Research and Applications

Time event ontology (TEO): to support semantic representation and reasoning of complex temporal relations of clinical events

Fang Li ,[1†] **Jingcheng Du** ,[1†] **Yongqun He,**[2] **Hsing-Yi Song,**[1] **Mohcine Madkour,**[3] **Guozheng Rao,**[4] **Yang Xiang,**[1] **Yi Luo,**[1] **Henry W. Chen** ,[1,5] **Sijia Liu,**[6] **Liwei Wang,**[6] **Hongfang Liu,**[6] **Hua Xu** ,[1] **and Cui Tao**[1]

[1]School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA, [2]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, USA, [3]Advanced Analytics, Cummins Inc, Columbus, Indiana, USA, [4]College of Intelligence and Computing, Tianjin University, Tianjin, China, [5]University of Texas Southwestern Medical Center, Dallas, Texas, USA, and [6]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

[†]Co-first authors.

Corresponding Author: Cui Tao, PhD, School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin St, Houston, TX 77030, USA (cui.tao@uth.tmc.edu)

ABSTRACT

Objective: The goal of this study is to develop a robust Time Event Ontology (TEO), which can formally represent and reason both structured and unstructured temporal information.

Materials and Methods: Using our previous Clinical Narrative Temporal Relation Ontology 1.0 and 2.0 as a starting point, we redesigned concept primitives (clinical events and temporal expressions) and enriched temporal relations. Specifically, 2 sets of temporal relations (Allen's interval algebra and a novel suite of basic time relations) were used to specify qualitative temporal order relations, and a *Temporal Relation Statement* was designed to formalize quantitative temporal relations. Moreover, a variety of data properties were defined to represent diversified temporal expressions in clinical narratives.

Results: TEO has a rich set of classes and properties (object, data, and annotation). When evaluated with real electronic health record data from the Mayo Clinic, it could faithfully represent more than 95% of the temporal expressions. Its reasoning ability was further demonstrated on a sample drug adverse event report annotated with respect to TEO. The results showed that our Java-based TEO reasoner could answer a set of frequently asked time-related queries, demonstrating that TEO has a strong capability of reasoning complex temporal relations.

Conclusion: TEO can support flexible temporal relation representation and reasoning. Our next step will be to apply TEO to the natural language processing field to facilitate automated temporal information annotation, extraction, and timeline reasoning to better support time-based clinical decision-making.

Key words: time event ontology, clinical event, temporal relational reasoning, Allen's interval algebra, basic time relations, clinical decision support

INTRODUCTION

Time is an important and pervasive concept of the real world.[1] In the clinical domains, temporal information elucidates the occurrence or changing status of medical events (eg, visits, laboratory tests, procedures). Accurate profiling of clinical timelines could benefit condition trajectory tracking, adverse reaction detecting, disease risk prediction, etc.[2–7] The widespread adoption of electronic health records (EHRs) provides great opportunities for accessing large amounts of clinical data. However, approximately 80% of the EHRs comprise unstructured data, and a wealth of temporal information is hidden therein. Due to the implicit nature of temporal expressions, often characterized by a considerable degree of under-specification,[8] automatically constructing a timeline of clinical events is quite challenging. Formal modeling of temporal concepts and relationships which could support subsequent temporal reasoning is a crucial prerequisite to overcoming this hurdle.

As an important research topic, temporal modeling has seen a lot of efforts over the past several decades. Among them, XML-based annotation scheme is a popular type. ISO-TimeML is a rich specification language for the event and temporal expressions,[9] which has been applied in several natural language processing (NLP) shared tasks, such as 2012 i2b2 Challenge[4] and Clinical TempEval in SemEval from 2015 to 2017.[10–12] It specifies 4 major data structures (TIMEX3, EVENT, SIGNAL, and LINK) and combines a broad range of syntactic and semantic rules to represent time, event, and temporal relations.[13] However, due to the high diversity of natural language, the strict calendar-based scheme is unable to represent some important types of time expressions (eg, Saturdays since March), and is not easily amenable to machine learning.[14] Targeting these limitations, Bethard and Parker (2016) proposed an alternate scheme named Semantically Compositional Annotation Scheme for Time Normalization (SCATE). By annotating time expression as compositional entities and defining several mathematical operators, SCATE can represent a wider variety of time expressions[15] and has been applied in SemEval 2018 as the time normalization standard.[16] To improve inter-annotator agreements (IAA) of temporal relation annotation by the existing schemes (such as TimeML), Ning et al (2018) proposed multi-axis modeling.[17] By anchoring events to different semantic axes, the modeling simplifies the task by (1) comparing only the events from the same (main) axis, and (2) using only start points of events. A pilot study on a subset of TimeBank-Dense showed a significant IAA improvement.

Compared with the XML-based annotation scheme, ontologies encoded by Web Ontology Language (OWL) have better expressiveness by defining concepts with machine-readable semantics.[18] Ontologies are systematic representations of knowledge, comprising a set of concepts and their formal relationships.[19,20] The use of description logics enables computational reasoning procedures to identify facts that are implied but not explicitly stated in the original data, which would play a significant role in assisting temporal reasoning. The time-related ontologies include Time Ontology in OWL,[21,22] SWRL ontology,[23] DARPA Agent Markup Language ontology of time,[24] and Reusable Time Ontology.[25] Newly constructed Time Ontology in OWL by W3C[22] provides a vocabulary to describe the temporal properties of resources and expresses facts about ordering relations among time instants and intervals. It covers temporal expressions and relations to a remarkable degree.

Though sharing the basic nature of universal time, temporal expressions and relations in clinical narratives have their own char-acteristics, such as greater association with periodic intervention (chemotherapy, rehabilitation training, etc.), with higher expectations for complete timeline reasoning. The existing ontologies are mainly designed for the general domain, not for clinical applications; they represent relations between time instants and intervals, without covering temporal relations between events. For these reasons, we propose Time Event Ontology (TEO) in this article. Figure 1 illustrates the overall framework of TEO design and reasoning support. TEO is extended from our previous efforts including Clinical Narrative Temporal Relation Ontology (CNTRO 1.0)[26–28] and CNTRO 2.0,[29] which were designed as clinical-narrative-oriented temporal relation modeling ontologies. The classes and properties were redesigned in depth. Two sets of temporal relation were used: (1) relationships adopted from Allen's interval algebra, which consists of a set of pairwise disjoint binary relations and is a widely used calculus for temporal reasoning,[30,31] applicable to situations where complete time information is available; (2) a newly proposed suite of Basic Time Relations (BTRs), to cover the common situations in which temporality is partially known. TEO's expressiveness was evaluated through annotating EHR data. With the support of a Java-based TEO reasoner, the answers for the time-related queries showed promising results. Our research contributions embrace the following aspects: (1) formally defined a robust ontology modeling temporal concepts and relations; (2) used real clinical notes to annotate and evaluate the coverage ability of the ontology; and (3) leveraged a Java-based TEO reasoner to realize complex timeline reasoning.

METHOD AND ONTOLOGY DESIGN

Meta-level design

Under the contemporary paradigm, an ontology comprises classes (concepts), individuals (instances), and properties.[18] A class is a set of entities within a domain that defines a group of individuals that share some common properties. Individuals are instances (concrete examples) of classes. Properties depict the characteristics of a class, stating relationships between individuals (object properties) or from individuals to data values (data properties).[32] For the object property, domain limits the individuals to which the property can be applied, and range limits the individuals that the property may have as its value. The third type of property, annotation properties, are used to associate additional information with ontologies, entities, and axioms.[18] TEO is defined using the OWL. Since TEO aims to model temporal expressions and relations of clinical events, *time* and *event* are the 2 main OWL classes. Class *time instant* is designed to represent a single time point, and class *granularity* is used to describe the basic level of the time unit. Class *time interval* is used to represent a continuous period, combined with class *duration* to describe its time length. A subclass *periodic time interval* is devised to describe a periodically or repeatedly occurred *time interval*. Moreover, a set of object properties are introduced to depict relations, particularly the temporal relations between classes, and a variety of data properties to describe the data features. Figure 2 illustrates the meta-level design of TEO.

Class design for modeling temporal expressions
Event

Class *event* is designed to represent time-oriented medical events, which includes any sort of occurrences, states, procedures or situations that occurs on a timeline.[26] Several subclasses are designed to
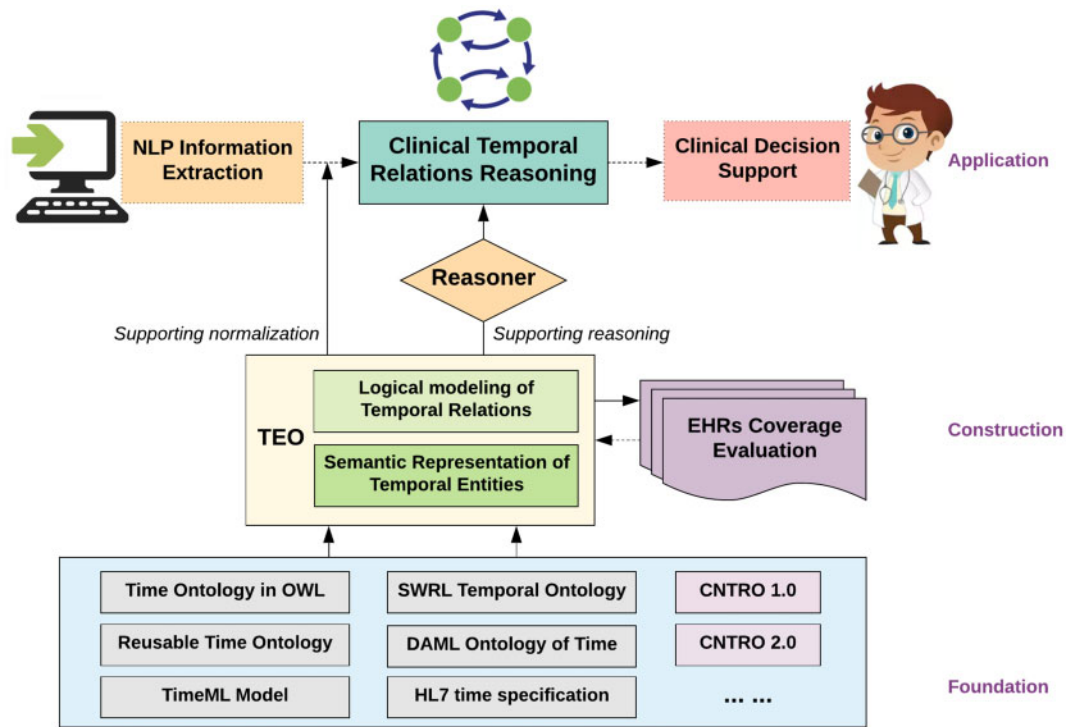
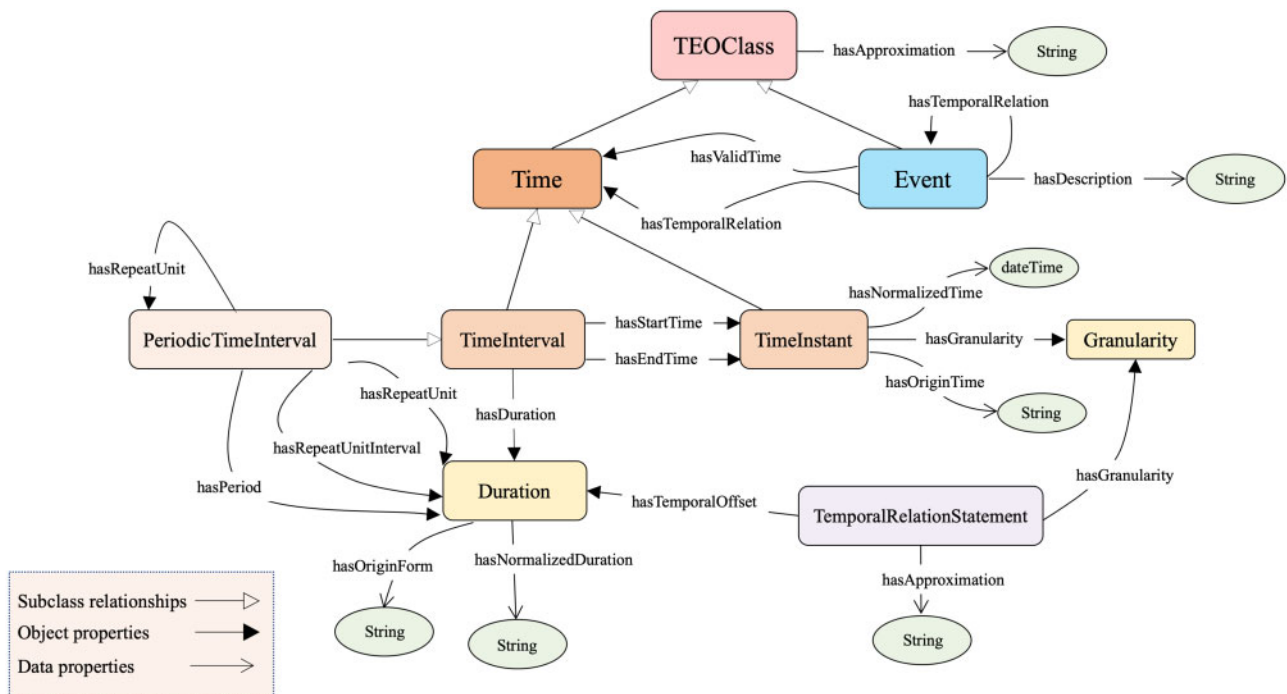**Figure 1.** Overview Framework of TEO Design and Reasoning Support.

Figure 2. Meta-level Design of TEO.

cover the common clinical events (eg, *clinical intervention*, *diagnosis* and *test*). More subclasses could be further defined based on the individual use case. Object property *hasValidTime* links an *event* individual with a *time* individual to describe its temporal dimension.

Time

Class *time* is designated to represent a multiplicity of temporal expressions. It is further subdivided into *time instant*, *time interval*, and a set of subclasses.

Time instant and granularity. Class *time instant* represents a single segment on the timeline that could be aligned to some unit of a calendar system. It could be a specific day (eg, "1992-03-05") or a specific year (eg, "1992"), etc. Object property *hasValidTime* links an *event* individual with a *time instant* individual. For example, "The subject received the flu vaccine on 26 April 2009," the <event> ("The subject received the flu vaccine") *hasValidTime* <time instant> ("26 April 2009").

In natural language, a *time instant* could be expressed in different formats. For example, "26 April 2009" might appear as "4/26/09" or "04-26-2009," To achieve normalization, a data property *hasNormalizedTime* is adopted to connect a *time instant* individual with a normalized value, which is in the form of "HH: mm: ss YYYY-MM-DD" following the xsd dateTime data type format.[33] For example, the <time instant> ("26 April 2009") *hasNormalizedTime* "2009-04-26." In a real-world setting, temporal information often appears in imprecise or uncertain forms, such as "early August" or "approximately 9 AM." Data property *hasApproximation* is used to describe this feature, for which "True" is the value for an "uncertain" instance, while "False" is the value (default) for a "certain" instance, for example, "at 8:00 AM on Jan 1st, 2017."

A *time instant* could be represented at different levels of temporal granularity (such as minute, hour, or day). To describe this feature, a parallel class, *granularity*, is defined to demonstrate the finest level of granularity. Common temporal granularities, such as <second>, <minute>, <day>, are introduced as its individuals. An instance of *time instant* is linked with an instance of *granularity* via the object property *hasGranularity*, for example, <time instant> ("8:30 AM") *hasGranularity* <granularity> ("Minute"), <time instant> ("Oct 10th, 2017") *hasGranularity* <granularity> ("Day").

To represent subcomponents of *time instant*, such as "morning" and "middle of the week," a set of subclasses are designed that include *instant of the day*, *instant of the week*. These subclasses are further divided into more specific subordinate classes. For instance, "*instant of the day*" has subclasses of *AM*, *PM*, *morning*, etc. In addition, another subclass, *time of the age*, is designed to record temporal information indicated by the age in years, and a data property *hasAgeValue* describes its data value. For example, in "At the age of 55, the patient was diagnosed with multiple sclerosis," <time of the age> ("At the age of 55") *hasAgeValue* "55."

Time interval, duration and periodic time interval. Class *time interval* is the region on the timeline that spans more than 1 single segment. It usually lasts for an amount of time (time length), for example, "albuterol sulfate aerosol inhalation for 30 minutes (9:30 AM–10:00 AM)." Class *duration* describes the time length of a *time interval* by means of the object property *hasDuration*. For example, in "monitor patient's heart rate for 72 hours starting from 2014-06-01," the <time interval> ("72 hours starting from 2014-06-01") *hasDuration* <duration> ("72 hours"). *Time interval* could also have a starting and ending point, which is described by the object properties *hasStartTime* and *hasEndTime*, respectively. Each of them links to an instance of *time instant* as in "the cardiac surgery lasted from 9:00 AM to 1:00 PM", the <time interval> ("from 9:00 AM to 1:00 PM") *hasStartTime* <time instant> ("9:00 AM").

Data property *hasNormalizedDuration* describes the normalized value of a *duration*, which is in the format of "0Y0M0W0D0H0m0s." For example, if the *duration* is "72 hours", the value of *hasNormalizedDuration* is "72H." Floating numbers are allowed to support precise expression. For example, if a duration lasts for 3-1/2 months, the value of *hasNormalizedDuration* is "3.5M." Similarly, if a duration is not certain, the value of data property *hasApproximation* would be "True."

It is very common that clinical events recur periodically or regularly, such as chemotherapy, blood glucose monitoring, and rehabilitation training. Modified from the HL7 time specification,[34] a subclass of *time interval*, *periodic time interval* is designed to represent each occurrence of a repeating interval. Further, 3 object properties are defined: *hasRepeatUnit* to describe the unit that occurred periodically, *hasPeriod* to describe the interval between the start time of 2 units, and *hasRepeatUnitInterval* to describe the interval between 2 repeat units of a *periodic time interval*. In addition, data property *hasRepeatTimes* describes the number of units that recurred in a *periodic time interval*, which is an integer. For example, "The patient received 30 minutes of aerobic exercise every day for 15 days from Feb 3, 2005" could be represented in the following resource description framework (RDF) triples:

```
 1  <event1>    rdf: type Event;
 2      rdfs: label "The patient received aerobic exercise";
 3      hasValidTime <tPeriodicInterval1>;
 4  <duration1>    rdf: type Duration;
 5      rdfs: label "30 minutes";
 6      hasNormalizedDuration "30m";
 7  <duration2>    rdf: type Duration;
 8      rdfs: label "every day";
 9      hasNormalizedDuration "1D";
10  <duration3>    rdf: type Duration;
11      rdfs: label "15 days";
12      hasNormalizedDuration "15D";
13  <tInstant1>    rdf: type Time Instant;
14      rdfs: label "Feb 3, 2005";
15      hasNormalizedTime "2005-02-03";
16      hasGranularity <Day>;
17  <tPeriodicInterval1> rdf: type Periodic Time Interval;
18      rdfs: label "30 minutes of aerobic exercise every day for 15
        days from Feb 3, 2005";
19      hasRepeatUnit <duration1>;
20      hasPeriod <duration2>;
21      hasDuration<duration3>;
22      hasStartTime <tInstant1>;
```

Property design for modeling relations

Properties are deliberately designed for TEO, including object properties (eg, *hasValidTime*, *hasTemporalRelation*, *hasGranularity*), data properties (eg, *hasApproximation*, *hasNormalizedTime*) and annotation properties (eg, skos: example). Some of them have been mentioned in the "Class design" section. Among them, the object property *hasTemporalRelation* serves as the parent property to model temporal relations of clinical events. Temporal relations include 2 main types: qualitative (eg, angina before headache) and quantitative (eg, angina 2h before headache). We proposed (1) two sets of temporal relations and (2) temporal relation statement to represent these 2 types of temporal relations, respectively.

Two sets of temporal relations

We extended Allen's interval algebra, which is originally designed for temporal relations between intervals, to cover temporal relations

between time and time, event and event, and time and event. An explicit temporal order relation relies on comparison of 4 time point pairs, namely the start and end time from 2 *time intervals/clinical events*: (T_{s1}, T_{s2}), (T_{s1}, T_{e2}), (T_{e1}, T_{s2}), (T_{e1}, T_{e2}), in which T_{s1}, T_{e1}, T_{s2}, T_{e2} represent the start time and end time of <Interval1> and <Interval2> or <Event1> and <Event2>, respectively. Allen's interval algebra defines 13 types of relations, that is, 6 pairs of invertible relations (*before/after*, *meets/metBy*, *overlaps/overlappedBy*, *starts/startedBy*, *finishes/finishedBy*, *during/contains*) and 1 symmetric relation (*equal*), and transitive axioms are used to hold between these relations. It applies to temporal reasoning when both start and end time of 2 *Intervals/Events* are known.

However, there are many occasions in which only 1 time point of each event or interval is known—then the strict Allen's algebra is not applicable. To address this issue, we designed BTRs, with which we only need to compare 1 pair of time points from each *Interval/Event*. There are 12 types of relations (4 pairs of time points and 3 possible temporal orders) in total, indicating the order of start or end time between 2 *time intervals/events*, including *startBeforeStart*, *startEqualEnd*, and *endBeforeStart*. For instance, *startBeforeStart* means that the start time of the first *interval/event* is before that of the second *interval/event*.

Figure 3 shows a graphical depiction of relations from Allen's interval algebra and BTRs. The upper part (Row 1) shows 13 relations from Allen's interval algebra, each relation derives from the comparison of 4 pairs of the starting/ending time points from interval a and b. The lower part (Row 2– Row 5) depicts the basic time relations, and each compares 1 pair of the start/end time point from a and b. Four basic relations together correspond to 1 Allen's interval relation; for example, the intersection region of 4 basic relations (startBeforeStart, endBeforeEnd, startBeforeEnd, and endBeforeStart) corresponds to "a before b" in Allen's. They could complement each other in the real clinical context for temporal reasoning.

Temporal relation statement

Although the Allen's and BTRs could represent the qualitative temporal order relations between *time/events*, it is out of their scope to describe the quantitative information of the time order. For example, in "Patient's bilirubin is elevated 2 weeks after the second cycle

of chemotherapy," the temporal relation between "Patient's bilirubin is elevated" and "the second cycle of chemotherapy" is "after" by using Allen's, but "2 weeks" could not be expressed. To address this issue, *TemporalRelationStatement* is designed to represent RDF triples, and the object property *hasTimeOffset* specifies the duration between events. Currently, *hasTimeOffset* is limited to the duration from the triple of "<event1> after/before <event2>". The example sentence could be represented using *TemporalRelationStatement* and *hasTimeOffset* in the following RDF triples:

1 <event1> rdf: type **Event**;
2 rdfs: label "Patient's bilirubin is elevated";
3 *after* <event2>;
4 <event2> rdf: type **Event**;
5 rdfs: label "the second cycle of chemotherapy";
6 <duration1> rdf: type **Duration**;
7 rdfs: label "2 weeks";
8 *hasNormalizedDuration* "2W";
9 <state1> rdf: type **TemporalRelationStatement**;
10 rdf: object <event1>;
11 rdf: predicate "after";
12 rdf: subject <event2>;
13 *hasTimeOffset* <duration1>;

Evaluation method

To evaluate the coverage of TEO, we annotated temporal information in clinical narratives from the Mayo Clinic using TEO.

Corpus and annotation process

A total of 6892 sentences that contained at least 1 TIMEX3 (ie, a phrase that contains time information)[35,36] were extracted from a 16-patient corpus (1996–2015) with the approval of the Institutional Review Board (IRB) of the Mayo Clinic. After removing incomplete or semi-structured sentences, and deidentifying all protected health information using the MITRE Identification Scrubber Toolkit (MIST),[37] 200 time-related sentences were selected for manual annotation. Two annotators (Annotator 1 [HS], Annotator 2 [JD]) annotated them for both classes and object properties using
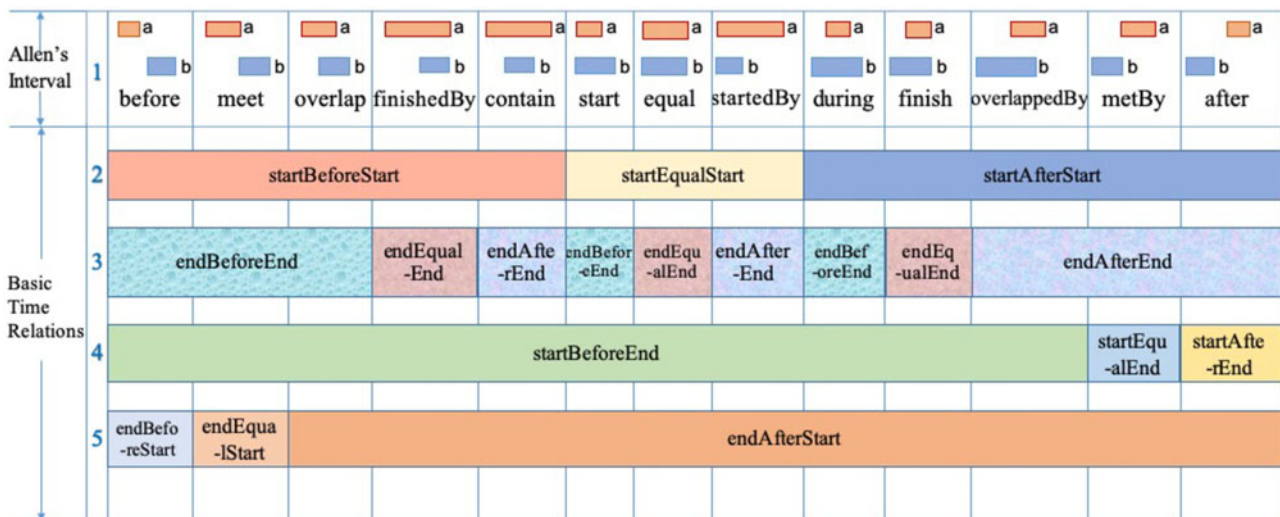

Figure 3. Allen's Interval Algebra and Basic Time Relations (BTRs).

the brat annotation tool.[38] They first annotated the classes in the sentences independently, then discussed and compared the annotation results of each sentence to remove annotation errors. An updated version of class annotation was created after the discussion. The annotation process of object properties was the same as that of the class.

Evaluation metrics

To facilitate the feasibility and simplicity of quantitative analysis, we mainly evaluated the intra-sentence temporal expressions and relations. The IAA evaluation was divided into the evaluation of class and object property. Annotation results from Annotator 1 were used as the gold standard. By comparing results from Annotator 2 with the gold standard, the precision, recall, and F1 measure were calcu-

lated. Precision is the fraction of annotations made by Annotator 2 that are true positive (TP). Recall is the fraction of annotations made by Annotator 1 that are TP. F1 is the harmonic mean of precision and recall. The calculations of agreement measures are listed below:

$$Precision = TP/(TP+FP)$$
$$Recall = TP/(TP+FN)$$
$$F1 = 2 \times precision \times recall/(precision + recall) = 2TP/(2TP+FP+FN)$$

RESULTS

Ontology metrics

The current TEO has 117 classes, 35 object properties, and 16 data properties, with 271 logical axioms and 188 declaration axioms. The hierarchical structure of classes is shown in Figure 4. The primary object properties and data properties, including their description information, are presented in Tables 1 and 2, respectively.

To promote community-driven feedback and adoption, TEO is published in https://sbmi.uth.edu/bsdi/TEO_1.0.0.owl.

Evaluation results

With respect to the IAA, by randomly selecting the result of Annotator 1 as the gold standard, F1 measures of time-related classes annotation were 77.05% and 81.22% (exact mapping and partial mapping), and that of object properties annotation was 94.62% after considering cases of semantic equivalence among the annotation results (eg, <event1> before <event2> was regarded as equal to <event2> after <event1>).

Concerning the coverage ability, TEO was found to faithfully represent 95.43% (940 instances) of 985 instances of temporal classes and 97.02% (684 relations) of 705 temporal relations. After completing coverage analysis, the 2 annotators discussed and reconciled annotation discrepancies. The finalized annotation of the selected 200 sentenced consisted of 1171 instances of temporal classes and 520 object properties. In addition, there were 162 temporal relations annotated. Among them, 16.05% (26 relations) were represented via BTRs, as opposed to Allen's algebra interval, indicating the value of the BTRs.
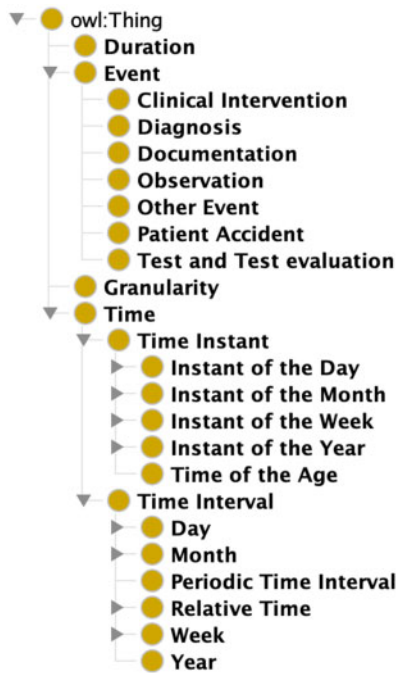

Figure 4. Hierarchical Structure of TEO Classes.

Table 1. Primary object properties of TEO

Object Property	Definition	Domain	Range
hasValidTime	Links an *event* with its specific timestamps	*Event*	*Time*
hasTemporalRelation	The superset of the temporal relations defined in the ontology	*Event* or *Time*	*Event* or *Time*
hasGranularity	Describes the granularity of a temporal element or temporal relation	*Time* or a triple	*Granularity*
hasDuration	Describes the *duration* of a *time interval* or *periodic time interval*	*Time Interval* or *Periodic Time Interval*	*Duration*
hasStartTime	Describes the start time of a *time interval* or *periodic time interval*	*Time Interval* or *Periodic Time Interval*	*Time Instant*
hasEndTime	Describes the end time of a *time interval* or *periodic time interval*	*Time Interval* or *Periodic Time Interval*	*Time Instant*
hasRepeatUnit	Describes the time unit that occurs periodically in a *periodic time interval*	*Periodic Time Interval*	*Duration* or *Periodic Time Interval*
hasRepeatUnitInterval	Describes the interval between 2 repeat units of a *periodic time interval*	*Periodic Time Interval*	*Duration*
hasPeriod	Describes the interval between the start time of 2 units in a *periodic time interval*	*Periodic Time Interval*	*Duration*

Table 2. Primary Data Properties of TEO

Data Property	Definition	Value/Format
hasApproximation	Describes the approximative or uncertain feature of a temporal expression	"True" or "False"
hasDescription	Adds a detailed description to the event, such as test results, report conclusions, settings of an experiment	String
hasNormalizedTime	Represents the normalized view of the time expression of the given instant	HH: mm: ss YYYY-MM-DD
hasNormalizedDuration	Captures the structured form of the duration	0Y0M0W0D0H0m0s
hasRepeatTimes	Describes the number of units that reoccur in a *periodic time interval*	Integer
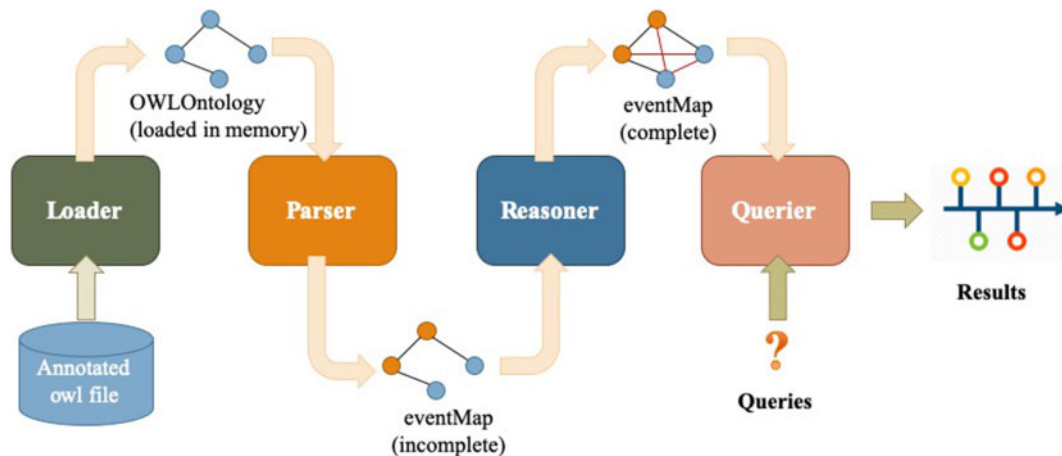
Figure 5. TEO Reasoner.

Box 1. An adapted drug adverse event report

"A 35-year-old man was *admitted to hospital with periorbital swelling, redness, and pain* on May 24, 2014. Then he was *diagnosed with periorbital cellulitis*. He *was treated with intravenous (IV) clindamycin*, and with *IV ciprofloxacin*, which *reduced the orbital redness and swelling*. However, on the second day following antibiotic treatment, he *developed nausea and right upper quadrant (RUQ) abdominal pain*, his *liver function tests (LFTs) began to increase*. A *diagnosis of idiosyncratic drug-induced liver injury (DILI) was made.*"

TEMPORAL INFORMATION REASONING

To demonstrate the temporal reasoning capability of TEO in a clinical setting, we developed a TEO reasoner and leveraged a drug adverse event case report as a use case. Prior work, such as the temporal reasoners in TempEval shared tasks,[39,40] allows users to query the temporal relations of events and to generate a timeline. TEO reasoner, however, offers its unique contribution to query the uncertain relationship between events with insufficient information. It is achieved by the support of the basic relations defined in TEO. The core part of TEO reasoner was based on a transition matrix that defines the property chains of all the basic relations (eg, *startBeforeStart*, *startEqualStart*).

More specifically, TEO reasoner is built upon OWL application programming interface[41] and HermiT OWL reasoner[42] using Java. It consists of 4 primary blocks (Figure 5): 1) the Loader, loads the OWL file which has been annotated for *events* and *time* with respect to TEO into memory; 2) the Parser, extracts the stated *events* and corresponding temporal information from the memory and builds an eventMap that links explicitly stated temporal information with events; 3) the Reasoner, infers indirect temporal relations among all *events* using explicitly stated temporal information in conjunction with a pre-defined transitive matrix and adds the inferred relations to the eventMap; 4) the Querier, provides multiple application programming interfaces to query temporal information for a specific event and the temporal relations as well as the timeline among events.

We adapted a drug adverse event report[43] to query complex temporal information, including timestamps, temporal relations, and timeline among events. The report was initially manually annotated with TEO using Semantator[44] and then loaded into the TEO Reasoner for inference. The report is shown in Box 1, in which the words in red italic are manually annotated as events.

We designed 4 types of queries that are frequently asked questions and can provide insights to clinical decision-making. The query types, clinical questions, queries using Querier application programming interface, and results are presented in Table 3.

DISCUSSION

Contributions

TEO integrated functionality from existing schemes and ontologies and proposed new patterns tailored to clinical narratives. Its expressiveness could be summarized as follows: (1) for entity repre-

Table 3. Use case of timeline reasoning: queries and results

Query Type	Clinical Questions and Queries	Results
Type 1	• To query temporal information of a specific event • (1) **Clinical question:** When was the patient admitted to the hospital? • (2) **Query:** *querier.getEventByName("admitted to hospital").validTime()*	[2014-05-24]
Type 2	• To query temporal relationships which are not directly available from the text but can be inferred through multiple relation triples • **Clinical question:** What is the temporal relation between "*admitted to hospital*" and "*liver function tests (LFTs) began to increase*"? • **Query:** *querier.getPossibleAllenRelationsBetweenTwoEvents ("admitted to hospital", "liver function tests (LFTs) began to increase")*	[BEFORE]
Type 3	• To query an uncertain relationship between events with insufficient information • **Clinical question:** Does "*ciprofloxacin*" treatment start before "*diagnosis of idiosyncratic drug-induced liver injury (DILI)*"? • **Analysis:** as the end time of the treatment "*ciprofloxacin*" is not contained in the text, the exact Allen's relations between these 2 events cannot be inferred. However, BTRs provide useful information. • **Query:** *querier.getBasicRelationsBetweenTwoEvents("ciprofloxacin", "diagnosis of idiosyncratic drug-induced liver injury (DILI)")*	[[STARTBEFORESTART]]
Type 4	• To query the timeline for all of the events • **Clinical question:** What events happened before "*diagnosis of idiosyncratic drug-induced liver injury (DILI)*"? • **Analysis:** by analyzing the events on the timeline prior to the specific event, we might be able to find events that may have a causal relationship with that specific event. The query can retrieve 2 timelines (the temporal relation between "clindamycin" and "ciprofloxacin" is unclear from the text). By examining the timelines, some events can be included for further analysis for the association with *DILI*, including "diagnosed with periorbital cellulitis" (disease progression), "clindamycin" and "ciprofloxacin" (drug-induced adverse events). • **Query:** *querier.getAllTimelines()*	["periorbital swelling, redness, and pain", "admitted to hospital", "diagnosed with periorbital cellulitis", "treated with intravenous (IV) clindamycin and with IV ciprofloxacin", "developed nausea and right upper quadrant (RUQ) abdominal pain", "liver function tests (LFTs) began to increase", "diagnosis of idiosyncratic drug-induced liver injury (DILI)"]. The whole timeline is illustrated in Figure 6.
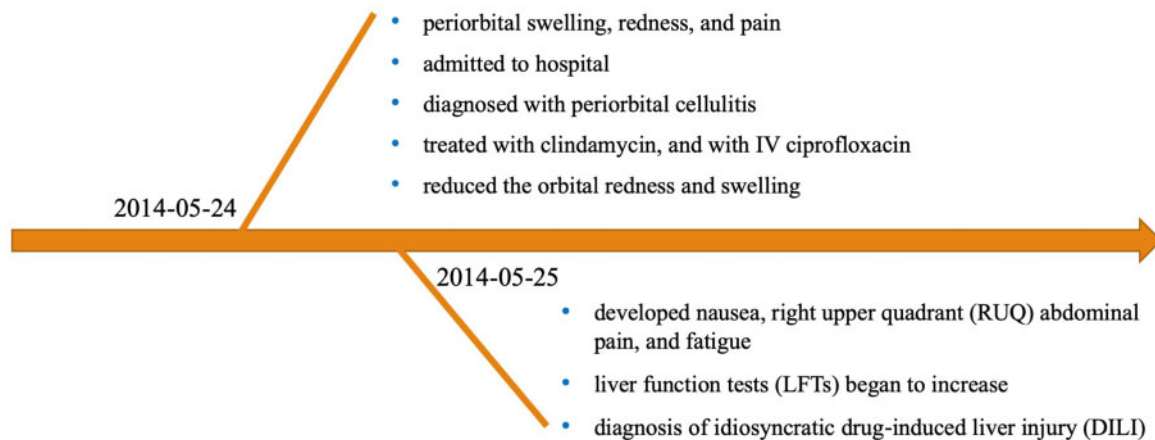
Figure 6. Timeline Reasoning Results.

sentation: a) it can specify periodic and recurring events (eg, "Ampicillin 250 mg q.i.d. for 5 days"), by leveraging class *periodic time interval* and related properties. This is 1 unique function of TEO; b) it can represent approximative and uncertain time (eg, "earlier in the week", "almost most of the day"), by using data property *hasApproximation* to encode time fuzziness; c) it can realize time normalization to reconcile semantic heterogeneity by using data properties *hasNormalizedTime* and *hasNormalizedDuration*; (2) for relation representation: a) it can describe both qualitative temporal order, by using Allen's interval algebra, and quantitative

feature of temporal order, by using *Temporal Relation Statement*; b) it can infer all possible temporal orders by using BTRs, when no sufficient temporal information is available; c) it can realize complex temporality reasoning over clinical events by applying these 2 suites of temporal relations, without rigid demands for information completeness. Moreover, as an ontology encoded in OWL, TEO has better expressiveness in supporting computational reasoning than the XML-based annotation schemes. Table 4 shows the feature comparison between TEO and important time-related schemas and ontologies.

Table 4. Comparison between TEO and other important time-related schemas and ontologies

Name	Type	Goal	Core Components	Major Time Types/ Classes	Temporal Relations	Strengths
ISO-TimeML [9,13,45]	Annotation scheme	To annotate event and temporal expression in natural language text	TIME3, Event, Signal, Link	4 types (Time, Date, Duration, Set)	Allen's interval	(1) Introduces temporal functions to allow intentionally specified expressions; (2) Identifies signals and language features determining the interpretation of temporal and event expressions
SCATE scheme [16]	Annotation scheme	To normalize temporal expressions	Time, Temporal operator	4 types (Timeline, Period, Interval, Repeating interval)	~	Uses mathematical operators and compositional mechanism to flexibly normalize a variety of temporal expressions
Time Ontology in OWL [22]	Ontology	To describe temporal properties of resources in the world or described in Web pages	Event, Time	18 temporal classes	Allen's interval	Supports representation of temporal ordering relationships
TEO	Ontology	To provide a formal conceptualization of temporal information in clinical data	Event, Time	109 temporal classes	(1) Allen's interval (2) Basic time relations	(1) Uses well-defined classes and properties to realize flexible and comprehensive representation of clinical temporal expressions; (2) Supports representation of quantitative and qualitative temporal ordering relationships

Abbreviations: OWL: web ontology language; SCATE, semantically compositional annotation scheme for time expressions; TEO: time event ontology.

Limitations and future efforts

Due to the diversity of temporal expressions, and the complex interplay of explicit and implicit inference required to understand temporal information, current TEO faces some limitations: (1) we have not defined the pattern that could represent an *event* individual with changing statuses, (eg, "The patient received rehabilitation training twice a day last year, and once every 2 days this year"). To assure reasoning definiteness, 1 *event* individual is allowed to link only 1 *time* individual via object property *hasValidTime*. Future efforts will be made to support 1 *event* individual connecting with multiple *time* individuals. (2) depicting negations of temporal relation (eg, "no later than," "not during") is currently out of the scope because the monotonicity assumption of OWL determines that negation as failure is not supported.[26] We will introduce new object properties (such as *not before*, *not after*) to improve negation expressiveness. In addition, TEO mainly relies on manual annotation as the first step, which is labor-intensive and time-consuming. It would be desirable to leverage NLP techniques to extract temporal information and assist automated TEO annotation. In all, we will increase the representation flexibility and machine amenability of TEO.

CONCLUSION

In this paper, we present a robust time ontology called TEO. Using CNTRO 1.0 and 2.0 as the starting point and referencing and reusing existing schemes and ontologies, the newly designed TEO has rich expressiveness of temporal entities and relations. With 2 sets of temporal relations (Allen's interval algebra and BTRs) and *Temporal Relation Statement*, it can specify both qualitative and quantitative temporal order relations. TEO can reason about complex time sequences of clinical events, which would disclose the embedded temporal information and facilitate full use of clinical narratives. In the future, we will combine TEO with NLP techniques (eg, encoding heuristic rules into NLP models) to improve the performance of temporal information annotation, extraction, and reasoning ultimately to empower clinical decision support with a precise timeline.

AUTHOR CONTRIBUTIONS

CT supervised the study; CT, HX, YH, HL, JD, FL, GR, and YX conceived and designed the research; CT, YH, JD, FL, YL, and HS participated in ontology construction; JD, HS, LW, SL, HC, FL, and MM performed the evaluations and analyzed the data; FL, MM, JD, HS, and CT drafted the original manuscript; HX, YH, HL, and YX contributed to manuscript revisions; all authors reviewed and approved the final manuscript.

REFERENCES

1. Combi C, Shahar Y. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. *Comput Biol Med* 1997; 27 (5): 353–68.
2. Zhou L, Friedman C, Parsons S, Hripcsak G. System architecture for temporal information extraction, representation and reasoning in clinical narrative reports. *AMIA Annu Symp Proc* 2005; 2005: 869–73.
3. Sun W, Rumshisky A, Uzuner O. Annotating temporal information in clinical narratives. *J Biomed Inform* 2013; 46: S5–12.
4. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Am Med Inform Assoc* 2013; 20 (5): 806–13.
5. Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *J Am Med Inform Assoc* 2015; 22 (4): 872–80.
6. Lin C, Karlson EW, Dligach D, *et al*. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. *J Am Med Inform Assoc* 2015; 22 (e1): e151–61.
7. Song X, Waitman LR, Yu A, Robbins DC, Hu Y, Liu M. Longitudinal risk prediction of chronic kidney disease in diabetic patients using temporal-enhanced gradient boosting machine: retrospective cohort study. *JMIR Med Inform* 2020; 8 (1): e15510.
8. Sun W, Rumshisky A, Uzuner O. Temporal reasoning over clinical text: the state of the art. *J Am Med Inform Assoc* 2013; 20 (5): 814–9.
9. Pustejovsky J, Castano JM, Ingria R, *et al*. TimeML: Robust specification of event and temporal expressions in text. *New Directions in Question Answering* 2003, AAAI Spring Symposium; March 24–26, 2003: 28–34; Stanford, CA.
10. Bethard S, Derczynski L, Savova G, Pustejovsky J, Verhagen M. Semeval-2015 task 6: Clinical tempeval. In: proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); June 4–5, 2015: 806–14; Denver, CO.
11. Bethard S, Savova G, Chen W-T, Derczynski L, Pustejovsky J, Verhagen M. Semeval-2016 task 12: Clinical tempeval. In: proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016); June 16–17, 2016: 1052–62; San Diego, CA.
12. Bethard S, Savova G, Palmer M, Pustejovsky J. SemEval-2017 Task 12: Clinical TempEval. In: Proceedings of the 11th International Workshop on Semantic Evaluation; August 3–4, 2017: 565–72; Vancouver, Canada.
13. Pustejovsky J. ISO-TimeML and the annotation of temporal information. *Handbook of Linguistic Annotation*. Berlin: Springer; 2017:941–68.
14. Laparra E, Xu D, Bethard S. From characters to time intervals: new paradigms for evaluation and neural parsing of time normalizations. *Trans Assoc Comput Linguist* 2018; 6: 343–56.
15. Bethard S, Parker J. A semantically compositional annotation scheme for time normalization. In: proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); May 23–28, 2016: 3779–86; Portorož, Slovenia.
16. Laparra E, Xu D, Elsayed A, Bethard S, Palmer M. SemEval 2018 Task 6: parsing time normalizations. In: proceedings of the 12th International Workshop on Semantic Evaluation; June 5–6, 2018: 88–96; New Orleans, LA.
17. Ning Q, Wu H, Roth D. A multi-axis annotation scheme for event temporal relations. arXiv preprint arXiv: 1804.07828 2018.
18. W3C. OWL Web Ontology Language Overview. Secondary OWL Web Ontology Language Overview; 2004. https://www.w3.org/TR/owl-features/ Accessed 28 March, 2017
19. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. *N Engl J Med* 2018; 379 (15): 1452–62.
20. Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis* 1993; 5 (2): 199–220.

21. Ermolayev V, Batsakis S, Keberle N, Tatarintseva O, Antoniou G. Ontologies of time: review and trends. *Int J Comput Sci Appl* 2014; 11 (3): 57–115.

22. W3C. Time ontology in OWL. Secondary time ontology in OWL 2017. https://www.w3.org/TR/owl-time/ Accessed 11 January, 2018

23. O'Connor MJ, Das AK. A method for representing and querying temporal information in OWL. In: proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies; January 20–23, 2010: 97–110; Valencia, Spain.

24. Hobbs JR, Pan F. An ontology of time for the semantic web. *ACM Trans Asian Lang Inf Process (TALIP)* 2004; 3 (1): 66–85.

25. Zhou Q, Fikes R. A reusable time ontology. In: proceedings of the AAAI Workshop on Ontologies for the Semantic Web; 2002: 1–6. https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-015.pdf

26. Tao C, Wei W-Q, Solbrig HR, Savova G, Chute CG. CNTRO: a semantic web ontology for temporal relation inferencing in clinical narratives. *AMIA Annu Symp Proc* 2010; 2010: 787–91.

27. Tao C, He Y, Yang H, Poland GA, Chute CG. Ontology-based time information representation of vaccine adverse events in VAERS for temporal analysis. *J Biomed Sem* 2012; 3 (1): 13.

28. Chen HW, Du J, Song H-Y, Liu X, Jiang G, Tao C. Representation of time-relevant common data elements in the cancer data standards repository: statistical evaluation of an ontological approach. *JMIR Med Inform* 2018; 6 (1): e7.

29. Tao C, Solbrig HR, Chute CG. CNTRO 2.0: a harmonized semantic web ontology for temporal relation inferencing in clinical narratives. *AMIA Summits Transl Sci Proc* 2011; 2011: 64–8.

30. Allen JF, Ferguson G. Actions and events in interval temporal logic. *J Logic Comput* 1994; 4 (5): 531–79.

31. Grüninger M, Li Z. The time ontology of allen's interval algebra. In: proceedings of the 24th International Symposium on Temporal Representation and Reasoning (TIME 2017); October 16–18, 2017: 1–16; Mons, Belgium.

32. Drummond N, Jupp S, Moulton G, Stevens R. A practical guide to building OWL ontologies using protege 4 and CO-ODE tools edition 1.2. Secondary a practical guide to building OWL Ontologies using protege 4 and CO-ODE tools edition 1.2 2018-03-06 2009. http://phd.jabenitez.com/wp-content/uploads/2014/03/A-Practical-Guide-To-Building-OWL-Ontologies-Using-Protege-4.pdf Accessed 5 February, 2018

33. W3C. XML Schema Part 2: Datatypes Second Edition. Secondary XML Schema Part 2: Datatypes Second Edition 2004. https://www.w3.org/TR/xmlschema-2/ Accessed 25 January, 2017

34. HL7. HL7 Time Specification. Secondary HL7 Time Specification. https://wiki.hl7.org/index.php? title=Datatypes_R2_Issue_6. Accessed 1 August, 2017

35. Group TW. Guidelines for temporal expression annotation for english for tempeval 2010; 2009. https://www.aclweb.org/anthology/S10-1010/ Accessed March 25, 2017

36. Sohn S, Wagholikar KB, Li D, *et al*. Comprehensive temporal information detection from clinical text: medical events, time, and TLINK identification. *J Am Med Inform Assoc* 2013; 20 (5): 836–42.

37. MIST: The MITRE Identification Scrubber Toolkit. Secondary MIST: The MITRE Identification Scrubber Toolkit. http://mist-deid.sourceforge.net/docs_1_3/html/index.html Accessed 3 January, 2016

38. Brat rapid annotation tool. Secondary Brat rapid annotation tool. https://brat.nlplab.org/ Accessed 5 January, 2016

39. Minard A-L, Speranza M, Agirre E, *et al*. Semeval-2015 task 4: Timeline: Cross-document event ordering. In: proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); June 4–5, 2015: 778–86; Denver, CO.

40. Llorens H, Chambers N, UzZaman N, Mostafazadeh N, Allen J, Pustejovsky J. Semeval-2015 task 5: QA tempeval-evaluating temporal information understanding with question answering. In: proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); June 4–5, 2015: 792–80; Denver, CO.

41. OWLAPI in GitHub. Secondary OWLAPI in GitHub. https://github.com/owlcs/owlapi Accessed 9 August, 2016

42. Group IS. HermiT OWL Reasoner. Secondary HermiT OWL Reasoner. http://www.hermit-reasoner.com/ Accessed 5 January 2017.

43. Radovanovic M, Dushenkovska T, Cvorovic I, *et al*. Idiosyncratic drug-induced liver injury due to ciprofloxacin: a report of two cases and review of the literature. *Am J Case Rep* 2018; 19: 1152–61.

44. Ontology Research Group of SBMI U. Semantator. Secondary Semantator. https://sbmi.uth.edu/ontology/project/semantator.htm Accessed 6 January, 2017

45. Pustejovsky J, Lee K, Bunt H, Romary L. ISO-TimeML: An International Standard for Semantic Annotation. LREC; 2010: 394–97.