omar Gutiérrez

June 7, 2018

Introduction

- A knowledge base (KB) is filled with facts.
  - Every fact is represented by a **relation** between a *subject* and *object*.
  - For example: *Homer* **isHusbandOf** *Marge*.
- Some popular KBs are **DBpedia**, **YAGO**, **Wikidata**, **Freebase**, etc.

How they are designed?

- The Resource Description Framework (RDF) is the most popular format for the semantic KBs.
- Every fact in KBs is known as triple.

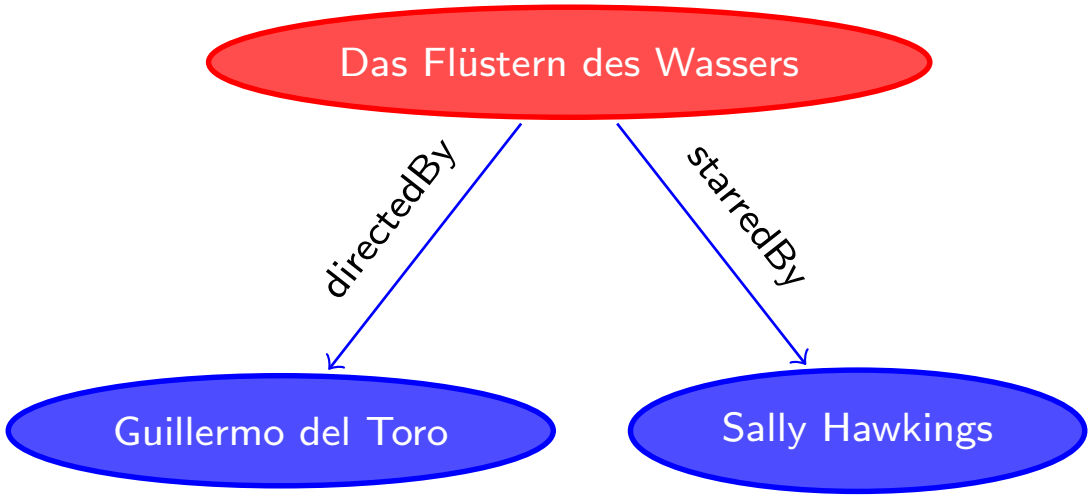

Figure 1: Very simple example of RDF graph

Open and Closed World Assumption

- **Open World Assumption** (OWA) is assumed in relational databases.
- **Closed World Assumption** (CWA) is assumed in semantic knowledge bases.
  - Semantic KBs do not contain negative evidence :(

Mining facts

- Let's say that we know the next facts
  - <Homer> **isHusbandOf** <Marge>
  - <Homer> **wasBornIn** <United States>
  - ...
  - <Marge> **wasBornIn** <?>
- Another example:
  - <Bart> **isSonOf** <Homer>
  - <Lisa> **isDaughterOf** <Homer>
  - ...
  - <Bart> **isBrotherOf** <?>

Conclusions

- Why this?
  - KBs, such as, DBpedia or Wikidata are *always growing*, we need to be sure that the incoming new facts are reliable.
  - We can find data modeled as KBs in several areas:
    - Medicine
    - Bioinformatics
    - Publishing
  - Because is interesting ;)