Ontology
Consumer Analysis Tool
Onto*CAT*

Valerie Cross and Anindita Pal

Computer Science and Systems Analysis

Miami University, Oxford OH

2006 Protégé Conference
Stanford University

Agenda

- **Motivation**
- **Perspectives on Ontology Evaluation**
- **Some Current Approaches**
- **Ontology Consumer Analysis Tool**
- **Some Experiments Using OntoCAT**
- **Conclusion**
- **Future Plans**

CAT on a Log

Evaluating

OWL on a Log

Note that OWL and CAT are not only on two separate logs
But also in two separate worlds!

Motivation

- **Ontologies the "backbone of the Semantic Web"**

- **Development and deployment 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**

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?**
  - ***Ontology Selection: Ontology Evaluation on the Real Semantic Web***
    (Sabou, Lopez, Motta,Uren EON 2006)

What kinds of selection criteria?

- *Popularity*
  - metrics account solely for the links between different ontologies.
  - same principle as Web search engines, often use a modified version of the PageRank algorithm.
- Semantic data richness
  - determine richness of the ontology's conceptualization
- *Topic coverage*
  - level to which ontology covers a certain topic.
  - ontology concept labels compared to a set of query terms representing the domain.

What are we evaluating?

- **From U.S. National Center for Ontology Research (NCOR) position paper at EON 2006:**
  - **well-defined ontology design techniques, i.e., *quality* of design**
  - **principled measurement methods, i.e., *quality* of evaluation**
  - **higher quality ontologies, i.e., *quality* of content**

Some Approaches to Evaluating Ontologies

- *One-T [Bouillon et al 2002] :*
  - **Ontology Group at Universidad Politécnica de Madrid (UPM)**
  - ***Content for completeness, consistency and correctness***

- *OntoClean [Guarino and Welty 2002] :*
  - **The Ontology Group at the Italian National Research Council (CNR).**
  - ***Methodologies to evaluate during its entire lifetime***
  - ***Formal analysis of taxonomies***

Some Approaches to Evaluating Ontologies

ONTOMETRIC [Lozano-Tello and Gómez-Pérez 2004]

- **Ontology Group at Universidad Politécnica de Madrid (UPM)**

- *method to quantify the suitability of ontologies for the users' systems,*

  - **uses a taxonomy of 160 ontology characteristics,**

    - Content, language, development methodology, built by software tool, cost of use.

  - **not fully automated, based on AHP (Saaty 1977)**

- *Application Use of ontology* **to assess application's performance, merits of**

  - competency questions,

  - use cases,

  - scenarios

Consumer Perspective Approach

- **Noy [2004] suggests for ontology re-use need more research from consumer perspective**
  - Somewhat analogous to reviewing Table of Contents and Index, number of pages, etc. for the usefulness of book before deciding whether to check out or purchase.

- **AKTiveRank [Alani and Brewster 2005]**
  - AKT (Advanced Knowledge Technologies) consortium of British universities: _Southampton_, Edinburgh, Aberdeen, Sheffield and The Open University.
  - ranks ontologies retrieved by an ontology search engine based on set of query terms and measures

- **OntoQA Analysis tool [Tartir 2005]**
  - LSDIS (Large Scale Distributed Information Systems) Lab, University of Georgia
  - analyzes ontology schemas and their populations and describes them through a set of metrics.

AKTiveRank

- **Ranks ontologies retrieved by search engine (EON 2005)**
  - **Class match: coverage of query terms**
  - **Centrality: more central a class**
  - **Density: degree of details**
  - **Semantic similarity measure: closeness of classes**
  - **Produces overall rank**
- **Extensions (EON 2006 and Protégé Conference)**
  - **Collect vocabulary for domain interest**
  - **Ranking based on number of class labels that match with terminology for domain of interest**
  - **New Centrality based on high "betweenness"**

OntoQA

- ## Schema:
    - **Relationship richness**
    - **Attribute richness**
    - **Inheritance richnness**

- ## Instances:
    - **Class Richness**
    - **Average Population**
    - **Connectivity**
    - **Cohesion**
    - **Importance**
    - **Relationship Richness**
    - **Fullness**

Ontology Consumer Analysis Tool

- **plug-in for OWL Protégé**

- **very parameterized**
  - **Intensional and extensional**
  - **View metrics interested in**
    - **Size**
    - **Structure**
  - **User selectable root for analysis**
  - **User selectable relation for establishing extensional structure**

WordNet

- **Princeton University**

- **Terminological ontology of English**
    - **Organizes nouns, verbs, adjectives and adverbs into synonym sets**
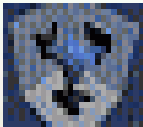   - **Simple intensional structure: 10 classes**
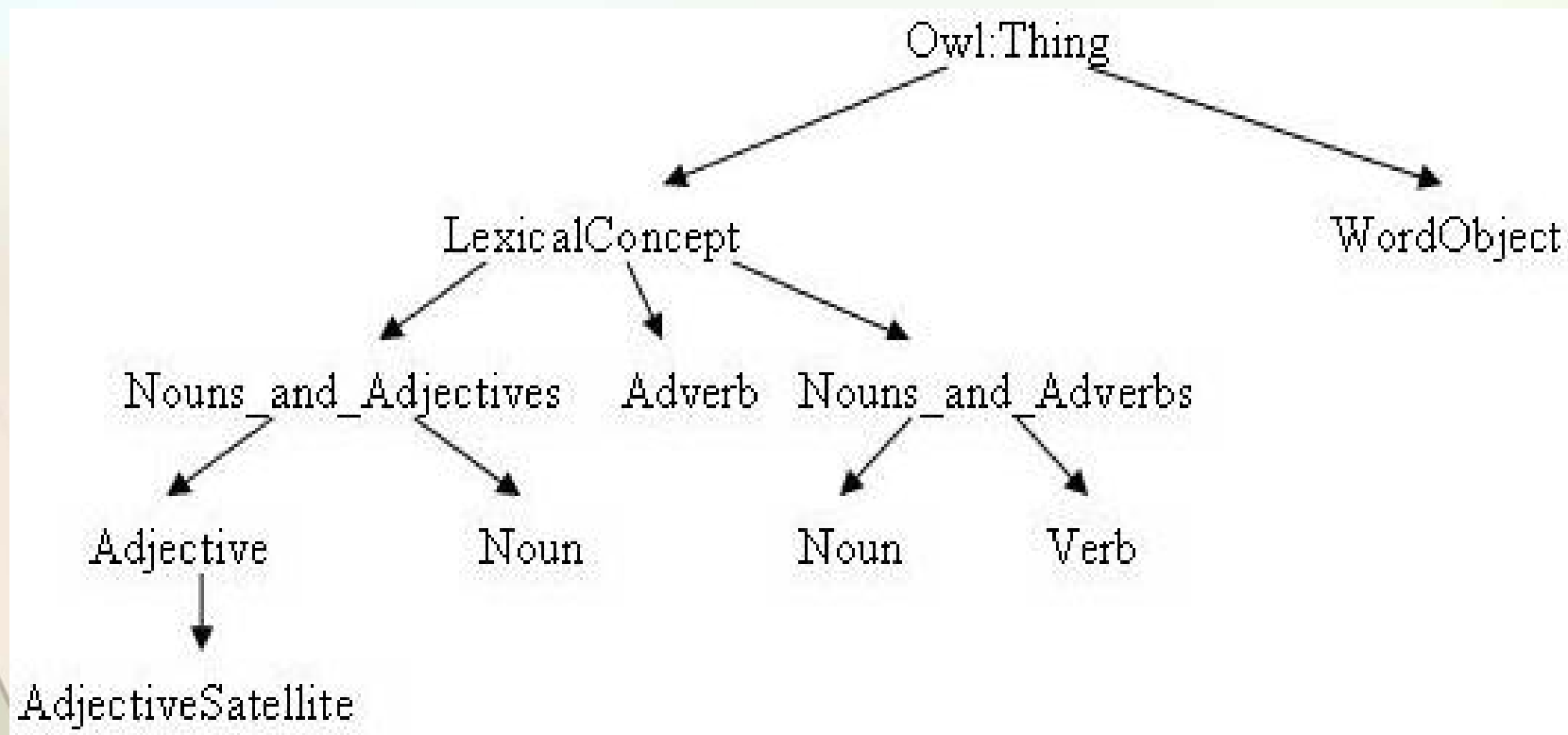
WordNet

- **Complex extensional structure based on hypernymOf /hyponymOf**
- **Example Root Instance "entity, physical thing", one of the nine noun roots**
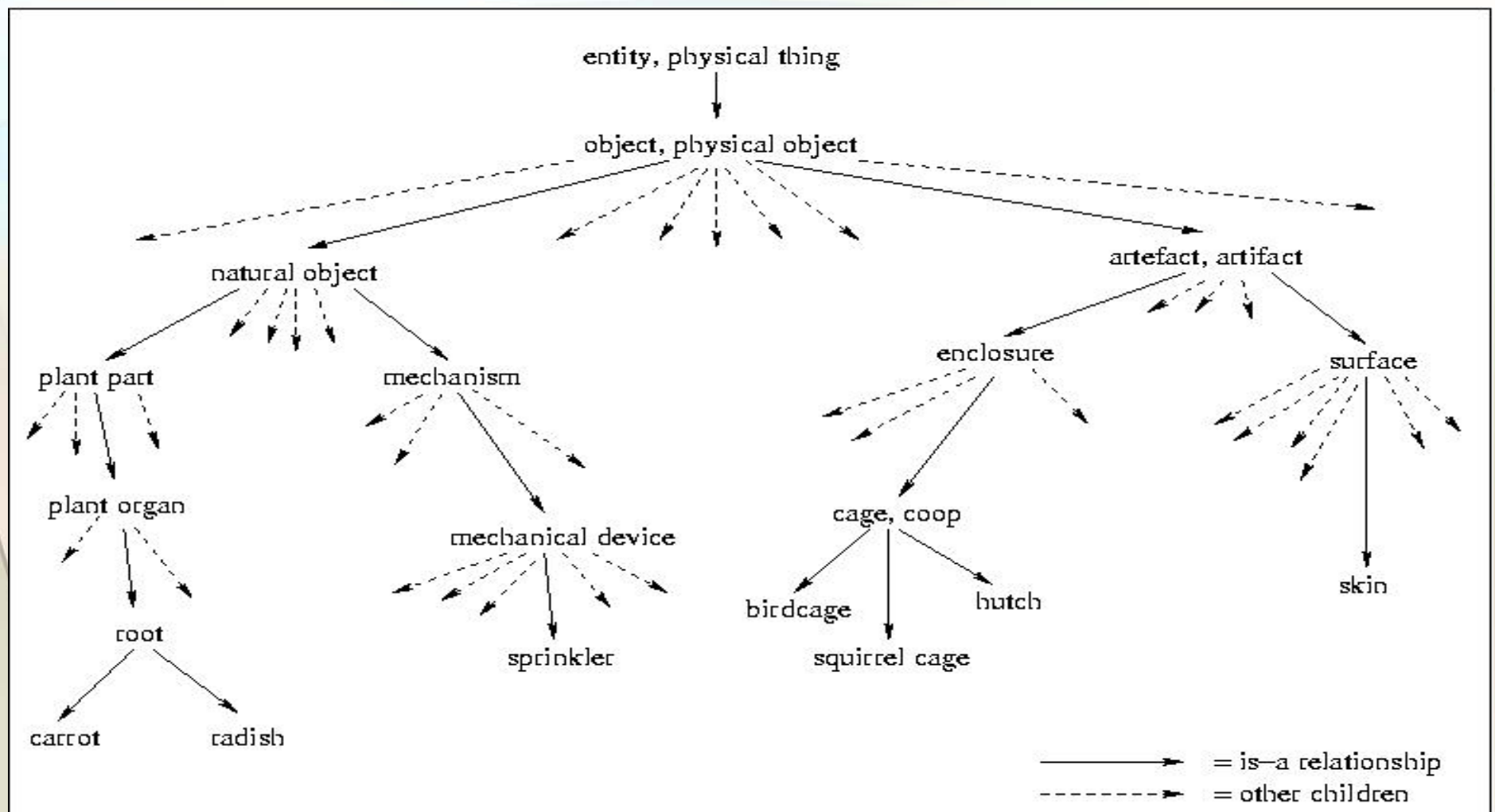
Onto*CAT* User Interface
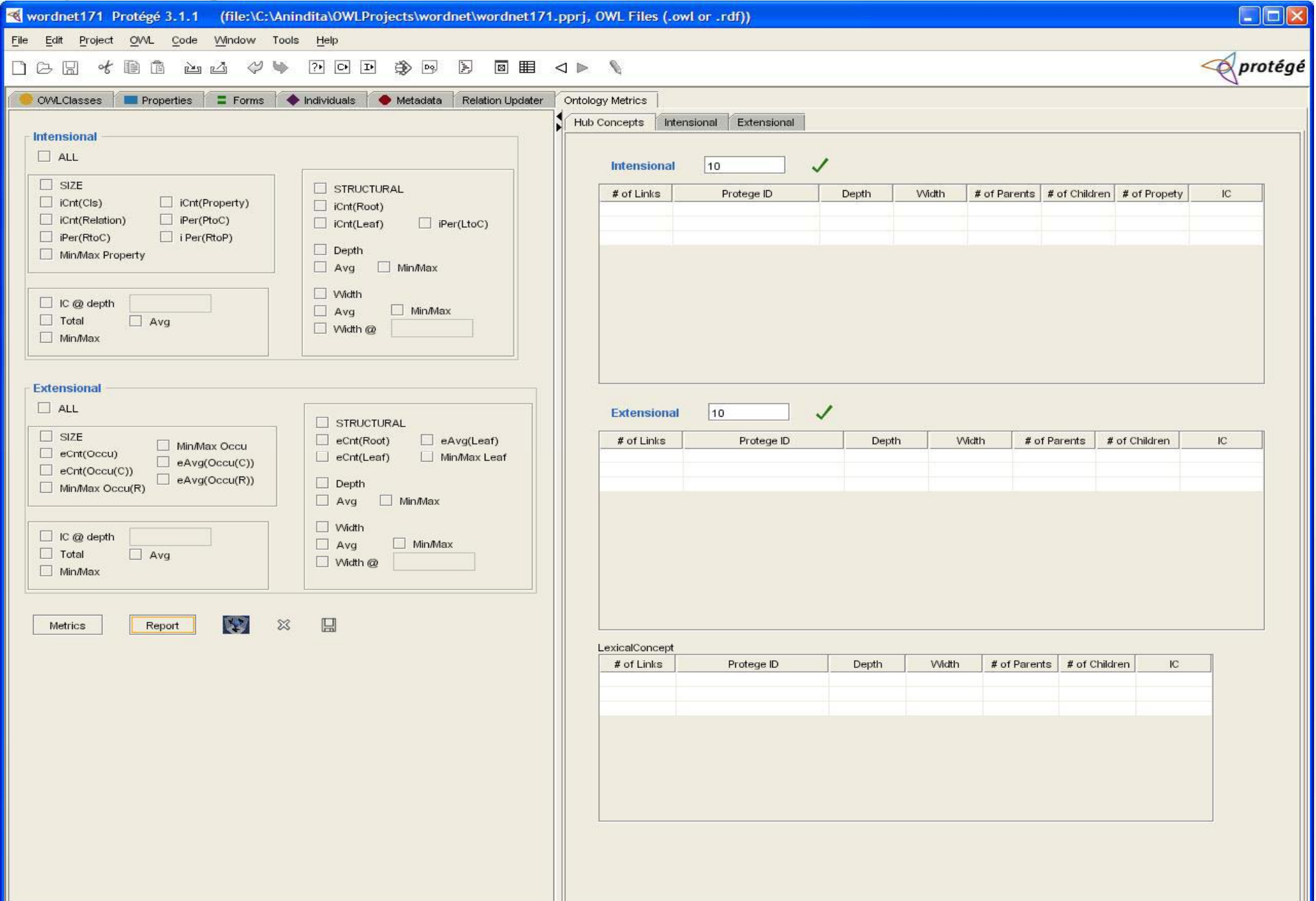
Onto*CAT* Buttons

- **Metrics Button**
  - **Display result of selected metrics**

- **Report Button**
  - **Report result of selected to file**

- **Button**
  - **Generate tree of hub concept to visualize**
  - **Click hub for individual hub visualization**

Onto*Cat* Selection Class/Extensional Relation

Onto*CAT* Hub Analysis

protégé

Ontology Metrics

Hub Concepts | Intensional | Extensional

Intensional 10 ✓

# of Links	Protege ID	Depth	Width	# of Parents	# of Children	# of Propety	IC
4	Noun	3	4	4	0	11	1.0
4	LexicalConcept	1	2	1	7	3	0.053
3	Nouns_and_Verbs	2	2	2	2	4	0.5
3	Adjective	3	4	3	1	5	0.684
3	Nouns_and_Adjectives	2	2	2	3	4	0.369
2	Verb	3	4	3	0	7	1.0
1	WordObject	1	2	1	0	4	1.0
1	AdjectiveSatellite	4	1	4	0	5	1.0
1	Adverb	2	2	2	0	3	1.0

Extensional 10 ✓

# of Links	Protege ID	Depth	Width	# of Parents	# of Children	IC
5	c100002956	2	5	1	4	0.431
4	c100002219	1	6	1	3	0.51
3	c100013067	1	6	1	2	0.313
3	c100001742	0	2	0	3	0.094
3	c100016840	0	2	0	3	0.51
1	c100002471	2	5	1	0	1.0
1	c100002579	2	5	1	0	1.0
1	c100002361	2	5	1	0	1.0
1	c100004911	1	6	1	0	1.0
1	c100002664	2	5	1	0	1.0

LexicalConcept

# of Links	Protege ID	Depth	Width	# of Parents	# of Children	IC
5	c100002956	2	5	1	4	0.431
4	c100002219	1	6	1	3	0.51
3	c100013067	1	6	1	2	0.313
3	c100001742	0	2	0	3	0.094
3	c100016840	0	2	0	3	0.51
1	c100002471	2	5	1	0	1.0
1	c100002579	2	5	1	0	1.0
1	c100002361	2	5	1	0	1.0
1	c100004911	1	6	1	0	1.0
1	c100002664	2	5	1	0	1.0

iPer(LtoC)

Max

eAvg(Leaf)
Min/Max Leaf

Max

Onto*CAT* Intensional Report

Extensional Hub Summary for WordNet

Extensional :						
#Links	Concept Name	Depth	Width	#Parent	#Child	IC
554	c107017569 (word_city, word_urban_center, word_metropolis)	6	18422	1	553	0.44
398	c101185314 (word_bird_genus)	4	9955	1	397	0.47
360	c101537097 (word_mammal_genus)	4	9955	1	359	0.47
356	c110146714 (word_herb, word_herbaceous_plant)	6	18422	1	355	0.38
343	c108833984 (word_writer, word_author)	4	9955	1	342	0.43
328	c100005303 (word_human, word_somebody, word_person, word_someone, word_soul, word_individual, word_mortal)	2	1300	2	326	0.19
321	c109528938 (word_asterid_dicot_genus)	5	17882	1	320	0.49

Onto*CAT* Root Summary

Root Occurrence in <Cls>

Root ID	# of Leaf	Avg Depth	Max Depth	Min Depth	Avg Width	Max Width	Level at Max	Min Width	Level at Min
c100023182	527	5.63	12	1	55.23	124	4	1	0,12
c100022634	7146	5.12	10	1	711.36	4389	5	1	0
c100020595	2367	5.18	11	1	264.66	618	3	1	0
c100022113	4718	5.15	11	1	535.66	1668	5	1	0
c100021905	821	4.33	8	1	125.11	448	4	1	0
c100001742	56006	7.33	17	1	3891.61	15406	7	1	0,17
c100016840	3312	6.47	12	2	337.61	1074	7	1	0
c100016993	7393	6.61	13	2	764.92	2163	6	1	0
c100025413	1156	5.16	10	1	141.36	399	4	1	0,10

UMLS Hub Summaries

Extensional :

#Links	Concept Name	Depth	Width	#Parent	#Child		IC
27	C1314803	3	554	2	25		0.32
25	C0376109	1	5	1	26		0.23
25	C0019829	4	3685	1	26		0.56
22	C0086692	3	554	2	20		0.50
22	C0000768	2	69	2	20		0.38
21	C0178359	2	69	1	20		0.47
21	C0027651	2	69	1	20		0.31
21	C0178332	3	554	1	20		0.47
20	C0178314	2	69	1	19		0.19
20	C0553730	5	8556	1	19		0.70

Table 6.22 Hub Summary of ICD9CM Ontology

Extensional :

#Links	Concept Name	Depth	Width	#Parent	#Child		IC
3328	C0305080	2	558	1	3327		0.30
2520	C0308208	2	558	1	2519		0.33
1032	C0201828	4	22792	1	1031		0.40
725	C0040676	4	22792	1	724		0.43
661	C0020289	4	22792	1	660		0.44
649	C0030016	4	22792	1	648		0.44
594	C0686939	3	8270	1	593		0.45
562	C0223075	4	22792	1	561		0.45
527	C0202295	4	22792	1	526		0.46
514	C0685935	3	8270	1	513		0.46

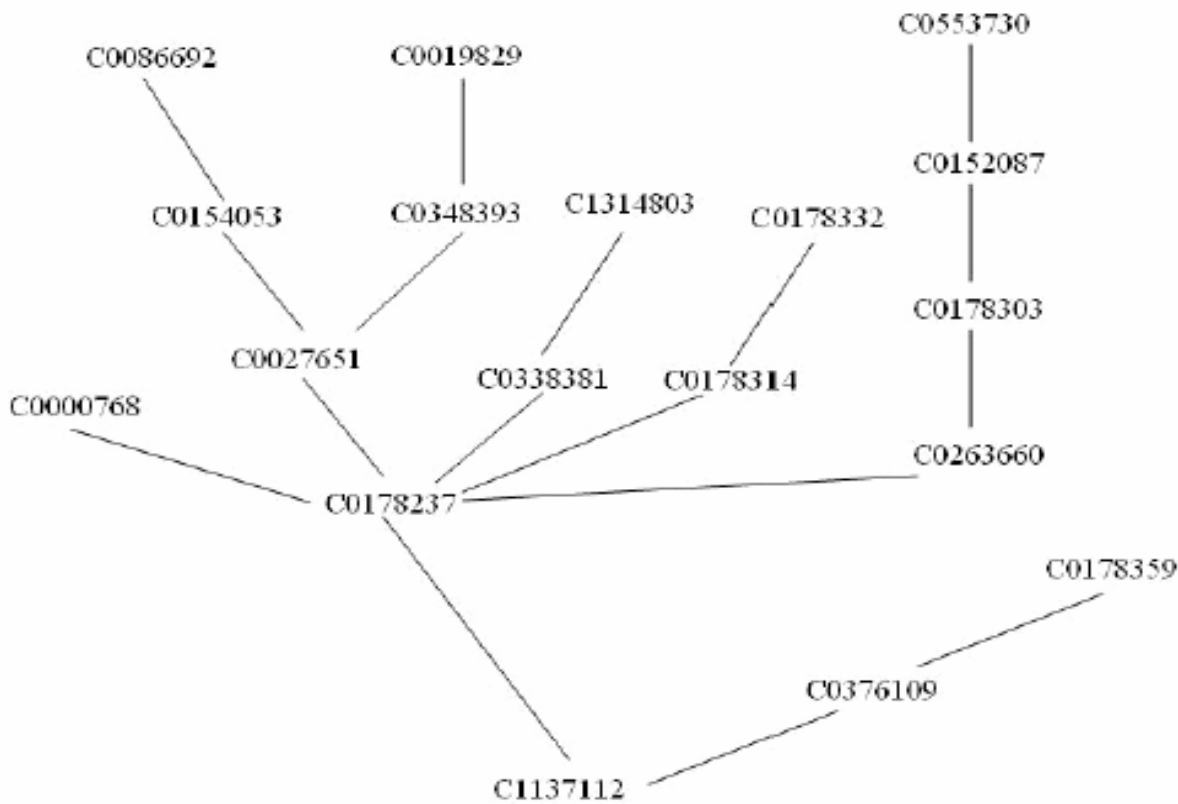Table 6.23 Hub Summary of SNMI Ontology

Visualizing Hubs

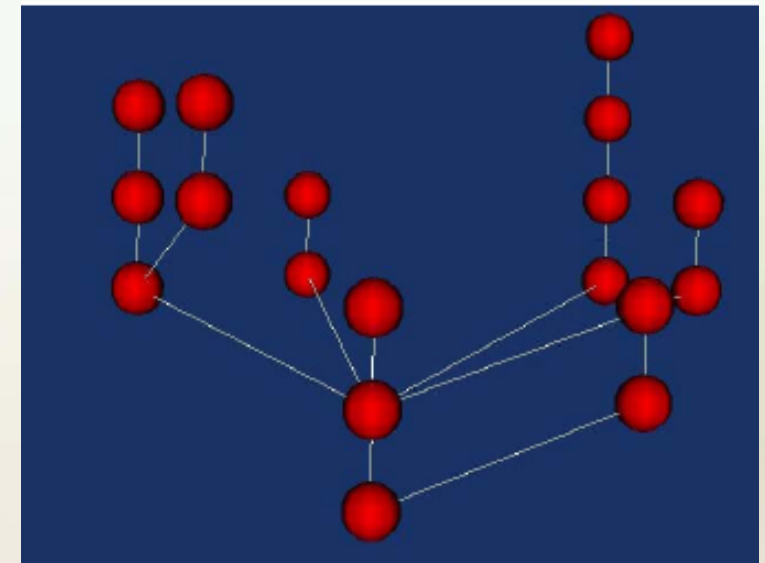Figure 6.5 ICD9CM Graphical View of Hubs with Connecting Concepts.

Figure 6.6 ICD9CM Information Visualization of Hubs with Connecting Concepts.

Summary

- **Many flavors of ontology evaluation or selection**
- **OntoCat - one of several tools to begin addressing needs of ontology evaluation for the purpose of re-use**
- **Structural and size analysis just one set of parameters.**
- **Challenge specifying parameters or structural properties for evaluation**
  - **user preference**
  - **purpose for reusing ontology**

Possible Future Work

- **Interface with filtering/selection approaches such as AKTiveRank before perform evaluation**

- **Comparison metrics/charts for multiple ontologies in addition to ranking**

- **Current Visualization**

  - **Hubs visualization Improvement**

  - **Individual hub visualization**

  - **Top-level summary**

  - **Bottom-up level summary**