

CNTRO: A Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives

Cui Tao, PhD Wei-Qi Wei, MM Harold R. Solbrig, BS
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 [1]. 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) [2] 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) [3] 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 ontol-

ogy 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 (Resource Description Framework) triples [4]. The information then becomes "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 [5]. There are two existing temporal ontologies in OWL, the Time ontology in OWL [6] and the SWRL Temporal ontology [7]. Both of the ontologies focus on the relationships between time instances and intervals and it is not obvious how these relationships can be applied to actual events themselves. What we care in the clinical domain, however, is the temporal relations between or the time line of clinical events. 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. This paper builds on previous threads for temporal representations and attempts to harmonize them into a unified model - an OWL based ontology of temporal relations for the purpose of clinical research. The purpose of this ontology is to allow temporal information of clinical data be semantically annotated and queried and to use inference to expose new temporal features and relations based on the semantic assertions and definitions of the temporal aspects in the ontology.

2 CNTRO

In this paper, we introduce CNTRO (Clinical Narrative Temporal Relation Ontology)¹, 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 [6], the SWRL Temporal ontology [7], Allen's temporal relations [10], the TimeML model [9], as well

¹<http://cntro.org>

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

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. CNTRO was developed based on these previous experiences combined with new ontological specifications that fit the needs of natural-language based clinical reports.

The major OWL classes of CNTRO 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 time line. In clinical reports, a time instant can be represented in different levels of granularity 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 levels of granularity. 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 xsd DateTime Data Type format [12]. 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 example 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 CNTRO 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 relation 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 month" and "in approximately two weeks". In CNTRO, 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 *equal*, *before*, *after*, *meet*, *overlap*, *contain*, *during*, *start*, and *finish*. 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 can 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*. The domain of *hasTemporalOffset* is *TemporalRelationStatement* and the range of it is *Duration*. This offset defines the relative timing of a pair of events. In order to model the sentence "patient's bilirubin is elevated 2 weeks after the second cycle of chemotherapy", 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.

```

1. <event1> rdf:type Event;
2.     rdfs:label "the second cycle
           of chemotherapy";
3.     hasTimeStamp <tInst1>;
4. <tInst1> rdf:type TimeInstant;
5.     hasOrigTime "June 10, 2004";
6.     hasNormalizedTime 2004-06-10;
7.     hasGranularity "day";

```

Figure 2: RDF Triple Representation for example 1 in Fig. 1

3 RDF Triple Representation for Temporal Information In Clinical Narratives

Once we have the CNTRO ontology, we can use it to model the temporal instances and temporal relations in clinical narratives. These instances can be stored as RDF triples [4] 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 CNTRO using RDF triples. These examples are chosen from real clinical notes and represent the major temporal expressions of natural clinical language.

Figure 2 shows the RDF triple representation of sentence 1 in Figure 1 and illustrates how to represent an event with a time stamp which is a time instant. For each data instance we want to annotate, we assign it an unique URI and also indicate which class it belongs to. In this example, we have an instance with URI *event1* which belongs to the Class *Event* as Line 1 shows. We use *rdfs:label* to indicate the description of the event such as Line 2 in Table 2 shows. Line 3 indicates that *event1* has a time stamp *tInst1*, which belongs to the *TimeInstant* class and has original value "June 10, 2004" as Lines 4 and 5 show respectively. The triples in italic are inferred values. Since the focus of this paper is to introduce the Clinical Narrative Temporal Relation ontology, we discuss the detailed information about temporal reasoning in [13].

Figure 3 shows the RDF triple representation of example 2 in Figure 1 and illustrates how to represent an event with an interval time stamp. Lines 1-2 define *event2*, which is a new instance of *Event*. Lines 3-4 indicate that *event2* has a time stamp which is a time interval. Lines 5-7 record the start time, end time, and duration of the interval. Lines 8-11 describe the detailed information about the start time. Lines 12-14 describe the detail information about the duration. Information about the end time is missing from the original document, but can be inferred by the start time and duration.

Figure 4 shows the RDF triple representation of example 3 in Figure 1 and illustrates how to represent an event with a time stamp that is a time phase. Since

```

1. <event2> rdf:type Event;
2.     rdfs:label "monitor patient's
           heart rate";
3.     hasTimeStamp <tIntvl>;
4. <tIntvl> rdf:type TimeInterval;
5.     hasStartTime <tInst2>;
6.     hasDuration <durat1>;
7.     hasEndTime <tInst3>;
8. <tInst2> rdf:type TimeInstant;
9.     hasOrigTime "today";
10.    hasNormalizedTime 2004-06-01;
11.    hasGranularity "day";
12. <durat1> rdf:type Duration;
13.     hasValue "72";
14.     hasUnit "hour";
15. <tInst3> hasNormalizedTime 2004-06-04;
16.     hasGranularity "day";

```

Figure 3: RDF Triple Representation for example 2 in Fig. 1

```

1. <event3> rdf:type Event;
2.     rdfs:label "take antibiotics";
3.     hasTimeStamp <tPhas1>;
4. <tPhas1> rdf:type TimePhase;
5.     hasStartTime <tInst4>;
6.     hasDuration <durat2>;
7.     hasEndTime <tInst5>;
8.     hasTimePeriod <tPeril>;
9. <tInst4> rdf:type TimeInstant;
10.    hasOrigTime "today";
11.    hasNormalizedTime 2004-06-01;
12.    hasGranularity "day";
13. <durat2> rdf:type Duration;
14.     hasValue "10";
15.     hasUnit "day";
16. <tInst5> hasNormalizedTime 2004-06-10;
17.     hasGranularity "day";
18. <tPeril> rdf:type TimePeriod;
19.     hasOrigForm "every 8 hours";

```

Figure 4: RDF Triple Representation for example 3 in Fig. 1

```

1. <event4> rdf:type Event;
2.     rdfs:label "see the patient back";
3.     hasTimeStamp <tInst6>;
4.     before <event5>;
5. <event5> rdf:type Event;
6.     rdfs:label "his third cycle
           of chemotherapy";
7. <tInst6> rdf:type TimeInstant;
8.     hasOrigTime "in approximately
           two weeks";
9.     hasNormalizedTime 2004-06-24;
10.    hasGranularity "day";
11.    hasModality "false";

```

Figure 5: RDF Triple Representation for example 4 in Fig. 1

```

1. <event6> rdf:type Event;
2.         rdfs:label "patient's bilirubin
                 is elevated";
3.         hasTimeStamp <tInst7>;
4. <event1> rdf:type Event;
5.         rdfs:label "the second cycle
                 of chemotherapy";
6. <state1> rdf:type TemporalRelationStatement;
7.         rdf:object <event6>;
8.         rdf:predicate after;
9.         rdf:subject <event1>;
10.        hasTemporalOffset <durat3>;
11. <durat3> rdf:type Duration;
12.        hasUnit "week";
13.        hasValue 2;
14. <tInst7> rdf:type TimeInstant;
15.        hasNormalizedTime 2004-06-24;
16.        hasGranularity "day";

```

Figure 6: RDF Triple Representation for example 5 in Fig. 1

TimePhase is a subclass of *TimeInterval*. The representations for start time (Lines 5, and 9-12), duration (Lines 6, and 13-15), and end time (Lines 7, and 16-17) are similar to a time interval. In addition, we defined a time period (Lines 8, and 18-19) to indicate how often the event repeats.

Figure 5 shows the RDF triple representation of example 4 in Figure 1 and illustrates how to represent a temporal relation. We first defined two events (Lines 1-6). In example 5 there are actually two temporal relations. Line 4 indicates that *event4* is *before event5*. And Lines 7-11 represent the time stamp (“in approximately two weeks”) of *event4*.

Figure 6 shows the RDF triple representation of example 5 in Figure 1 and illustrates how to represent a temporal relation using reification. Lines 1-5 define the two events². In Line 6, we defined *state1*, which is an instance of *TemporalRelationStatement*. Lines 7-9 define the temporal relation between the two events by defining the object, predicate, and subject of *state1*. Line 10 defines that *state1* 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 Lines 11-13 define the offset which is an instance of the *Duration* class.

4 Evaluation, Summary, and Discussion

The CNTRO 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 contain temporal information.

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

³with protocols approved by Mayo Clinic IRB

We first compared the expressiveness capabilities of CNTRO with the two existing temporal ontologies in OWL: the Time ontology [6] and the SWRL Temporal ontology [7]. 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 CNTRO 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 CNTRO 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 CNTRO is able to represent the relation (*after*) between the two events (*patient's bilirubin is elevated* and *the second cycle of chemotherapy*). The other two ontologies, on the other hand, only focus on temporal relations between time instants, but events. (3) **Reification**. The CNTRO ontology defines a *TemporalRelationStatement* class which enables representing properties of a temporal relation using reification. For example, Lines 6-13 in Figure 6 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 clinical reports. The CNTRO ontology captures the relative time information in its original form and at the same is able to represent the calculated absolute time in the normalized form such as Lines 11 and 16 in Figure 5 show. (5) **Uncertainty**. Often temporal information is represented with uncertainty in clinical notes such as sentence 4 in Figure 1 show. CNTRO also keeps track of the uncertainty to make sure it can be taken into consideration in answering temporal questions. Currently we capture uncertainties in a relatively simple way. If the original document states that the value is uncertain, we capture it and returns it to users. How to describe the uncertainty in a systematic way while still support meaningful reasoning powers, however, is a non-trivial problem. While OWL can provide means for including numeric uncertainty measures or level of uncertainties as data type properties, there is no standardized way of representing uncertainties. In order to adequately represent uncertainties in OWL, some language extension is necessary. We are currently investigating on adopting this previous work and using OWL to represent temporal uncertainties.

We also used CNTRO to annotate the temporal information and relations in these sentences. We were able to 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. (4) **Negations**. Sometimes clinicians use negations of temporal relations in clinical narratives, such as “no later than”, “not during”, and “not before”. OWL’s monotonicity assumption determines that negation as failure is not supported. Without a not operator, new temporal relation properties such as *not_before*, *not_after* have to be introduced and semantically defined, like what the SWRL Temporal Built-In Ontology does.

5 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 Semantic-web 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 CNTRO ontology to Mayo Clinic’s Text Analysis and Knowledge Extraction System (cTAKES) [14]. We will extend and improve cTAKES and use it as an automatic annotator for temporal information [15] and annotate information with respect to the CNTRO ontology. Sec-

ondly, 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 partially supported by the National Science Foundation under Grant #0937060 to the Computing Research Association for the CIFellows Project and also funded in part by U01-HG04599, the Mayo Clinic eMERGE study.

References

- [1] Shahar Y. Timing Is Everything: Temporal Reasoning and Temporal Data Maintenance in Medicine. In: Proceedings of Artificial Intelligence in Medicine. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM’99). Aalborg Denmark; 1999. p. 30–46.
- [2] OWL Web Ontology Language Reference;. <http://www.w3.org/TR/owl-ref/>.
- [3] A Semantic Web Rule Language Combining OWL and RuleML;. <http://www.w3.org/Submission/SWRL/>.
- [4] Resource Description Framework (RDF);. <http://www.w3.org/RDF/>.
- [5] Zhou L, Parsons S, Hripcsak G. The Evaluation of a Temporal Reasoning System in Processing Clinical Discharge Summaries. JAMIA. 2008;15(1):99–106.
- [6] Time Ontology in OWL;. <http://www.w3.org/TR/owl-time/>.
- [7] The SWRLTab’s Valid-Time Temporal Ontology;. <http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl>.
- [8] Zhou L, Melton G, Parsons S, Hripcsak GA. A temporal constraint structure for extracting temporal information from clinical narrative. Biomedical Informatics. 2006 August;39(4):424–439.
- [9] Markup Language for Temporal and Event Expressions;. <http://www.timeml.org/site/index.html>.
- [10] Allen JF. Maintaining knowledge about temporal intervals. Communications of the ACM. 1983;26(11):832–843.
- [11] HL7 Time Specification;. <http://www.hl7.org/>.
- [12] XML Schema Date/Time Datatypes;. <http://www.w3.org/TR/xmlschema-2/>.
- [13] Tao C, Sobrig HR, Sharma DK, Wei WQ, Savova G, Chute CG. Time-Oriented Question Answering from Clinical Narratives Using Semantic-Web Techniques. In: the 9th International Semantic Web Conference (ISWC 10). Shanghai, China; 2010. (submitted).
- [14] cTAKES on Open Health Natural Language Processing (OHNLP) Consortium;. <http://www.ohnlp.org>.
- [15] Savova G, Bethard S, Styler W, Martin JH, Palmer M, Masanz J, et al. Towards temporal relation discovery from the clinical narrative. In: Proceedings in the American Medical Informatics Association (AMIA) Annual Symposium. San Francisco, California; 2009. .