

# Developing a Taxonomy of Semantic Relations in the Oil Spill Domain for Knowledge Discovery

Yejun Wu  
Associate Professor  
School of Library and Information Science  
Louisiana State University  
wuyj@lsu.edu

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

## Abstract

The paper presents the rationale, significance, method and procedure of building a taxonomy of semantic relations in the oil spill domain for supporting knowledge discovery through inference. Difficult problems during the development of the taxonomy are discussed and partial solutions are proposed. A preliminary functional evaluation of the taxonomy for supporting knowledge discovery was performed. The study proposes more research problems than solutions.

## Introduction

Human beings are naturally interested in semantic relations between entities, such as the influence of diabetes on human health, the impact of the 2008 financial crisis on global economy, and the impact of the 2010 Gulf of Mexico Oil Spill Incident on coastal states. Semantic relations between entities are usually represented as verb phrases. People in different domains tend to be interested in different topics and their relations. For instance, economists discuss economic events (e.g., the end of quantitative easing may *raise* interest rates), and medical professionals care about drugs and diseases (e.g., a drug is used to *treat* a disease).

The goal of this study is to develop a three-to-four-level taxonomy of semantic relations in the oil spill domain for knowledge discovery purpose (Wu and Yang, 2015). The reasons why the oil spill domain is selected are two-fold. One, the 2010 Gulf of Mexico Oil Spill Incident (White House, 2012) has impacted many aspects of the coastal environment of the Gulf of Mexico and the people living in the coastal states. Government officials, Gulf-based researchers and the general public wanted to get a general understanding of the impact. The other, an oil spill topic map was created to help people understand the impact (Wu and Dunaway, 2013). About 5,000 entity-relationship tuples have been collected from oil spill related literature (Wu, 2013), and can be the appropriate data for this study. A knowledge discovery system that facilitates inference of impacts through chains of semantic relations is desired. A three-to-four-level taxonomy of semantic relations is expected to be fine-grained enough to support knowledge discovery through inference. The top-level taxonomy of semantic relations is expected to be complete and universal so that it can be useful to other domains.

## Significance of the Study

Semantic relations have many applications in information retrieval, question answering, and knowledge organization (such as ontology construction). Bertaud et al. (2007) found that using verbs (i.e., *to show*, *to confirm*) in MEDLINE (the National Library of Medicine premier bibliographic database) queries can improve the retrieval of findings. Green (1996) identified an inventory of 26 basic relations structured by investigating the general relationships underlying the 1,250+ verbs, and hypothesized that frame-based index should have the potential of contributing to precision and recall. Semantic relations have proved valuable in question-answering (Wang et al., 1985). Ontologies represent entities and their relations, so semantic relations are an important part of ontology development.

Semantic relations also facilitate knowledge discovery through inference. Swanson and Smalheiser (1999) discovered numerous undiscovered implicit relationships within the biomedical literature. For example, if one article reports that *substance A causes disease B* and another reports that *disease B causes disease C*, then we can infer that *substance A might cause disease C*. Semantic relations facilitate the grouping of relations and support inference of relations through specified patterns of relation chains.

The taxonomy of the oil spill domain is expected to be useful to support information retrieval, question answering, and knowledge discovery in this domain. The method and lessons learned from this study can also be useful to build semantic relations taxonomies in other domains.

## Theoretical and Practical Background

There are two types of semantic relations: (1) relations between concepts, senses or meanings, and (2) relations between words, terms, and expressions or signs that are used to express the concepts (such as synonyms, homonyms, and BT/NT/RT in thesauri) (Hjørland, 2007). It is common to mix both kinds of relations, and this study does not plan to distinguish these two types of relations. This study focuses on the relations between entities that are expressed as verb phrases, therefore verb classes are highly relevant.

Levin's verb classes and FrameNet's frames are two comprehensive verb classification schemes. The grouping of Levin's 193 verb classes is based on argument syntax whereas the grouping of FrameNet's 230 semantic frames is based on lexical semantics (Baker and Ruppenhofer, 2002). Both schemes provide useful resources for this study. FrameNet classifies predicates into frames based on a shared semantics, whereas in Levin's verb classes, predicates belong to classes based on same syntactic behavior (alternation patterns) that make some semantic sense (Baker and Ruppenhofer, 2002), therefore FrameNet is more useful to develop the semantic relation taxonomy in this study. For example, in Levin's verb classes, "ameliorate" and "americanize" are in the same class (Levin, 1993; Lawler, 2015). Such a grouping does not support inference of semantic relations between entities. However, Levin's verb classes are still useful resource for the development of the semantic relation taxonomy in this study.

Green (1996) developed an inventory of 28 general relational structures after investigating 1,250+ verbs. The inventory is expressed as frames in eight groups. One example group is action. Another example group is link hierarchy, comparison, whole-

part, balance, and path. The grouping of frames provides a useful model for this study even though each group does not have a category label. At an abstract level, Spradley (1979) proposes nine types of universal semantic relationships for conducting domain analysis in ethnographic studies: strict inclusion, spatial, cause-effect, rationale, location for action, function, mean-end, sequence, and attribution. The nine types of relationships provide a good foundation for developing the top-level taxonomy in this study.

In addition to the studies of general semantic relations, there are verb lists in specific domains. For example, Broom's taxonomy of action verbs classifies verbs in six categories of cognitive activities: knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom et al. 1956). The Unified Medical Language System (UMLS) Semantic Network defines 54 semantic relations in two big categories (i.e., is a, associated with) and five sub-categories (i.e., physically related to, spatially related to, functionally related to, temporarily related to, conceptually related to) (UMLS 2013). The Open Biological and Biomedical Ontology (OBO) Foundry provides an OBO relation ontology, which is a list of 385 verbs in the biological and biomedical domain (OBO, 2002; Xiang et al., 2011).

## **Methodology**

We have collected 898 verb phrases from about 5,000 entity-relationship tuples that were extracted from over 300 oil spill related documents (Wu, 2013). The goal of the study is to develop a three-to-four-level taxonomy of semantic relations in this domain for supporting knowledge discovery. A combination of top-down and bottom-up approach is used to develop the taxonomy since it is the best practice in taxonomy construction as discussed in knowledge organization literature (Wang, Chaudhry, and Khoo, 2010; Ramos and Rasmus, 2003; Cisco and Jackson, 2005; Holgate, 2004). A bottom-up approach builds up important categories from the concepts that are extracted from source content. Automated technologies such as concept extraction and clustering can automate bottom-up analysis (Ramos and Rasmus, 2003), but offers little control over the meaning and arrangement of higher-level categories (Cisco and Jackson, 2005). A top-down approach starts at the general, conceptual levels, and establishes a general framework for the taxonomy based on the objectives of the taxonomy (Ramos and Rasmus, 2003). Therefore, it offers control over the top and higher level categories of the taxonomy (Cisco and Jackson, 2005). A combination of the top-down and bottom-up approach develops the higher level categories in the taxonomy first, classifies semantic relation terms into lower-level categories, and refines the lower-level categories according to the constraints of the higher level categories. The higher-level categories can also be adjusted and refined according to the need of governing the lower-level categories.

Various taxonomic and linguistic resources were used during the development of the taxonomy. Levin's verb classes and FrameNet provide a good foundation for verb classification and clustering. WordNet contains over 21,000 verb word forms and approximately 84,000 word meanings (Fellbaum, 1990), which is also useful linguistic resource for this task.

The top level of the taxonomy was initially built using Spradley's nine categories of universal semantic relations, Green's eight groups of frames, and Hjørland's (2007) list

of important semantic relations. The top level was adjusted when the second and third levels were developed.

The second level of the taxonomy was initially built using Green's 28 frames, UMLS' five sub-categories, FrameNet's 230 frames, and Levin's 193 verb classes. The second level was revised during bottom-up clustering of verb phrases. Clustering the verb phrases based on synonymy without the guidance of higher level categories proved to be unsuccessful.

The bottom level (i.e., the third and occasionally the fourth level) is composed of lists of verb phrases under each second-level category, just like UMLS's bottom level verb phrases. The verb phrases under each second-level category should have some degree of shared semantics or synonymy. FrameNet, Levin's verb classes, and WordNet are all helpful resources to classify the verb phrases. Since people would like to know the impact of the 2010 Gulf of Mexico Oil Spill Incident, verb phrases that represent impact is a focus of the taxonomy. Occasionally a fourth level can occur when there is a need. The following procedure describes the specific steps of the development process.

### **Procedure**

Some best practices and guidelines for taxonomy design are introduced in the literature (Ramos and Rasmus, 2003; Cisco and Jackson, 2005; Lambe 2007; Hedden, 2010). Those guidelines were referenced before and during the development of the Oil Spill Relation Taxonomy, and the following procedure was developed and followed.

- Step 1: Normalizing all the verb phrases by converting them to their original forms.
- Step 2: Cluster the verb phrases based on synonymy of terms. This step generates the preliminary bottom-level categories. 15 big clusters were built for the 896 verb phrases. There is an "all other" cluster that contains orphans or singletons that do not belong to any of the 14 specific clusters.
- Step 3: Consult taxonomic and linguistic resources relevant to verbs and semantic relations (such as FrameNet, Levin's verb classes, WordNet, and dictionaries), build a preliminary taxonomy with one or two top-level categories using a top-down approach.
- Step 4: Load the clusters, one by one, into the preliminary taxonomy with one or two-level categories. Build middle level categories using a combination of bottom-up and top-down approach. Consult the dictionaries, taxonomic and linguistic resources when needed. This is a muddy middle game, and is an iterative process.
- Step 5: Audit the categories from a top-down perspective, adjust (i.e., split, merge, revise, add) the categories when necessary. Each sub-category of a category is a facet of that category. Maximum mutual exclusiveness between sub-categories and between categories is pursued during this process.

The outcome of the procedure is the preliminary taxonomy. The taxonomy with major categories and a couple of instances under most bottom-level categories is provided in the Appendix.

## Difficult Problems and Partial Solutions

Various difficult scenarios were encountered during the development process. Three major difficult problems with our partial solutions are discussed below although no perfect solutions are suggested. The purpose of the discussion is to initiate more discussion and study of these problems instead of drawing conclusions by offering solutions to the problems.

The first is the muddy middle game in building middle level categories, which is rarely discussed in the literature. The problem happens when a relation term is given but no lower-level category is available or appropriate, therefore a new bottom-level and very likely a middle-level category needs to be created, which requires creative and logic thinking. However, sometimes, it can be really difficult to figure out what category a relation term belongs to. For example, when “*be subject of*” was given, we could not figure out an appropriate bottom-level and middle-level category for it. We put it aside until “*be about*” was encountered. This indicates that, when there is no category available for a term, clustering can be delayed until more synonymous terms are encountered, then a cluster may emerge easily. However, clustering is a bottom-up approach which does not guarantee deterministic categories. This may cause fluidity or instability of bottom-level and middle-level categories.

The second is the possible inconsistency between local validity and global validity due to contextual or partial membership. A term can be a member of a lower-level category partially or contextually. The membership or classification has local validity. Partial membership is a classification based on partially overlapped semantics. Contextual membership is a classification based on a certain context. A term can belong to a lower-level category partially or contextually, and a lower-level category can belong to a higher-level category partially or contextually. However, the term may not be classified into the higher-level category because the context has changed or the overlap of semantics is lost during the transitivity of membership or classification. When this happens, the membership does not have global validity. Figure 1 describes the loss of membership due to partially overlapped semantics during the transitivity of partial membership. Term C partially belongs to category B, B partially belongs to category A, but C does not belong to A. Polysemous and homonomous terms can also contribute to contextual and partial membership due to their partially overlapped or non-overlapped semantics. Semantic analysis of the terms is conducted and scope notes are added to the terms to specify their contextual semantics in order to avoid the inconsistency between local validity and global validity.

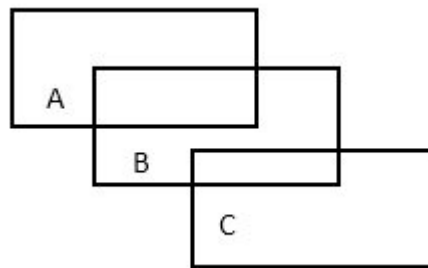


Figure 1. Loss of membership due to partially overlapped semantics.

The third is the possible poly-hierarchical structure due to classification based on multiple competing facets. For instance, verb “*sample*” can be classified into the category of Membership based on its feature facet (e.g., X is sampled from a population), and can also be classified into the category of Evaluation based on its function facet (e.g., X is sampled for evaluating its toxicity). Sometimes, it is difficult to figure out what facet should be used to classify a relation term because S. J. Ranganathan’s five facets (i.e., personality, Matter, Energy, Space, and Time) does not seem to apply to semantic relation terms. Interestingly, it is unknown whether facet analysis of relation terms should be performed at all. However, classifying a relation term into multiple categories is not ideal because it may cause confusion in knowledge discovery through inference. Our partial solution to this problem is to think of the nature of the relation term in its application context of “Topic A <relation term> Topic B,” or to replace the generic term (e.g., “sample”) with a term with more context (e.g., “be sampled from” or “be sampled for”).

### **Preliminary Evaluation**

Validation or evaluation of a taxonomy is mostly subjective and qualitative work based on a list of criteria. A taxonomy is a classification scheme which organizes concepts and things in a hierarchically ordered, systematic and abstract structure (Ramos and Rasmus, 2003; Lambe, 2007). So the criteria of evaluating a classification scheme can also be applied to evaluating a taxonomy. Taylor (1992, 322-333) proposed the following general criteria for judging a successful classification system: (1) inclusive and comprehensive knowledge of a whole field, (2) systematic division of subjects and organization of related topics, (3) flexible, hospitable and expansible structure, (4) clear and descriptive terminology with consistent meaning for both the user and the classifier. Lambe (2007, 201) proposed nine key criteria for usable, robust taxonomy structures: “intuitive (is easy to navigate and use), unambiguous (does not offer alternates), hospitable (can accommodate all content), consistent and predictable (provides context), relevant (reflects user perspective), parsimonious (no redundancy or repetition), meaningful (provides context), durable (will not need frequent change), balanced (even levels of detail or depth).” However, Lambe (2007, 201) pointed out that “these criteria are best treated as heuristics for an effective taxonomy rather than hard and fast rules” and there are three stages in validating a taxonomy: structural validation, validation with people (i.e., domain experts, users), and validation with content (i.e., categorizing content into the taxonomy).

Not all of these criteria are easy to be used to evaluate a taxonomy. Most of these criteria are subjective and qualitative, and are supposed to be used by domain experts, linguists, and users as evaluators. Validation with content is a functional validation method. This is analogous to a thesaurus evaluation method proposed by Soergel (1974), who proposed to test a thesaurus by indexing and retrieval experiments, such as “indexing 1,000 to 2,000 documents with the aid of the thesaurus” (Soergel, 1974, 411). A taxonomy has its functions. A taxonomy, in a corporate setting, serves the functions of (1) navigating through resources of the corporate, (2) providing tools for representing documents of the corporate, (3) serving as a sense-making tool or visual representation of the knowledge base of the corporate (Gilchrist, 2001; Abbas, 2010). Wang et al. (2010, 2014) designed an organizational taxonomy for navigation purpose, and evaluated its

navigation effectiveness using scenario-based navigation exercises and post-exercise interviews. The functional evaluation method can be an effective and relatively objective method to evaluate the functions of the designed taxonomy.

We have not found any discussion of the evaluation of a relation taxonomy (as opposed to subject/topic taxonomies) from literature. The general criteria for judging a successful taxonomy can be applied, but can be expensive to implement if domain experts and users are to be invited to evaluate the taxonomy. The Oil Spill Relation Taxonomy is designed not for navigating information resources, but for supporting knowledge discovery through inference. Therefore we decided to do some quick functional evaluation by discovering some examples of inferred knowledge from the oil spill topic map research data (Wu, 2013).

The logic of using the Oil Spill Relation Taxonomy to make inference is described below. If we can follow Swanson and Smalheiser's (1999) idea of discovery through inference and find a series of statements from the oil spill research data in the following general pattern, the taxonomy can facilitate knowledge discovery through inference.

A <R1> B,

B <R2> C,

C <R3> D,

Inferred knowledge: A <may/might R4> D.

Here A, B, C, & D are topics or concepts. R1, R2, R3, & R4 are relation terms and/or categories in the relation taxonomy. Following this general pattern, we found the following examples from the data:

Example 1:

Gulf Coast communities <experience> income loss,

income loss <cause> worse depression,

depression <cause> corrosive social cycle,

Inferred knowledge: Gulf Coast communities <may experience> corrosive social cycle.

Example 2:

oil <kill> Arctic phytoplankton,

Arctic phytoplankton <be consumed by> Arctic cod,

Arctic cod <be consumed by> ringed seal (*phoca hispida*),

Inferred knowledge: oil <may kill> ringed seal (*phoca hispida*).

The inference examples sheds light on the knowledge discovery function of the Oil Spill Relation Taxonomy. No efforts have been made to develop a series of specific inference patterns or to discover many of such examples from the data.

In addition to the preliminary functional evaluation, some structural evaluation was conducted. From the perspective of balance, one of the nine criteria for judging a successful taxonomy, the Oil Spill Relation Taxonomy does not have a balanced structure yet. Some categories (such as Act, Impact) are bigger and deeper than others. It is unknown whether the imbalance reflects the reality of semantic relations in the oil spill domain which focuses on impact or the balance criteria applies to any semantic relation taxonomy. More study on this topic is needed.

A taxonomy should be in a semi-permeable state in order to maintain modernity and validity (Faith, 2013). Out of the nine key criteria for judging a successful taxonomy, durability and expansibility can probably be evaluated in a non-expensive way. The durability and expansibility of the Oil Spill Relation Taxonomy is to be tested by classifying the relation terms in the OBO Relation Ontology into the Oil Spill Relation Taxonomy.

The OBO Relation Ontology (OBO, 2002) is a list of 397 relation terms in the biological and biomedical domain. The Oil Spill Relation Taxonomy has some biological and biomedical relation terms, but their scope is broader and shallower than those in OBO. Therefore the two taxonomies should have some overlap but also much difference. It is expected that some categories in the Oil Spill Relation Taxonomy may be revised and some new categories may be added when classifying the OBO relation terms into the Oil Spill Relation Taxonomy. By examining the number of revised and newly added categories, we can have a sense of the durability and expansibility of the taxonomy.

### **Summary and Future Work**

A preliminary semantic relation taxonomy in the oil spill domain (i.e., the Oil Spill Relation Taxonomy) was developed for supporting knowledge discovery through inference using a combination of top-down and bottom-up approach. Several difficult problems were discussed, including the muddy middle game in building middle level categories, the possible inconsistency between local validity and global validity due to contextual or partial membership, and the possible poly-hierarchical structure due to classification based on multiple competing facets. Partial solutions to these problems were suggested, but more discussion and study of these problems are needed.

The taxonomy was built for supporting knowledge discovery through inference, not for organizing and navigating information resources, therefore a preliminary functional evaluation was performed to examine its functionality of supporting knowledge discovery. Several examples were found from the oil spill topic map research data to demonstrate this functionality. Developing specific, systematic inference patterns for knowledge discovery can be a topic for future study.

Many issues remain to be studied in the future. In addition to the difficult problems during the development of the relation taxonomy, facet analysis of relation terms is an interesting topic because S. J. Ranganathan's five facets do not seem to apply to relation terms. Systematic evaluation of taxonomy needs more research. Practical, non-expensive, systematic evaluation approaches are needed. The evaluation approaches may be related to the difficult problems identified in taxonomy development process. Once we know how to evaluate the effectiveness of a taxonomy, we probably can solve some of the problems in the development process and build an effective taxonomy. This study has proposed more research problems than solutions.

### **Acknowledgements**

The study was partially supported by the Gulf of Mexico Research Initiative (GRI) Year One Block Grant. We would like to thank Steven Buras in the School of Library and Information Science at the Louisiana State University for clustering the oil spill relation terms.



## References:

- Abbas, June. 2010. *Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies, and Other Schemas*. New York: Neal-Schuman Publishers, Inc.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1990. "The Berkeley FrameNet Project." In *Proceedings of the 17th International Conference on Computational Linguistics*, 86-90.
- Baker, Collin F. and Josef Ruppenhofer. 2002. "FrameNet's frames vs. Levin's verb classes." In *Proceedings of the 28<sup>th</sup> Annual Meeting of Berkeley Linguistics Society*. <http://www.icsi.berkeley.edu/pubs/ai/framesvs02.pdf>
- Bertaud, Valerie, W. Said, Nicolas Garcelon, Franck Marin, and Regis Duvaufferrier. 2007. "The Value of Using Verbs in Medline Searches." *Medical Informatics and the Internet in Medicine* 32(2):117-122.
- Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. 1956. "Taxonomy of educational objectives: The classification of educational goals." *Handbook I: Cognitive Domain*. New York, Toronto: Longmans, Green. Retrieved from <http://www.edpsycinteractive.org/topics/cognition/bloom.html> (accessed 1/10/2015).
- Cisco, Susan L. & Wanda K. Jackson. 2005. "Creating Order out of Chaos with Taxonomies." *The Information Management Journal*, May/June 2005, 45-50.
- Faith, Ashleigh. 2013. "Linguistic Analysis of Taxonomy Facet Creation and Validation." *Key Words* 21(1): 11-14.
- Fellbaum, Christiane 1990. "English Verbs as a Semantic Net." *International Journal of Lexicography* 3(4): 278-301.
- Gilchrist, Alan. 2001. "Corporate Taxonomies: Report on a Survey of Current Practice." *Online Information Review* 25(2): 94-102
- Green, Rebecca. 1996. "A Relational Thesaurus: Modeling Semantic Relationships Using Frames." *Annual Review of OCLC Research*, 94-97.
- Hedden, Heather. 2010. *The Accidental Taxonomist*. Medford, New Jersey: Information Today, Inc.
- Hjørland, Birger. 2007. Semantic Relations. Retrieved from [http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/semantic\\_relations.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/semantic_relations.htm) (accessed 1/20/2015).
- Holgate, Lisle. 2004. "Creating and Using Taxonomies to Enhance Enterprise Search." *EContent* 21(7):S10-S11.

Lambe, Patrick. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Oxford, UK: Chandos Publishing.

Lawler, John M. 2015. A reverse-engineered list of Levin Categories. Retrieved from <http://www-personal.umich.edu/~jlawler/levin.verbs> (accessed 1/20/2015).

Levin, Beth. 1993. *English Verb Class and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

OBO (2002). OBO relation ontology. University of Michigan Medical School. <http://www.OBO.org/browser/term.php?o=RO&iri=http://www.w3.org/2002/07/owl%23ObjectProperty&graph=http://purl.obolibrary.org/obo/merged/RO> (accessed 1/5/2015).

Ramos, Laura and Daniel Rasmus. "Best Practices in Taxonomy Development and Management." Giga Information Group. 8 January 2013. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.201.4848&rep=rep1&type=pdf> (accessed 4/20/2015)

Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles, California: Melville Publishing Company.

Spradley, James P. 1979. *The Ethnographic Interview*. New York: Holt, Rinehart and Winston.

Swanson, Don R. and Neil R. Smalheiser. 1999. Implicit text linkages between Medline records: using Arrowsmith as an aid to scientific discovery. *Library Trends*, 48(1), 48-59.

Taylor, Arlene G. 1992. *Introduction to Cataloging and Classification* (8<sup>th</sup> ed). Englewood, Colorado: Libraries Unlimited, Inc.

UMLS (2013). Current Relations in the Semantic Network. Retrieved from [http://www.nlm.nih.gov/research/umls/META3\\_current\\_relations.html](http://www.nlm.nih.gov/research/umls/META3_current_relations.html) (accessed 1/5/2015).

Wang, Yih-Chen, James Vandendorpe, & Martha Evans. 1985. "Relational Thesauri in Information Retrieval." *Journal of the American Society for Information Science* 36(1):15-27.

Wang, Zhonghong, Abdus S. Chaudhry, & Christopher Khoo. 2010. "Support from Bibliographic Tools to Build an Organizational Taxonomy for Navigation: Use of a General Classification Scheme and Domain Thesauri." *Knowledge Organization* 37(4): 256-269.

Wang, Zhonghong, Christopher S.G. Khoo, and Abdus Sattar Chaudhry. 2014. "Evaluation of the Navigation Effectiveness of an Organizational Taxonomy Built on a

General Classification Scheme and Domain Thesauri.” *Journal of the Association for Information Science and Technology* 65(5): 948-963.

White House. 2012. Deepwater BP Oil Spill. Available at <http://www.whitehouse.gov/deepwater-bp-oil-spill> (accessed 4/20/2014).

Wu, Yejun. 2013. Oil Spill Topic Map Data. 28 November 2013.  
<https://data.gulfresearchinitiative.org/data/Y1.x101.000:0001/>

Wu, Yejun and David J. Dunaway. 2013. “Creating a large topic map by integrating Wandora and Ontopia.” *Library Hi Tech* 31(1):64-75.

Xiang, Zuoshuang, Chris Mungall, Alan Ruttenberg, and Yongqun He. 2011. “OBO: A Linked Data Server and Browser for Ontology Terms.” In *Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO)*, July 28-30, 2011, Buffalo, NY, USA. 279-281. <http://ceur-ws.org/Vol-833/paper48.pdf>

## **Appendix. A Preliminary Oil Spill Relation Taxonomy (Main Categories and Example Instances)**

(Category labels in upper-case noun phrases, instances in lower-case verb phrases)

1 ACTION	compensate public for
1.1 CONFRONTATION	pay for
defy	1.3.2 FUNDING
1.2 EVALUATION	finance
1.2.1 ASSESSMENT	fund
assess	1.3.3 RECEIPT
evaluate	receive
1.2.2 COGITATION	1.4 GOVERNMENTAL ACTION
consider	authorize
reflect	regulate
1.2.3 DECISION	1.5 INTENTION
decide	aim to
1.2.4 DETERMINATION	intend
be determined in	1.6 IMPLEMENTATION
determine	1.6.1 CLOSURE/OPEN
1.2.5 EMPHASIS	close
concentrate on	re-open
focus on	1.6.2 COMMUNICATION
1.2.6 EVIDENCE	communicate
be evident in	respond to
reveal	1.6.3 CONTROLLING
1.2.7 EXPECTATION	control
expect	1.6.4 CREATION
predict	create
1.2.8 NONSUPPORTIVE JUDGEMENT	establish
criticize	1.6.5 DETECTION
ignore	detect
1.2.9 RECOMMENDATION	discover
recommend	1.6.6 EDUCATION
suggest	teach
1.2.10 REQUIREMENT	1.6.7 MANAGING
request	manage
require	run
1.2.11 SELECTION	1.6.8 MONITORING
select	monitor
1.2.12 SUPPORTIVE JUDGEMENT	observe
agree with	1.6.9 OBTAINING
grant	access
1.3 FINACIAL ACTION	obtain
1.3.1 COMPENSATION	1.6.10 PERFORMANCE

conduct	be identified as
perform	identify
1.6.11 PRACTICE	1.9.5 METHOD OF
drill in	be compiled with
practice	be quantified as
1.6.12 RECOVERY	1.9.6 SPECIFICATION
recover	explain
remediate	specify
1.6.13 RESCUE	1.9.7 USE
search for	use
1.6.14 RESEARCH	utilize
research	1.10 MOVEMENT
study	1.10.1 EMISSION
1.7 INSTRUMENT	release
1.7.1 ANALYSIS	spill
analyze	1.10.2 GATHERING
be analyzed to determine	accumulate
1.7.2 DIAGNOSIS	gather
diagnose	1.10.3 MOVING
1.7.3 MEASUREMENT	flow for
be calibrated for	move
measure	1.10.4 PLACING
1.8 LEGAL ACTION	deliver
1.8.1 ACCUSATION	transport
accuse	1.10.5 REMOVING
sue	eliminate
1.8.2 LEGAL JUDGEMENT	remove
violate	1.11 PERCEPTION
waive	be aware of
1.8.3 LEGISLATION	see
be amended by	1.12 PROVIDING SUPPORT
1.8.4 TESTIFICATION	1.12.1 COLLABORATION
pledge	collaborate with
testify	cooperate with
1.9 Method/Manner	1.12.2 EMPLOYMENT
1.9.1 CATEGORIZATION	assign
be used to categorize	employ
classify	1.12.3 FACILITATION
1.9.2 DEFINITION	aid
define	facilitate
1.9.3 EXAMINATION	1.12.4 SUPPLY
check	offer
examine	provide
1.9.4 IDENTIFICATION	1.13 STATEMENT

argue	4.2.2 FUNCTION (GENERAL)
state	be suited for
2 ASSOCIATION	function in
2.1 CORRELATION	4.2.3 INTAKE FUNCTION
be correlated for	absorb
be highly/strongly correlated with	uptake
2.2 RELATEDNESS	5 IMPACT
be linked to	5.1 INFLUENCE
be related to	5.1.1 INFLUENCE (GENERAL)
3 EQUIVALENCE/COMPARABILITY	affect
3.1 COMPARISON	impact
be more than	5.1.2 ALLOWANCE
compare	allow
3.2 CORRESPONDENCE	permit
correspond with	5.1.3 CHANGE
3.3 EQUIVALENCE	change
3.3.1 EQUAL TO	stabilize
be an alternative to	5.1.4 CONTAMINATION
be substituted for	contaminate
3.3.2 PARTNER OF	pollute
be partner of	5.1.5 DAMAGE
3.4 SIMILARITY/DIFFERENCE	damage
3.4.1 DIFFERENCE	destroy
differ from	5.1.6 EXPERIENCE
differentiate among	experience
3.4.2 SIMILARITY	undergo
be close to	5.1.7 HARM
be similar to/in	harm
4 FEATURE/FUNCTION	weaken
4.1 FEATURE	5.1.8 INCREASE
4.1.1 CHARACTERIZATION	improve
characterize	increase
have feature	5.1.9 INTERFERENCE
4.1.2 COMMUNITY FEATURE	5.1.9.1 COMPLICATION
be as equally resilient as	5.1.9.2 DISRUPTION
4.1.3 GEOGRAPHICAL FEATURE	disturb
be native of	interrupt
4.1.4 PSYCHOLOGICAL FEATURE	5.1.9.3 INTERACTION
hate	interact with
surprise	want to comply with
4.2 FUNCTION	5.1.10 INVOLVEMENT/PARTICIPATION
4.2.1 BIOLOGICAL FUNCTION	be engaged in
metabolize	involve
stimulate	5.1.11 KILL

be lethal to	contribute to
kill	play a key role in
5.1.12 MODIFICATION	5.2.1.4 PRODUCING
alter	generate
update	produce
5.1.13 PREVENTION	5.2.1.5 RESULT
avoid	be conclusion of
prevent	be result of
5.1.14 PROTECTION	5.2.2 RATIONALE
protect	5.2.2.1 REASON
safeguard	6 POSSESSION
5.1.15 RESTRICTION	6.1 HAVING
5.1.15.1 IMPEDIMENT	have
impede	own
inhibit	7 RELIANCE
5.1.15.2 LIMIT	7.1 CONDITION
limit	7.1.1 BASIS/PREREQUISITE
restrict	/FOUNDATION
5.1.16 REDUCTION	be based on
decrease	rely on
reduce	7.2 IMPORTANCE
5.1.17 RISK	be critical in
be at risk	be essential to
threaten	8 SEQUENCE
5.1.18 TREATMENT	8.1 CHRONOLOGICAL SEQUENCE
5.1.18.1 BIOLOGICAL & CHEMICAL	8.1.1 BEGINNING/CONTINUANCE/
TREATMENT	ENDING
biodegrade	begin
oxidize	end
5.1.18.2 MEDICAL TREATMENT	8.1.2 OCCURRENCE
anesthetize	occur during/while
treat (disease, patient)	happen
5.1.18.3 PHYSICAL TREATMENT	8.1.3 PRECEDING
be treated with	be previously
wash away	occur before
5.1.18.4 REPAIRMENT	8.2 DEVELOPMENTAL SEQUENCE
repair	4.2.1 DERIVATIVE OF (CREATION)
5.2 CAUSE-EFFECT	derive mainly from
5.2.1 PRODUCTION	4.2.2 DEVELOPMENTAL FORM OF
5.2.1.1 BRING ABOUT	develop
5.2.1.2 CAUSE	mature in
cause	8.3 FEEDING SEQUENCE
lead to	8.3.1 FOOD CHAIN
5.2.1.3 CONTRIBUTION	be food source for

consume (eat)

8.4 PROCEDURAL SEQUENCE

8.4.1 FOLLOWING

be ready for  
follow

8.4.2 PLANNING

plan  
schedule

8.5 RANK

be above

8.6 RULE-BASED SEQUENCE IN  
GAMES

8.7 SOURCE-PRODUCT SEQUENCE

be obtained from  
be refined to

9 SPATIAL RELATIONSHIP

9.1 CENTER-PERIPHERY

9.1.1 SURROUNDING

border

9.2 LOCATIVE

9.2.1 LOCATION OF

be at bottom of  
be located in

9.3 PATH

9.3.1 SPATIAL CONNECTION

9.3.2 TRAVERSE

10 STRICT INCLUSION

10.1 HIERARCHY

10.1.1 IKO

is a  
be regarded as

10.2 MEMBERSHIP

10.2.1 INSTANCE OF

sample

10.3 PART-WHOLE

10.3.1 BRANCH/TRIBUTARY OF

branch of  
tributary of

10.3.2 CONTAINING

be richly endowed with  
contain

10.3.3 INCLUSION

include  
include significant factor of

10.3.4 INGREDIENT OF

be component

10.3.5 KIND OF

be a kind of  
have rig type

10.3.6 PART OF

consist of  
be integrated into