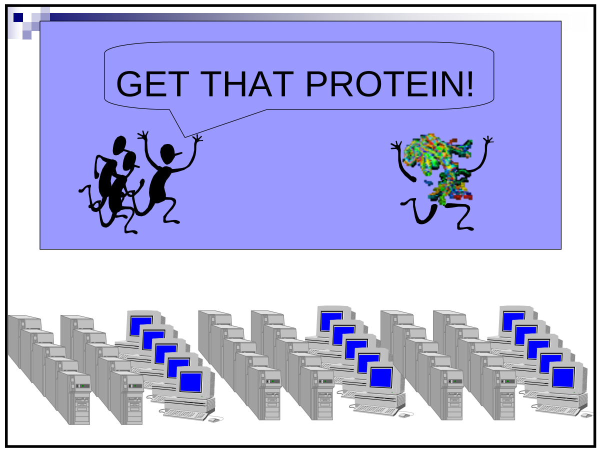
# Towards a Semantic Web

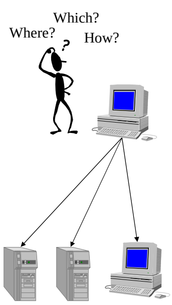
Patrick Lambrix
Linköpings universitet




## Locating relevant information

Vision: Web services

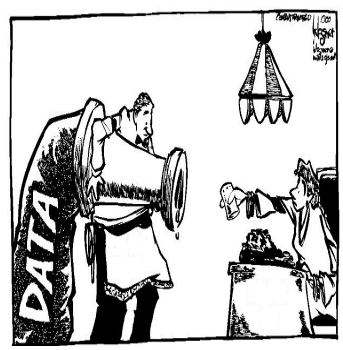
- Databases and tools (service providers) announce their service capabilities
- Users request services which may be based on task descriptions
- Service matchers find relevant services (composition) based on user needs and user preferences, negotiate service delivery, and deliver results to user




## Retrieving relevant information

Vision:

Based on the meaning of the query:
- only relevant information is retrieved
- all relevant information is retrieved




## Integrating information

Vision:

Integrate data sources that are heterogeneous in content, data quality, data models, access methods, terminology

Genomics → DISCOVERY
Clinical trials → DISCOVERY
Disease information → DISCOVERY
Metabolism, toxicology → DISCOVERY
Chemical structure → DISCOVERY
Target structure → DISCOVERY
Disease models → DISCOVERY

## Today: syntactic Web

- A library of documents (web pages) interconnected by links
- A common portal to applications accessible through web pages, and presenting their results as web pages

A place where computers do the presentation (easy) and people do the linking and interpreting (hard).

## Semantic Web

W3C: Facilities to put machine-understandable data on the Web are becoming a high priority for many communities. The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.

## What is the problem?

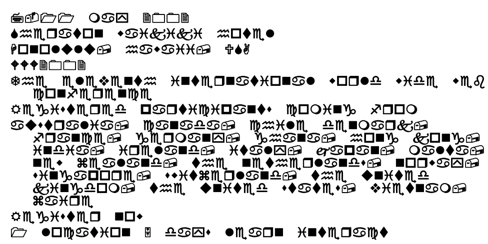

Example based on example on slides by P. Patel-Schneider

## What information can we see…

Date: 13-15 June, 2005
Location: Linköping
Sponsors: IEEE, CERC, LiU
14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborating Enterprises (WETICE-2005)
Welcome to WETICE-2005

…

## What information can a machine see…



## Use XML markup with “meaningful” tags

<date> 13-15 June 2005 </date>

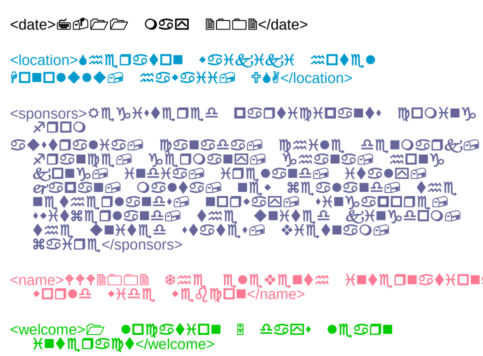
<location> Linköping </location>

<sponsors>IEEE, CERC, LiU </sponsors>

<name> 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborating Enterprises (WETICE-2005) </name>

<welcome> Welcome to WETICE-2005 </welcome>

## Machine sees …



## But what about …

<date> 13-15 June 2005 </date>

<place> Linköping </place>

<sponsors>IEEE, CERC, LiU </sponsors>

<conf> 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborating Enterprises (WETICE-2005) </conf>

<introduction> Welcome to WETICE-2005 </introduction>

## Machine sees …

<date> </date>

<sponsors> </sponsors>

## Adding "Semantics" – first approach

External agreement on meaning of annotations

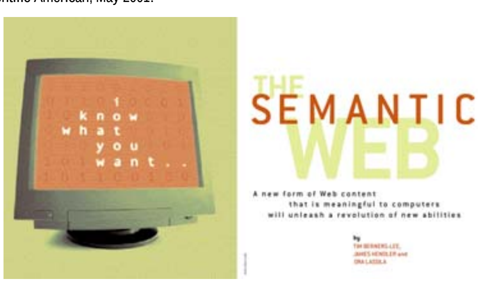
- Agree on the meaning of a set of annotation tags
- Problems with this approach:
  - Inflexible
  - Limited number of things can be expressed

## Adding "Semantics" – second approach

Use on-line 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

Scientific American, May 2001:




- First step towards the vision:
  adding semantic annotation to web resources

## Semantic annotations based on ontologies

- Locating information
  - Web service descriptions use ontologies
  - Users use ontologies when formulating requests
  - Service matchers find services based on meaning
- Retrieving relevant information
  - Reduce non-relevant information (precision)
  - Find more relevant information (recall)
- Integrating information
  - Relating similar entities in different databases

# Ontologies

## Ontologies

- Definition
- Use
- Components
- Knowledge representation

## Ontologies

*"Ontologies define the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary."*

*(Neches, Fikes, Finin, Gruber, Senator, Swartout, 1991)*

## Definitions

- Ontology as specification of a conceptualization
- Ontology as philosophical discipline
- Ontology as informal conceptual system
- Ontology as formal semantic account
- Ontology as representation of conceptual system via a logical theory
- Ontology as the vocabulary used by a logical theory
- Ontology as a meta-level specification of a logical theory

(Guarino, Giaretta)

## Definitions

- An ontology is an explicit specification of a conceptualization (Gruber)
- An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base. (Swartout, Patil, Knight, Russ)
- An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base. (Bernaras, Lasergoiti, Correra)
- An ontology is a formal, explicit specification of a shared conceptualization (Studer, Benjamins, Fensel)

## Example

GENE ONTOLOGY (GO)

```
immune response
  i- acute-phase response
  i- anaphylaxis
  i- antigen presentation
  i- antigen processing
  i- cellular defense response
  i- cytokine metabolism
     i- cytokine biosynthesis synonym cytokine production
        …
     p- regulation of cytokine biosynthesis
       …
  …
  i- B-cell activation
    i- B-cell differentiation
    i- B-cell proliferation
  i- cellular defense response
  …
  i-  T-cell activation
     i- activation of natural killer cell activity
       …
```

## Example Ontologies

- Knowledge representation ontology: frame ontology
- Top level ontologies: TLO, Cyc
- Linguistic ontologies: GUM, WordNet
- Engineering ontologies: EngMath, PhysSys
- Domain ontologies: CHEMICALS, Gene Ontology, Open Biomedical Ontologies

## Ontologies used …

- for communication between people and organizations
- for enabling knowledge reuse and sharing
- as basis for interoperability between systems
- as repository of information
- as query model for information sources

Key technology for the Semantic Web

## Biomedical Ontologies - efforts

OBO – Open Biomedical Ontologies
http://www.obofoundry.org/
(over 50 ontologies)

” The mission of OBO is to support community members who are developing and publishing ontologies in the biomedical domain. It is our vision that a core of these ontologies will be fully interoperable, by virtue of a common design philosophy and implementation, thereby enabling scientists and their instruments to communicate with minimum ambiguity. In this way the data generated in the course of biomedical research will form a single, consistent, cumulatively expanding, and algorithmically tractable whole. This core will be known as the "OBO Foundry". .”

## OBO Foundry

1. open and available
2. common shared syntax
3. unique identifier space
4. procedures for identifying distinct successive versions
5. clearly specified and clearly delineated content
6. textual definitions for all terms
7. use relations from OBO Relation Ontology
8. well documented
9. plurality of independent users
10. developed collaboratively with other OBO Foundry members

## Biomedical Ontologies - efforts

National Center for Biomedical Ontology
http://bioontology.org/index.html
Funded by National Institutes of Health

”The goal of the Center is to support biomedical researchers in their knowledge-intensive work, by providing online tools and a Web portal enabling them to access, review, and integrate disparate ontological resources in all aspects of biomedical investigation and clinical practice. A major focus of our work involves the use of biomedical ontologies to aid in the management and analysis of data derived from complex experiments.”

## Biomedical Ontologies - efforts

- Gene Ontology Consortium (GO): molecular function, biological process, cellular component
- Standards and Ontologies for Functional Genomics (SOFG): meeting and website
- Proteomics Standards Initiative
- Plant Ontology consortium

## Biomedical Ontologies - efforts

- International Health Terminology Standards Development Organisation

  http://www.ihtsdo.org

SNOMED CT (**S**ystematized **No**menclature of **Med**icine-**C**linical **T**erms)

## Ontologies in biomedical research

- many biomedical ontologies

  e.g. GO, OBO, SNOMED-CT
- practical use of biomedical ontologies

  e.g. databases annotated with GO

**GENE ONTOLOGY (GO)**

```
immune response
  i- acute-phase response
  i- anaphylaxis
  i- antigen presentation
  i- antigen processing
  i- cellular defense response
  i- cytokine metabolism
     i- cytokine biosynthesis
        synonym cytokine production
        …
     p- regulation of cytokine
        biosynthesis
     …
  …
  i- B-cell activation
    i- B-cell differentiation
    i- B-cell proliferation
  i- cellular defense response
  …
  i-  T-cell activation
     i- activation of natural killer
        cell activity
     …
```

## Components

- concepts

  - represent a set or class of entities in a domain

    *immune response*
  - organized in taxonomies

    (hierarchies based on e.g. *is-a* or *is-part-of*)

    *immune response* is-a *defense response*
- instances

  - often not represented in an ontology

  (instantiated ontology)

## Components

- relations

  R: C1 x C2 x … x Cn

  *Protein hasName ProteinName*

  *Chromosone hasSubcellularLocation Nucleus*

## Components

- axioms

  'facts that are always true'

*The origin of a protein is always of the type 'gene coding origin type'*

*Each protein has at least one source.*

*A helix can never be a sheet and vice versa.*

## Different kinds of ontologies

- Controlled vocabularies

  Concepts
- Taxonomies

  Concepts, is-a
- Thesauri

  Concepts, predefined relations
- Data models (e.g. EER, UML)

  Concepts, relations, axioms
- Logics

  Concepts, relations, axioms

## Taxonomy - GeneOntology

id: GO:0003674 name: molecular_function
def: "Elemental activities, such as catalysis or binding, describing the actions of a gene product at the molecular level. A given gene product may exhibit one or more molecular functions."

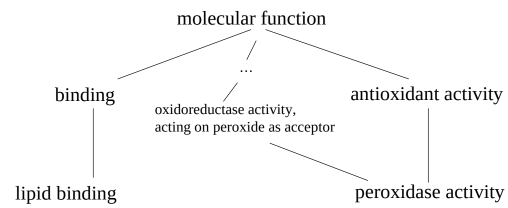
id: GO:0015643 name: binding
def: "The selective, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule."
is-a: GO:0003674 ! molecular_function

id: GO:0008289 name: lipid binding
is_a: GO:0015643 ! binding

id: GO:0016209 name: antioxidant activity
def: "Inhibition of the reactions brought about by dioxygen (O2) or peroxides. Usually the antioxidant is effective because it can itself be more easily oxidized than the substance protected."
is_a: GO:0003674 ! molecular_function

id: GO:0004601 name: peroxidase activity
def: "Catalysis of the reaction: donor + H2O2 = oxidized donor + 2 H2O."
is_a: GO:0016209 ! antioxidant activity
is_a: GO:0016684 ! oxidoreductase activity, acting on peroxide as acceptor

## Taxonomy - GeneOntology




## Thesaurus

- graph
- fixed set of relations

(synonym, narrower term, broader term, similar)

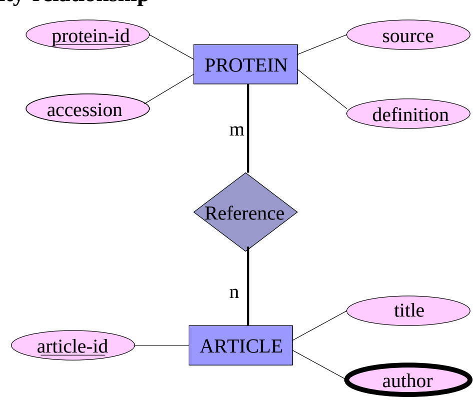
## Thesaurus - WordNet

thesaurus, synonym finder
- => wordbook
  - => reference book, reference, reference work, book of facts
    - => book
      - => publication
        - => print media
          - => medium
            - => means
              - => instrumentality, instrumentation
                - => artifact, artefact
                  - => object, inanimate object, physical object
                    - => entity
        - => work, piece of work
          - => product, production
            - => creation
              - => artifact, artefact
                - => object, inanimate object, physical object
                  - => entity

## OO Data models

- EER
  entity types, attributes, relationships, cardinality constraints, taxonomy
- UML
  classes, attributes, associations, cardinality constraints, taxonomy, operations
- Taxonomy/inheritance – semantics?
- Intuitive, lots of tools, widely used.

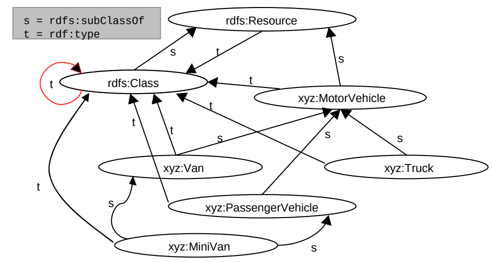
## Entity-relationship



## RDF + RDF Schema

- Basic construct: sentence: *Subject Predicate Object*
  - Encoded in XML
  - Can be seen as ground atomic formula
  - Represented as graph
- RDF Schema
- Editors, query tools exist

## RDF Schema - example




# Logics

- Formal languages
- Syntax, semantics, inference mechanisms

# Logics

Reasoning services used in

- Ontology design

  Check concept satisfiability, ontology satisfiability and (unexpected) implied relationships
- Ontology aligning and merging

  Assert inter-ontology relationships.

  Reasoner computes integrated concept hierarchy/consistency.
- Ontology deployment

  Determine if a set of facts are consistent w. r. t. ontology.

  Determine if individuals are instances of ontology concepts.

  Query inclusion.

  Classification-based querying.

## Description Logics

- A family of KR formalisms tailored for expressing knowledge about concepts and concept hierarchies
- Based on FOPL, supported by automatic reasoning systems
- Basic building blocks: concepts (concepts), roles (binary relations), individuals (instances)
- Language constructs can be used to define new concepts and roles (axioms).
  - Intersection, union, negation, quantification, …
- Knowledge base is Tbox + Abox
  - Tbox: concept level - axioms: equality and subsumption (is-a)
  - Abox: instance level - axioms: membership, relations
- Reasoning services
  - Satisfiability of concept, Subsumption/Equivalence/Disjointness between concepts, Classification, Instantiation, Retrieval

## Description Logics

Intersection

Signal-transducer-activity $\cap$ binding

Negation

$\neg$ Helix

Quantifiers

$\exists$ hasOrigin.Mitochondrion

$\forall$ hasOrigin.Gene-coding-origin-type

## DAML+OIL / OWL

- DAML+ OIL almost equivalent to SHIQ
- DAML+ OIL supports the full range of XML Schema data types
- OWL updated DAML+OIL

## DAML+OIL Class Constructors

| Constructor | DL Syntax | Example |
|---|---|---|
| intersectionOf | $C_1 \sqcap \ldots \sqcap C_n$ | Human $\sqcap$ Male |
| unionOf | $C_1 \sqcup \ldots \sqcup C_n$ | Doctor $\sqcup$ Lawyer |
| complementOf | $\neg C$ | $\neg$Male |
| oneOf | $\{x_1 \ldots x_n\}$ | {john, mary} |
| toClass | $\forall P.C$ | $\forall$hasChild.Doctor |
| hasClass | $\exists P.C$ | $\exists$hasChild.Lawyer |
| hasValue | $\exists P.\{x\}$ | $\exists$citizenOf.{USA} |
| minCardinalityQ | $\geqslant n P.C$ | $\geqslant$2hasChild.Lawyer |
| maxCardinalityQ | $\leqslant n P.C$ | $\leqslant$1hasChild.Male |
| cardinalityQ | $= n\ P.C$ | =1 hasParent.Female |

- XMLS datatypes as well as classes
- Arbitrarily complex nesting of constructors
  - E.g., Person $\sqcap$ $\forall$hasChild.(Doctor $\sqcup$ $\exists$hasChild.Doctor)



## DAML+OIL Axioms

| Axiom | DL Syntax | Example |
|---|---|---|
| subClassOf | $C_1 \sqsubseteq C_2$ | Human $\sqsubseteq$ Animal $\sqcap$ Biped |
| sameClassAs | $C_1 \equiv C_2$ | Man $\equiv$ Human $\sqcap$ Male |
| subPropertyOf | $P_1 \sqsubseteq P_2$ | hasDaughter $\sqsubseteq$ hasChild |
| samePropertyAs | $P_1 \equiv P_2$ | cost $\equiv$ price |
| sameIndividualAs | $\{x_1\} \equiv \{x_2\}$ | {President_Bush} $\equiv$ {G_W_Bush} |
| disjointWith | $C_1 \sqsubseteq \neg C_2$ | Male $\sqsubseteq$ $\neg$Female |
| differentIndividualFrom | $\{x_1\} \sqsubseteq \neg\{x_2\}$ | {john} $\sqsubseteq$ $\neg${peter} |
| inverseOf | $P_1 \equiv P_2^-$ | hasChild $\equiv$ hasParent$^-$ |
| transitiveProperty | $P^+ \sqsubseteq P$ | ancestor$^+$ $\sqsubseteq$ ancestor |
| uniqueProperty | $\top \sqsubseteq \leqslant 1 P$ | $\top \sqsubseteq \leqslant$1hasMother |
| unambiguousProperty | $\top \sqsubseteq \leqslant 1 P^-$ | $\top \sqsubseteq \leqslant$1isMotherOf$^-$ |

- Axioms (mostly) reducible to subClass/PropertyOf



## OWL

- OWL-Lite, OWL-DL, OWL-Full: increasing expressivity
- A legal OWL-Lite ontology is a legal OWL-DL ontology is a legal OWL-Full ontology
- OWL-DL: expressive description logic, decidable
- XML-based
- RDF-based (OWL-Full is extension of RDF, OWL-Lite and OWL-DL are extensions of a restriction of RDF)

## OWL-Lite

- **Class**, subClassOf, equivalentClass
- intersectionOf (only named classes and restrictions)
- **Property**, subPropertyOf, equivalentProperty
- domain, range (global restrictions)
- inverseOf, TransitiveProperty (*), SymmetricProperty, FunctionalProperty, InverseFunctionalProperty
- allValuesFrom, someValuesFrom (local restrictions)
- minCardinality, maxCardinality (only 0/1)
- **Individual**, sameAs, differentFrom, AllDifferent

(*) restricted

## OWL-DL

- **Type separation (**class cannot also be individual or property, property cannot be also class or individual), Separation between DatatypeProperties and ObjectProperties
- **Class –complex classes**, subClassOf, equivalentClass, *disjointWith*
- <u>intersectionOf</u>, *unionOf, complementOf*
- **Property**, subPropertyOf, equivalentProperty
- domain, range (global restrictions)
- inverseOf, TransitiveProperty (*), SymmetricProperty, FunctionalProperty, InverseFunctionalProperty
- allValuesFrom, someValuesFrom (local restrictions), *oneOf, hasValue*
- <u>minCardinality, maxCardinality</u>
- **Individual**, sameAs, differentFrom, AllDifferent

(*) restricted

## Defining ontologies is not so easy ...

The Celestial Emporium of Benevolent Knowledge, Borges

"On those remote pages it is written that animals are divided into:
a. those that belong to the Emperor
b. embalmed ones
c. those that are trained
d. suckling pigs
e. mermaids
f. fabulous ones
g. stray dogs
h. those that are included in this classification
i. those that tremble as if they were mad
j. innumerable ones
k. those drawn with a very fine camel's hair brush
l. others
m. those that have just broken a flower vase
n. those that resemble flies from a distance"

Slide from talk by C. Goble

## Defining ontologies is not so easy ...

Dyirbal classification of objects in the universe

- Bayi: men, kangaroos, possums, bats, most snakes, most fishes, some birds, most insects, the moon, storms, rainbows, boomerangs, some spears, etc.
- Balan: women, anything connected with water or fire, bandicoots, dogs, platypus, echidna, some snakes, some fishes, most birds, fireflies, scorpions, crickets, the stars, shields, some spears, some trees, etc.
- Balam: all edible fruit and the plants that bear them, tubers, ferns, honey, cigarettes, wine, cake.
- Bala: parts of the body, meat, bees, wind, yamsticks, some spears, most trees, grass, mud, stones, noises, language, etc.

Slide from talk by C. Goble

## Ontology tools

- Ontology development tools
- Ontology merge and alignment tools
- Ontology evaluation tools
- Ontology-based annotation tools
- Ontology storage and querying tools
- Ontology learning tools

# Further reading

Starting points for further studies

## Further reading ontologies

- KnowledgeWeb ( http://knowledgeweb.semanticweb.org/ ) and its predecessor OntoWeb ( http://ontoweb.aifb.uni-karlsruhe.de/ )
- Lambrix, Tan, Jakoniene, Strömbäck, Biological Ontologies, chapter 4 in Baker, Cheung, (eds), *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, 85-99, Springer, 2007. ISBN: 978-0-387-48436-5.

(general about ontologies)

- Lambrix, Towards a Semantic Web for Bioinformatics using Ontology-based Annotation, *Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises*, 3-7, 2005. Invited talk.

(ontologies for semantic web)

- OWL, http://www.w3.org/TR/owl-features/ , http://www.w3.org/2004/OWL/