Yejun Wu and Rujiang Bai. 2017. An Event Relationship Model for Knowledge Organization and Visualization. NASKO, Vol. 6. pp. 216-230.

Yejun Wu
School of Library and Information Science
Louisiana State University, USA

Rujiang Bai
University Library
Shandong University of Technology, China

An Event Relationship Model for Knowledge Organization and Visualization

Abstract

An event is a specific occurrence involving participants, which is a typed, n-ary association of entities or other events, each identified as a participant in a specific semantic role in the event (Pyysalo et al. 2012; Linguistic Data Consortium 2005). Event types may vary across domains. Representing relationships between events can facilitate the understanding of knowledge in complex systems (such as economic systems, human body, social systems). In the simplest form, an event can be represented as *Entity A <Relation> Entity B*. This paper evaluates several knowledge organization and visualization models and tools, such as concept maps (Cmap), topic maps (Ontopia), network analysis models (Gephi), and ontology (Protégé), then proposes an event relationship model that aims to integrate the strengths of these models, and can represent complex knowledge expressed in events and their relationships.

1. Introduction and Significance

When we gain more understanding of a domain (such as medicine, oil spill), we accumulate knowledge in that domain, and want to organize and visualize the knowledge for learning and discovery. What knowledge organization and visualization models and tools are effective for managing an accumulated set of knowledge that is still growing?

Common knowledge organization models include concept maps, topic maps, classification schemes, thesauri, and ontologies. Concept maps are widely used in education to represent relationships between concepts (Novak and Gowin 1984). Topic maps are an information and knowledge organization tool that represent the relationship between topics and between the topic and the occurrences of the topic in information resources (Hatzigaidas et al. 2004). Concept maps and topic maps are useful knowledge organization tools for expressing the relationship between concepts (or entities), but have limited power in expressing the relationships between events. Classification schemes and thesauri can express the hierarchical relationship between concepts (or entities), such as class-member, broader term, narrower term. The related term (RT) relationship in thesauri is non-specific. “A knowledge base contains different kinds of knowledge, typically an ontology, facts, rules and constraints… A computational ontology provides a symbolic representation of objects, classes of objects, properties and relationships between objects used to explicitly represent knowledge about an application domain. It is the cornerstone of a knowledge representation” (Chein and Mugnier 2009, 2). However, ontologies are too formal and rigid for representing concepts and their relations, because an ontology, as “a formal explicit specification of a shared conceptualization” (Gruber 1993), explicitly represents concepts in a domain and their relations in a formal and consistent way, and captures consensual knowledge (Studer et al. 1998). The knowledge we would like to represent and visualize in this paper is in the form of “*who did what to whom” with some conditions (such as time, place, and method).* In its simplest form, it can be expressed as *who <did what to> whom*. In an abstract from, it can be expressed as *Entity A <relation> Entity B* where *relation* is normally a verb phrase. In its linguistic form, it can be expressed as *NP <VP> NP* where NP is a noun phrase and VP is a verb phrase.

In this simplest form of knowledge encapsulated in an event, if entities and relations are not typed, they may bring about ambiguity in their meanings. For example, in the statement of “Tiffany has a Jaguar,” it is unclear whether Tiffany is a person or an organization, and whether Jaguar is an animal or a car, and whether “has” means an ownership relationship. Therefore we need a model to effectively organize, represent and visualize typed entities and typed relations. In order to represent an event unambiguously, we need to represent a minimum number of triples, such as

Entity A1 <Relation R1> Entity B1 (which is abbreviated as A1 <R1> B1),

Entity A1 <is a kind of> Entity Class A (which is abbreviated as A1 <ISA> A),

Entity B1 <is a kind of >Entity Class B (which is abbreviated as B1 <ISA> B),

Relation R1 <is a kind of> Relation Class R (which abbreviated as R1 <ISA> R).

Furthermore, in order to represent the relationship between two events, we need to represent two events and their relationship in the following simplest form:

(A1 <R1> B1) <R12> (A2 <R2> B2).

We propose an event relationship model, which aims to not only overcome the limitations of the common knowledge organization models introduced above when representing events (composed of entities and their relations, and other participants), but also has the power of representing the relationship between events. Events are “typed, n-ary associations of entities or other events, each identified as participating in a specific role" (Pyysalo et al. 2012), therefore an event has the power of encapsulating the relationship between entities and/or events. The event relationship model has the power of representing the relationship between events, which can be useful to represent knowledge of complex systems (such as economic system) which is often composed of a series of events.

We also propose a visualization prototype of the event relationship model, which is to be built based on current graphic representation formalisms. It has the power of visualizing the relationships between events, and can be useful for various purposes (such as knowledge understanding, knowledge discovery and decision-making).

2. Related Work

The following work has inspired the event relationship model that is proposed here: knowledge modelling tools [such as entity-relationship models (Chen 1976), object modelling techniques (Rumbaugh et al. 1991)], extended relational data model (ERDM), object oriented data model (OODM) (Rob and Coronel 2002), schema theory which plays a central role in knowledge construction and learning (Basque et al. 2008), typology of knowledge in educational sciences [which specifies four basic types of knowledge entities: facts, concepts, procedures, and principles (Basque et al. 2008)], typology of relationships, including *instantiation (is-a)*, *composition (is part of)*, *specialization (a kind of)*, *input/product*, *precedence*, and *regulation* (Basque et al. 2008), events which are expressed with frames or templates in Message Understanding Conference (Ralph and Sundheim 1996), event types (Linguistic Data Consortium 2005), and semantic relation taxonomy (Wu and Yang 2015).

3. Research Questions and Methodology

The following three research questions are discussed in this paper:

(1) For managing an accumulated set of knowledge that is still growing, are current knowledge organization and visualization models and tools effective in representing and visualizing simplified knowledge in the form of *Entity A <Relation> Entity B*?

(2) Are current knowledge organization and visualization models and tools effective in representing and visualizing relationships between two events in the simplified form of *(A1 <R1> B1) <R12> (A2 <R2> B2)*?

(3) How can we organize, represent, and visualize entities and their relationships, and events and their relationships?

We first briefly evaluate several knowledge organization and visualization models and tools for representing simplified knowledge in the form of *Entity A <Relation> Entity B*, such as Cmap (for concept map), Ontopia (for topic maps), Gephi (for graph network analysis), and Protégé (for ontology). We then propose the event relationship model. When applicable and possible, we use the oil spill data set in the form of *Entity A <Relation> Entity B*, which contains approximately 5,000 concepts and 1,000 relationships extracted from about 300 documents comprised of conference presentations, journal articles, news reports, and authoritative Web pages (Wu, Lehman and Dunaway 2015).

4. A Brief Evaluation of Knowledge Organization and Visualization Models and Tools

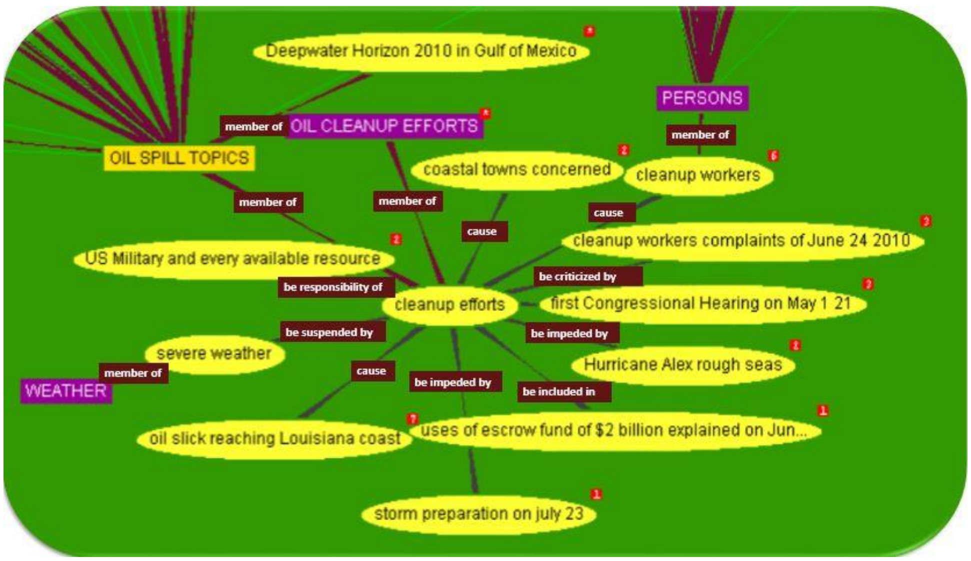
This section briefly reviews several knowledge organization and visualization models and tools on their effectiveness in knowledge organization, representation and visualization. The evaluation criteria include the following factors: (1) the availability of typed entities and typed relations that are used to represent meanings of entities and relationship unambiguously, (2) the availability of the construct of *Entity A <Relation> Entity B* in the model, (3) the size of knowledge network that is to be managed, (4) the flexibility of adding and deleting knowledge statements, (5) visualization power, (6) the capability of visualizing relationships between objects.

4.1 Cmap (for concept maps) and Ontopia (for topic maps)

A concept map is a graphical tool for organizing and representing knowledge. It includes concepts, and relationships between concepts (Novak and Cañas 2008). A topic map has three constructs: 1) "topics," which represent "subjects of discourse;" 2) "associations," which represent relationships between the subjects; and, 3) "occurrences," which are information resources relevant to a given topic. Topic maps have built into them the relationship between topics and between the topic and the occurrences of the topic in information resources (Hatzigaidas et al. 2004).

One strength of concept maps and topic maps lies in that they can take a batch of knowledge statements in the form of *Entity A <Relation> Entity B* as input, and users can add knowledge statements into the input file easily. However, their visualization power is limited. A network of topics can be cluttered when a large number of topics are presented. For example, Figure 1 shows 12 topics associated with *cleanup efforts* and the screen is almost full. The display would be very cluttered if the number of associated topics doubled. This is a weakness of Ontopia (Wu, Lehman and Dunaway 2015). Concept maps have the similar weakness. A second weakness is the imprecise meaning of the relationships (or links) represented in these models since non-typed relations (or links) can have various meanings (Basque et al. 2008). A third weakness is the ambiguity around the type of entities (or concepts). Entities, actions performed on entities, conditions applied to actions, and statements of properties about the entities are often not distinguished, which results in imprecise graph interpretation (Basque et al. 2008).

Figure 1. Graphical topic map for topic "cleanup efforts"



4.2 Gephi (for graph and network analysis)

Gephi is an open source software for network exploration and manipulation. It can import, visualize, spatialize, filter, and manipulate all types of networks and large networks (i.e., over 20,000 nodes). It can explore networks in an interactive way and display a dynamic network (Bastian et al. 2009). Gephi has many network visualization strengths. It can spatialize the nodes so that the nodes are not cluttered. Gephi can also spatialize and display chains of relationships. Figure 2 shows how Gephi spatializes the nodes (i.e., concepts or entities) and displays the clusters of nodes in the network. Figure 3 shows lengthy chains of relationships between concepts/entities, which can be used for knowledge discovery through inference. As a network visualization software, Gephi does not focus on knowledge organization functions. Its weaknesses include the imprecise meaning of non-typed relations (or links) represented in the models, and the ambiguity around the type of entities or concepts.

Figure 2. Visualization of Oil Spill Knowledge by Gephi

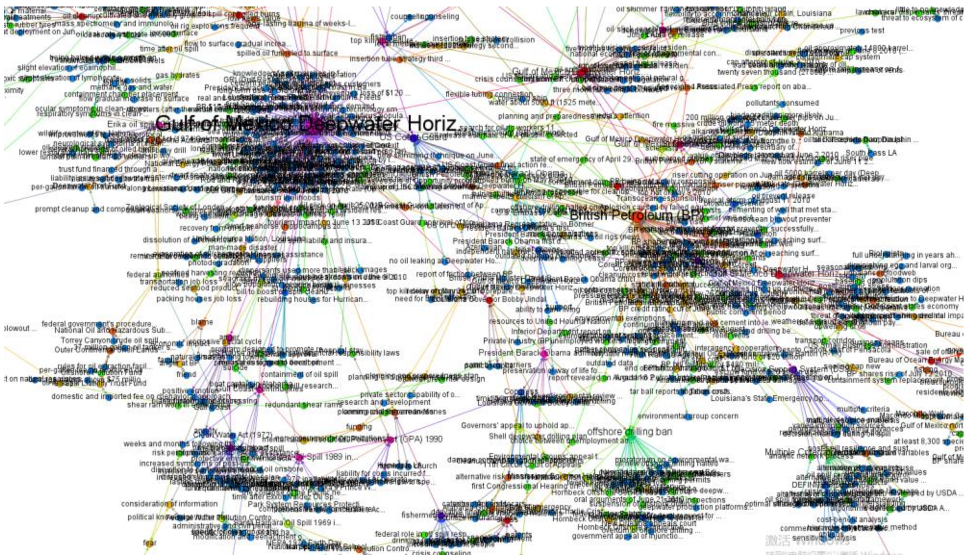
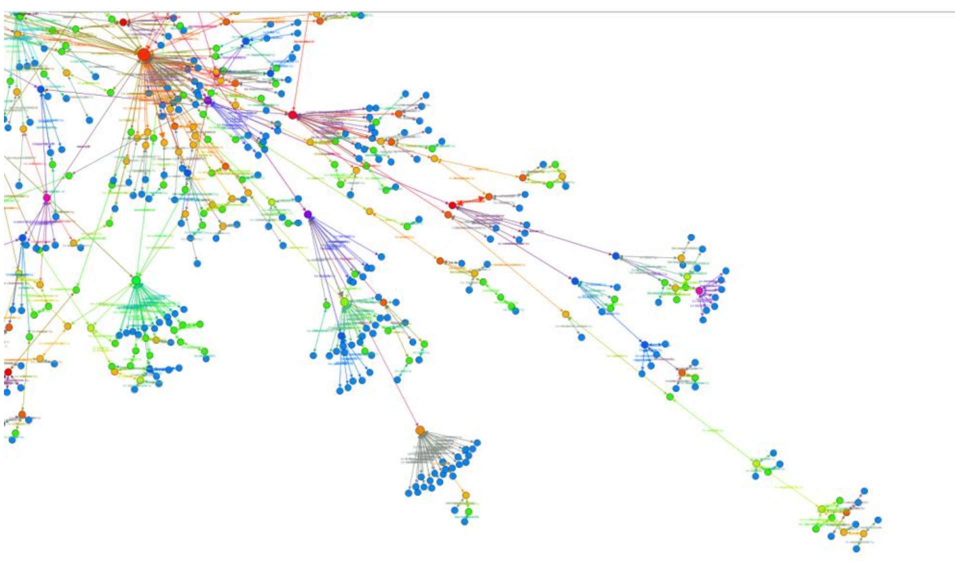


Figure 3. Visualization of Long Chains of Relationships in Oil Spill Data by Gephi



4.3 Protégé (for ontology)

When we have complete understanding of an application domain and want to represent concepts in the domain and their relations in a formal, explicit and consistent way, ontology is the appropriate model. Protégé is an open-source ontology editor (http://protege.stanford.edu). Figure 4 shows Protégé's interface for creating an ontology of wines. The strength of ontology lies in that it explicitly describes a domain's concepts, properties and attributes of concepts, constraints on properties and attributes (such as role, name, type, cardinality, allowed values), and instances. A weakness is that creating an ontology requires complete understanding of the domain and can be very time-consuming. Protégé has various visualization plugins for ontology data display (http://protegewiki.stanford.edu/wiki/Visualization), but is not as powerful as Gephi.

Figure 4. Wines Ontology by Protégé (Source: Noy & Tu 2003)

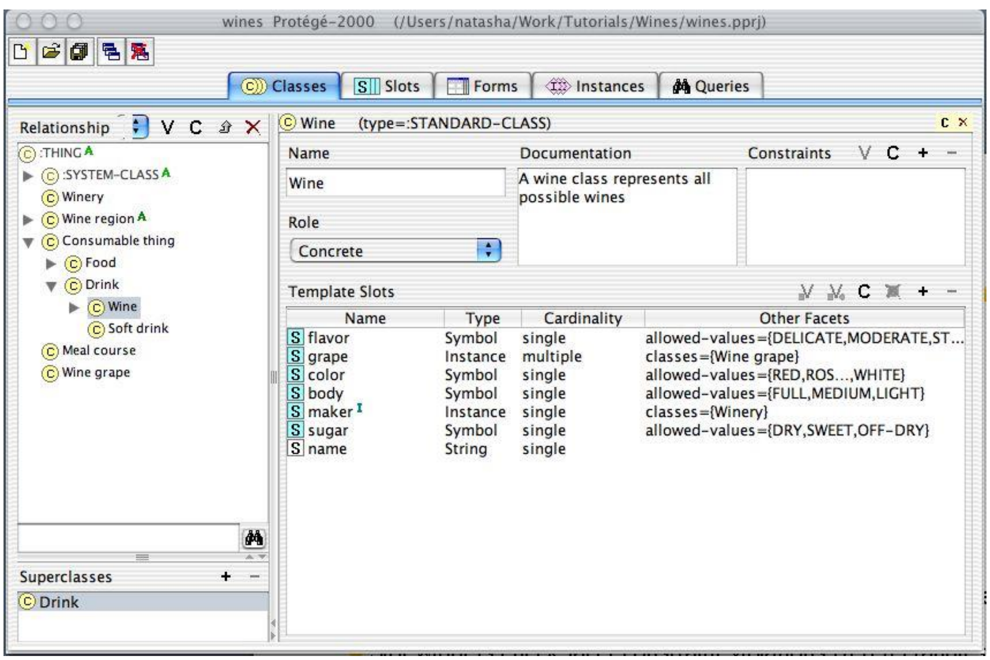


Table 1. A Comparison of Knowledge Organization and Visualization Models and Tools

Criteria	Concept Map (CMap)	Topic Map (Ontopia)	Network Analysis (Gephi)	Ontology (Protégé)
Typed entities	no	no	no	yes
Typed relations	no	no	no	no
Construct of *Entity A <R> Entity B*	medium	medium	weak	medium
Size of knowledge network	medium	medium	large	large
Flexibility of adding & deleting knowledge	strong	strong	strong	medium
Visualization power	medium	medium	strong	medium
Capability of visualizing relationships	medium	medium	strong	medium

Table 1 shows a comparison of the models and tools according to the evaluation criteria. To answer our first research question, we find that the four models and tools are weak in representing and visualizing simplified knowledge in the form of *Entity A <Relation> Entity B*. To answer our second research question, we find that neither of the four models and tools can represent and visualize relationships between two events in the simplified form of *(A1 <R1> B1) <R12> (A2 <R2> B2)*. To answer the third research question, we need a knowledge organization and visualization model that can take the strengths of concept maps (Cmap), topics maps (Ontopia), network analysis (Gephi), and ontology (Protégé), and minimize their weaknesses. We aims to integrate their strengths in the event relationship model, which is presented below.

Yejun Wu and Rujiang Bai. 2017. An Event Relationship Model for Knowledge Organization and Visualization. NASKO, Vol. 6. pp. 216-230.

5. Proposal: An Event Relationship Model

Our proposal has three parts: (1) developing a generic event description model by investigating event types in multiple domains (such as news, biomedical sciences), (2) developing an event relationship model based on the event description model, and (3) developing a visualization prototype of the event relationship model.

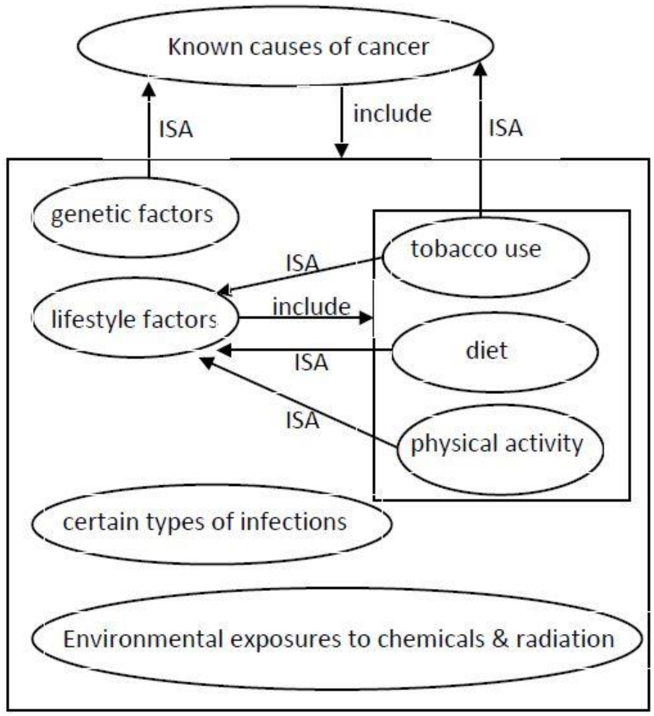
5.1 A generic event description model

A generic event description model is to be developed by investigating event types (or frames) in various domains (such as news, biomedical sciences) from event extraction research literature and text corpora (especially those annotated with events). For example, news events are typically presented as "*who did what to whom, through what methods (instruments), when, where and why*" (Atkinson and Piskorski 2011, 749). Such a template can be simplified as 5W1H (*who, what, whom, when, where, how*) (Wei 2012). Various domains may have various event templates, but the simplest, core frame (or template) is "*who did what to whom*," expressed as "*Entity A <relation> Entity B*."

Borrowing the idea from the extended relational data model (ERDM), an event can be implemented as an object in object-oriented modeling. An event can involve multiple entities and multiple relations. Figure 5 defines six basic kinds of relations between entities in an event. For example, the event in the following statement describes the relationship ("*can cause*") between a list of four entities and one entity ("*cancer*"). It can be represented as E3 (or E4) in Figure 5. A list (or taxonomy) of entities can be implemented as a compound object. Figure 6 visualizes the knowledge in this statement.

Example statement: "*You can learn more about the known causes of cancer, including genetic factors; lifestyle factors such as tobacco use, diet, and physical activity; certain types of infections; and environmental exposures to different types of chemicals and radiation*" (American Cancer Society 2015).

Figure 6. An Example of Knowledge Represented using ERDM.



(Note: ISA: “is a” or “is a kind of” relationship. Not all ISA relations are visualized.)

Figure 5. Relations between entities in an event

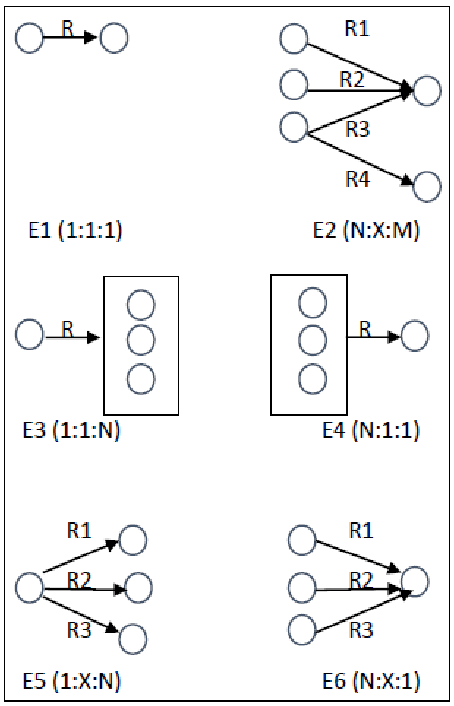
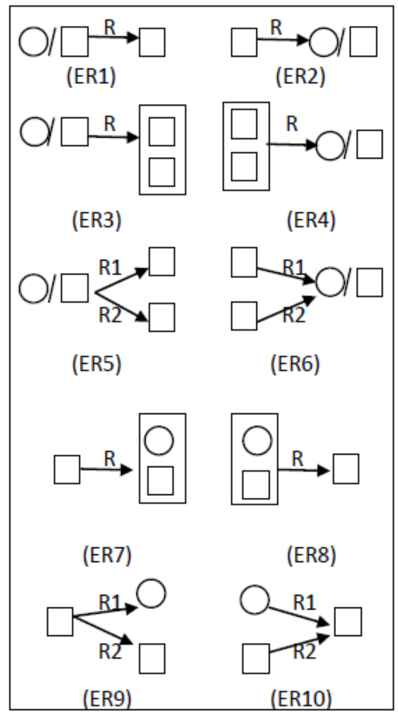


Figure 7. Relations between events



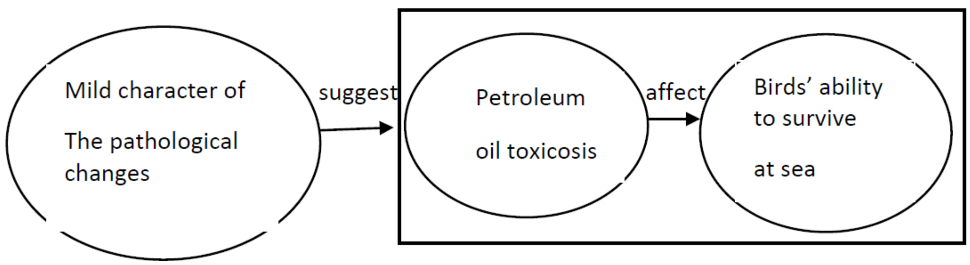
[Note: a circle (○) represents an entity; a square (□) represents an event; a slash (/) means "or." R represents a relation. N:X:M means multiple relations between multiple (N, M) entities. Figure 7 presents an <u>incomplete</u> list of relations between events for simplicity reason.]

5.2 An event relationship model

The event relationship model describes the relationship between two events, or between an entity and an event. A generic event relationship model is to be developed by investigating the various kinds of relationships between events from domain corpora (especially those annotated with events). An incomplete list of 10 kinds of preliminary relationships between events is summarized in Figure 7. For example, the events in the following statement can be represented using E1 (in Figure 5) and ER1 (in Figure 7). Figure 8 shows the representation of the knowledge in the statement. The event relationship model is to be implemented using extended relational data model (ERDM). The relationships are to be typed according to a semantic relation taxonomy (Wu and Yang 2015). Since an event can encapsulate entities, events, and the relationship between entities and/or events, the model can express complex relationships between entities and/or events.

Example statement: "*The mild character of the pathological changes suggests that petroleum oil toxicosis causes multiple sublethal changes that have an effect on the ability of the birds to survive at sea*" (Balseiro et al. 2005).

Figure 8. An example representation and visualization of the event relationship model



5.3 A visualization prototype of the event relationship model

An event is represented as a frame. For usability reason, the core frame is represented and visualized as "*Entity A <relation> Entity B*," and other frame slots (or participants) can be represented and visualized as attributes of the core event frame. The visualization model can be implemented using the Data-Driven Documents (D3) Javascript (D3.js) technology. D3.js is a JavaScript library for manipulating documents based on data. "D3.js reduces overhead and allows greater graphical complexity at high frame rates… With minimal overhead, D3.js is extremely fast, supporting large datasets and dynamic behaviors for interaction and animation" (Bostock, 2015). The prototype has Ontopia's flexibility in the description of events and batch input, has Gephi's network visualization power (such as spatialized display and long chains of relationships), has some of Protégé's entity description power (such as a taxonomy of entities, types of entities, attributes of entities), has typed relationships, but does not require all the formal specifications of entities in an ontology such as Protégé. The prototype is under construction.

6. Summary

We explored knowledge organization and visualization models and tools that are effective for managing an accumulated set of knowledge that is still growing. The knowledge, in its simplest form, is represented as *Entity A <Relation> Entity B*. We evaluated several models and tools, such as concept maps (Cmap), topic maps (Ontopinetwork visualization (Gephi), and ontology (Protégé). By integrating the strengths of these models and tools and minimizing their weaknesses, an event relationship model is proposed. The model can express complex relationships between entities and/or events in the simplified form of *(A1 <R1> B1) <R12> (A2 <R2> B2*.

References:

American Cancer Society (2015). What causes cancer? http://www.cancer.org/cancer/cancercauses/ (accessed 12/20/2016).

Balseiro, A., Espí, A., Márquez, I. et al. (2005). Pathological features in marine birds affected by the Prestige's oil spill in the North of Spain. *Journal of Wildlife Diseases* 41(2): 371-378.

Basque, Josianne, Paquette, Gilbert, Pudelko, Beatrice et al. (2008). Collaborative knowledge modeling with a graphical knowledge representation tool: a strategy to support the transfer of expertise in organizations. In Alexandra Okada et al. (eds), *Knowledge Cartography: Software Tools and Mapping Techniques*. London: Springer.

Bastian, Mathieu, Heymann, Sebastien, and Jacomy, Mathieu (2009). Gephi: an open source software for exploring and manipulating networks. AAAI 2009.

Björne, Jari, Ginter, Filip, and Pyysalo, Sampo et al. (2010). Scaling up biomedical event extraction to entire PubMed. *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing (ACL)*, Uppsala, Sweden, July 15. 28-36.

Bostock, Mike (2015). Data-Driven Documents. http://d3js.org (Retrieved 10/5/2015).

Chen, Peter P.-S. (1976). The entity-relationship model: toward a unified view of data. A*CM Transactions on Database Systems*, 1(1), 9-36.

Chien, Michel and Marie-Laure Mugnier (2009). *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*. London: Spinger-Verlage.

Elloumi, Samir, Jaoua, Ali, Ferjani, Fethi et al. (2012). General learning approach for event extraction: case of management change event. *Journal of Information Science*, 39(2): 211-24.

Gruber, T.R. (1993). A translation approach to portable ontology specification. *Knowledge Acquisition* 5(2): 199-220.

Hatzigaidas, A., Anastasia P., George T. et al. (2004). Topic map existing tools: A Brief Review. In *Proceedings of the International Conference on Theory and Applications of Mathematics and Informatics* 2004, Thessaloniki, Greece, 185-201.

Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

Linguistic Data Consortium (2005). ACE (Automatic Content Extraction) English annotation guidelines for events. https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf

Novak, J. D. & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. Institute for Human and Machine Cognition Technical Report CmapTools 2006-01 Rev 2008-01. http://cmap.ihmc.us/docs/theory-of-concept-maps

Novak, J. D. and Gowin, D. B. (1984). *Learning How to Learn*. Cambridge: Cambridge University Press.

Noy, Natasha & Tu, Samson (2003). Developing medical informatics ontologies using Protégé (tutorial), *AMIA* 2003, http://bmir-stage.stanford.edu/amia2003/index.html

Pyysalo, Sampo, Ohta, Tomoko, Rak, Rafal et al. (2012). Overview of the ID, EPI and REL tasks of BioNLP Shared Task 2011. *BMC Bioinformatics* 13 (Suppl 11). doi: 10.1186/1471-2105-13-S11-S2.

Paquette, G. (2006). Building graphical knowledge representation languages – from informal to interoperable executable models. Paper presented at the *i2LOR-06 Conference*, November 8-10, Montreal.

Ralph, Grishman and Sundheim, Beth (1996). Message understanding conference-6: a brief history. *COLING*, Vol. 1. 466-471.

14

Reichart, R. and Barzilay, R. (2012). Multi event extraction guided by global constraints. *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Montreal, Canada, June 3-8. 70-79.

Rob, P. and Coronel, C. (2002). Database Systems: Design, Implementation, & Management (5th ed.). Course Technology, Thomas Learning, Inc.

Rumbaugh, J., Blaha, M., Premerlani, W. et al. (1991). *Object-Oriented Modelling and Design*. Englewood Cliffs, NJ: Prentice Hall.

Sasaki, Yutaka, Thompson, Paul, Cotter, Philip et al. (2008). Event frame extraction based on a gene regulation corpus. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, Manchester, August. 761-768.

Schreiber, G., Wielinga, B., and Breuker, J. A. (1993). *KADS – A Principled Approach to Knowledge-based System Development*. San Diego, CA: Academic Press.

Studer, R., Benjamins, V.R., and Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data Knowledge Engineering* 25(1–2): 161-197.

Wu, Yejun, Lehman, Amanda, and Dunaway, David (2015). Evaluations of a large topic map as a knowledge organization tool for supporting self-regulated learning. *Knowledge Organization* 42(6): 386-398.

Wu, Yejun and Yang, Li (2015). Construction and Evaluation of an Oil Spill Semantic Relation Taxonomy for Supporting Knowledge Discovery. *Knowledge Organization* 42(4): 222-231.