OntoModeler - Semi-Automated Method for Constructing Domain Specific Ontology

By: William L. Sousan

KEWI – Knowledge Engineering & Web Intelligence Group – UNO
http://kewi.unomaha.edu
**Goals**

- Create usable ontology from text corpus using ontology construction tool
- User interacts with ontology construction tool - incrementally builds and refines model
- Semi-automated learning process - tool learns the user's goals
- Maintains and updates user's preferences in an ontology meta-model (Goal Tree)
- Presently using cyber-attack domain for evaluation
Cyberattack Domain – SPEC

- Characterize Cyberattack events by their SPEC – Social, Political, Economical, Cultural motivations and attack characteristics
- Common set of concepts that makeup attacks such as Attacker, Victim, Consequences, etc…
- Determine what types of entities under each concept typically occur for a given SPEC motivation – presents need for Ontology
Upper-Level Ontology

- Develop Upper-Level/Seed Ontology
- Use OntoModeler to extend and refine this ontology
OntoModeler Process – Part 1

- Build Domain-Specific collection of cyberattack events described in news reports and information reports
- Use NLP tools to extract Noun & Verb Phrases, and Subject-Predicate-Object (SPO) Sequences
- Create initial Upper-Level Seed Ontology where each concept has a least 1 Term
- Present ranked lists of SPO’s to user ranked by semantic relevance of either subject or object term to each seed ontology concept
### Example Ranked SPO List by Object

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Semantic Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Indian Government</td>
<td>Set</td>
<td>Information</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Intruder</strong></td>
<td><strong>Steal</strong></td>
<td><strong>valuable Information</strong></td>
<td>1.00</td>
</tr>
<tr>
<td>Hacker</td>
<td>Steal</td>
<td>Information</td>
<td>1.00</td>
</tr>
<tr>
<td>computer System</td>
<td>Steal</td>
<td>credit card Information</td>
<td>1.00</td>
</tr>
<tr>
<td>recent cyber Attack</td>
<td>Target</td>
<td>critical Information</td>
<td>1.00</td>
</tr>
<tr>
<td>famous Inability</td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Hacker</strong></td>
<td></td>
<td></td>
<td>0.89</td>
</tr>
<tr>
<td>Outsider</td>
<td></td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>Datum</td>
<td>Include</td>
<td>customers' name Credit</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Sachs</td>
<td>Note</td>
<td>Directive</td>
<td></td>
</tr>
<tr>
<td><strong>Hacker</strong></td>
<td></td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>cyber Attack</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hacker</strong></td>
<td></td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td>Attack</td>
<td>Target</td>
<td>gmail Account</td>
<td>0.45</td>
</tr>
<tr>
<td>Employee</td>
<td>Begin</td>
<td>well-crafted e-mail Message</td>
<td>-0.45</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SPO’s in Blue** are patterns for finding Attack-Agents in Subject Slot

**Ranked By Comparing Object to Concept of “Information”**

**Sorted By Semantic Relevance of Object Term to “Information Concept”**
OntoModeler Process – Part 2

- User select relevant SPO’s to create Term Extraction Patterns (TEP’s) for extracting relevant terms for a given seed concept
- Term Pools (TP/s) are created from the user of the TEP’s
- TP’s are then conceptualized by untangling the terms in into concepts and arranging them into an Hierarchical structure
- This process is performed iteratively to build and refine the ontology
Structure - Process

1. Define Seed Concepts
   - Cyberattack
   - Attack-Agent
   - Attack-Victim
   - Attack-Consequences

   1. Define Seed Concepts

2. Build TEP List
   - Table of TEP’s
     - `<*>` = Attack-Agent
     - `<*>` => Steals => Information
     - `<*>` => Launch => Attack
     - `<*>` => Target => Information
     - ...

3. Create TP’s from TEP’s

4. Conceptualize Term Pool
   - Term Pool
     - Hacker
     - Hacktivist
     - Intruder
     - ...

   - Table of TEP’s
     - `<*>` = Attack-Agent
     - `<*>` => Steals => Information
     - `<*>` => Launch => Attack
     - `<*>` => Target => Information
     - ...

   - Term Pool
     - Hacker
     - Hacktivist
     - Intruder
     - ...

KEWI Group - University of Nebraska,
Omaha
Current Experiments

- Created seed ontology consisting of 3 concepts – Attack-Agent, Attack-Victim, and Attack-Consequences
- 145 documents => extracted 11,239 noun phrases, 639 verb phrases, and 3,538 SPO’s.
- Created and used TEP’s to create TP’s and conceptualized each of the 3 TP’s into a single ontology
- However, we only performed one iteration of the process – future work to add support for multiple iterations
Example Output
Preliminary Evaluation

- Some terms/concepts are not relevant to their respective seed concepts
- Cycles exist in the graph as a result of problems within the hierarchy creation
- Some concepts that are separate should be semantically linked
- Current Prototype only performs single iteration, need implement the iterative refine process