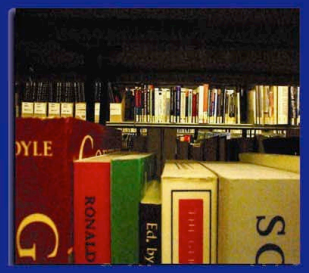
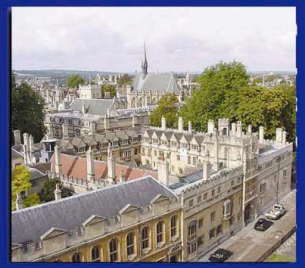




XML

Summer School 2005

A Practical Introduction to Ontologies & OWL

Session 1: Introduction

Duncan Hull and Nick Drummond

MANCHESTER 1824

Acknowledgements

- Tutorial developed by Biomedical Informatics Group in Manchester (in alphabetical order)
  Mike Bada, Sean Bechhofer, Carole Goble, Matthew Horridge, Ian Horrocks, Alan Rector, Jeremy Rogers, Robert Stevens, Chris Wroe

- Protégé team at Stanford
- Co-ode project, Funding agencies JISC, EPSRC

Tutorial outline

1. Introduction, Theory and Motivation to The Web Ontology Language (OWL)
2. Primitive classes in OWL
3. Defined classes and additional constructs
4. Common errors and how to correct them, Q&A session
5. 1.5 hours per session

About you - About us

- Clinicians / Information managers / digital libraries / healthcare / pharmaceutical / Software / Knowledge engineers…???
- Assume you have no prior exposure to RDF/OWL, ontologies or Protégé - probably not true!

- Duncan: biomedical ontologies, background in software engineering and biology
- Nick: Research Associate on www.co-ode.org in Medical Informatics

Overview: Session 1

- Introduction to Ontologies
- The Pizza Ontology
- Informal Modelling – Card Sorting
- OWL Overview
- Classes vs Individuals
- Protégé Introduction
- Notes about the Exercises

MANCHESTER 1824

XML Summer School 2005

Introduction to Ontologies

Hard Work using the Syntactic Web…

Find images of Tim Bray and Alan Rector…

Google™ Image Search

Tim Bray

Sun Microsystems

Rev. *Alan* M. Gates, Associate *Rector* of the Church of the Holy Spirit, Lake Forest, Illinois

Prof *Alan Rector*, University of Manchester, Medical Informatics Group

MANCHESTER 1824

XML Summer School 2005

Impossible (?) using the Syntactic Web...

- Complex queries involving *background knowledge*
  - Find information about "animals that use sonar but are not either bats or dolphins", e.g., Barn owl
- Locating information in *data repositories*
  - Travel enquiries
  - Prices of goods and services
- Finding and using "*web services*"
  - Book me a holiday next weekend somewhere warm, not too far away, and where they speak French or English

What is the Problem?

Consider a typical web page:

Markup consists of:

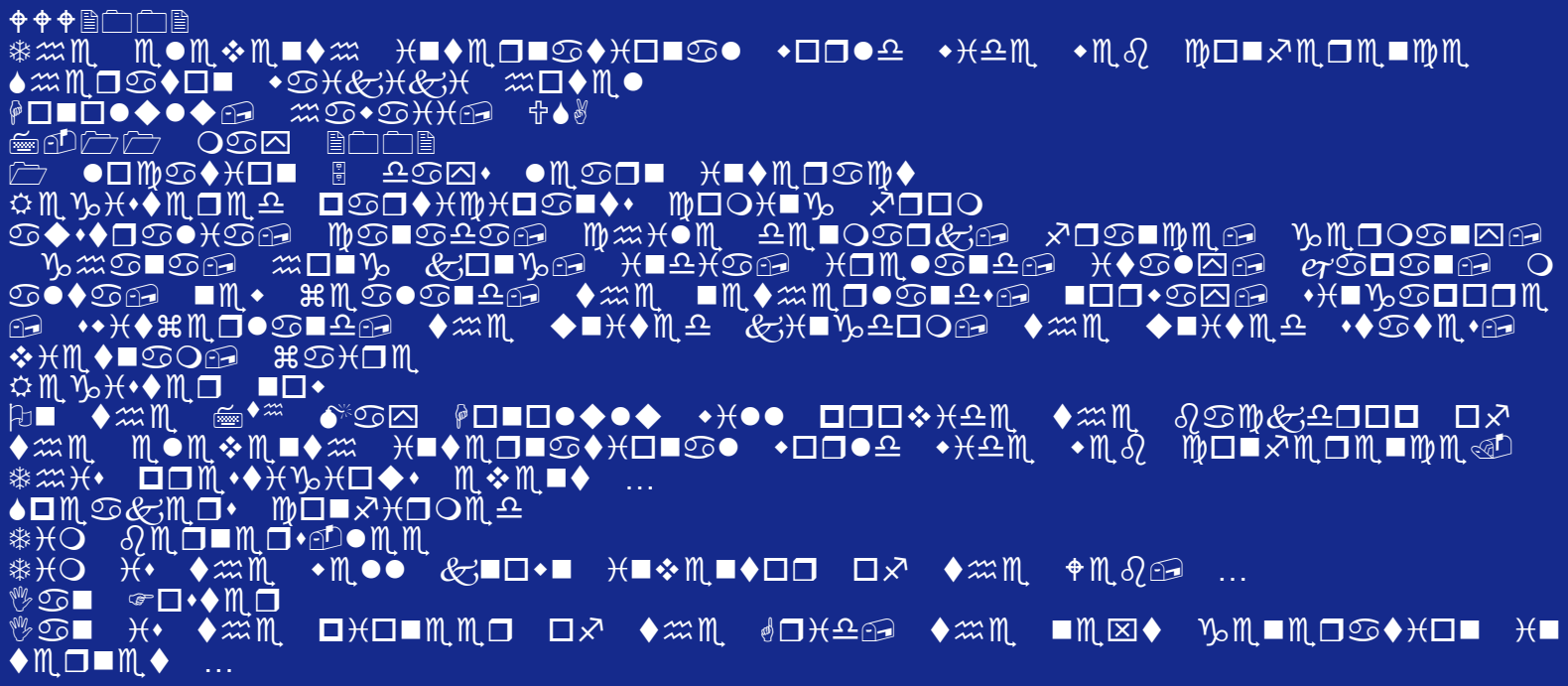
- rendering information (e.g., font size and colour)
- Hyper-links to related content

Semantic content is accessible to humans but not (easily) to computers…

What information can we see…

WWW2002
The eleventh international world wide web conference
Sheraton waikiki hotel
Honolulu, hawaii, USA
7-11 may 2002
1 location 5 days learn interact
Registered participants coming from
australia, canada, chile denmark, france, germany, ghana, hong kong, india, ireland, italy, japan, malta, new zealand, the netherlands, norway, singapore, switzerland, the united kingdom, the united states, vietnam, zaire
Register now
On the 7th May Honolulu will provide the backdrop of the eleventh international world wide web conference. This prestigious event …
Speakers confirmed
Tim berners-lee
Tim is the well known inventor of the Web, …
Ian Foster
Ian is the pioneer of the Grid, the next generation internet …

What information can a machine see...



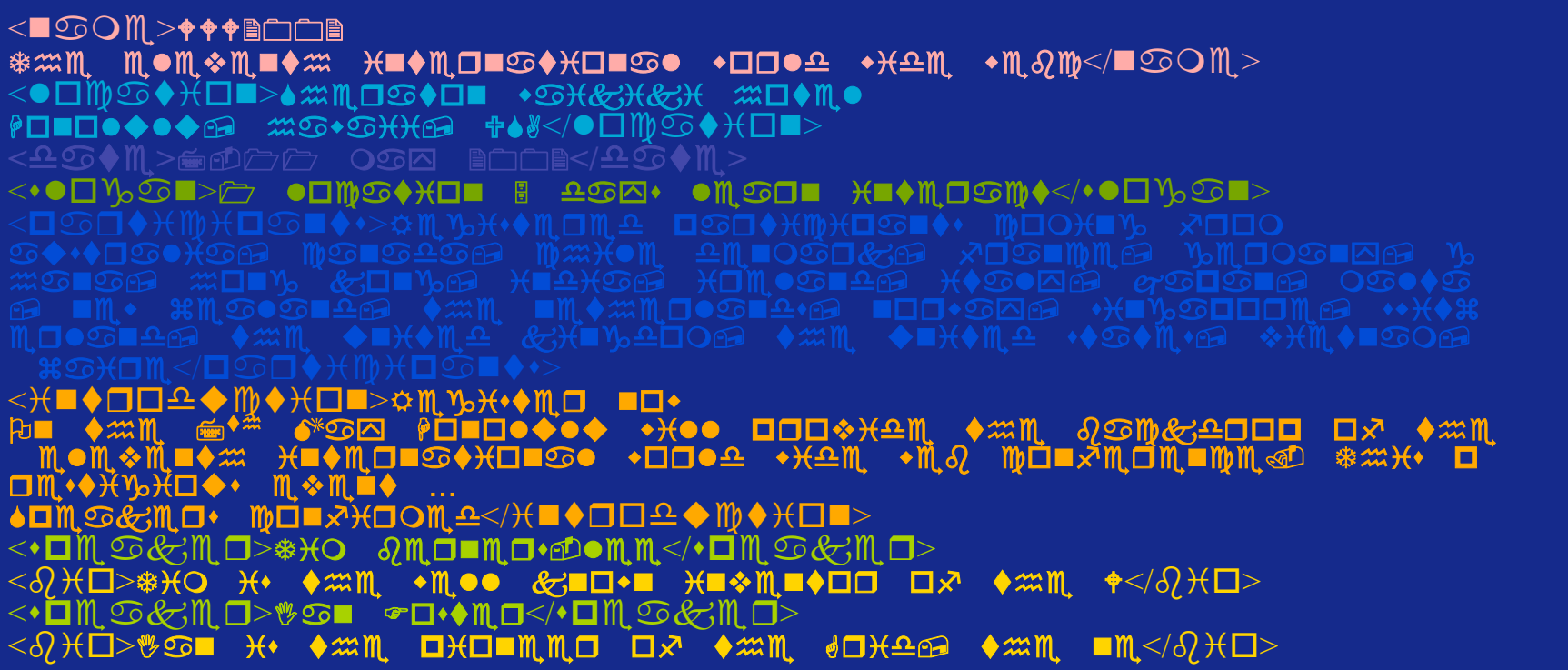
Solution: XML markup with "meaningful" tags?

<name>Jeni Tennison</name>
<location>Manchester, UK</location>
<date>[illegible]</date>
<slogan>[illegible]</slogan>
<participants>[illegible]</participants>
<introduction>[illegible]</introduction>
<speaker>[illegible]</speaker>
<bio>[illegible]</bio>...

But What About...

<conf>⌖⌖⌖🗎🗀🗀🗎❄♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■☋♦♓❒■☋● ⬩❒❒●♎ ⬩♓♎♏ ⬩♏♌♍❒■
</conf>
<place>♦♒♏❒☋♦❒■ ⬩☋♓&♓&♓ ♒❒♦♏●⁉❒■❒●◆●◆🖃 ♒☋⬩☋♓♓🖃 ✝♦✌
</place>
<date>🖮🖃🖁🖁 ○☋🖂 🗎🗀🗀🗎</date>
<slogan>🖁 ●❒♍☋♦♓❒■ 🕯 ♎☋🖂⬩ ●♏☋❒■ ♓■♦♏❒☋♍♦</slogan>
<participants>☼♏♑♓⬩♦♏❒♏♎ ❒☋❒♦♓♍♓❒☋■◆⬩ ♍❒○♓■♑ ➶❒❒○☋◆⬩♦❒☋●♓☋ 🖃 ♍☋■☋♎☋🖃 ♍♒♓●♏ ♎♏■○☋❒&🖃 ➶❒☋■♍♏🖃 ♑♏❒○☋■🖂🖃 ♑♒☋■☋🖃 ♒❒ ■♑ &❒■♑🖃 ♓■♎♓☋🖃 ♓❒♏●☋■♎🖃 ♓♦☋●🖂🖃 ☙☋❒☋■🖃 ○☋●♦☋🖃 ■♏⬩ ⌘♏☋ ●☋■♎🖃 ♦♒♏ ■♏♦♒♏❒●☋■♎⬩🖃 ■❒❒⬩☋🖂🖃 ⬩♓■♑☋❒❒❒♏🖃 ⬩⬩♓♦⌘♏❒●☋■♎🖃 ♦♒♏ ◆■♓♦♏♎ &♓■♑♎❒○🖃 ♦♒♏ ◆■♓♦♏♎ ⬩♦☋♦♏⬩🖃 ❖♓♏♦■☋○🖃 ⌘☋♓❒♏
</participants>
<introduction>☼♏♑♓⬩♦♏❒ ■❒⬩⚐■ ♦♒♏ 🖮♦♒ 💣☋🖂 ⁉❒■❒●◆●◆ ⬩♓●● ❒❒❒ ❖♓♎♏ ♦♒♏ ♌☋♍&♎❒❒❒ ❒➶ ♦♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■☋♦♓❒■☋● ⬩❒❒●♎ ⬩♓♎♏ ⬩♏♌ ♍❒■➶♏❒♏■♍♏🖃 ❄♒♓⬩ ❒❒♏⬩♦♓♑♓❒◆⬩ ♏❖♏■♦ …
♦❒♏☋&♏❒⬩ ♍❒■➶♓❒○♏♎</introduction>
<speaker>❄♓○ ♌♏❒■♏❒⬩🖃●♏♏</speaker>
<bio>❄♓○ ♓⬩ ♦♒♏ ⬩♏●● &■❒⬩■ ♓■❖♏■♦❒❒ ❒➶ ♦♒♏ ⌖♏♌🖃…

Machine sees...



Need to Add "Semantics"

- *External agreement* on meaning of annotations
  - E.g., Dublin Core "interoperable online metadata standards" author, title, date of creation or publication, subject coverage etc
    - Agree on the meaning of annotation tags
  - Problems with this approach (RDFS - later)
    - Inflexible, Limited number of things can be expressed
- Use *Ontologies* to specify meaning of annotations
  - Ontologies provide a vocabulary of terms
  - New terms can be formed by combining existing ones
  - Meaning (*semantics*) of such terms is formally specified
  - Can also specify relationships between terms in multiple ontologies

History of the Semantic Web

- Web was "invented" by Tim Berners-Lee (amongst others), a physicist working at CERN
- TBL's original vision of the Web was much more ambitious than the reality of the existing (syntactic) Web…

- TBL (and others) have since been working towards realising this vision, which has become known as the "*Semantic Web*"
  - E.g., article in May 2001 issue of Scientific American…

Scientific American, May 2001:

Beware of the Hype

Beware of the Hype

- Hype seems to suggest that Semantic Web means: "semantics + web = AI"
  - "A new form of Web content that is meaningful to computers will unleash a revolution of new abilities"
- More realistic to think of it as meaning: "semantics + web + AI = more useful web"
  - Realising the complete "vision" is too hard for now (probably)
  - But we can make a start by adding semantic annotation to web resources (e.g. pages, services) using ontologies

Ontology: Origins and History

- BEWARE!!! Ontology is a heavily overloaded term with (ironically) several different meanings in different disciplines (Philosophy, Linguistics and Computer Science)
- E.g. In Philosophy, ontology deals with the *nature and organisation of reality*
- Science of Being
  (Aristotle, Metaphysics, IV, 1)

- Tries to answer the questions:

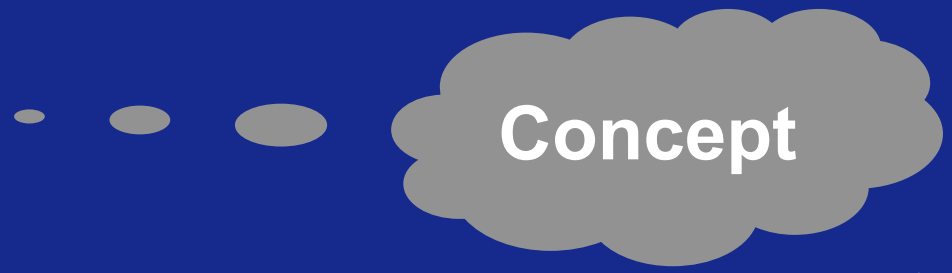
  What characterizes being?

  Eventually, what is being?

MANCHESTER 1824

XML Summer School 2005

Ontology in Linguistics



"a concept, is the mediator that relates the symbol to its object" John F Sowa

activates

"semiotic triangle"

Relates to

Form / Symbol

Stands for

Referent

"Tank"

?

[Ogden, Richards, 1923]

Ontology in Computer Science

- A definition from Computer Science
  Rudi Studer(98): An ontology is a

Formal, explicit specification of shared conceptualisation

Machine readable, for computers, not just humans!

Concepts and properties are explicitly defined

Consensual knowledge

Model of something in the world (e.g. pizzas, genes, car parts etc)

(Thanks to Marta Sabou for this slide)

Structure of an Ontology

Ontologies typically have two distinct components: Terminology

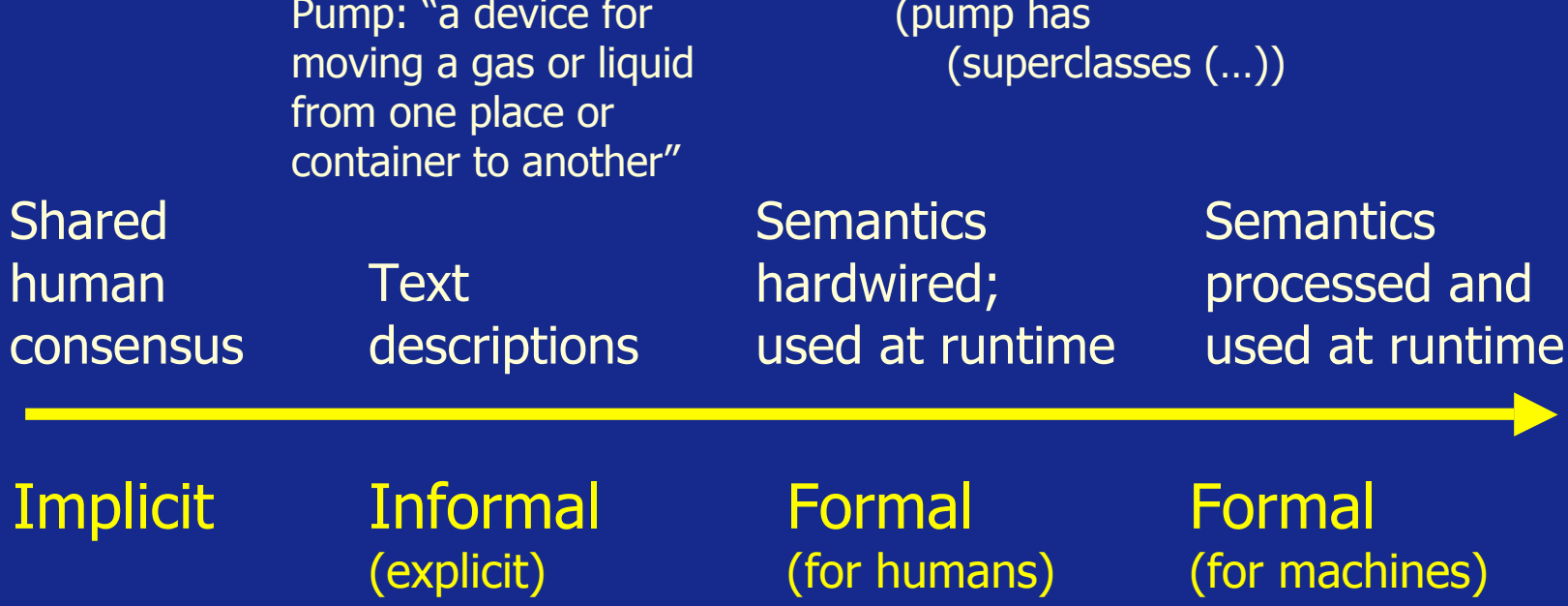
- Names for important concepts in the domain
  - Elephant is a concept whose members are a kind of animal
  - Herbivore is a concept whose members are exactly those animals who eat only plants or parts of plants
  - Adult_Elephant is a concept whose members are exactly those elephants whose age is > 20 years
- Background knowledge/constraints on the domain: Axioms
  - Adult_Elephants weigh at least 2,000 kg
  - All Elephants are either African_Elephants or Indian_Elephants
  - No individual can be both a Herbivore and a Carnivore

[Mike Uschold, Boeing Corp]

A semantic continuum

Pump: "a device for moving a gas or liquid from one place or container to another"

(pump has
(superclasses (...))



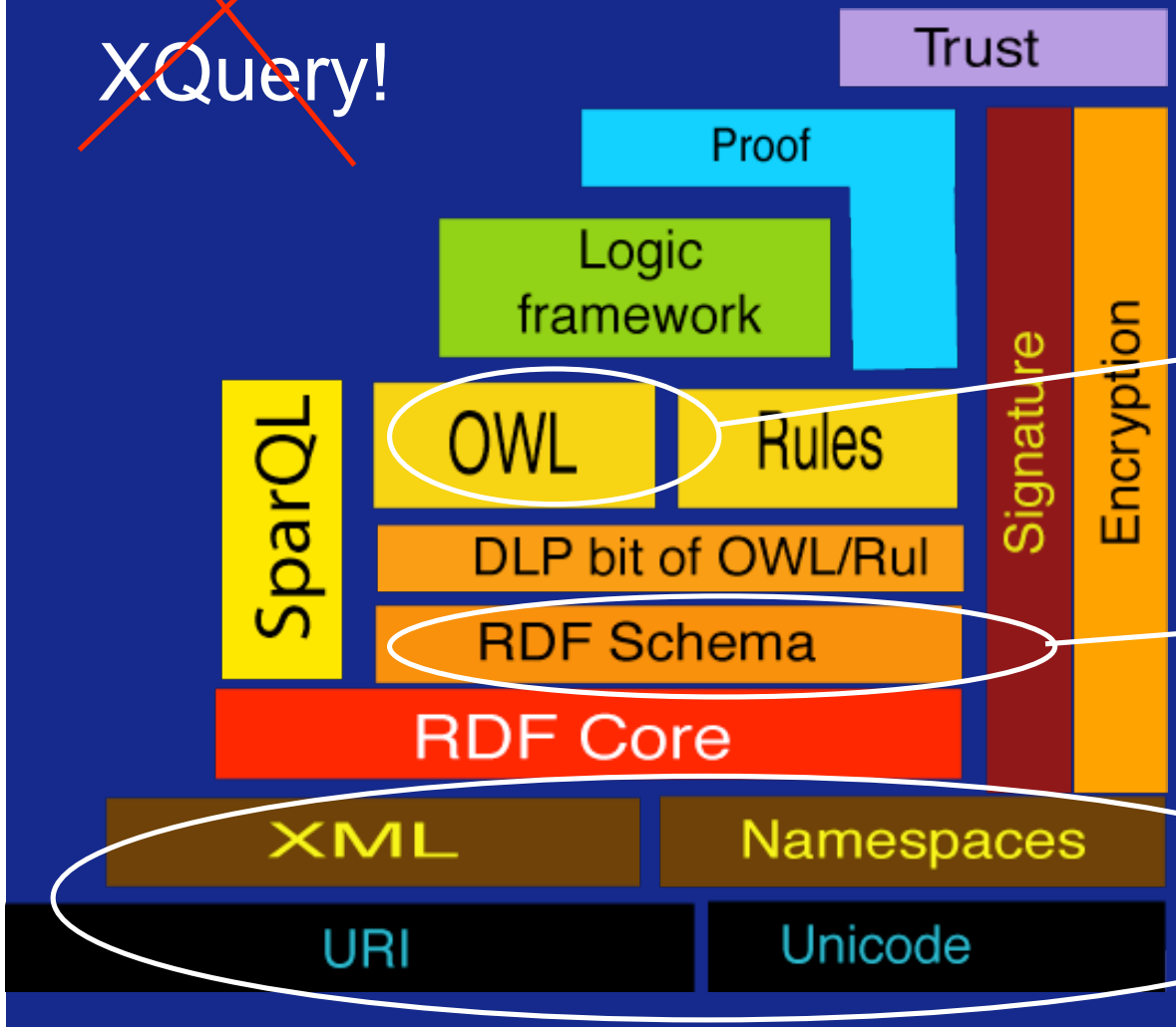
→ Further to the right →

- Less ambiguity
- Better inter-operation
- More robust – less hardwiring
- More difficult

Semantic Web Stack (wrong!)

~~XSLT!~~

~~XQuery!~~

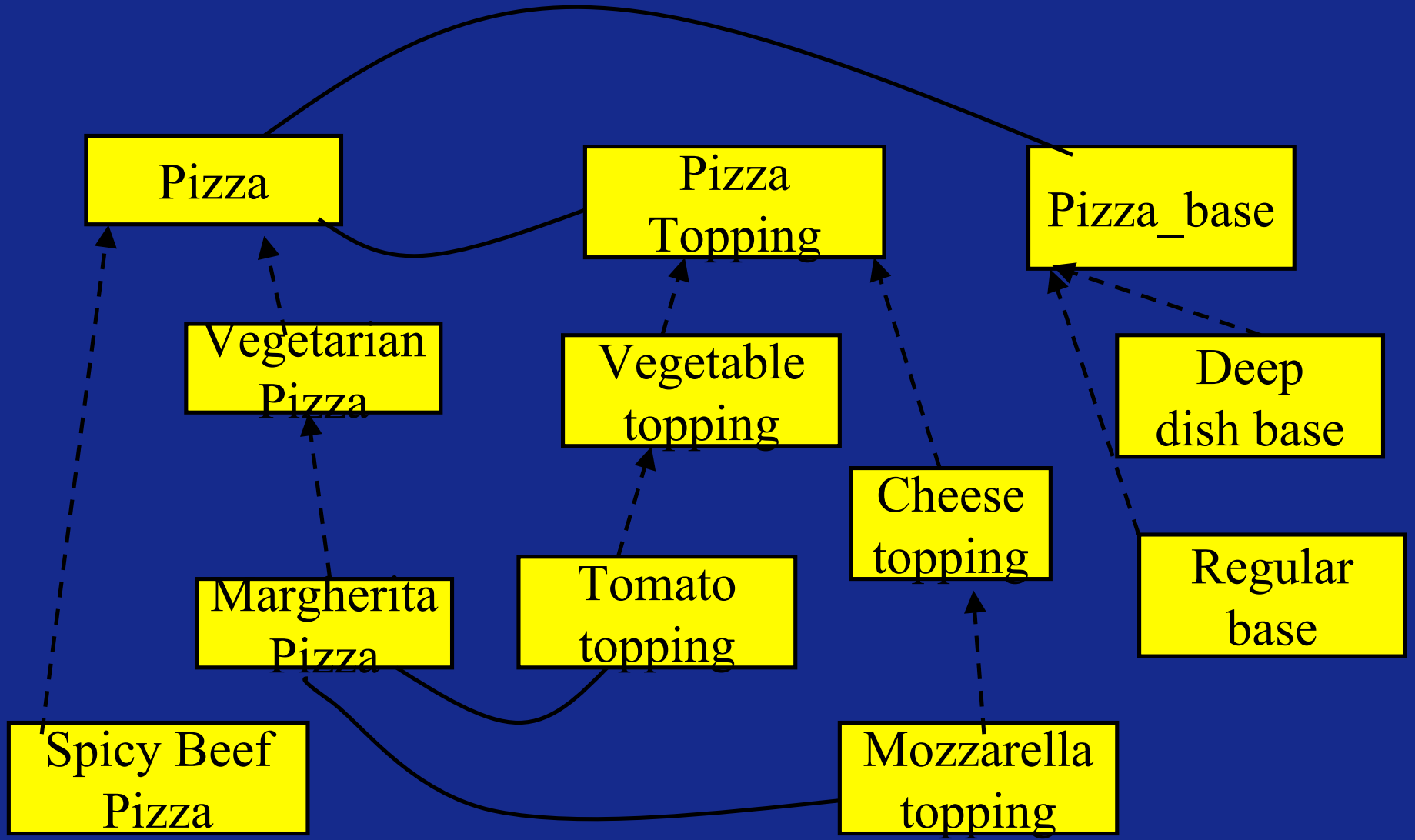


Today!

e.g. Dublin core

XML summer school

A simple ontology: Pizzas



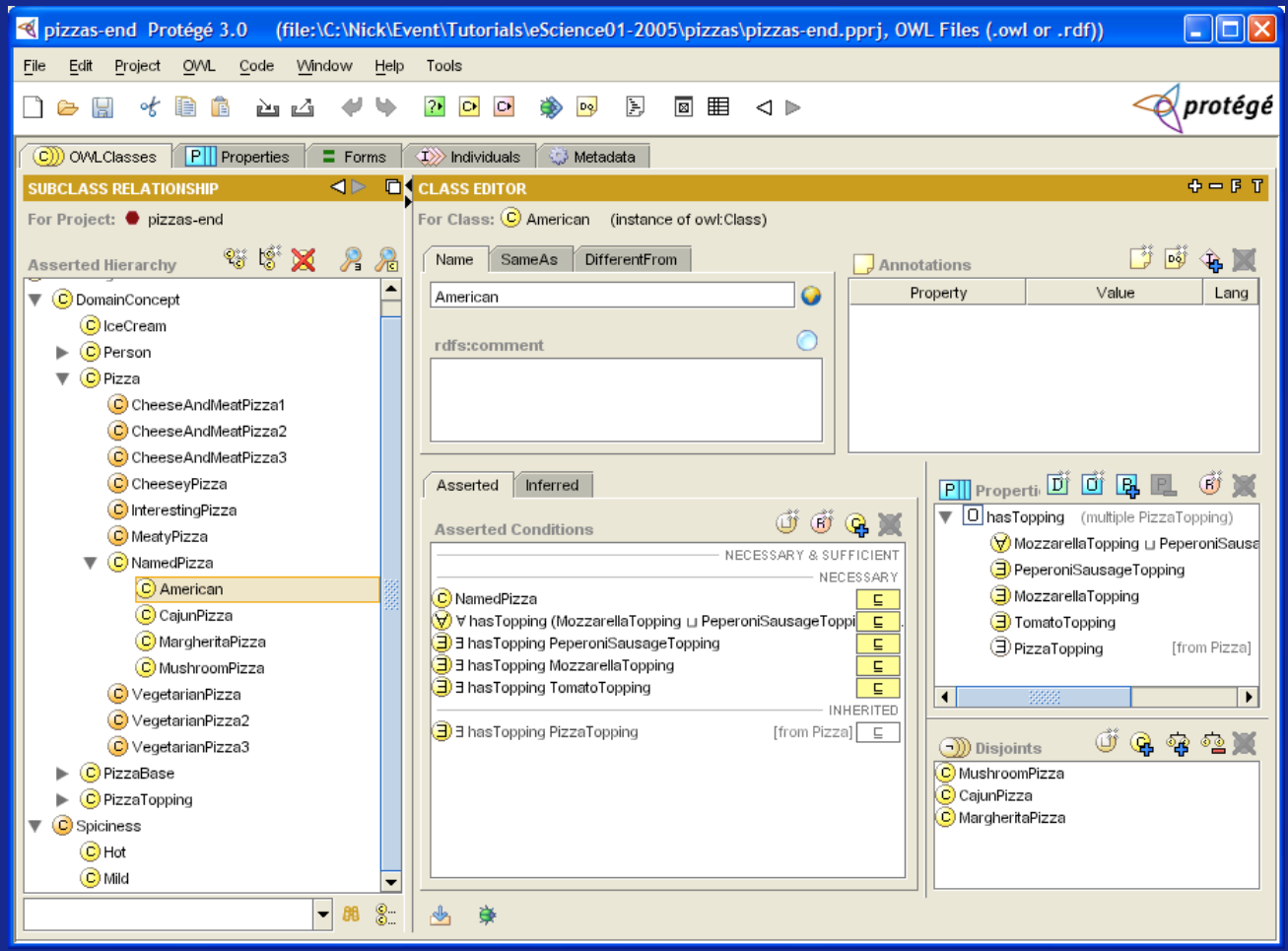
Ontology Design and Deployment

- Given key role of ontologies in the Semantic Web, it will be essential to provide *tools* and *services* to help users:
  - Design and maintain high quality ontologies:
    - *Meaningful* — all named classes can have instances
    - *Correct* — captured intuitions of domain experts
    - *Minimally redundant* — no unintended synonyms
    - *Richly axiomatised* — (sufficiently) detailed descriptions
  - Store (large numbers) of *instances* of ontology classes, e.g.: Annotations from web pages
  - Answer *queries* over ontology classes and instances:
    - Find more general/specific classes
    - Retrieve annotations/pages matching a given description
  - *Integrate* and align multiple ontologies

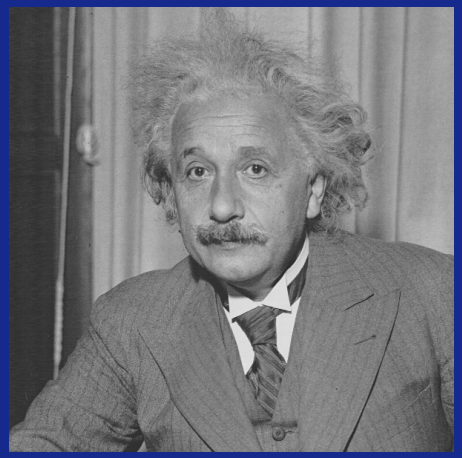
MANCHESTER 1824

XML Summer School 2005

Example Ontology



Why it's hard (1)



- Clash of intuitions
  - Subject Matter Experts (you?) motivated by custom & practice
    - Prototypes & Generalities
  - Mathematicians, Logicians, computer scientists motivated by logic, computability & novelty
- Transparency & predictability vs Rigour & Completeness
- Potential users caught in the muddled middle

Why it's hard (2)

- Confusion of terminology and usage
  - Religious wars over words and assumptions
  - Legacy issues (e.g. SNOMED-CT)
- The intersection of
  - Linguistics
  - Cognitive science
  - Software engineering
  - Philosophy
- Human Factors – cost, time, distribution e.g. www.geneontology.org/

Vocabulary

- "Class" ≈ "Concept" ≈ "Category" ≈ "Type"
- "Instance" ≈ "Individual"
- "Entity" ≈ "object", Class or individual
- "Property" ≈ "Slot" ≈ "Relation" ≈ "Relationtype" ≈ "Attribute" ≈ "Semantic link type" ≈ "Role"
  - but be careful about "role"
    - Means "property" in Description Logic-speak
    - Means "role played" in most ontologies
      - E.g. "doctor_role", "student role" …

The Pizza Ontology

The Manchester Pizza Finder

MANCHESTER 1824

XML Summer School 2005

Our Domain

- Pizzas have been used in Manchester tutorials for years.
- Pizzas were selected as a domain for several reasons:
  - They are fun
  - They are internationally known
  - They are highly compositional
  - They have a natural limit to their scope
  - They are fairly neutral
    - Although arguments still break out over representation
    - Even pizzas can do this - its an inevitable part of knowledge modelling
    - ARGUING IS NOT BAD!

You are the Expert

- Most often it is **not** the domain expert that formalises their knowledge – because of the complexity of the modelling it is normally a specialist "knowledge engineer"
  Hopefully, as tools get easier to use, this will change
- Having access to experts is critical for most domains
- Luckily, we are all experts in Pizzas, so we just need some material to verify our knowledge…

Reference Materials

- Having references to validate decisions, and act as provenance can be useful for maintaining an ontology
- Mistakes, omissions and intentions can be more easily traced if a reference can be made
  - When building, we highly recommend documenting your model as you go – keeping provenance information is a good way of doing this
- We have provided you with a pizza menu and several cards with ingredients on

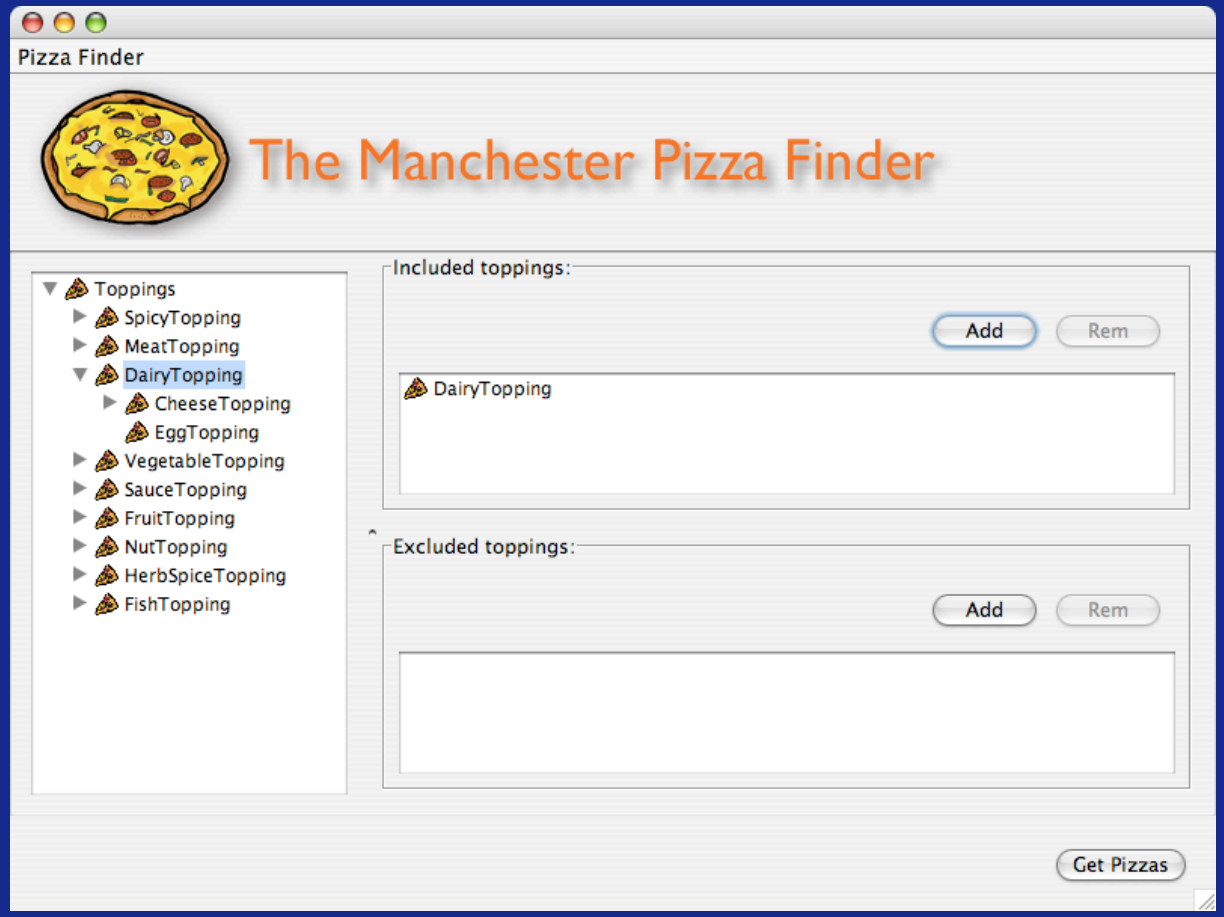
Our Ontology

- When building an ontology we need an application in mind – ontologies should not be built for the sake of it
- Keep the application in mind when creating concepts – this should help you scope the project

- The PizzaFinder application has been developed so that you can plug your ontology in at the end of the day and see it in action

  Let us know your ideas for extending the application

Our Application



www.co-ode.org/downloads/pizzafinder/

Exercise 1: Card Sorting

Card Sorting - Issues

- different viewpoints
  - Tomato – Vegetable or Fruit?
  - culinary vs biological
- Ambiguity
  - words not concepts
- Missing Knowledge
  - What is peperonata?
- multiple classifications (2+ parents)
- lots of missing categories (superclasses?)
- competency questions
  - What are we likely to want to "ask" our ontology?
  - bear the application in mind

OWL Overview

OWL...

- Is the Web Ontology Language
- Is part of the W3Cs Semantic Web framework
- Is a W3C standard
- Is used for capturing knowledge in a machine understandable way
- Has explicit semantics
- Is more semantically rich than RDFS

OWL Flavours

- OWL comes in 3 flavours
  - Lite – partially restricted to aid learning curve
  - DL = Description Logic
    Description Logics are a decidable fragment of First Order Logic (FOL)
  - Full – unrestricted use of OWL constructs, but undecidable
- we will be using OWL DL – because it is decidable, we can use a reasoner (an inference engine) to help us build our model

OWL Syntax

- OWL is a syntax independent language that has several common representations
  - Abstract Syntax
  - N3
  - RDF/XML

OWL Syntax: Abstract Syntax

- One of the clearer human-readable syntaxes

```
Class(SpicyPizza complete
   annotation(rdfs:label "PizzaTemperada"@pt)
   annotation(rdfs:comment "Any pizza that has a spicy topping
                            is a SpicyPizza"@en)
   Pizza
   restriction(hasTopping someValuesFrom(SpicyTopping))
)
```

OWL Syntax: N3

- Recommended for human-readable fragments

```
default:SpicyPizza
      a owl:Class ;
      rdfs:comment "Any pizza that has a spicy topping is a
                                    SpicyPizza"@en ;
      rdfs:label "PizzaTemperada"@pt ;
      owl:equivalentClass
              [ a owl:Class ;
                owl:intersectionOf (default:Pizza [ a owl:Restriction ;
                            owl:onProperty default:hasTopping ;
                            owl:someValuesFrom default:SpicyTopping
                          ])
              ] .
```

OWL Syntax: RDF/XML

- Recommended for serialisation

```
<owl:Class rdf:ID="SpicyPizza">
   <rdfs:label xml:lang="pt">PizzaTemperada</rdfs:label>
   <rdfs:comment xml:lang="en">Any pizza that has a spicy topping is a
   SpicyPizza</rdfs:comment>
   <owl:equivalentClass>
     <owl:Class>
       <owl:intersectionOf rdf:parseType="Collection">
         <owl:Class rdf:about="#Pizza"/>
         <owl:Restriction>
           <owl:onProperty>
             <owl:ObjectProperty rdf:about="#hasTopping"/>
           </owl:onProperty>
           <owl:someValuesFrom rdf:resource="#SpicyTopping"/>
         </owl:Restriction>
       </owl:intersectionOf>
     </owl:Class>
   </owl:equivalentClass>
 </owl:Class>
```

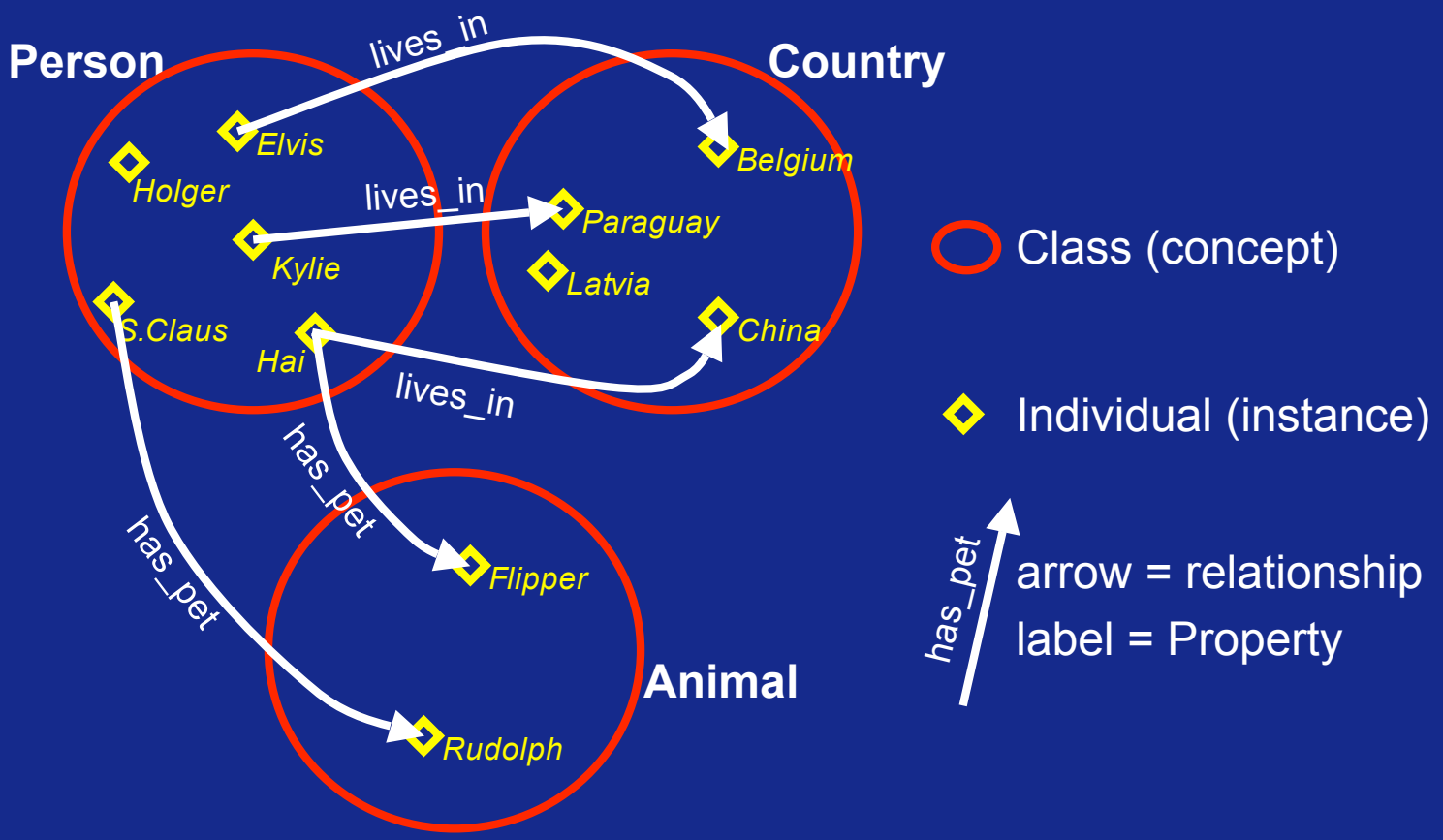
OWL: Creating/Editing

- Even with the expertise gained at the XML Summer School, editing the RDF/XML by hand is probably not recommended
- Ontologies range in size, but because of their explicit nature they require verbose definitions

- Thankfully we have tools to help us reduce the syntactic complexity
- However, the tools are still in the process of trying to reduce the semantic complexity

- Building ontologies in OWL is still hard

OWL Constructs



OWL Constructs: Classes

Eg Mammal, Tree, Person, Building, Fluid, Company

- Classes are sets of Individuals
- aka “Type”, “Concept”, “Category”
- Membership of a Class is dependent on its logical description, not its name
- Classes do not have to be named – they can be logical expressions – eg **things that have colour Blue**
- A Class should be described such that it is **possible** for it to contain Individuals (unless the intention is to represent the empty class)
- Classes that cannot possibly contain any Individuals are said to be **inconsistent**

OWL Constructs: Properties

Eg hasPart, isInhabitedBy, isNextTo, occursBefore

- Properties are used to relate Individuals
- We often say that Individuals are related **along** a given property
- Relationships in OWL are binary:

  Subject → predicate → Object

  Individual a → hasProperty → Individual b

  nick_drummond → givesTutorial → xml_summer_school_OWL_tutorial_2005
- N-ary relationships can be modelled with workarounds in OWL, but this design pattern will not be discussed today

OWL Constructs: Individuals

Eg me, you, this tutorial, this room

- Individuals are the objects in the domain
- aka “Instance”, “Object”
- Individuals may be (and are likely to be) a member of multiple Classes

Individuals vs Classes

Individuals vs Classes

Can it have kinds?

If so, it is a Class

Otherwise, it is an Individual

Egs:

- "Kinds of dog" makes sense
- "Kinds of pizza" makes sense
- "Kinds of Nick Drummond" does not make sense
- "Kinds of bacon" makes sense
- "Kinds of the one piece of bacon in this morning's breakfast" does not make sense

Individuals vs Classes

When you say something about it…

…if you made a new concept, it is a Class

…if you just stated a fact about it, it is an Individual

Egs:

- "Big dog" is a new subclass of dog
- "Rover is big" just says something about Rover
- "Mouldy Pizza Ingredient" is a new subclass of Pizza Ingredient
- "The mouldy bit of cheese on the pizza I bought last Friday" is a fact about the bit of cheese in your bad pizza experience

Clues in English

- Proper nouns (almost always) indicate individuals
  - Nick Drummond, Manchester, England, …
- Articles + singular indicate individual
  - "the book there on the shelf" – an individual
  - "a book" – an unspecified individual
- Plurals usually indicate classes
  - "the books" – probably a class (Although possibly an individual aggregation)
  - Perversely, we conventionally name classes in the singular

More clues in English

- a '…that…' clause usually indicates a class
  - "The people that drive buses"
- a '…which…" clause depends on local usage
  - Some English stylebooks would have 'which' clauses used only for individuals, others say there is no real difference between 'that' and 'which'
  - MS Word usually asks for 'that' with plurals (classes) and 'which' with singulars
- No perfect guide, must take case by case.

Leaf nodes are not Individuals

- Depends on ontology – may be very detailed:
  - Golden_retriever_bitch_from_karmella_kennels_from_2003_litter
  - Individual in that class "Halo"
- Even if there is only one possible individual, a leaf node is not an individual:
  - postie_for_the_Isle_of_Sheppey
  - There'll be a new postie each time the previous one gets blown off a cliff
- Only individuals if there *could never be* kinds

Keeping the Ontology Re-usable

- If we make leaf nodes individuals, we close off any extension to more granular kinds
  - The ontology is only specific to our immediate needs
  - Extensions require radical surgery

Pizza Toppings

Are your pizza toppings Classes or Individuals?

We will not be creating Individuals in this tutorial

- We are trying to demonstrate the benefits of using a reasoner to assist modelling
- Individuals create scalability issues with reasoning

Protégé and Protégé-OWL

MANCHESTER 1824

XML Summer School 2005

Protégé...

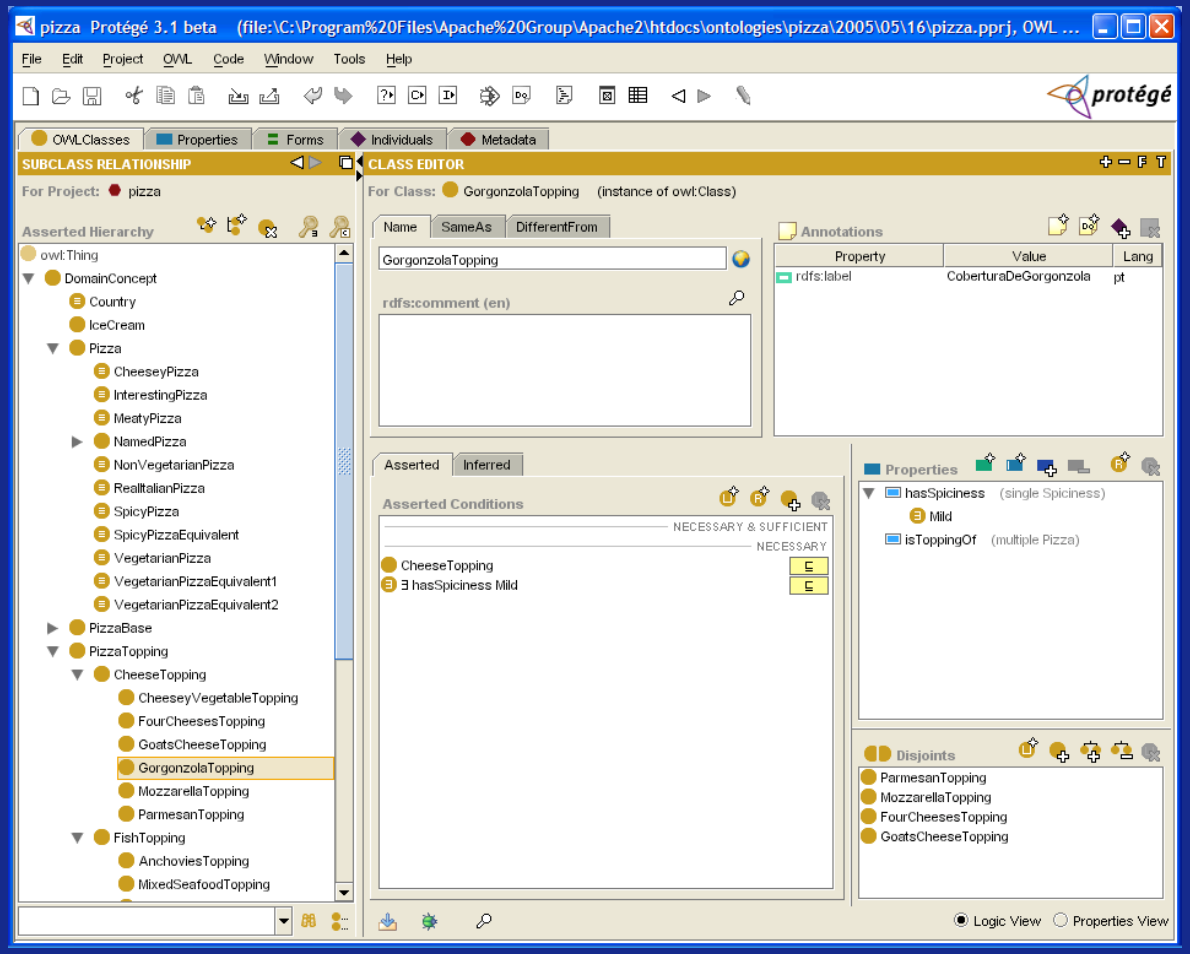
- Is a knowledge modelling environment
- Is free, open source software
- Is developed by Stanford Medical Informatics
- Has a large user community (approx 30k)

Protégé…

- core is based on Frames (object oriented) modelling
- has an open architecture that allows other modelling languages to be built on top
- supports development of plugins to allow backend / interface extensions
- now supports OWL through the Protégé-OWL plugin

So let's have a look…

Protégé-OWL



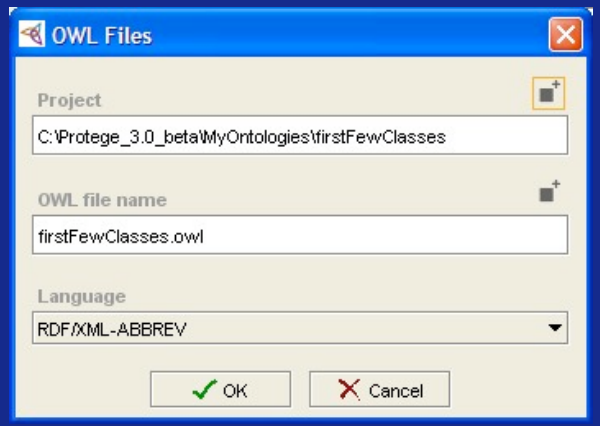
Saving OWL Files

OWL = easy to make mistakes = save regularly

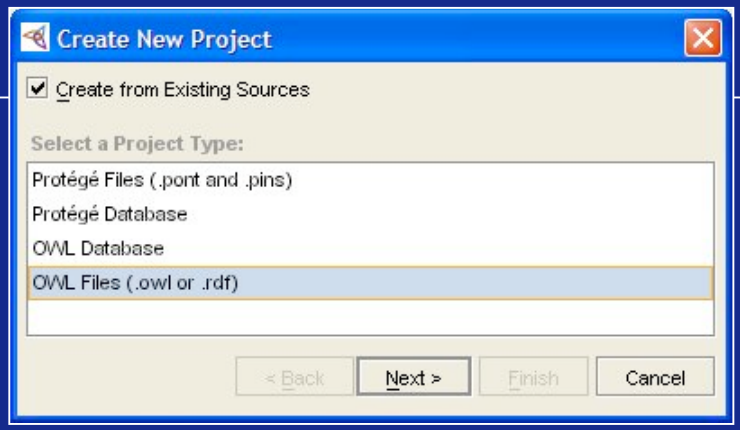
1. *Select File → Save Project As*
   A dialog (as shown) will pop up

2. *Select a file directly by clicking the button on the top right*
   You will notice that 2 files are created
   .pprj – the project file
   this just stores information about the GUI and the workspace
   .owl – the OWL file
   this is where your ontology is stored in RDF/OWL format

3. *Select OK*



Loading OWL files



1. *If you only have an OWL file:*
   - ***File→ New Project***
   - *Select **OWL Files** as the type*
   - *Tick **Create from existing sources***
   - ***Next** to select the .owl file*

2. *If you've got a valid project file*:*
   - ***File → Open Project***
   - *select the .pprj file*

** ie one created on this version of Protégé - the s/w gets updated once every few days, so don't count on it unless you've created it recently– safest to build from the .owl file if in doubt*

MANCHESTER 1824

XML Summer School 2005

Exercises

- You should each have a set of exercises to work through
- The idea of the practical sessions is to allow you to build your own Pizza ontology
- If you get stuck, or something goes wrong, there are demonstration OWL files for each step as a backup