Ontology Research Group Overview

Dr. Valerie Cross

Sriram Ramakrishnan        En Yu
Ramanathan Somasundaram    Yi Sun

Miami University

OCWI C’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

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

Scientific America, May 2001

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

OCWIC 2007
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?
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
Ontology Dimensions

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

- a catalog
- a glossary
- a set of text files
- a thesaurus
- a collection of taxonomies
- a collection of frames
- a set of general logical axioms

Complexity: Without automated reasoning → With automated reasoning
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**
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.

Other roots for noun hierarchies:
- psychological_feature, abstraction,
- state, event, act, possession, group,
- phenomenon

Diagram:
- entity, physical thing
- object, physical object
- natural object
- plant part
  - plant organ
    - root
      - carrot
      - radish
  - mechanism
    - sprinkler
  - mechanical device
- enclosure
  - cage, coop
    - birdcage
    - hutch
    - squirrel cage
- artefact, artifact
- surface
  - skin

Solid lines = is-a relationship
Dashed lines = other children
Gene Ontology - BP

GO contains three structured, controlled vocabularies describing gene products: biological processes (BP), cellular components (C), and 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
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

<table>
<thead>
<tr>
<th>Size</th>
<th>owl:Thing</th>
<th>Apparel_and_Luggage_and_Personal_Care_Products</th>
<th>Building_and_Construction_and_Maintenance_Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total #Cls [iCnt(C)]</td>
<td>16500</td>
<td>254</td>
<td>87</td>
</tr>
<tr>
<td>Total #Property [iCnt(P)]</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total Roots [iCnt(Roots)]</td>
<td>56</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Total Leaves [iCnt(leaves)]</th>
<th>14317</th>
<th>219</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Leaves [iAv(leaves)]</td>
<td>0.86</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>Max Depth [iMaxDepth]</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Max Width [iMaxWidth]</td>
<td>14317</td>
<td>219</td>
<td>70</td>
</tr>
<tr>
<td>Depth of iMaxWidth</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average Width [iAvWidth]</td>
<td>4125</td>
<td>84.66</td>
<td>29</td>
</tr>
</tbody>
</table>

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
- J WordNet
- Test Corpus
- General Corpus
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

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

Interval rank (SpindleViz)

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

OCWIC 2007
Research Possibilities....

Semantic Similarity

• Develop a task/application-specific evaluation procedure
• Continue with mathematical analysis and properties
• Apply to 3D visualization
• 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 University’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