Tools and Resources that Build and Exploit Semantic Knowledge

Dan Moldovan
Lymba Corporation
Richardson, Texas

Resources, Tools and Applications of Semantic Knowledge

Resources
- XWN
- XWN - KB
- EventNet

Tools
- Semantic Parser – Polaris
- Semantic Calculus
- Ontology Builder – Jaguar

Applications
- Reasoning on KB
- KB-driven Question Answering
WordNet-based Resources

WordNet (http://wordnet.princeton.edu)
- Plain text definitions for English concepts
- {tennis, lawn tennis} (a game played with rackets by two or four players who hit a ball back and forth over a net that divides the court)

eXtended WordNet (http://xwn.hlt.utdallas.edu)
- Syntactic parses of WordNet's glosses
- Semantic disambiguation of content words (nouns, verbs, adjectives and adverbs) in glosses pointing to their synsets.
- Logical forms of glosses
  - tennis_NN_1(x1) ↔ game_NN_1(x1) & play_VB_1(e1,x6,x1) & with_IN(e1,x3) & racket_NN_4(x3) & by_IN(e1,x6) & two_JJ(x4,x6) & or_CC(x10,x4,x5) & four_JJ(x5,x6) & player_NN_1(x6) & hit_VB_1(e2,x6,x7) & ball_NN_1(x7) & back_and_forth_RB_1(x11,e2) & over_IN(e2,x8) & net_NN_5(x8) & divide_VB_3(e3,x8,x9) & court_NN_4(x9)
XWN-KB

- Semantically rich upper Knowledge Base based on WordNet
- Utilizes information from rich definitional glosses of WordNet concepts
- Associates generic slots (properties) for WordNet concepts
- Generates rich *World Knowledge Axioms* extracted using Semantic Patterns (Semantic Calculus) on WordNet glosses
- Features axioms that have been propagated up the WordNet hierarchy (in order to support generalization of axioms)

### Example Gloss

- Tennis is a game played with rackets by two or four players who hit a ball back and forth over a net that divides the court

- ISA (Tennis, game)
- AGT (two or four players, play)
- THM (game, play)
- INS (rackets, play)
- MEA (two or four, players)

- AGT (two or four players, hit)
- THM (a ball, hit)
- MNR (back and forth, hit)
- LOC (over a net that divides the court, hit)
- AGT (a net, divides)
- THM (the court, divides)
Semantic Cluster of a Gloss

<table>
<thead>
<tr>
<th>Synset ID: 00457626</th>
<th>Name: tennis, lawn_tennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>tennis</td>
<td>player</td>
</tr>
<tr>
<td>ISA game</td>
<td>MEA two or four</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>play</th>
<th>hit</th>
<th>divide</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT player</td>
<td>AGT player</td>
<td>AGT net</td>
</tr>
<tr>
<td>THM game</td>
<td>THM ball</td>
<td>THM court</td>
</tr>
<tr>
<td>INS racket</td>
<td>MNR back and forth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOC over a net</td>
<td></td>
</tr>
</tbody>
</table>

Synset ID: 00457626 Name: tennis, lawn_tennis

Axioms from XWN-KB

- World Knowledge Axioms from XWN-KB
  - Using specific Semantic Relation Types
    - Axioms are extracted from WordNet glosses
    - Aggregated on individual WordNet concepts
    - Finally, propagated up the WordNet hierarchy
  - Examples
    - A “player” can-
      - achieve reputation, play cards, steal ball, etc. (75 activities in WN)
    - A “terrorist group” can
      - use terror, oppose peace, dispense money, etc.
    - A “bank” can
      - issue currency, prepare periodic statement, provide credit, etc.
EventNet

- EventNet is a generalized network of events built on top of WordNet
- EventNet is represented by frames linked to the nominal and verbal WordNet event synsets and connected by probabilistic relations
- The frames include slots corresponding to different properties, states, roles, and relations that an event has, with different values and probabilities of occurrence in text
- Information on events, their properties, and how they are related is important for many applications

Event Net Definitions

- An event is something specific that happens somewhere at some time.
  - kill = cause to die; put to death, usually intentionally or knowingly
- The events/actions are defined and characterized by different properties. This man killed several people when he tried to rob a bank.
  - Actor
  - Undergoer
- The relations represent interactions/dependencies between events
  - This man killed several people when he tried to rob a bank.
  - Reason
  - Temporal-Overlapping Event
Knowledge Understanding and Organization

- Generalize knowledge (from roles to meta-roles, from concepts to classes, etc)
- Cluster properties, relations and values (features, states, roles, relations)
- Gather statistics from all properties
- Functional knowledge for the Killing [kill#v#1] event from all resources

<table>
<thead>
<tr>
<th>Functional Roles</th>
<th>XWN</th>
<th>Wikipedia</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor</td>
<td>person (3/6)</td>
<td>person (4/4)</td>
<td>person (7/10=0.7) substance(2/10=0.2) animal (1/10=0.1)</td>
</tr>
<tr>
<td></td>
<td>substance(2/6) animal(1/6)</td>
<td></td>
<td>animal (1/10=0.1)</td>
</tr>
<tr>
<td>Undergoer</td>
<td>animal(5/6)</td>
<td>person (6/8)</td>
<td>person (9/14=0.63) animal(5/14=0.36) cell (1/14=0.07)</td>
</tr>
<tr>
<td></td>
<td>person(2/8) cell(1/8)</td>
<td></td>
<td>cell (1/14=0.07)</td>
</tr>
<tr>
<td>Instrument</td>
<td>substance(2/3)</td>
<td>weapon (1/1)</td>
<td>substance (2/4=0.5) weapon (1/4=0.25) phenomenon (1/4=0.25)</td>
</tr>
<tr>
<td></td>
<td>phenomenon(1/3)</td>
<td></td>
<td>phenomenon (1/4=0.25)</td>
</tr>
</tbody>
</table>

EventNet Network for Kill – Definition View

Roles
- Person
- Animal
- Substance
- Phenomenon
- Weapon

Features
- Kill
- Verb

Properties
- Open
- Shot
- Without looking
- Crushing
- Gas chamber

Definition
- Cause to die, put to death, usually intentionally or knowingly

Relations
- Sub-class
- Sub-class
- Sub-class
- Sub-class
- Sub-class

Massacre
- Murder
- Kill_off

Manner
- 1.0
EventNet Network for *Kill* – Temporal View

**States**
- In progress
- Finished

**Properties**
- April 20, 1999, 11:19a.m.-12:05p.m.
- Unknown

**Relations**
- Shooting
- Passive
- Acquiring
- Injuring
- Moving
- Defending

EventNet Network for *Kill* – Causal View

**Factors**
- School bullying
- Outdoing other massacres
- Anniversary of event
- Planning
- Acquiring
- Shooting
- Taking antidepressant

**Reasons**
- Outdoing other massacres
- Anniversary of event
- Psychiatric treatment and medication

**Outcomes/Results**
- Injuries (0.63)
- Death (0.32)
- Emergency call (0.03)
- Rescue operations (0.03)
EventNet in Reasoning

- Generate **probabilistic axioms** for an event
- Generate logic forms for the axioms
- Axioms (and logic form representations) for killing:

### EventNet Temporal Succeeding Events:
- shooting (0.27)
- killing (0.26)
- passive (0.13)
- injuring (0.09)
- other

#### Axiom:
- There is a 0.27 probability for a killing to be followed by a shooting
- There is a 0.26 probability for a killing to be followed by another killing

#### Logic forms:
- \( \text{Kill}_V(e1, x1, x2) \rightarrow \text{Shoot}_V(e2, x1, x3) \) & TEMPORAL-AFTER(e1, e2, 0.27)
- \( \text{Kill}_V(e1, x1, x2) \rightarrow \text{Kill}_V(e3, x1, x4) \) & TEMPORAL-AFTER(e1, e3, 0.26)

### EventNet Causal Factors:
- planning (0.46)
- acquiring (0.46)
- shooting (0.02)
- other

#### Axiom:
- The acquisition of a weapon or planning to buy or build a weapon is very likely (45.58%) to be followed by a sequence of killings with those weapons

#### Logic from:
- \( \text{Acquire}_V(e4, x1, x2) \rightarrow \text{Kill}_V(e5, x1, x4) \) & INSTRUMENT(x2, e4) & CAUSE (e4, e5, 0.46)

EventNet in Hypotheses Generation

- Generate **hypotheses** for an event (What happen before, after, during an event or between two events?)
- Find **evidences** for the events or intermediate steps in the text collection
- What happened before the Virginia Tech. Massacre from April 16, 2007?

**EventNet: Causal Factors slot:**
1. Planning the attack (0.46)
2. Acquiring the instrument (0.25)
3. Shooting (0.02)

**Hypotheses:**
1. The killer planned the attack
2. The killer purchased guns
3. The killer shot the guns before

**Evidences from web:**
1. The suicide note video show him planning to shot people and himself
2. Cho **bought** one gun, a Walther P32, in February at a pawnshop on Main Street. Cho **purchased** a Glock 9 mm pistol here for $535, 30 rounds of ammunition included on March 12 at Roanoke Firearms about 40 miles away.
3. none
Applications of EventNet

- In QA answer questions related to events
  - What happen before or after an event?
  - What will happen next?
  - What happen between the two events? What is required to get from one point to another point?
  - What happen during an event?
- Find evidences for events that are linked through cause relations to prior events.
- EventNet can act as a Predictor
- EventNet can act as Hypothesis Generator
- EventNet can generate axioms

Semantic Parser - Polaris

- Extracts semantic relations between concepts in text.
- Relations between verbs and their arguments, between nouns, between nouns and modifiers, between clauses.
- Set of semantic relations needs to be defined carefully, (26 relations)
- Polaris uses a hybrid approach that includes machine learning and hand-coded rules
## List of Semantic Relations

<table>
<thead>
<tr>
<th>Code</th>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS</td>
<td>Possession</td>
<td>X is a possession of Y, Y owns/has X</td>
<td><img src="example1.png" alt="Example" /></td>
</tr>
<tr>
<td>PW</td>
<td>Part-Whole/ Meronymy</td>
<td>X is a part of Y</td>
<td><img src="example2.png" alt="Example" /></td>
</tr>
<tr>
<td>KIN</td>
<td>Kinship</td>
<td>X is a kinship of Y; X is related to Y by blood or by marriage</td>
<td><img src="example3.png" alt="Example" /></td>
</tr>
<tr>
<td>ASO</td>
<td>Association</td>
<td>X is associated with Y; X and Y can be people or groups</td>
<td><img src="example4.png" alt="Example" /></td>
</tr>
<tr>
<td>SRC</td>
<td>Source/Origin</td>
<td>X is the origin or previous location of Y</td>
<td><img src="example5.png" alt="Example" /></td>
</tr>
<tr>
<td>ISA</td>
<td>ISA</td>
<td>X is a (kind of) Y</td>
<td><img src="example6.png" alt="Example" /></td>
</tr>
<tr>
<td>SYN</td>
<td>Synonymy/ Name</td>
<td>X is a synonym/name/equal for/to Y</td>
<td><img src="example7.png" alt="Example" /></td>
</tr>
<tr>
<td>PRO</td>
<td>Property Type</td>
<td>X is a property type of Y</td>
<td><img src="example8.png" alt="Example" /></td>
</tr>
<tr>
<td>VAL</td>
<td>Property/ Attribute/ Value</td>
<td>X is a property/attribute/value of Y</td>
<td><img src="example9.png" alt="Example" /></td>
</tr>
</tbody>
</table>

## List of Semantic Relations

<table>
<thead>
<tr>
<th>Code</th>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNT</td>
<td>Quantification/ Extent</td>
<td>X is a quantification of Y; Y can be an entity or event</td>
<td><img src="example10.png" alt="Example" /></td>
</tr>
<tr>
<td>AGT</td>
<td>Agent</td>
<td>X is the agent for Y; Y is prototypically a person.</td>
<td><img src="example11.png" alt="Example" /></td>
</tr>
<tr>
<td>EXP</td>
<td>Experiencer</td>
<td>X is an experiencer of Y; involves cognition and senses; X is a person</td>
<td><img src="example12.png" alt="Example" /></td>
</tr>
<tr>
<td>INS</td>
<td>Instrument</td>
<td>X is an instrument in Y</td>
<td><img src="example13.png" alt="Example" /></td>
</tr>
<tr>
<td>THM</td>
<td>Theme/Patient/ Result/Consumed</td>
<td>X is the theme/patient/result/consumed in/from/of Y</td>
<td><img src="example14.png" alt="Example" /></td>
</tr>
<tr>
<td>RCP</td>
<td>Recipient/ Receiver</td>
<td>X is the recipient of Y; X is an animated entity</td>
<td><img src="example15.png" alt="Example" /></td>
</tr>
<tr>
<td>TPC</td>
<td>Topic/ Content</td>
<td>X is the topic/focus of cognitive communication Y</td>
<td><img src="example16.png" alt="Example" /></td>
</tr>
<tr>
<td>INT</td>
<td>Intent</td>
<td>X is the intent/goal/reason of Y</td>
<td><img src="example17.png" alt="Example" /></td>
</tr>
</tbody>
</table>
List of Semantic Relations

<table>
<thead>
<tr>
<th>Code</th>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>STI</td>
<td>Stimulus</td>
<td>X is the stimulus of Y; Perceived thorough senses</td>
<td>[XY] [YX] Mary [heard] [the train] while [smelling] [the roses].</td>
</tr>
<tr>
<td>MNR</td>
<td>Manner</td>
<td>X is the manner in which Y happens</td>
<td>[XY] John [head] [carefully]; [ran] [quickly]; [spoke] [bably]</td>
</tr>
<tr>
<td>LOC</td>
<td>Location/Direction/Source/Path/Goal</td>
<td>X is the location of Y or where Y take place</td>
<td>[XY] There is [a cat] on [the roof]; [YX] The hurricane [passes] through [Galveston].</td>
</tr>
<tr>
<td>TMP</td>
<td>Temporal</td>
<td>X is the time of Y (when Y take place)</td>
<td>[XY] John [woke up] at [noon]</td>
</tr>
<tr>
<td>CAU</td>
<td>Cause</td>
<td>X causes Y; X and Y are events, states</td>
<td>[XY] [Drinking] causes [accidents].</td>
</tr>
<tr>
<td>MAK</td>
<td>Make-Produce</td>
<td>X makes Y</td>
<td>[XY] [GM] manufactures [cars].</td>
</tr>
<tr>
<td>JST</td>
<td>Reason/Justification</td>
<td>X is the reason/motive/justification for Y</td>
<td>[XY] [The severity of the crime] justifies [the harsh sentence]; [YX] [He is innocent] by reason of [insanity]</td>
</tr>
<tr>
<td>PRP</td>
<td>Purpose</td>
<td>X is the purpose for Y; Y did something because this person wanted X</td>
<td>[XY] John [swims] for [fun]; Mary [works] part-time [to earn some extra money]</td>
</tr>
<tr>
<td>FL</td>
<td>Influence</td>
<td>X caused something to happen to Y</td>
<td>[XY] [The war] had an impact on [the Economy]</td>
</tr>
</tbody>
</table>

Example

John went to the park yesterday because he saw hot air balloons taking off from there.

Agent (John, went)
Location (to the park, went)
Temporal (yesterday, went)
Cause (saw, went)
Experiencer (He, saw)
Stimulus (hot air balloons taking off from there, saw)
Property (hot, air)
Part-Whole (hot air balloons, balloons)
In-A (hot air balloons, balloons)
Agent (hot air balloons, taking off)
Location (from there, taking off)
Lymba’s Semantic Calculus Increases Knowledge

Document Collection → Semantic parser → Lymba Semantic Calculus → Smarter Semantic Store

Example Ontology Stats:
- 8726 concepts, 5148 relations
- ~2 million semantic triples
- RDFS rules
- ~3.1 million semantic triples
- 7 Lymba Semantic Calculus rules
- ~5.2 million semantic triples

26 Generic Semantic Relations

More than 60 rules

- transitive
  PW(X,Y), PW(Y,Z) → PW(X,Z)
- reflexive
  KIN(X,Y) → KIN(Y,X)
- special combinations
  AGT(X,Y), LOC(Y,Z) → LOC(X,Z)
  IFL(X,Y), PW(Y,Z) → IFL(X,Z)

Building Knowledge Bases from Text

Documents → Seeds → Ontology (structured knowledge) → Jaguar – KAT Ontology Builder

Ontology + pointers to text → Knowledge Base (ontology + contextual knowledge + pointers to text)

Functionality

1. Produce ontologies
2. Link concepts & relations to text
3. Visualize ontology
4. Edit ontology
5. Enhance an existing ontology
6. Merge two ontologies into a consistent ontology
7. Ontological search of documents (search documents using ontology)
Types of Knowledge

- **Universal (or ontological)**
  - Represented in Hierarchies
  - Simple binary relations between concepts
  - “It has always been a thoroughly commercial cinema, concentrating on crowd pleasing genres, like comedy and….”

- **Contextual**
  - Represented in individual (semantic) contexts
  - Groups of relations centered on a common concept
  - “Pixar will release a large budget movie in May for….”
Knowledge Bases

- Ontology/KB creation overview
  - Knowledge Extraction from Text
    - Pattern recognition, Semantic Parsing
  - Knowledge Representation and Storage
    - Contextual vs. Universal
    - RDF Database, OWL/RDF, XML, Relational Database
  - Knowledge Base Maintenance
    - Conflict Resolution
    - Ontology Merging
    - User Interaction- Ontology Editing

Jaguar Modules – Oracle SES Ontology Creation

Text Processing

Input: Documents, Seeds
- Proprietary seed expansion algorithm
- Extract “concepts” of interest
- Extract binary relations (universal)
- Use Semantic Parser to obtain contextual knowledge

Output: Concepts, Contexts, Binary Relations

Input Text Examples:
- All Date attributes use the format mm/dd/yyyy.
- “date attribute” ISA “attribute”
- “mm/dd/yyyy” ISA “format”
- “mm/dd/yyyy” PRO “date attribute”
Jaguar Modules – Oracle SES Ontology Creation

**Classification/ Hierarchy Creation**

*Input: Concepts, Binary Relations*
- Use WordNet as the upper ontology
- Classify each concept against every other using defined procedures, obtaining set of ISA relations
- Add all ISA and other binary relations to the hierarchy using *conflict resolution*

*Output: Hierarchy of relations*

---

**Knowledge Base Maintenance**

- Knowledge Base Merging
- Visualization
- Knowledge Base Editing
  - User Interaction
  - Modifications
Reasoning on Knowledge Bases

- A Knowledge Base combined with a semantic model is Semantic Calculus enables reasoning via smart "Semantic Queries".

Semantic KB

Semantic Calculus

KB Reasoner

Document Collection

Semantic Parser

Ontology Builder

Ontology

?? Semantic Query ??:
- "What are the properties of a document?"
- PROPERTY(X, document)

Semantic KB

EventNet

XWN KB

Querying the Knowledge Base

?? Part-Whole(X, cluster) ??

- Part-Whole(X, Y), Part-Whole(Y, Z) -> Part-Whole(X, Z)
- Part-Whole(X, Y), IS-A(Z, X) -> Part-Whole(Z, Y)
  - Part-Whole(application, cluster)
  - Part-Whole(data, cluster)
  - Part-Whole(plug-in, cluster)
  - Part-Whole(hummingbird, cluster)
Motivation for KB Driven Question Answering

- Few if any semantic models (and the tools that implement those models) can accurately and completely capture all the information in text.
- KB’s and ontologies that are created from sources other than the underlying document collection may need to be mapped to the concepts in the collection.
- For search applications a mix of information retrieval and semantic querying from a KB is required to robustly and accurately satisfy the needs of an end user.
- KB Driven Question Answering combines an information retrieval approach with semantic querying of a Knowledge Base.

KB Driven Question Answering

- Example: What component encrypts communication?

  - component
    - business component
    - communication channel
    - connector
    - external component
    - oracle content server semicomponent
    - oracle Secure Enterprise Search component
    - oracle Secure Enterprise Search SEF component
    - path component
    - query component
    - secure enterprise search component
    - ssl component
      - administration tool
      - crawler
      - java ssl
      - oracle SES Command Line tool
      - securing component
      - authentication methods component
      - authorization and authentication
      - password
      - SSL
KB Driven Question Answering

ISA(X, component) AND (encrypt OR protect) AND (communication)

X=SSL

- The KB guides the question answering system but does not restrict it, allowing for a robust implementation.

Semantic Knowledge in Question Answering

NLP Tools

Search Engine

Semantic Index

Question Processing

Passage Retrieval

Answer Processing

Answer

- Semantic DB

- Query Formulation

- Keyword Expansion

- Answer Likelihood

- Answer Candidate Selection

- Answer Likelihood

- Answer Redundancy
Combining KB’s and Question Answering

- Document Collection
- Semantic Parser
- Ontology Builder
- Semantic Calculus
- Ontology
- IR Index
- Question Answering
- KB Reasoner
- Semantic KB
  - EventNet
  - XWN KB

Summary

- WordNet is a good start for developing Semantic Knowledge.
- Knowledge can be extracted from text automatically.
- Ontology building process can be automated.
- Semantic technologies is a fruitful area of research that opens many applications, and new businesses.
- Many (hundreds) commercial companies are interested in Semantic Technologies.