Expanding an Oil Spill Semantic Relations Taxonomy with a Foreign Policy Semantic Relations Taxonomy

Yejun Wu
School of Library and Information Science
Louisiana State University
267 Coates Hall, Baton Rouge, LA 70803
wuyj@lsu.edu

Li Yang
School of Computer Science
Southwest Petroleum University
Chengdu, Sichuan, China 610500
yangli0027@163.com

ABSTRACT
This paper presents the significance, method and findings of expanding an oil spill semantic relations taxonomy with the CAMEO foreign policy semantic relations taxonomy for supporting knowledge discovery. The oil spill relations taxonomy is enriched by integrating the CAMEO relations taxonomy. Context scope notes may need to be added to CAMEO relation terms (especially the polysemous terms) when they are merged into the oil spill relations taxonomy. A term may be classified into different categories due to different context.

Keywords
Semantic relations, taxonomy, CAMEO, oil spill.

INTRODUCTION
Human beings are naturally interested in relations between entities, such as the influence of climate change on the earth and the impact of the 2010 Gulf of Mexico Deepwater Horizon oil spill incident on coastal states. Semantic relations between entities are expressed using verb phrases. Since there are a large number of entities in the world and a large number of relations between the entities, it can be difficult to define all the possible relations between all the entities in the world. However, people in a particular domain constantly discuss topics in the domain and their relations. Consequently, there are domain-specific entity ontologies (such as the ASIS&T Thesaurus of Information Science, Technology, and Librarianship) and domain-specific semantic relations taxonomies (such as the Unified Medical Language System (UMLS) Semantic Network).

The fundamental goal of this study is to develop a comprehensive cross-domain semantic relations taxonomy for supporting knowledge discovery through inference. We have developed a semantic relations taxonomy in the interdisciplinary oil spill domain (Wu and Yang, 2015a). The taxonomy has been augmented by integrating the OBO Relation Ontology (OBO, 2002), which is in the biological and biomedical domain. The ontology is topically related to the oil spill domain. A feasible approach to developing a comprehensive semantic relations taxonomy is to expand the oil spill semantic relations taxonomy by integrating semantic relations taxonomies from other domains. One of such taxonomies is the Conflict and Mediation Event Observation (CAMEO) relations taxonomy in the foreign policy or international relation domain (Schrodt et al., 2008). International relation is generally not related to the oil spill domain. The objective of this study is to expand the oil spill relations taxonomy with the CAMEO relations taxonomy, and to learn some interesting lessons.

SIGNIFICANCE OF THE STUDY
Semantic relations are useful in information retrieval, question answering, knowledge organization (such as ontology construction), and knowledge discovery. Bertaud et al. (2007) found that using verbs (i.e., to show, to confirm) in MEDLINE queries can improve information retrieval effectiveness. Wang et al. (1985) proved semantic relations valuable in question-answering. Ontologies represent entities and their relations, so semantic relations are an indispensable part of ontology development. Semantic relations also facilitate knowledge discovery through inference. Swanson and Smalheiser (1999) discovered many implicit relationships within the biomedical literature. For example, if one paper reports that substance A causes disease B and another reports that disease B may result in disease C, then it can be inferred that substance A might result in disease C. A semantic relations taxonomy may support inference of relations between entities through specified patterns of relation chains.

A comprehensive semantic relations taxonomy is expected to be useful in information retrieval, question answering, ontology development, and knowledge discovery in many domains and cross-domains. The method and lessons...
learned from this study can also be useful to build cross-domain semantic relations taxonomies.

**METHODOLOGY**

The oil spill semantic relations taxonomy is a four-level taxonomy of 898 verb phrases which were extracted from over 300 oil spill related documents (Wu, 2013). The taxonomy was developed using a combination of top-down and bottom-up approach. Various taxonomic and linguistic resources were used during the development of the taxonomy. The top two levels of the taxonomy was built using Spradley (1979)’s nine categories of universal semantic relations, Green (1996)’s 28 frames, Hjorland’s (2007) list of important semantic relations, UMLS’s five-subcategories of semantic relations, Levin (1993)’s 193 verb classes and FrameNet’s 230 frames (Baker et al., 1990). The top two levels of the taxonomy was revised during bottom-up clustering of verb phrases.

The oil spill semantic relations taxonomy was augmented by integrating the OBO Relation Ontology (OBO, 2002), which is a list of 397 relation terms in the biological and biomedical domain. The oil spill relations taxonomy has some biological and biomedical relation terms, but their scope is broader and shallower than those in OBO. OBO terms are very often deeper and detailed, such as be downstream of sequence of, be attached to part of. The oil spill relations taxonomy is found to be fully expandable and has high degree of durability because only a small number of categories was revised and added (Wu and Yang, 2015b). The augmented oil spill relations taxonomy has the following top 10 categories: strict inclusion, possession, spatial relationship, sequence, equivalence/comparability, association, reliance, impact, action, and feature/function.

The CAMEO relations taxonomy has 2,168 verbs in the following 20 categories (Schrodet et al., 2008): (01) make public statement, (02) appeal, (03) express intent to cooperate, (04) consult, (05) engage in diplomatic cooperation, (06) engage in material cooperation, (07) provide aid, (08) yield, (09) investigate, (10) demand, (11) disapprove, (12) reject, (13) threaten, (14) protest, (15) exhibit military posture, (16) reduce relations, (17) coerce, (18) assault, (19) fight, (20) attack with weapons of mass destructions. A total of 333 non-verbs (such as gunfire, confidence, collision), verbs not in original form (such as making, wrapped up), and erroneous words (such as satify, reassesment) were removed. The remaining 1,835 verb phrases were classified into the augmented oil spill relations taxonomy.

We can see that the two taxonomies have very different categories – the top categories in the oil spill relations taxonomy are broad whereas CAMEO’s top categories are specific. Occasionally a CAMEO category can be merged into a sub-category of the augmented oil spill relations taxonomy directly. Very often CAMEO categories cannot be merged into the augmented oil spill relations taxonomy directly because the category labels reflect the context of international relations. Instead, the CAMEO relation terms are very often classified into multiple categories of the oil spill relations taxonomy. When classifying a relation term into a category, the nature of the relation term in its application context of “Topic A <relation term> Topic B” is considered. It is unknown whether facet analysis of relation terms should be performed because S. J. Ranganathan’s five facets (i.e., personality, Matter, Energy, Space, and Time) does not seem to apply to semantic relations (Wu and Yang, 2015a).

**PRELIMINARY FINDINGS**

Table 1 shows the distribution of CAMEO relation terms in the expanded taxonomy. The same 10 top categories of the augmented oil spill relations taxonomy absorbs the CAMEO taxonomy with small addition and revision of categories, and also provides an overview of the CAMEO’s 20 categories. It indicates that more than 95% of the terms are classified into three categories: sequence, impact, and action. The major sub-categories under the three categories reveals the major categories of relations in the international relation domain, that is, evaluation, movement, implementation, influence, and chronological and procedural sequence.

<table>
<thead>
<tr>
<th>Category</th>
<th># of terms</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01. Strict Inclusion</td>
<td>16</td>
<td>0.9%</td>
</tr>
<tr>
<td>02. Possession</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>03. Spatial relationship</td>
<td>6</td>
<td>0.3%</td>
</tr>
<tr>
<td>04. Sequence</td>
<td>82</td>
<td>4.5%</td>
</tr>
<tr>
<td>Chronical sequence</td>
<td>(35)</td>
<td></td>
</tr>
<tr>
<td>Procedural sequence</td>
<td>(32)</td>
<td></td>
</tr>
<tr>
<td>05. Equivalence/Comparability</td>
<td>19</td>
<td>1.0%</td>
</tr>
<tr>
<td>06. Association</td>
<td>8</td>
<td>0.4%</td>
</tr>
<tr>
<td>07. Reliance</td>
<td>1</td>
<td>0.1%</td>
</tr>
<tr>
<td>08. Impact</td>
<td>540</td>
<td>29%</td>
</tr>
<tr>
<td>Influence</td>
<td>(519)</td>
<td></td>
</tr>
<tr>
<td>Cause-effect</td>
<td>(21)</td>
<td></td>
</tr>
<tr>
<td>09. Action</td>
<td>1,129</td>
<td>62%</td>
</tr>
<tr>
<td>Movement</td>
<td>(189)</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>(290)</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>(329)</td>
<td></td>
</tr>
<tr>
<td>10. Feature/Function</td>
<td>34</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

Table 1. Distribution of CAMEO relation terms in expanded taxonomy.

Since the international relation domain is generally complementary to the oil spill domain, the major benefit of using CAMEO to expand the augmented oil spill relations taxonomy is enriching the oil spill taxonomy. For example, there is a sub-category of “treatment” (of spilled oil) in the oil spill taxonomy, which includes physical treatment, repairment, medical treatment, biological and chemical treatment. CAMEO contributes political treatment, social treatment, and military treatment to that category, thus
making it more complete. For another example, CAMEO has a “maintain” category, which contains conserve, preserve, save, and spare. This category is merged into the augmented oil spill relations taxonomy directly, making it more complete.

Both CAMEO relations taxonomy and the oil spill relations taxonomy are contextualized classification schemes. The major challenge of merging relation terms of one domain into a relations taxonomy of another domain is the different context of the terms. Context scope notes may need to be added to the terms (especially the polysemous terms) to help specify their meanings when they are merged into a taxonomy of a different domain and when they have different contextual meanings. As the result, a term can be classified into different categories due to different context. For example, “burn” in international relations is mainly a destructive action, but is a constructive action in cleaning spilled oil. In the augmented oil spill relations taxonomy, “burn” is classified under “treatment” since it is an approach to cleaning spilled oil. The destructive sense of “burn” can be classified into action – destructive action. However, blindly adding scope notes to all terms can make the merged taxonomy difficult to read. Therefore, scope notes should be added when necessary.

**SUMMARY AND FUTURE WORK**

The CAMEO relations taxonomy was used to expand the oil spill relations taxonomy. As the result, the oil spill relations taxonomy is enriched and becomes more complete. Since both taxonomies are contextualized classifications, context scope notes may need to be added to CAMEO relation terms when they are merged into the oil spill relations taxonomy. A term may be classified into different categories due to different context. In the future, facet analysis of relation terms and functional evaluation of comprehensive relations taxonomies for supporting knowledge discovery remain interesting topics.

**REFERENCES**


