

Modeling Entity and Event Relations in Scientific Documents for Supporting Knowledge Discovery and Organization

Yejun Wu

School of Library and Information Science,
Louisiana State University, USA
wuyj@lsu.edu

ABSTRACT

Background. Scientific documents often contain knowledge about *what one entity did to another entity under what conditions (such as time, place, and method)*, which is related to another statement of *what one entity did to another entity under what conditions*. Such knowledge can be represented as relations between entities and events. Here *what one entity did to another entity under what condition* is defined as an event, which expresses the relationship between two entities under a condition.

Objective. The objective of this paper is to design a model of entity and event relationship that can be used to represent knowledge identified from scientific documents and to facilitate knowledge discovery and organization.

Method. The paper first presents a brief literature review on causal relationships, then evaluates four existing knowledge organization models and five event ontologies for their commonalities and differences in representing entity relationships and event relationships. The paper then proposes a combined entity and event relationship model based on the strengths of the existing event ontologies. Five main kinds of entity and event relationships are identified from an oil spill document set.

Results. The three domain event ontologies, CIDOC CRM, Event Ontology and NewsML-G2, are only useful in serving specific purposes. The two generic event ontologies, DOLCE+DnS and Event Model F, must be enriched to be useful for representing knowledge for discovery. An entity and event model is proposed based on the strengths of these event models for representing knowledge in scientific documents.

INTRODUCTION

Scientific documents often contain knowledge about *what one entity did to another entity under what conditions (such as time, place, and method)*, which can be expressed as a tuple of entity relationships under a condition. Scientific documents may also express ideas of a tuple of entity relationships related to another tuple of entity relationships, which can be represented as a tuple of event relationships. The term “event” has many meanings. According to lexico.com, an event is “a thing that happens,” and has three meanings: (1) “a planned public or social occasion”, (2) “each of several particular contests making up a sports competition”, (3) “a single occurrence of a process, for example, the ionization of one atom”. The first two meanings can be consolidated as scheduled events. There are also phenomena that have happened, for example, historical events. “People conventionally refer to an action or occurrence taking place at a certain time at a specific location as an event” (Shaw, Troncy & Hardman, 2009, p. 153). A typical definition of 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).

This paper presents a simplified definition of events, which does not emphasize their geographical location. An event is defined here as a single occurrence involving at least one participant, which is a typed, n-ary association of typed entities under optional conditions. It can be a happened event, a scheduled event, or a single occurrence of a process. This is a generalized definition of event, which includes a static description of an entity (such as a jaguar is an animal), so it may not necessarily imply a state change of the entity.

The objective of this paper is to develop a model of entity and event relationship that can be used to represent knowledge identified from scientific documents and to facilitate knowledge discovery and organization. Swanson and Smalheiser (1999) discovered numerous previously unknown implicit relationships in the biomedical literature. For example, if one article reports that substance *A* causes condition *B* and another reports that condition *B* causes disease *C*, then we can infer that substance *A* could cause disease *C*. The entity and event relationship model aim to represent the knowledge extracted from the knowledge statements and to support inference for knowledge discovery.

A typical knowledge statement is expressed as *what an entity did what to another entity under what conditions (such as time, place, and method)*. In its simplest form, the knowledge statement can be expressed as *Entity A <relation> Entity B* where a relation is normally a verb phrase. It can be implemented as an entity-relationship model (Chen, 1976). In its linguistic form, it can be expressed as *NP <VP> NP* where *NP* is a noun phrase and *VP* is a verb phrase. Although causal verbs are desired, here *VP* includes stative verbs (such as *is*). Entities and relations are typed, and their types may vary across domains. 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. To support inference, we need a model that represents typed entities and typed relations unambiguously. In its simplest form, the entity and event relationship model needs to represent the following minimum number of triplets that are used to represent a knowledge statement. The form of representation of the triplets is similar to Resource Description Framework (RDF):

- Entity A1 <Relation R1> Entity B1 (abbreviated as A1 <R1> B1)
- Entity A1 <is a kind of> Entity Class A (abbreviated as A1 <ISA> Type A)
- Entity B1 <is a kind of> Entity Class B (abbreviated as B1 <ISA> Type B)
- Relation R1 <is a kind of> Relation Class R (abbreviated as R1 <ISA> Type R)

Then the entity-relationship knowledge statement can be expressed as:

(Type A) A1 <(Type R) R1> (Type B) B1

Based on the definition of events given above, an event has the power of encapsulating the relationship between entities. The above entity-relationship statement can be encapsulated in an event, and conditions can be associated with the event, such as:

[(Type A) A1 <(Type R) R1> (Type B) B1] <RC (RC1)> (Type C) Condition
 where RC is a condition relation.

An example condition is an *if* statement, such as *if temperature is above 50F*. In its simplest form, two events and their relationship can be represented as:

$$\begin{aligned} &[(\text{Type A}) A1 <(\text{Type R}) R1 > (\text{Type B}) B1] <(\text{Type R}) R12 > \\ &[(\text{Type A}) A2 <(\text{Type R}) R2 > (\text{Type B}) B2] \end{aligned}$$

This paper proposes an entity-event relationship model, which can represent the relationship between entities and events. It is designed based on existing event ontologies and can be used to represent knowledge which is composed of a series of events and event relationships.

The following research questions are explored:

- Q1. For managing accumulated knowledge extracted from scientific documents for supporting knowledge discovery through inference, are current event ontologies effective in representing simplified knowledge in the form of $(\text{Type A}) A1 <(\text{Type R}) R1 > (\text{Type B}) B1$?
- Q2. Are existing event ontologies effective in representing relationships between two events in the simplified form of $[(\text{Type A}) A1 <(\text{Type R}) R1 > (\text{Type B}) B1] <(\text{Type R}) R12 > [(\text{Type A}) A2 <(\text{Type R}) R2 > (\text{Type B}) B2]$?
- Q3. How can we model entity and event relationships to represent knowledge in the two forms addressed above, and more complex forms that involve multiple relationships between multiple entities/events?

METHODOLOGY

Since the proposed entity-event relationship model aims to support knowledge discovery through inference, and causal relationships are the most important relationships in supporting inference, this article first conducts a brief literature review on causal relationships. To answer these research questions, this article then briefly evaluates existing knowledge organization models and event models or ontologies, then proposes the entity and event relationship model based on the strengths of the existing event ontologies.

The term *ontology* has a philosophical origin and is defined as the study of being, of what there is (Bricker, 2014). In computer and information science, “an ontology formally represents knowledge as a set of concepts within a domain and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about concepts”¹ (IGI Global, 2020). The latter definition is used in this article. Since a formal ontology often has rigorous definitions and specifications for entities such as attributes and values, whereas this article does not address these specifications, a more appropriate term would be *knowledge-base*. However, since the event ontologies that are to be evaluated in this article use the term *ontology*, to be consistent, *ontology* is used. When applicable, the oil spill data set in the form of $(\text{Type A}) A1 <(\text{Type R}) R1 > (\text{Type B}) B1$ is used to explain the model. The data set contains approximately 5,000 concepts and 1,000 relationships extracted from about 300 documents comprising conference presentations, journal articles, news reports, and authoritative Web pages related to the 2010 Gulf of Mexico oil spill incident (Wu, Lehman, & Dunaway, 2015).

¹ What is Ontology? <https://www.igi-global.com/dictionary/interoperability-medical-devices-information-systems/21117>

BRIEF LITERATURE REVIEW ON CAUSAL RELATIONSHIPS

This section presents a brief literature review of causal relationships, including the expression of causal relationships and detection of causal rules, causal chains, visualization of causal relationships, and using mechanisms to assist inference based on causal relationships.

Identifying causal relationships from text and detecting causal rules from data is the first step in using causal relationships to support knowledge discovery. Khoo, Chan and Niu (2002) investigated how the cause-effect relation is expressed in text, and found that there is a wide variety of linguistic expressions for explicitly indicating cause and effect. Galton et al. (2015) developed an algorithm that detects causal rules from spatio-temporal data. Experiments with real data concerning the movement of fish in a large Australian river system found that the algorithm was able to identify causal rules compatible with perpetuation that explain a significant proportion of the fish movements whereas perpetuation takes environmental conditions as the drivers of ongoing process of movement.

A causal chain is a chain of entities linked by causality, and according to Mizoguchi et al. (2011, p.65) “is composed of one or more pairs of entities/events such as a causal event and an effect event, in which the latter has been caused by the former. The effect becomes another cause that causes another effect in the case of multiple-pair chains”. They introduced the core idea of a disease as causal chain and clinical imbalance, and defined “a disease as a dependent continuant constituted of one or more causal chains of clinical disorders” (p. 63). Their model of diseases can be considered as a concretization of the Ontology for General Medical Science view of disease as disposition. Petrochemical plants have strong interdependencies between various units. One unit’s failure often triggers cascade failures as a chain reaction. Hu (2016) aimed to describe the interdependency in the complex system and to discover the root cause of an abnormal event. He used Granger causality test to study the fault interdependency by analyzing the relationship between process parameters of petrochemical units in a petrochemical plant and establishing a cause and effect diagram of the process parameters. A fault quantitative cause and effect diagram can be established based on the degree of the causal relationship of the process parameters. The root cause of the alarm can be identified by using the diagram to find the path with the biggest quantitative value of causal relationship, which can be considered as the most probable fault propagation path in the petrochemical units according to an alarm.

Visualizing events and causes facilitate the understanding and organization of events. Allen, Wu and Luo (2005) described events as simple state transitions, causes as the factors which make state changes, and a state change as a change in one or more attributes. They built a browser to explore the network of events and causes described in a science text, and presented the network of causal links as a type of interactive hypertext map, called *an interactive causal schematic*. Allen (2011) proposed temporally constrained causal relationships as a key for organizing interwoven historical events, and described a panel-oriented visualization interface that show causal chains of events and causes of events leading up to the Civil War.

Allen (2018) described a Semantic Modeling Framework for structured description of mechanisms, which are entities and activities organized to produce regular changes from start to finish. Mechanisms have four basic features: a phenomenon, parts, causings, and organization.² Having a priori candidate mechanisms is helpful in inferring causal relationships in a complex data set such as in medical diagnosis (Russo & Williamson, 2007). On the one

² Mechanisms in Science, <https://plato.stanford.edu/entries/science-mechanisms/> (Section 2).

hand, the proposed entity-event model can be used to build mechanisms for an application. On the other, mechanisms can be used to assist inference.

EVALUATION OF KNOWLEDGE ORGANIZATION MODELS AND EVENT ONTOLOGIES

This paper builds upon previous work in which Wu and Bai (2017) briefly evaluated four knowledge organization and visualization models and tools for representing simplified knowledge in the form of *Entity A* <*Relation*> *Entity B*. They are: Cmap for building concept map (Novak & Cañas, 2008), Ontopia for creating topic maps (Hatzigaidas et al., 2004), Gephi for graph network analysis (Bastian et al., 2009), and Protégé for creating ontologies (<http://protege.stanford.edu>). The four models and tools were found weak in representing and visualizing simplified knowledge in the form of *(Type A) A1* <*(Type R) R1*> *(Type B) B1*. None of them can represent and visualize relationships between two events in the simplified form of [*(Type A) A1* <*(Type R) R1*> *(Type B) B1*] <*(Type R) R12*> [*(Type A) A2* <*(Type R) R2*> *(Type B) B2*] (Wu & Bai, 2017).

This section briefly reviews several event models or ontologies, which define classes and properties for modeling events and their relationships. The purpose of the evaluation is to understand the commonalities and differences between event ontologies and to shed light on how to design a new ontology based on their strengths. The evaluation criteria include (1) application domain of the ontology, (2) availability of entity classes and relation classes that are used to represent meanings of entities and relationships unambiguously, (3) ability to represent entity relationship and event relationship and (4) scope and richness of relationship.

This paper focuses on four kinds of structural relationships and studied five event models. The former covers temporal, mereological, causal, and correlation relationships because they can be used to support knowledge discovery through inference.

Event Models Studied

CIDOC Conceptual Reference Model (CRM)

- The CIDOC CRM³ is “a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information... it defines and is restricted to the underlying semantics of database schemata and document structures used in cultural heritage and museum documentation in terms of a formal ontology” (ICOM/CIDOC, 2018, p. i). According to the website of CIDOC CRM (<http://www.cidoc-crm.org>), the CIDOC CRM provides “definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.” So, its entity classes and relations are restricted to cultural heritage and museum contexts. CIDOC CRM provides a taxonomy of entities and a taxonomy of relations (“property” in its original term). The model has a focus on “time” and defines the *E2 Temporal Entity* class, which “comprises all phenomena, such as the instances of *E4 Periods*, *E5 Events and states*, which happen over a limited extent in time” (ICOM/CIDOC, 2018, p. 2). The model defines several types of relationships, such as: Temporal: *P4 has time-span (is time-span of)*, *P82 at some time within*, *P114 is equal in time to*, *P117 occurs during (includes)*, *P118 overlaps in time with (is overlapped in time by)*, *P119 meets in time with (is met in time by)*, *P120 occurs before (occurs after)*

³ <http://www.cidoc-crm.org/version/version-6.2.3>

- Mereological: *P46 is composed of* (forms part of), *P148 has component* (is composed of)
- Correlation: *P15 was influenced by* (influenced), *P69 has association with* (is associated with)
- Causal: *P123 resulted in* (resulted from)

Event Ontology

The Event Ontology⁴ is “centered around the notion of event,” which is “an arbitrary classification of a space/time region, by a cognitive agent. An event may have actively participating agents, passive factors, products, and a location in space/time” (Raimond & Abdallah, 2007). The ontology was developed in the Centre for Digital Music at the University of London to be used in conjunction with music-related ontologies (Shaw et al., 2009). It has only three classes (i.e., Event, Factor, Product) and 18 properties. It uses the timeline ontology, which defines a timeline concept with 22 classes and 40 properties (such as *after*, *at*, *before*, *during*). Mereological relationships are very limited.

NewsML-G2

NewsML-G2⁵ is a standard for exchanging news event information, developed by the International Press Telecommunication Council (IPTC). The information about an event consists of *what* (an event name), *where*, *when* and *who* is involved. “Large news events may be split into a series of smaller, manageable events, arranged in a hierarchy which can express parent-child and peer relationships” (IPTC, 2019). Hence, there are four types of event relationships: broader, narrower, related, and sameAs. “The <dates> property contains the dates and times of the event, expressed using the child elements of <start>, <end>, and <duration>” (IPTC, 2019).

DOLCE+DnS Ultralite Ontology (DUL)

DUL⁶ is a lightweight, foundational ontology for modeling either physical or social contexts. It models any physical, social, or mental process, event, or state. It provides “a set of upper level concepts that can be the basis for easier interoperability among many middle and lower level ontologies” (DUL, 2010). It defines dozens of generic entity classes (such as Concept, Diagnosis, Design, Description, Entity, Event) and some generic relationships, such as:

- Temporal: *hasTimeInterval*, *isTimeIncludedIn*, *precedes*, *hasEventDate*, *hasIntervalDate*, *directlyFollows*, *directlyPrecedes*, *follows*
- Mereological: *hasPart*, *hasComponent*, *hasConstituent*, *isPartOf*, *isComponentOf*, *isConstituentOf*, *expands*, *includesEvent*, *isEventIncludedIn*
- Correlation: *isRelatedTOConcept*, *isRelatedToDescription*, *associatedWith*, *expresses*, *hasCommonBoundary*, *farFrom*

DUL defines two relations for expressing conditions: *isPreconditionOf*, *isPostconditionOf*.

⁴ <http://motools.sourceforge.net/event/event.html>

⁵ <https://www.iptc.org/std/NewsML-G2/guidelines/>

⁶ Ontology URL: <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl>

Event Model F

Event Model F is “a formal model of events designed to facilitate interoperability in distributed event-based systems. The model is based on DUL, the foundational ontology and provides comprehensive support to represent time and space, objects and persons, as well as mereological, causal, and correlative relationships between events” (Scherp et al., 2010). Based on DUL, the F model defines some generic relationships, such as:

- Temporal: *SpatioTemporalConstraint*, *TemporalConstraint*
- Mereological: *component*, *EventCompositionDescription*, *EventCompositionConstraint*, *RelevantComposition*, *EventCompositionSituation*
- Causal: *cause*, *effect*, *eventCausalityDescription*, *EventCausalitySituation*, *RelevantCausality*
- Correlation: *correlate*, *EventCorrelationDescription*, *EventCorrelationSituation*, *RelevantCorrelation*

Comparison of Event Models

All of the five models above can represent event relationships. All them, except NewsML-G2, define entity classes and can represent entity relationships. NewsML-G2 has only four types of event relationships. A comparison of the models in terms of domain, relationship scope, and availability of four types of relationship is presented in Table 1.

The comparison of the five event ontologies shows that generally event ontologies define some classes (such as Event, Entity), properties, and relationships. There are great differences in the classes, properties, and relationships between the ontologies. The three domain ontologies show greater differences among themselves than between them and the two generic, upper-level ontologies. Despite the differences, the domain ontologies serve their own purposes.

In Table 1, absolute temporal relation is distinguished from relative temporal relation. Relative temporal relations are more important than absolute relations in describing the events involved in a process (such as the ionization of one atom), and so is more useful for the proposed entity and event model.

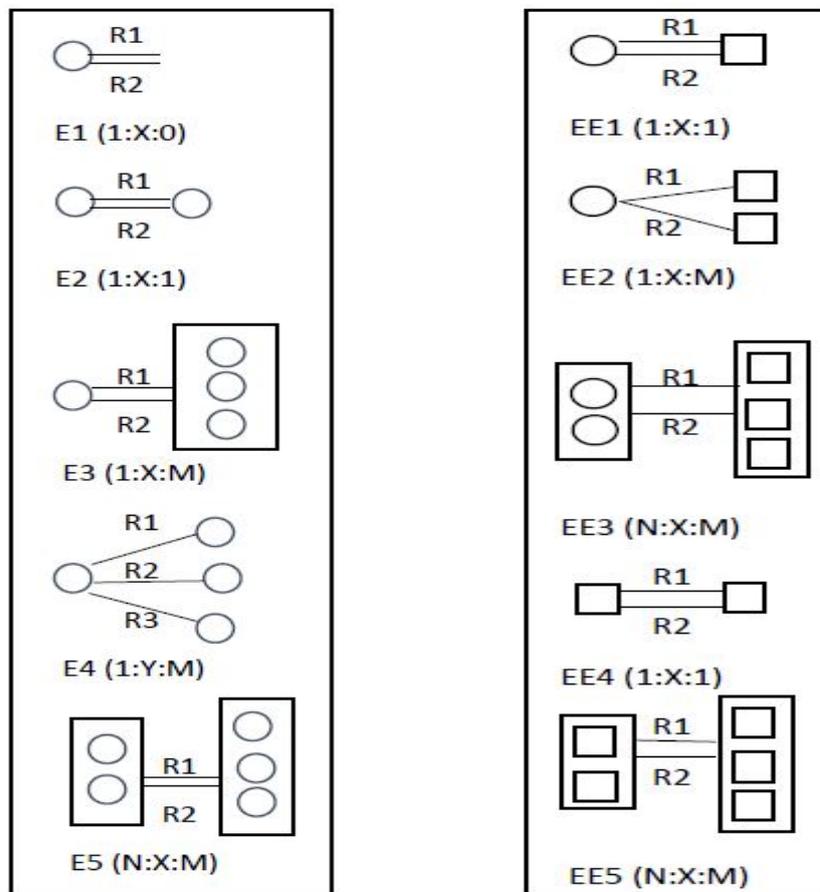
The first two research questions can be answered now. The three domain event ontologies can represent simplified knowledge in the form of (Type A) A1 <(Type R) R1> (Type B) B1, and can represent relationships between two events in the simplified form of [(Type A) A1 <(Type R) R1> (Type B) B1] <(Type R) R12> [(Type A) A2 <(Type R) R2> (Type B) B2], but their entity classes, event class, and relations are so narrow (domain-specific) and rich that they are only useful in serving their specific applications. The two generic event ontologies provide a skeleton of the event class, entity classes and relations. They are so wide and skeletal that they must be enriched to be useful for representing knowledge for discovery.

FrameNet⁷ defines more than 1,000 semantic frames, which are typed events (such as cooking, revenge). Since science is a very broad domain, and scientific documents may contain numerous types of events, defining typed events seems impossible. From the comparison above, it is clear that DUL and Event Model F can be used to derive a skeleton of the entity and event model, but some specific relations defined in the three domain event ontologies can be used in the new model. The next section describes the proposed entity-event relationship model.

⁷ <https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet>

Table 1. A comparison of five event models

Model	Domain	Relations Scope	Temporal		Mereological, Causal, Correlation Relations
			Absolute	Relative	
CIDOC-CRM	Cultural heritage	narrow, rich	Yes	Yes	mereological (limited) causal (limited) correlation (limited)
Event Ontology	Music	narrow, rich	Yes	Yes	mereological (limited)
NewsML-G2	News	narrow, limited	Yes	No	
DUL	Generic, Upper	wide, skeleton	Yes	Yes	mereological, correlation
F	Generic, Upper	wide, skeleton	Yes	Yes	mereological, causal, correlation



Note: Relations are directional but are not displayed. A circle represents an entity. A square box represents an event. A rectangular box represents a group (or container) of entities or events.

Figure 1. Entity relationship (E1-E5 in left figure), entity & event relationship (EE1-EE5 in right figure)

PROPOSAL: AN ENTITY AND EVENT RELATIONSHIP MODEL

This section first describes the main kinds of entity and event relationships that are identified in scientific documents, and then describes the model, followed by the implementation of the model. Finally, two examples of using the model to support knowledge organization and discovery are given to show the functionality of the model.

Main Kinds of Entity and Event Relationships

The following lists five main kinds of entity and event relationships identified from the oil spill data set:

1. *Static description of entity*

Example sentence: “In the aftermath of oil spills, amphipod and echinoderm populations may completely disappear or be severely reduced” (Buckburn, et al., 2014, p. 26).

In this statement, the entity is *amphipod and echinoderm populations*, and its status is *may completely disappear or be severely reduced*. The condition is *in the aftermath of oil spills*. No other entities are related to this entity. This can be expressed as Entity A1 <Relation R1> Null (Condition).

2. *Entity-entity relationship*

Example sentence: “Trapped birds get weighed down by the oil” (Rebello, 2010).

In this statement, two entities are *trapped birds* and *the oil*. Their relation is *get weighed down by*. This can be expressed as Entity A1 <Relation R1> Entity B1.

3. *Entity-multiple entities relationship*

Example sentence: “In a cross-sectional study of clean-up workers and volunteers, [authors] identified neurological, dermal, ocular, and respiratory symptoms, as well as lumbar pain, in individuals involved in clean-up activities” (Sathiakuma 2010, p. 46).

In this statement, Entity *individuals involved in clean-up activities* is related to multiple entities *neurological, dermal, ocular, and respiratory symptoms*, and *lumbar pain*. This can be expressed as Entity A1 <Relation R1> Entity B1, B2, etc.

4. *Entity-event relationship*

Example sentence: “The mild character of the pathological changes suggests that petroleum oil toxicosis causes multiple sublethal changes that ...[affect] the ability of the birds to survive at sea” (Balseiro et al., 2005).

In this statement, Entity *the mild character of the pathological changes* is related to an event, in which Entity *petroleum oil toxicosis* is related to *the ability of the birds to survive at sea*. This can be expressed as Entity A1 <Relation R1> [Entity B1 <Relation R2> Entity B2].

5. *Event-event relationship*

Example sentence: “Suspension-feeding invertebrates such as mollusks metabolize hydrocarbons slowly, so walrus would be exposed to contamination by eating them” (Nuka Research and Planning Group, 2010, p. 59).

In this statement, Event *Suspension-feeding invertebrates such as mollusks metabolize hydrocarbons slowly* causes Event *walrus would be exposed to contamination by eating them*. This can be expressed as [Entity A1 <Relation R1> Entity B1] <Relation R12> [Entity A2 <Relation R2> Entity B2] in the simplest format.

Other kinds of entity and event relationships can also be derived from these five main kinds, such as multiple entities-multiple entities relationship, entity-multiple events relationship, multiple entities-multiple events relationships, event-multiple events relationship, and multiple events-multiple events relationships. Figure 1 illustrates these kinds of relationships.

Description and Implementation of the Model

Generally, the entity-event relationship model needs to define some entity classes and an event class. Entity classes are generally defined in a classification scheme or taxonomy. It can be difficult to define a taxonomy for all sciences that is useful for supporting inference. There are comprehensive library classification systems, such as the Library of Congress Classification and the Dewey Decimal Classification. They are designed for classifying books, and may be too general to support fine-grained inference. However, any fine-grained domain-specific taxonomy, or cross-domain special collection taxonomy, such as the Oil Spill Taxonomy (Wu, 2016) can be plugged into the model.

The model does not define an Event class and its properties (such as event name, time, location, participants) as in the five event ontologies. Instead, it defines an event as a container in the language of object-oriented programming. Borrowing the idea from the object-oriented data model (OODM) and the extended relational data model (ERDM) (Rob & Coronel, 2002), an event can be implemented as an object in object-oriented modeling. The event container can encapsulate entities and their relationships. Conditions of each event can also be encapsulated in the event. Allen and Jones (2018) proposed developing a semantic model layer to describe the interaction of the Basic Formal Ontology (BFO2) entities which as a pure upper ontology, focuses on identifying types of entities and allows domain applications to extend the entity types. They described a programming environment called the eXtended Formal Ontology (XFO) for implementing the model layer. Such a programming environment may be useful in the implementation of the entity-event model.

A relation taxonomy is needed in supporting fine-grained inference. Such a taxonomy has been designed for the oil spill document collection (Wu & Yang, 2015). Mereological, temporal, causal, and correlation relations are the most important relations for facilitating inference. The existing event ontologies define some of these relations and can be incorporated into the relation taxonomy.

The proposed entity-event relationship model has the power of representing knowledge in the entity-relationship model, and supporting knowledge discovery through inference based on the relations between entities, between events, and between entities and events. The following example shows how to use the model to represent knowledge as relationships between entities and events:

Example sentence: “Suspension-feeding invertebrates such as mollusks metabolize hydrocarbons slowly, so walrus would be exposed to contamination by eating them” (Nuka Research and Planning Group, 2010).

The knowledge in this sentence can be represented as the following events and their relationships:

- Event 1: (Type) mollusks <(Type) IS A> (Type) suspension-feeding invertebrates.
- Event 2: (Type) suspension-feeding invertebrates <(Type) metabolize slowly> (Type) hydrocarbons,
- Event 3: (Type) walrus <(Type) eat> (Type) suspension-feeding invertebrates,

Event 4: (Type) walrus <(Type) would be exposed to> (Type) contamination.
Event 2 and Event 3 <cause> Event 4. This corresponds to EE5 in Figure 1. In this example, every entity has a type (or class) in an entity taxonomy, and every relation has a type (or class) in a relation taxonomy.

The following example, adapted from Wu and Yang (2015), shows how to use the model to facilitate knowledge discovery through inference:

(Type) oil <(Type) kill> (Type) Arctic phytoplankton,
(Type) Arctic phytoplankton <(Type) be consumed by> (Type) Arctic cod,
(Type) Arctic cod <(Type) be consumed by> (Type) ringed seal,
Inferred knowledge: (Type) oil <(Type) may kill> (Type) ringed seal.

To implement knowledge discovery through inference, a reasoning mechanism or inference engine needs to be designed. Having a priori candidate mechanisms is helpful in inferring causal relationships in a complex data set such as in medical diagnosis (Russo & Williamson, 2007). Allen (2018) described a Semantic Modeling Framework for structured description of mechanisms. An inference engine applies logical rules to a knowledge base to deduce new knowledge. For example, Shanahan (1999) proposed the event calculus, which is a logic-based formalism for representing actions and their effects. An expert system has a knowledge base and an inference engine. By applying similar technologies, mechanisms and inference engines can be built for the proposed model to enable knowledge discovery through inference.

CONCLUSION

This paper aims to propose an entity and event relationship model for representing knowledge extracted from scientific documents. It is designed based on existing event ontologies, and can be used to represent knowledge which is composed of a series of events and event relationships. Five event ontologies were evaluated for their effectiveness in representing entity relationship and event relationship. They are CIDOC Conceptual Reference Model (CRM), Event Ontology, NewsML-G2, DOLCE+Dns Ultralight Ontology (DUL), and Event Model F. The three domain event ontologies (i.e., CIDOC CRM, Event Ontology, NewsML-G2) can represent event relationship, and two event ontologies (i.e., DUL and Event Model F) can represent entity relationship, but their entity classes, event class, and relations are so narrow (domain-specific) and rich that they are useful in serving their own specific purposes. The two generic event ontologies (i.e., DUL, Event Model F) provide a skeleton of the event class, entity classes and relations. They are so broad that they must be enriched to be useful for representing knowledge for discovery.

The proposed entity-event relationship model defines five main kinds of entity and event relationships: static description of entity, entity-entity relationship, entity-multiple entity relationship, entity-event relationship, and event-event relationship. The model does not define an Event class and its properties (such as event name, time, location, participants) as in the five event ontologies. Instead, it defines an event as an object or container in the language of object-oriented programming. The event container can encapsulate entities and their relationships. Conditions of each event can also be encapsulated in the event. An entity taxonomy and a relationship taxonomy can be plugged into the model to organize knowledge and support knowledge discovery. An inference engine needs to be designed to execute reasoning.

The limitation of the model is that it can only represent knowledge in a binary relationship of entities. For n-ary relationship (where n is bigger than 2), it has to be divided into multiple binary relationships. A better approach for representing n-ary relationship needs to be designed in the future.

ACKNOWLEDGEMENT

The author would like to thank Dr. Robert B. Allen for his comments on the abstract of a presentation at the 2019 Workshop on Ontology & Rich Semantics (RICH 2019). This paper was developed based on the abstract of the presentation.

REFERENCES

- ACM. (2012). *The 2012 ACM computing classification system*.
<https://www.acm.org/publications/class-2012>
- Allen, R.B. (2011). Visualization, causation, and history. In *Proceedings of iConference 2011 Seattle, Washington, USA, 8-11 February 2011* (pp. 538–545). ACM.
<http://boballen.info/RBA/PAPERS/ICONF2011/VCH.pdf>
- Allen, R.B. (2018). Issues for using semantic modeling to represent mechanisms. Paper presented at the *Workshop on Ontology and Rich Semantics 2018*.
<https://arxiv.org/abs/1812.11431>
- Allen, R.B., & Jones, T.K. (2018). XFO: Toward programming rich semantic models.
arXiv:1805.11050, 2018. <http://doi.org/10.13140/RG.2.2.28704.69127>
- Allen, R.B., Wu, Y., & Luo, J. (2005). Interactive causal schematics for qualitative scientific explanations. In *Proceedings of 8th International Conference on Asian Digital Libraries (ICADL), 12-15 December 2005, Bangkok, Thailand* (LNCS 3815, pp. 411-415). Springer.
- Balseiro, A., Espí, A., Márquez, I., Pérez, V., Ferreras, M.C., Marín J.F., & Prieto, J.M. (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.
<https://www.jwildlifedis.org/doi/pdf/10.7589/0090-3558-41.2.371>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of Third International AAAI 2009 Conference on Weblogs and Social Media, 17-20 May 2009, San Jose, USA* (pp. 361-362). <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>
- Bricker, P. (2014). Ontological commitment. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Center for the Study of Language and Information, Stanford University.
<https://plato.stanford.edu/entries/ontological-commitment/>
- Chen, P.P.S. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems, 1*(1), 9-36.
- DUL (2010). Ontology: DOLCE+DnS Ultralite ontology overview. Retrieved from
http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite
- Galton, A., Duckham, M., & Both, A. (2015). Extracting causal rules from spatio-temporal data. In *Proceedings of the 12th International Conference on Spatial Information Theory (COSIT), 12-16 October 2015, Santa Fe, New Mexico, USA* (pp. 23-43).
https://doi.org/10.1007/978-3-319-23374-1_2
- Hatzigaidas, A., Papastergiou, A., Tryfon, G., & Maritsa, D. (2004). Topic map existing tools: A brief review. In *Proceedings of the 2004 International Conference on Theory and*

- Applications of Mathematics and Informatics, Thessaloniki, Greece* (pp. 185-201).
<http://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.392.4225&rep=rep1&type=pdf>
- Hu, J. (2016). Fault propagation analysis and causal fault diagnosis for petrochemical plant based on Granger causality test [video file]. Presented at *AIChE Spring Meeting and Global Congress on Process Safety, 12 April 2016, Houston, Texas, USA*.
<https://www.aiche.org/academy/videos/conference-presentations/fault-propagation-analysis-and-causal-fault-diagnosis-petrochemical-plant-based-on-granger-causality>
- ICOM/CIDOC (2018). *Definition of the CIDOC conceptual reference model, version 6.2.3, May 2018*. <http://www.cidoc-crm.org/version/version-6.2.3>
- IPTC (2019). *NewsML-G2 guidelines*. <https://www.iptc.org/std/NewsML-G2/guidelines/>
- Khoo, C., Chan, S., & Niu, Y. (2002). The many facets of the cause-effect relation. In Green R., Bean C.A., Myaeng S.H. (Eds.), *The Semantics of Relationships* (pp. 51-76). Springer.
- 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>
- Mizoguchi, R., Kozakil, K., Kou, H., Yamagata, Y., Imai, T., Waki, K., & Ohe, K. (2011). River flow model of disease. In *Proceedings of International Conference on Biomedical Ontology, 28-30 July 2011, Buffalo, NY, USA* (pp. 63-70).
https://www.researchgate.net/profile/Riichiro_Mizoguchi/publication/267303683_River_Flow_Model_of_Diseases/links/546169280cf27487b4527593/River-Flow-Model-of-Diseases.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.php>
- Nuka Research & Planning Group, LLC. (Nov 2010). *Oil spill prevention and response in the U.S. Arctic Ocean: Unexamined risks, unacceptable consequences*.
<https://fdocuments.us/document/oil-spill-prevention-and-response-in-the-us-weblaw-nuka-research-and-planning.html>
- Pyysalo, S., Ohta, T., Rak, R., Sullivan, D., Mao, C., Wang, C., ... Ananiadou, S. (2012). Overview of the ID, EPI and REL tasks of BioNLP Shared Task 2011. *Bioinformatics*, 13(S11). <https://doi.org/10.1186/1471-2105-13-S11-S2>
- Raimond, Y., & Abdallah, S. (2007). *The event ontology*.
<http://motools.sourceforge.net/event/event.html>
- Rebello, S. (2010, August 13). Migratory birds could be affected. *Hindustan Times*.
<https://www.hindustantimes.com/india/migratory-birds-could-be-affected/story-QNq5PqdGe2cWkOW1fIzhfJ.html>
- Rob, P., & Coronel, C. (2002). *Database systems: Design, implementation, & management* (5th ed.). Boston, MA: Course Technology.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in Philosophy of Science*, 21(2), 157-170.
- Sathiakuma, N. (2010). Short-term physical effects. In M.A. McCoy & J.A. Salerno (Eds.), *Assessing the effects of the Gulf of Mexico oil spill on human health: A summary of the June 2010 Workshop* (pp. 45-50). The National Academies Press.
- Shaw, R. Troncy, R., & Hardman, L. (2009). LOD: Linking open descriptions of events. In *Proceedings of ASWC 2009, 6-9 December 2009, Shanghai, China* (pp. 153-167). Springer. https://link.springer.com/chapter/10.1007/978-3-642-10871-6_11

- Shanahan, M. (1999). The event calculus explained. In M.J. Wooldridge & M. Veloso (Eds.), *Artificial intelligence today* (Lecture Notes in Artificial Intelligence, 1600, pp. 409-430). Springer.
- Scherp, A., Franz, T., Saathoff, C., & Staab S. (2010). Ontology: Event Model F overview. http://ontologydesignpatterns.org/wiki/Ontology:Event_Model_F
- Swanson, D., & Smalheiser, N. (1999). Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. *Library Trends* 48(1), 48-59.
- Wu, Y. (2016). Oil spill taxonomy. In Y. Wu (Ed.), *Oil spill impacts: Taxonomic and ontological approaches* (pp. 29-170). CRC Press.
- Wu, Y., & Bai, R. (2017). An event relationship model for knowledge organization and visualization. In *Proceedings of the North American Symposium on Knowledge Organization (NASKO), 15-16 June 2017, Champaign, Illinois* (pp. 216-230). <https://journals.lib.washington.edu/index.php/nasko/article/view/15241>
- Wu, Y., Lehman, A., & Dunaway, D. (2015). Evaluations of a large topic map as a knowledge organization tool for supporting self-regulated learning. *Knowledge Organization* 42(6), 386-398.
- Wu, Y., & Yang, L. (2015). Construction and evaluation of an oil spill semantic relation taxonomy for supporting knowledge discovery. *Knowledge Organization* 42(4), 222-231.