Ontology-Based Information Systems

Ian Horrocks
<ian.horrocks@comlab.ox.ac.uk>
Information Systems Group
Oxford University Computing Laboratory

What is an Ontology?

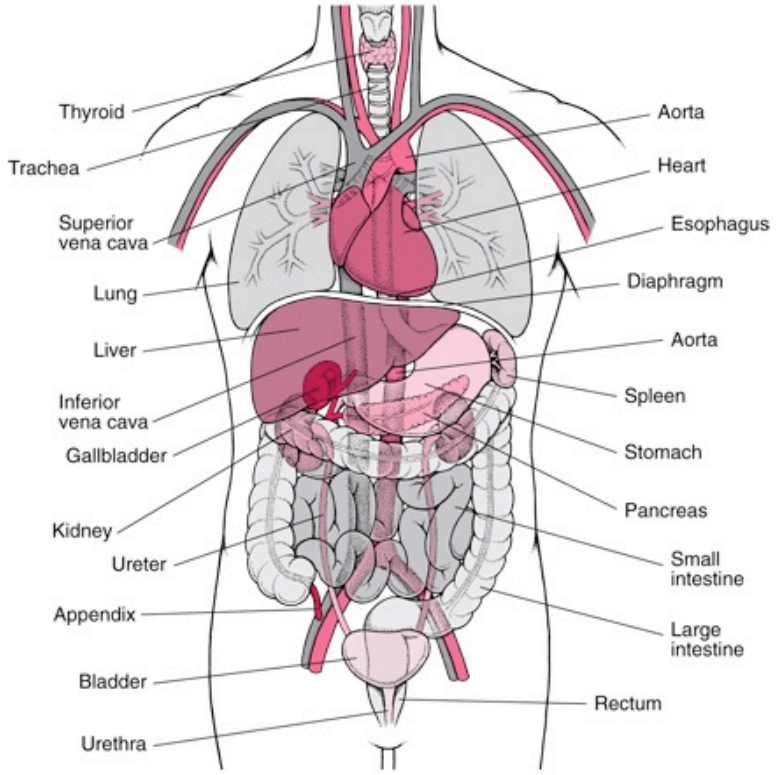
What is an Ontology?

A model of (some aspect of) the world

What is an Ontology?

A model of (some aspect of) the world

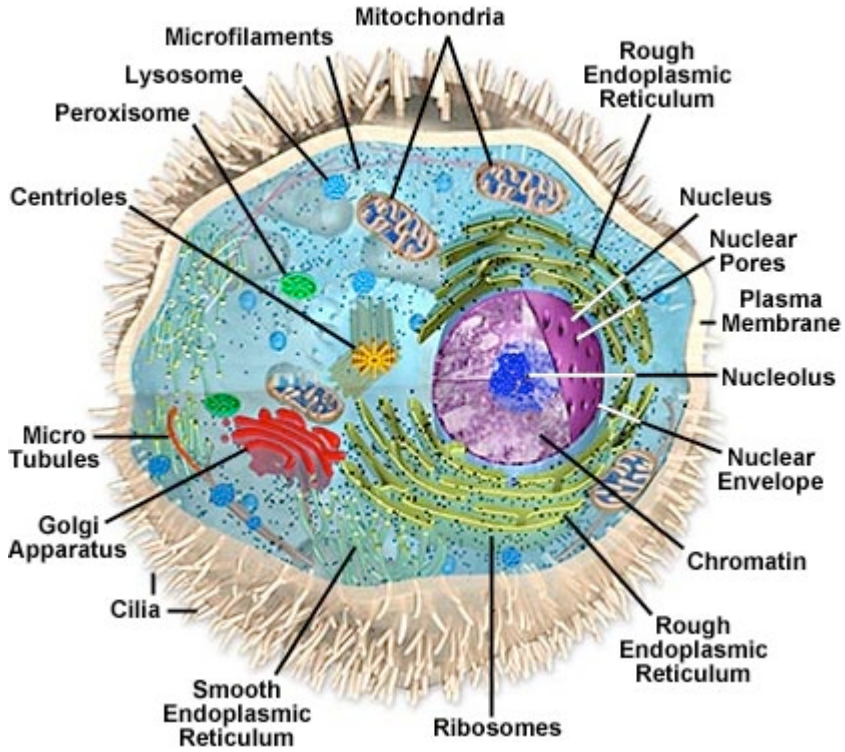
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy



What is an Ontology?

A model of (some aspect of) the world

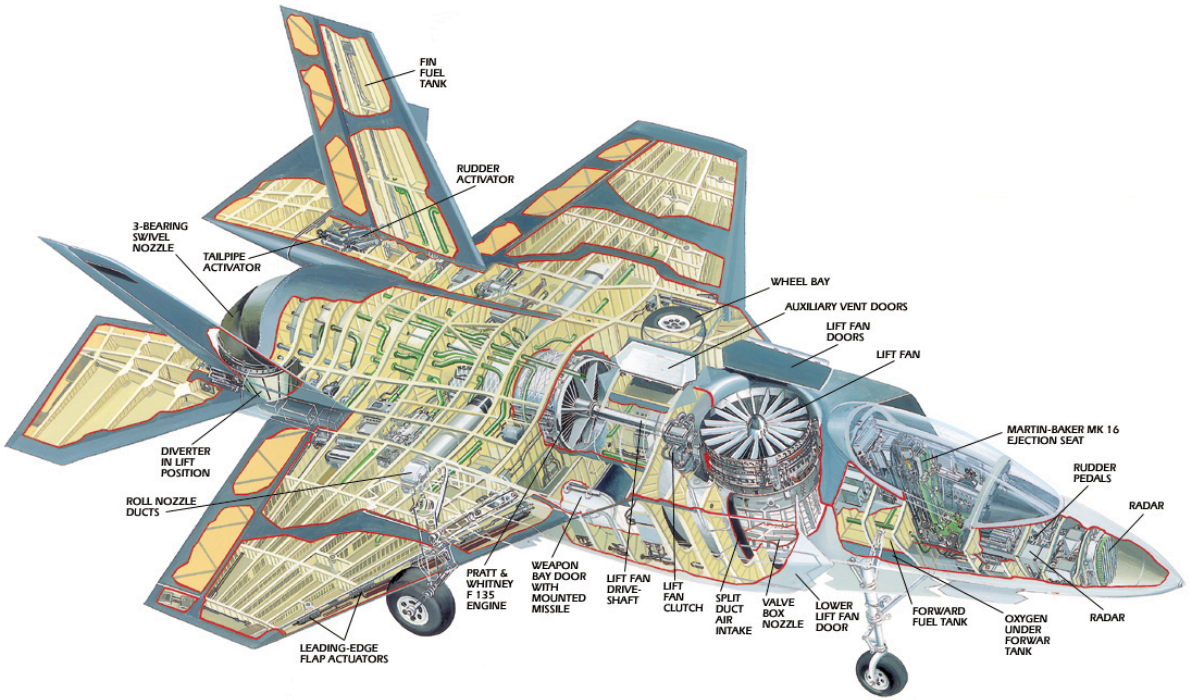
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology



What is an Ontology?

A model of (some aspect of) the world

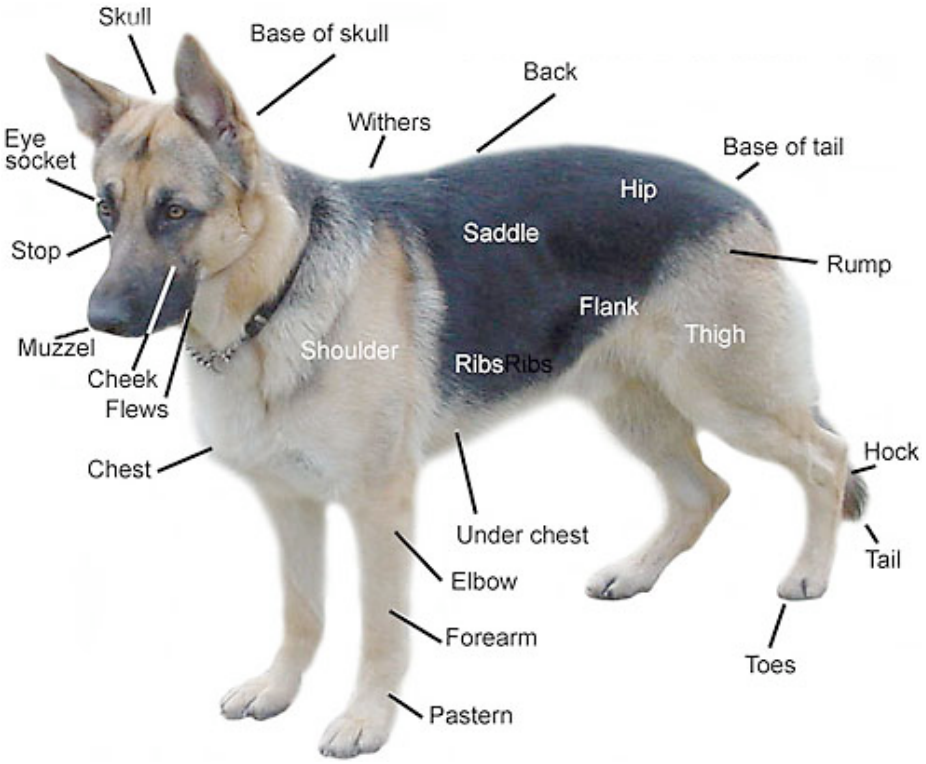
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace



What is an Ontology?

A model of (some aspect of) the world

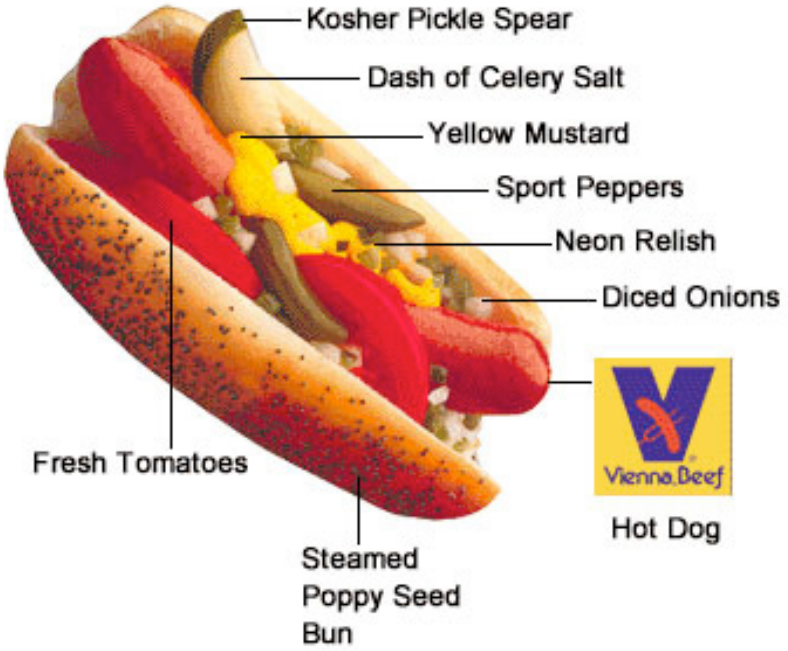
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs



What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs
  - Hotdogs
  - …

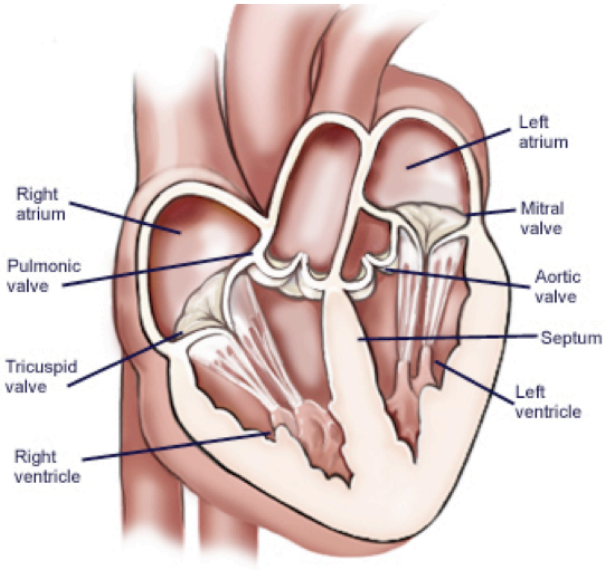


What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Heart is a muscular organ that is part of the circulatory system



What is an Ontology?

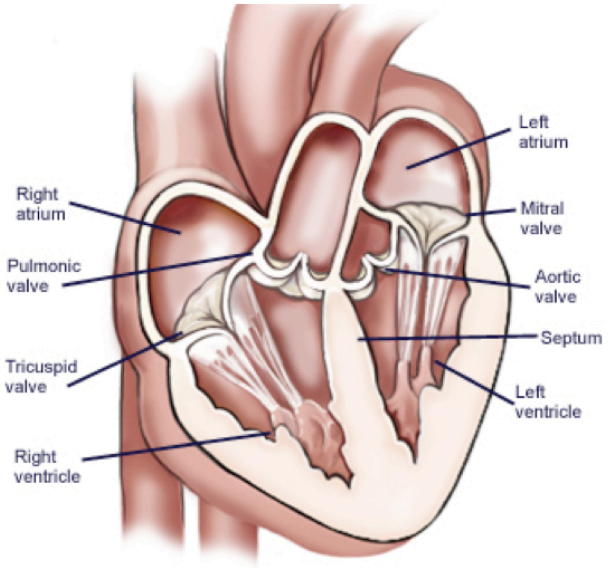
A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Heart is a muscular organ that is part of the circulatory system

- **Formalised** using suitable logic

$$\forall x.[\mathsf{Heart}(x) \rightarrow \mathsf{MuscularOrgan}(x) \wedge \exists y.[\mathsf{isPartOf}(x,y) \wedge \mathsf{CirculatorySystem}(y)]]$$



The Web Ontology Language OWL

- Motivated by **Semantic Web** activity

  Add meaning (semantics) to web content by annotating with terms defined in ontologies

- Developed by W3C WebOnt working group
  - Based on earlier languages **RDF**, **OIL** and **DAML+OIL**
  - Became a **recommendation** on 10 Feb 2004

- Supported by **tools and infrastructure**
  - APIs (e.g., OWL API, Thea, OWLink)
  - Development environments (e.g., Protégé, TopBraid Composer)
  - Reasoners & Information Systems (e.g., Pellet, HermiT, Quonto)
- Based on a **Description Logic** ($\mathcal{SHOIN}$)

Description Logics (DLs)

- Fragments of **first order logic** designed for KR
- Desirable computational properties
  - **Decidable** (essential)
  - Low complexity (desirable)
- Succinct and **quantifier free syntax**

$$\forall x.[\mathsf{Heart}(x) \rightarrow \mathsf{MuscularOrgan}(x) \wedge \exists y.[\mathsf{isPartOf}(x,y) \wedge \mathsf{CirculatorySystem}(y)]]$$

$$\mathsf{Heart} \sqsubseteq \mathsf{MuscularOrgan} \sqcap \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem}$$

Description Logics (DLs)

DL **Knowledge Base** (KB) consists of two parts:

- Ontology (aka **TBox**) axioms define terminology (schema)

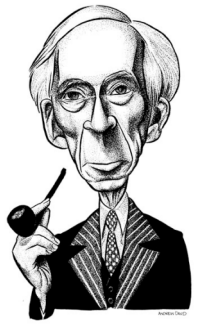
$$\begin{aligned} \mathsf{Heart} &\sqsubseteq \mathsf{MuscularOrgan} \sqcap \\ &\quad \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \\ \mathsf{HeartDisease} &\equiv \mathsf{Disease} \sqcap \\ &\quad \exists \mathsf{affects}.\mathsf{Heart} \\ \mathsf{VascularDisease} &\equiv \mathsf{Disease} \sqcap \\ &\quad \exists \mathsf{affects}.(\exists \mathsf{isPartOf}.\mathsf{CirculatorySystem}) \end{aligned}$$

- Ground facts (aka **ABox**) use the terminology (data)

$$\begin{aligned} \mathsf{John} : \mathsf{Patient} &\sqcap \\ &\exists \mathsf{suffersFrom}.\mathsf{HeartDisease} \end{aligned}$$

Why Care About Semantics?

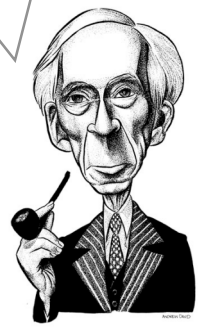
Why should I care about semantics?



Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

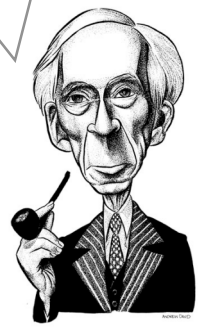


Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

That's OK, but I don't get paid for philosophy.



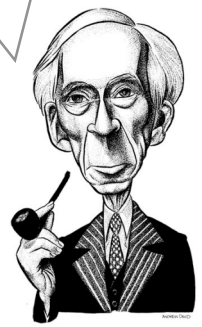
Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

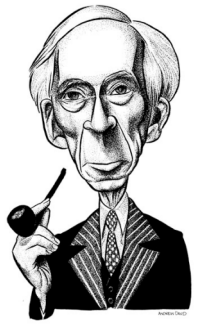
That's OK, but I don't get paid for philosophy.

From a practical POV, in order to specify and test ontology-based information systems we need to precisely define relationships (like entailment) between logical statements.



Why Care About Semantics?

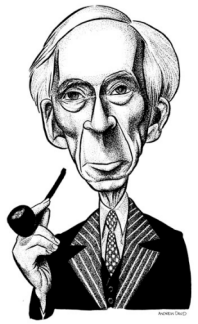
In FOL we define the semantics in terms of models (a model theory). A model is supposed to be an analogue of (part of) the world being modeled. FOL uses a very simple kind of model, in which “objects” in the world (not necessarily physical objects) are modeled as elements of a set, and relationships between objects are modeled as sets of tuples.



Why Care About Semantics?

In FOL we define the semantics in terms of models (a model theory). A model is supposed to be an analogue of (part of) the world being modeled. FOL uses a very simple kind of model, in which “objects” in the world (not necessarily physical objects) are modeled as elements of a set, and relationships between objects are modeled as sets of tuples.

Note that this is exactly the same kind of model as used in a database: objects in the world are modeled as values (elements) and relationships as tables (sets of tuples).



What are Ontologies Good For?

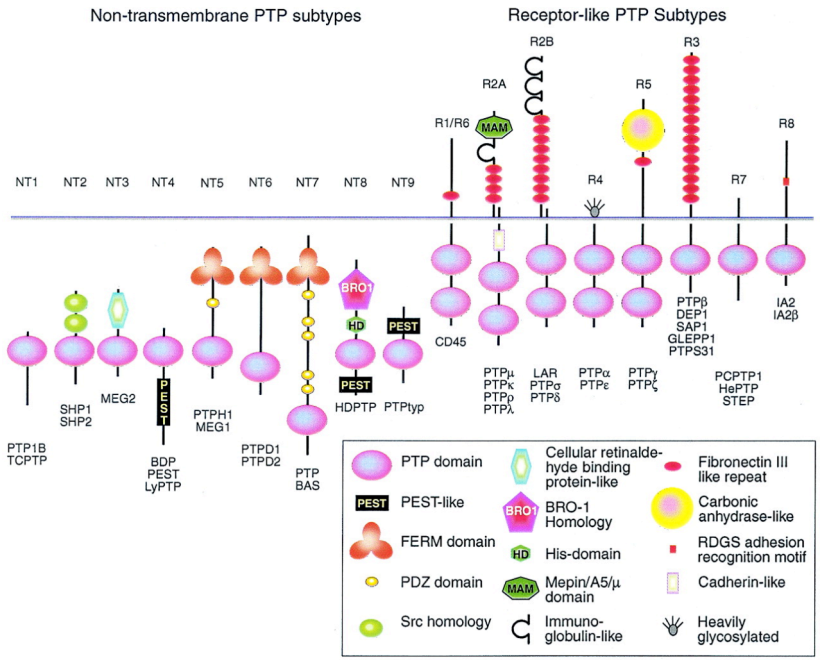
- Coherent **user-centric view** of domain
  - Help identify and resolve disagreements
- Ontology-based **Information Systems**
  - View of data that is independent of logical/physical schema
  - Queries use terms familiar to users
  - Answers reflect knowledge & data, e.g.:

    "Patients suffering from Vascular Disease"
  - Query navigation/refinement
  - Incomplete and semi-structured data
  - Integration of heterogeneous sources

Now... *that* should clear up a few things around here

e-Science

- E.g., for “in silico” investigations and “**hypothesis testing**”
  - Comparing data (e.g., on proteins) to (model of) biological knowledge
  - Characteristics of proteins captured in an ontology $\mathcal{O}$
  - Abox populated with e.g., data from **gene sequencing experiments**



e-Science

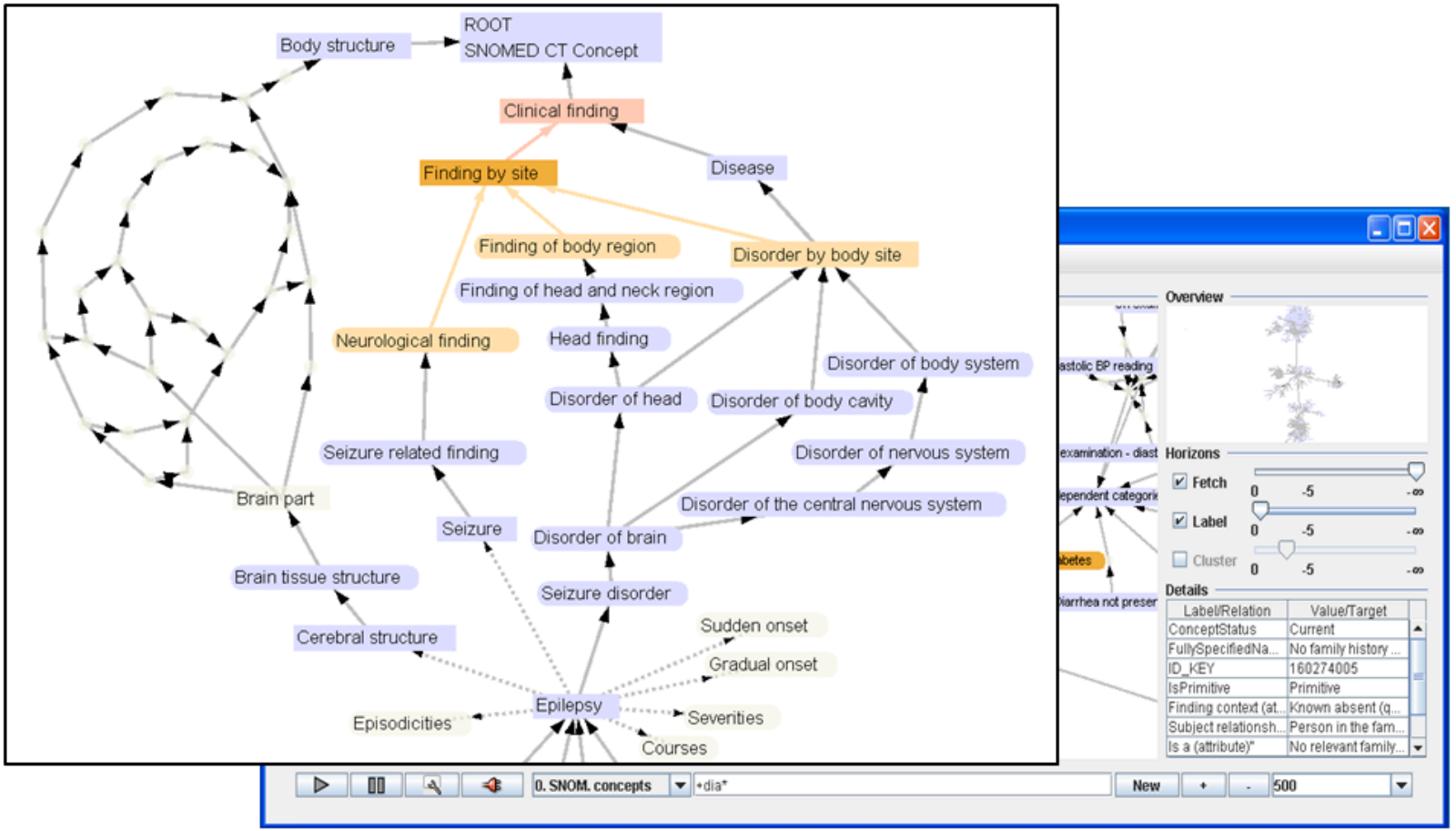
- E.g., for “in silico” investigations and “**hypothesis testing**”
  - Comparing data (e.g., on proteins) to (model of) biological knowledge
  - Characteristics of proteins captured in an ontology $\mathcal{O}$
  - Abox populated with e.g., data from **gene sequencing experiments**
  - Expert compares hypotheses with query answers
    - E.g., all human phosphotases are of type p1, …, pi
  - Result may be, e.g., discovery of new kinds of protein
    - And these may be potential **drug targets** if unique to a pathenogen
  - Result may also be discovery of errors in model
    - Which may reflect **gaps/errors in existing knowledge**

Healthcare

- UK NHS has a **£6.2 billion** "Connecting for Health" IT programme
- Key component is **Care Records Service** (CRS)
  - "Live, interactive patient record service accessible 24/7"
  - Patient **data distributed** across local centres in 5 regional clusters, and a national DB
    - Detailed **records** held by local service providers
    - Diverse **applications** support radiology, pharmacy, etc
    - Applications exchange **messages** containing "semantically rich clinical information"
    - **Summaries** sent to national database
  - **SNOMED-CT** ontology provides common **vocabulary** for data
    - Clinical data uses terms drawn from ontology

SNOMED

- Over **400,000 concepts**



SNOMED

- Over **400,000 concepts**
- **Schema only** — no instances
- Language used is a (well known) **fragment of OWL**
- NHS version extended with 1,000s of additional classes
  - **OWL reasoner** (FaCT++) used to classify and check ontology
    - Currently takes ≈ 10 minutes
  - 180 **missing subClass relationships** were found, e.g.:
    - Periocular_dermatitis subClassOf Disease_of_face
    - Fibrin_measurement subClassOf Coagulation_factor_assay

SNOMED

- Vocabulary is **extensible** at point of use: “post coordination”
  - Users (e.g. clinicians) may add/define new vocabulary
  - Terminology service (reasoner) used to insert in ontology
- Typical new term:
  - **almond_allergy** ≡ “allergy caused_by almond”
  - OWL reasoner (FaCT++) used to classify new term
    - Takes <10 ms
  - Classified as a kind of “**nut allergy**”
    - Clearly of **crucial importance** to recognise patients with allergy caused by almond as kinds of patient with nut allergy

Columbia Presbyterian Medical Center

- Ontology used in analysis of results in path lab
- OWL reasoner used to check this ontology

- Several **errors and omissions found** that:

  *"**would have led to missed test results**"*

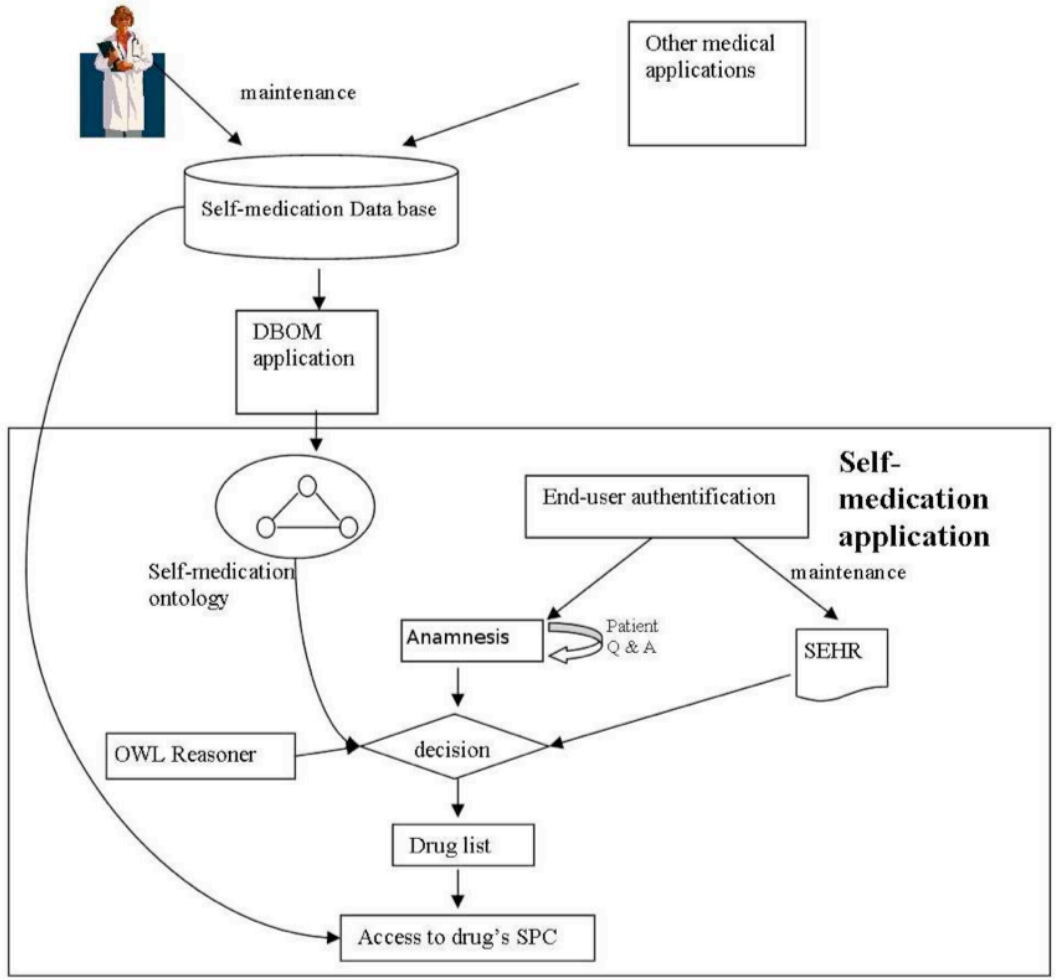
- Result: improvement in **improvement in patient care**

Online Self-Medication Advice

- Self-medication is pervasive, but can be **hazardous**
  - **180 deaths** in the USA in 2006
- French project to provide **on-line advice**
  - Will be made available to **20 million** customers of French health insurance companies
  - Patients have their own **simple health care record** (SEHR)
  - **Diagnosis system** considers symptom descriptions, SEHR, Q&A and self-medication KB
  - Uses an ontology for vocabulary and knowledge (axioms) about treatments, contra-indications, side-effects, etc.
    - E.g., do not take x if patient suffers from y; side-effects of x may include z

Online Self-Medication Advice

- Self-me ardous
  - 180 c
- French
  - Will b French healt
  - Patie d (SEHR)
  - Diag s, SEHR, Q&A
  - Uses check for contr
    - E cts may in



Online Self-Medication Application

- Data taken from **drug terminologies**, e.g.:
  - European Pharmaceutical Market Research Association (EphMRA)
  - Anatomical Therapeutic Chemical (ATC)
- Data transformed into **OWL ontology**
  - Expert uses reasoner to check and enhance ontology
- **OWL reasoner** also used to check and enhance data
  - Combined with induction and interaction with expert
  - Corrected missing/incorrect information on interactions, contra-indications, allergies, side-effects, etc.
  - Quality of data improved by factor of 8%

Thank you for listening

Any questions?

Resources:

- This talk:
  - http://www.comlab.ox.ac.uk/people/ian.horrocks/Seminars/
- OWL 2 Proposed Recommendation:
  - http://www.w3.org/2007/OWL/wiki/OWL_Working_Group#Deliverables