Medical Informatics

Lecture 8: Ontologies

Dr Areti Manataki

Nanjing Medical University

Introduction

- Combining information from different databases can be highly desirable.
  - But this task is not always that straightforward.
- Alternatively, we could hold all possible data in a single, global, web-based relational database.
  - But how about data ownership and control, or even freedom of expression?
- The semantic web is an approach proposed by the World Wide Web Consortium (W3C) towards linked data over the web.

The Semantic Web

- What is the Semantic Web?

Web of connected, machine-readable data

- What are the main aims of the Semantic Web?
  - Automated query-answering
  - Automated use of the data: reasoning, planning, acting, etc.

In the next lectures

- Fundamental concepts
  - Ontologies
  - Resources
  - Linked data
- Representation
  - RDF data model
  - Turtle serialisation for RDF
- Querying
  - SPARQL

What is an ontology?

- The philosophy of being

- A definitive account of what exists

- So we could, in theory, create a single ontology that describes the world.

  - But relevant knowledge is highly subjective: which concepts are important, how these are organised, what terms to use

  - Ontologies are designed by individuals: central control is impossible and undesirable

What is an ontology?

- Ontological differences are desirable and essential:
  - Freedom of expression
  - Ability to adapt to task
  - Changing environment

Ontology definitions

- *An explicit specification of a conceptualisation* (Gruber, 1993)

- *A formal, explicit specification of a shared conceptualisation* (Studer et al, 1998)

- *An explicit representation of a shared understanding of the important concepts in some domain of interest* (Kalfoglou, 2002)

- *A set of types, properties and relationships* (Wikipedia, 2016)

But what does this mean?

So what is an ontology?

- Essentially: a way of encoding domain knowledge

- Something like an enhanced dictionary, where you can look up the meaning of different concepts and find relations between them

- Several ontologies out there:

  - Music Ontology: basic concepts and properties for describing music, e.g. artists, albums, tracks and performances

  - BIO Ontology: terms useful for finding out more about people and their backgrounds, e.g. birth, parent, divorce, etc.
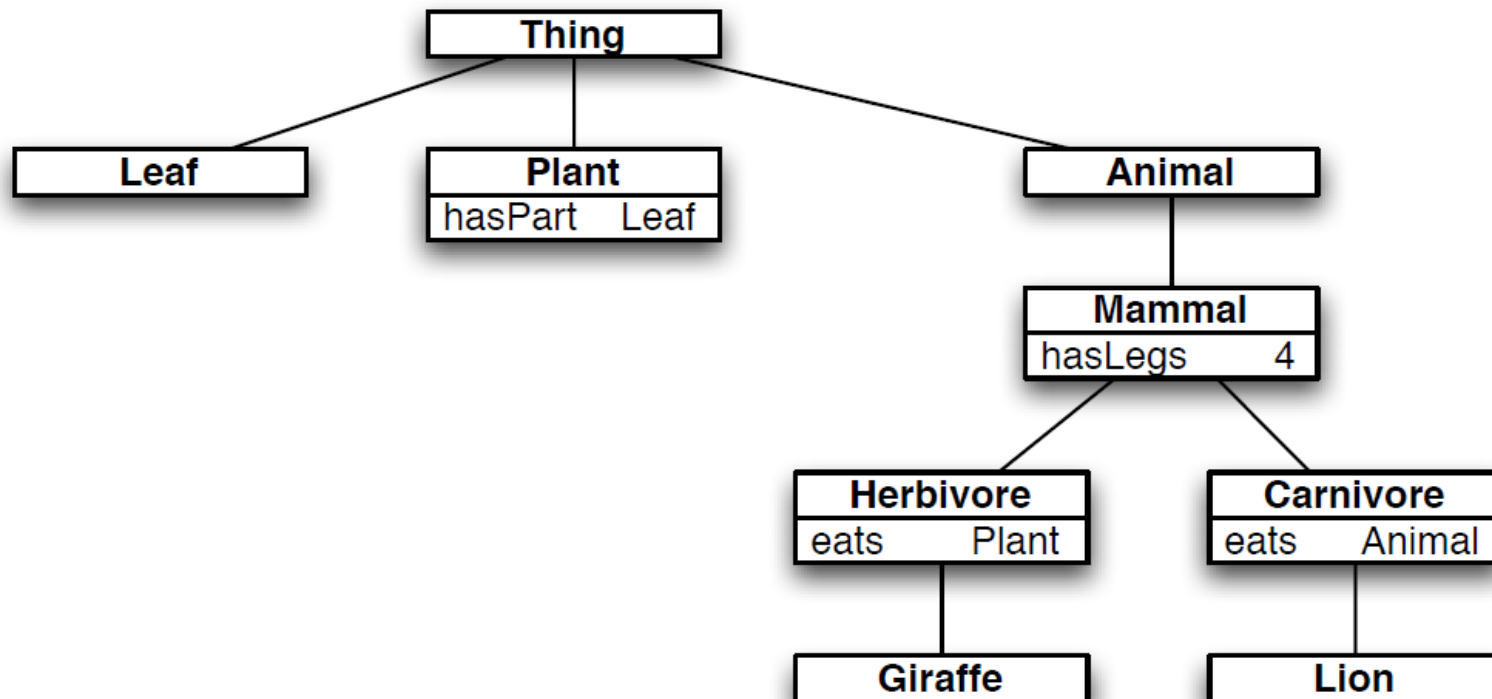
Why Semantic Web ontologies?

- Ontologies allow us to attach meanings to data
  - "William Burnes is the father of Robert Burns"
    - Here, "father" could be assigned the meaning of male parent, stepfather, father-in-law, etc.
- Ontologies allow us to infer new knowledge from existing data
  - If we have that
    "William Burnes is the father of Robert Burns"
       and
    "Father is a sublcass of Parent"
  - …then we can deduce that
    "William Burnes is the parent of Robert Burns".

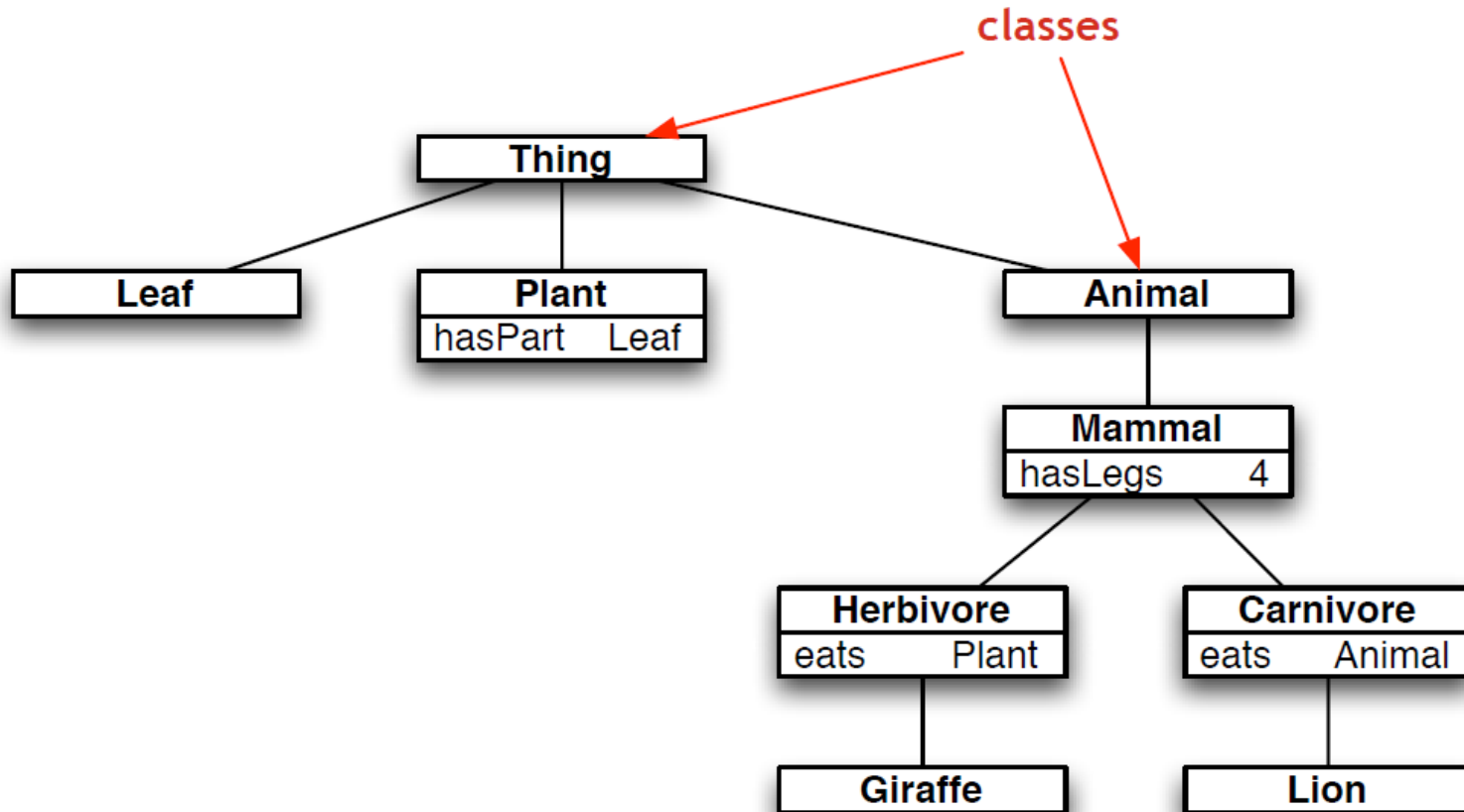# Hierarchies and Frames

- Frames are a way of describing classes or concepts or types

- Usual to think of classes in terms of sets of individuals

- Frames contain slots with values, which can be restricted in various ways:

  - integer, boolean or literal values

  - enumerated values

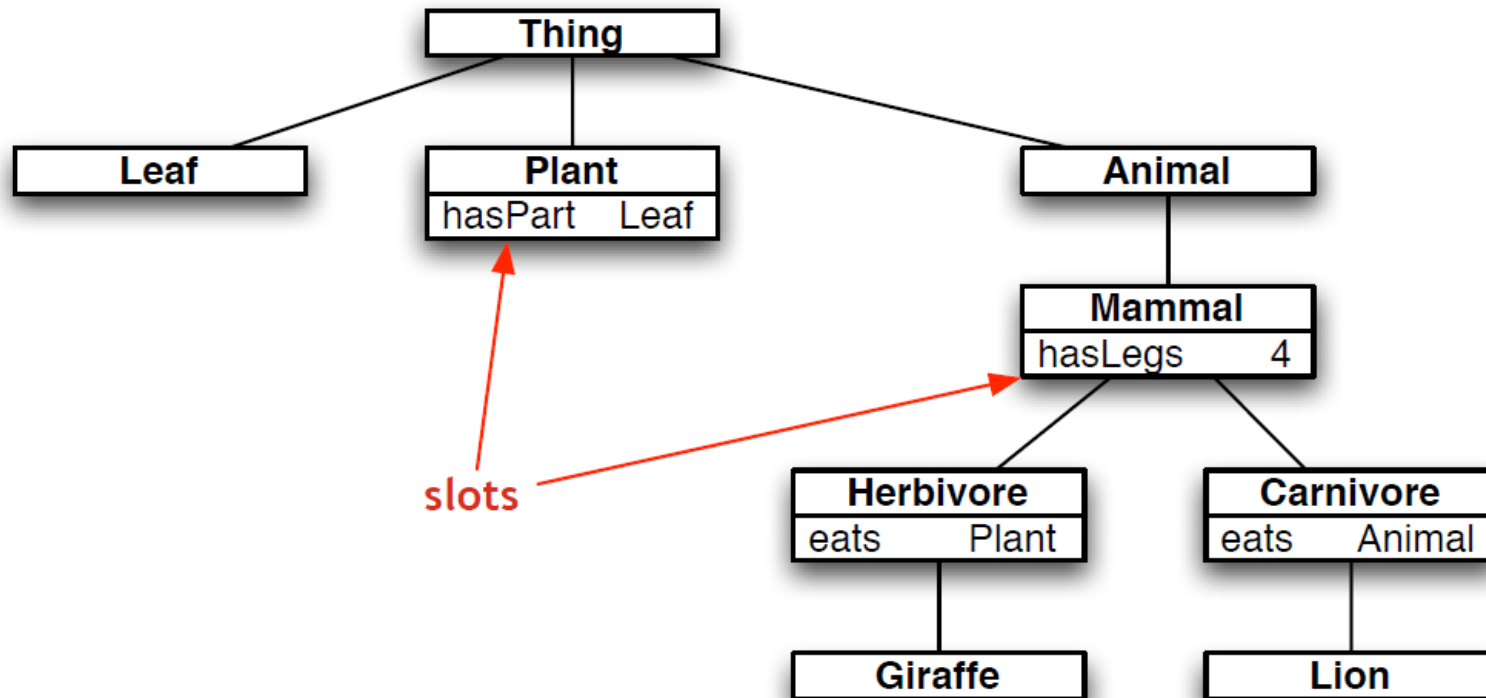 - instances of a specified class
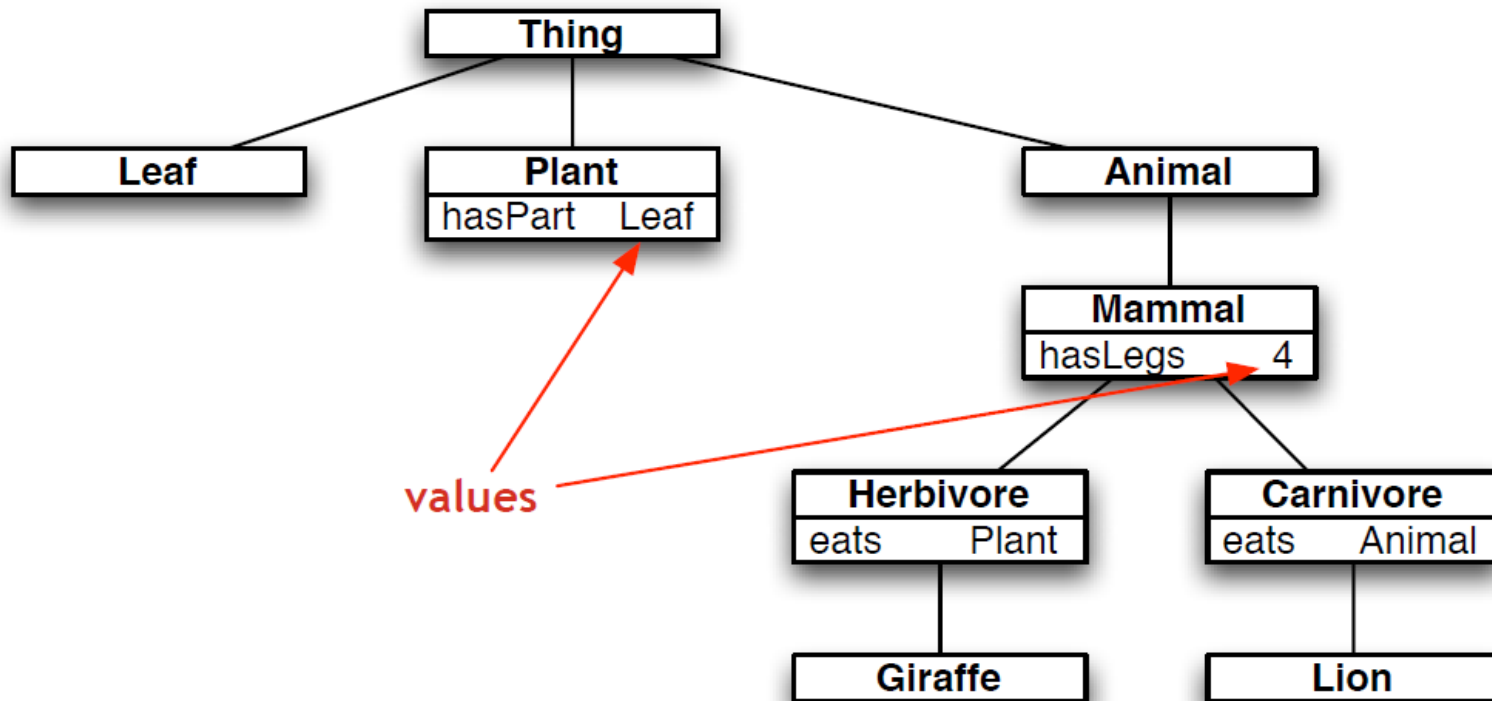
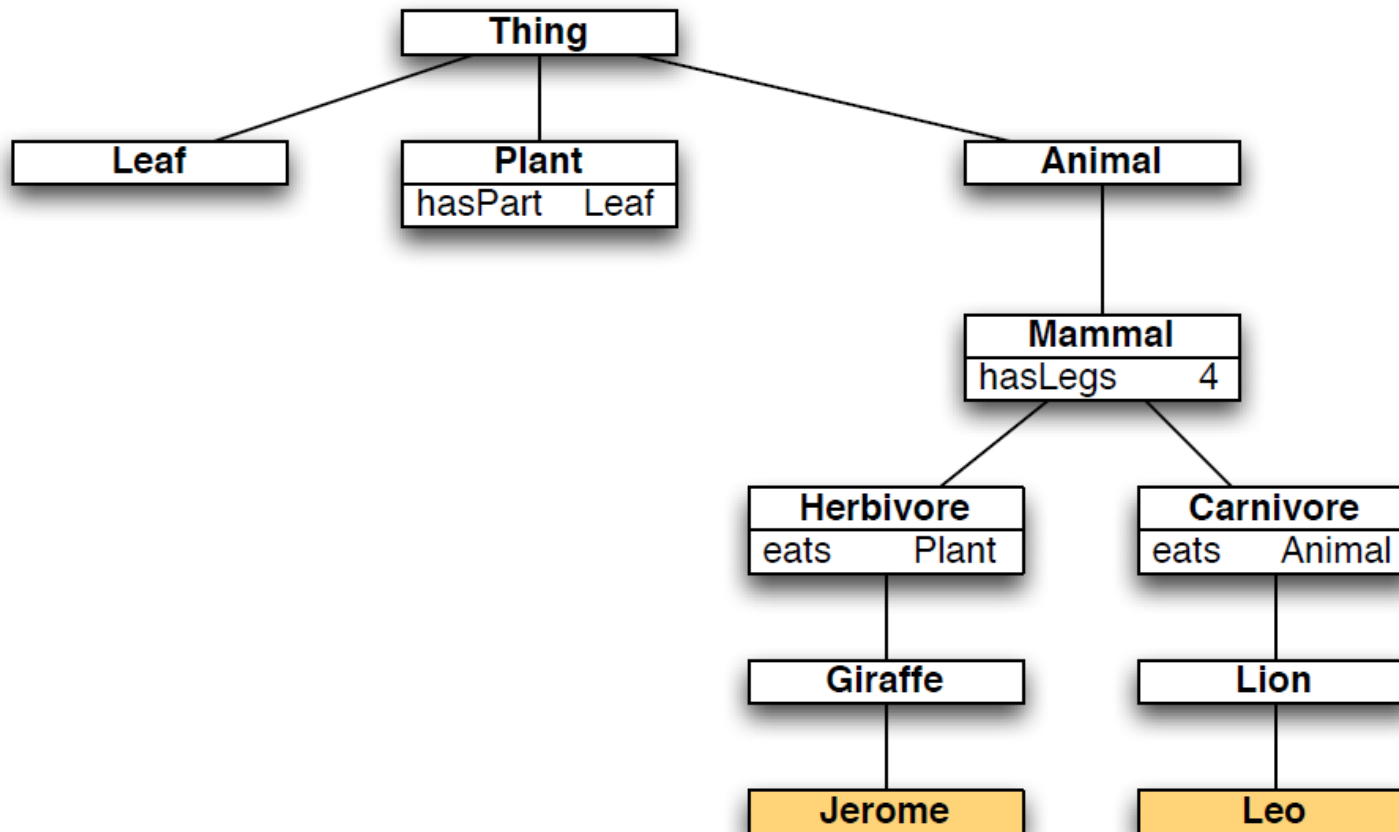# Hierarchies and Frames

Hierarchies and Frames

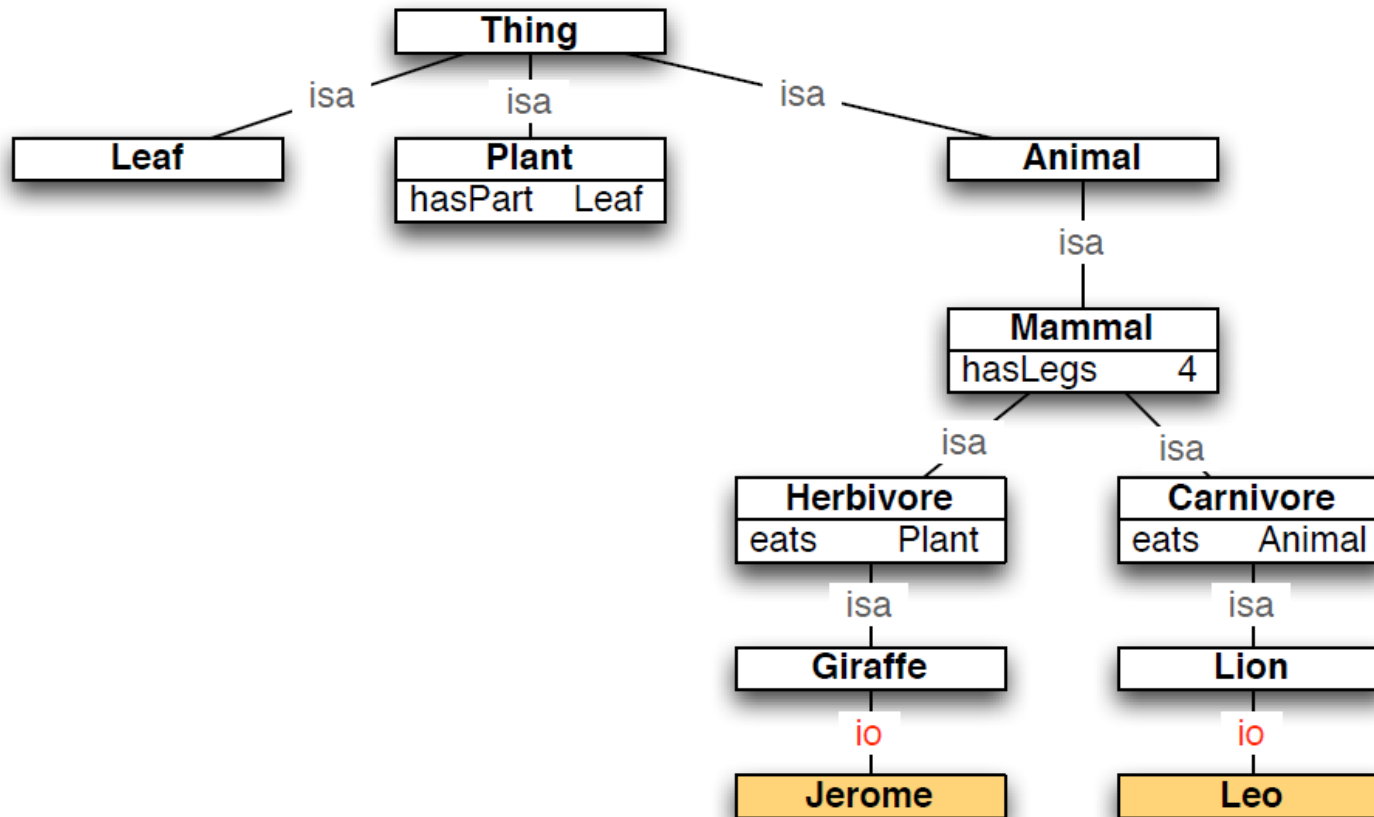# Hierarchies and Frames

Hierarchies and Frames

Classes and Individuals

Classes and Individuals

- Ambiguity about nature of the edge in the graph. Reflected in English:
  - A lion *is a* carnivore
  - Jerome *is a* giraffe
- Two different relations / labels:
  - ISA: taxonomic — a lion is a kind of carnivore
    - Lion ⊆ Carnivore
  - IO: instance-of / membership — Jerome is a member of the class of giraffes
    - Jerome ∈ Giraffe

Classes and Individuals

Inheritance

How many legs does Jerome have? 4

⇒ Jerome is an instance of Giraffe

⇒ Every instance of Giraffe is an instance of Herbivore

⇒ Every instance of Herbivore is an instance of Mammal

⇒ Mammals have 4 legs

So the attribute of having 4 legs is inherited by Giraffe from Mammal

Assertion vs Terminology

- **Assertions**: simple facts about the world
  - Joe is married to Sue
  - Bill has a brother with no children
- **Terminology**: general properties of concepts
  - *parent* is a superclass of *father*
  - *brother* is *sibling* restricted to *males*
- The KRYPTON system (Brachman, Fikes Levesque, 1983) proposed dividing a knowledge representation system into two main components:
  - ABox (assertions)
  - TBox (terminological structure)

Ontology Components

- Possible components include:
  - Individuals
  - Classes
  - Attributes
  - Relations
  - Functions
  - Axioms
  - Planning rules

Ontology Components – Individuals

- Individuals are instances or objects

- These are:
  - usually concrete,
    e.g. areti_manataki, uk_prime_minister,
    uoe_student_1389203
  - or sometimes abstract, e.g. numbers and words

- Two individuals may be equivalent
  - uk_prime_minister and david_cameron

- It is not always clear whether something ought to be an individual or a class
  - uk_prime_minister

Ontology Components – Classes

- Classes are used to group things together and they represent
    - concrete concepts, e.g. Mother, Bed, Muffin
    - abstract concepts, e.g. Love, Speed, Anger
- In most representations, members of classes must be individuals.
    - For instance, Helen, Anna and Rhona are individuals of class Mother.
- Subclasses and superclasses: Classes subsumed by, or subsuming other classes
- Taxonomies are ontologies that consist only of a class hierarchy.

Ontology Components – Attributes

- Attributes are aspects, properties, features, characteristics, or parameters that classes and individuals can have.
  - For example, the class Mother has attributes age and address
- Attributes can have different types of values:
  - boolean values (true/false)
  - specific values (integers, individuals or other literals)
  - classes
  - complex data types (e.g. enumerated lists)

Ontology Components – Relations

- Relations describe how classes relate to one another.
- Relations can be:
  - binary: relations between two classes,
  - the first argument is the domain of the relation, and the second argument is the range
    - MotherOf(Mother, Child)
  - n-ary: relations with n arguments, where n is unlimited
    - course(Course_Name,Lecturer,Level,Credits,Year)

Ontology Components – Functions

- Functions are relations where the n-th element of the relation is unique for the n-1 preceding elements.
  - plus(Addend, Addend, Result)
  - In the above function, if the two Addends are instantiated, there is only one possible value for Result.
- The functional nature of relations is often indicated by using the representation:
  plus(Addend, Addend) = Result

Ontology Components – Axioms

- Axioms model sentences that are always true and they describe how new facts can be derived from existing ones in the ontology.
  - sibling(X,Y) ∧ male(X) → brother(X,Y)
    - Note on notation: ∧ can be understood as "and", while → can be understood as "then"
- Given the above example, it is not necessary to store all the facts about brothers: if information exists about gender of individuals and sibling relations, then information about brothers can be derived when required.

Ontology Components – Planning rules

- Rules describe how the world may be changed.
- They consist of antecedents (things that must be true before the rule is applied) and consequents (things that are made true by applying the rule).

in_stock(Item, Time1) ∧ has_money(Person, Amount, Time1) ∧ cost(Item,Price) ∧ Amount ≥ Price
→ has(Item,Person, Time2) ∧ has_money(Person, New_Amount, Time2) ∧ New_Amount = Amount - Price

Medical Ontologies

- Gene Ontology
  - http://www.geneontology.org/
  - It aims to address the need for consistent descriptions of gene products across databases.
  - It represents information about biological processes, cellular components and molecular functions.
- Disease Ontology
  - http://disease-ontology.org/
  - It provides descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts.

Medical Ontologies

- SNOMED-CT
  - https://www.snomed.org/snomed-ct
  - It is a collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.
  - It is the most comprehensive, multilingual clinical healthcare terminology in the world.

- ICD10
  - https://bioportal.bioontology.org/ontologies/ICD10
  - It is a classification of diseases and related health problems.
  - It is widely used worldwide.

Conclusions

- Semantic web data:
  - connected
  - machine-understandable
- Ontologies allow us to encode domain knowledge.
- Ontology components may include classes, attributes, relations, etc.
- There are several ontologies in the medical field.

Acknowledgements

The content of these slides was originally created for the Medical Informatics course from The University of Edinburgh, which is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license.

These lecture slides are also licensed under a CC BY-SA 4.0 license.