Introduction to the Semantic Web

Semantic tecnologies – a quick overview

Fulvio Corno

Politecnico di Torino

Semantic Web

http://www.w3.org/2001/sw/

- Web second generation
  - Web 3.0
- "Conceptual structuring of the Web in an explicit machine-readable way" (Tim Berners-Lee)
- In other words…

…let the machine do most of the work!!!

‘Official’ introduction (1/2)

- The Semantic Web is **a web of data**. There is lots of data we all use every day, and its not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?
- Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

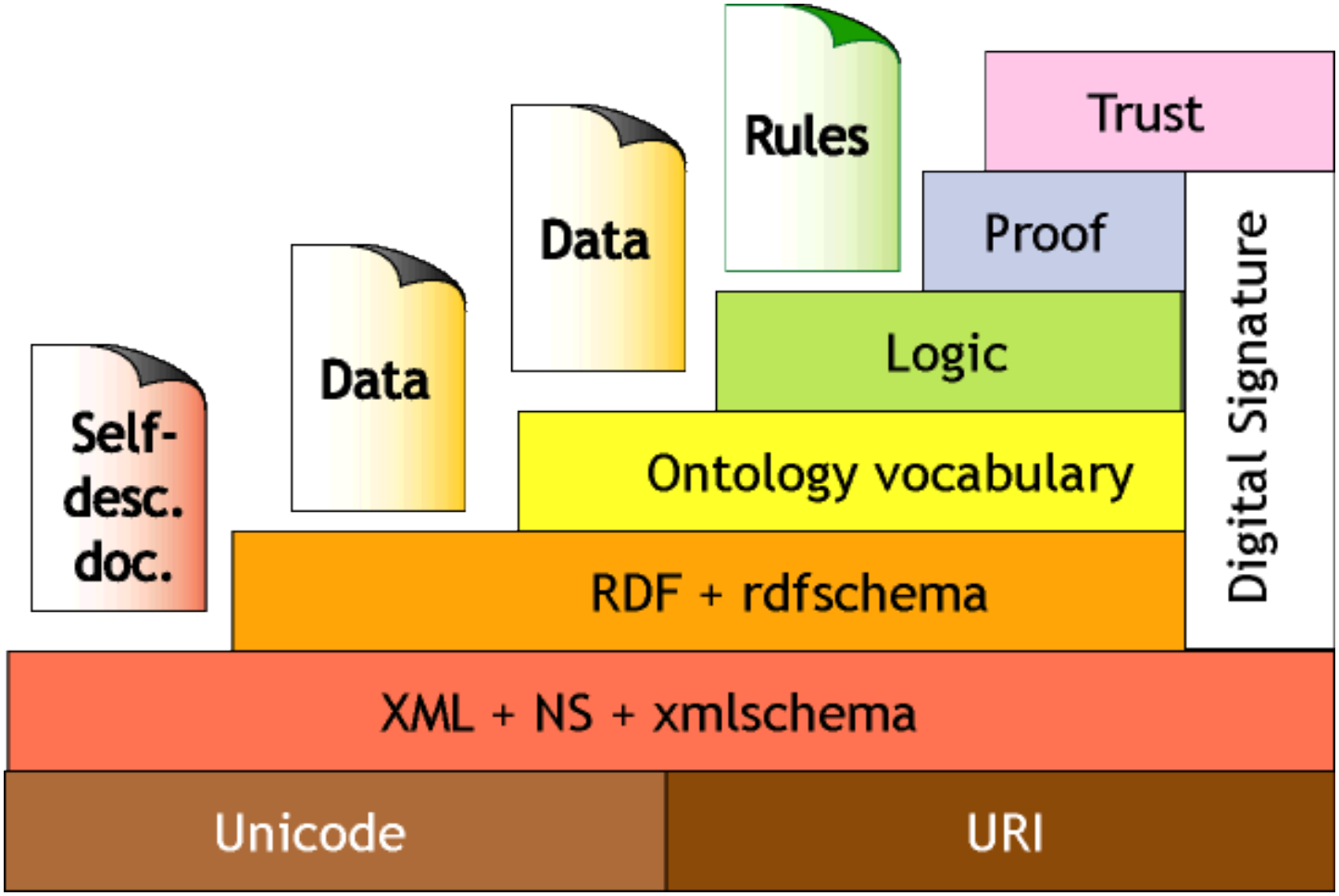
‘Official’ introduction (2/2)

- The Semantic Web is about two things. It is about **common formats** for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about **language** for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

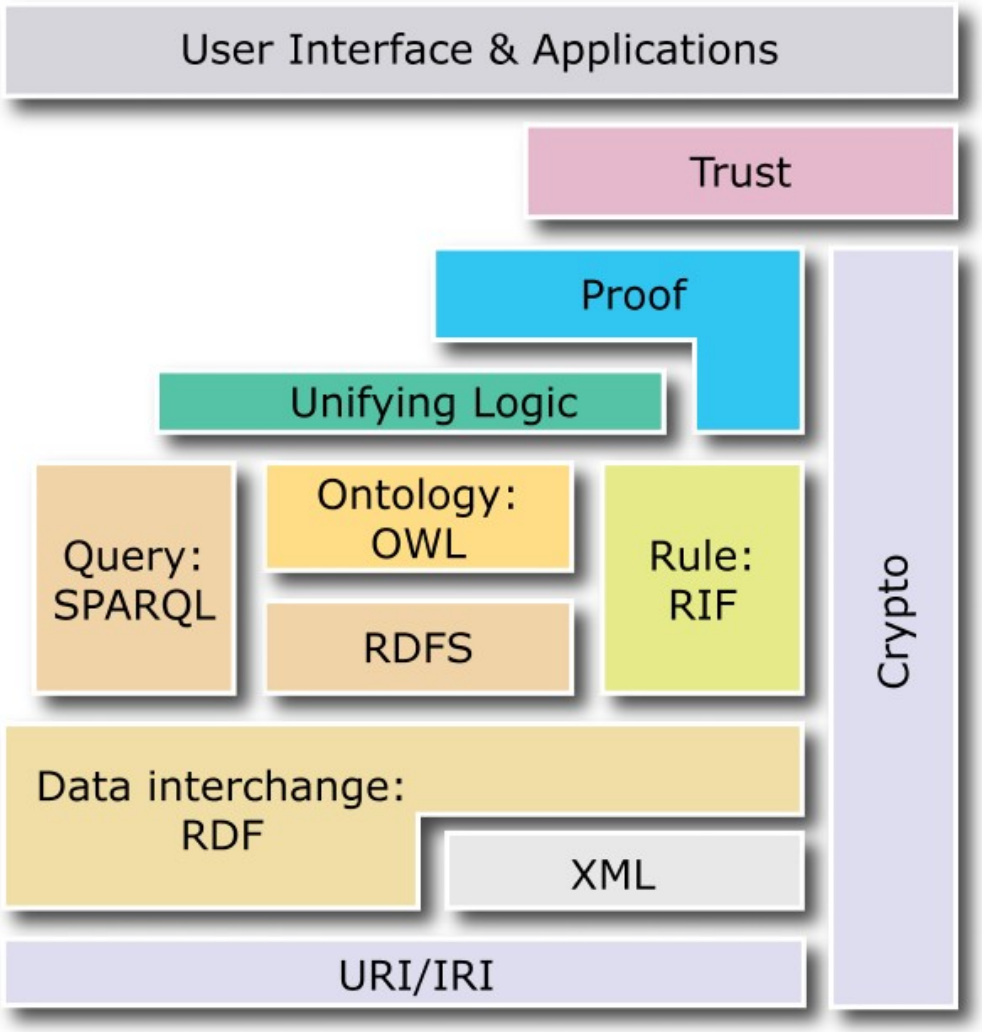
Key principles

- The Semantic Web is the Web
  - Same base technologies, Evolutionary
  - Decentralized (incomplete, inconsistent)
- Provide explicit statements regarding web resources
  - Authors, original information providers
  - Intermediaries (humans and/or machines)
- Information consumers determine consequences of the statements
  - Distributed ‘reasoning’

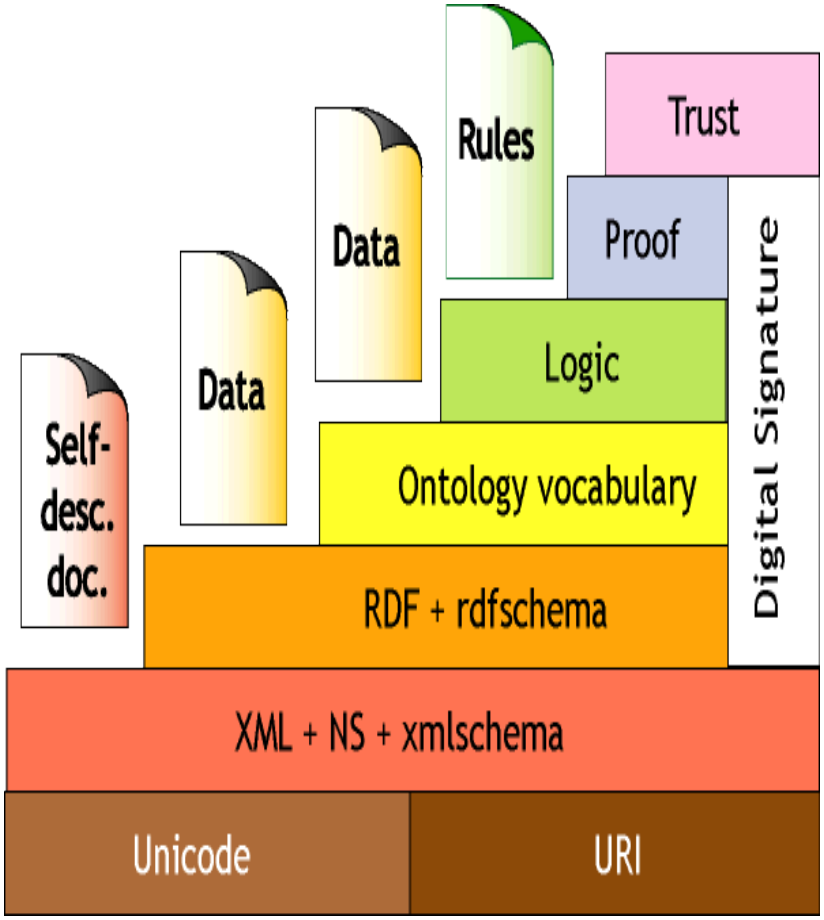
Technology stack (old: pre-2008)

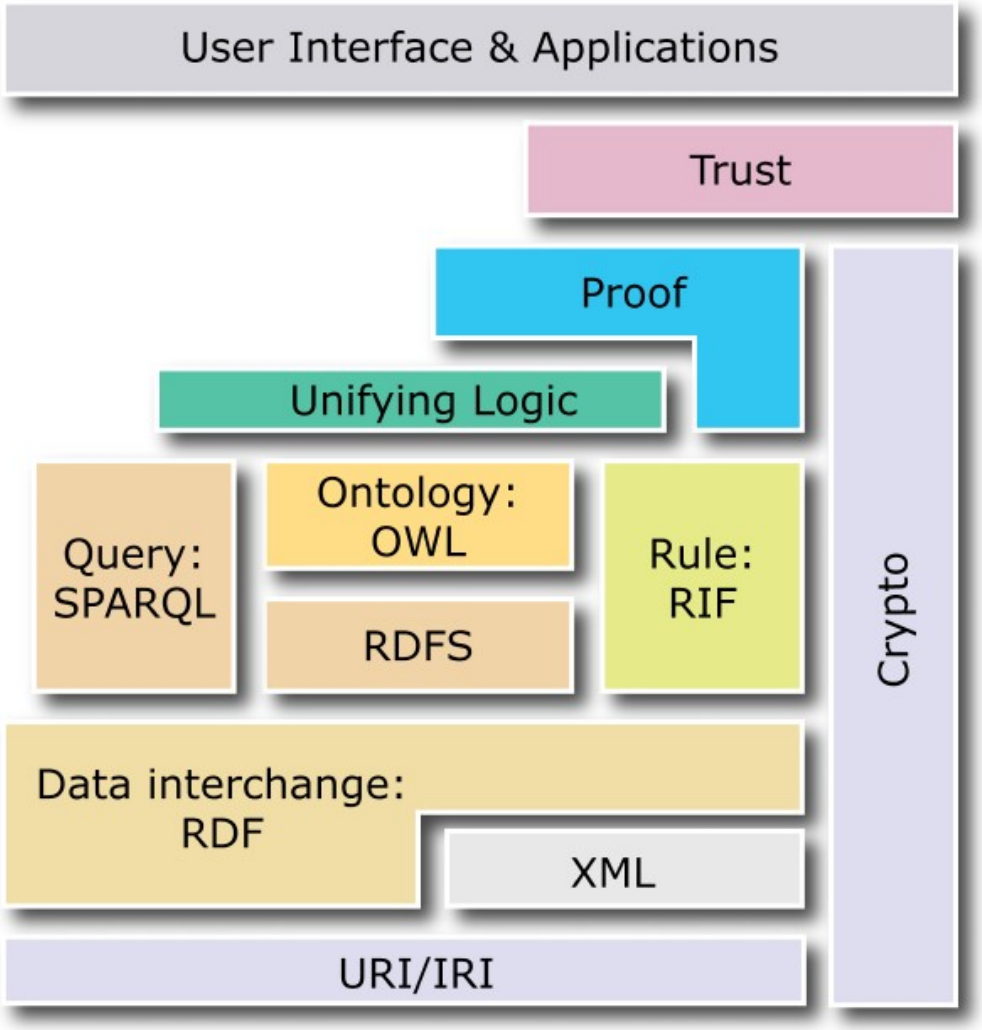


Technology stack (current: 2008)



Comparison... current trends



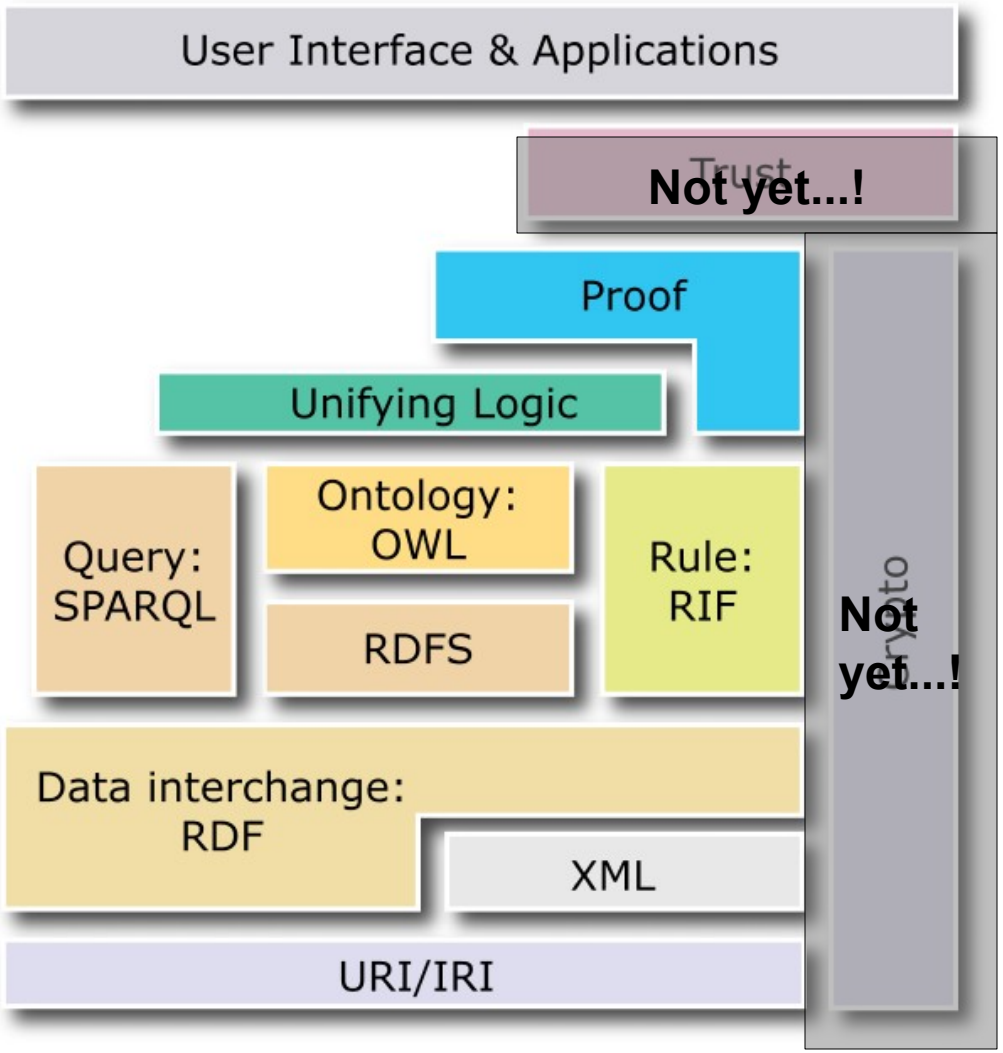


Main technologies

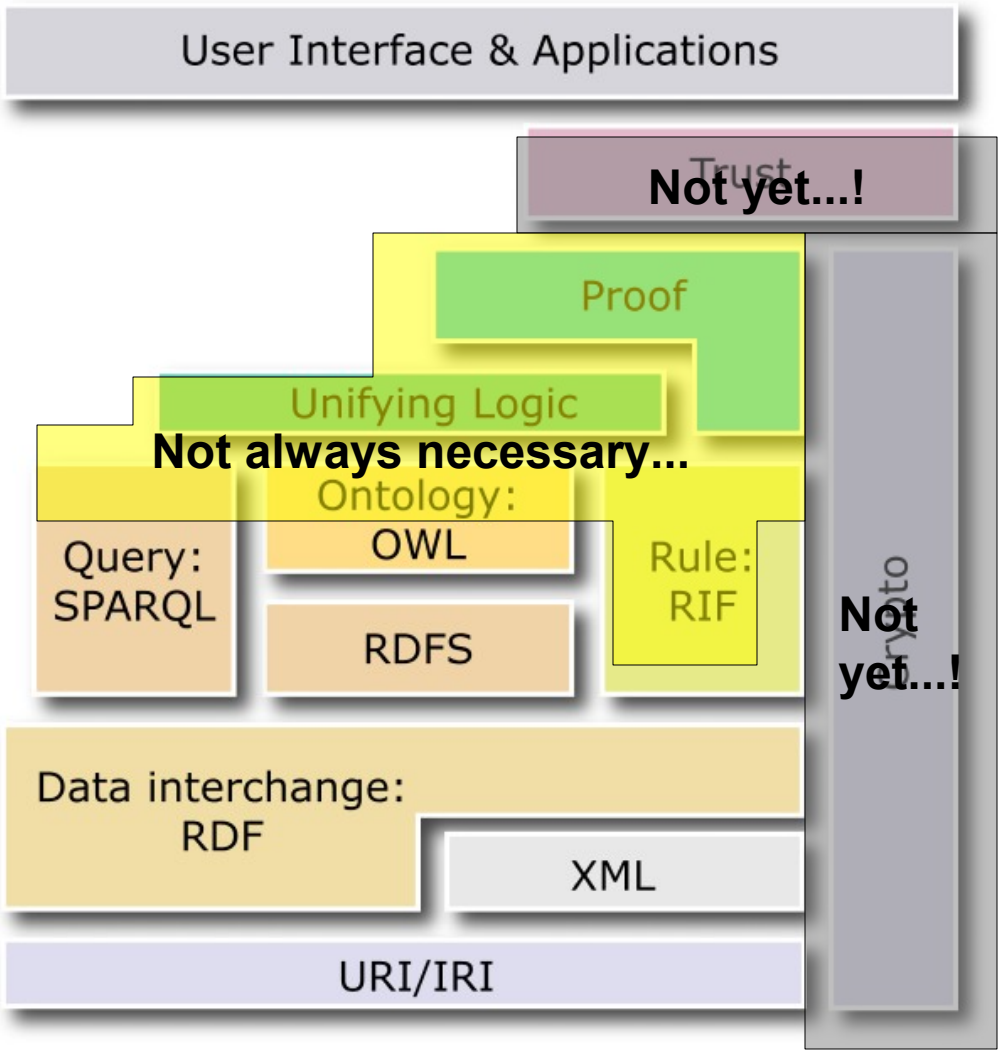
- Resource Description Framework (RDF) [2004]
- Gleaning Resource Descriptions from Dialects of Languages (GRDDL) [2007]
- SPARQL Query Language for RDF [2008]
- Web Ontology Language (OWL) [2004]

http://www.w3.org/2001/sw/

The real world...



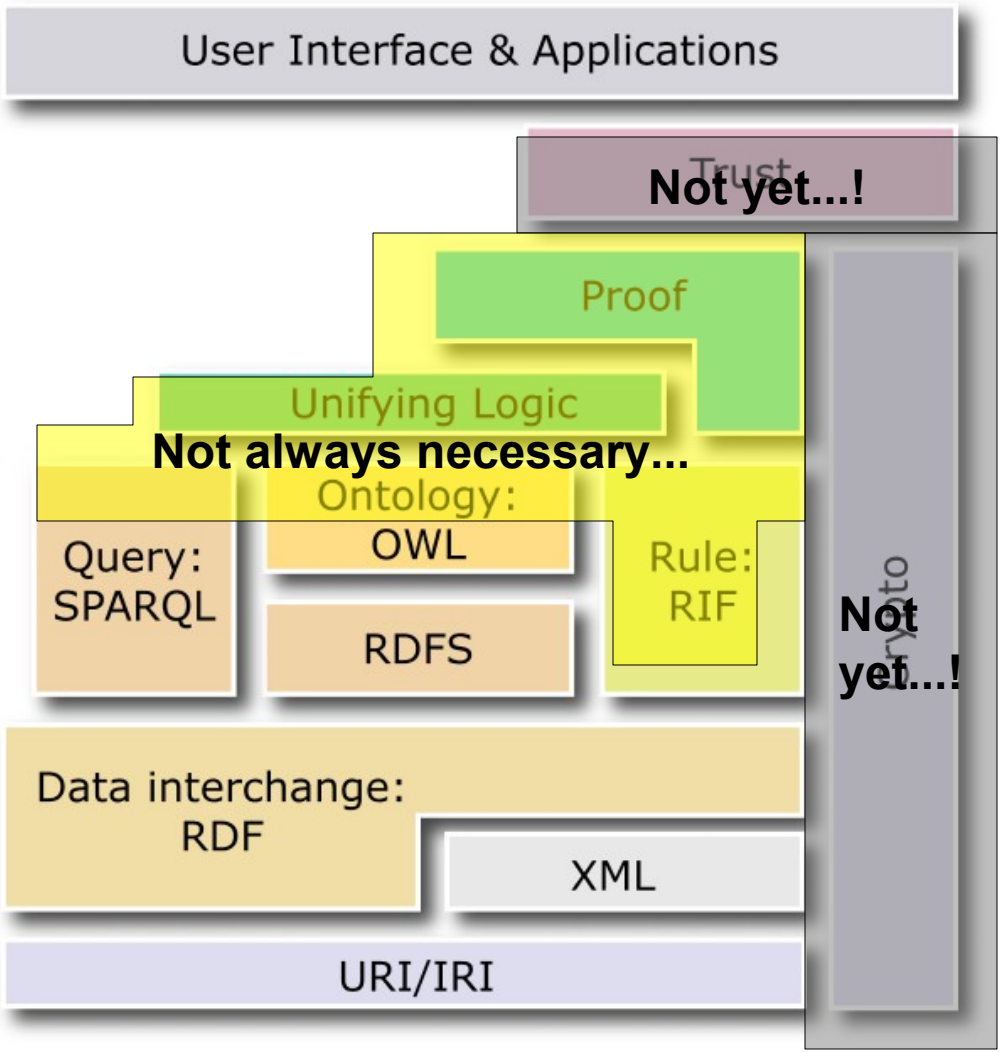
The real world...



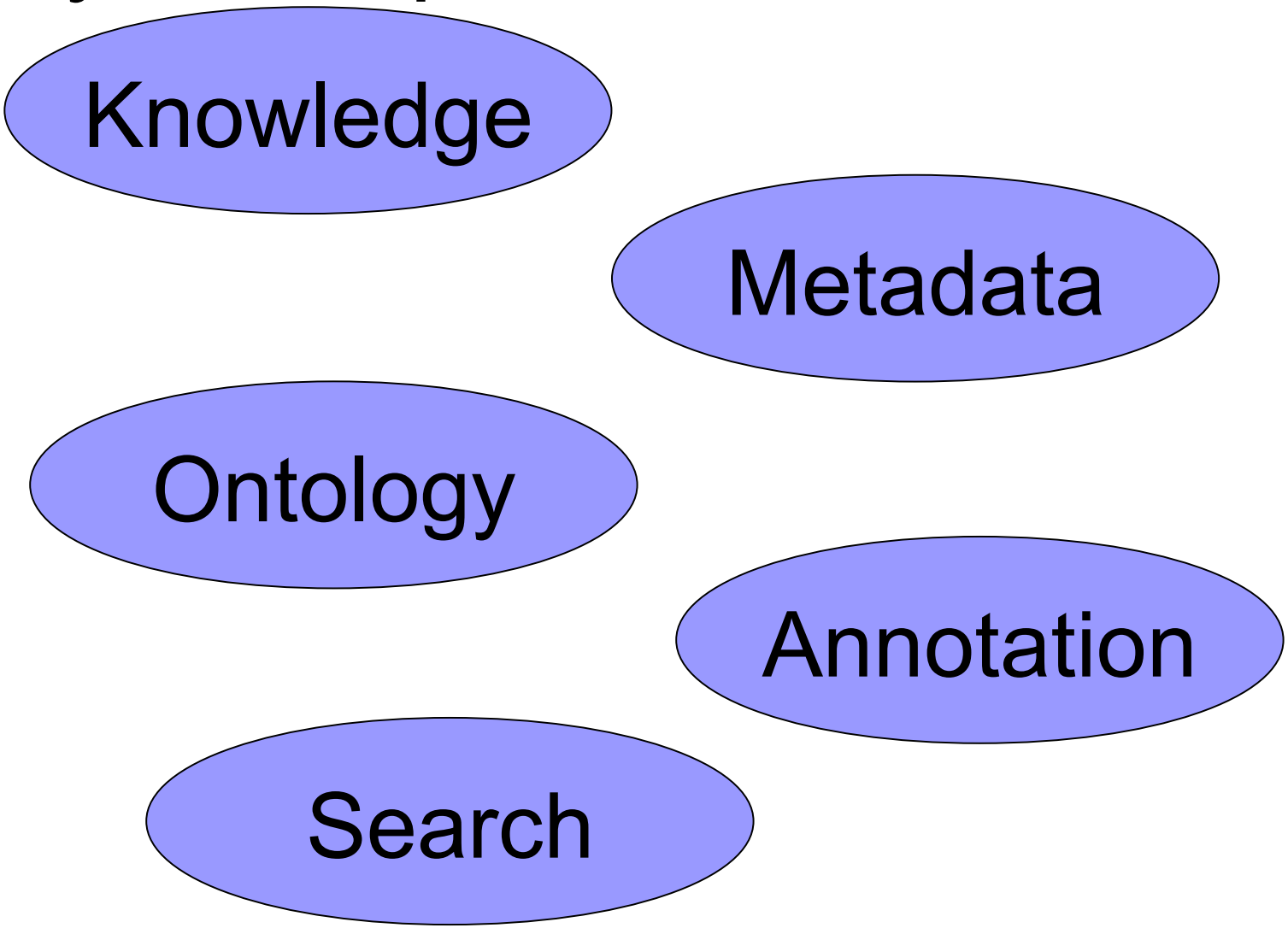
The real world...

Information retrieval

Statistics



Key concepts



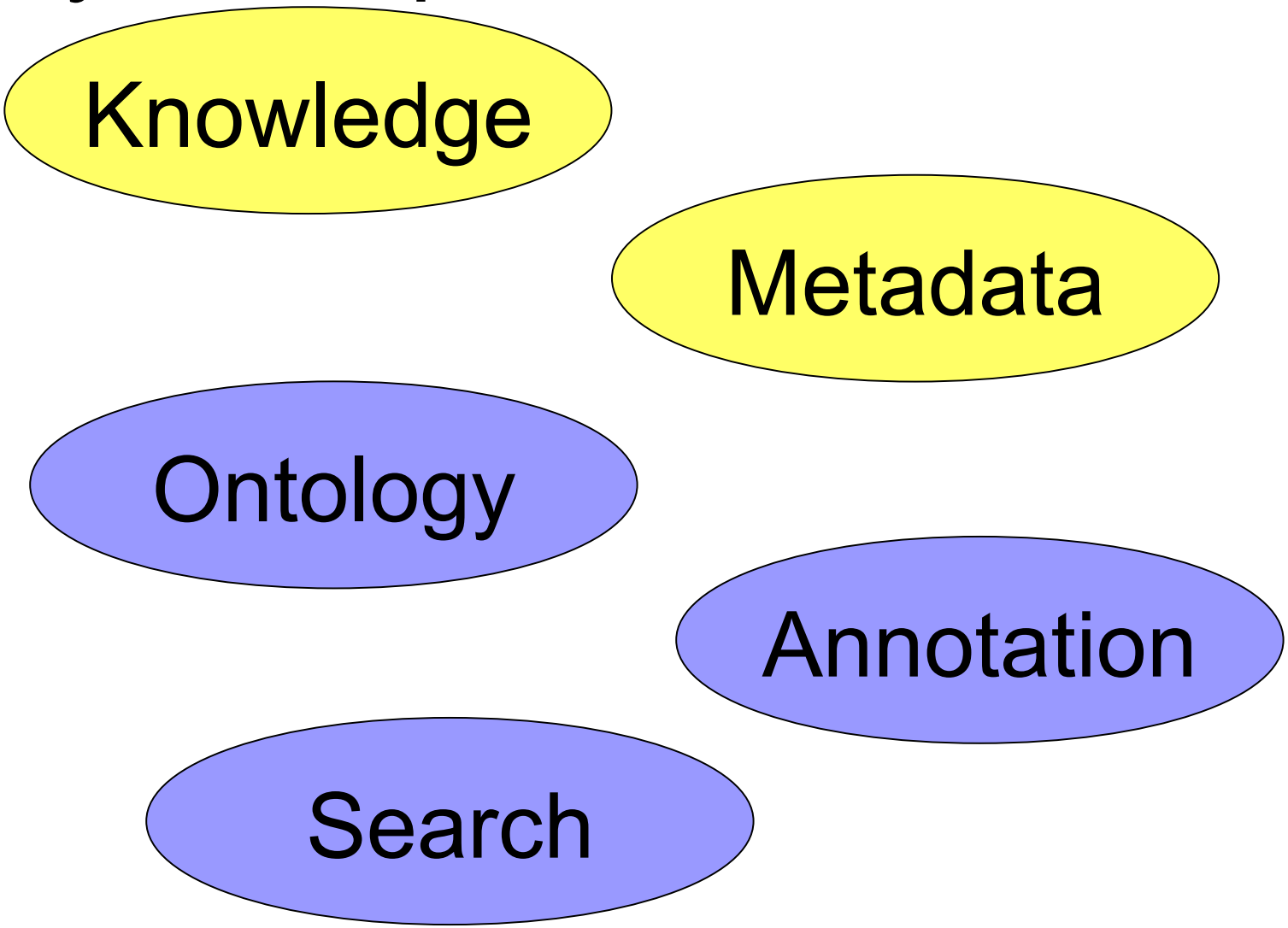
Metadata and Knowledge Representation

Semantic tecnologies – a quick overview

Fulvio Corno

Politecnico di Torino

Key concepts



Metadata

- Structured data about data
- Higher level information that describes
  - content
  - quality
  - structure
  - accessibility
  - …

  of a specific data set

Problems

- Even the author, sometimes, has difficulty in correctly classifying the contents
- Choice of “topics” may be done at different levels, all of them “correct”
- Choice of “keywords” is afflicted by synonyms and homonyms
- Typographical errors often appear in the original text

Classification techniques

- Controlled vocabularies
- Taxonomies
- Thesauri
- Facets
- Ontologies
- Folksonomies

Controlled vocabulary

- A **closed list of named subjects**, which can be used for classification
- Composed of **terms** [particular name for a particular concept] (similar to *keywords*)
- Terms are not concepts
  - A single term may be the name of one (or more) concept(s)
  - A single concept may have multiple names
  - Ambiguity avoided by forbidding duplicate terms

Controlled vocabulary

- Purpose:
  - to avoid authors defining **meaningless terms**, terms which are too **broad**, or terms which are too **narrow**
  - to prevent different authors from misspelling and choosing slightly different forms of the same term
- The simplest form of controlled vocabulary is simply a **list of terms** and nothing more.

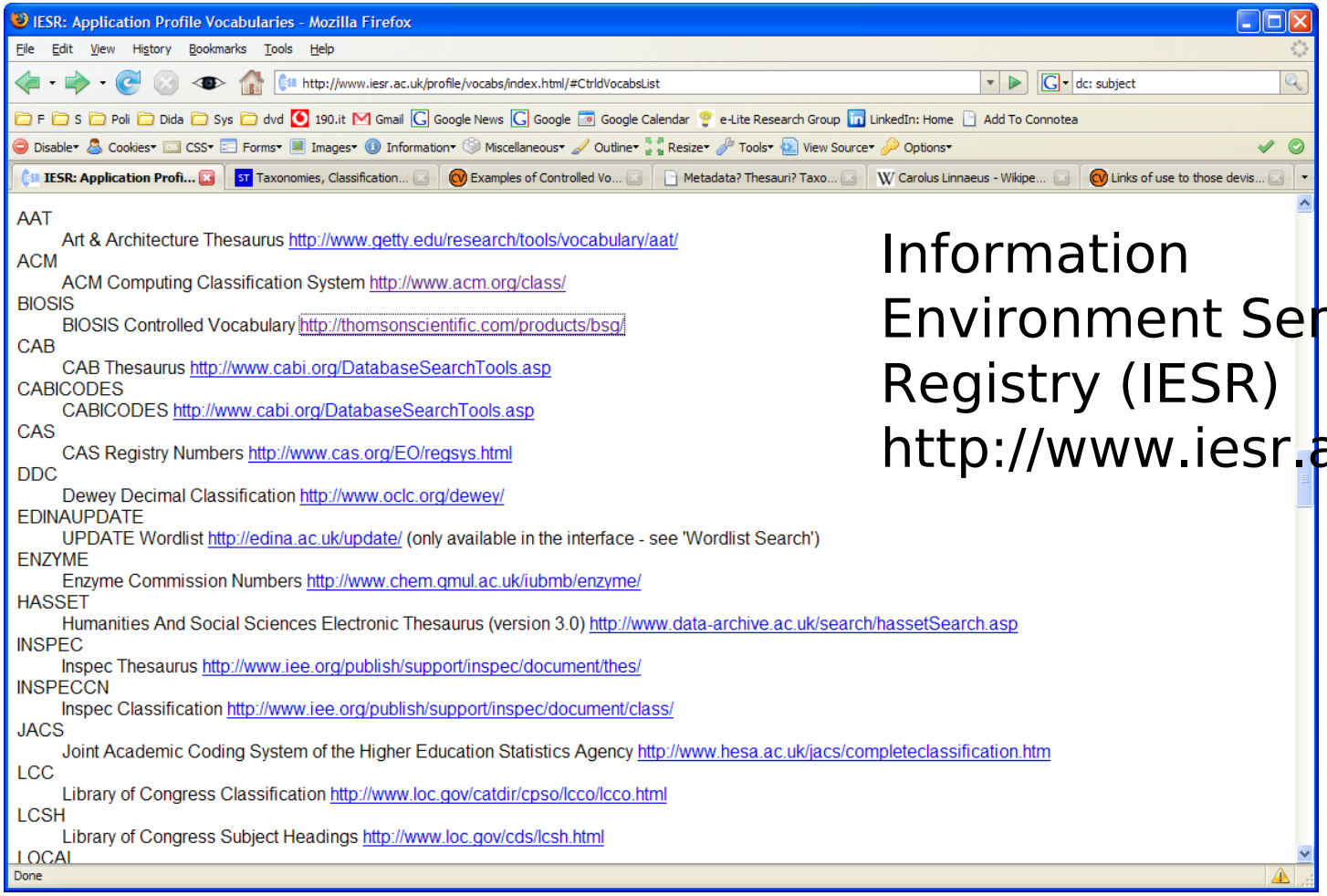
Problems solved

- Homograph
  - group of words that share the same spelling but have different meanings
- Homonym
  - group of words that share the same spelling or pronunciation (or both) but have different meanings

Problems solved

- Synonym
  - different words with identical or at least similar meanings
- Polysemy
  - the capacity for a word to have multiple meanings

Example



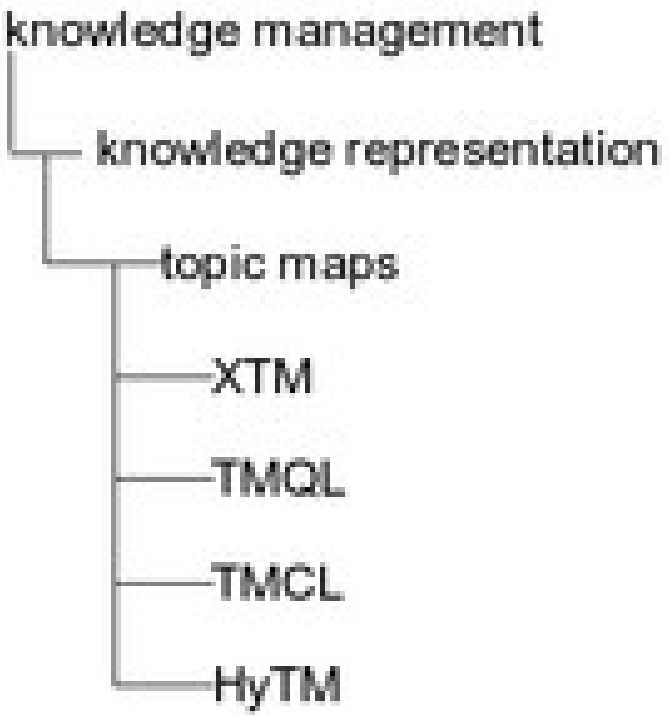
Information Environment Service Registry (IESR) http://www.iesr.ac.uk/

http://www.iesr.ac.uk/profile/vocabs/index.html/#CtrldVocabsList

Taxonomy

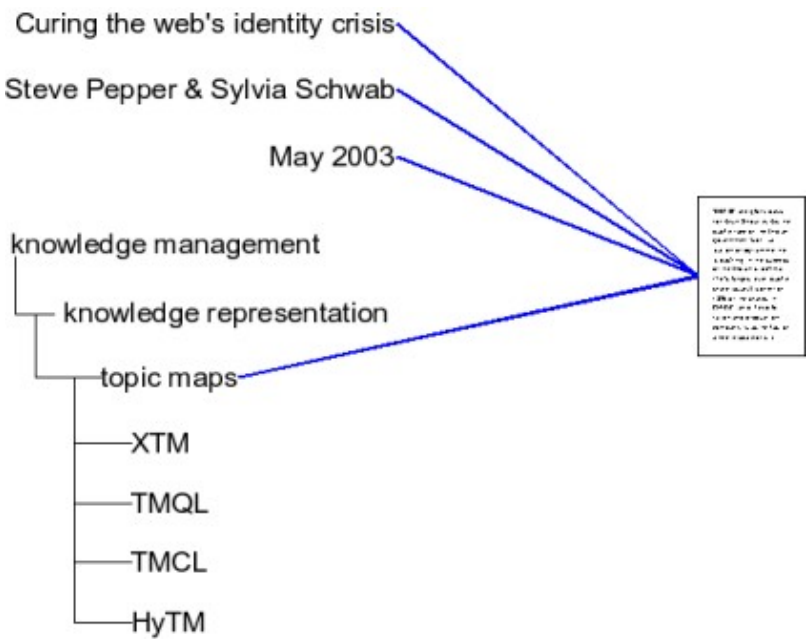
- Subject-based classification that arranges the terms in the controlled vocabulary into a **hierarchy**
  - Dating back to the work of Carl Linnæus in the 18th century on zoological and botanical classification and naming system for species
- Benefit: they allow related terms to be **grouped together and categorized** in ways that make it easier to find the correct term to use for searching or to describe an object

Taxonomy



- It is clear that "topic maps" and "XTM" are related
- Easier to classify documents
- Easier to choose search keywords

Taxonomies and metadata

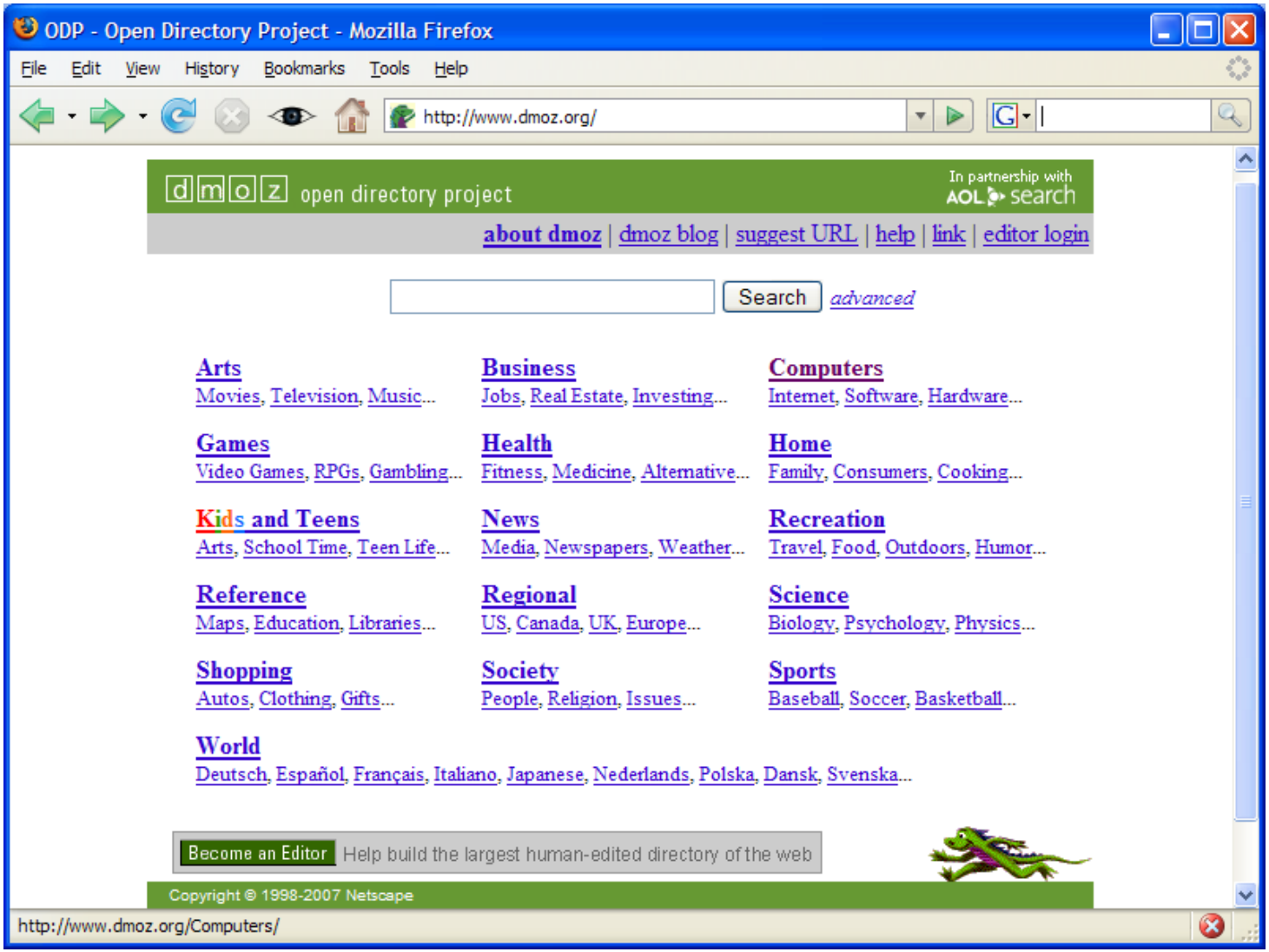


- Metadata are stored as usual with the resource
- The “subject” will contain only controlled terms
- Controlled terms belong to a hierarchy, shared by all papers

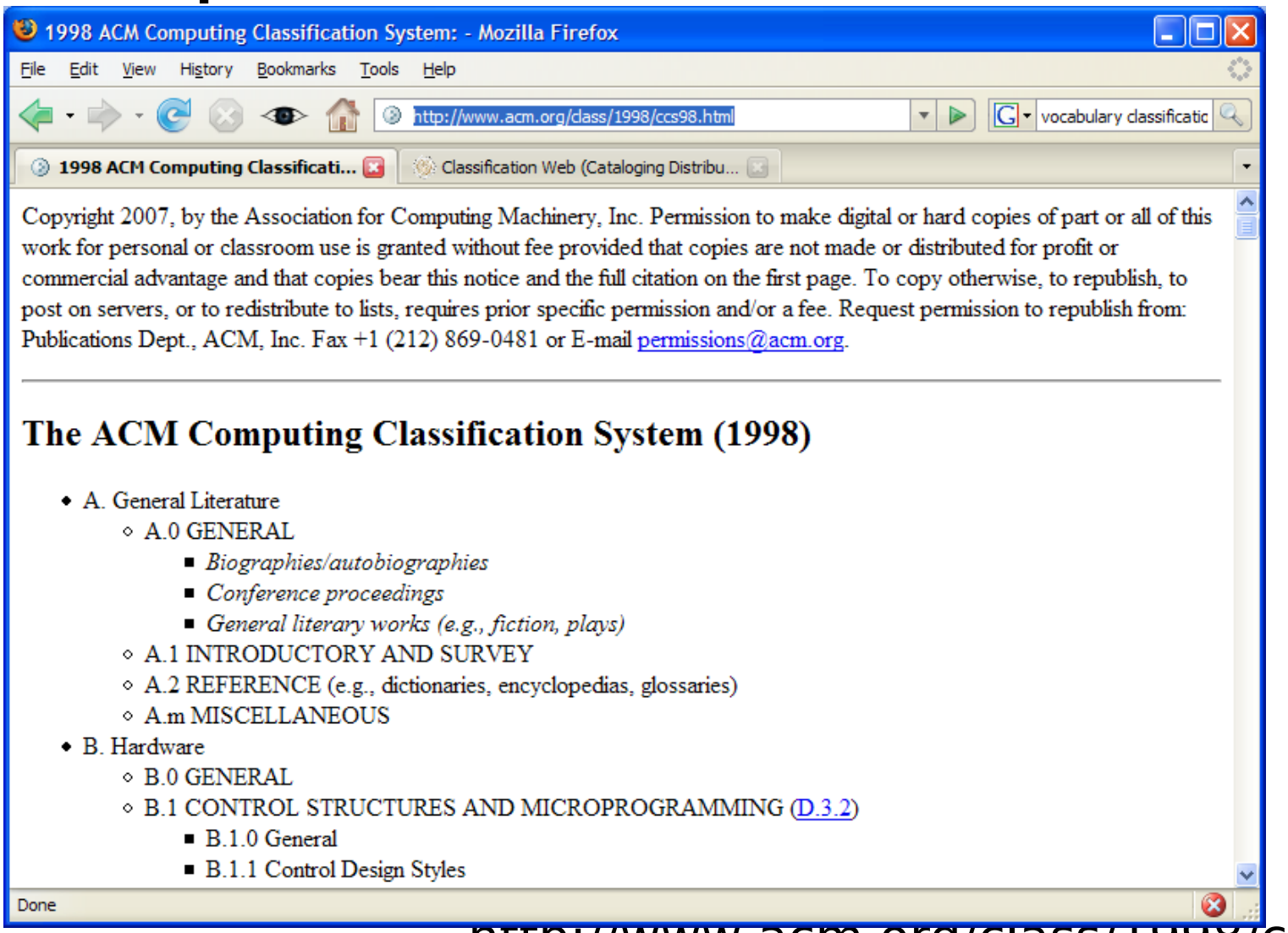
Limitations

- A taxonomy may **not** express:
  - "XML Topic Maps" is synonymous with "XTM"
  - difference between "XTM" and "topic maps".
  - "topic navigation maps" is synonymous with "topic maps", but should no longer be used\
  - relationship between topic maps and subject-based classification, or the semantic web.
  - relationship between XTM and XML and HyTM and SGML.
  - similarity between HyTM and XTM, and their difference from TMQ

Example


ODP - Open Directory Project - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.dmoz.org/
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | suggest URL | help | link | editor login
Search advanced
Arts
Movies, Television, Music...
Business
Jobs, Real Estate, Investing...
Computers
Internet, Software, Hardware...
Games
Video Games, RPGs, Gambling...
Health
Fitness, Medicine, Alternative...
Home
Family, Consumers, Cooking...
Kids and Teens
Arts, School Time, Teen Life...
News
Media, Newspapers, Weather...
Recreation
Travel, Food, Outdoors, Humor...
Reference
Maps, Education, Libraries...
Regional
US, Canada, UK, Europe...
Science
Biology, Psychology, Physics...
Shopping
Autos, Clothing, Gifts...
Society
People, Religion, Issues...
Sports
Baseball, Soccer, Basketball...
World
Deutsch, Español, Français, Italiano, Japanese, Nederlands, Polska, Dansk, Svenska...
Become an Editor Help build the largest human-edited directory of the web
Copyright © 1998-2007 Netscape
http://www.dmoz.org/Computers/

Example



http://www.acm.org/class/1998/ccs98.html

Thesaurus

- **Extends taxonomies**: subjects are arranged in a hierarchy
- **Other statements** can be made about the subjects
- Two ISO standards
  - ISO2788 for **monolingual** thesauri
  - ISO5964 for **multilingual** thesauri

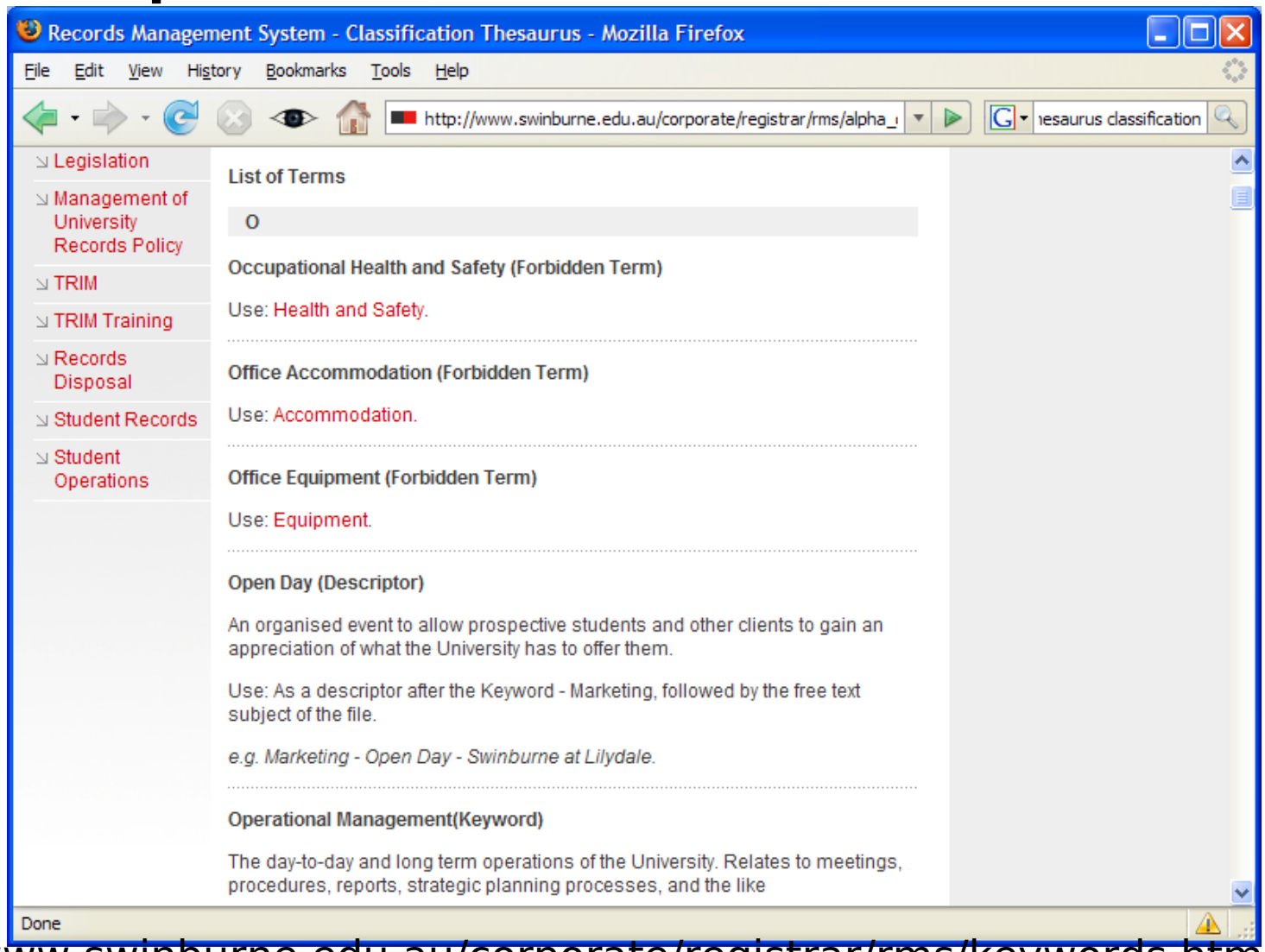
Thesaurus relationships

- BT – broader term
  - Refers to a term with wider or less specific meaning
  - Some systems allow multiple BTs for one term, while others do not
  - Inverse property: NT - narrower term
  - A taxonomy only uses BT and NT
- SN – scope note
  - String explaining its meaning within the thesaurus
  - Useful when the precise meaning of the term is not obvious from context.

Thesaurus relationships

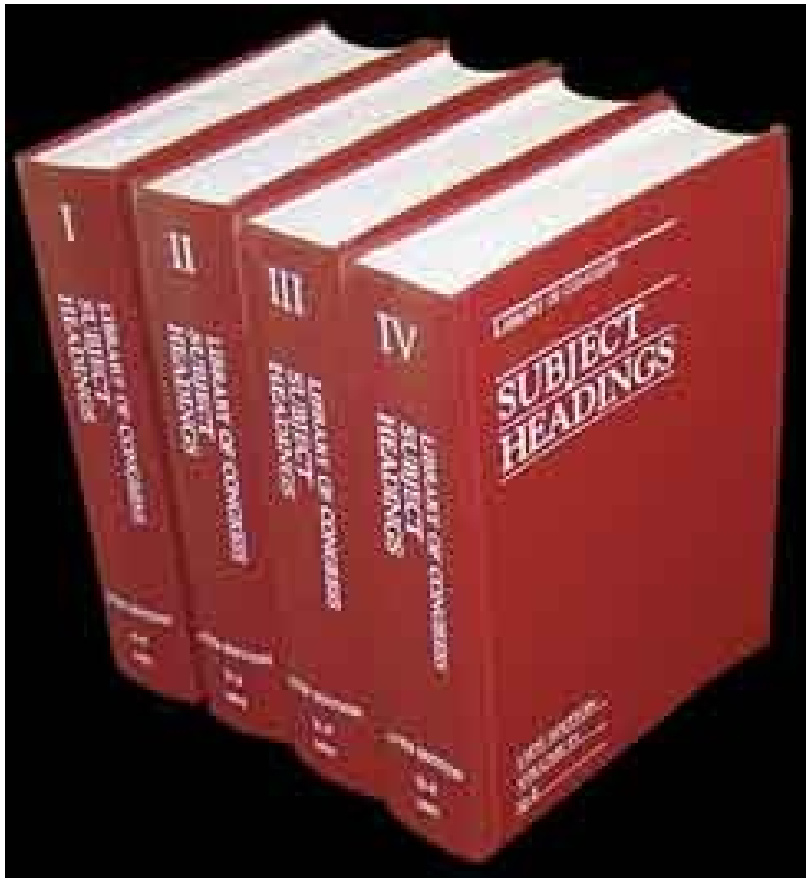
- USE
  - Another term that is to be preferred instead of this term
  - Implies that the terms are synonymous
  - Inverse property: UF
- TT – top term
  - The topmost ancestor of this term
  - The BT of the BT of the BT...
- RT – related term
  - A term that is related to this term, without being a synonym of it or a broader/narrower term.

Example



http://www.swinburne.edu.au/corporate/registrar/rms/keywords.htm

Example

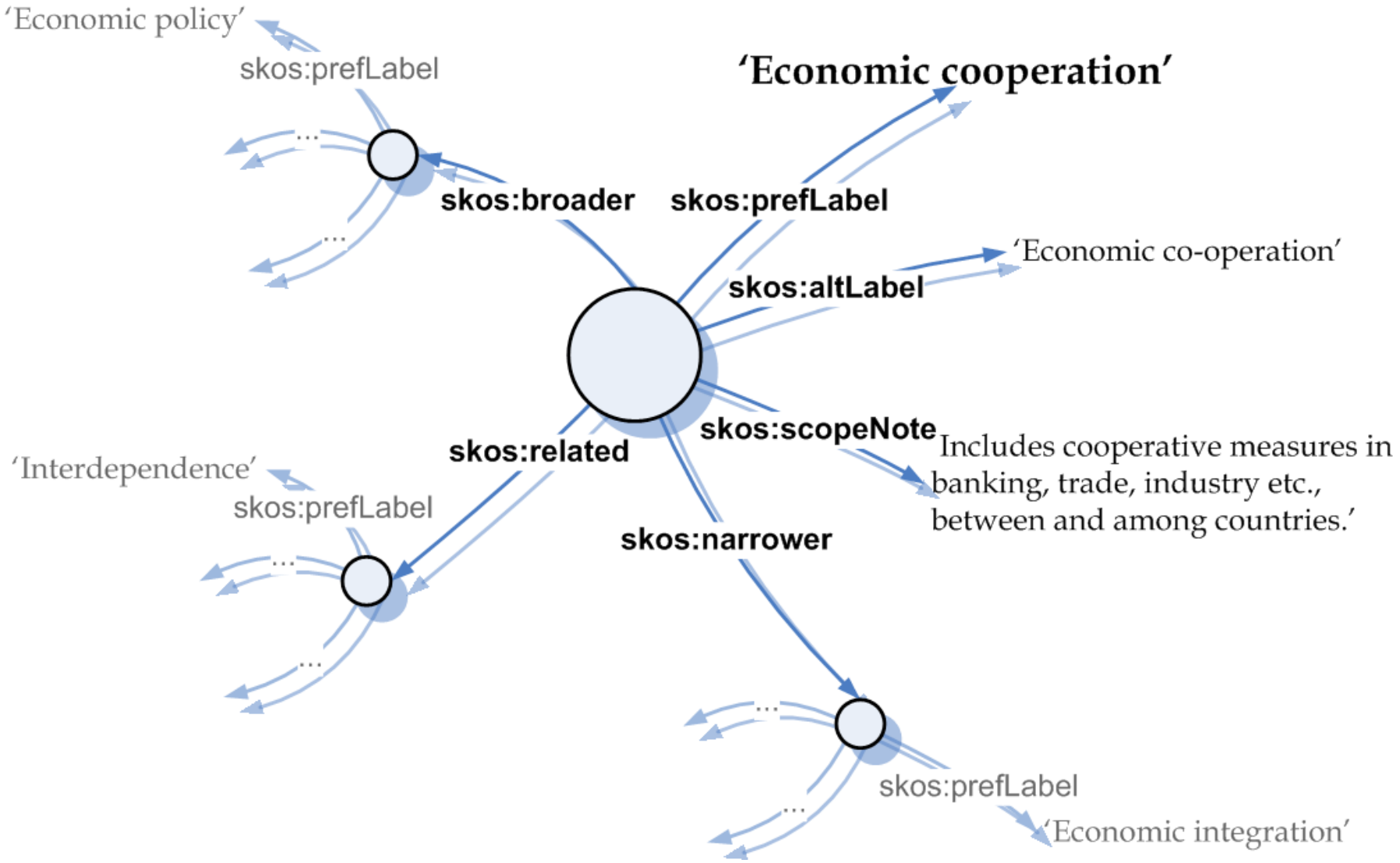


- Library of Congress Subject Headings
- Since 1985 it became a taxonomy
- http://www.loc.gov/lexico/servlet/lexico/

SKOS

- SKOS: Simple Knowledge Organization System
  - http://www.w3.org/2004/02/skos/
  - «SKOS provides a standard way to represent knowledge organisation systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.»

SKOS: example



Facets

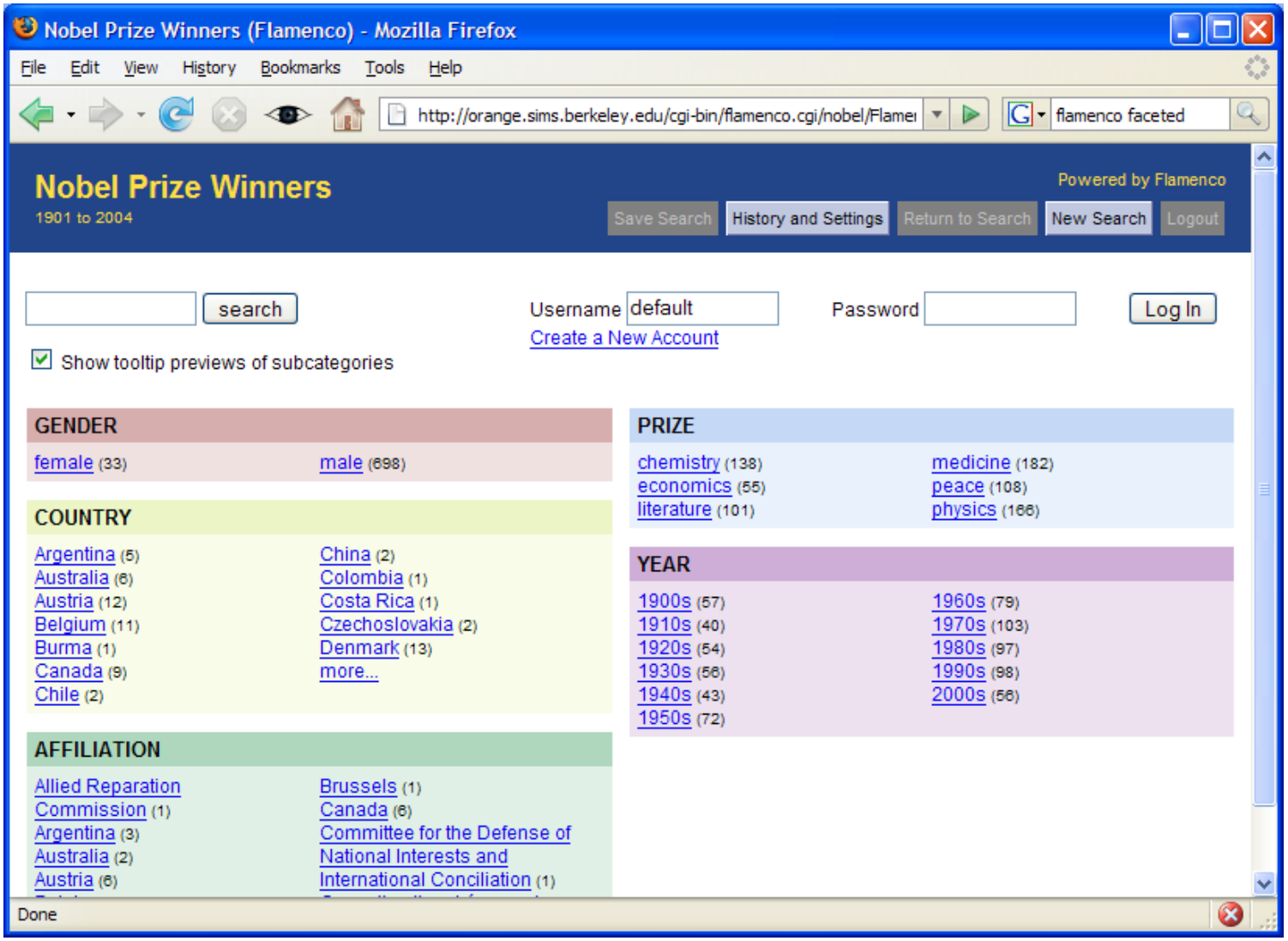
- Proposed by S.R. Ranganathan in the 1930s
- Identifies a number of **facets** into which the terms are divided.
  - Facets can be thought of as different axes along which documents can be classified
  - Each facet contains a number of terms
    - Usually with at thesaurus-like organization
    - A term is only allowed to belong to a single facet
- A document is classified by picking one term from each facet to describe the document along all the different axes

Advantages

- Multi-dimensionality
- Persistence
- Scalability
- Flexibility

Example

http://flamenco.berkeley.edu/



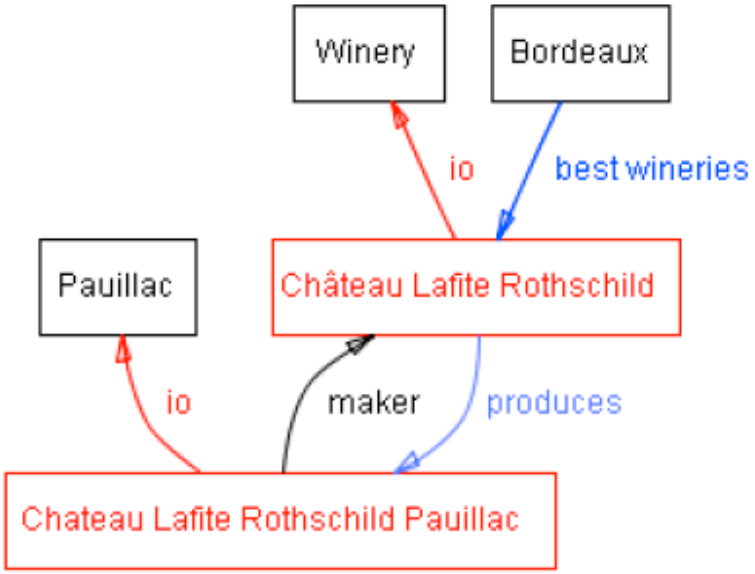
Example

http://intellisemantic.myvnc.com:8080/IntelliFacet/

Ontology

- Model for describing the world that consists of a set of **types**, **properties**, and **relationships**
- Extends the other subject-based classification approaches
  - Has **open vocabularies**
  - Has open **relationship types**
    - Not just BT/NT + USE/UF

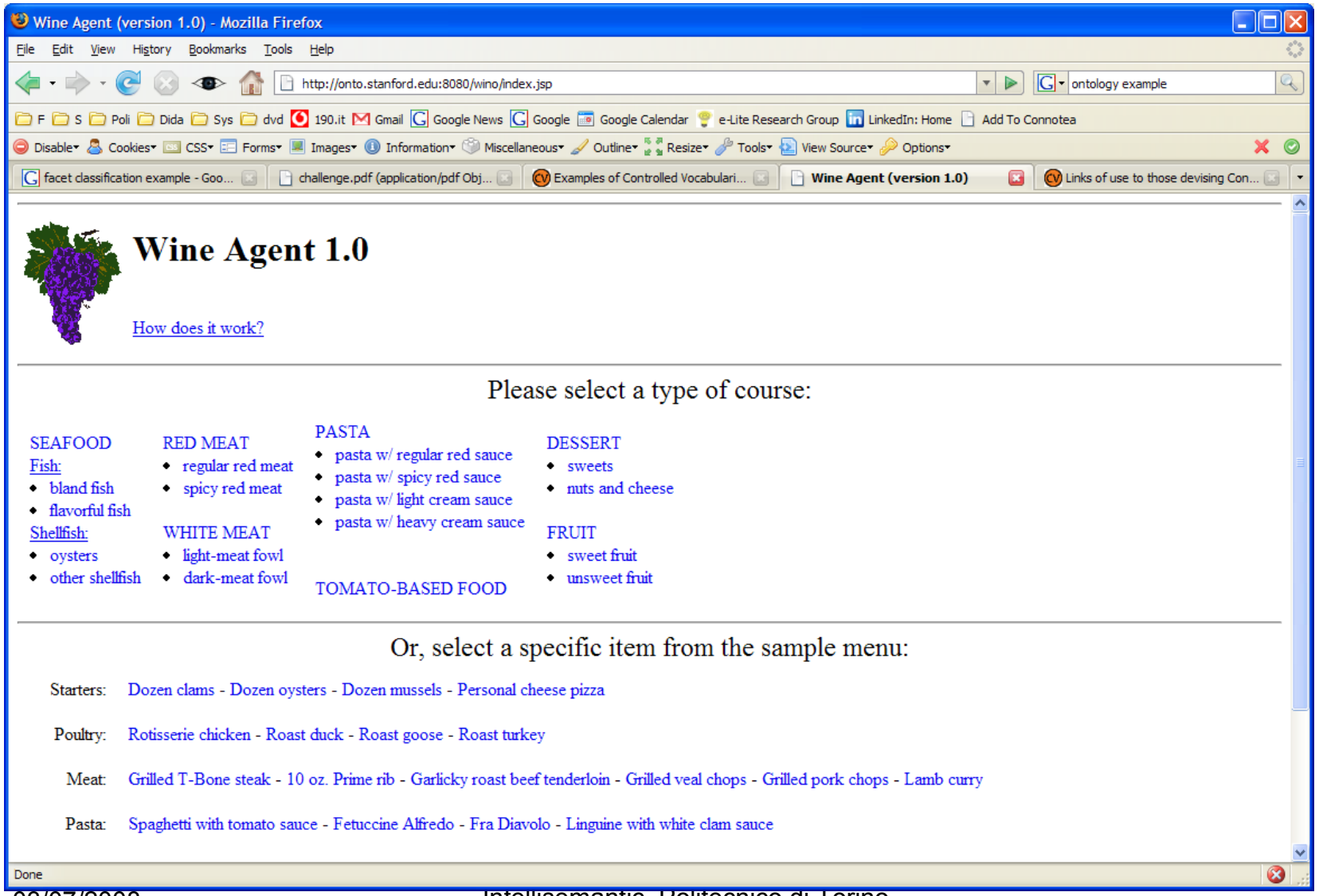
Ontology structure



- Concepts
- Relationships
  - Is-a
  - Other
- Instances

Example

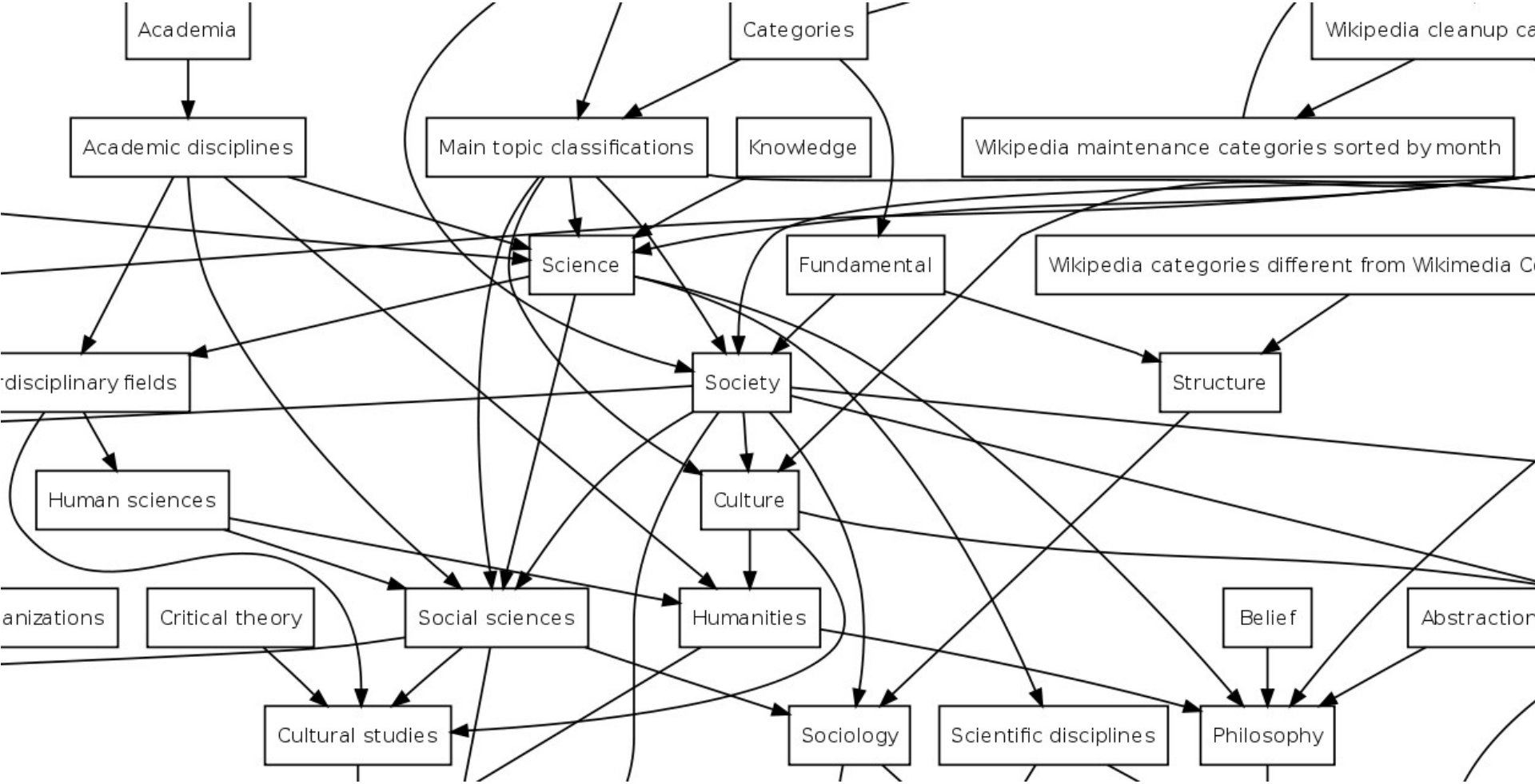
http://onto.stanford.edu:8080/wino/index.jsp



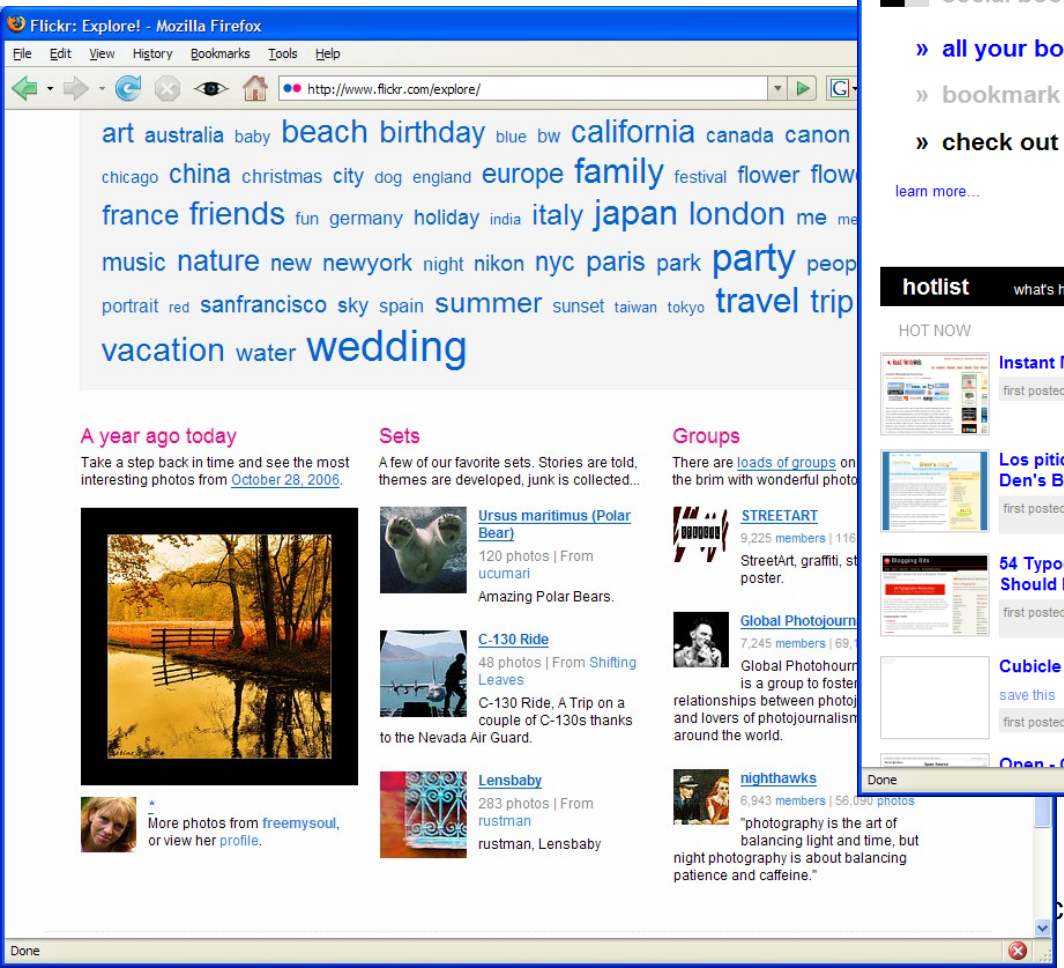
Folksonomy

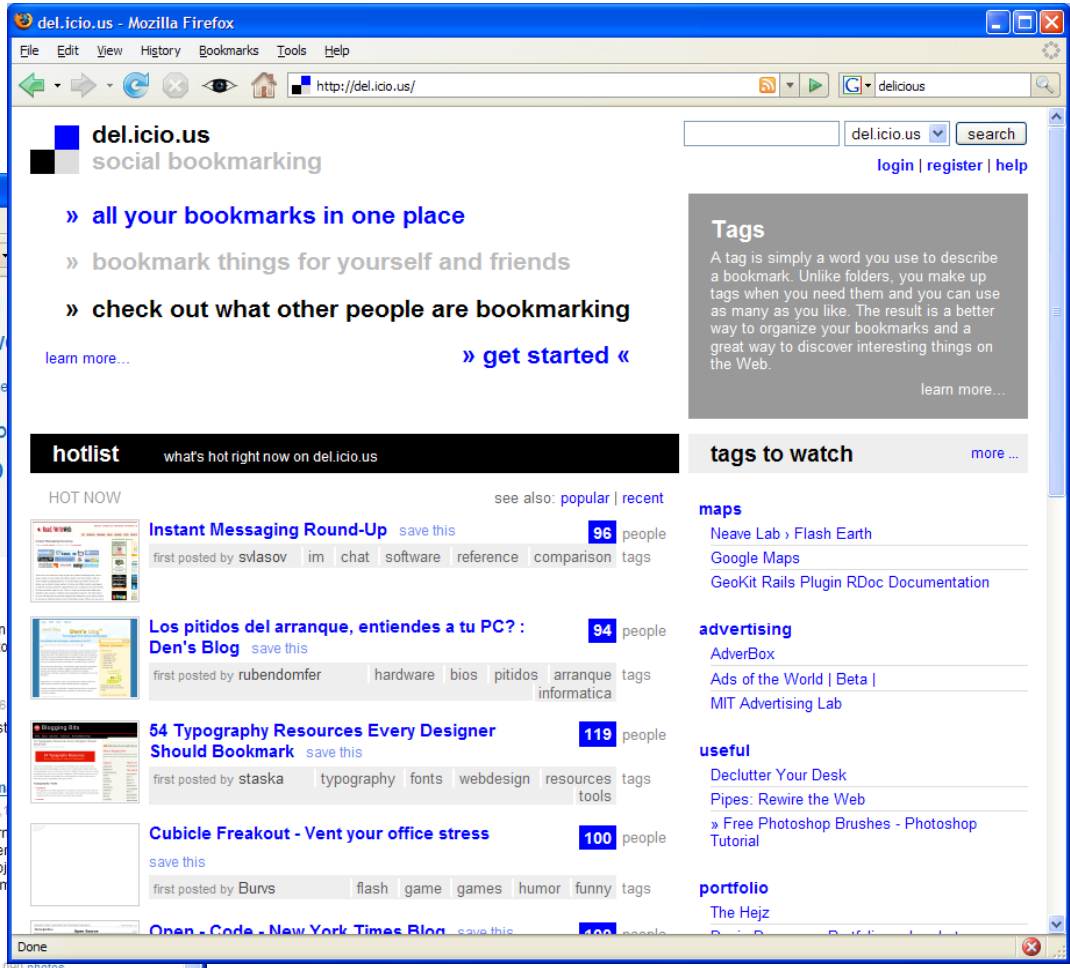
- Internet-mediated social environments
- Collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as Web pages, online photographs, and Web links
- Users can discover who created a given folksonomy tag, and see the other tags that this person created
  - reward: better user's capacity to find related content

Example (wikipedia)



Example (flickr - del.icio.us)





References

- http://www.controlledvocabulary.com/links.html
- http://en.wikipedia.org/wiki/Controlled_vocabulary
- http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html
- http://www.w3.org/2004/02/skos/
- http://en.wikipedia.org/wiki/SKOS

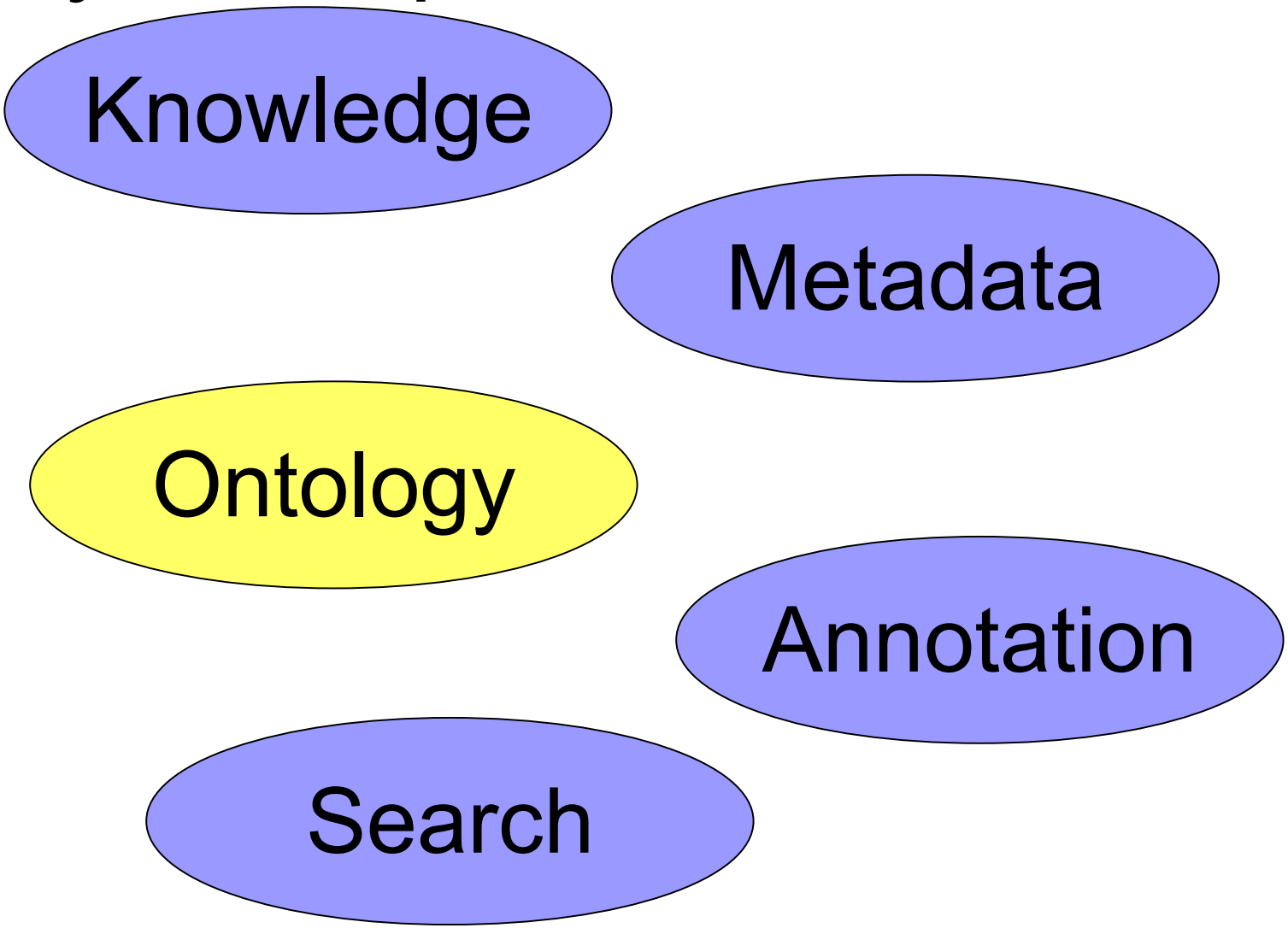
Ontologies

Semantic tecnologies – a quick overview

Fulvio Corno

Politecnico di Torino

Key concepts



What are the problems?

- People, Organizations, Software Systems must communicate
  - Among themselves
  - To the others
- Obstacles
  - Different needs
  - Different background
  - Different viewpoints
  - Different assumptions

How to solve them?

- resolve terminological confusion
- resolve conceptual confusion
- come to a **shared understanding**
- A shared understanding serves for:
  - providing a unifying framework for different viewpoints
  - providing the basis for communication between different people from different contexts

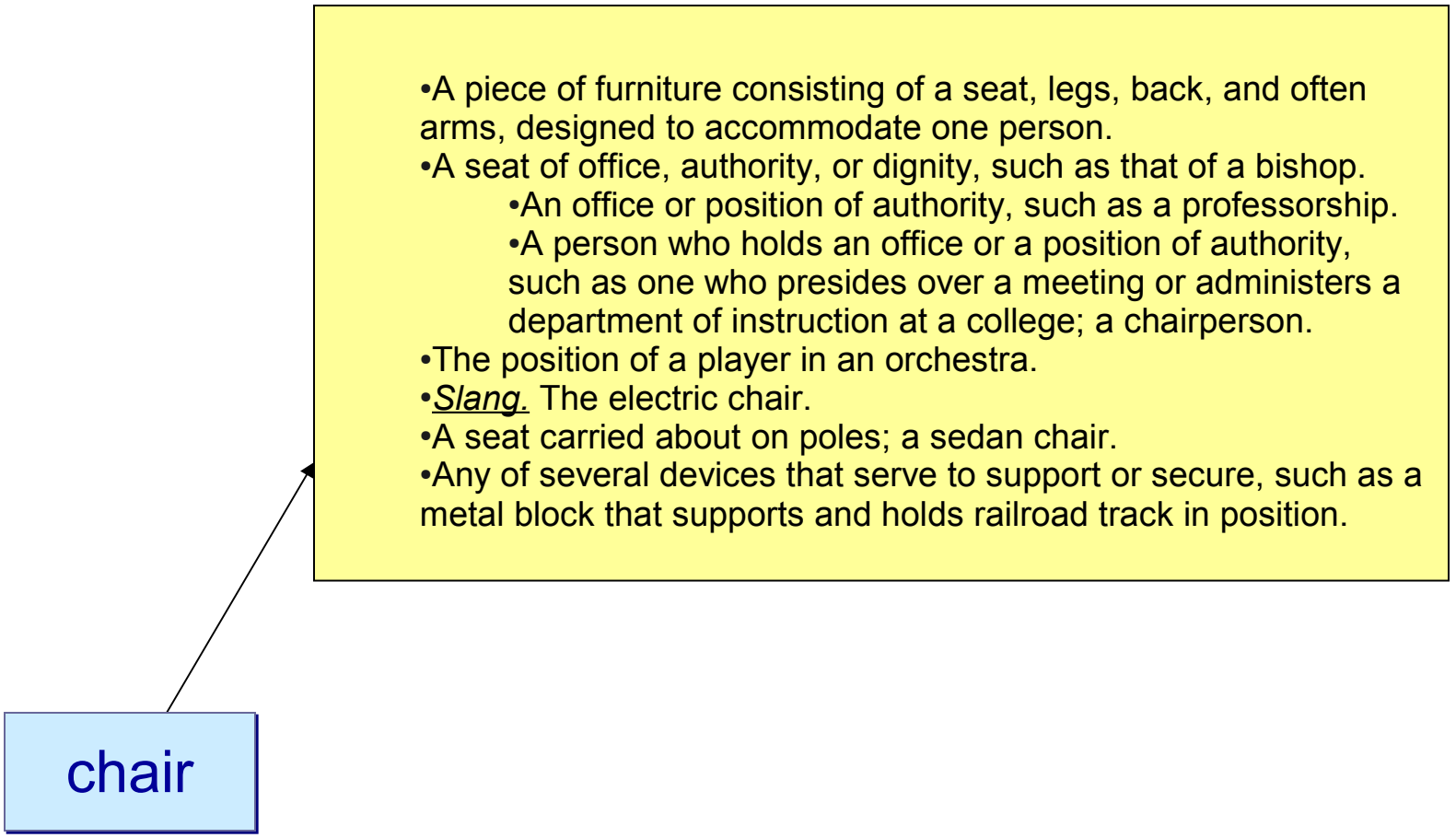
Ontologies: summary

- An ontology is an explicit description of a domain
  - concepts
  - properties and attributes of concepts
  - constraints on properties and attributes
  - individuals (often, but not always)
- An ontology defines
  - a common vocabulary
  - a shared understanding

Example

chair

Example


•A piece of furniture consisting of a seat, legs, back, and often arms, designed to accommodate one person.
•A seat of office, authority, or dignity, such as that of a bishop.
•An office or position of authority, such as a professorship.
•A person who holds an office or a position of authority, such as one who presides over a meeting or administers a department of instruction at a college; a chairperson.
•The position of a player in an orchestra.
•Slang. The electric chair.
•A seat carried about on poles; a sedan chair.
•Any of several devices that serve to support or secure, such as a metal block that supports and holds railroad track in position.
chair

Example

A piece of furniture consisting of a seat, legs, back, and often arms, designed to accommodate one person.

chair

Example

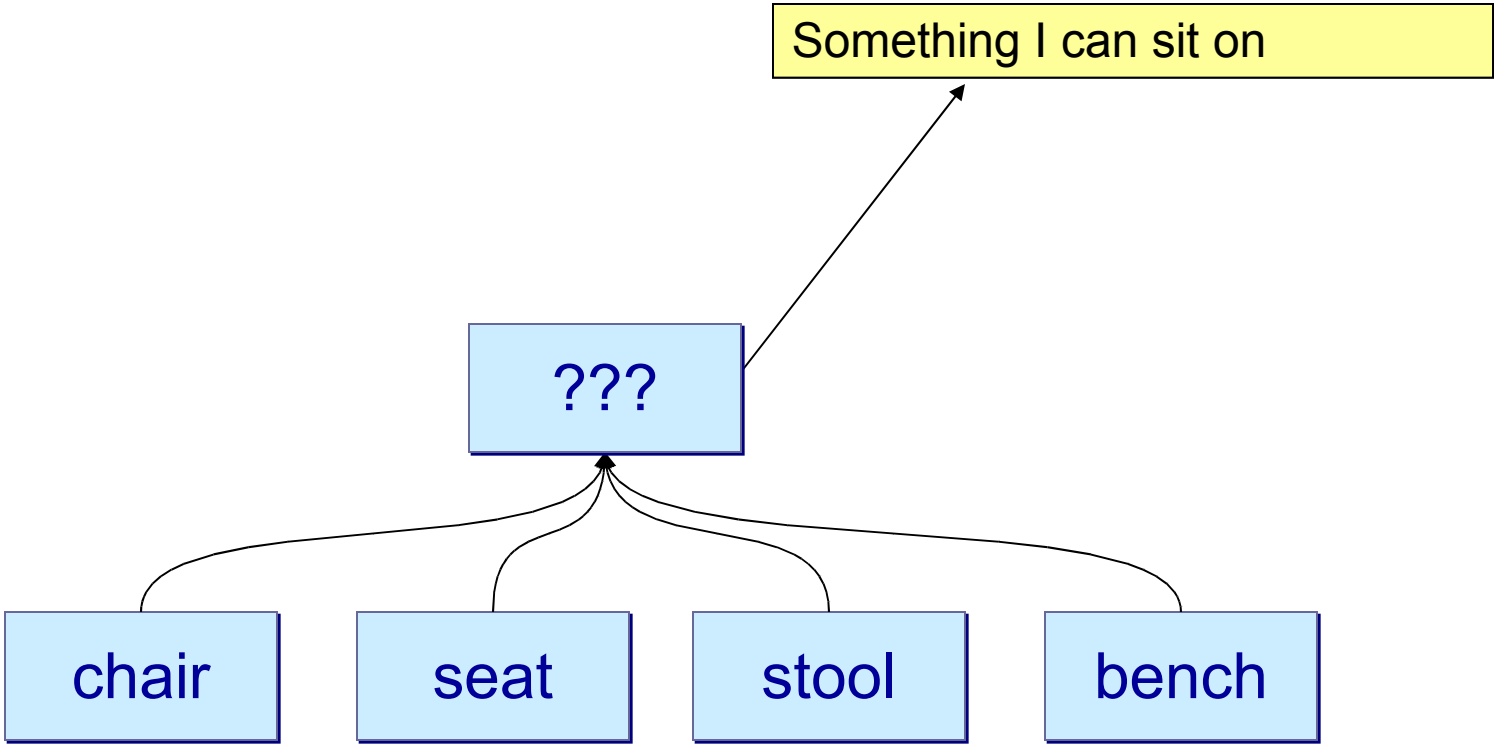
chair

seat

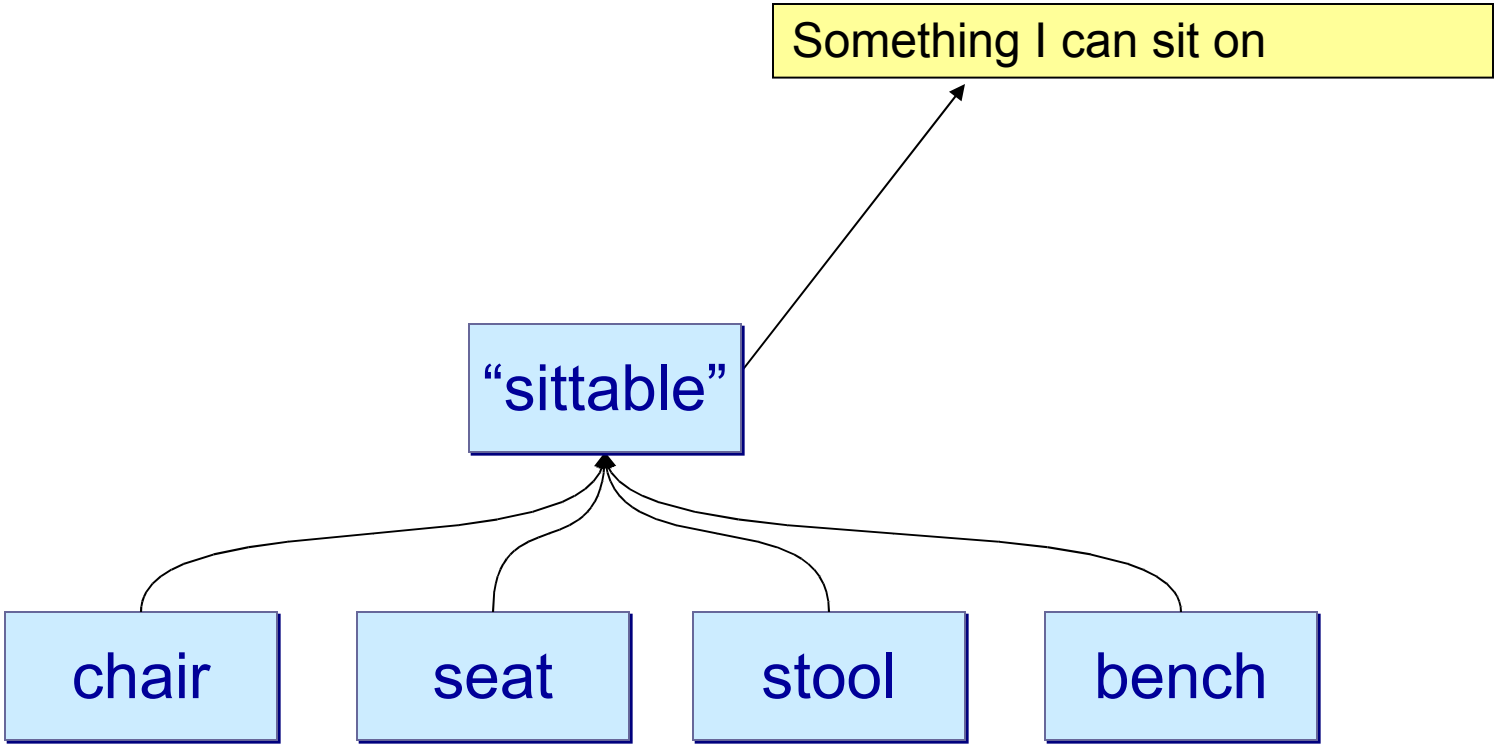
stool

bench

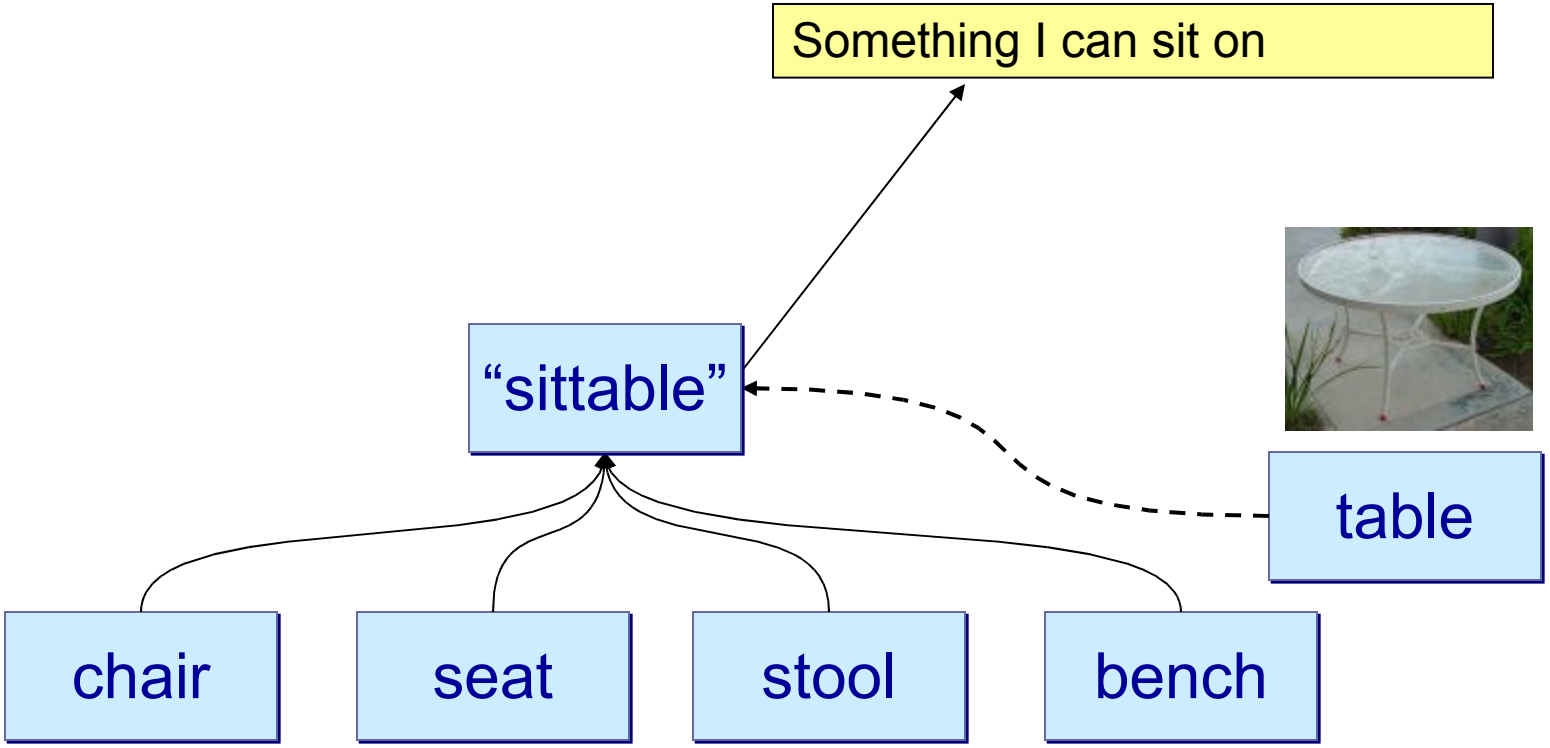
Example



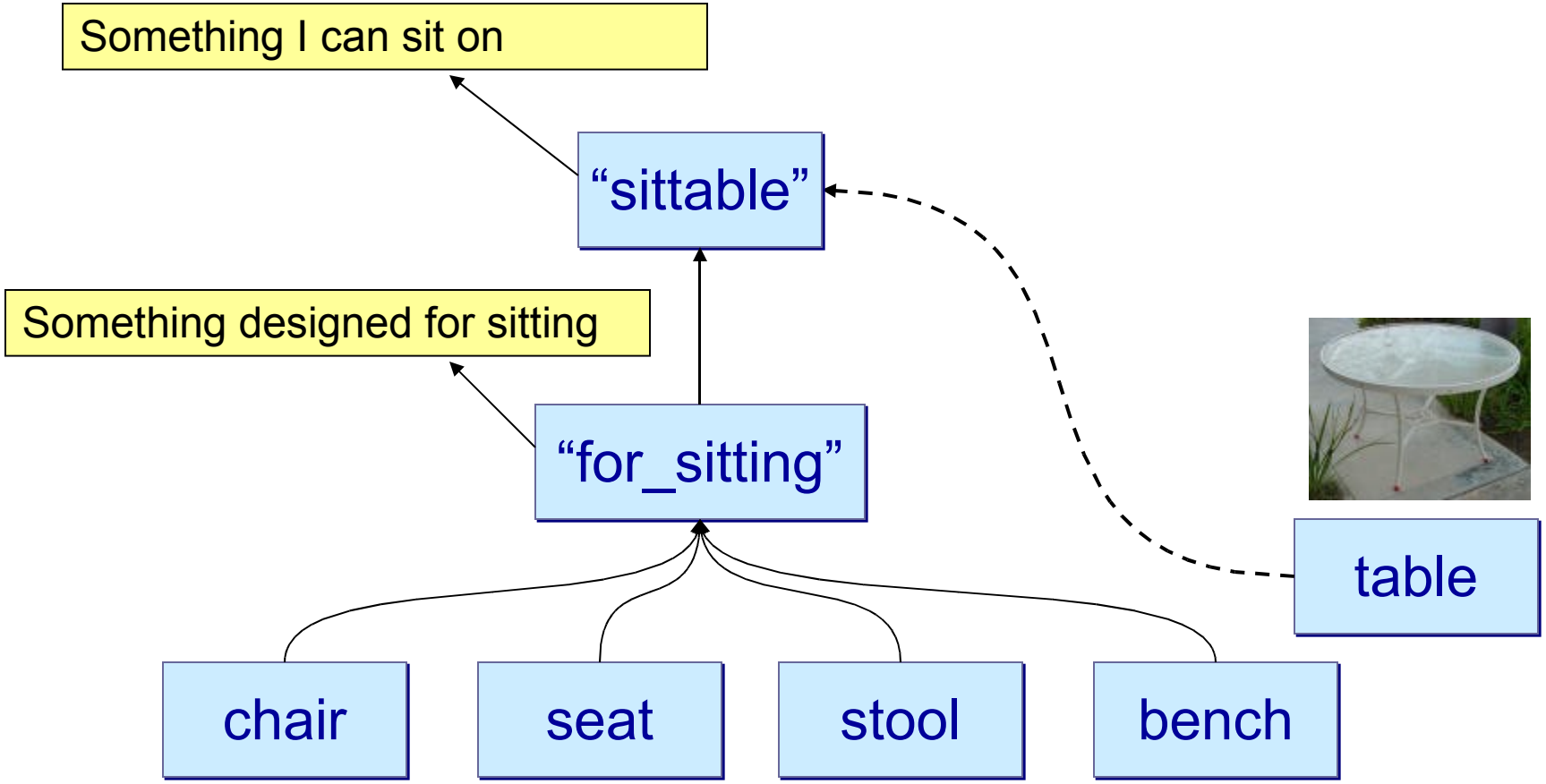
Example



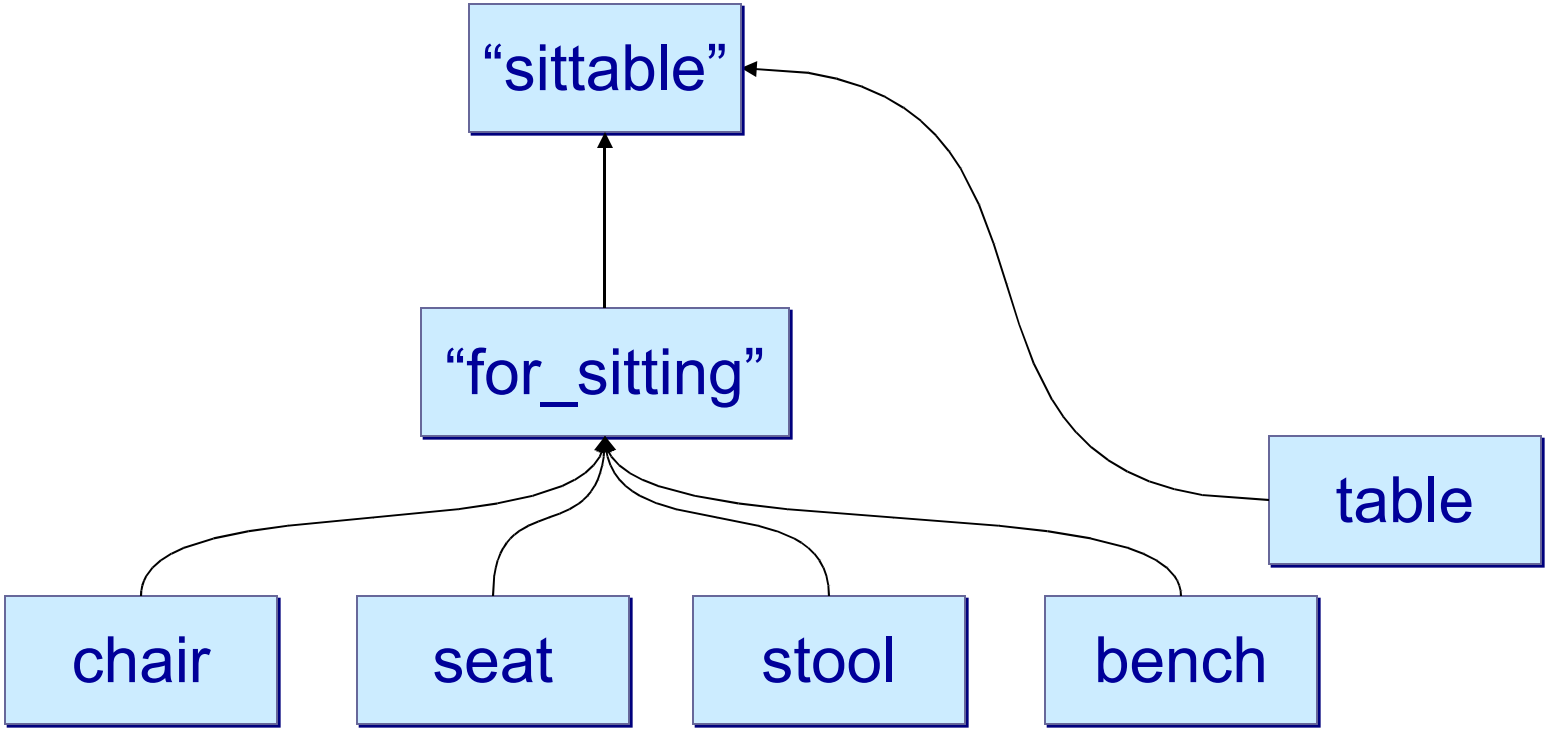
Example


Something I can sit on
“sittable”
table
chair
seat
stool
bench

Example


Something I can sit on
“sittable”
Something designed for sitting
“for_sitting”
table
chair
seat
stool
bench

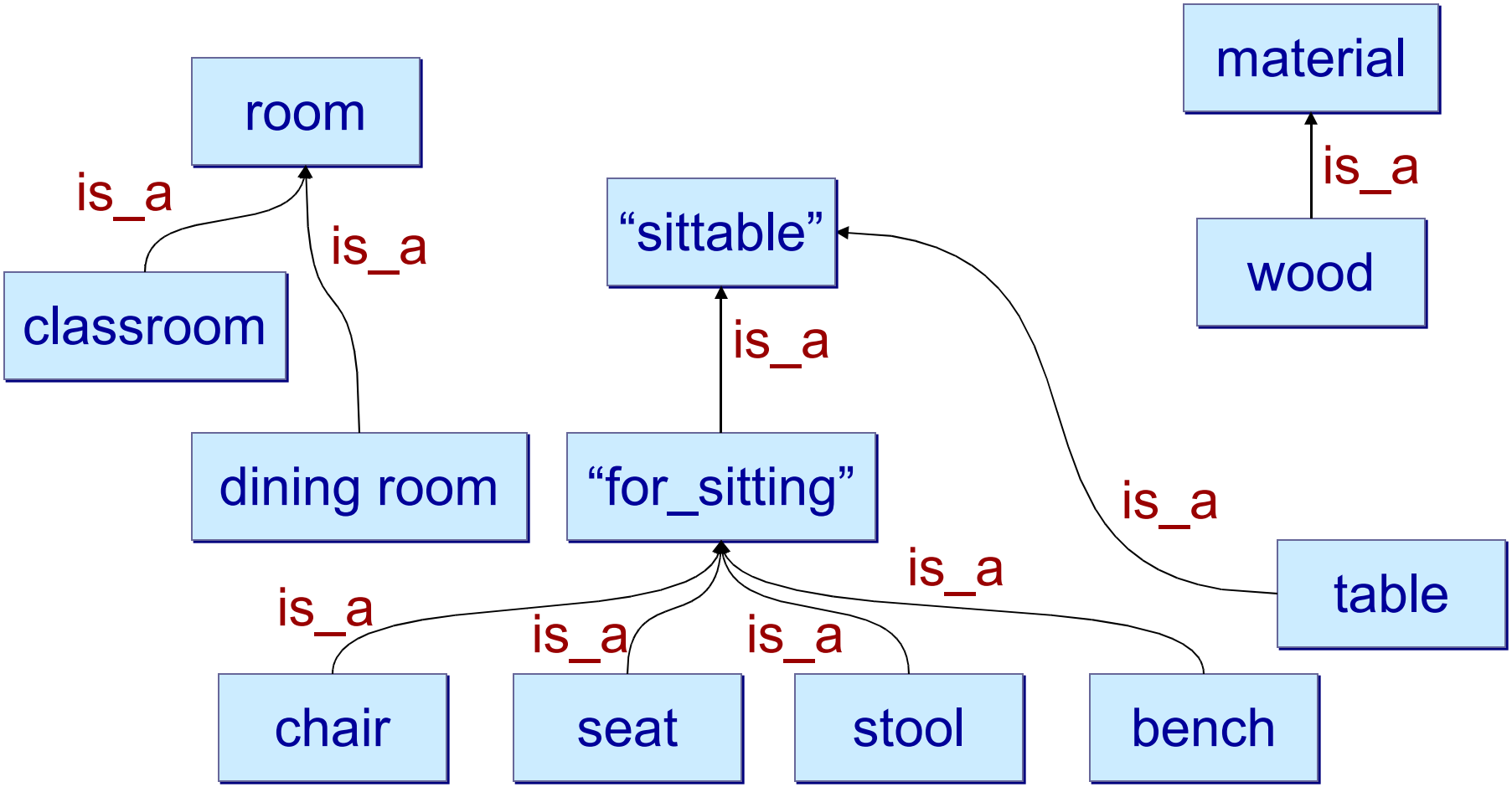
Ontology structure



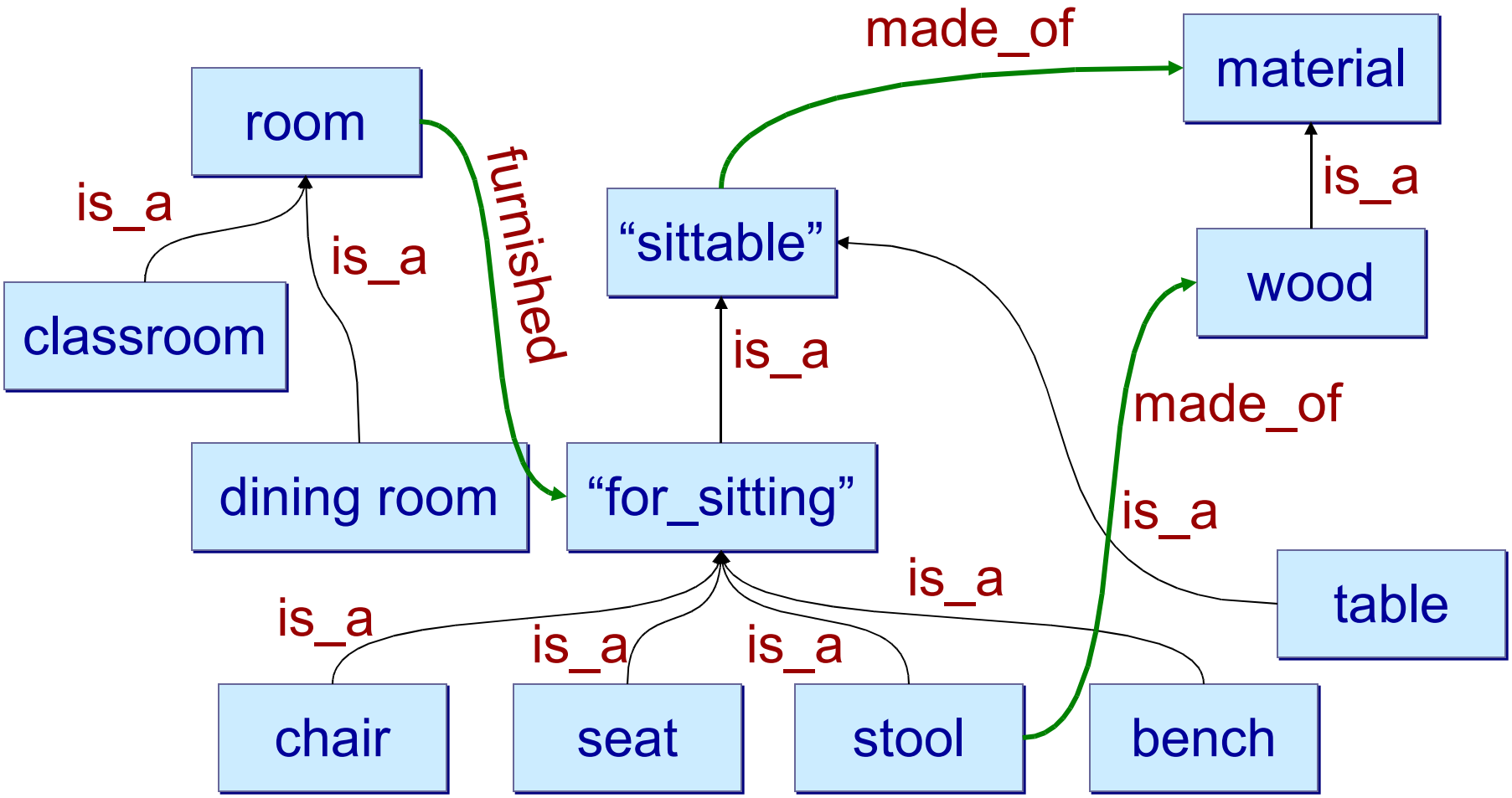
Ingredients

- Concepts
  - shorthand name (internal use)
  - synthetic title (to be displayed)
  - definition (real unambiguous shared definition)
- Relationships among concepts
  - is_a
  - other
- Annotations

Relationships



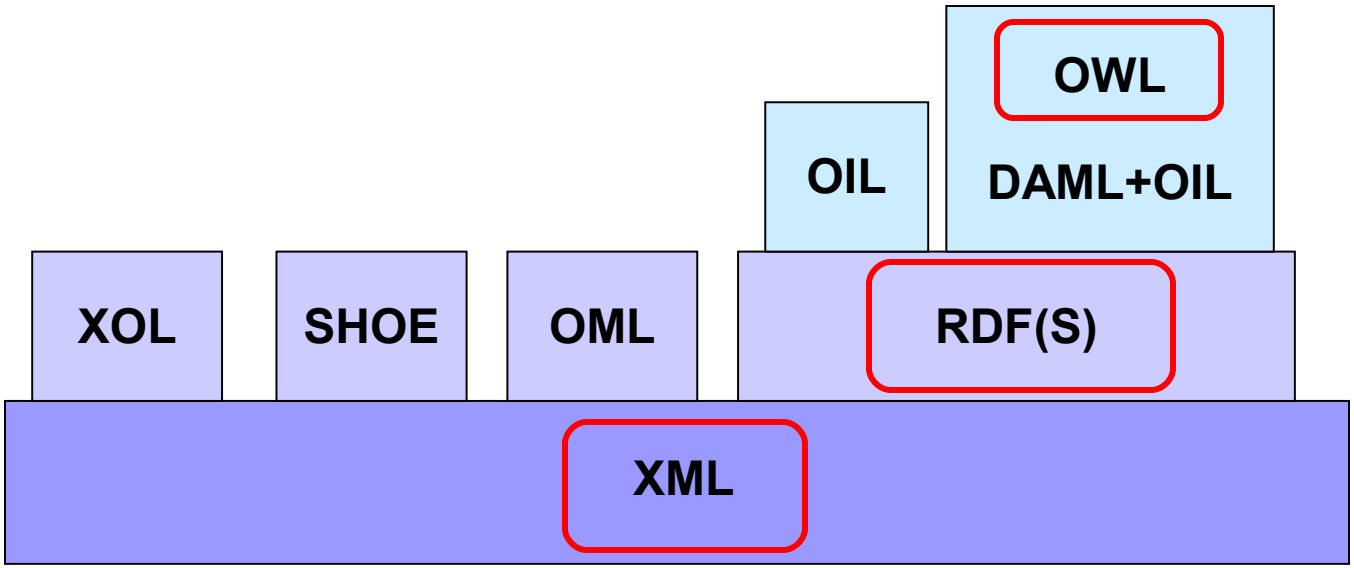
Relationships



Ontology building blocks

- Ontologies generally describe:
  - Individuals
    - the basic or "ground level" objects
  - Classes
    - sets, collections, or types of objects
  - Attributes
    - properties, features, characteristics, or parameters that objects can have and share
  - Relationships
    - ways that objects can be related to one another

Ontology languages



RDF

- Resource Description Framework
  - Resource = URI (retrievable, or not)
- RDF is structured in **statements**
- A statement is a **triple**:
  - Subject predicate object
    - Subject: a resource
    - Predicate: a verb / property / relationship
    - Object: a resource, or a literal string

RDF examples

- http://elite.polito.it/people#FulvioCorno http://purl.org/vocab/relationship/employedBy http://www.polito.it
- http://elite.polito.it/people#FulvioCorno hasProfession http://reliant.teknowledge.com/DAML/Mid-level-ontology.owl#Professor

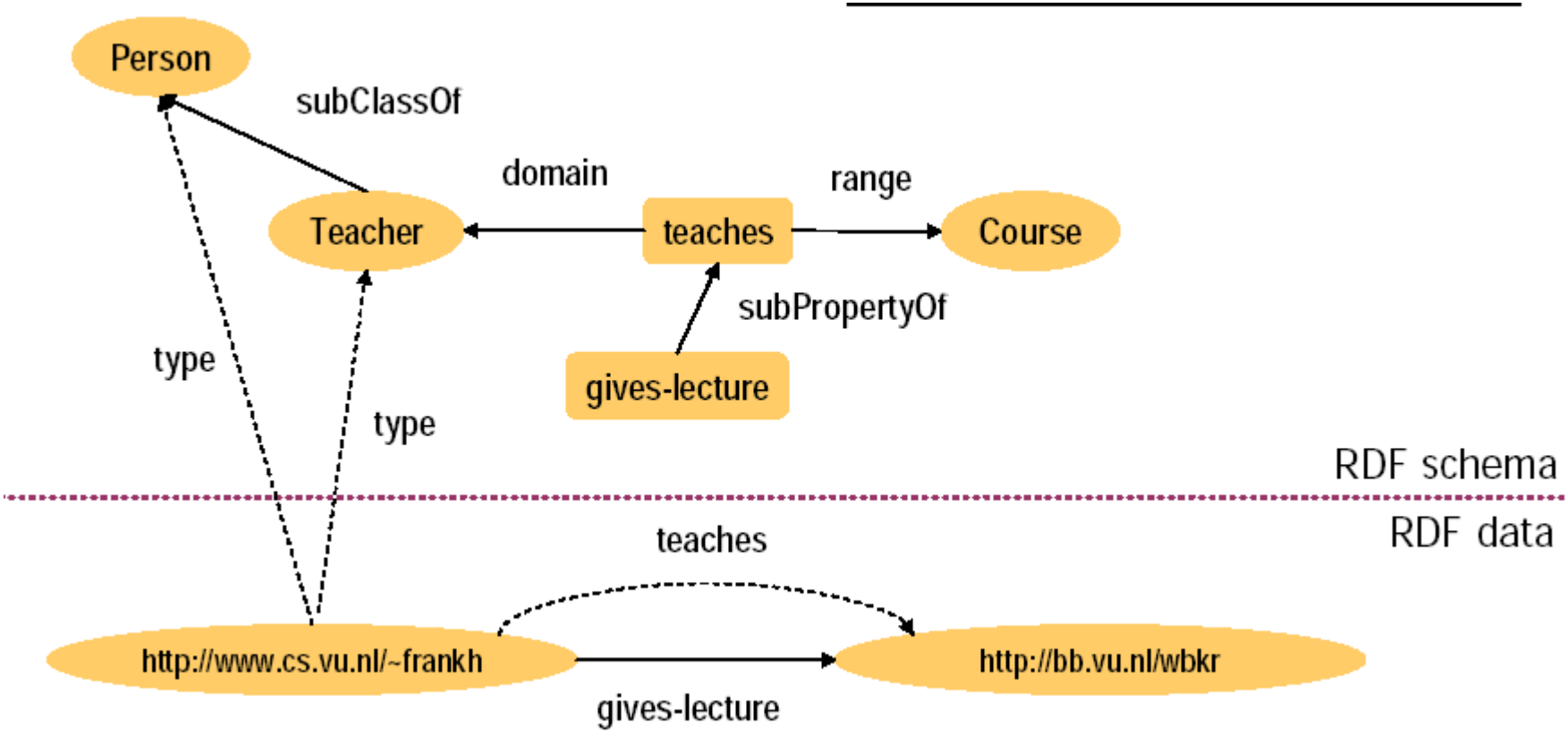
RDF in XML

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
        <rdf:Description
              rdf:about="http://en.wikipedia.org/wiki/Tony_Benn">
                 <dc:title>Tony Benn</dc:title>
                 <dc:publisher>Wikipedia</dc:publisher>
        </rdf:Description>
</rdf:RDF>
```

RDF from multiple dictionaries

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
        <rdf:Description
rdf:about="http://en.wikipedia.org/wiki/Tony_Benn">
                <dc:title>Tony Benn</dc:title>
                <dc:publisher>Wikipedia</dc:publisher>
                <foaf:primaryTopic>
                     <foaf:Person>
                          <foaf:name>Tony Benn</foaf:name>
                     </foaf:Person>
                </foaf:primaryTopic>
        </rdf:Description>
</rdf:RDF>
```

RDF and RDF schema

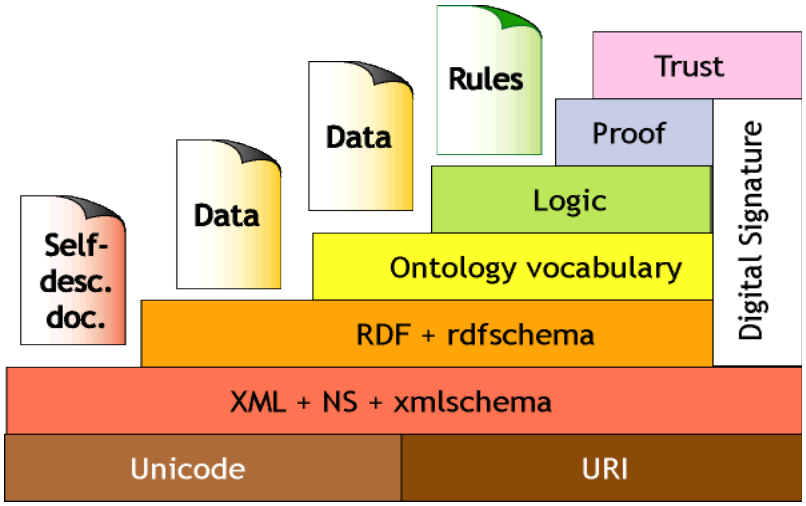


Formal ontologies

- A formal ontology
  - is defined using a language with a formal semantics
    - well defined meaning of the language constructs
    - well known rules to combine them
  - is bound to a given logic theory
    - first-order logic
    - description logic
    - ...

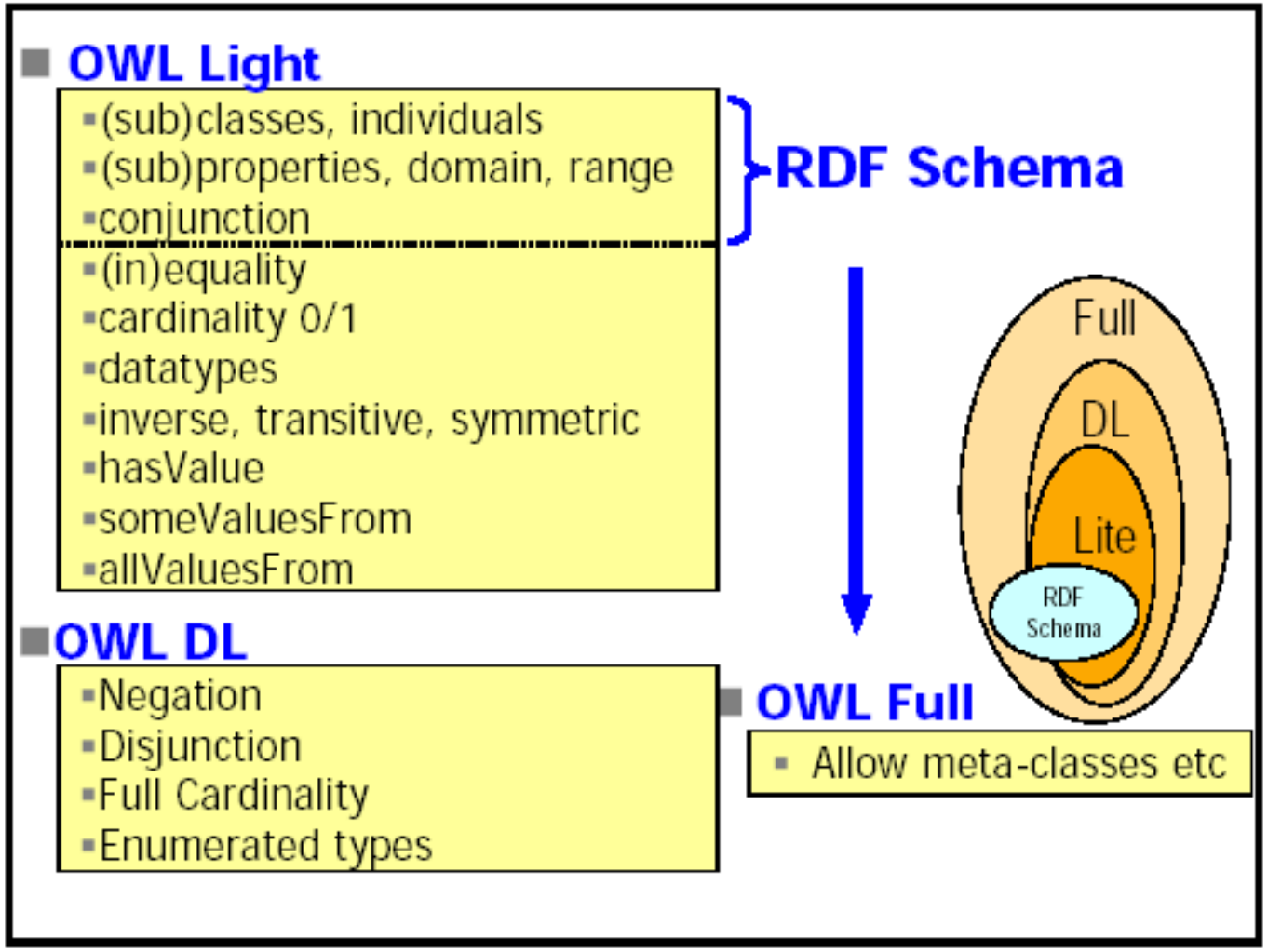
OWL - Introduction

- 4th level on the semantic web cake
- built on top of
  - XML
  - RDF/S



- Three versions
  - Lite
  - DL (maps to Description Logic)
  - Full (not fully tractable)
- Serializable as XML

OWL: Ontology Web Language



OWL-DL

- Based on Description Logic
- Well defined formal semantics
  - well defined rules to treat sentence meaning
  - well defined assumptions on the world being modeled
- Well known reasoning/inferencing algorithms
  - tractable, conclusions can be derived in finite time
- Widely available reasoning systems
  - no need to re-invent the wheel

Building blocks in OWL

- Ontology declaration (XML syntax)

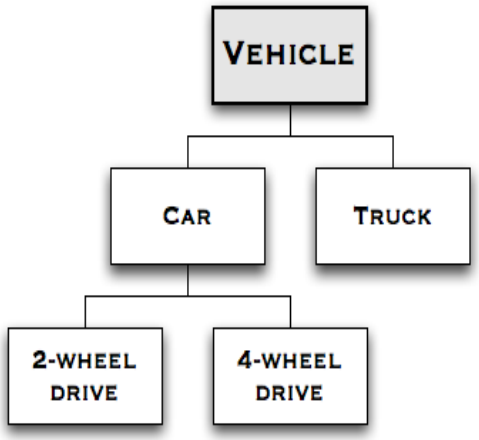
```
<rdf:RDF xmlns:owl =http://www.w3.org/2002/07/owl#"
xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:xsd ="http://www.w3.org/2001/XMLSchema#">
```

- Ontology metadata (information about the ontology)

```
<owl:Ontology rdf:about="">
    <rdfs:comment>An example OWL ontology</rdfs:comment>
    <owl:priorVersion
    rdf:resource="http://www.mydomain.org/uni-ns-old"/>
    <owl:imports
    rdf:resource="http://www.mydomain.org/persons"/>
    <rdfs:label>University Ontology</rdfs:label>
</owl:Ontology>
```

Classes

- Every class is a descendant of
  - owl:Thing
- Are defined using
  - owl:Class




```
<owl:Class rdf:ID="Vehicle"/>
```

- Equivalence
  - owl:equivalentClass

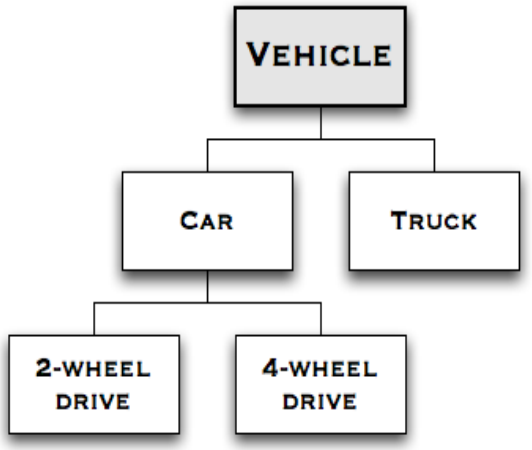
```
<owl:Class rdf:ID="Car">

   <owl:equivalentClass
rdf:resource="#Automobile"/>

</owl:Class>
```

Subsumption

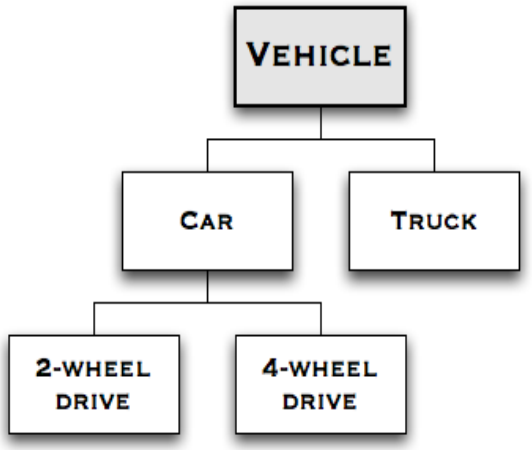
- Provided by
  - owl:subClassOf




```
<owl:Class rdf:ID="2-Wheel-Drive">
   <rdfs:subClassOf rdf:resource="#Car"/>
</owl:Class>
```

Partitions

- Disjoint partition
  - owl:disjointWith



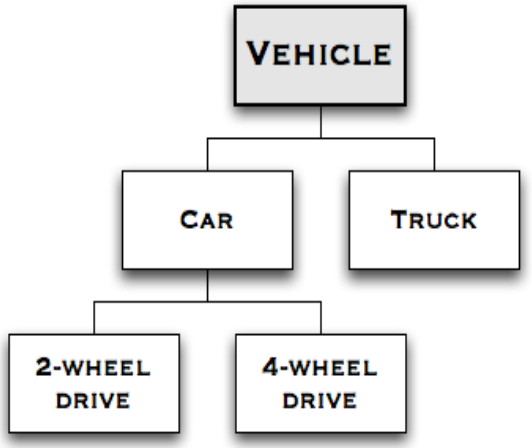

```
<owl:Class rdf:about="#2-Wheel-Drive">

   <owl:disjointWith
   rdf:resource="#4-Wheel-Drive"/>

</owl:Class>
```

Partitions

- Exhaustive partition
  - owl:oneOf




```
<owl:Class rdf:ID="Car">
   <owl:oneOf rdf:parseType="Collection">
      <owl:Thing rdf:about="#2-Wheel-Drive"/>
      <owl:Thing rdf:about="#4-Wheel-Drive"/>
   </owl:oneOf>
</owl:Class>
```

Attributes

- Known as “properties”
- Subdivided in
  - Datatype properties
    - Attributes that specify a class features by means of data (XSD datatype)
    - phone, title, age
  - Object properties
    - Attributes that define relationships between classes (Relations)
    - isTaughtBy(Class(course), Class(professor))

Attributes - datatypeProperties

- Allow to describe a specific aspect of a concept
  - Based on XSD data types
  - The range specifies the data type
  - The domain specifies the class to which the property is referred

```
<owl:DatatypeProperty rdf:ID="age">

   <rdfs:domain rdf:resource="#Person"/>

   <rdfs:range  rdf:resource="http://www.w3.org/
2001/XMLSchema#nonNegativeInteger"/>

</owl:DatatypeProperty>
```

Relationships

- Directed (from one concept to another, no viceversa)
- Defined through object properties
  - domain
    - the class(es) from which the relation departs
  - range
    - the relation destination(s)
- Subsumption between relationships is possible

Relationships

- Example

```
<owl:ObjectProperty rdf:ID="isTaughtBy">

   <rdfs:domain rdf:resource="#course"/>

   <rdfs:range
rdf:resource="#academicStaffMember"/>

   <rdfs:subPropertyOf rdf:resource="#involves"/>

</owl:ObjectProperty>
```

Instances (Individuals)

- Defined by means of
  - rdf:Description + rdf:Type
- No unique name assumption
  - if two instances have a different name or ID this does not imply that they are different individuals.

```
<academicStaffMember rdf:ID="949352">
   <uni:age rdf:datatype="&xsd;integer">
      39
   </uni:age>
</academicStaffMember>

<rdf:Description rdf:ID="949353">
<rdf:type rdf:resource="#academicStaffMember"/>
</rdf:Description>
```

Advanced constructs

- OWL supports several advanced constructs to define classes and relationships
- Intensional definition of classes
  - by defining constraints on attribute values (either object or datatype properties)

```
<owl:Class rdf:about="#academicStaffMember">
   <rdfs:subClassOf>
      <owl:Restriction>
      <owl:onProperty rdf:resource="#teaches"/>
      <owl:someValuesFrom
         rdf:resource="#undergraduateCourse"/>
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
```

Advanced constructs

- Cardinality
  - Used to fix the number of instances that can be related
  - example:
    - A department should have at least 10 members

```
<owl:Class rdf:about="#department">
 <rdfs:subClassOf>
   <owl:Restriction>
   <owl:onProperty rdf:resource="#hasMember"/>
   <owl:minCardinality
      rdf:datatype="&xsd;nonNegativeInteger">
      10
   </owl:minCardinality>
   </owl:Restriction>
 </rdfs:subClassOf>
</owl:Class>
```

Special properties (1/2)

- owl:TransitiveProperty
  - defines a transitive property, such as “has better grade than”, “is taller than”, or “is ancestor of”.
- owl:SymmetricProperty
  - defines a symmetric property, such as “has same grade as” or “is sibling of”.
- owl:FunctionalProperty
  - defines a property that has at most one value for each object, such as “age”, “height”, or “directSupervisor”.

Special properties (2/2)

- owl:InverseFunctionalProperty
  - defines a property for which two different objects cannot have the same value
  - example
    - the property “isTheSocialSecurityNumberFor”
    - a social security number is assigned to one person only.

OWL class constructors

Constructor	DL Syntax	Example	Modal Syntax
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male	$C_1 \wedge \ldots \wedge C_n$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer	$C_1 \vee \ldots \vee C_n$
complementOf	$\neg C$	$\neg$Male	$\neg C$
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	{john} $\sqcup$ {mary}	$x_1 \vee \ldots \vee x_n$
allValuesFrom	$\forall P.C$	$\forall$hasChild.Doctor	$[P]C$
someValuesFrom	$\exists P.C$	$\exists$hasChild.Lawyer	$\langle P \rangle C$
maxCardinality	$\leqslant nP$	$\leqslant$1hasChild	$[P]_{n+1}$
minCardinality	$\geqslant nP$	$\geqslant$2hasChild	$\langle P \rangle_n$

OWL axioms

Axiom	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human $\sqsubseteq$ Animal $\sqcap$ Biped
equivalentClass	$C_1 \equiv C_2$	Man $\equiv$ Human $\sqcap$ Male
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male $\sqsubseteq \neg$Female
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{President_Bush} $\equiv$ {G_W_Bush}
differentFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	{john} $\sqsubseteq \neg$ {peter}
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter $\sqsubseteq$ hasChild
equivalentProperty	$P_1 \equiv P_2$	cost $\equiv$ price
inverseOf	$P_1 \equiv P_2^-$	hasChild $\equiv$ hasParent$^-$
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+ \sqsubseteq$ ancestor
functionalProperty	$\top \sqsubseteq \leqslant 1 P$	$\top \sqsubseteq \leqslant 1$hasMother
inverseFunctionalProperty	$\top \sqsubseteq \leqslant 1 P^-$	$\top \sqsubseteq \leqslant 1$hasSSN$^-$

Open questions...

- Reasoning is only feasible for OWL-DL
  - limitations
    - vocabulary partition
      - a resource can only be a class, an object property, a datatype property (a class cannot be at same time an individual)
    - explicit typing
      - partitioning must be stated explicitly
    - no transitive cardinality restrictions
      - transitive properties cannot have cardinality restrictions
- Some semantics is still missing
  - e.g. how to define what a given relation means?

SWRL (quick glance)

- SWRL
  - Semantic Web Rule Language
- Can be used to define new relationships (thus provides their operational semantics)
- Builds on top of RDF
- Example:
  - Base predicates
    - `Father(?x,?y)`
    - `Mother(?x,?z)`
  - New relations (defined through SWRL)
    - `Father(?x,?y)→Parent(?x,?y)`
    - `Mother(?x,?z)→Parent(?x,?y)`

SWRL (quick glance)

- Example 2
  - DomoMLPlus:Room(?x) ^
    DomoMLPlus:hasWall(?x, ?z) ^
    DomoMLPlus:hasWallOpening(?z, ?y)^
    DomoMLPlus:Window(?y)→
    DomoMLPlus:hasWindows(?x, true)

No ontology, no semantics

- **Reuse** of existing taxonomies as ontologies
- Publicly **available** ontologies
  - Horizontal ontologies (SUMO, MILO, WordNet, ...)
  - Vertical ontologies (domain-specific)
- Semi-automatic ontology **learning** from resources
- Manual development / **enrichment**