Ontology Maintenance Support

Text, Tools, and Theories

Chris Welty
IBM Research

Outline

- Opening joke
- Motivation
- Maintenance
- Support
  - Tools
  - Theories
  - Text Analysis

Motivation

- Given: Ontologies matter
  - Does quality matter?

Does quality matter?

- Good quality ontologies cost more
  - Coverage, correctness, richness, commitment [Kashyap, 2003]
  - Organization, meta-level consistency [Guarino & Welty, 2000] [Rector, 2002]
  - **Required** for some applications
- Improvements in quality can improve performance [Welty, et al, 2004]
  - 18% *f*-improvement in search
  - Cleanup cost ~1mw/3000 classes
  - BUT ... low quality ontology still improved base

Motivation

- Given: Ontologies matter
  - Does quality matter? **Sometimes**
- Problem: How to create them
- Bigger problem: how to *maintain* them
  - From SE: 80% of the cost is maintenance [Schrobe, 1996]

Software Maintenance

- Fixing Bugs
- Testing
- Enhancing

Ontology Maintenance

- Fixing Bugs
  - Inconsistent
  - Inaccurate
  - Inefficient
- Testing
  - Regression tests
  - Test Suites
  - Meta tag sets for test content
  - Ablation tests
- Enhancing
  - Tweaking
    - Richness
    - Correctness
    - Organization
    - Meta-level consistency
    - Efficiency
  - Extending
    - Improving coverage
    - Extending commitment
    - Integration
  - Refactoring

A looming problem

- Prediction
  - Ontology maintenance will become **the** significant problem as ontologies become more mainstream
  - Will follow the SE model (80% of cost)
- Observation/Conjecture
  - High quality ontologies are easier to maintain

Tool Support

- Hierarchical view of classes
- Hierarchical view of properties
- Consistency Reasoning
  - But….no “segmentation faults”
- Inferential Reasoning
- View non-tree taxonomies
- View relations between classes
- Global axioms
- View meta-level
- Basic Upper-level Theories
  - Space, Time, Parts, …
- Assistance for integration

Theory Support

- Meta-level analysis
  - OntoClean [Guarino & Welty, 2000]
- Good organizing principles
  - R-Normalization [Rector, 2002]
- Well-founded upper levels
  - Dolce [Gangemi, et al., 2003]
  - DAML-Time [Hobbs, 2003]
  - RCC [Randell, Cui & Cohn, 1992]

OntoClean

- Draw ***fundamental notions*** from Formal Ontology
- Establish a set of useful ***meta-properties***, based on behavior wrt above notions
- Explore the way these meta-properties combine to form relevant ***property kinds***
- Explore the ***taxonomic constraints*** imposed by these property kinds
  - Expose common modeling pitfalls

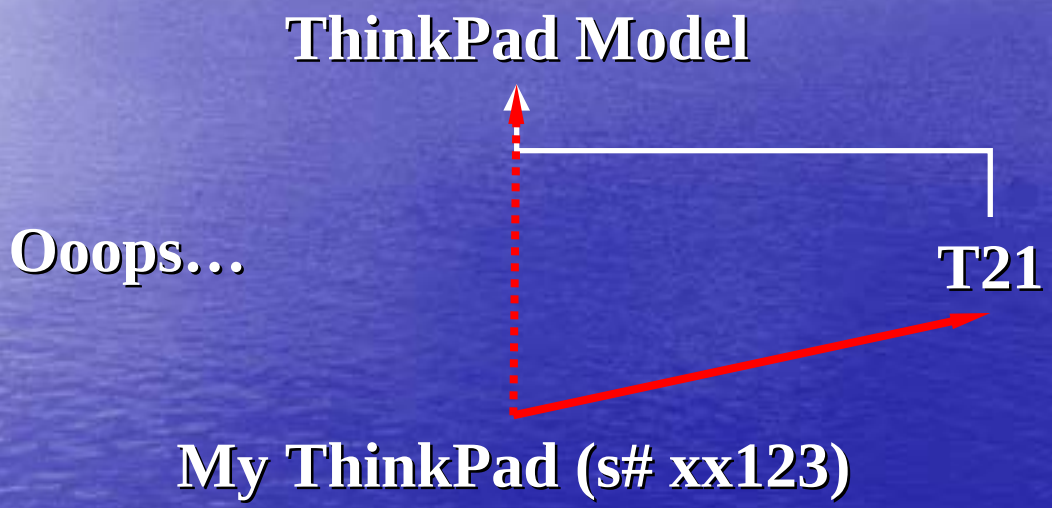
Overloading Subsumption

Common modeling pitfalls

- Instantiation
- Constitution
- Composition
- Disjunction
- Polysemy

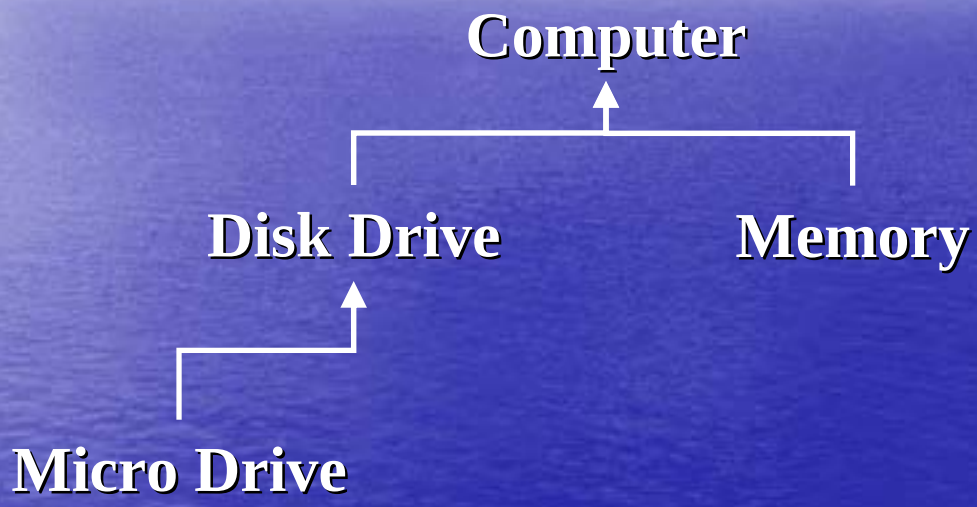
Instantiation

Does this ontology mean that **My ThinkPad** ***is a*** **ThinkPad Model?**



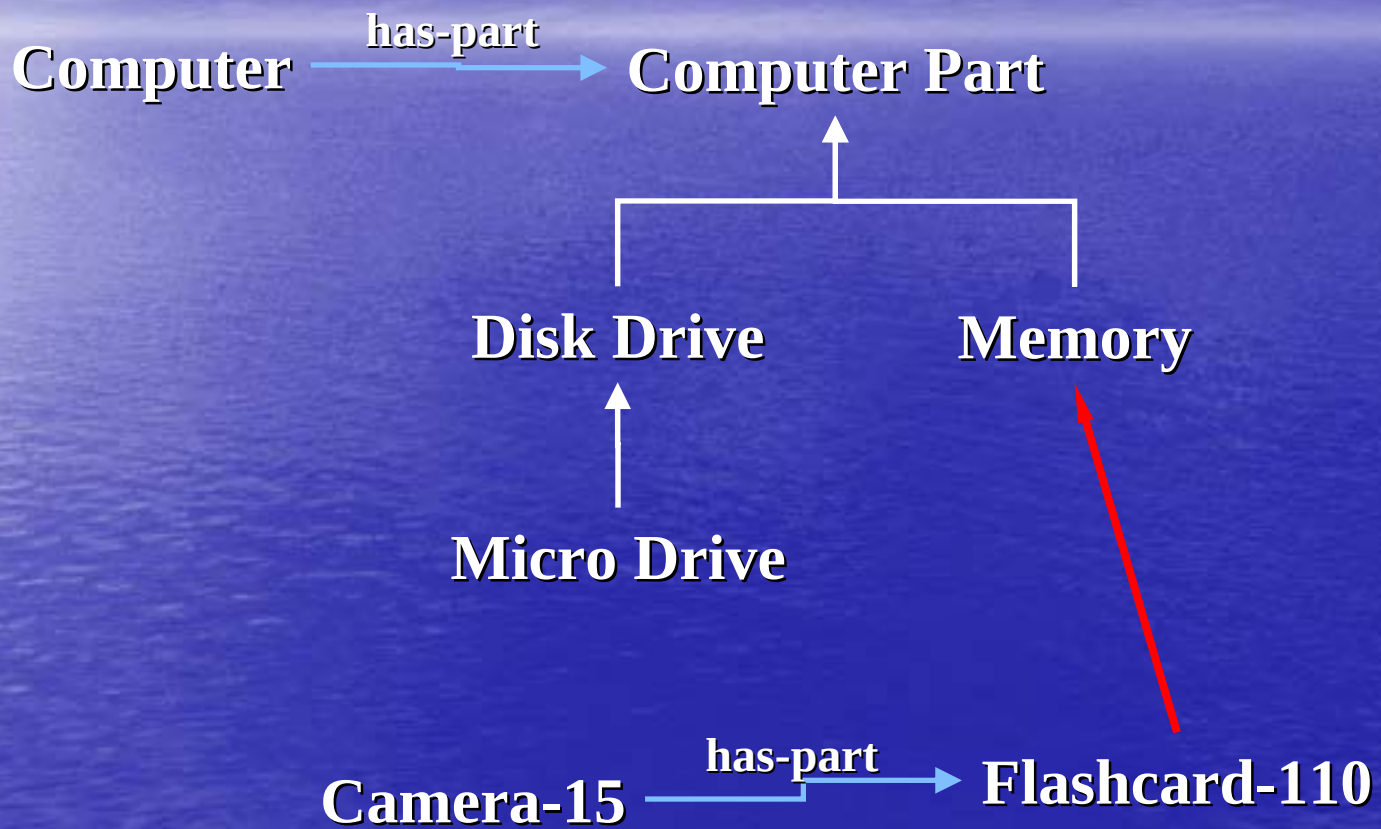
Question: What ThinkPad models do you sell?
Answer should NOT include My ThinkPad -- nor yours.

Composition



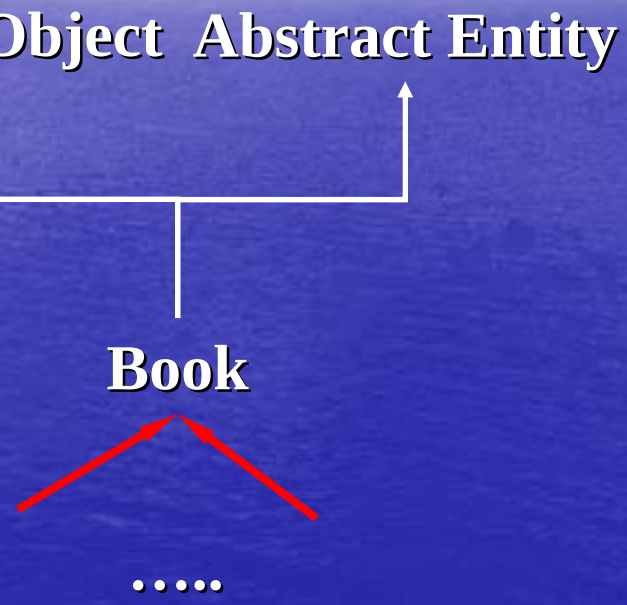
Question: What Computers do you sell?
Answer should NOT include Disk Drives or Memory.

Disjunction



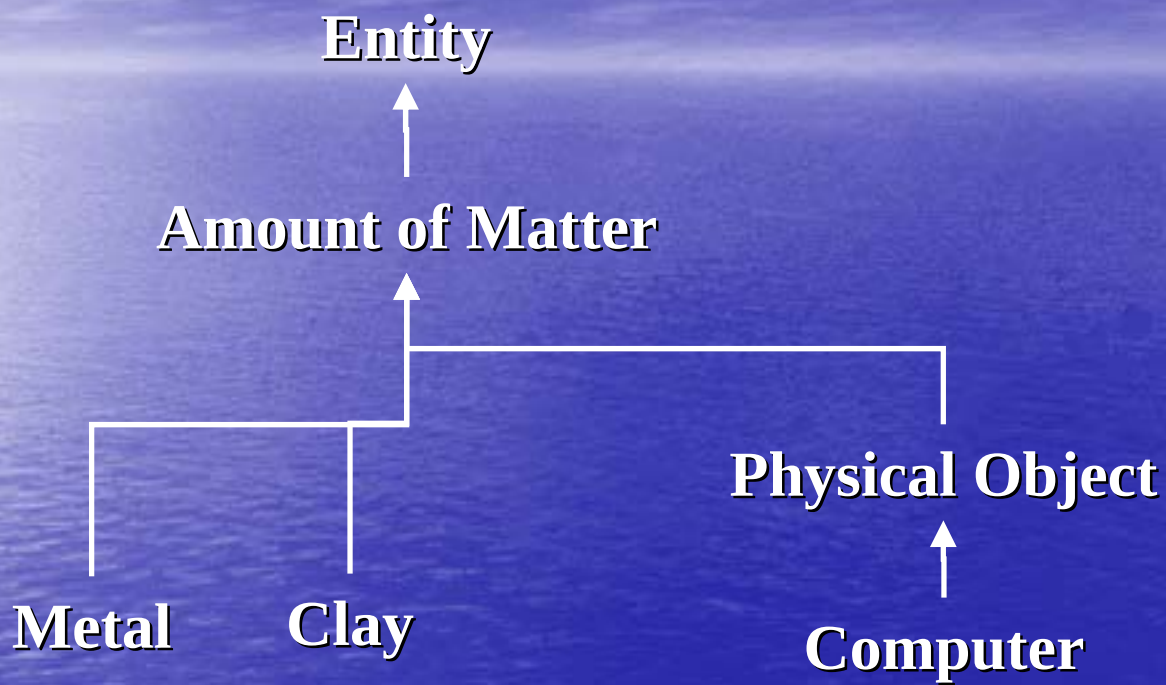
Unintended model: flashcard-110 is a computer-part

Polysemy

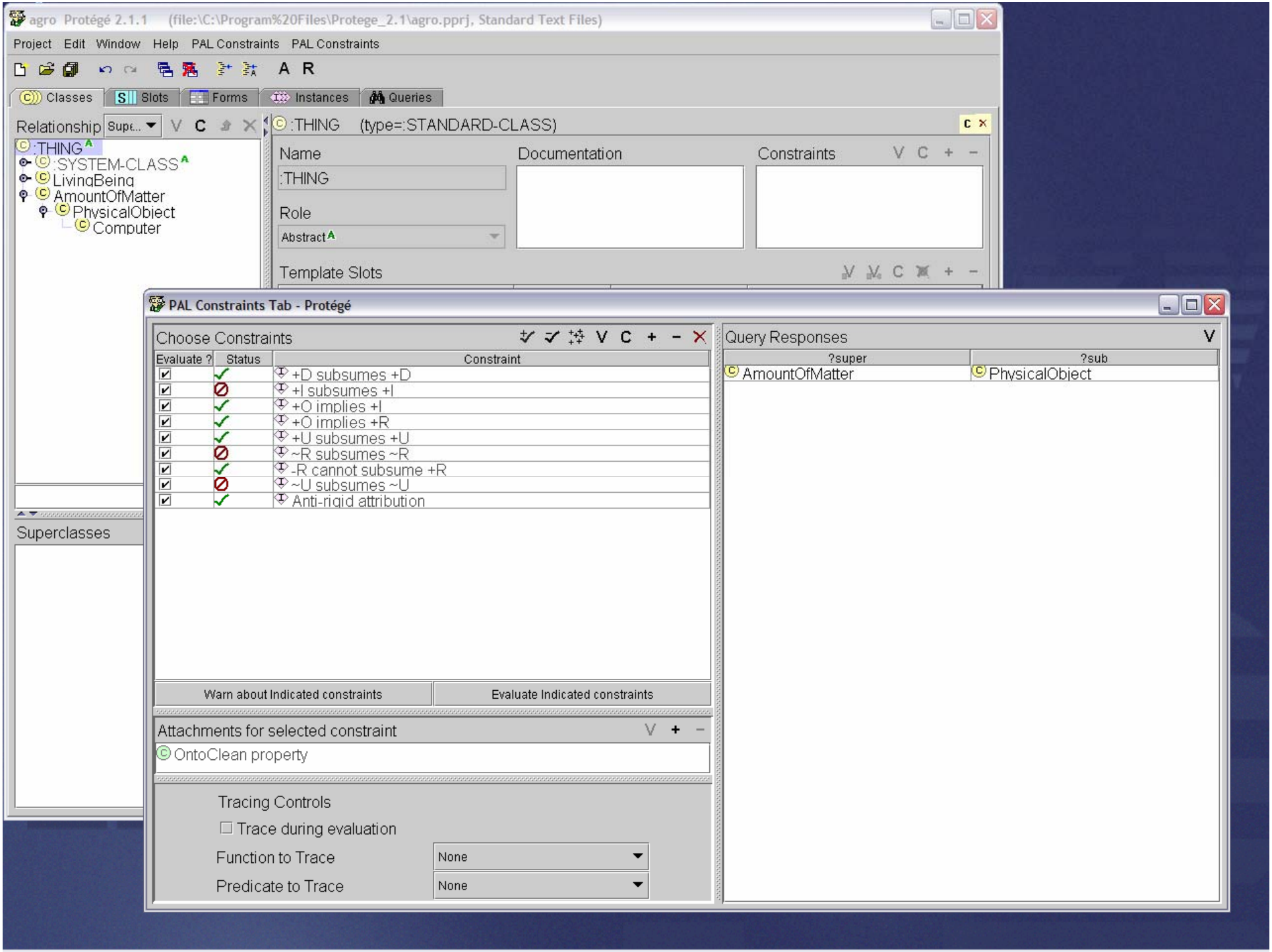


Question: How many books do you have on Hemingway?
Answer: 5,000

Constitution



Question: What types of matter will conduct electricity? Answer should NOT include computers.


agro Protégé 2.1.1 (file:\C:\Program%20Files\Protege_2.1\agro.pprj, Standard Text Files)
Project Edit Window Help PAL Constraints PAL Constraints
Classes Slots Forms Instances Queries
Relationship
:THING
:SYSTEM-CLASS
LivingBeing
AmountOfMatter
PhysicalObject
Computer
:THING (type=:STANDARD-CLASS)
Name
:THING
Documentation
Constraints
Role
Abstract
Template Slots
Superclasses
PAL Constraints Tab - Protégé
Choose Constraints
Evaluate ?
Status
Constraint
+D subsumes +D
+I subsumes +I
+O implies +I
+O implies +R
+U subsumes +U
~R subsumes ~R
-R cannot subsume +R
~U subsumes ~U
Anti-rigid attribution
Warn about Indicated constraints
Evaluate Indicated constraints
Attachments for selected constraint
OntoClean property
Tracing Controls
Trace during evaluation
Function to Trace
None
Predicate to Trace
None
Query Responses
?super
?sub
AmountOfMatter
PhysicalObject

Text Analysis Support

- Document Classification
  - Subject hierarchies
  - Identify relevant concepts
- Information Extraction
  - Find individuals
  - Glossary extraction [Park, 2004]

Concept-specific Ontology Building through Search

- Human expert knows what she is interested in: anchor concept
- Find relations and other related concepts for the anchor concept
- Active acquisition of knowledge sources through search
  - Concept-defining knowledge source: glossaries or dictionaries
  - Up-to-date knowledge source: web documents
- Very useful for recognizing missing terms

Domain Term Recognition

- Nominal Expressions
  - acute radiation syndrome
  - intercontinental and submarine-launched ballistic missile
  - highly enriched uranium
- New Domain Word Identification
  - agroterrorism, astrobiology, biocomputation
- Generic Premodifier Filtering
  - average radial first harmonic runout
  - absolute amazement/zero

Domain Term Aggregation

- Abbreviations
  - 5-HT-3R --- 5-hydroxytryptamine type 3 receptor
  - D2T2 --- Dye Diffusion Thermal Transfer
  - nAchRs --- nicotinic acetylcholine receptors
- Aliases : T1 .. *{known as|called}* T2
  - Zomig, formerly known as 311C90
  - Eleutherococcus senticosus (ES), also known as Siberian ginseng or ciwuija

Domain Term Aggregation

- Spelling errors or alternative spellings
  - anesthesia --- anaesthesia
- Orthographic variants
  - audio/visual input --- audio-visual input
  - electro-magnetic clutch --- electromagnetic clutch
  - Passenger airbag --- passenger air bag
- Morphological variants
  - multiprocessing ps/2 --- multiprocessing ps/2s
  - CD ROM --- CD ROMs, CD-ROMs

Related Concept Recognition

- A term G is related to term T if
  - T and G share some words
    - Ballistic missile -- medium-range ballistic missile
  - T and G often appear together in same sentences
- Select a set of semantically related terms with higher domain specificity

Relation Extraction (IS-A)

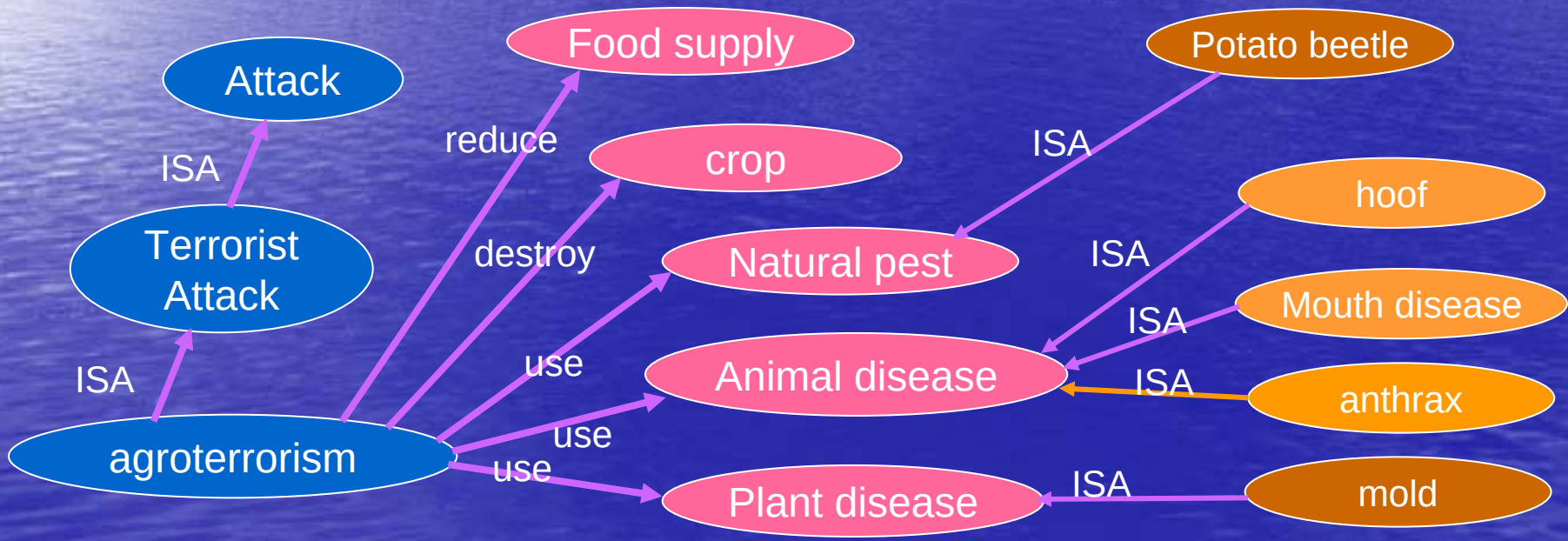
- Structurally Suggested ISA Relation
  - Ballistic missile. A guided rocket-powered delivery vehicle for use against ground targets
  - Position defense. The type of defense in which ….
  - Hyperspectral imagery. A term used to describe the imagery derived from ..
- Lexically Suggested ISA Relation
  - Ballistic missile ---ISA--- guided rocket-powered delivery vehicle
  - guided rocket-powered delivery vehicle ---ISA--- delivery vehicle

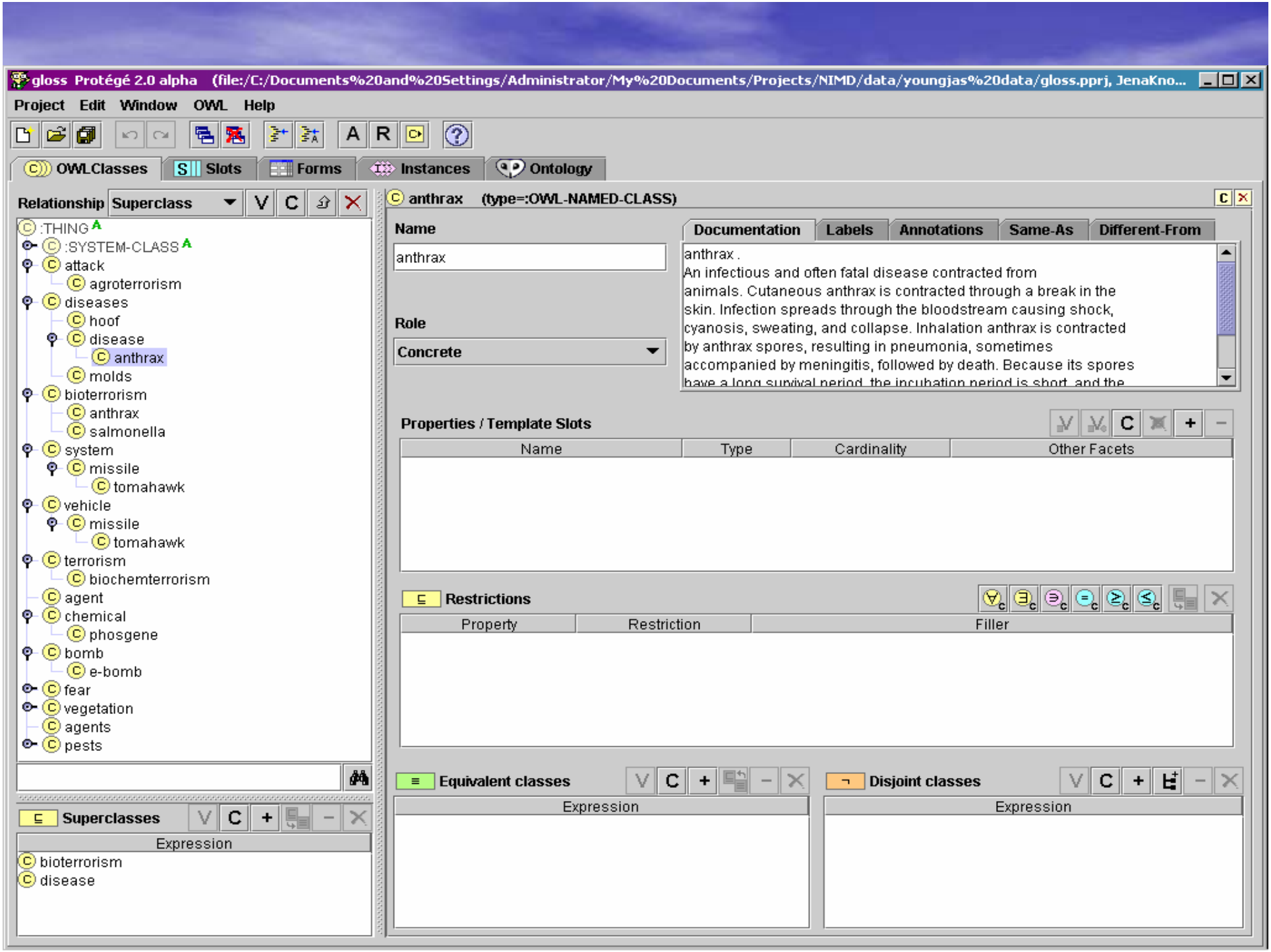
Lexical Patterns for IS-A

- T *is a kind/type of* H
- (T_1, T_2, ..., T_n) *and/or other* H
  - rescue, meteorological information, navigational aid, communications facilities and other services
- H_1, H_2, H_3 *{such as/including}* (T_1,T_2,...) *and/or* T
  - conditions such as fractures, wounds, sprains, strains, dislocations, concussions, and compressions

Ontology Construction

agroterrorism. Terrorist attacks aimed at reducing the food supply by destroying crops using natural pests such as the potato beetle, animal diseases such as hoof and mouth disease and anthrax, molds and other plant diseases.




gloss Protégé 2.0 alpha (file:/C:/Documents%20and%20Settings/Administrator/My%20Documents/Projects/NIMD/data/youngjas%20data/gloss.pprj, JenaKno...
Project Edit Window OWL Help
OWLClasses Slots Forms Instances Ontology
Relationship Superclass
:THING
:SYSTEM-CLASS
attack
agroterrorism
diseases
hoof
disease
anthrax
molds
bioterrorism
anthrax
salmonella
system
missile
tomahawk
vehicle
missile
tomahawk
terrorism
biochemterrorism
agent
chemical
phosgene
bomb
e-bomb
fear
vegetation
agents
pests
Superclasses
Expression
bioterrorism
disease
anthrax (type=:OWL-NAMED-CLASS)
Name
anthrax
Role
Concrete
Documentation Labels Annotations Same-As Different-From
anthrax .
An infectious and often fatal disease contracted from animals. Cutaneous anthrax is contracted through a break in the skin. Infection spreads through the bloodstream causing shock, cyanosis, sweating, and collapse. Inhalation anthrax is contracted by anthrax spores, resulting in pneumonia, sometimes accompanied by meningitis, followed by death. Because its spores
Properties / Template Slots
Name Type Cardinality Other Facets
Restrictions
Property Restriction Filler
Equivalent classes
Expression
Disjoint classes
Expression

Conclusions

- Ontology maintenance is a critical problem
- Need support
  - Tools help
  - Theories help
  - Text analysis helps
- All together helps more
  - Embedded in Protégé

...Bring Research into Practice...