Welcome to Knowledge Graphs

- Knowledge base

knowledge base

noun

1. a store of information or data that is available to draw on.
2. the underlying set of facts, assumptions, and rules which a computer system has available to solve a problem.

= technology used to store complex structured/unstructured **info**

- stores answers to questions or solutions to problems

- Vs. Database:
  collection of data representing facts

Knowledge Graphs Intro

- What is a knowledge graph:
  = a **network** of real-world entities (objects, events, concepts) and their **relationships**. Commonly stored in a graph database and visualized as a graph structure (a.k.a. semantic network)

- What is the Semantic Web:
  = an extension of the current web in which information is given **well-defined meaning**, better enabling computers and people to work in cooperation.

Semantic Web Technologies

- Semantic Web technologies enable people to:
  - Create data stores on the Web
  - Build vocabularies
  - Write rules for handling data
- We will be looking at RDF, SPARQL and OWL

- We want to give the data further meaning!

WWW was/is? tailored to…?

- AAA principle = Anybody can write Anything about Any topic
- Comprised of Web pages(++) – What are they?
- No central point of control
- No central repository

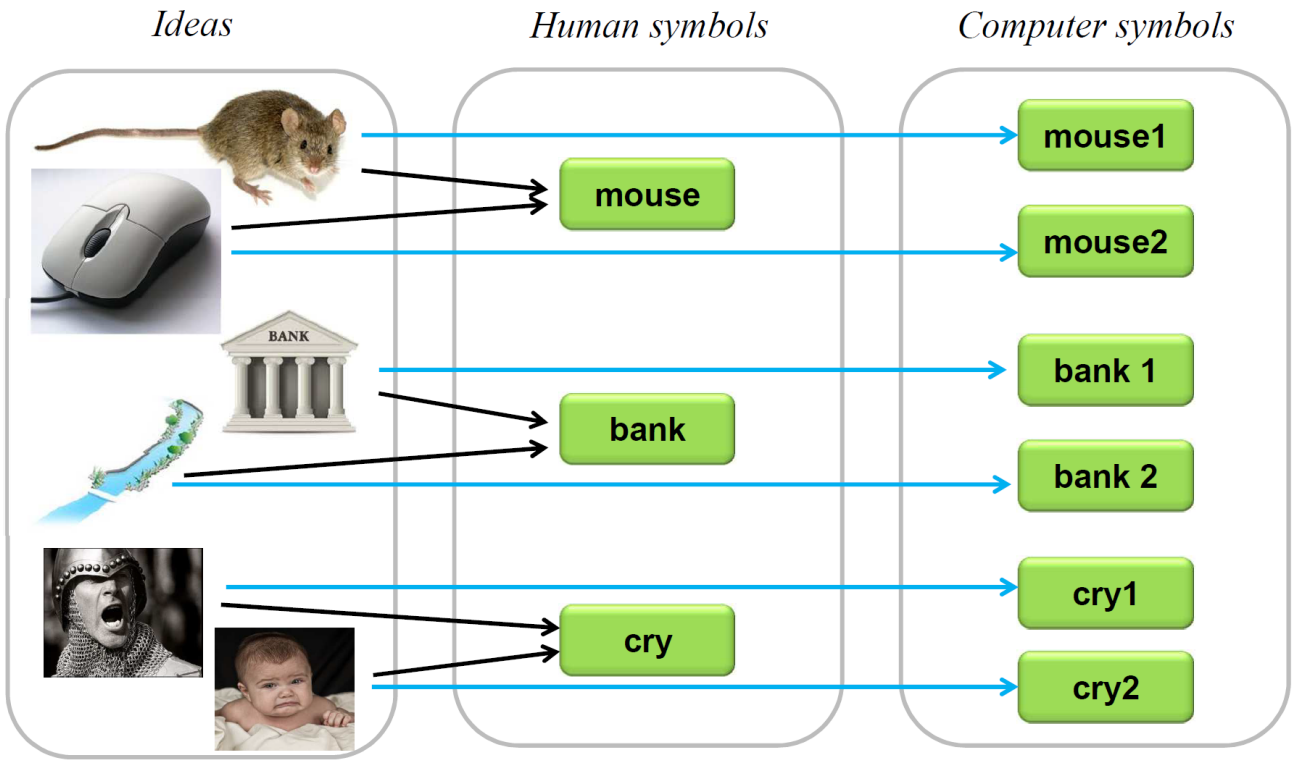
- Web page is a document – can you see any problems?
  (in other words, can you identify any limits?)

Ambiguity in natural language

- Different words/expressions for the same context
  => have we heard of a synonym or a metaphor before?

- This is the issue for machines parsing strings
- Main issue for information retrieval
- Are people aware of any solutions/workarounds?

A couple of things difficult for a machine



- Disambiguating

Non-unique Naming Assumption

- Train
  = working out
  = public transportation
  = machine learning

- It is evident that you can have multiple names/definitions for a thing
- The computer needs to know this,
  and *by needs to know*, we mean **EXPLICITLY**
- This is handled with URIs

RDF

- The Resource Description Framework (RDF) is the **base language** of the Semantic Web. It's a language used for describing data, metadata, and even other data languages.
- RDF uses a **graph data format**, in contrast to relational data formats (such as most databases) and hierarchical data formats (such as XML).
- Any data model or data language that uses RDF is a part of the Semantic Web.

RDF & URIs

- The RDF graph is based on the idea that every data item should have a unique Web identifier
  called a URI (Uniform Resource Identifier),

  and that

- **every** data item *can be* connected to every other item.

- A URI is different from a URL (Uniform Resource Locator) in that a URI may refer to either a Web name or a location; a URL may refer only to actual Web locations

Resource Description Framework

- A model for representing **metadata**
- A model for encoding **semantic relationships** between items of data so that these relationships can be interpreted computationally.
- A general method to decompose knowledge into **small pieces** with **rules about the meaning** of those pieces.
- A method to describe facts in a short form.
- Everything is a Resource in the form of a URI*

*there tends to be an(or more) exception(s)
– think about it and let me if/when it comes up

Triples (Conceptual)

- All the data in RDF is described in statements/triples:

  subject – predicate – object

From the previous example

per145 hasID “145”

per145 hasFirstName “Albert”

per145 hasSurname “Einstein”

per145 livesIn NY

Open World Assumption

- “Any piece of information can come at any point in the future”
- Translation: You live, you learn. It is ok not to know everything

Open World Assumption Example

- Heraklion and Chania are in Crete. Athens is not in Crete.

- Is Heraklion in Crete?
- Is Athens in Crete?
- Is Thessaloniki in Crete?