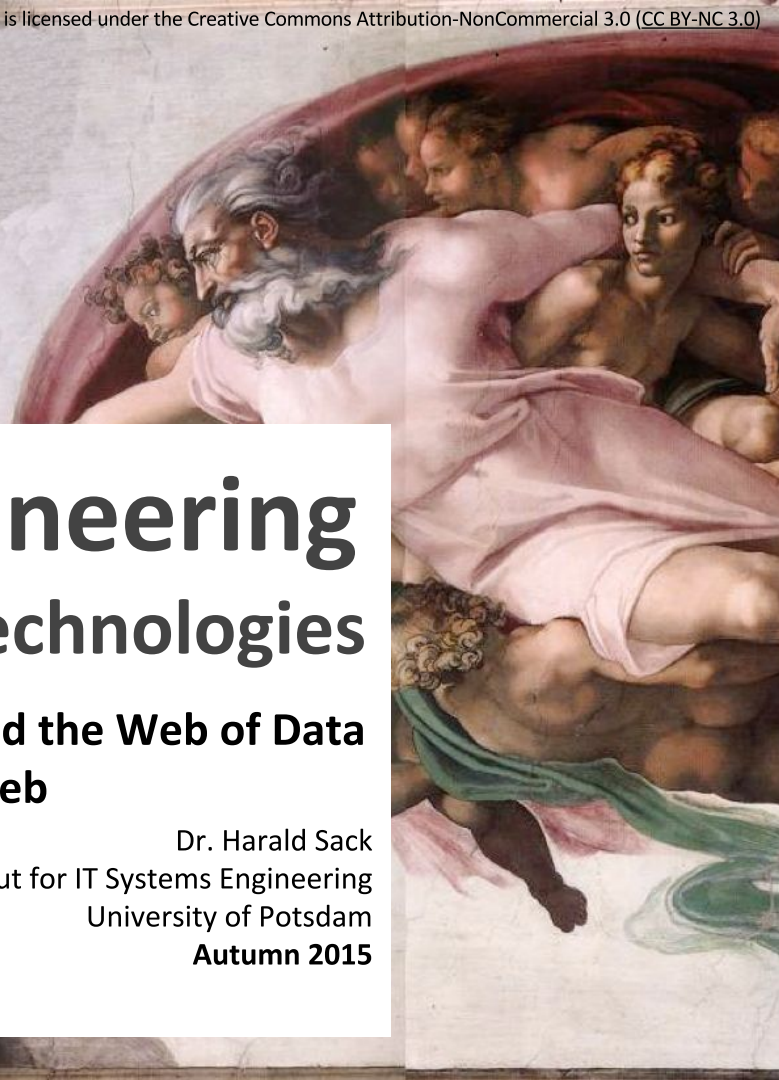
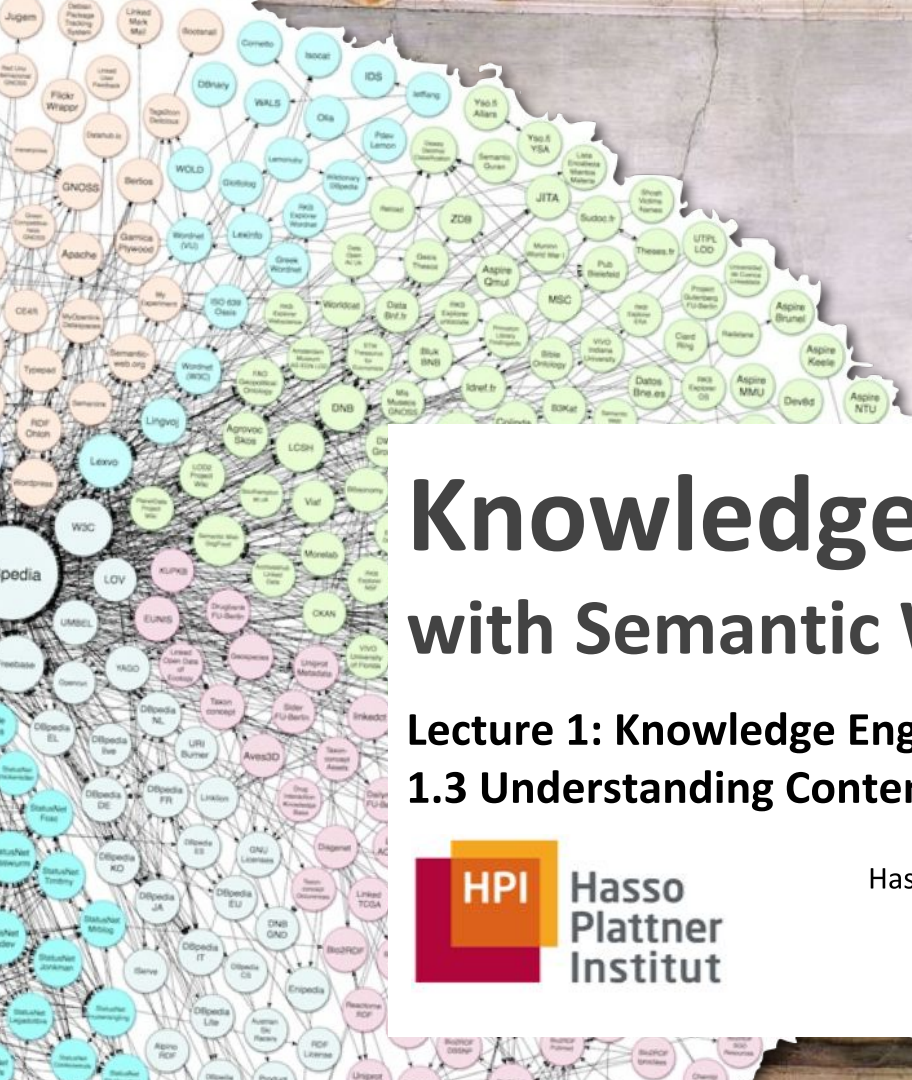




Knowledge Engineering with Semantic Web Technologies

Lecture 1: Knowledge Engineering and the Web of Data 1.3 Understanding Content on the Web



Dr. Harald Sack
Hasso-Plattner-Institut for IT Systems Engineering
University of Potsdam
Autumn 2015

Third Generation: The Web of Data

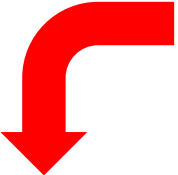
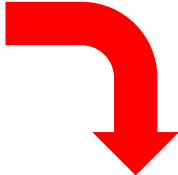
Data Centered Processing

- The **Web of Data** is an upgrade of the Web of Documents
- It's the Web as a huge decentralised database (knowledge base) of machine-accessible data

„The web of **human-readable document** is being merged with a web of **machine understandable data**. The potential of the mixture of humans and machines working together and communication through the web could be immense.“

Tim Berners-Lee, [The World Wide Web: A very short personal history](#), May 1998

From the World Wide Web to the Web of Data

- 
- Precondition:
 - Content can be **read and interpreted correctly**
(= "understood") by machines
- 

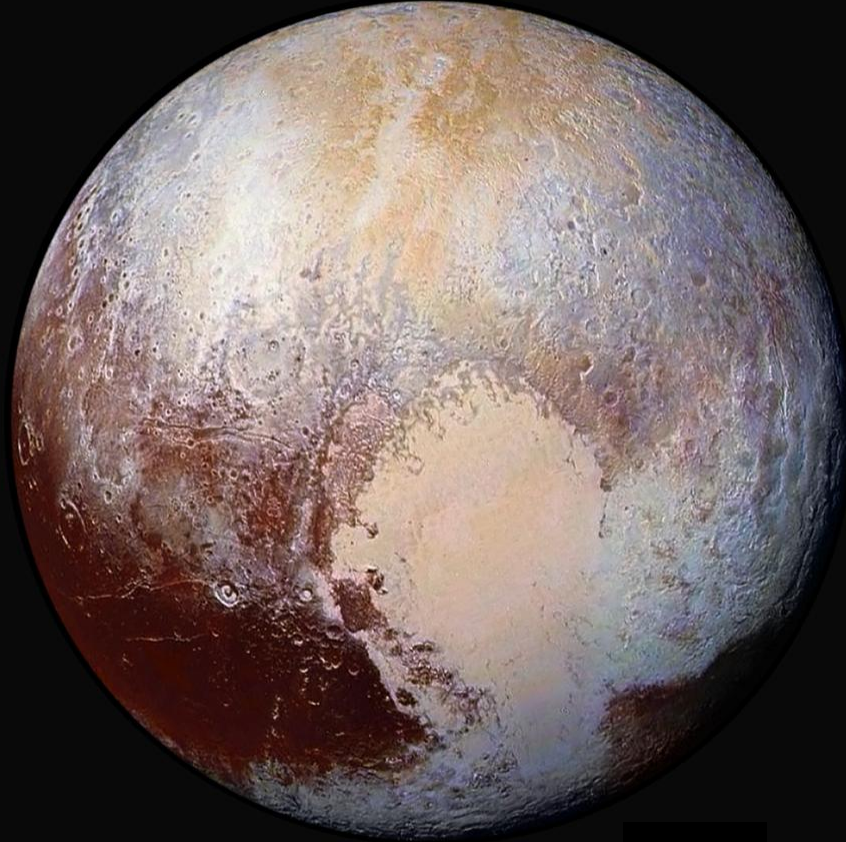
Natural Language Processing

- Technologies of **traditional Information Retrieval**
- Statistical models & machine learning

Semantic Web Technologies

- Natural language web content is **explicitly annotated with semantic metadata**
- Semantic metadata encode the **meaning** of the content and can be **read and interpreted correctly by machines**

Understanding Content on the Web



PLUTO
Never Forget!
(1930–2006)

Understanding Content on the Web



Text: "Pluto"

Entity Mapping
Disambiguation

Pluto

a Disney cartoon character

Pluto

a Roman god

Pluto

a song by Björk

HMS Pluto

a ship

...

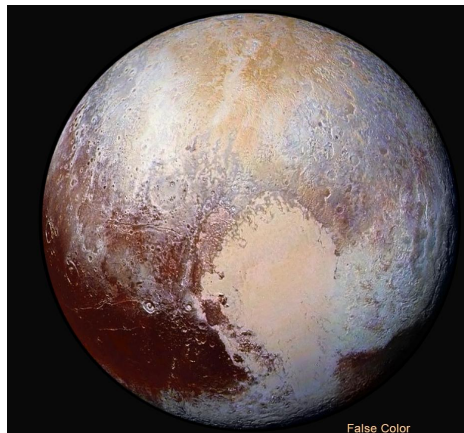
Pluto

a dwarf planet

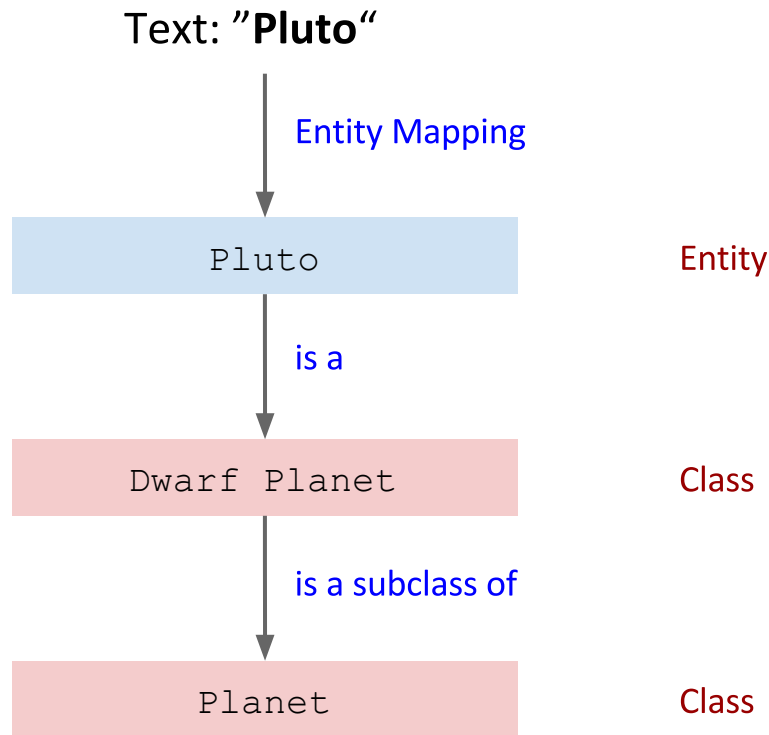
Disambiguation

- solution of linguistic ambiguities

Understanding Content on the Web



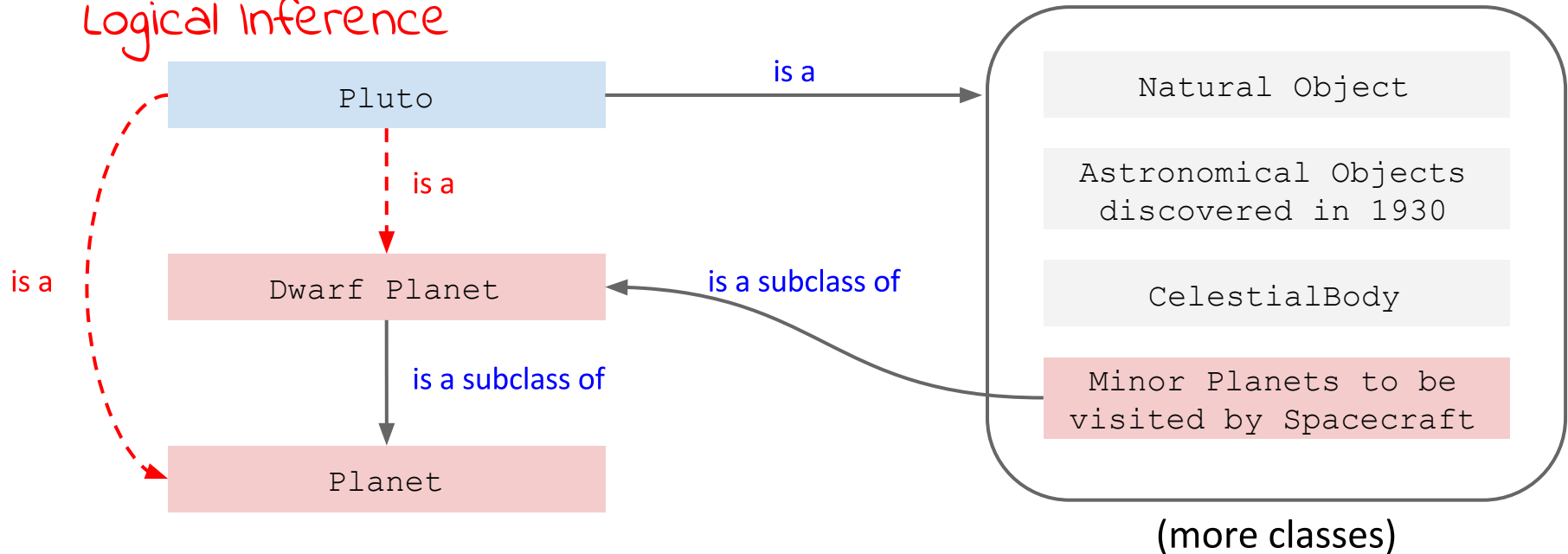
- The **Meaning (Semantics)** of entities and classes must be defined explicitly.



Understanding Content on the Web

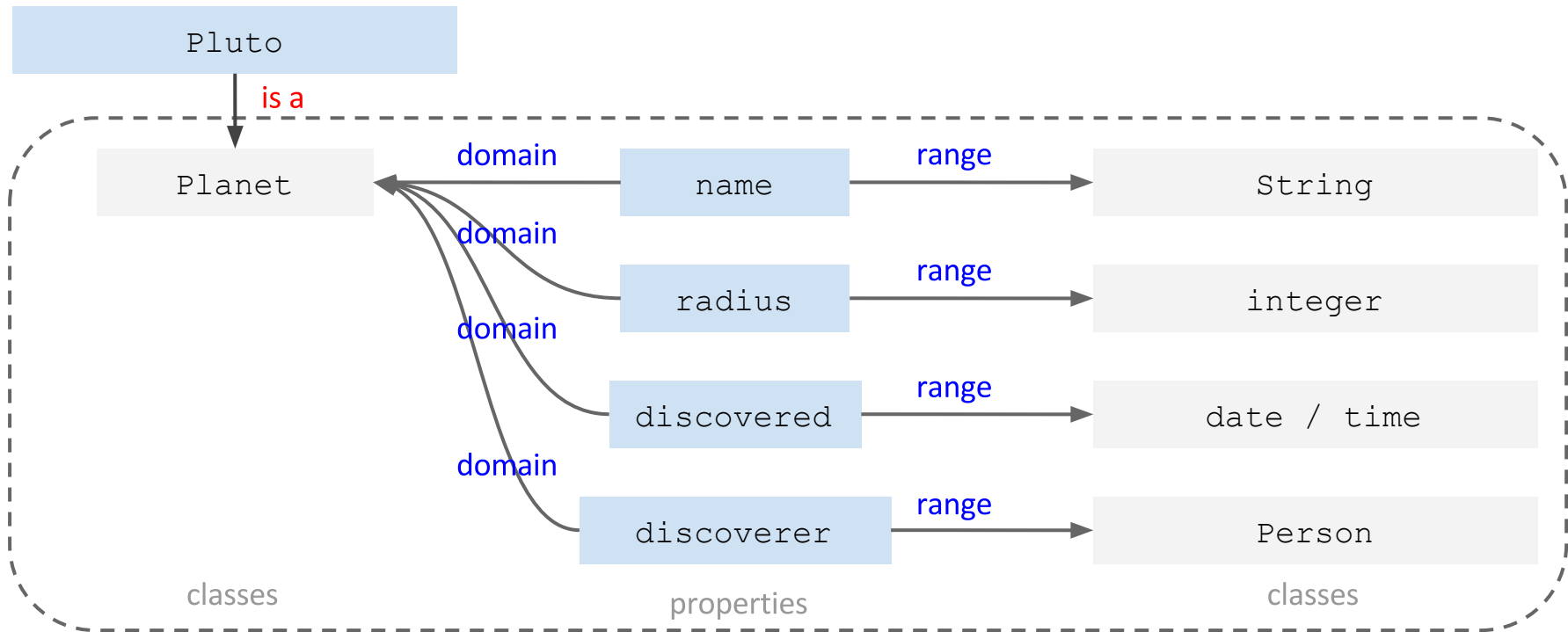
- The Meaning (Semantics) is expressed with the help of knowledge representations (**Ontologies**)

Logical Inference



Understanding Content on the Web

- The Meaning (Semantics) is expressed with the help of knowledge representations (**Ontologies**)



The Semantic Web - A Web of Data

- The meaning of information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies)**.
- Thus it will be possible,
 - to process the meaning of information automatically
 - to relate and integrate heterogeneous data
 - to deduce implicit (not evident) information from existing (evident) information in an automated way
- The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions**.

The Semantic Web - A Web of Data



„The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation“

Tim Berners-Lee, James Hendler, Ora Lassila: [The Semantic Web](#), Scientific American, 284(5), pp. 34-43(2001)

The Semantic Web Technology Stack (not a piece of cake...)

Most apps use only a subset of the stack

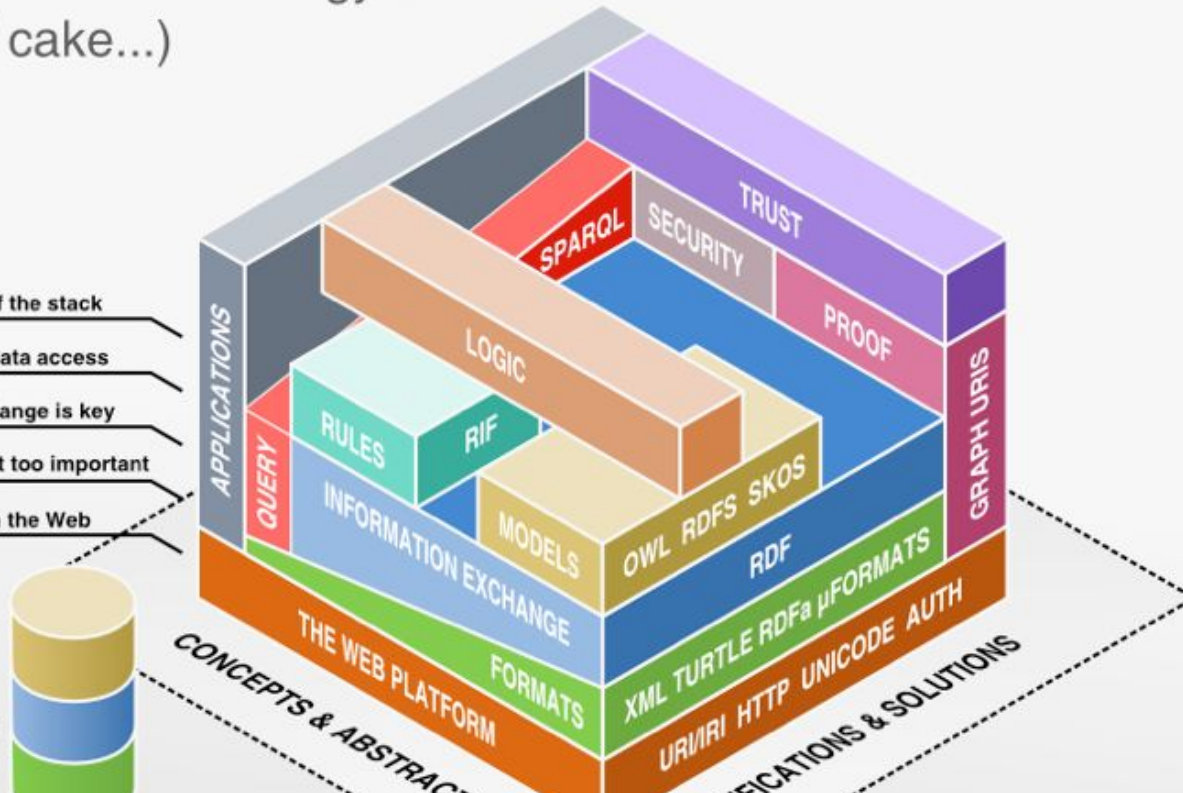
Querying allows fine-grained data access

Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

Linked Data uses a small selection of technologies



Next: 04 - Semantic Web Technology and the Web of Data

OpenHPI - Course Knowledge Engineering with Semantic Web Technologies

Lecture 1: Knowledge Engineering and the Web of Data