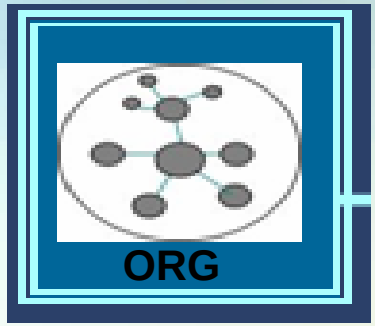
Ontology Research Group Overview



Dr. Valerie Cross

Sriram Ramakrishnan **En Yu**
Ramanathan Somasundaram **Yi Sun**

Miami University

OCWIC'2007
February 17,
Deer Creek Resort

Outline

- Motivation for Research Agenda
  - Semantic Web
  - What are Ontologies?
  - Some Ontologies Used in our Research
- Overview of Current Ontology Research Areas
  - Semantic Similarity in Ontologies
  - Ontology Evaluation
  - Ontology Learning
  - Ontology Querying
  - Ontology 3D Visualization
- Possibilities …

Building the Semantic Web

THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new abilities

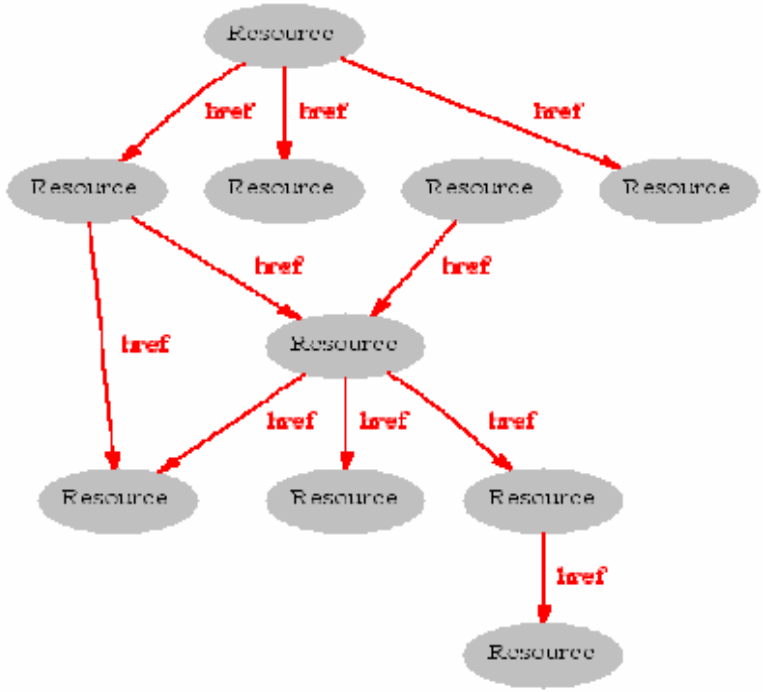
by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTO CREDIT HERE

Scientific America, May 2001

http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

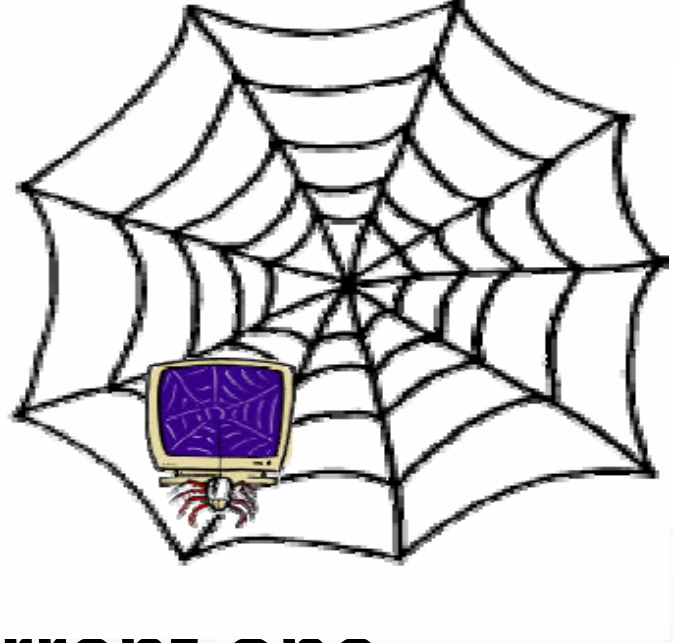
Today's "Syntactic Web"



- A place where computers do the presentation (easy) and people do the linking and interpreting (hard)
- Why not get computers to do more of the hard work?

Ian Horrocks, ICCL Summer School Course 2005

The *Semantic Web*



is an extension of the current one,

in which information is given well-defined *meaning*,

better enabling computers and people to work in cooperation.

T. Berners-Lee, J. Hendler, O. Lassila

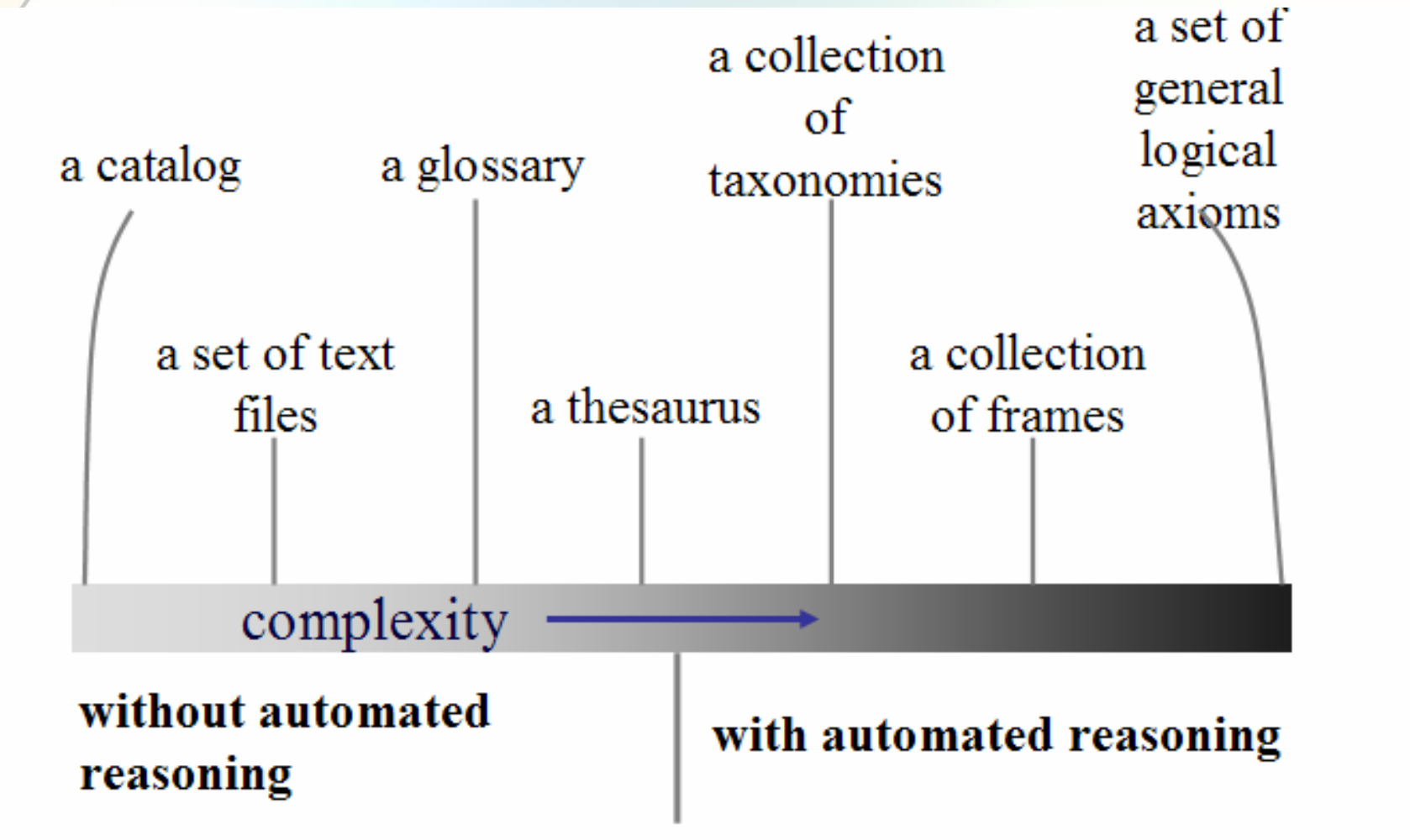
Ontology in Computer Science

- An ontology is an engineering artefact consisting of:
  - A vocabulary used to describe (a particular view of) some domain
  - An explicit specification of the intended meaning of the vocabulary.
    - almost always includes how concepts should be classified
  - Constraints capturing additional knowledge about the domain
- Ideally, an ontology should:
  - Capture a shared understanding of a domain of interest
  - Provide a formal and machine manipulable model of the domain

Ian Horrocks, ICCL Summer School Course 2005

Ontology Dimensions

[Welty, Uschold, Gruninger, Lehmann & McGuinness, 1999]



Where are ontologies used?

- e-Science, e.g., Bioinformatics
  - The Gene Ontology
  - The Protein Ontology (MGED)
  - “in silico” investigations relating theory and data
- Medicine
  - Terminologies
- Databases
  - Integration
  - Query answering
- User interfaces
- Linguistics
- The Semantic Web

Ian Horrocks, ICCL Summer School Course 2005

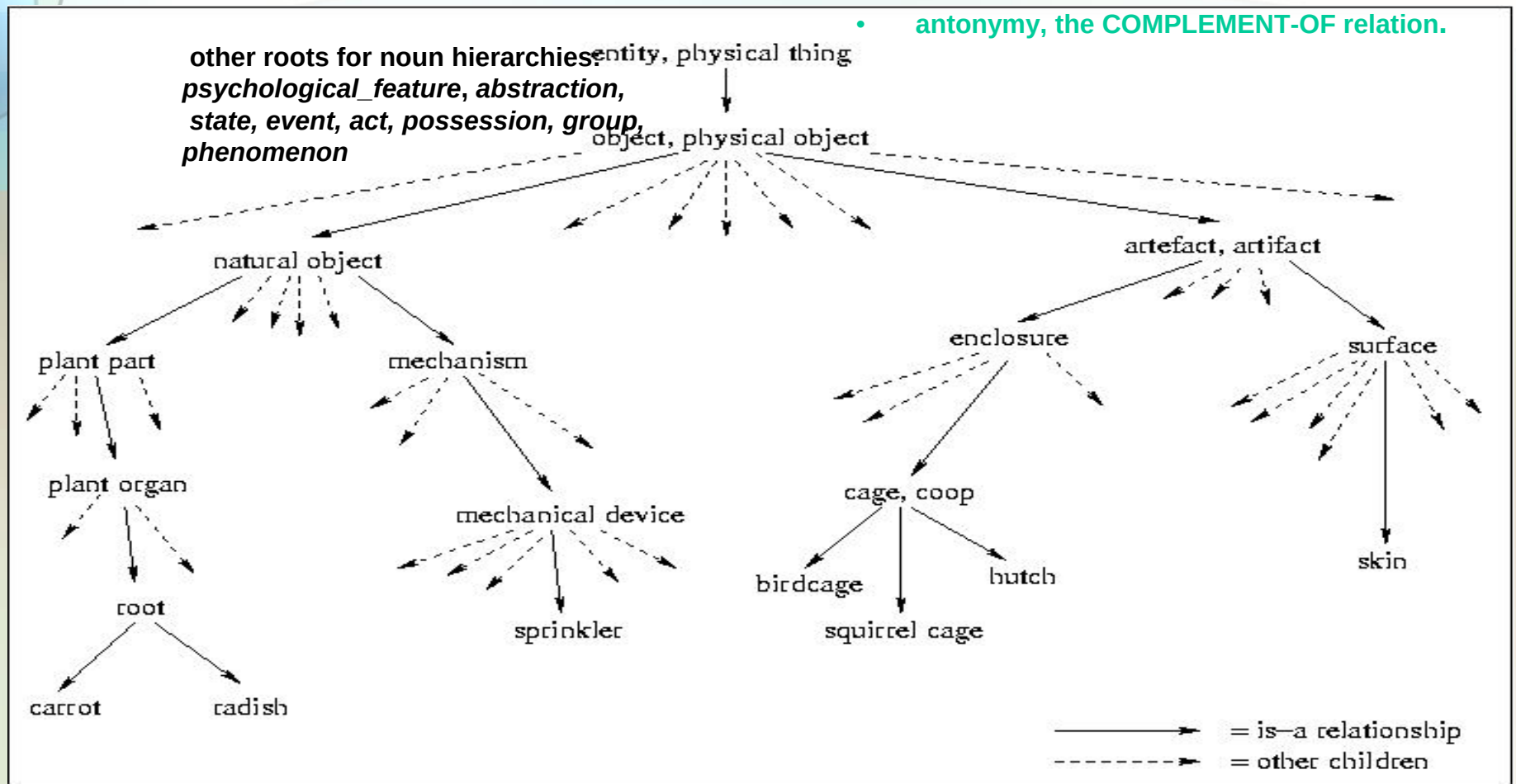
Current Ontologies Used in Our Research

- WordNet
- UMLS
- Gene Ontology
- SEURAT's Argument Ontology
- PSCS Ontologies
  - UNSPSC
  - ecl@ass

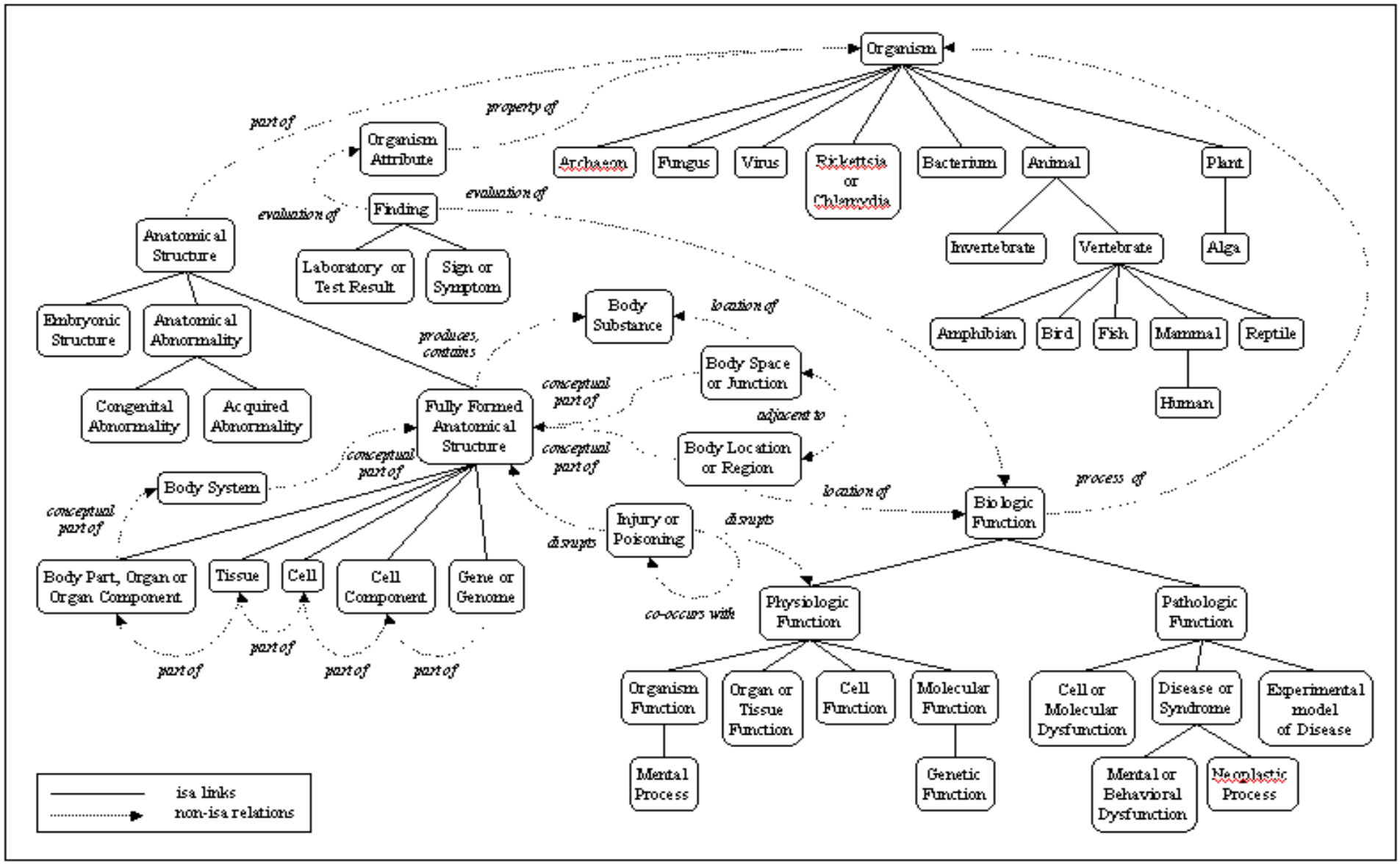
WordNet – *Is a* Hierarchy

- Four parts of speech – nouns, verbs, adjectives and adverbs.
- Concept Node = Synonyms + Definition (gloss)
- ~ 111,400 concepts.
- ~ 13 types of relationships.
- 9 noun "*is a*" hierarchies with an average depth of 13.
- 628 verb *is a* hierarchies with an average depth of 2.

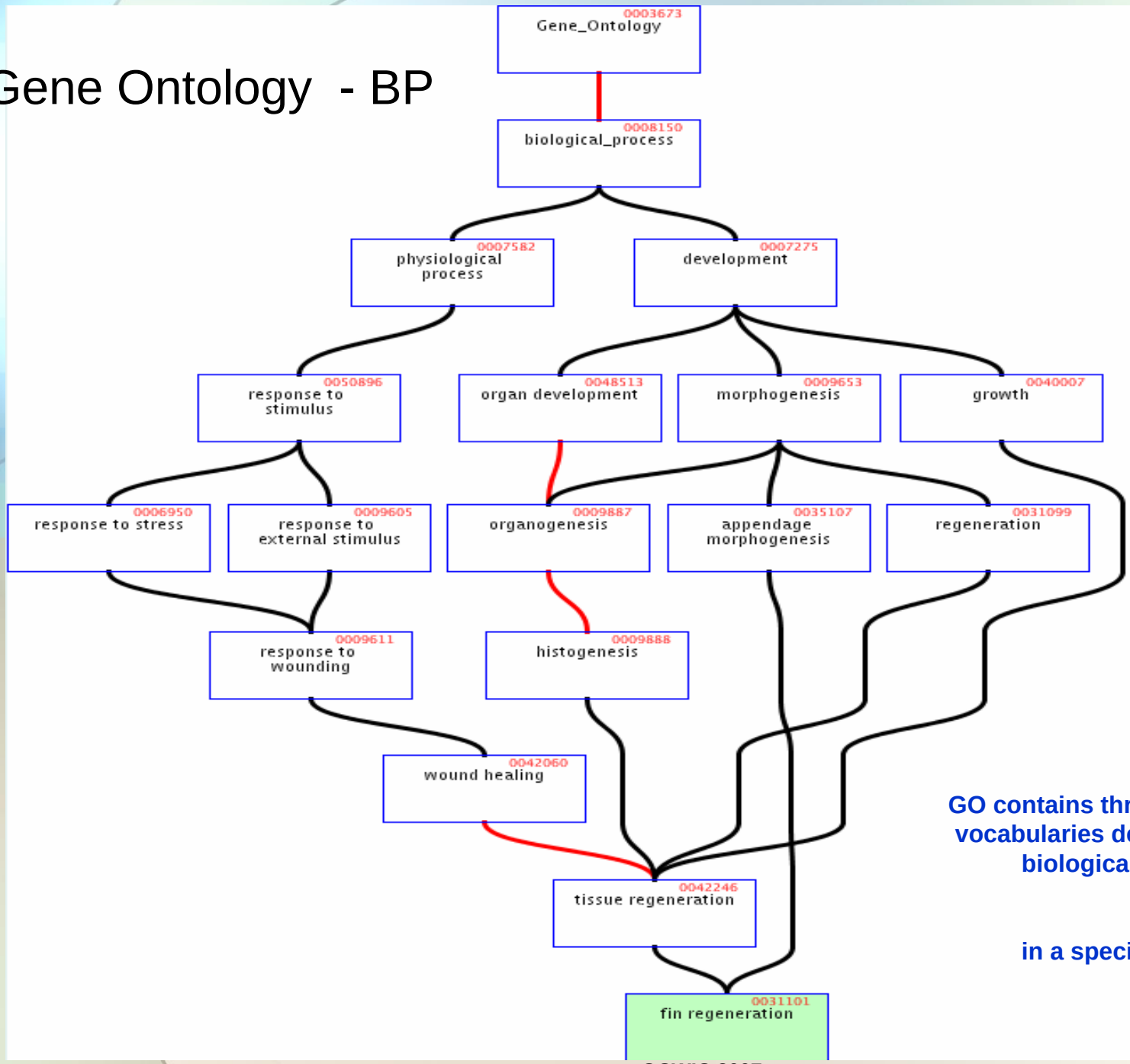
- synonymy relation implicit in each node
  - (through synsets)
- hyponymy (IS-A) relation, its inverse, hypernymy;
- six meronymic (PART-OF) relations (and inverses):
  - COMPONENT-OF,
  - MEMBER-OF
  - SUBSTANCE-OF
- antonymy, the COMPLEMENT-OF relation.



UMLS Semantic Network



Gene Ontology - BP



GO contains three structured, controlled vocabularies describing gene products
biological processes (BP),
cellular components (C
molecular functions (M
in a species-independent manner.

Semantic Distance, Similarity, Relatedness

- **Semantic relatedness**
  - General term involving many relationships
    - car-wheel (meronymy)
    - hot-cold (antonymy)
    - pencil-paper (functional)
    - penguin-Antarctica (association)
    - vehicle-car (subsumption)
- **Semantic similarity**
  - More specific term involving likeness
    - bank-trust company (synonymy)
- **Distance**
  - Inverse of either one
    - reldist(x)=semantic relatedness^{-1}(x)
    - simdist(x)=semantic similarity^{-1}(x)

Applications of Semantic Relatedness

- Interoperability between Ontologies
  - Merging, Linking, agent communication
- Information Retrieval
  - To improve automated assignment of indexing based descriptors
- Semantic Vocabulary Integration
  - Choose closest related concepts while translating in and out of the multiple vocabularies/ontologies
  - Semantic Vocabulary Interoperability Project at
    - http://cgsb2.nlm.nih.gov/~kashyap/projects/SVIP

Semantic Similarity Research
(Youbo Wang)

- Comparison of semantic similarity measures in literature using experimental approach for different domain ontologies and types of ontologies
- Create a testbed Acquire different domain ontologies and different types of ontologies (WordNet and UMLS)
  - Develop plug-ins for Protégé for all the existing semantic relatedness measures
  - Develop comparison with human judgment for new domain ontologies such as medical

Motivation for Ontology Evaluation

- Ontologies the “backbone of the Semantic Web”
- Development and deployment of ontology-based software solutions requires considerable time and effort
- Numerous existing ontologies in libraries available on the WWW
- Why reinvent the wheel? Reuse of ontologies important to SW success

What is ontology evaluation?

- **Ontology evaluation - key problem in the field of ontology development and reuse.**
- **Selection vs. Evaluation**
  - **Two separate tasks?**
  - **How related?**
  - **When does it occur?**
  - **Selection → Evaluation? Evaluation→Selection?**
  - ***Ontology Selection: Ontology Evaluation on the Real Semantic Web***

    (Sabou, Lopez, Motta,Uren EON 2006)

Ontology Consumer Analysis Tool

- Objective is to examine from Consumer perspective to
  - Reuse or adapt ontology
- Suite of metrics from
  - conceptual modeling,
  - software development,
  - information systems development
  - information retrieval
- Used on WordNet, UMLS, UNSPSC and ecl@ss

OntoCAT
(Anindita Pal)

- plug-in for OWL Protégé
- very parameterized
  - Intensional (classes) and extensional (instances)
  - View metrics interested in
    - Size
    - Structure
  - User selectable root for analysis
  - Hub concept analysis
  - User selectable relation for establishing extensional structure

OntoCAT intensional metrics UNSPSC

Size	owl:Thing	Apparel_and_ Luggage_and_ Personal_Care_ Products	Building_and_ Construction_and_ Maintenance_ Services
Total #Cls [iCnt(C)]	16500	254	87
Total #Property [iCnt(P)]	2	2	2
Total Roots [iCnt(Roots)]	56	5	1
		Clothing, Footwear, Luggage_and_handbags _ and_ packs_and_cases, Personal_ care_products, Sewing_supplies _and_ accessories	Building_ construction_ and_ support_ and_ maintenance_ and_ repair_ services
Total Leaves [iCnt(leaves)]	14317	219	70
Average Leaves [iAv(leaves)]	0.86	0.86	0.80
Max Depth [iMaxDepth]	4	3	3
Max Width [iMaxWidth]	14317	219	70
Depth of iMaxWidth	4	3	3
Average Width [iAvWidth]	4125	84.66	29

RDF/RDF(S) version of UNSPSC is developed by Michel Klein and was obtained from http://www.cs.vu.nl/~mcaklein/unspsc

Ontology Learning
(Vishal Bathija)

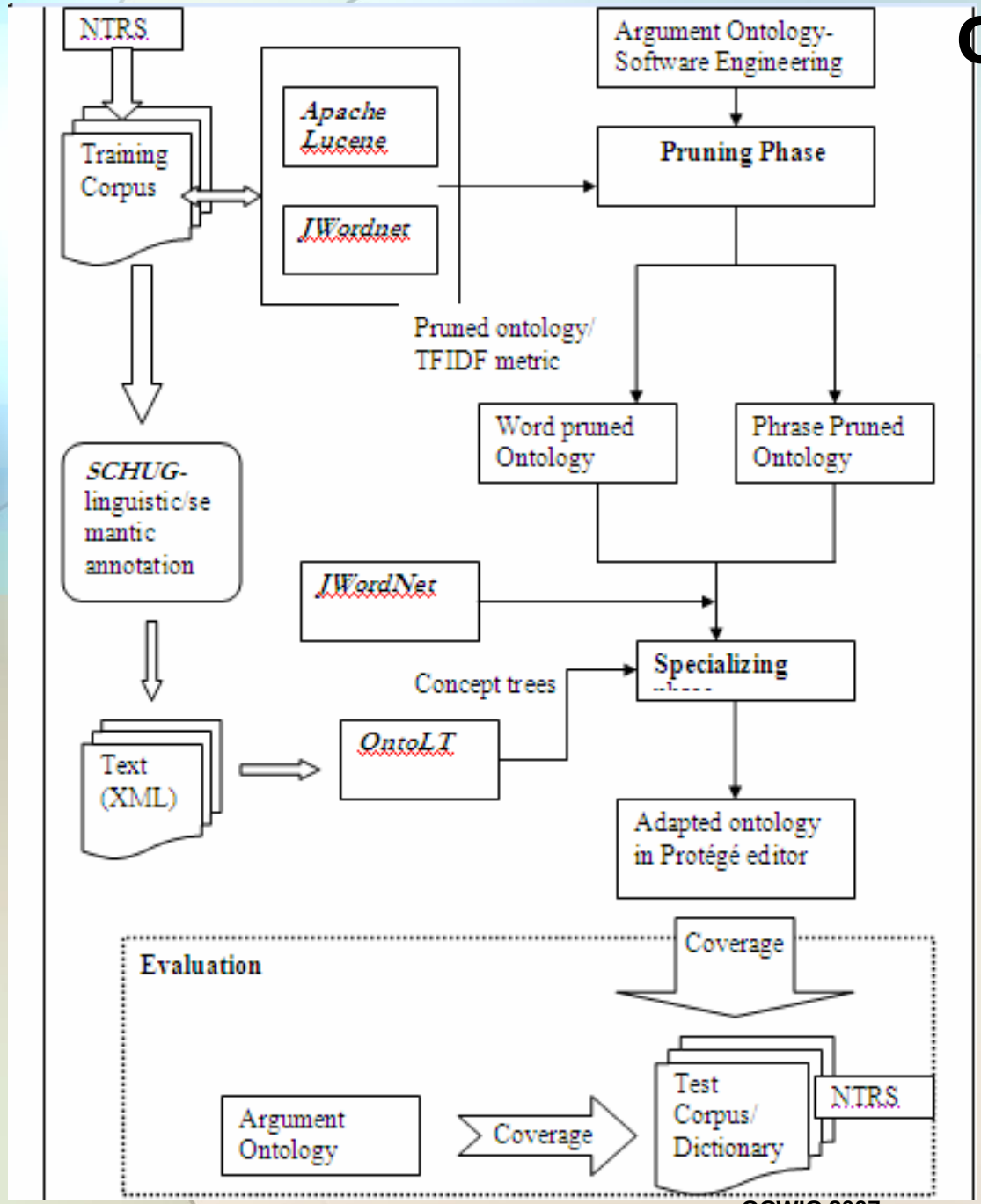
- **Objective: Adapt the SEURAT Argument Ontology from the Software Engineering Domain to the Spacecraft Engineering Domain.**
- **Developed an Ontology Adaptation Methodology**
  - **Pruning:**

    Remove from the existing Argument Ontology concepts not relevant to Engineering Design (ED) by analyzing text corpuses specific to domain
  - **Adapting and Specializing:**

    Adding new concepts and relations (is-a only) to transform pruned ontology into the ED criteria ontology.
  - **Evaluation**

    Evaluate the adapted ontology based on performance measures used in ontology learning research literature

Ontology Adaptation Software Architecture



Input Resources:
- Argument Ontology
- Training Corpus
- Test Corpus
- General Corpus

Software Resources:
- Apache Lucene
- JWordNet
- Test Corpus
- General Corpus

Ontology Querying
(Yi Sun)

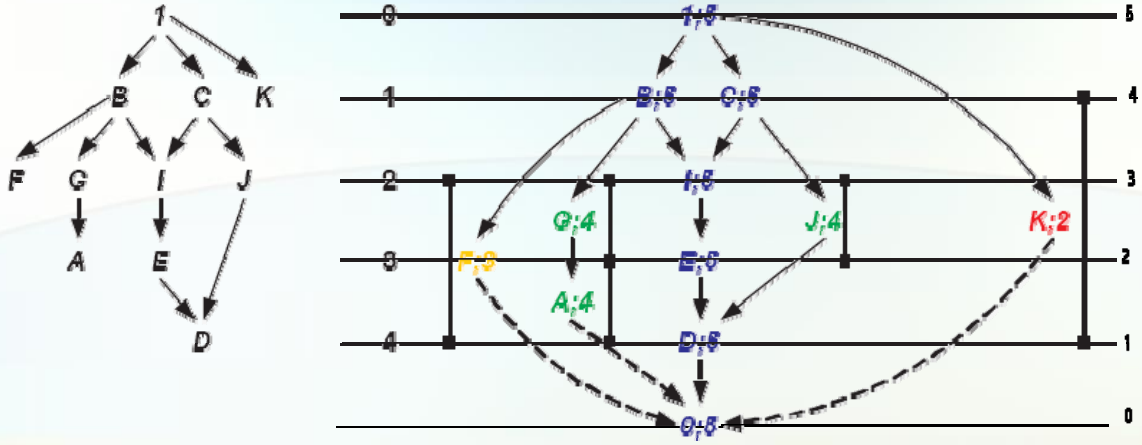
QUOTA

- **QUerying with Ontological Terminologies and their Annotations uses previous research on:**
  - **Semantic Similarity**
  - **Ontology metrics**
  - **Gene Ontology**
  - **Annotations from model organism databases.**

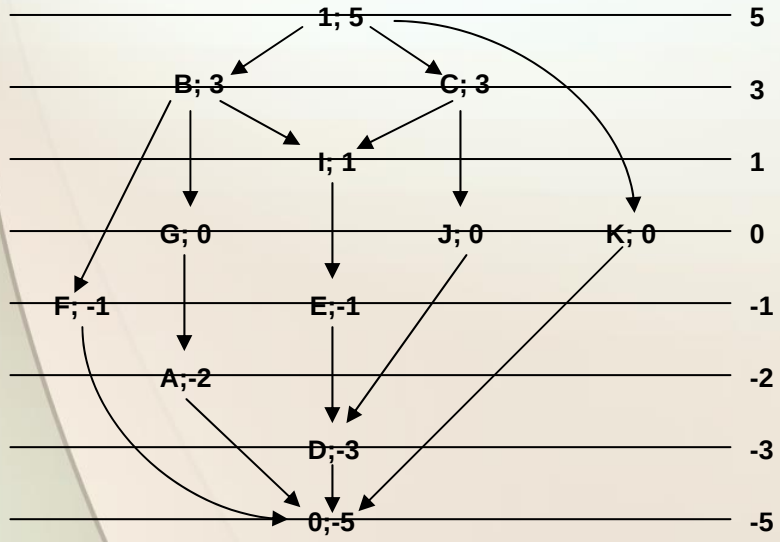
Ontology 3D Visualization
(Ramanathan Somasundaram)

- **Objective: build a package of 3D visualization algorithms to produce visualizations of different types of ontologies for different purposes**
  - **Experiment with different weight functions**
  - **Apply the weight functions to different force directed layout algorithms**
  - **Designing filtering view for different tasks**
  - **Parallelizing the execution**

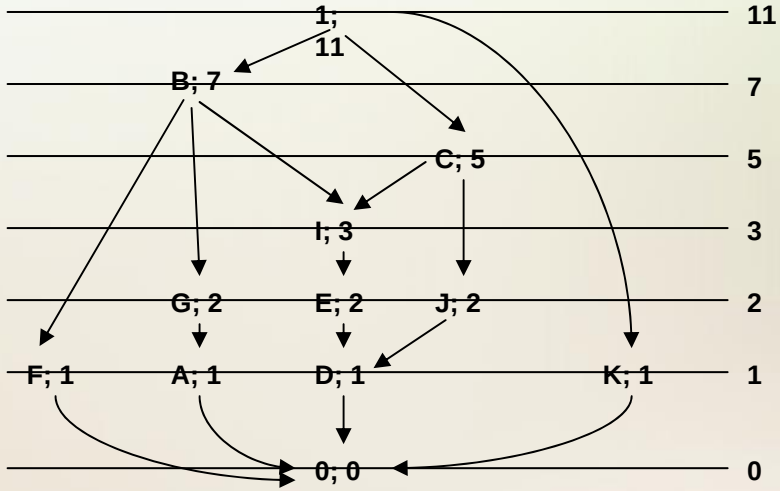
Visualization using weight function



Interval rank (SpindleViz)



Balanced Height Weight Function, $W_{BH}(C)$



|extent (c)| Weight Function, $W_{NS}(C)$

Research Possibilities….

Semantic Similarity

- Develop a task/application-specific evaluation procedure
- Continue with mathematical analysis and properties
- Apply to 3D visualziation
- Investigate semantic relatedness measures in assessing ontology quality

Research Possibilities

- **Ontology Evaluation**
  - **Interface with “candidate” selection approaches before perform detailed analysis**
  - **Comparison metrics/charts/visualization for multiple ontologies for “candidates”**
  - **Visualization to help consumers “see” ontology for reuse and comparison**
    - **Hubs visualization Improvement**
    - **Individual hub visualization**
    - **Top-level summary / visualization**
    - **Bottom-up level summary / visualization**
  - **Combine and aggregate analysis results to provide consumers with summaries characterizing each ontology and a vocabulary for “critiquing” an ontology after all the metric analysis**

Research Possibilities

- **Ontology Learning**
  - **More experimentation with OntoLT user-specified rules to produce concept trees**
  - **Use of DR sublanguage to parse the training text collections looking for keywords of the domain sublanguage**
  - **Limitation of Ontology adaptation software**
    - **Dependency on SCHUG**
    - **Find other open source software or develop software to provide the SCHUG functionality**
  - **Domain expert user interface to provide input to the pruning and merging process**

Interested?

- Send me email: crossv@muohio.edu
- Be happy to send some references
- Consider applying to Miami Unversity's masters program in computer science
- Most graduate students fully funded as a first year teaching assistant and a second year research assistant
- Pick up Miami folder at OCWIC if you don't have one.


Thank You