A Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives

Cui Tao, PhD  Wei-Qi Wei, MM  Guergana Savova, PhD  Christopher G. Chute, MD, DrPH
Mayo Clinic College of Medicine, Rochester, MN

Abstract

Using Semantic-Web specifications to represent temporal information in clinical narratives is an important step for temporal reasoning and answering time-oriented queries. Existing temporal models are either not compatible with the powerful reasoning tools developed for the Semantic Web, or designed only for structured clinical data and therefore are not ready to be applied on natural-language-based clinical narrative reports directly. We have developed a Semantic-Web ontology which is called Clinical Narrative Temporal Relation ontology. Using this ontology, temporal information in clinical narratives can be represented as RDF (Resource Description Framework) triples. More temporal information and relations can then be inferred by Semantic-Web based reasoning tools. Experimental results show that this ontology can represent temporal information in real clinical narratives successfully.

1 Introduction

Time is essential in clinical research. The temporal dimension in medical data analysis allows clinical activities such as (1) uncovering temporal patterns at the disease and patient level and better understanding of disease progression, (2) explaining past events such as the possible causes of a clinical situation, and (3) predicting future events such as possible complexities based on a patient’s current status.

Managing time-stamped data and explicitly representing temporal relationships is an important step toward querying and inferring useful temporal assertions. In this research, we introduce an ontology in the Web Ontology Language (OWL) [1] format for modeling temporal information in clinical narratives. Using OWL to represent temporal assertions brings us many benefits. First, the Semantic Web and the Web Ontology Language provide a standard mechanism with explicit and formal semantic knowledge representation. Secondly, the Semantic Web offers powerful reasoning capabilities. OWL is built on formalisms that adhere to Description Logic (DL) forms and therefore allows reasoning and inference. In addition, the Semantic Web Rule Language (SWRL) [2] can be used to add rules to OWL and enable Horn-like rules that can be used to infer new knowledge from an OWL ontology and reason about OWL individuals. Thirdly, once we have an ontology that can represent temporal assertions in the clinical domain precisely, we can annotate temporal expressions and relations with respect to the ontology and store the instances as RDF triples [3]. The information then become “machine-understandable”. Tools and services such as reasoners, editors, querying systems, and storage mechanisms that have been developed by the Semantic Web community can be directly applied to the temporal data.

Many previous efforts have been made for modeling temporal information. Most of these research efforts have focused on temporal information stored in structured databases [4]. There are two existing temporal ontologies in OWL, the Time ontology in OWL [5] and the SWRL Temporal ontology [6]. The first one is a general time ontology whereas the second one has been applied more by clinical research [7]. Both of the ontologies are designed only for structured data stored in databases and therefore only focus on timing events with points anchored in absolute time.

In addition to structured data, however, many important temporal assertions are interweaved in free-text based reports such as clinical notes. Models such as Temporal Constraint Structure (TCS) [8] and the TimeML model [9] target on modeling temporal information represented in natural language. These models, however, are not compatible to OWL and the semantic-web based tools especially the reasoners to infer new temporal knowledge.

2 Clinical Note Temporal Relation Ontology

In this paper, we introduce the CN TemporalRelation (Clinical Narrative Temporal Relation) Ontology1, an OWL ontology that can model temporal information found not only in structured databases, but also in natural-language based clinical reports. We investigated existing conceptual models for temporal information such as Time ontology in OWL [5], the SWRL Temporal ontology [6], Allen’s temporal relations [10], Temporal Constraint Structure (TCS) [8], the TimeML model [9], as well as the HL7 time specification [11]. We also evaluated actual clinical notes and summarized the temporal-relation notations that are commonly used in these clinical notes. The CN TemporalRelation ontology was developed based on these previous experiences combined with new ontological specifications that fit the needs of natural-language based clinical reports.

1http://informatics.mayo.edu/LexGrid/downloads/CNTemporalRelation.owl
1. the second cycle of chemotherapy was on June 10, 2004
2. monitor patient’s heart rate for 72 hours starting from today (note date:2004-06-01)
3. take antibiotics every 8 hours for 10 days starting from today (note date:2004-06-01)
4. see the patient back in approximately two weeks prior to his third cycle of chemotherapy (note date:2004-06-10)
5. patient’s bilirubin is elevated 2 weeks after the second cycle of chemotherapy

Figure 1: Examples of Temporal Relations from Clinical Notes

The major OWL classes of the CNTemporalRelation ontology includes: Event, Time, Duration, Granularity, Precision, and TemporalRelationStatement.

We defined an Event class which describes any sort of occurrence, state, perception, procedure, symptom or situation that occurs on a time line in clinical narratives.

The Time class is the superclass of all the OWL temporal representation classes: TimeInstant, TimeInterval, TimePhase, and TimePeriod. An OWL TimeInstant is a specific point of time on the timeline. In clinical reports, a time instant can be represented in different granularities such as year, month, and day. In the ontology, we defined an OWL object property called hasGranularity to specify the granularity of each time instant. For example, the granularity value of the time instant “June 10, 2004” is day and the granularity value of the time instant “Dec. 2004” is month. The OWL class Granularity specifies the possible time units for different granularities. A time instant may also be represented in different formats. For example, 6/10/04, 06-10-2004, June, 10, 04, or Jun. 10, 2004 can all be used to represent a date 2004-06-10. We implemented a normalizer that converts commonly used time notations to the ISO 8601 [12] format. In the ontology, we defined two data properties hasOrigTime and hasNormalizedTime that keep track of the time instant in its original form and in the normalized form respectively.

An OWL TimeInterval represents a duration of time. It could have two relations (OWL object properties), hasStartTime and hasEndTime. Each of them links to instances of TimeInstant. A TimeInterval could also have a Duration. An instance of the Duration class represents the time length of a TimeInterval. We use an OWL data type property hasValue and an OWL object property hasUnit to describe a Duration. For example, in sentence 2 in Figure 1, the event is “monitor patient’s heart rate”, the duration is 72 hours (hasValue is “72” and hasUnit is “hour”), and the startTime is “today”.

Many clinical events recur periodically. Adopted and modified from the HL7 time specification [11], two OWL classes, TimePhase and TimePeriod, are defined in the CNTemporalRelation ontology to represent intervals of time that recur periodically. A TimePhase represents each occurrence of the repeating interval and a TimePeriod specifies a reciprocal measure of the frequency at which the TimePhase repeats. The class TimePhase is a subclass of TimeInterval, therefore, we can also specify a StartTime, an EndTime, and a Duration. In addition, a relation (OWL ObjectProperty), hasTimePeriod, is defined to specify the relationship between a TimePhase and a TimePeriod. For example, in sentence 3 in Figure 1, “every 8 hours for 10 days starting from today” is a TimePhase. Its StartTime is “today”. Its Duration is “10 days”. And its TimePeriod is “every 8 hours”.

We also define the certainty of a Time instance. For example, a physician can describe a time notation with ambiguities such as “early next week” and “in approximately two weeks”. In the CNTemporalRelation ontology, we defined a class called “Modality” which serves as a flag to indicate whether a time representation is approximated or not.

Each event can have a time stamp described by a Time instance. The OWL object property hasTimeStamp is defined to specify the time stamp of an event. In addition, the ontology also defines a set of temporal relations such as after, before, is_included, include, simultaneous, overlap, overlappedBy, during, and continuesThrough. These relations are defined as OWL object properties and can be used to describe temporal relations between two events, or an event and a Time instance. For example, in sentence 4 in Figure 1, “see the patient” is an event and “third cycle of chemotherapy” is another. And the temporal relation between these two events is before.

We also use TemporalRelationStatement class to describe temporal relations between two events or between an event and a Time instance. The TemporalRelationStatement class is a sub-class of rdf:Statement, we can define temporal subject, object, and predicate of a TemporalRelationStatement. Using TemporalRelationStatement to describe a temporal relation enables defining properties of the relation by reification. For example, we can add an offset time frame to the relation by using an OWL object property called hasTemporalOffset. This offset defines the relative timing of a pair of events. In order to model sentence 5 in Figure 1, for example, we can use a TemporalRelationStatement to represent “patient’s bilirubin is elevated” (object after predicate “the second cycle of chemotherapy” (subject), and then add “2 week” as an instance of TemporalOffset to this TemporalRelationStatement instance.

3 RDF Triple Representation for Temporal Information In Clinical Narratives

After the CNTemporalRelation ontology has been defined, we can use this ontology to model the temporal instances and temporal relations in clinical narratives. These instances can be stored as RDF triples [3] in either an RDF file or in an RDF triple store. An RDF triple contains a subject, a predicate, and an object. A predicate in a triple represents the relationship from the subject to the object. In this section, we use a few examples to illustrate how to represent temporal assertions with respect to the CNTemporalRelation Ontology using RDF triples. These examples are chosen from real clinical notes and represent the major temporal
an event with a time stamp that is a time phase. Since an event such as Row 2 in Table 1 shows. Row 3 indicates that event1 has a time stamp which is a time interval. Rows 5-7 record the start time of the interval and Rows 10-13 record the duration of time interval. The end time of the event is missing but can be inferred as Rows 14-16 show.

Table 2 shows the RDF triple representation of sentence 2 in Figure 1 and illustrates how to represent an event with an interval time stamp. Rows 1-2 define event2, which is a new instance of Event. Rows 3-4 indicate that event2 has a time stamp which is a time interval. Rows 5-7 record the start time of the interval and Rows 10-13 record the duration of time interval. The end time of the event is missing but can be inferred as Rows 14-16 show.

Table 3 shows the RDF triple representation of sentence 3 in Figure 1 and illustrates how to represent an event with an interval time stamp that is a time phase. Since TimePhase is a subclass of TimeInterval. The representations for start time (Rows 5-9), duration (Rows 10-13), and end time (Rows 14-16) are similar to a time interval. In addition, we defined expressions of natural clinical language.

Table 1 shows the RDF triple representation of sentence 1 in Figure 1 and illustrates how to represent an event with a time stamp that is a time instants. For each data instants we want to annotate, we assign it a unique URI and also indicate which class it belongs to. For example, Row 1 in Table 1 indicates that an instance with URI event1 belongs to Class Event. We use rdfs:label to indicate the description of the event such as Row 2 in Table 1 shows. Row 3-4 indicates that event1 has a time stamp timeInstant1, which belongs to the TimeInstant class and has original value “June 10, 2004” as Rows 4 and 5 show respectively. The triples in italic are inferred values and will be discussed in the next section.

Table 2 shows the RDF triple representation of sentence 2 in Figure 1 and illustrates how to represent an event with an interval time stamp. Rows 1-2 define event2, which is a new instance of Event. Rows 3-4 indicate that event2 has a time stamp which is a time interval. Rows 5-7 record the start time of the interval and Rows 10-13 record the duration of time interval. The end time of the event is missing but can be inferred as Rows 14-16 show.

Table 4 shows the RDF triple representation of sentence 4 in Figure 1.
a time period (Rows 17-19) to indicate how often the event repeats.

Table 4 shows the RDF triple representation of sentence 4 in Figure 1 and illustrates how to represent a temporal relation. We first defined two events (Rows 1-4). In sentence 5 there are actually two temporal relations. Row 5 indicates that event4 is before event5. And Rows 6-8 represent the time stamp ("in approximately two weeks") of event4.

Table 5 shows the RDF triple representation of sentence 5 in Figure 1 and illustrates how to represent a temporal relation using reification. Lines 1-4 define the two events2. In Row 5, we defined statement1, which is an instance of TemporalRelationStatement. Rows 6-8 define the temporal relation between the two events by defining the object, predicate, and subject of statement1. Row 9 defines that statement1 has a temporal offset. This offset defines the relative timing of the pair of events, i.e., how long after event1 happened, event6 happened. And rows 10-12 define the offset which is an instance of the Duration class.

4 Temporal Information Inference

Using the CNTemporalRelation ontology, we can model temporal information and relations that are literally indicated in clinical reports and represent them in RDF triples. With these triples, we can then infer more temporal information for the events. Since the focus of this paper is to introduce the Clinical Narrative Temporal Relation ontology, here we only briefly introduce how the inferencing is done.

Our first step is to normalize the temporal representations and assign granularities and certainties. We use the information extraction technology developed by the Brigham Young University (BYU) Data Extraction Group (DEG) [13] to recognize different time notations. Once the time notation has been recognized, our software normalizer tries to convert a time notation/expressions from one format to the normalized form (e.g., Rows 5-7 in Table 1) or a relative time notation to an absolute normalized form (e.g., Rows 7-9 in Table 2). It can also assign an appropriate granularity (e.g., Row 7 in Table 1) and certainty (e.g., Row 11 in Table 4) to a time instance.

With normalized temporal representations of the time stamps, we can apply temporal operations on them and infer new time-oriented knowledge. We adopted the temporal operation functions provided by Java and the SWRL Temporal Built-In Library [14]. Using these functions, we can compare two normalized time stamps and to determine their relationships. For example, Line 6 in Table 1 shows the normalized time stamp for event1 and Line 8 in Table 2 shows the normalized start time for event2. By comparing these two time instants, the system can infer that event1 happened after event2 started. For a time interval, we can also infer duration, start time, or end time, if two of the three components are given in the original source such as Rows 14-16 in Table 2 show. The “add” operation in the SWRL Temporal Built-In library also provides us an option to calculate the absolute time stamp of an event given its temporal distance to another time instant is given. For example, in Table 4, we know that event4 happened “in approximately two weeks” from today, information in Rows 9-11 in Table 4 can then be inferred.

We can also infer new temporal information based on the ontology definitions of the relations. For example, we defined that before is transitive. Therefore, given that event A is before event B, and event B is before event C, we can infer that event A is before event C. We defined that before and after are inverse properties of each other. Therefore, given that event A is before event B, we can infer that event B is after event A, and vice versa. We also defined that simultaneous is a symmetric property. Therefore, given that event A is simultaneous with event B, we can infer that event B is also simultaneous with event A.

5 Evaluation, Summary, and Discussion

The CNTemporalRelation ontology was evaluated on real clinical notes from Mayo Clinic3. We randomly selected five clinical notes for different patients created by different physicians. From these notes, we extracted 153 sentences that contained temporal information.

We first compared the expressiveness capabilities of the CNTemporalRelation ontology with the two existing temporal ontologies in OWL: the Time ontology [5] and the SWRL Temporal ontology [6]. Since these two ontologies are designed only for structured data in databases, they mainly focus on timing events with points anchored in absolute time. In the 153 sentences we extracted, however, only 64 of them fall in this category. To cover the temporal assertions in natural-language based clinical narratives, we have added the following major expressiveness capabilities to the CNTemporalRelation ontology. (1) Periodic Time Interval. In clinical narratives, there are many events that recur periodically. It is important to be able to represent periodic time intervals. Two OWL classes, TimePeriod and TimePhase, have been defined to represent periodic time intervals in the CNTemporalRelation ontology. (2) Relation between Two Events. In many cases in clinical notes, physicians describe the relations between two events without indicating the time stamps of the events. Sentence 5 in Figure 1 shows an example. The CNTemporalRelation ontology is able to represent the relation (after) between the two events (patient’s bilirubin is elevated and the second cycle of chemotherapy). (3) Reification. The CNTemporalRelation ontology defines a TemporalRelationStatement class which enables representing properties of a relation using reification. For example, Rows 5-12 in Table 5 show how to add a modifier “2 weeks” to the relation after by using reification with TemporalRelationStatement class and hasTemporalOffset property. (4) Relative Time. Relative time such as “today”, “tomorrow”, “two months ago”, or “in 3 weeks” is very commonly used in clin-

---

2 event1 has been defined previously in Table 1, but we still show it here to make the table self-explanatory.

3 with protocols approved by Mayo Clinic IRB
ical reports. The CNTemporalRelation ontology captures the relative time information in its original form and at the same time is able to represent the calculated absolute time in the normalized form such as Lines 8 and 9 in Table 4 show. (5) **Uncertainty.** Often temporal information is represent with uncertainty in clinical notes such as sentence 4 in Figure 1 show. The CNTemporalRelation ontology also keeps track of the uncertainty to make sure it can be taken into consideration in answering temporal questions.

We also used the CNTemporalRelation ontology to annotate the temporal information and relations in these sentences. We were successfully annotate 178 events, 98 time instances, 10 time intervals, 53 time phases, and 170 temporal relations. For 142 out of the 153 sentences we extracted, we were able to represent the temporal information and relation precisely without losing any temporal-related information.

For the rest 11 sentences, we believe that we can improve the model to capture the temporal information more precisely. We capture the problems into 4 categories: (1) **Range.** In 2 test sentences, the physicians used a time range to describe a time instant. For example, “stent removal in one-to-two weeks”. We need to improve our ontology to be able to represent a range like this. (2) **Domain Timing Event.** In 6 test sentences, daily living based events (e.g., bed time, breakfast, lunch, and dinner) were used to describe a specific time. We need to improve the ontology to capture the temporal relations between these events. (3) **Timing-Event-Dependent Change.** It is important to monitor the change between two time points or two timing events. For example, in “Most recent ultrasound in May 2007 showed no change comparing to Nov last year”, we can annotate two timing events, "ultrasound in May 2007" and “ultrasound in Nov last year”. But we were not able to annotate “no change” between these two events.

6 Conclusion and Future Work

In this paper, we introduced a semantic-web ontology for temporal relation in clinical narratives. This ontology models temporal information such as timing events, time instants, time intervals, durations, and temporal relations. Based on this ontology, temporal information in clinical narratives can be annotated and represented in RDF. More temporal information and relations can then be inferred by using Semanticweb reasoning tools. Our experimental results indicate that the ontology can successfully represent most of the temporal-related information in real clinical notes.

In addition to the improvements we discussed in the previous section, there are several directions we would like to pursue. First, we would like to connect the CNTemporalRelation ontology to Mayo Clinic’s Text Analysis and Knowledge Extraction System (cTAKES) [15]. We will extend and improve cTAKES and use it as an automatic annotator for temporal information [16] and annotate information with respect to the CNTemporalRelation ontology. Secondly, we want to scale up the data collection and investigate more on reasoning temporal information in clinical narratives. We would also like to address the consistency issues and object identification problem over heterogeneous sources. Thirdly, we plan to evaluate the ontology for other types of medical text such as pathology reports, surgical reports, and radiology reports. Finally, we would like to develop a user-friendly querying mechanism for physicians and clinicians to ask time-oriented clinical questions.

Acknowledgment This research is supported by the National Science Foundation under Grant #0937060 to the Computing Research Association for the CIfellows Project.

References