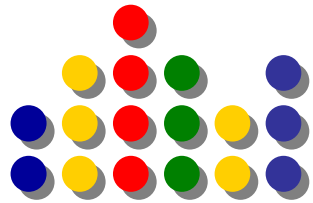
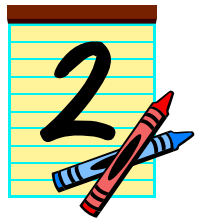
6th International Enformatika Conference
26-28 October 2005, Budapest, Hungary.

On the move to semantic Web services



Jorge Cardoso
Department of Mathematics and Engineering
University of Madeira (Portugal)

Semantic Web Processes Overview



- Introduction
- Semantic Web Processes Life cycle
- Web services Semantic Annotation
- Web services Discovery
- Semantic Process Composition
- Web service QoS
- Ontologies, Ontology Languages and Editors
- Projects/approaches: OWL-S, METEOR-S
- Conclusions


Syntactic Web
and Semantic
Web?

Syntactic Web

- World Wide Web
  - Primarily composed of documents written in HTML
    - HTML is a set of “markup” symbols
    - Useful for visual presentation
    - Designed only for human consumption
  - Humans can read Web pages and understand them
    - but their inherent meaning is not shown in a way that allows their interpretation by computers

Semantic Web

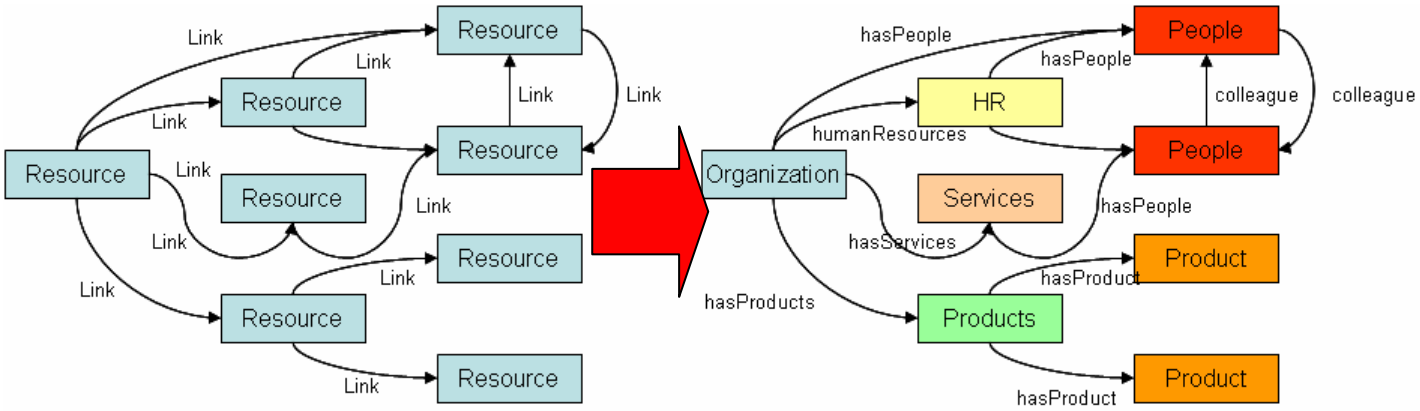
- Define Web information in a way that it can be used by computers
  - Not only for display purposes
  - But also for interoperability and integration between systems
- Challenge
  - Enable machine-to-machine exchange and automated processing
- One Solution
  - Provide the information in such a way that computers can understand it.
- This is the objective of the semantic Web
  - Make possible the processing of Web information by computers

Semantic Web

- “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” (Berners-Lee, Hendler et al. 2001).
- The next generation of the Web will combine:
  - Existing Web technologies
  - Knowledge representation formalisms (Grau 2004)

Semantic Web

- Currently the Web is in evolution



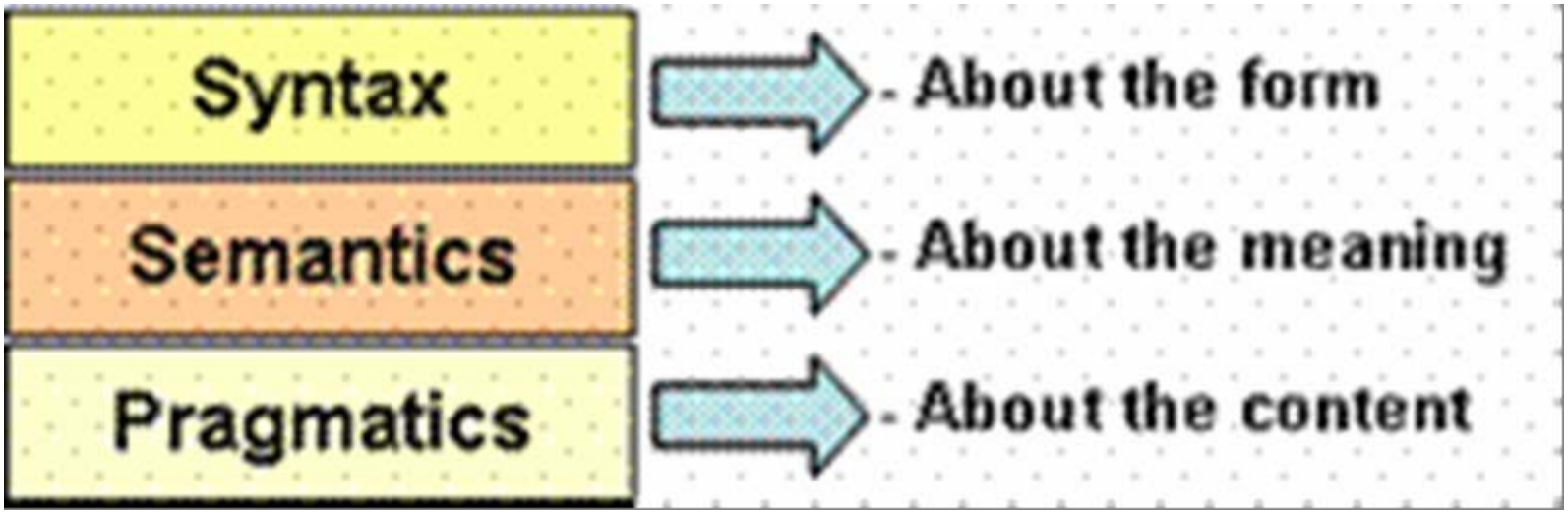
- Syntactic Web
  - Resources are linked together forming the Web
  - No distinction between resources or links
- Semantic Web
  - Resources and links have meaning
  - New standards and languages are being investigated and developed.

Semantic Web

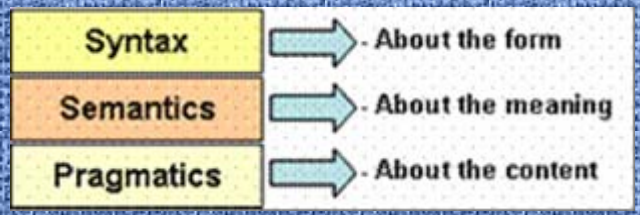
- To give meaning to Web resource and links, the research community has developed semantic standards such as
  - Resource Description Framework (RDF)
    - RDF is a standard for creating descriptions of information
    - What XML is for syntax, RDF is for semantics.
    - Provides a clear set of rules for providing simple descriptive information.
  - Web Ontology Language (OWL)
    - Is an extension of RDF
    - Provides a language for defining structured Web-based ontologies which allows a richer integration and interoperability of data among communities and domains.

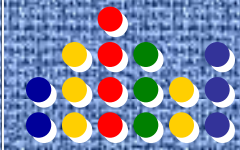
Semantic Web

Semiotics



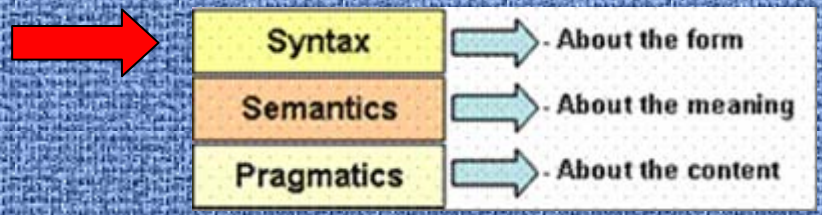
Semiotics

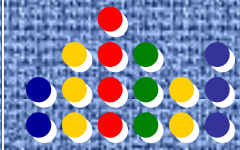




- General science of signs
  - Such as icons, images, objects, tokens, and symbols – and how their meaning is transmitted and understood.
  - A sign is generally defined as something that stands for something else.
  - The human language is a particular case of semiotics.
- Semiotics is composed of three components:
  - Syntax, semantics, and pragmatics.

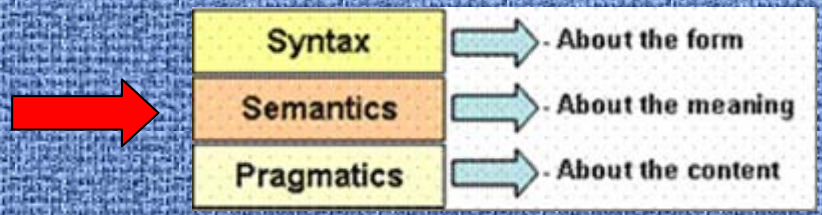
Syntax





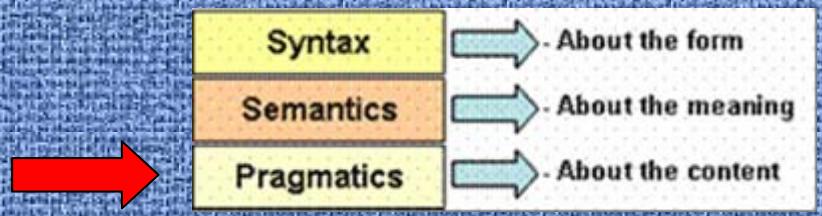
- **D**eals with the formal or structural relations between signs (or tokens) and the production of new ones.
  - For example, grammatical syntax is the study of which sequences of symbols are well formed according to the recursive rules of grammar.
    - If a program is syntactically correct according to its rules of syntax
    - The compiler will validate the syntax and will not generate error messages.
  - Does not ensure that the program is semantically correct.

Semantics

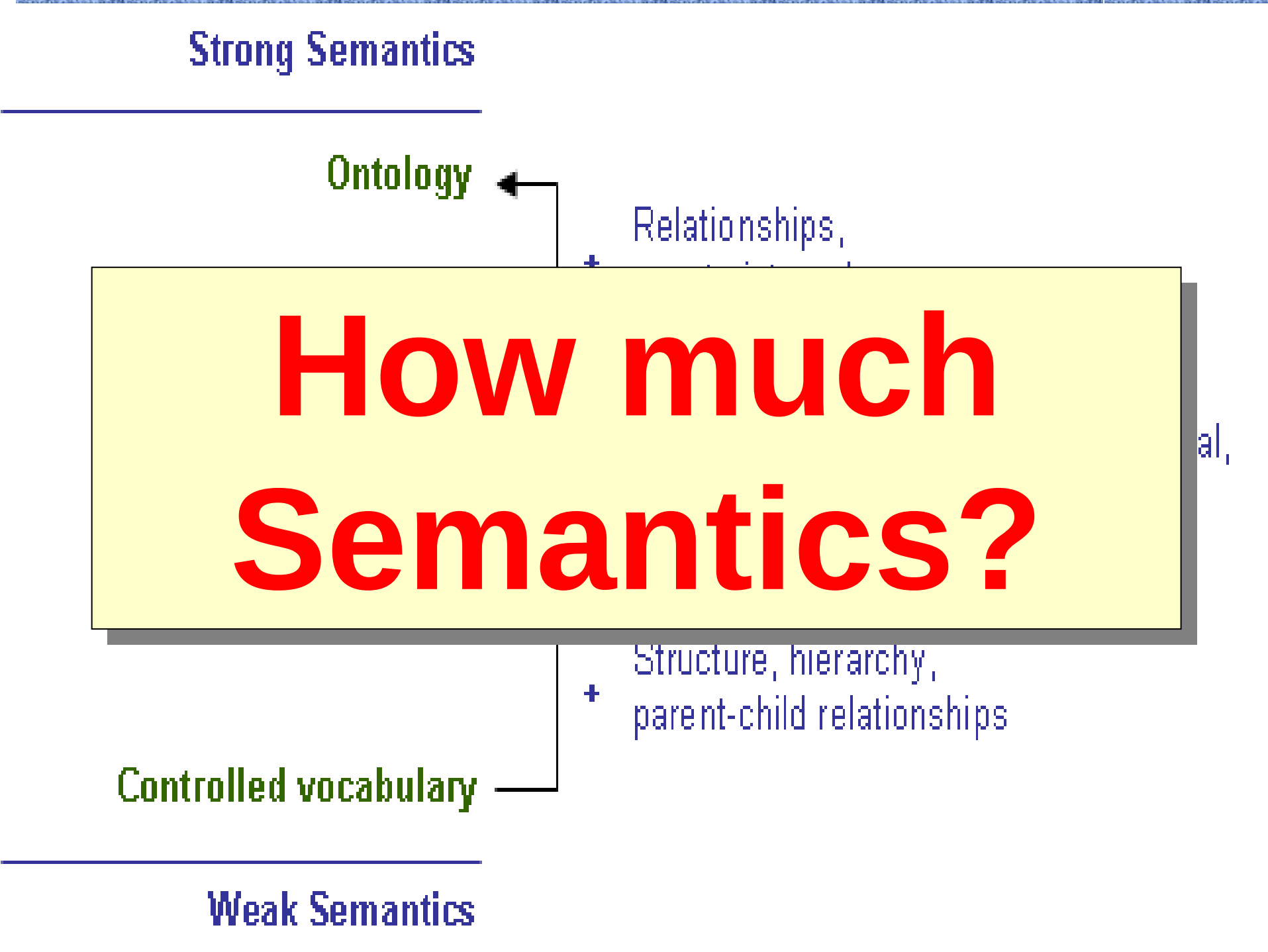


- **T**he study of relations between the system of signs (such as words, phrases, and sentences) and their meanings.
- Semantics <> Syntax
  - Semantics: what something means
  - Syntax: formal structure/patterns in which something is expressed

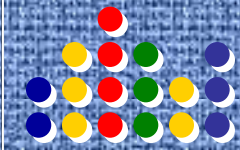
Pragmatics

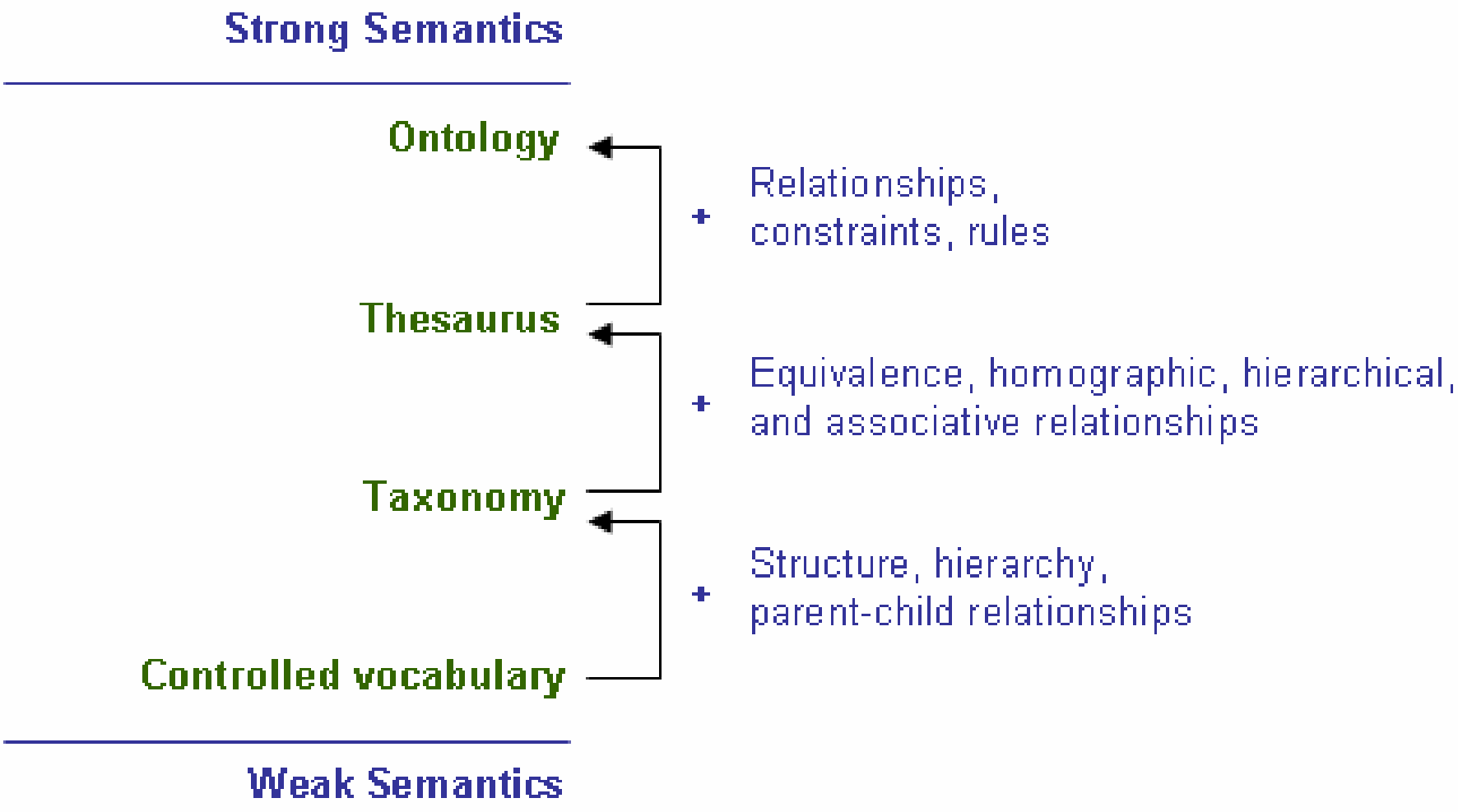


- The study of natural language understanding
  - Specifically the study of how context influences the interpretation of meaning.
- The context may include
  - Social, environmental, and psychological factors.
- Pragmatics <> Semantics
  - Pragmatics: origin, uses, and effects of signs within the content or context
  - Semantics: meaning of signs

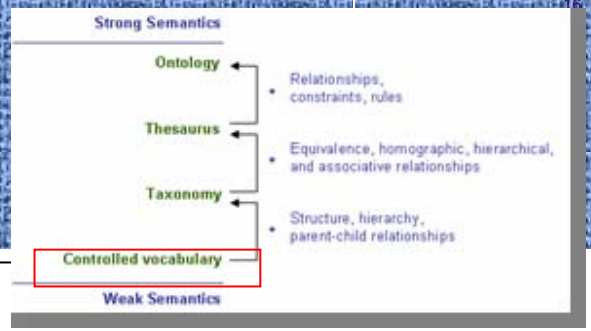

Strong Semantics
Ontology
Relationships,
al,
How much Semantics?
Structure, hierarchy,
parent-child relationships
Controlled vocabulary
Weak Semantics

Levels of semantics





Controlled vocabularies

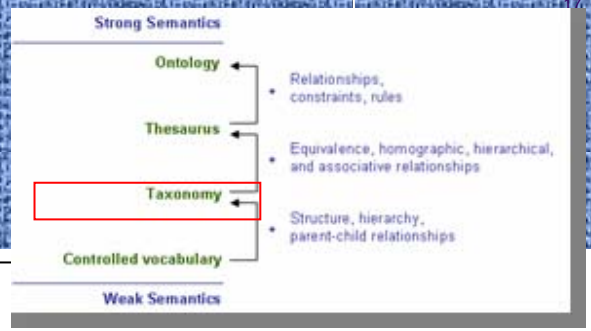


- The weaker end of the semantic spectrum
  - Is a list of terms (e.g., words, phrases, or notations)
  - Enumerated explicitly
  - Unambiguous, non-redundant
  - Limit choices to an agreed upon set of terms
- Objective
  - Prevent users from defining their own terms
  - User terms can be ambiguous, meaningless, or misspelled

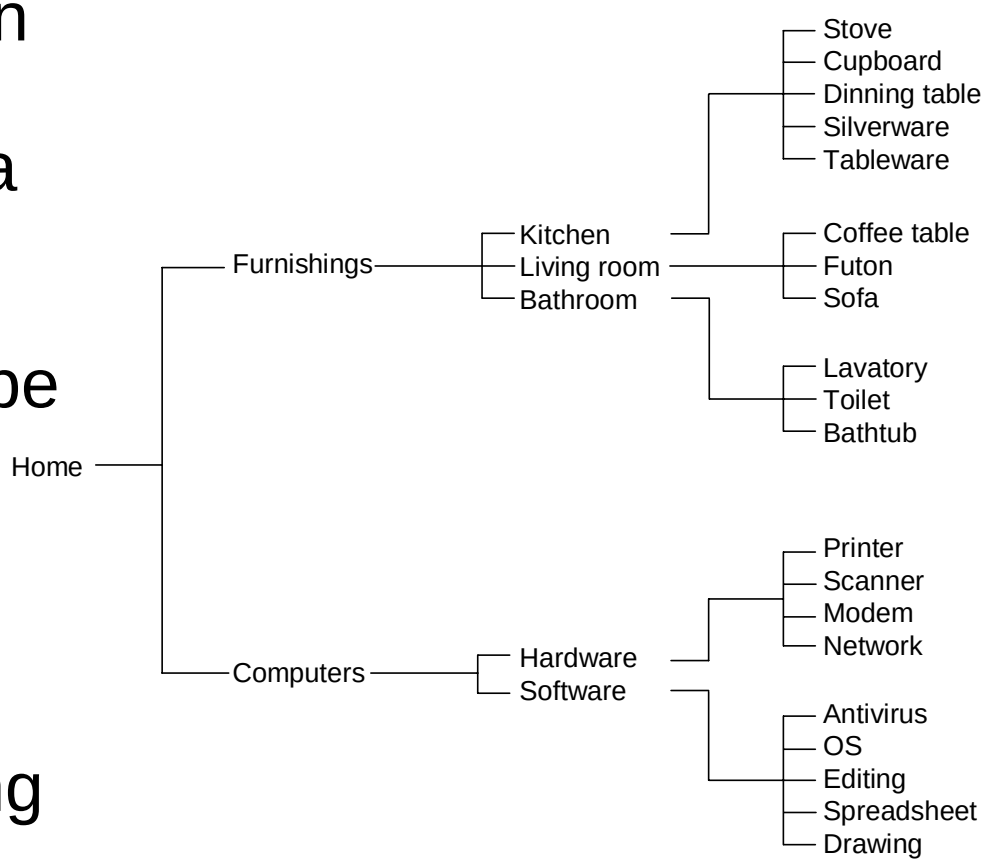
amazon.com

Uses controlled vocabulary to search for products.
Books, **Popular Music**, **Music Downloads**, **Classical Music**, **DVD**, **VHS**, **Apparel**, **Yellow Pages**, **Restaurants**, etc.

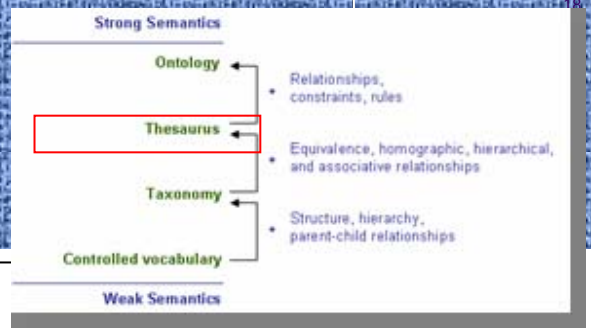
Taxonomy



- Subject-based classification
- Arranges the terms in a controlled vocabulary into a hierarchy without doing anything further
- Classifies terms in the shape of a hierarchy or tree.
  - Contains parent-child relationships
  - “is subclass of” and “is superclass of”.
- Describes a word by making explicit its relationship with other words

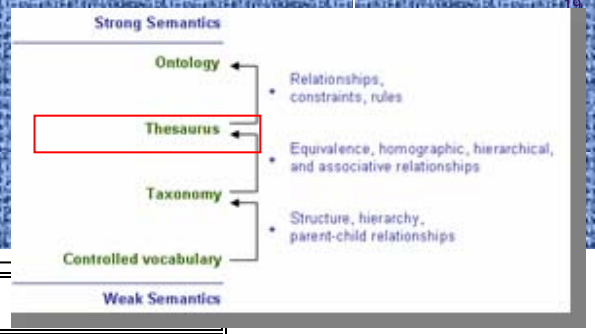


Thesaurus



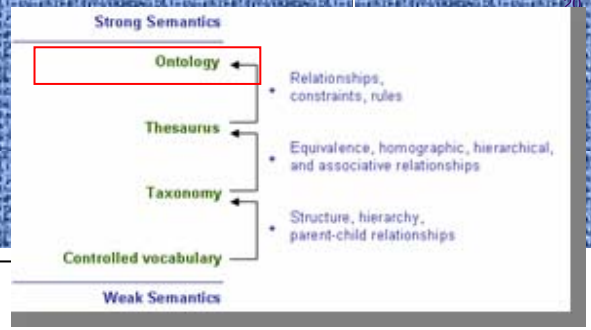
- A networked collection of controlled vocabulary terms with conceptual relationships between terms
- An extension of a taxonomy by allowing
  - Terms to be arranged in a hierarchy
  - Relationships to be made about the terms
- Types of relationships.
  - **Equivalence**.
    - Term t1 has the same or nearly the same meaning as a term t2.
  - **Homographic**.
    - Term t1 is spelled the same way as a term t2, but has a different meaning
  - **Hierarchical**.
    - Degrees or levels of “is subclass of” and “is superclass of” relationships.
  - **Associative**.
    - Link terms that are closely related in meaning semantically but not hierarchically. Ex: “is related to”, term t1 “is related to” term t2.

Thesaurus



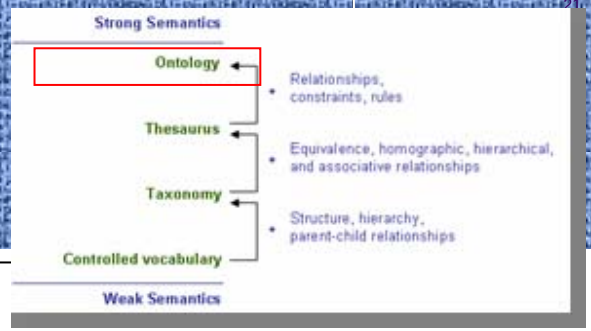
Relationship	Term
Used for	Grade point Average Scholastic Achievement School Achievement
Narrower than	Academic Overachievement Academic Underachievement College Academic Achievement Mathematics Achievement Reading Achievement Science Achievement
Broader than	Achievement
Related to	Academic Achievement Motivation Academic Achievement Prediction Academic Aptitude Academic Failure Academic Self Concept Education Educational Attainment Level School Graduation School Learning School Transition

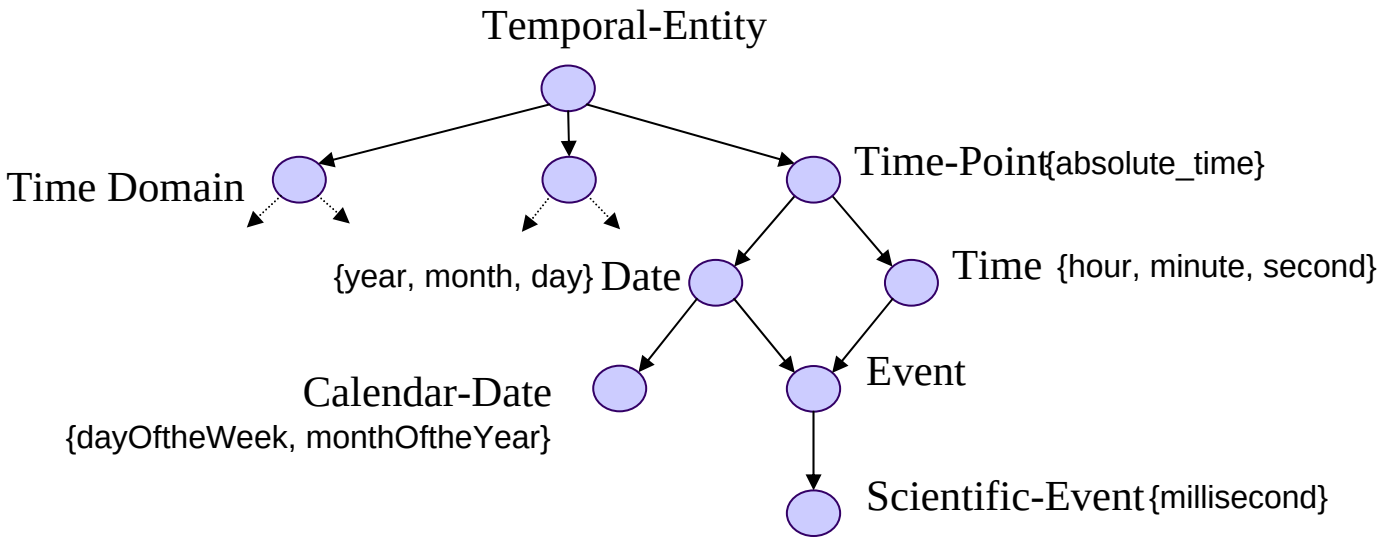
Ontologies



- Ontologies are similar to taxonomies but use richer semantic relationships among terms and attributes
  - Are a shared conceptualization of the world.
  - Provide a common understanding of a particular domain.
  - Consist of definitional aspects such as high-level schemas and assertional aspects such as entities, attributes, interrelationships between entities, domain vocabulary and factual knowledge – all connected in a semantic manner (Sheth 2003).
- Uses of ontologies:
  - Assist in communication between human beings
  - Achieve interoperability among software systems

Ontologies





- Create an **agreed-upon vocabulary** and semantic structure for exchanging information about a domain

Examples of Ontologies

Ontology Container Information

Title:	"The MGED Ontology"	Classes:	100
Creator:	Chris Stoeckert and Helen Parkinson	Properties:	82
Subject:	An ontology for microarray experiments	Individuals:	143
Description:	Concepts, definitions, terms, and resources for standardized description of a microarray experiment with an emphasis on study design, biomaterials and their treatment.	Axioms:	0
Date:	August 7 2002		
Version:	"1.6"		

Namespaces used

default

1

2

Classes

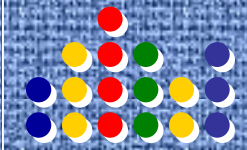
Age #2 Allele #2 Assay Atmosphere #2 BarrierFacility #2 Bedding #2 BehavioralStimulus #2 BibliographicReference #2 BiologicalFactor #1 BiomaterialDescription #2 BiomaterialManipulation #2 BiomaterialMeasurement #2 BiomaterialPreparation #2 BiomaterialState #2 Biosample #2 Biosource #2 BiosourceOntologyEntry #2 BiosourceProperty #2 BiosourceProvider #2 CellLine #2 CellType #2 ClinicTreatmentHistory #2 ClinicalHistory #2 ClinicalInformation #2 Compound #2 CompoundBasedTreatment #2 ConcentrationUnit #2 Contact #2 ContaminantOrganism #2 CurrentDiseaseHistory #2 DatabaseEntry #2 DensityRange #2 DevelopmentalStage #2 DiseaseState #2 DistanceUnit #2 EnvironmentalHistory #2 FamilyHistory #2 Gender #2 Gene #2 Generations #2 GeneticModification #2 GeneticVariation #2 Genotype #2 GrowthCondition #2 Haplotype #2 Hardware #2 HardwareVariation #1 Histology HistoryFactor #1 Host #2 Humidity #2 Individual #2 IndividualGenericCharacteristics #2 Infection #2 LabeledExtract #2 Light #2 MGEDOntology #1 MassUnit #2 Measurement #2 Medium #2 MethodologicalFactor #1 Modification #2 NormalizationDesign #1 Nutrients #2 OILED_DATATYPE #1 OntologyEntry #2 OperatorVariation Organism #2 OrganismPart #2 Organization #2 PastMedicalHistory #2 PathogenTests #2 Person #2 PhysicalCharacteristics Preservation #2 Protocol #2 ProtocolVariation #1 QualityControlDesign #1 QuantityUnit #2 ReplicateDesign #1 Resource #2 Sex #2 Software #2 SoftwareVariation #1 SomaticModification #2 Starvation #2 StrainOrLine #2 Study #1 StudyDesign #1 StudyFactor #1 TargetedCellType #2 Temperature #2 TemperatureUnit #2 TimeUnit #2 Treatment #2 URI #2 Unit #2 User #2 VolumeUnit #2 Water #2

Examples of Real Ontologies

MGED Ontology

- The MGED Ontology
  - Provide standard terms for the annotation of microarray experiments.
  - Terms will enable unambiguous descriptions of how the experiment was performed.
  - 212 classes, 101 properties.
- The MGED Ontology is being developed within the microarray community to provide consistent terminology for experiments.
- This community effort has resulted in a list of multiple resources for many species.
  - Approximately 50 other ontologies for different species
- The concepts are structured in DAML+OIL and available in other formats (rdfs)

The MGED Ontology is Structured in DAML+OIL using OILed 3.4



Ontology Container Information

Title:	"The MGED Ontology"
Creator:	Chris Stoeckert and Helen Parkinson
Subject:	An ontology for microarray experiments
Description:	Concepts, definitions, terms, and resource standardized description of a microarray experiment with an initial focus on study d biomaterials and their treatment.
Date:	August 7,2002
Version:	"1.5"

Namespaces used

default
file:/Applications/OilEd3.4 Folder/ontologies/MGEDontology
1
http://www.cbil.upenn.edu/Ontology/MGEDontology.daml#
2
http://www.cbil.upenn.edu/Ontology/MGEDontology.rdfs#

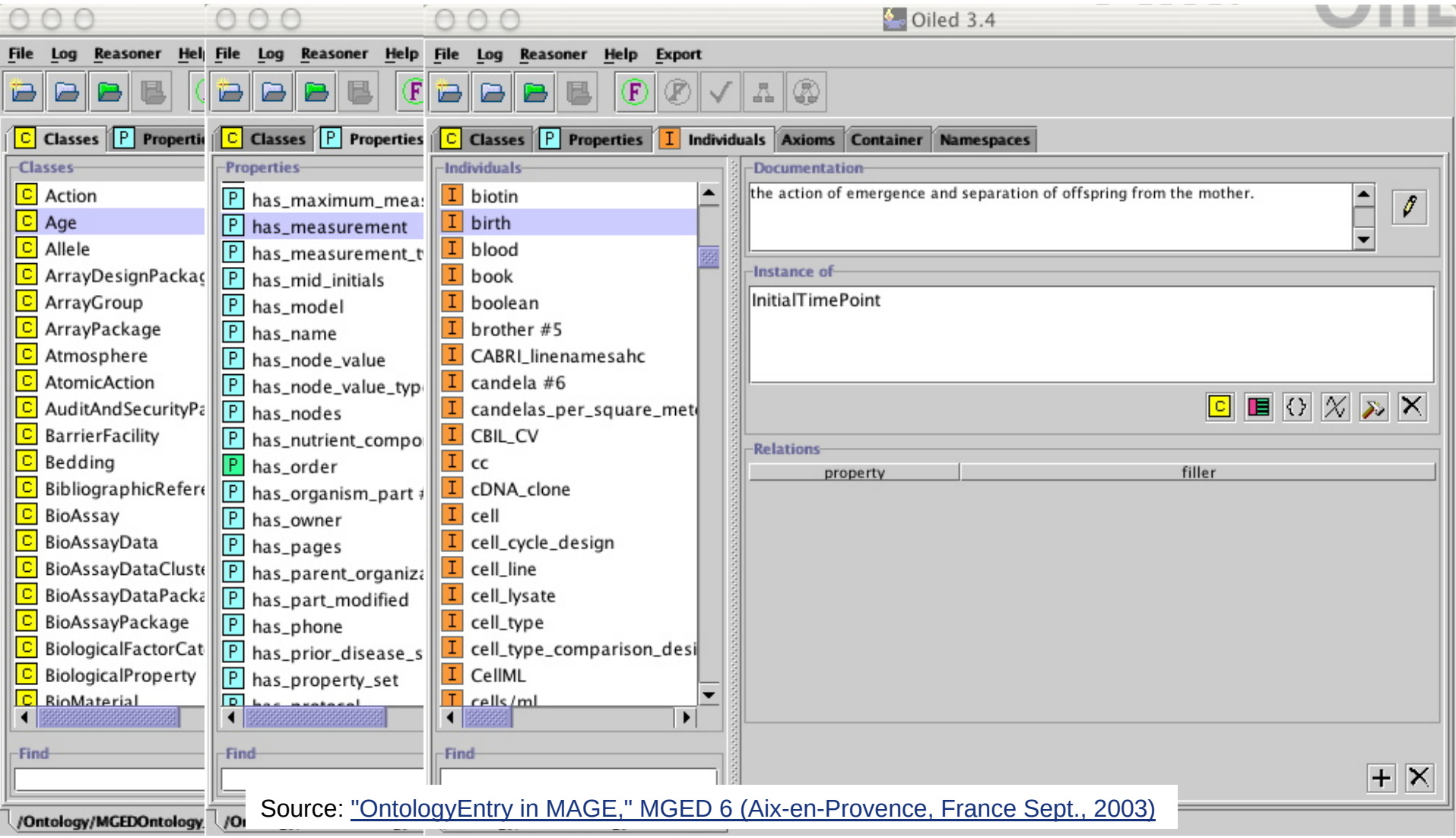
Classes

Age #2 Allele #2 Assay Atmosphere #2 BarrierFacility #2 Beddin
BiomaterialDescription #2 BiomaterialManipulation #2 Biomateria
Biosource #2 BiosourceOntologyEntry #2 BiosourceProperty #2
ClinicalHistory #2 ClinicalInformation #2 Compound #2 Compoun
CurrentDiseaseHistory #2 DatabaseEntry #2 DensityRange #2 D
FamilyHistory #2 Gender #2 Gene #2 Generations #2 GeneticM
Hardware #2 HardwareVariation #1 Histology HistoryFactor #1 H
LabeledExtract #2 Light #2 MGEDontology #1 MassUnit #2 Mea
NormalizationDesign #1 Nutrients #2 OILED_DATATYPE #1 Ont
PastMedicalHistory #2 PathogenTests #2 Person #2 PhysicalCh
QualityControlDesign #1 QuantityUnit #2 ReplicateDesign #1 Re

- BiosourceProperty
  - Age
  - BiosourceOntologyEntry
    - CellLine
    - CellType
    - ClinicalInformation
    - DevelopmentalStage
    - DiseaseState
    - GeneticVariation
    - IndividualGeneticCharacterstics
      - Allele
      - Genotype
      - Haplotype
    - Organism
    - OrganismPart
    - StrainOrLine
    - TargetedCellType
  - BiosourceProvider
  - Individual
  - Sex

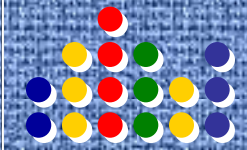
Source: "The MGED Ontology is an Experimental Ontology," 5th Annual Bio-Ontologies meeting (Edmonton, Canada Aug. 2002)

MGED Ontology consists of classes, properties, and individuals (instances)



Source: "OntologyEntry in MAGE," MGED 6 (Aix-en-Provence, France Sept., 2003)

MGED Ontology: BiomaterialDescription: BiosourceProperty: Age



class **Age #1**

namespace:
http://www.cbil.upenn.edu/Ontology/MGEDontology.rdfs#
documentation:
The time period elapsed since an identifiable point in the life cycle of an organism. If a developmental stage is specified, the identifiable point would be the beginning of that stage. Otherwise the identifia
type:
primitive
superclasses:
BiosourceProperty #1
constraints:
restriction initial_time_point #
restriction has_measurement
used in properties:
initial_time_point #1

class **Measurement #1**

namespace:
http://www.cbil.upenn.edu/Ontology/MGEDontology.rdfs#
documentation:
Measured values and units.
type:
primitive
superclasses:
MGEDontology #2
constraints:
restriction value #1 has-class thing
restriction has_units #1 has-class Unit #1
restriction measurement_type #1 has-class one-of (change #1 absolute #1)
known subclasses:
BiomaterialMeasurement #1
used in classes:
Age #1
BiomaterialMeasurement #1
BiomaterialPreparation #1
ClinicalHistory #1
CompoundBasedTreatment #1
GrowthCondition #1
used in properties:
measurement_type #1

Source: "The MGED Ontology is an Experimental Ontology," 5th Annual Bio-Ontologies meeting (Edmonton, Canada Aug. 2002)

Examples of Real Ontologies
OBO

- OBO (Open Biological Ontologies)
  - Is an umbrella organization for structured shared controlled vocabularies and ontologies for use within the genomics and proteomics domains.

The ontologies must be **open** and can be used by all without any constraint other than that their origin must be acknowledged and they can not be altered and redistributed under the same name.

The ontologies are in, or can be instantiated in, a **common shared syntax**. This may be either the GO syntax, extensions of this syntax, or OWL.

The ontologies share an **unique identifier space**.

The ontologies include textual **definitions** of their terms.

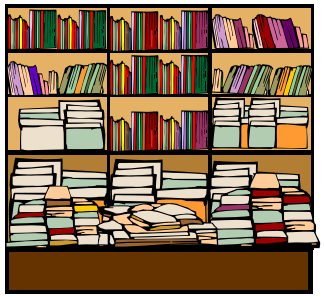
The ontologies are **orthogonal** to other ontologies already lodged with OBO.

Examples of Real Ontologies

GO Ontology

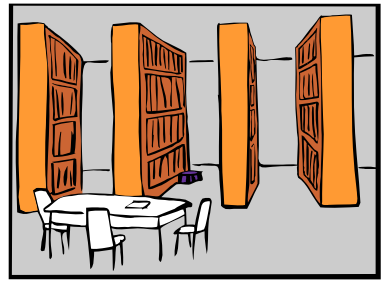
- Gene Ontology (**GO**)
  - Describes gene products in terms of their
    - Associated biological processes,
    - cellular components and
    - Molecular functions in a species-independent manner.

GO format - flat files, XML, MySQL



Component ontology
1379 terms
212 KB

Process ontology
8151 terms
4.82 MB



Function ontology
7278 terms
1.16 MB

function.ontology

<molecular_function ; GO:0003674

%antioxidant activity ; GO:0016209

%glutathione dehydrogenase (ascorbate) activity ; GO:0045174 ; EC:1.8.5.1 ; MetaCyc:1.8.5.1-RXN ; synonym:dehydroascorbate reductase % electron carrier activity ; GO:0009055 % glutathione disulfide oxidoreductase activity ; GO:0015038 % oxidoreductase activity\, acting on sulfur group of donors\, quinone or similar compound as acceptor ; GO:0016672

%glutathione-disulfide reductase activity ; GO:0004362 ; EC:1.8.1.7 ; MetaCyc:1.8.1.7-RXN ; MetaCyc:GLUTATHIONE-REDUCT-NADPH-RXN ; synonym:glutathione reductase (NADPH) activity ; synonym:glutathione-disulphide reductase activity % electron transporter activity ; GO:0005489 % glutathione disulfide oxidoreductase activity ; GO:0015038 % oxidoreductase activity\, acting on NADH or NADPH\, disulfide as acceptor ; GO:0016654

%peroxidase activity ; GO:0004601, GO:0016685, GO:0016686, GO:0016687 ; EC:1.11.1.7 ; MetaCyc:PEROXID-RXN ; synonym:eosinophil peroxidase activity ; synonym:lactoperoxidase activity ; synonym:myeloperoxidase activity % oxidoreductase activity\, acting on peroxide as acceptor ; GO:0016684

%thioredoxin-disulfide reductase activity ; GO:0004791 ; EC:1.8.1.9 ; MetaCyc:1.8.1.9-RXN ; MetaCyc:THIOREDOXIN-REDUCT-NADPH-RXN ; synonym:thioredoxin disulfide reductase activity ; synonym:thioredoxin reductase (NADPH) activity ; synonym:thioredoxin-disulphide reductase activity % electron transporter activity ; GO:0005489 % oxidoreductase activity\, acting on NADH or NADPH\, disulfide as acceptor ; GO:0016654