Ontology Consumer Analysis Tool Onto*CAT*

Valerie Cross and Anindita Pal
Computer Science and Systems Analysis
Miami University, Oxford OH

SEBIZ WORKSHOP
University of Georgia
November 6, 2006

Agenda

- Motivation
- Perspectives on Ontology Evaluation
- Some Current Approaches
- Ontology Consumer Analysis Tool
- Some Evaluations Using OntoCAT
- Summary
- Future Plans

Motivation

- 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)

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*

Earlier 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***

Earlier Approaches to Evaluating Ontologies

- *ONTOMETRIC [Lozano-Tello & Gómez-Pérez 2004]*
  - **Ontology Group at Universidad Politécnica de Madrid**
  - ***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)**
    - **drawback is its *usability***: complicated and time-consuming to specify characteristics of an ontology and assessing some characteristics is subjective.
- 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.
  - ontology summarization, e-pinions for ontologies, views
- 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 (2006 EON, Protégé Conference, ISWC)
  - Collect vocabulary for domain interest
  - Ranking based on number of class labels that match with terminology for domain of interest, parameterized partial vs. exact match
  - New Centrality using JUNG's "betweenness" measure Elinimated original Centrality measure since somewhat redundant with Density

Ontology Consumer Analysis Tool

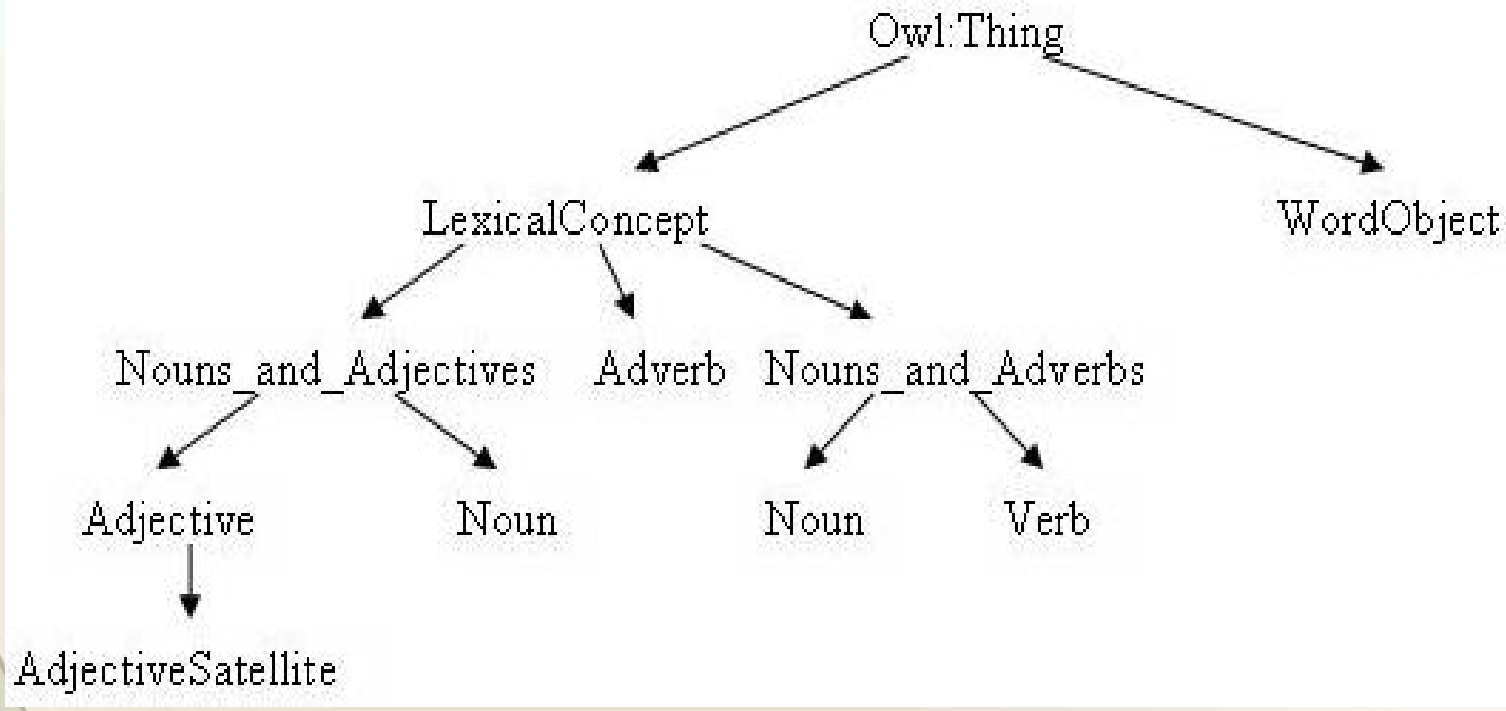
- Envision as another tool that may be used to further analyze and compare after selecting set of ontologies that are potential candidates
- Currently working on methods to become more “Consumer oriented”
  - summarizations of and combinations of the various metrics
  - Visualization of various metrics to help consumer understand result comparisons

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

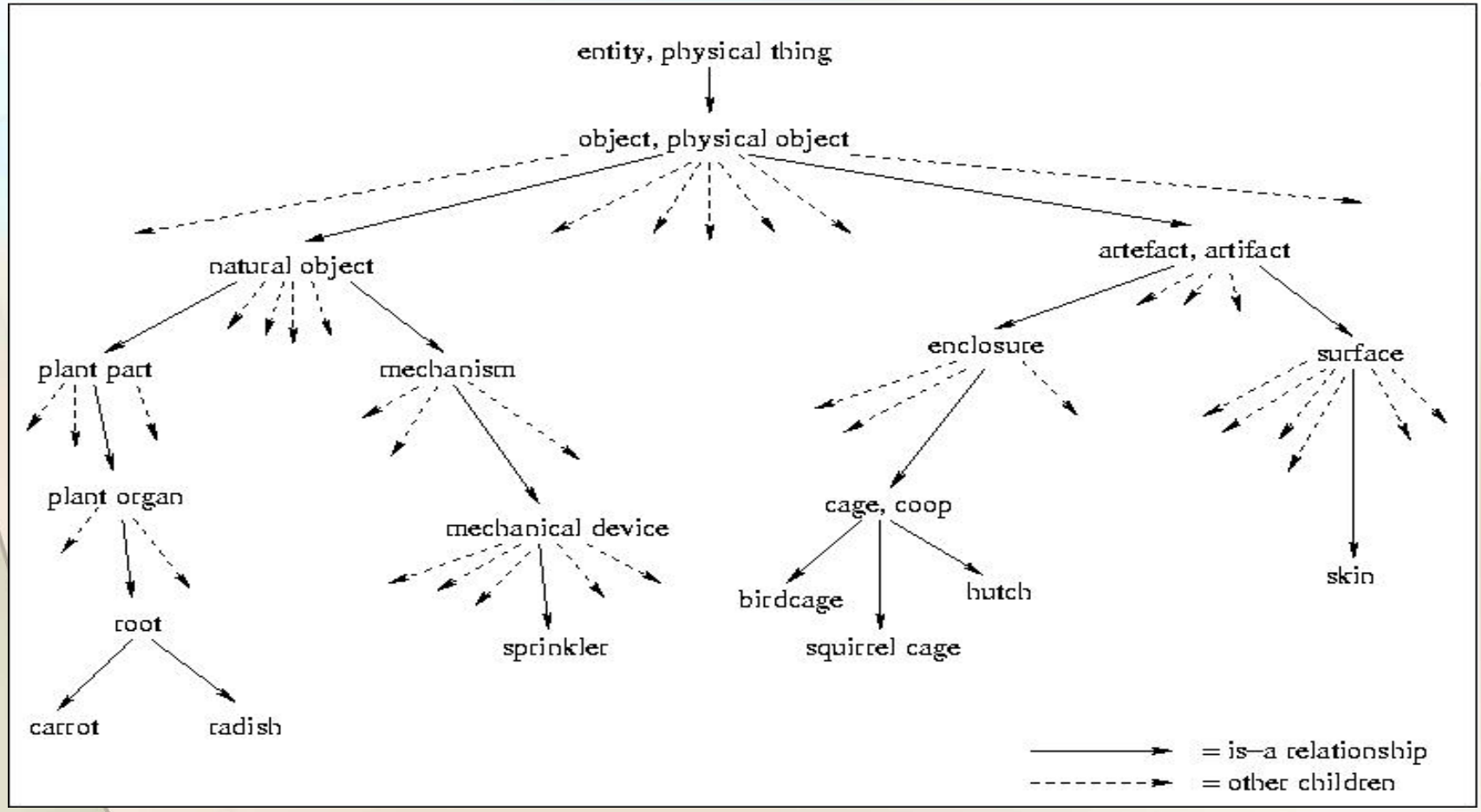
WordNet

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



WordNet

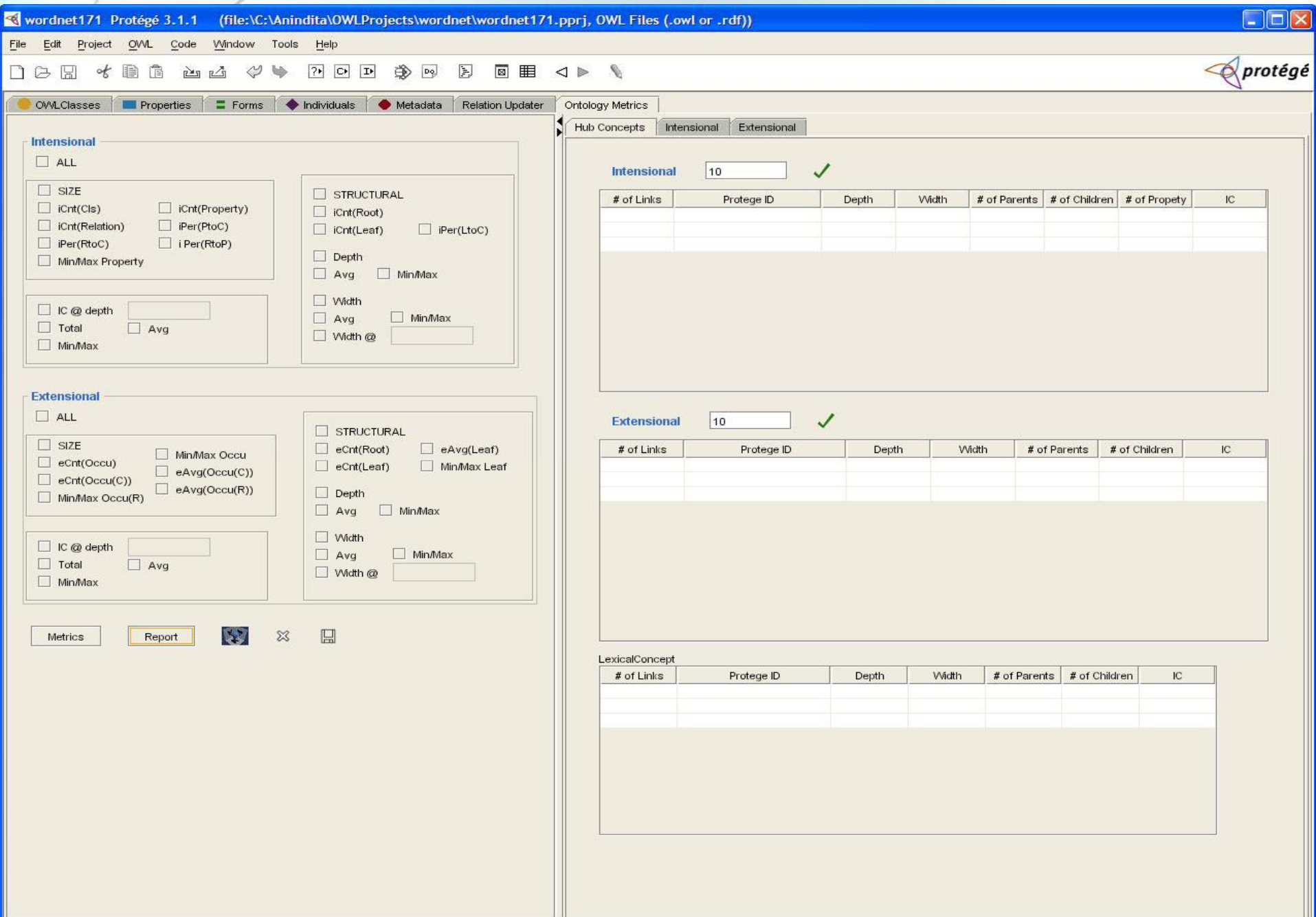
- **Complex extensional structure based on hypernymOf /hyponymOf**
- **Example Root Instance “entity, physical thing”, one of the nine noun roots**



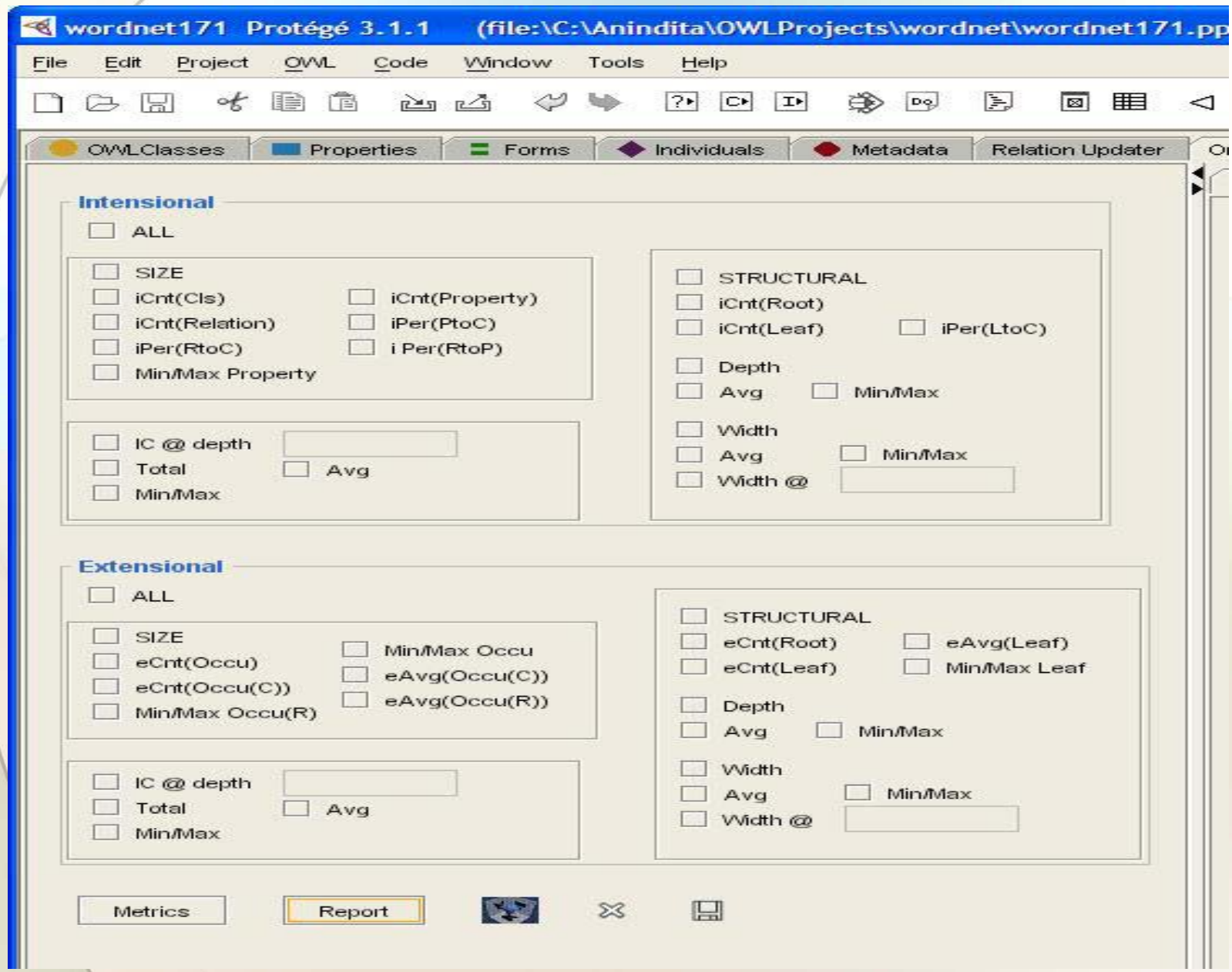
Ontology Consumer Analysis Tool

- Metrics categorized into
  - Intensional: Measures on definition (classes)
  - Extensional: Measures actual occurrences (instances)
  - Size and Structural
- Summary
  - Hub Concepts
    - Concepts with maximum links in & out
  - Root Concepts

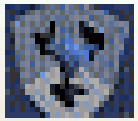
Onto*CAT* User Interface



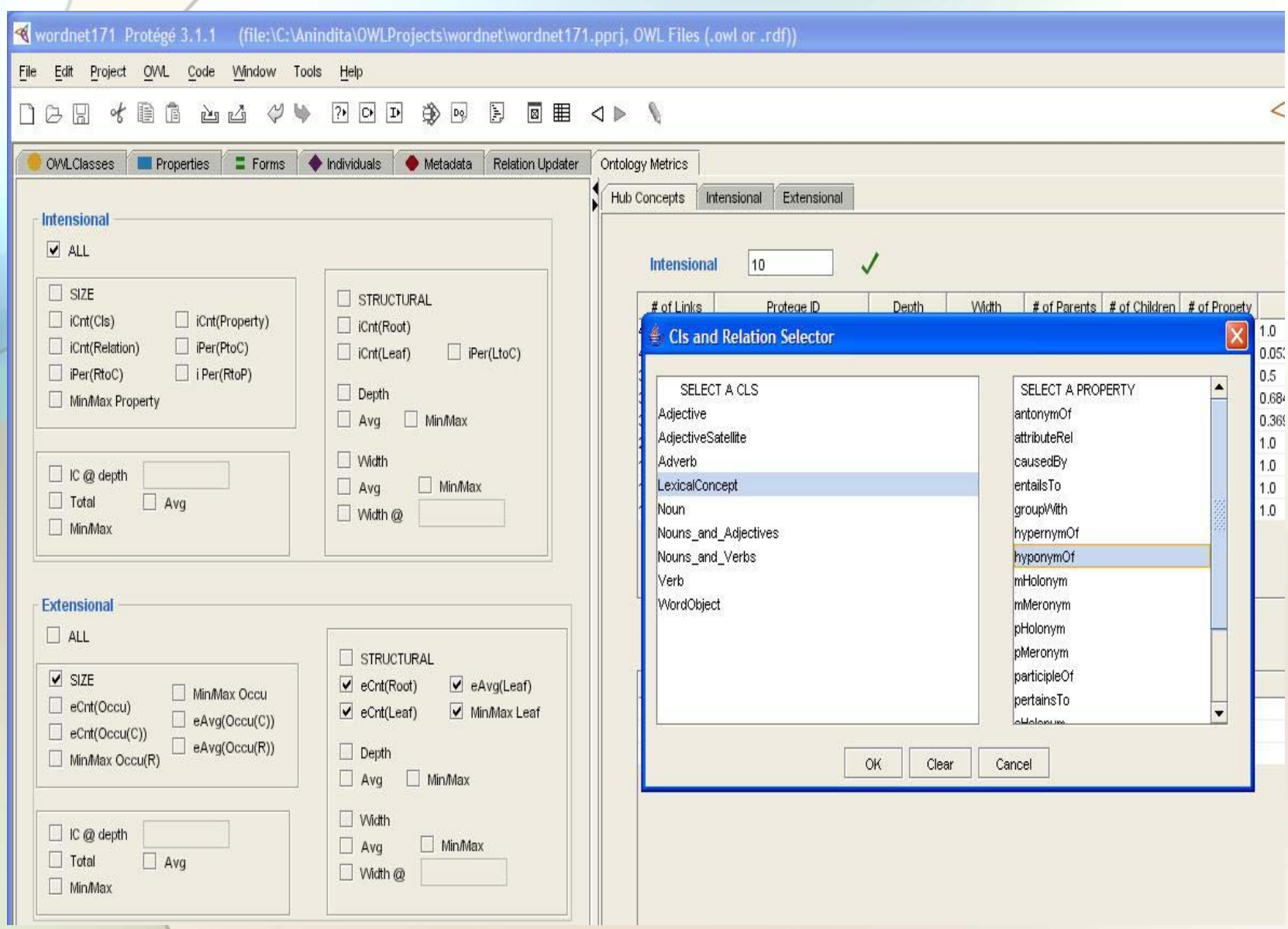
Ontology Analysis Tool



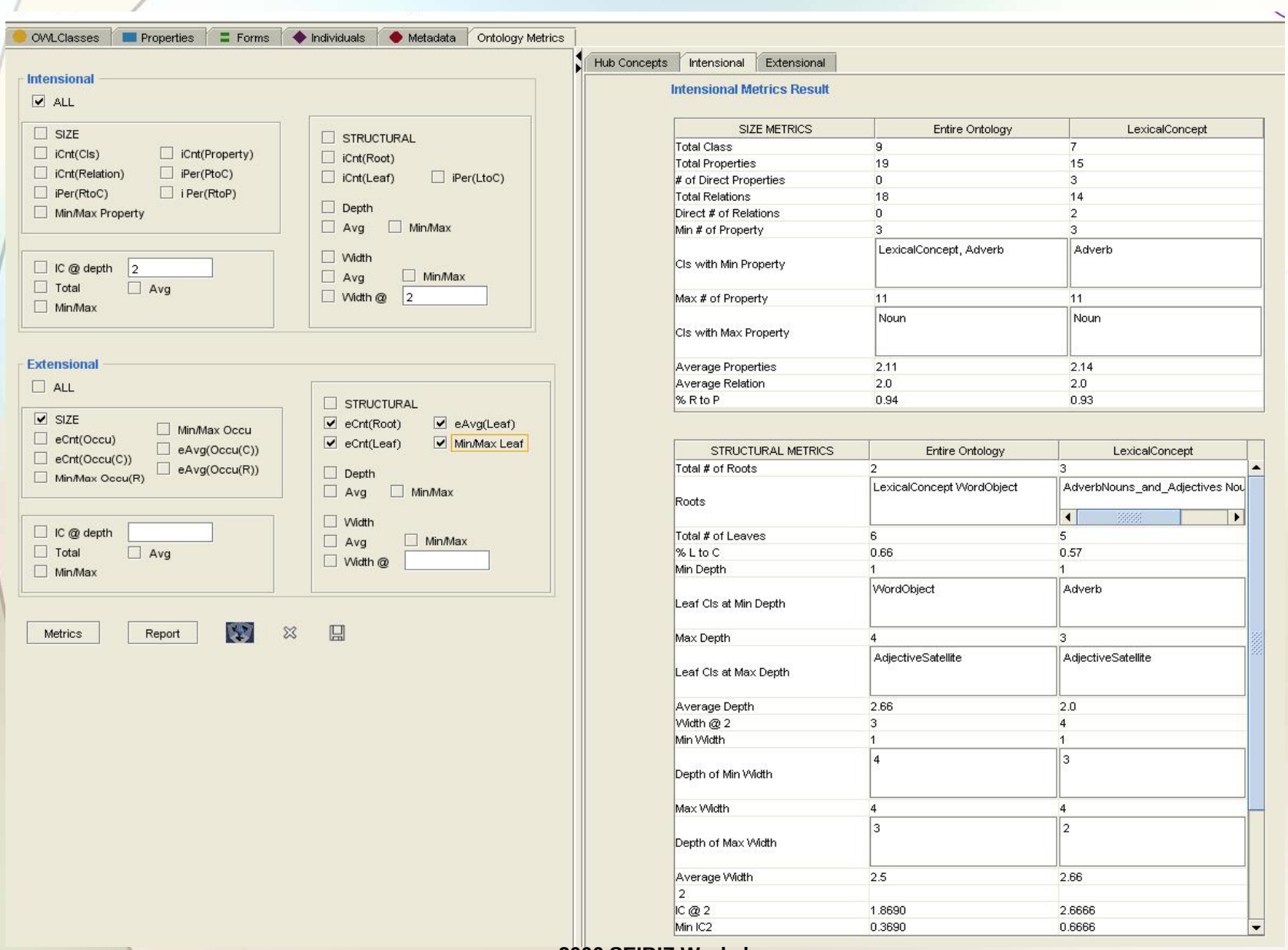
Onto*CAT* Buttons

- **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



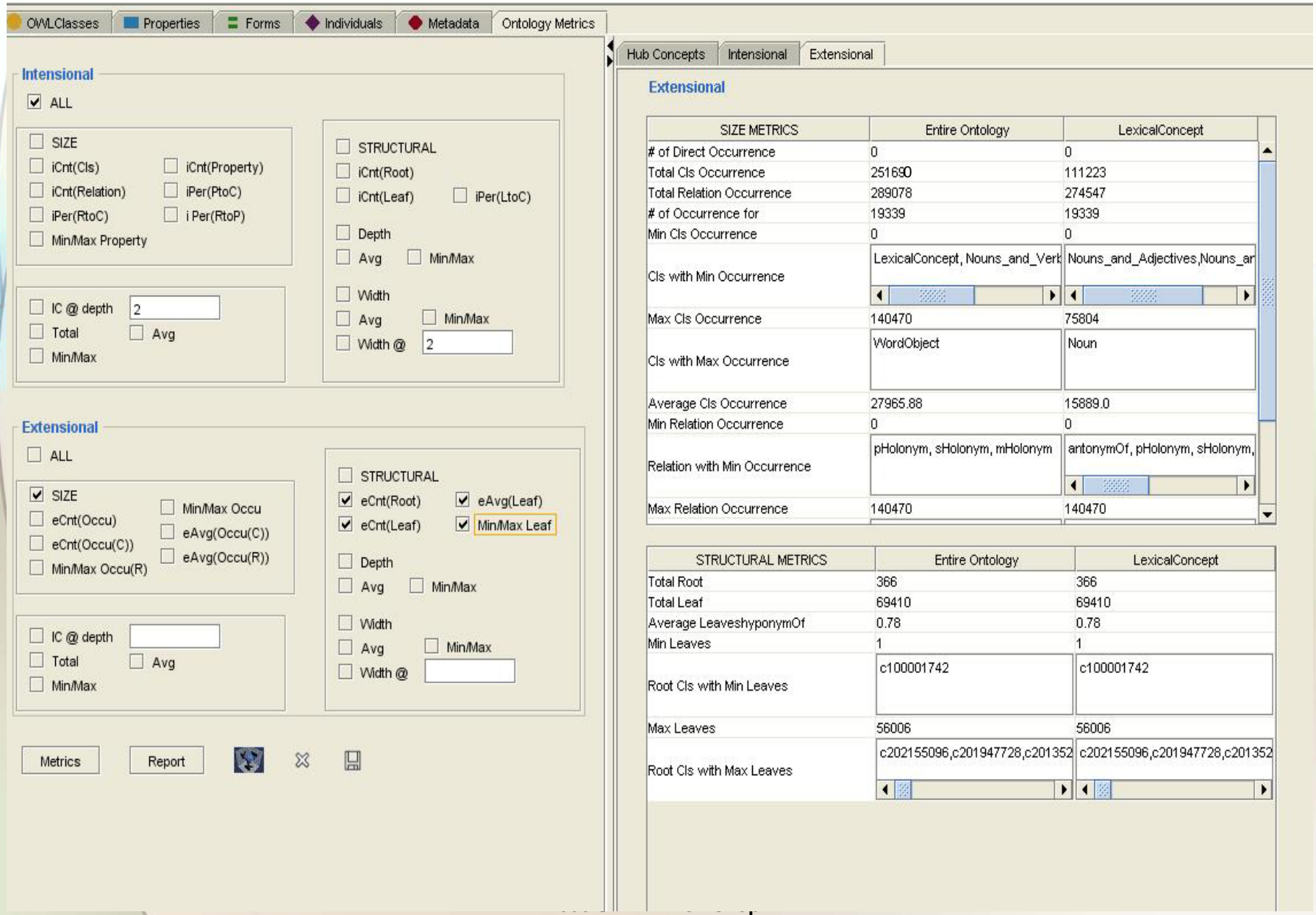
Onto*CAT* Intensional Report



SIZE METRICS	Entire Ontology	LexicalConcept
Total Class	9	7
Total Properties	19	15
# of Direct Properties	0	3
Total Relations	18	14
Direct # of Relations	0	2
Min # of Property	3	3
Cls with Min Property	LexicalConcept, Adverb	Adverb
Max # of Property	11	11
Cls with Max Property	Noun	Noun
Average Properties	2.11	2.14
Average Relation	2.0	2.0
% R to P	0.94	0.93

STRUCTURAL METRICS	Entire Ontology	LexicalConcept
Total # of Roots	2	3
Roots	LexicalConcept WordObject	AdverbNouns_and_Adjectives Nou
Total # of Leaves	6	5
% L to C	0.66	0.57
Min Depth	1	1
Leaf Cls at Min Depth	WordObject	Adverb
Max Depth	4	3
Leaf Cls at Max Depth	AdjectiveSatellite	AdjectiveSatellite
Average Depth	2.66	2.0
Width @ 2	3	4
Min Width	1	1
Depth of Min Width	4	3
Max Width	4	4
Depth of Max Width	3	2
Average Width	2.5	2.66
2		
IC @ 2	1.8690	2.6666
Min IC2	0.3690	0.6666

Onto*CAT* Extensional Report



Onto*CAT* Root Summary for WordNet Nouns

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
c100023162	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	15405	7	1	0,17
c100016840	3312	6.47	12	2	337.61	1074	7	1	0
c100016993	7993	6.61	13	2	764.92	2163	6	1	0
c100025413	1156	5.16	10	1	141.36	399	4	1	0,10

E-COMMERCE ONTOLOGIES

- **Standardized vocabulary of product and services terminology referred to as Product and Service Categorization Standards (PSCS)**
- **Example PSCS developed into intensional ontologies**
  - UNSPSC
    - hierarchical classification of all products and services for use throughout the global marketplace.
    - In between coarse taxonomies for customs purposes and expressive descriptive languages for products and services
  - eCl@ss,
    - offers a standard for information exchange between suppliers and their customers.
  - Both important horizontal standards since cover a broad range of industries

Previous Study (Hepp 2005)

- **Previous study "to assess the quality and maturity of products and services categorization standard" proposed metrics target four aspects:**
  - size, growth, and maintenance volume,
  - degree of balance among segments, hierarchical order, and the breadth of coverage,
  - size and expressiveness of the property library,
  - specificity of property assignment in class-wise property lists.
- **Example metrics**
  - *"size of segments"* corresponds to OntoCAT's iCnt(C)(cj-root), the number of classes for a root class.
  - *"size"* corresponds to OntoCAT's iCnt(C), the number of classes for the entire intensional ontology.
  - *"property list size"* corresponds to iCnt(P), the number of properties defined for the entire intensional ontology

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

OntoCAT Root summary for the UNSPSC ontology.

Concept Name	Total Classes	Total Leaf	Max Depth	Min Depth	Avg Depth	Max Width	Level @ max	Min Width	level @ min	Avg Width
Laboratory_and_Measuring_and_Observing_and_Testing_Equipment	1103	1008	3.00	3	3	1008	3	3	1	367.66
Structures_and_Building_and_Construction_and_Manufacturing_Components_and_Supplies	696	615	3.00	3	3	615	3	1	1	232
Chemicals_including_Bio_Chemicals_and_Gas_Materials	614	508	3.00	3	3	508	3	14	1	204.66
Drugs_and_Pharmaceutical_Products	611	514	3.00	3	3	514	3	15	1	203.66
Commercial_and_Military_and_Private_Vehicles_and_their_Accessories_and_Components	496	417	3.00	3	3	417	3	10	1	165.33
Communications_and_Computer_Equipment_and_Peripherals_and_Components_and_Supplies	416	369	3.00	3	3	369	3	3	1	138.66
Industrial_Manufacturing_and_Processing_Machinery_and_Accessories	395	342	3.00	3	3	342	3	13	1	131.66
Distribution_and_Conditioning_Systems_and_Equipment_and_Components	340	312	3.00	3	3	312	3	4	1	113.33
Power_Generation_and_Distribution_Machinery_and_Accessories	305	275	3.00	3	3	275	3	5	1	101.66
Farming_and_Fishing_and_Forestry_and_Wildlife_Contracting_Services	280	237	3.00	3	3	237	3	8	1	9333
Industrial_Production_and_Manufacturing_Services	279	236	3.00	3	3	236	3	9	1	93
Politics_and_Civic_Affairs_Services	279	241	3.00	3	3	241	3	8	1	93
Domestic_Appliances_and_Supplies_and_Consumer_Electronic_Products	275	246	3.00	3	3	246	3	7	1	91.66

OntoCAT intensional metrics ecl@ss

Size	owl:Thing	Structural	owl:Thing
Total #Cls [iCnt(C)]	76975	Total Roots [iCnt(Roots)]	25684
Total #Property [iCnt(P)]	5527	Total Leaves [iCnt(leaves)]	51317
Total #Relation [iCnt(R)]	2293	Average Leaves [iAv(leaves)]	0.66
% R to P [Per(R of P)]	0.41	Average Depth [iAvDepth]	3.4
Max #Property [iMaxTotal(P to C)]	5527	Width @ 1 [iWidth($depth_k$)]	25684
		Max Depth [iMaxDepth]	5
		Min Depth [iMinDepth]	1
		Max Width [iMaxWidth]	26162
		Depth of iMaxWidth	2
		Min Width [iMinWidth]	4533
		Depth of iMinWidth	3

Used http://www.heppnetz.de/eclassowl as input to OntoCAT

OntoCAT ecl@ss Root Summary

Concept Name	Total Classes	Total Leaf	Max Depth	Min Depth	Avg Depth	Max Width	Level @ max	Min Width	level @ min	Avg Width
C_AAG961003-tax	10823	5038	4	1	3.94	5292	3	20	1	2655.75
C_AAB572002-tax	5317	2181	4	1	3.8	2624	3	35	1	1329.25
C_AAB072002-tax	3983	1669	4	1	3.82	1973	3	19	1	995.75
C_AAD302002-tax	3585	1317	4	1	3.71	1756	3	37	1	896.25
C_AAF876003-tax	2927	1315	4	1	3.88	1444	3	20	1	731.75
C_AAC473002-tax	2653	1186	4	1	3.88	1320	3	7	1	663.25
C_AAC350002-tax	2431	1024	4	1	3.82	1192	3	24	1	607.75
C_AAB315002-tax	2127	850	4	1	3.77	1041	3	23	1	531.75
C_AAA183002-tax	2065	832	4	1	3.79	1019	3	14	1	516.25
C_AAA862002-tax	1927	739	4	1	3.73	932	3	32	1	481.75
C_AAA647002-tax	1603	589	4	1	3.68	763	3	39	1	400.75
C_AAD111002-tax	1519	580	4	1	3.74	750	3	10	1	379.75
C_AAF397003-tax	1451	499	4	1	3.62	680	3	46	1	362.75
C_AAT090003-tax	1239	445	4	1	3.64	577	3	43	1	309.75
C_AAD025002-tax	1041	318	4	1	3.57	502	3	19	1	260.25
C_AAW154003-tax	1007	417	4	1	3.79	490	3	14	1	251.75
C_AKJ313002-tax	977	403	4	1	3.79	477	3	12	1	244.25
C_AAD640002-tax	879	329	4	1	3.7	420	3	20	1	219.75
C_AKK397002-tax	863	286	4	1	3.62	416	3	16	1	215.75
C_AAC286002-tax	701	253	4	1	3.68	339	3	12	1	175.25
C_AAN560003-tax	515	214	4	1	3.8	253	3	5	1	128.75
C_AKJ644002-tax	509	121	4	1	3.41	242	3	13	1	127.25
C_AAE587002-tax	493	189	4	1	3.73	240	3	7	1	123.25
C_AAD170002-tax	451	175	4	1	3.74	221	3	5	1	112.75
C_AAC168002-tax	405	131	4	1	3.58	191	3	12	1	101.25

Comparisons

- Both defined as intensional ontologies
- UNSPSC
  - 16500 classes and only two properties.
  - 56 root classes and 14317 leaves, Percent leaves 88%
  - more wide than deep with an average depth of 4 and average width of 4125
  - a uniform maximum and minimum depth of 3.
  - root classes have all leaves at the same level
  - maximum width occurs at the maximum depth, equivalent to the number of leaves for the root class.
- ecl@ass
  - 76975 classes and 5527 properties
  - 25684 root classes and 51317 leaves, Percent leaves 66%
  - more wide than deep with an average depth of 3.4 and average width of 20526
  - **Unlike UNSPCS maximum width occurs not at the greatest depth but at depth 3 for all roots.**
  - **Like UNPSCS, the minimum width varies and always occurs at depth 1 for each root.**

Summary

- Many flavors of ontology evaluation or selection
  - Creating “candidate” set of ontologies for reuse with initial evaluation
  - Detailed analysis of “candidate” set using metric analysis
- OntoCat - one of numerous tools to address needs of ontology evaluation
  - Structural and size analysis
  - Both intensional and extensional
  - Root selection parameters
  - Root and hub summaries
  - Initial experiment with hub visualizations
- Experiments on numerous domains:
  - WordNet
  - UMLS vocabularies
  - E-commerce ontologies UNSPSC and ecl@ss

Possible Future Work

- 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