Ontologies (mini)Workshop

---- IDS internal training

Kody Moodley & Stuti Nayak

Institute of Data Science
30-10-2018

Learning objectives

- Introduction [20min]
  - What is an ontology?
  - What are some uses of ontologies relevant to IDS research?
- Hands-on [40min]
  - Basics of authoring an ontology using ontology editing software
  - Example application of ontologies: Disease Expert System
    - **Entity disambiguation**
    - **Data integration**
    - **Reasoning**

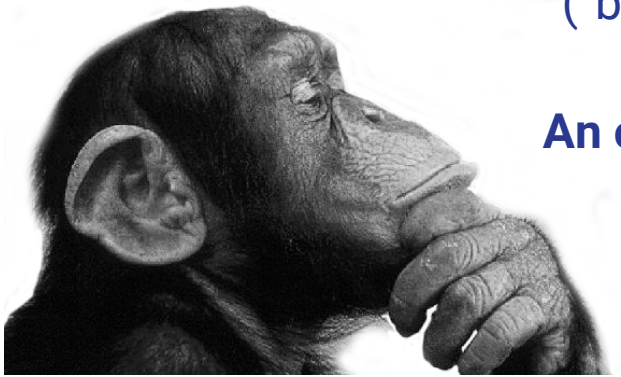
Introduction to Ontologies

What are they?

Ontology

Onto (“being” or “that which is”)

Logia (“logical discourse”)

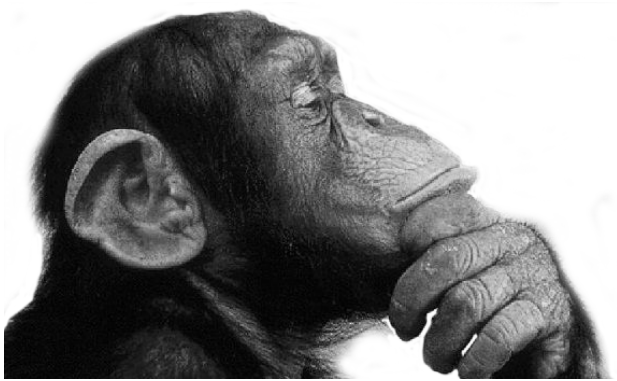


An ontology is a description of the world or some part of it (concerned with defining things)

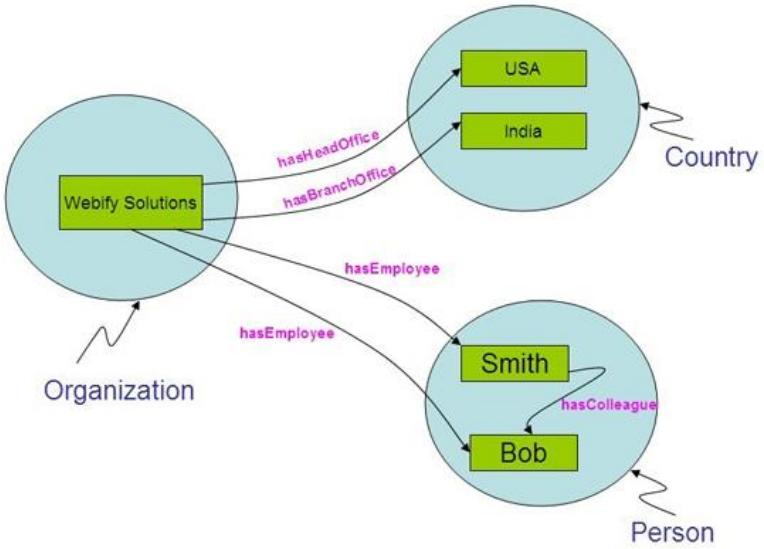
What are they?

Ontology in Computer Science

"A **formal** specification of knowledge about a domain that enables **machines** to derive implicit information from it"



What are they?



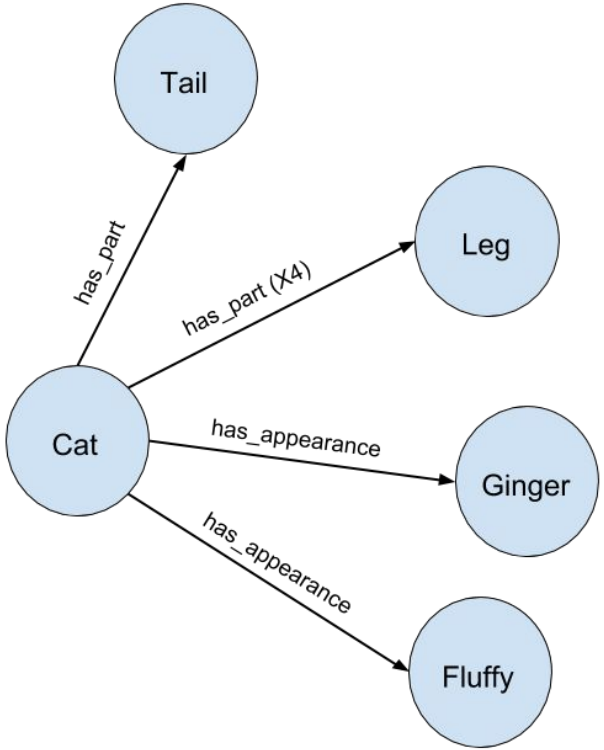
OR

1. Organization $\sqsubseteq \exists$hasHeadOffice.Country
2. Organization $\sqsubseteq \exists$hasBranchOffice.Country
3. Organization $\sqsubseteq \exists$hasEmployee.Person
4. Organization(webify_solutions)
5. Country(usa)
6. Country(india)
7. Person(smith)
8. Person(bob)
9. hasColleague(smith, bob)

Logical sentences (axioms)

Classes, relations, instances

Why make ontologies formal?



- What does “has_part (x4)” mean?
  - Cat has 4 parts in total, one of which is a leg?
  - Cat has 4 legs?
- Cat refers to a specific cat or *all* cats?

From syntax to semantics

All men are mortal

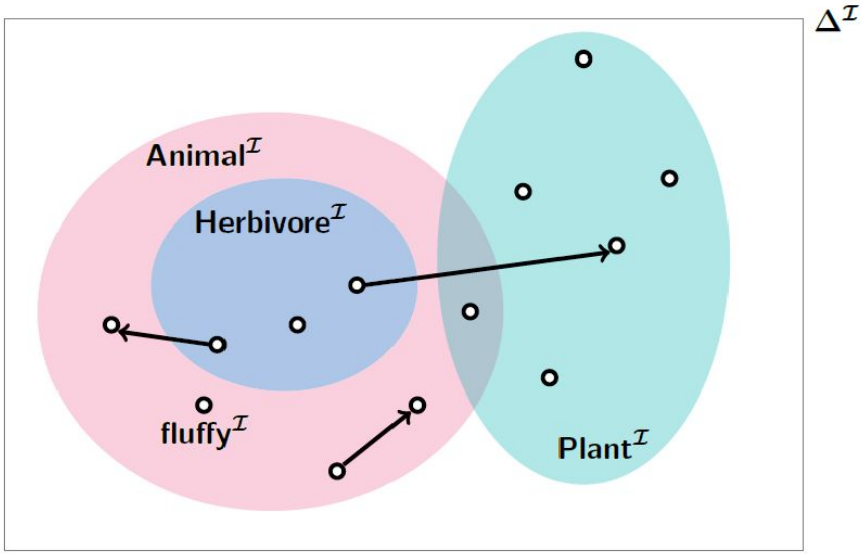
Socrates is a man

Modus ponens (rule of inference):

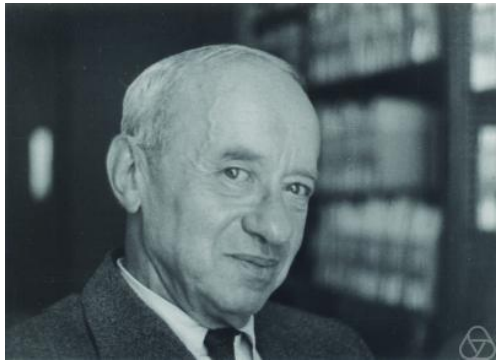
$P \rightarrow Q$
P
Therefore Q

Syntactical reasoning

From syntax to semantics



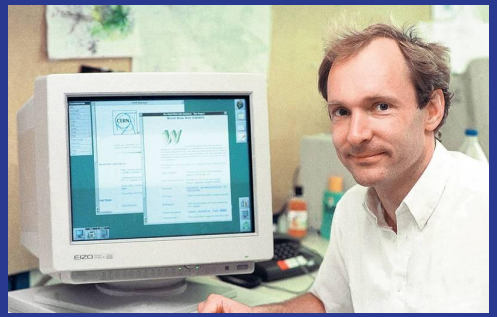
Animal	$\sqsubseteq$	¬Plant	NO
Herbivore	$\sqsubseteq$	Animal	YES
Herbivore	$\sqsubseteq$	∀eats.Plant	NO
Animal(fluffy)			YES
Herbivore(fluffy)			NO



Alfred Tarski (c. 1933)

Semantic reasoning

The Semantic Web Vision



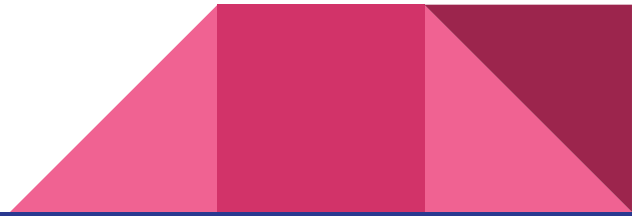
Tim Berners-Lee

Inventor of the World Wide Web

What is the Semantic Web?

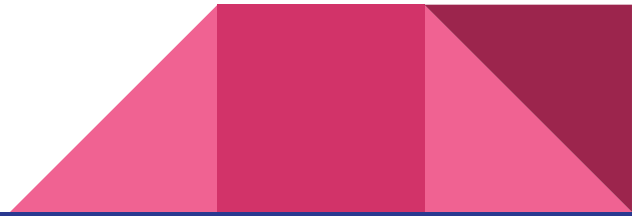
"...an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee et al., 2001



OWL - The Web Ontology Language(s)

- W3C recommendation in 2004 (OWL 1)
- Description Logics (DLs) underpin OWL
- Latest major version: OWL 2
- Profiles: OWL 2 Full, OWL 2 DL, OWL 2 RL, OWL 2 EL
- Structural specification (no commitment to concrete syntax)
- Concrete syntax examples: RDF/XML and OWL/XML (.rdf or .owl), Turtle (.ttl)



What does an OWL statement look like?

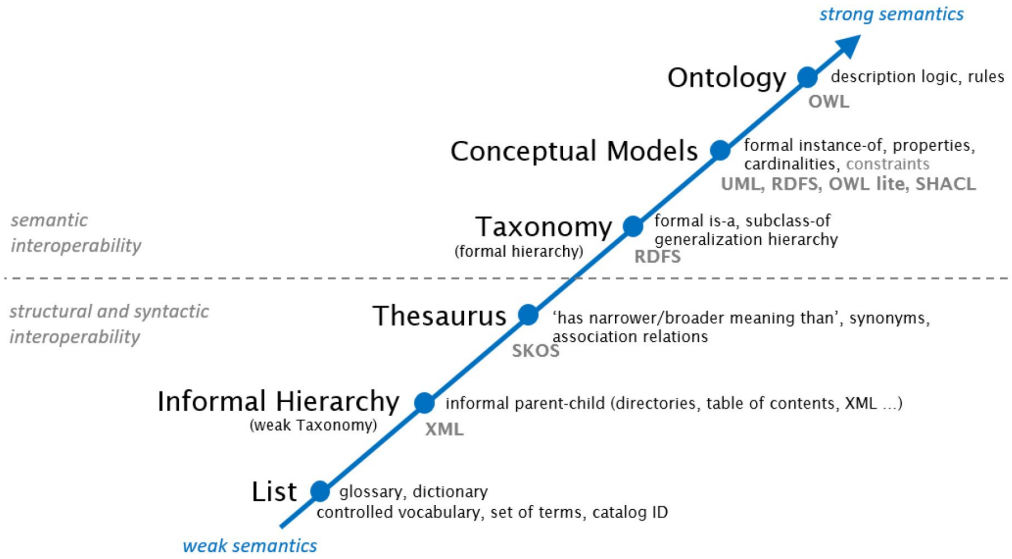
English: “An actress is a female person that plays a role in a movie”

DL: $\text{Actress} \equiv \text{Female} \sqcap \exists \text{playsARoleIn}.\text{Movie}$

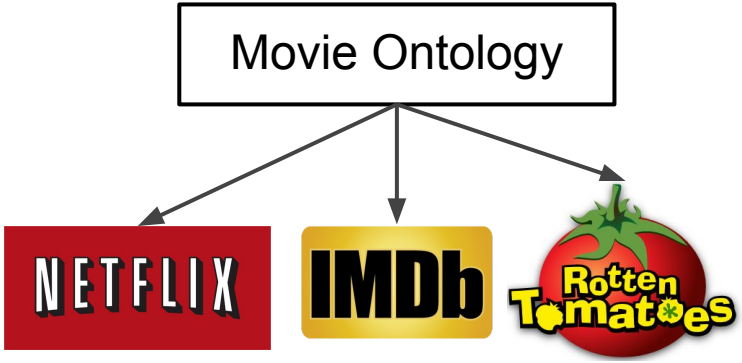
OWL (XML syntax):

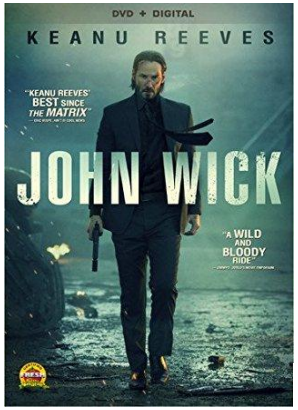
```
<owl:Class rdf:about="http://movies.com/movieontology.owl#Actress">
    <owl:equivalentClass>
      <owl:Class>
        <owl:intersectionOf rdf:parseType="Collection">
          <rdf:Description rdf:about="http://movies.com/movieontology.owl#Female"/>
          <owl:Restriction>
            <owl:onProperty rdf:resource="http://movies.com/movieontology.owl#playsARoleIn"/>
            <owl:someValuesFrom rdf:resource="http://movies.com/movieontology.owl#Movie"/>
          </owl:Restriction>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
  </owl:Class>
```

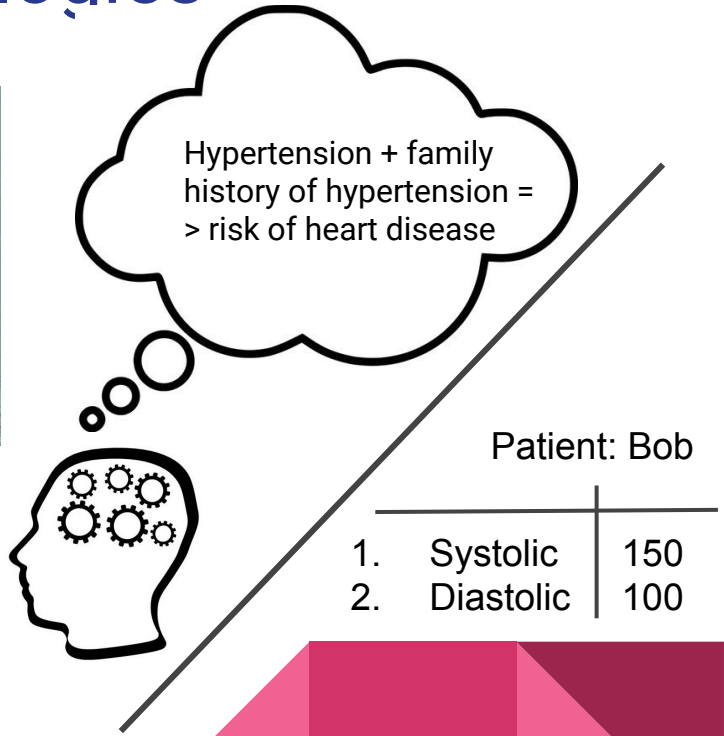
Species of ontology “in the wild”



Summary: general uses of ontologies







share common understanding of a domain

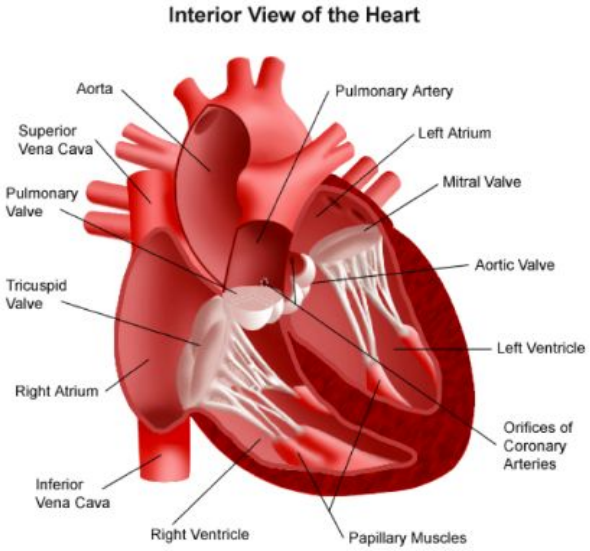
enable reuse of domain knowledge

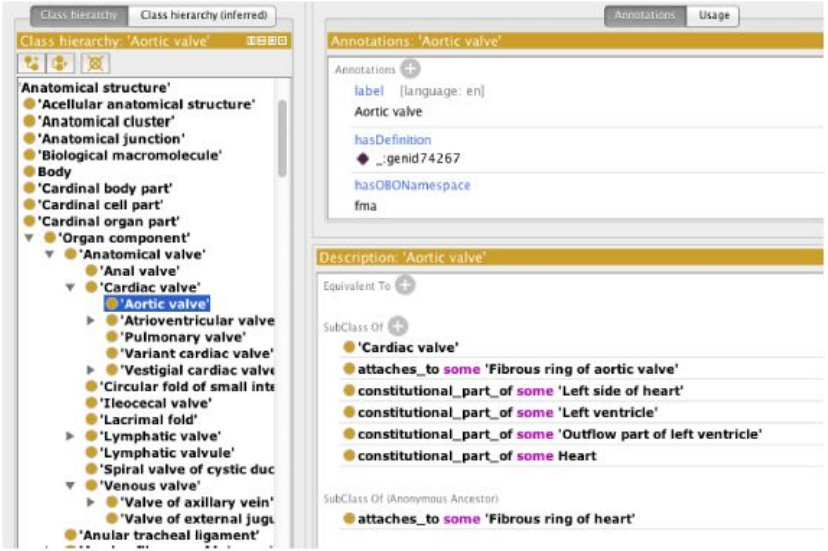
make domain assumptions **explicit**

separate domain & operational knowledge

Ontologies in Life Sciences

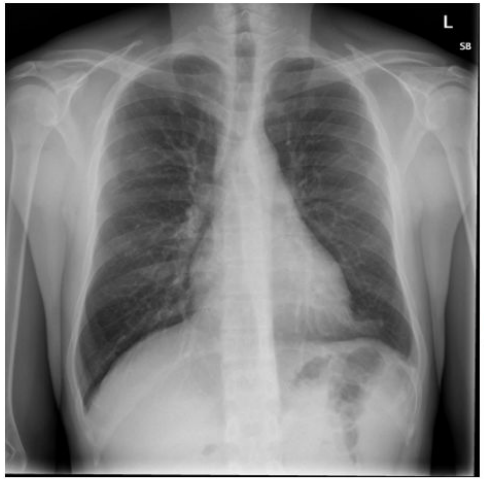
SNOMED (General), Foundational Model of Anatomy (FMA), NCI (Cancer)





Source: Meghyn Bienvenu, CNRS

Ontologies in Medicine



Comparison: None

Indication: Burmese male has complete TB treatment

Findings: Both lungs are clear and expanded with no infiltrates. Basilar focal atelectasis is present in the lingula. Heart size normal. Calcified right hilar XXXX are present

Impression: No active disease.

MeSH

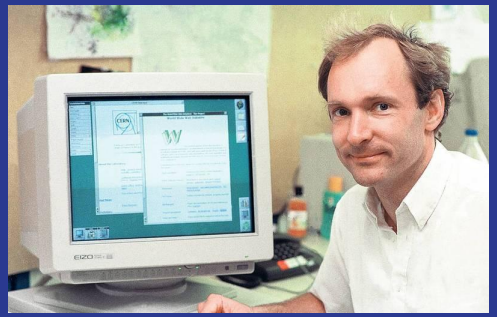
Major

Pulmonary Atelectasis / lingula / focal
Calcinosis / lung / hilum / right

CHEST 2V FRONTAL/LATERAL XXXX, XXXX XXXX PM

- Querying medical records (e.g. find patients eligible for medical trials)
  - Myocardial infarction vs. MI. vs. heart attack vs. 410.0
- Supporting annotation and visualisation tools (patient scans and x-rays)

The Linked Data Paradigm



Tim Berners-Lee

Inventor of the World Wide Web

OWL vs Resource Description Framework (RDF)

- OWL **defines** terms, Resource Description Framework (RDF) provides mechanism for **associating** terms with data entities to categorise & link them
- Linked data:
  - Give things **unique names (URIs)**
  - Make these names **dereferencable** on the Web (give them URLs)
  - At these URLs **provide info** about the things
  - **Include links** to other URIs so we can discover related things

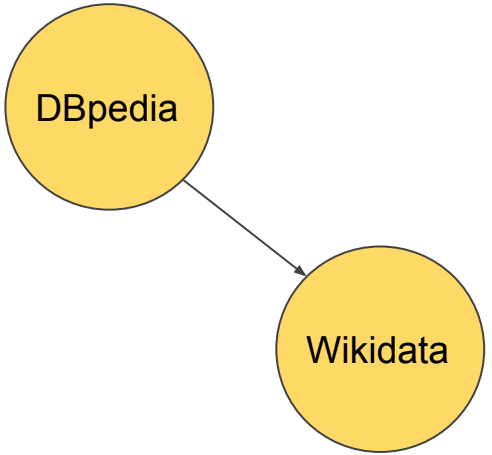
What do RDF triples look like?

“Angelina Jolie is an actress, she acts in the movie Lara Croft: Tomb Raider”...and “she has breast cancer.”

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpedia: <http://dbpedia.org/page/>
PREFIX mov: <http://movies.com/movieontology.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX wikidata: <https://www.wikidata.org/wiki/>

- dbpedia:Angelina_Jolie mov:playsARoleIn dbpedia:Lara_Croft:_Tomb_Raider .
- dbpedia:Angelina_Jolie rdf:type mov:Actress .
- dbpedia:Lara_Croft:_Tomb_Raider rdf:type mov:Movie .

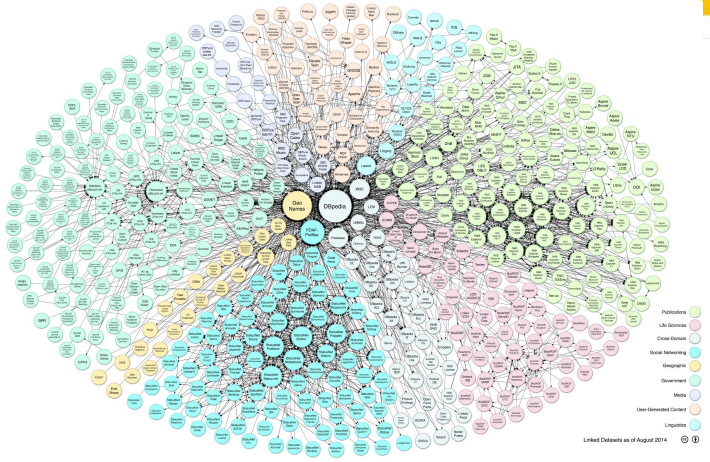
- dbpedia:Angelina_Jolie obo:RO_0002200 obo:DOID_1612 .
- obo:RO_0002200 rdfs:label “has phenotype”^xsd:string .
- obo:DOID_1612 rdfs:label “Breast Cancer”^xsd:string .
- obo:DOID_1612 owl:sameAs wikidata:Q128581 .

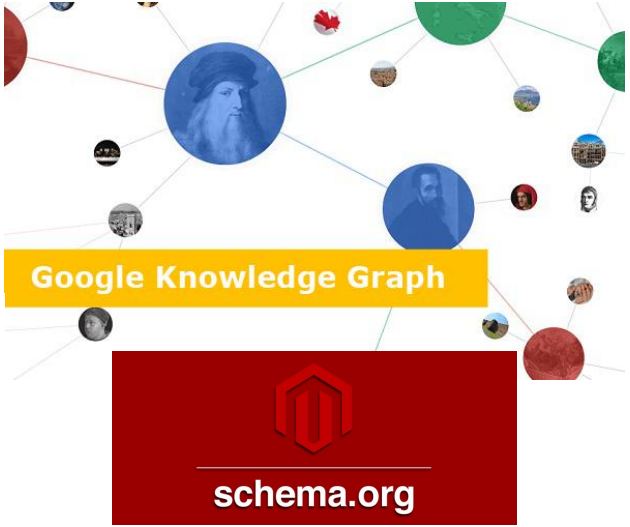


Knowledge Graphs: Bio2RDF, Google, LOD

BIO2RDF

Linked Data for the Life Sciences





Linked Data & Knowledge Graphs

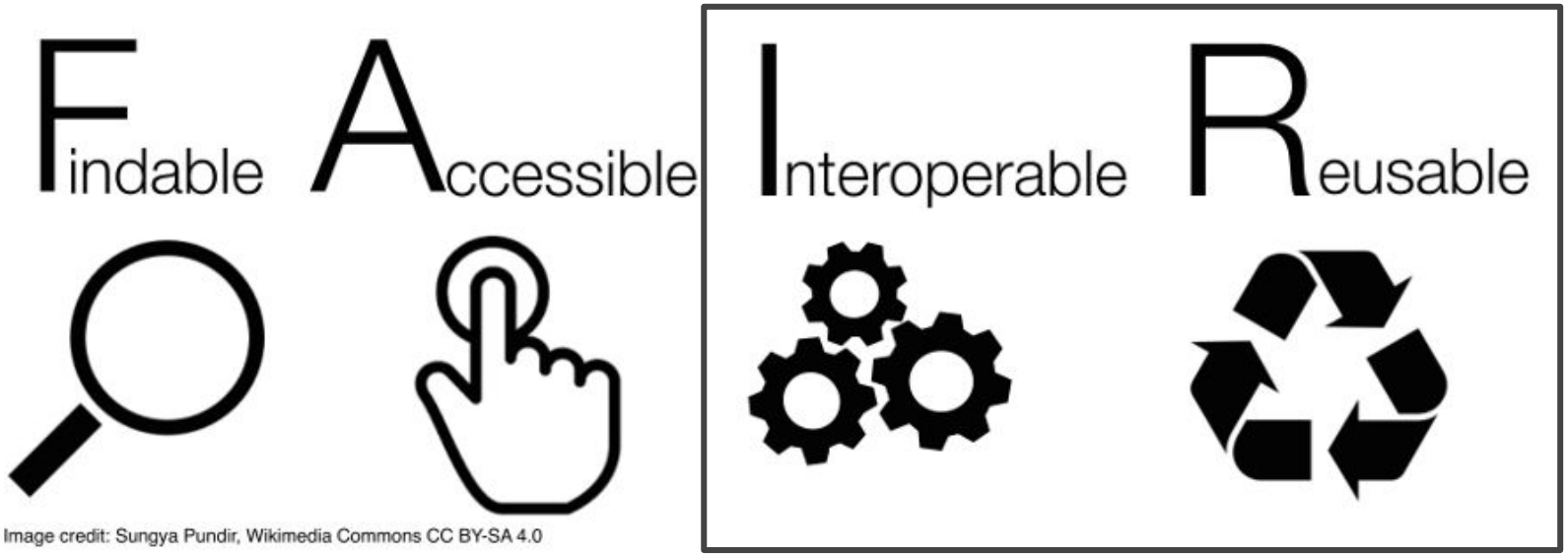
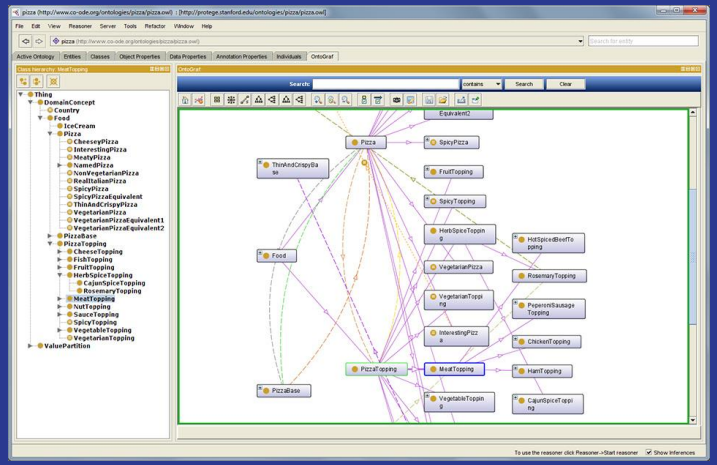


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

Hands-on task 1:

Construct a simple ontology in Protege

Building your first ontology!



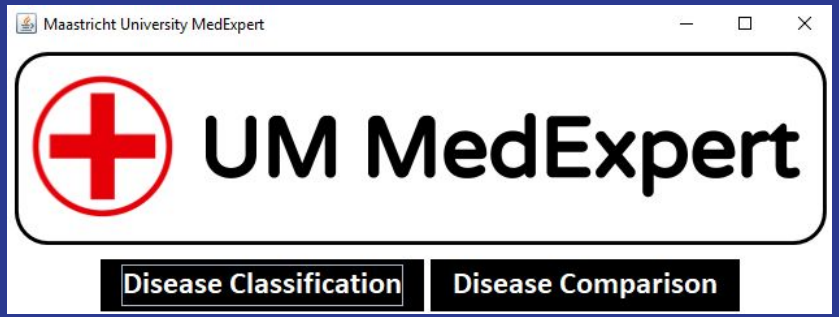
Protégé

Developed at Stanford University

Hands-on task 2:

Ontology based inference

Ontology inference



Medical Expert System

Made purely for educational purposes

Further reading

- Semantic Web
- OWL
- Linked Data
- OWL reasoners
- Description Logics
- Protege