Knowledge Representation and Reasoning for Complex Time Expression in Clinical Text (Postprint)

Authors: Danyang, Hu, Meng, Wang, Feng, Gao, Xu Fangfang, Jinguang, Gu, Fangfang Xu

Date: 2022-11-28T00:00:00+00:00

Abstract

Temporal information is pervasive and crucial in medical records and other clinical text, as it formulates the development process of medical conditions and is vital for clinical decision making. However, providing a holistic knowledge representation and reasoning framework for various time expressions in the clinical text is challenging. In order to capture complex temporal semantics in clinical text, we propose a novel Clinical Time Ontology (CTO) as an extension from OWL framework. More specifically, we identified eight time-related problems in clinical text and created 11 core temporal classes to conceptualize the fuzzy time, cyclic time, irregular time, negations and other complex aspects of clinical time. Then, we extended Allen' s and TEO' s temporal relations and defined the relation concept description between complex and simple time. Simultaneously, we provided a formulaic and graphical presentation of complex time and complex time relationships. We carried out empirical study on the expressiveness and usability of CTO using real-world healthcare datasets. Finally, experiment results demonstrate that CTO could faithfully represent and reason over 93% of the temporal expressions, and it can cover a wider range of time-related classes in clinical domain.

Full Text

Preamble

RESEARCH PAPER ChinaXiv 合作期刊

Knowledge Representation and Reasoning for Complex Time Expression in Clinical Text

Danyang Hu[1,2,3],, *Meng Wang*[1,2,3],, Feng Gao[1,2,3], Fangfang Xu[1,2,3]†, Jinguang Gu[1,2,3]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[2]Key Laboratory of Rich-Media Knowledge Organization and Service of Digital Publishing Content, National Press and Publication Administration of the People’ s Republic of China, Beijing 10038, China
[3]Institute of Big Data Science and Engineering, Wuhan University of Science and Technology, Wuhan 430065, China

Keywords: Clinical text; Temporal ontology; Temporal relations; OWL; Negation of temporal relation

Citation: Hu, D.Y., Wang, M., Gao, F., et al.: Knowledge Representation and Reasoning for Complex Time Expression in Clinical Text. Data Intelligence 4(3), 573-598 (2022). doi: 10.1162/dint_a_{00152}

Received: Oct. 10, 2021; **Revised:** March 15, 2022; **Accepted:** May 10, 2022

ABSTRACT

Temporal information is pervasive and crucial in medical records and other clinical text, as it formulates the development process of medical conditions and is vital for clinical decision making. However, providing a holistic knowledge representation and reasoning framework for various time expressions in clinical text is challenging. In order to capture complex temporal semantics in clinical text, we propose a novel Clinical Time Ontology (CTO) as an extension of the OWL framework. More specifically, we identified eight time-related problems in clinical text and created 11 core temporal classes to conceptualize fuzzy time, cyclic time, irregular time, negations, and other complex aspects of clinical time. We then extended Allen’ s and TEO’ s temporal relations and defined relational concept descriptions between complex and simple time.

Simultaneously, we provided both formulaic and graphical presentations of complex time and complex temporal relationships. We carried out an empirical study on the expressiveness and usability of CTO using real-world healthcare datasets. Experimental results demonstrate that CTO can faithfully represent and reason over 93% of temporal expressions and cover a wider range of time-related classes in the clinical domain.

*These authors contributed equally to this work.
†Corresponding author: Fangfang Xu (E-mail: xuff@wust.edu.cn; ORCID: 0000-0003-1057-6444).

1. INTRODUCTION

Time-related information is an essential component of inpatient medical records, doctor prescriptions, and other clinical text. It describes the sequence of symptoms and treatments received by patients and is crucial for clinical decision-making and medical research [1, 2]. However, natural language expressions about time and temporal relation concepts are diverse and can vary with different contexts and writers. Current research on temporal knowledge representation in text has mainly focused on news or open domains [3, 4, 5]. However, due to distinct features in the medical domain, research results from the news domain or open domain cannot be applied directly to the medical domain [6].

Temporal knowledge representation is challenging and has been discussed extensively. Generally, OWL-Time [7] can be used to represent instants and intervals, and possibly Allen's 13 temporal relations [8]. The reasoning engine CHRONOS [9] can be used for relation inference, inconsistency detection, and path consistency checking. However, these ontologies do not cover some important features, such as irregular time, granular time, and modality. These ontologies designed for general domains are not adequate enough for specific domains such as healthcare [10]. To fill the gap between general tools and the clinical domain, Tao et al. [11] proposed the temporal ontology CNTRO. After that, Li et al. [12] extended it and proposed the TEO ontology. They leveraged a Java-based TEO reasoner to realize complex timeline reasoning, which can represent granular time, instants, intervals, and incomplete time. However, there is also uncertain and irregular time in the clinical domain, such as "take pills around 6 a.m." and "The patient received rehabilitation training twice a day last year, and once every 2 days this year." The former describes an uncertain time—there is no precise time point for the occurrence of an event, but there is still a rough range (for example, "4 p.m." is not in the range). The latter describes irregular time and represents the phased change of the event.

Another challenge is the representation of inference queries with complex time. For example, if we know that "Patient A had surgery yesterday and was advised by his doctor to do 20 minutes of rehab around 9:00 am every morning," then for the question "What should Patient A do at 9:05 am this morning?" Humans can quickly answer it. Nevertheless, all of the ontologies mentioned above cannot obtain the answer. Based on this, we propose Clinical Time Ontology (CTO), which aims to represent complex temporal information in clinical texts.

The contributions of this paper can be summarized as follows:

- We defined 11 core time classes and proposed a robust ontology for modeling complex temporal concepts and relations.
- We introduced the concept of complex time-related classes and complex temporal relations using temporal logic language and depicted them in the form of diagrams and formulas.
- We invited three annotators to annotate real data and evaluated the results using IAA metrics and the ontology' s coverage ability. The results show that the CTO ontology can cover a wider range of time-related classes and temporal relationships in medical record data, and it can faithfully represent 93% of time-related classes and temporal relationships.

The remainder of this paper is organized as follows: Section 2 discusses related work; Section 3 summarizes eight types of time-related problems in the clinical domain and introduces 11 core time classes; Section 4 elaborates on CTO ontology' s design principles; Section 5 presents the definition of complex time classes and complex time relations in terms of semantics; Section 6 depicts datasets, experiments, and comparisons with other ontologies about representation; Section 7 summarizes this paper.

2. RELATED WORK

The representation of temporal information has been discussed extensively in the Semantic Web. In this paper, we discuss general-purpose temporal ontologies and medical domain temporal ontologies separately.

2.1 General Domain Temporal Ontologies

Hobbs et al. [13] proposed the DAML ontology of time that integrates first-order predicate calculus for topological temporal relations such as intervals, events, dates, and times. Based on this, the OWL-Time ontology [7] provides vocabularies for instants and intervals, duration, datetime, and Allen' s temporal relations [8] in terms of time ontologies. However, OWL-Time focuses on definitive, accurate timestamp-based time representation and cannot cater to complicated time expressions (e.g., periodic time, fuzzy time, and multi-granularity time representation).

The SWRL temporal ontology [14] defined OWL entities to represent interval-based time information in OWL and provided OWL entities to represent all entities defined by the model. It contains propositions, effective time (instants and intervals), granularity, and duration, and also contains a set of SWRL built-in functions that can be used to reason about temporal information. Eleftherios et al. [9] proposed CHRONOS, a system for reasoning about temporal information in OWL ontologies. CHRONOS provides both qualitative and quantitative information representation, enabling the inference of implied relations, detection of inconsistencies, and ensuring path consistency. Vadim et al. [15] proposed the PSI-Time Ontology, which can represent concepts of some complex time such as relativist and absolutist duration, periodic time intervals, interval phases, open time, linear time, and the relations point-to-point, point-to-interval, and interval-to-interval. However, they do not cover all time-related aspects in the clinical domain and are difficult to link to structured or unstructured clinical data.

2.2 Clinical Domain Temporal Ontologies

Tao et al. [11] proposed the Clinical Narrative Temporal Relations Ontology (CNTRO), which aims to fill the semantic gap in temporal modeling in the clinical domain. It provides vocabularies such as Event, Time, Duration, Granularity, Precision, and TemporalRelationStatement. Li et al. [12] extended CN-

TRO to develop a novel Temporal Event Ontology (TEO), which can formally represent and reason about structured and unstructured temporal information, supporting flexible temporal relationship representation and reasoning. However, TEO is still insufficient in representing and reasoning about fuzzy time, irregular time, and complex relationships.

The representation of complex time (e.g., fuzzy time, irregular collection time, etc.), subjectivity, negation, and complex temporal relationships is a significant challenge for temporal ontologies. Based on this, we propose Clinical Time Ontology (CTO), which aims to achieve inference and query of complex temporal information problems and mine more relationships between them.

3. TEMPORAL EXPRESSIONS IN CLINICAL TEXT

Temporal information is pervasive in clinical data. However, it is nontrivial to capture temporal information in unstructured text from clinical data, which is pivotal to healthcare decision-making. For example, a patient' s self-description and previous medical history can be obtained from medical record data. It could take the form of a paragraph in natural language, and healthcare providers need to digest this paragraph to understand the chronological events it implies to make a diagnosis or treatment plan.

Formulating a semantic framework for time-related expressions in clinical text is a first step toward automated reasoning and decision support. In this section, we first categorize the time-related expressions in typical clinical text, and then we provide a set of key time concepts.

3.1 Complicated Expressions of Time in Clinical Text

We refer to TIDES [16] for a basic classification of temporal expressions, then further refine the categorization of temporal expressions in the clinical domain based on our observation of over 3.5K anonymized real-world clinical records. We identified different types of time descriptions from clinical text according to various temporal aspects, including completeness, accuracy, repetition, subjectivity, etc.

Regarding completeness, there can be complete or incomplete time; regarding accuracy, there can be accurate or fuzzy time; regarding repetition, there can be cyclic or irregular time. Based on this, we divided time into five types (e.g., absolute time, lack information time, fuzzy time, cyclic time, and irregular time).

We also identified three qualitative questions: questions of subjectivity (e.g., whether it is worth believing), questions of temporal relationships (e.g., before, after, etc.), and questions of negation of temporal relationships (e.g., not before, not after, etc.). In general, we summarized eight problems, which are shown in Table 1 . “Absolute time”is a point or duration that can be represented separately on the time axis. “Lack information time” means the sentence describing the time information about the event is incomplete and includes two subcategories: "relative time" and "incomplete time." The former means we need to find the reference time and obtain the real time by calculation. The latter is commonly used to describe a time interval but without complete information (e.g., lacking start time or end time). "Fuzzy time" indicates that there exist approximate and vague modifiers in the sentence, and it may be an instant or an interval. "Cyclic time" means the event is repeated and the occurrence is regular, and it may be a collection of multiple instants or intervals. "Irregular time" means the event is repeated but not regular, and its items may be instants, intervals, or collections.

3.2 Time Types

The temporal expressions listed in Table 1 can be addressed using a combination of syntactic forms and/or semantic concepts. Syntactically, all temporal expressions can use one, two, or more timestamps to denote the relevant instance, interval, or repetitions; additionally, temporal concepts and relations can have modifiers to express beliefs or polarity. Semantically, a temporal concept can be absolute, fuzzy, relative, or periodic, etc. Based on the first five problems in Table 1, we categorized time. Problems 6 and 8 motivate the subjective and negation modifiers. We analyzed the combinations of the above syntactical and semantic axes of temporal concepts and relations and organized them in Table 2 with correlations to the temporal expression types listed in Table 1.

"Absolute Instant" is a point in time with no fuzzy (e.g., about, several) and relative modifiers (e.g., ago, after). "Relative Instant" is a point in time but with relative modifiers. "Fuzzy Instant" is a point in time but with fuzzy modifiers. "Absolute Interval" is a duration in time whose start and end times are "Absolute Instant" instances. "Relative Interval" is a duration in time in which at least one of the start or end times is a "Relative Instant" instance. "Fuzzy Interval" is a duration in time, and the condition is satisfied as long as one of its three components (duration quantity, start time, and end time) has a fuzzy modifier, even though its start time may make it resemble a "Relative Interval" instance. "Incomplete Interval" is a duration in time with only one of the three components, regardless of whether the component has a modifier.

"Periodic Instant Collection" is a collection of points in time commonly used to describe the time for events in a cycle, where the occurrence time is also a point in time. "Irregular Instant Collection" is a collection of points in time commonly used to describe the time for events with irregular occurrence times, where the occurrence time is a point in time, and its items can also be some "Periodic Instant Collection" instances. "Periodic Interval Collection" is similar to "Periodic Instant Collection," but the difference is that the occurrence time of the former is a duration. "Irregular Interval Collection" is similar to "Irregular Instant Collection," but the occurrence time is a duration, and its items can also be some "Periodic Interval Collection" instances.

We divided the time appearing in the clinical field into the 11 main types mentioned above, and there is no intersection between them. For negation and subjective modifiers, we treat them as properties of temporal instances.

4. CTO ONTOLOGY

We developed the Clinical Time Ontology (CTO) for semantic representation and reasoning over the temporal aspects discussed in Section 3. In this section, we elaborate on its module design.

4.1 Overview

Temporal concepts cannot be utilized alone, so we need to bind them with the concept of Events. Therefore, we defined the class Event to represent clinical events, with subclasses such as HospitalizationEvent, DischargeEvent, etc. We can extend the list of subclasses as needed. According to Table 2, we can obtain 11 classes, such as AbsoluteInstant, RelativeInstant, FuzzyInstant, AbsoluteInterval, etc. They are all subclasses of Time. Class Event and class Time can be linked by the object property hasTime. We referred to the OWL-Time and TEO ontologies when designing our ontology.

We have some common and different concepts. For example, class AbsoluteInstant is the same as Instant in OWL-Time, and class TimeQuantity is the same as Duration in TEO. However, class IrregularInstantCollection does not exist in their ontologies. Figure 1 [Figure 1: see original paper] shows the taxonomy of CTO and the relationship with OWL-Time and TEO (using sameAs).

4.2 Module Design

We divided time into four modules: Instant, Interval, InstantCollection, and IntervalCollection. Figure 1 illustrates this, and more details are provided below:

4.2.1 Instant Class Instant has three subclasses: AbsoluteInstant, RelativeInstant, and FuzzyInstant, designed for absolute instant time, relative instant time, and fuzzy instant time, respectively. There exist several properties that describe time at a granular level, such as second, minute, and hour. Class TimeUnit is designed for representing time' s precision, meaning the smallest granularity unit of time.

The object property prefixAnchor is designed to give a qualitative representation of the relationship between the moment and its reference moment. The object property relativeQuantity can quantitatively represent the length of difference. For example, "Hospitalized on July 1, 2021, and started showing significant symptoms three days ago." "Three days ago" is a relative time description, which we can set to . "July 1, 2021" is an AbsoluteInstant instance and serves as the prefixAnchor of . The relativeQuantity of is "3 days." Figure 2 [Figure 2: see original paper] illustrates this case.

For fuzzy time, the object property leftAnchor can indicate its earliest possible occurrence time, and rightAnchor can indicate the latest possible occurrence time. For example, “Early January” is a fuzzy description, and we can set its leftAnchor to for “January 1” and its rightAnchor to for “January 10.” Figure 3 [Figure 3: see original paper] illustrates this case.

4.2.2 Interval Class Interval has four subclasses: AbsoluteInterval, RelativeInterval, FuzzyInterval, and IncompleteInterval. The first two both have object properties startTime, endTime, and durationQuantity. The differences between them are that their startTime’ s and endTime’ s ranges are not the same class.

Object properties minDurationQuantity and maxDurationQuantity are designed for fuzzy intervals, indicating the minimum possible duration and the maximum possible duration, respectively. For example, “it started on January 1 and lasted for about half a month.” Figure 4 [Figure 4: see original paper] illustrates this case.

Class IncompleteInterval is a special case of Interval, which has only startTime or endTime and does not have durationQuantity.

4.2.3 InstantCollection and IntervalCollection The temporal description of medication and check-up events in clinical medical texts is often cyclical. For example, “From today, review daily at 2:00 pm for five days.” In addition, some irregular time collections still exist in medical texts, such as “The patient received rehabilitation training twice a day last year and once every 2 days this year.” Each occurrence may be an instant, interval, or collection. Therefore, we designed class InstantCollection and class IntervalCollection. Class InstantCollection has two subclasses: PeriodicInstantCollection and IrregularInstantCollection, and class IntervalCollection has two subclasses: PeriodicIntervalCollection and IrregularIntervalCollection.

The object property periodicQuantity can quantitatively represent the cycle, and the data property periodicCount can represent the number of cycles. The object property eachTime is designed to represent the time of each occurrence, whose range can be class Instant or Interval. We can use periodicQuantity with a TimeQuantity instance to represent “daily” and use durationQuantity with a TimeQuantity instance to represent “for five days.” Figure 5 [Figure 5: see original paper] illustrates this case.

We can implement the representation of irregular time in two steps as follows: (1) create some Time instances; (2) connect them using object properties belongTo or hasSubset and data property itemIndex. The Time instance can be not only Instant or Interval but also InstantCollection or IntervalCollection. We can set the second sentence mentioned above to . There are two stages, and both are cyclic time. They can be represented by two instances of PeriodicInstantCollection: and . Then, we use object property hasSubset to connect them.

Figure 6 [Figure 6: see original paper] illustrates this case.

4.2.4 Subjective Time In clinical texts, it is widespread for patients to describe the occurrence time of events themselves. In this case, the time in the text is subjective and may not be the real time of the event. To represent subjectivity, each Time instance can use the data property subjective with True or False value to indicate whether it is a subjective description by the patient or an objective record, allowing physicians to make judgments about that time information. Figure 7 [Figure 7: see original paper] illustrates this.

4.2.5 Temporal Relationship Besides the aforementioned cases, there are also cases that indicate the temporal relationship between events, e.g., "receiving treatment B is earlier than taking medicine A." Allen's 13 temporal relationships can be used to represent partial situations; however, they cannot cover all cases, such as point-to-collection, duration-to-collection, and collection-to-collection situations. Therefore, to fill this gap, we extended them with additions such as containAll, containNull, and their inverse properties. Figure 8 [Figure 8: see original paper] illustrates a sample.

4.2.6 Negation The application of chronological relations sometimes appears with negation. For example, "taking medicine A is not earlier than the end of treatment B." In medical record data, as long as we cannot know or infer that "taking medicine A" occurs before "treatment B," we can assume that A is not before B. Based on the closed-world assumption (CWA), we designed the algorithm not, which can be used to represent the negation of a temporal relationship, such as not before, not after, and not overlap, etc.

5. CTO SEMANTICS

This section introduces class concepts and temporal relationship concepts semantically and graphically. Formally, Baader and Hanschke [17] proposed an extension to ALC, named ALC(D). We describe the complex time classes and complex temporal relationships in the scheme of this language, and the reasoner can implement inductive reasoning according to the descriptions.

5.1 Complex Time Classes Definition

We defined some new classes for representing complex time, e.g., PeriodicInstantCollection, PeriodicIntervalCollection, etc. Their descriptions are shown below.

The description of concept PeriodicInstantCollection means its instance has start-time (Instant instance), periodic-quantity (Quantity instance), and each-time (Instant instance). It is worth noting that it may have end-time (Instant instance) or periodic-count (Integer instance). Quantity denotes the quantification distance between one time and another. Pair means the two objects are used to calculate the relationship (one-by-one combinations between all time-related types).

5.2 Complex Temporal Relation Definition

We extended Allen’ s 13 temporal relationships to cover more comparisons between temporal types, such as startPoint, containAll, etc. We can determine whether the time of two events is completely covered, intersected, or unrelated based on whether there exists a relationship allDuring, someDuring, or nullDuring between the times of the two events. before, after, and meet are special cases of nullDuring, but there are more than just these three cases of nullDuring; there is also a case similar to overlap, but it may be that each element of the two collections alternates without intersecting each other. These three temporal relationships are mainly for Collection’ s subclasses. startPoint and finishPoint are mainly for Instant’ s subclasses; however, start and finish are mainly for Interval’ s and IntervalCollection’ s subclasses.

All temporal relationships are shown in Figure 9 [Figure 9: see original paper]. We use different colors to distinguish between a, b objects, simple and complex time classes, as well as Allen’ s, TEO’ s, and CTO’ s temporal relationships.

In Table 3 , we used three auxiliary functions: AllItems(), CalTime(), and StartTime(). The semantics of these functions are provided in Appendix A.

5.3 Negation Temporal Relation Definition

If the reasoning machine can infer that there exists a temporal relation r1 between time A and time B but cannot conclude that there exists temporal relation r2, then according to CWA, there is a negation relation not r2 between the two times. For example, if the reasoner cannot reason out before , not before can be concluded. All temporal relationships can have their negations. Table 4 introduces some examples.

6. EMPIRICAL STUDY

To verify the ability of the CTO ontology to express temporal information in the clinical field, we selected a set of case texts from people with mental health conditions. We used an anonymized realistic dataset from a hospital in China. We randomly selected 300 patients, which contained approximately 3000 Chinese statements. We then arranged annotators to annotate the temporal information and conducted a quantitative analysis using the Inter-Annotator Agreement (IAA) [18] metric. The IAA metric is commonly used to evaluate annotation consistency and the difficulty of annotation, usually focusing on Precision, Recall, and F1. Simultaneously, we analyzed the characteristics of other ontologies and compared the expressiveness of CTO with other ontologies in terms of complex time and temporal relations.

6.1 Evaluation Results

For IAA, three annotators were asked to annotate temporal information in 300 patient cases jointly. The evaluation consists of two parts: (1) evaluation of classes and object properties (without temporal relations), and (2) evaluation of temporal relationships. For evaluation of classes and object properties, they first annotated classes, and the annotation methods were discussed together. After that, they discussed and agreed on the Gold Standard. The annotation of object properties was the same, and the base data is the Gold Standard.

The mean value of the F1 metric for time-related classes annotation was 77.18% and 83.06% (exact mapping and partial mapping). The average F1 metric of object properties (without temporal relationship object properties) was 87.34%. Considering that complex time information can be represented in various ways, such as representing a time interval using startTime/endTime + durationQuantity or startTime + endTime, if such errors are ignored, the average score is 93.61%.

For evaluation of temporal relations, we asked three annotators to annotate the temporal relations between time entities using the Gold Standard obtained from the previous sub-process. We developed a reasoning machine for the experiment evaluating temporal relations and regarded the results as the Gold Standard for the second evaluation task. Note that we treat pairs of tokens that have the inverse property as true positive as well, such as before and after . The mean values of Precision, Recall, and F1 are 96.98%, 91.97%, and 94.39%, respectively. We also calculated the three metrics for each temporal relationship to analyze which temporal relations are more difficult to express. The results are shown in Figure 10 [Figure 10: see original paper].

From Figure 10, it can be seen that it is easy to reach a consensus on the representation of simple relations, such as after, overlap, endBeforeEnd, etc. However, there are difficulties in representing more complex temporal relations, such as containAll, finishPoint, etc. After analysis, we found that manual annotation shows omissions for this type of data but still maintains a high F1 score, which indicates that the CTO ontology can also achieve better results in representing temporal relations.

Regarding the ability of ontologies to cover temporal instances, we found that CTO ontologies can cover almost all types of temporal descriptions. As can be seen from Figure 11 [Figure 11: see original paper] and Figure 12 [Figure 12: see original paper], complex time types (PeriodicInstantCollection, PeriodicIntervalCollection, IrregularInstantCollection, IrregularIntervalCollection) account for approximately 16.77% (594 of 3542) of all time types, and the extended temporal relationships (containAll, containSome, containNull, startPoint, finishPoint) account for 21.19% of all temporal relationships (629 of 2969). We provide two sample texts from clinical records and their relevant knowledge graphs annotated with CTO in Appendix B.

6.2 Expressiveness Analysis

In this section, OWL-Time, PSI-Time, and TEO ontology are compared with the CTO ontology regarding their ability to represent the eight problems presented in Section 3. Table 5 shows the results, and we use the mark √ for those ontologies capable of obtaining problem results by querying.

For Problem-1 (Absolute Time Problems), all of them can be queried to obtain the exact time, even the hour, minute, and second.

For Problem-2 (Lack Information Time Problems (Relative Time and Incomplete Time)), OWL-Time mainly focuses on absolute instant and interval and cannot represent relative quantities and certainly cannot represent incomplete time, as TEO noted. If we want to obtain the time referred to by "three days before hospitalization," we need to implement a representation of relative quantities as well as reference objects, but they do not have this capability and therefore cannot answer questions relating to such times.

For Problem-3 (Fuzzy Time Problems), when we want to compare the temporal relationships between events that occurred at more ambiguous times, it is inappropriate to compare their absolute times. For example, the temporal relationship between time A "at about 8:10 a.m. on November 1, 2021" and time B "at 8 a.m. on November 1, 2021, which lasted about 9 minutes" could be during, after, or finishPoint. It is, of course, related to the range of fuzzy values. By introducing the maximum and minimum values as left and right anchors, we can compare them in terms of temporal relationships and obtain a range of values for the end time. OWL-Time and PSI-Time do not consider the concept of fuzzy time and therefore cannot be queried for answers; TEO does consider fuzzy time, but it only marks time by True and False and does not analyze it quantitatively.

For Problem-4 (Cyclic Time Problems), we can query PSI-Time, TEO, and CTO ontologies to obtain the start time, each occurrence, and the frequency of the periodic time to answer questions such as "When is the next check?" "How many times has it been done so far?" etc.

For Problem-5 (Irregular Time Problems), one of the characteristic points of the CTO ontology is that an item of the IrregularCollection can still be a collection. Thus, it is possible to implement a stage-by-stage representation of time, e.g., "once a day for the first session, once a week for the second session," etc. Not only can we obtain internal time-related calculations such as "How long between the first and second sessions?" and "Which session is now?" but we can also compare the temporal relationship with other times.

For Problem-6 (Subjective Problems), we use a True or False flag, which is the value of the subjective property, to provide a basis for users to refer to when reasoning or querying results.

For Problem-7 (Temporal Relation Problems), all of them can obtain the answer.

Both TEO and CTO ontologies extend the temporal relations, and the CTO ontology defines different temporal relations between multiple types of time.

For Problem-8 (Negation Problems), the CTO ontology regards case data as a closed world, and if there does not exist relation r between times a and b, then (a, not r, b) is satisfied. Thus, the CTO ontology can answer that question based on CWA.

Overall, the CTO ontology is superior to other temporal ontologies in terms of answering time-related clinical questions.

7. CONCLUSION AND FUTURE WORK

In this paper, we identified eight time-related problems in the clinical domain and proposed the CTO ontology with a taxonomy of complicated temporal concepts to solve these problems, which covers the fuzziness, relativity, irregularity, and cyclical aspects of time expressions in clinical text. In addition, we also marked whether the time is subjective so that we can deal with it in the future. Simultaneously, we defined the concept description of complex time-related classes and extended Allen' s and TEO' s temporal relationships to support temporal relations for complex time. More importantly, we provided definitions of the various temporal relationships between all time-related types (Instant, Interval, Collection) in both formulaic and pictorial forms. Experimental results show that complex time classes account for 16.77% of all classes in clinical medical texts and the extended temporal relationships account for 21.19% of all temporal relations, and the CTO ontology can represent more than 93% of temporal information and temporal relationships. In summary, we proposed a robust temporal ontology capable of expressing a variety of complex temporal information in clinical texts.

The CTO ontology can be improved or extended in the following aspects: (1) The reasoning about fuzzy time currently uses the method of setting left and right anchors, and the problem of selecting left and right anchor points is also tricky. In the future, we hope to use probabilistic methods to achieve probability-level representation and reasoning capabilities. (2) Building a patient history knowledge graph using the CTO ontology automatically is also on the agenda. By this means, we can further validate and improve its use in realistic scenarios.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No. U1836118), the Open Fund of Key Laboratory of Content Organization and Knowledge Services for Rich Media Digital Publishing (ZD2021-11/01), and the Natural Science Foundation of Hubei Province Educational Committee (B2019009).

Prof. Zhisheng Huang of Vrije Universiteit Amsterdam provided the data for this study.

AUTHOR CONTRIBUTIONS

Danyang Hu (HuDanyang wust@163.com) and Meng Wang (wangmeng@wust.edu.cn) planned and contributed equally to the overall work. Both are co-responsible for analyzing case data, designing the research, implementing the approach, and writing the manuscript. Feng Gao (feng.gao86@wust.edu.cn) contributed to the design of the research, the analysis of the results, and the revising of the manuscript. Fangfang Xu (xuff@wust.edu.cn) contributed to the design of the research and the analysis of the results. Jinguang Gu (simon@wust.edu.cn) proposed the research problems and contributed to the design of the research.

REFERENCES

[1] Olex, A.L., McInnes, B.T.: Review of temporal reasoning in the clinical domain for timeline extraction: Where we are and where we need to be. Journal of Biomedical Informatics 118, 103784 (2021)

[2] Moharasan, G., Ho, T.B.: Extraction of temporal information from clinical narratives. Journal of Healthcare Informatics Research 3(2), 220–244 (2019)

[3] Piskorski, J., Zavarella, V., Atkinson, M., et al.: Timelines: Entity-centric event extraction from online news. In: Text2Story@ ECIR, pp. 105–114 (2020)

[4] Ding, K., Zhou, F., Miao, J., et. al: Temporal relation recognition method for news events based on cross event theory. Computer Engineering 34(6), 6 (2017).

[5] Zhao, X., Jin, P., Yue, L.: Discovering topic time from web news. Information Processing & Management 51(6), 869–890 (2015)

[6] Zhou, X., Zhou, Q., Li H.: Automatic annotation of temporal information in Chinese medical narrative records. Chinese Journal of Biomedical Engineering 31(3), 434–439 (2012)

[7] Hobbs, J.R., Pan, F.: Time ontology in OWL. W3C working draft 27. Available at: https://www.w3.org/TR/owl-time/ (2006). Accessed on 2021

[8] Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM 26(1), 832–843 (1983)

[9] Anagnostopoulos, E., Batsakis, S., Petrakis, E.G.M.: CHRONOS: A reasoning engine for qualitative temporal information in OWL. Procedia Computer Science 22, 70–77 (2013)

[10] Madkour, M., Benhaddou, D., Tao, C.: Temporal data representation, normalization, extraction, and reasoning: A review from clinical domain. Computer Methods and Programs in Biomedicine 128, 52–68 (2016)

[11] Tao, C., Wei, W., Solbrig, H.R., et al.: CNTRO: a semantic web ontology for temporal relation inferencing in clinical narratives. In: AMIA annual symposium proceedings. American Medical Informatics Association p. 787 (2010)

[12] Li, F., Du, J., He, Y., et al.: Time event ontology (TEO): to support semantic representation and reasoning of complex temporal relations of clinical events. Journal of the American Medical Informatics Association 27(7), 1046-1056 (2020)

[13] Hobbs, J.R., Ferguson, G., Allen, J., et al.: A daml ontology of time. Available at: http://www.cs.rochester.edu/ (2002). Accessed on 2021

[14] SWRL Temporal Ontology. Available at: http://protege.cim3.net/cgibin/. Accessed on 2021

[15] Ermolayev, V., Keberle, N., Matzke, W.E.: An ontology of environments, events, and happenings. In: 2008 32nd Annual IEEE International Computer Software and Applications Conference, pp. 539-546 (2008)

[16] Ferro, L., Gerber, L., Mani, I., et al.: Standard for the annotation of temporal expressions-tides. The MITRE Corporation, McLean-VG-USA (2005)

[17] Baader, F., Philipp, H.: A scheme for integrating concrete domains into concept languages. In: Proceedings of the 12th International Joint Conference on Artificial Intelligence, pp. 446-451 (1991)

[18] Artstein, R.: Inter-annotator agreement. Handbook of linguistic annotation, Springer, Dordrecht (2017)

APPENDIX A. SEMANTICS FOR THE AUXILIARY FUNCTIONS IN TEMPORAL RELATION DEFINITION

The AllItems(a.Collection) function can get all elements in a Collection instance, which is convenient for calculation, especially with cyclic collections that need to be calculated to obtain the real time of each occurrence. The CalTime(a.Instant, q.Quantity) function can get the calculated real time point of relative time or the result of temporal addition. The StartTime(a.Time) function provides an interface for obtaining the start time of all time types, which is similar to the EndTime(a.Time) function.

APPENDIX B. USAGE EXAMPLES FOR TEMPORAL REPRESENTATION AND QUERY

Figures B.13-B.16 illustrate the details, where the words in red italic are manually annotated as events and times. We use lines with arrows to establish a correspondence between statements containing temporal information and the class to which that temporal information belongs.

Figure B.13 shows sample text that includes Absolute Instant Time, Absolute Interval Time, Relative Time, Fuzzy Time, and Incomplete Time.

Figure B.14 shows sample context that includes Cyclic Time, Irregular Time, Subjective Time, and apparent Temporal Relation.

Figure B.15 illustrates the representation for the text shown in Figure B.13 using CTO.

Figure B.16 illustrates the representation for the text shown in Figure B.14 using CTO.

Figure B.13 is an example case data that contains some events (e.g., hospitalization, discharge, medication, and paroxysm events) and some types of time (e.g., absolute instant time, absolute interval time, relative time, fuzzy time, and incomplete time). Figure B.15 illustrates the representation.

Figures B.13 and B.15 present the time representation for problems 1-3. For problems 4-7, we found some short sentences from the dataset. Figure B.14 shows the corpora, and Figure B.16 illustrates the representation.

AUTHOR BIOGRAPHY

Danyang Hu is currently a graduate student at the School of Computer Science and Technology, Wuhan University of Science and Technology. Her research interests include Knowledge Graph and Knowledge Representation. ORCID: 0000-0002-4242-099X

Meng Wang is currently a graduate student at the School of Computer Science and Technology, Wuhan University of Science and Technology. His research interests include Data Mining and Knowledge Graph. ORCID: 0000-0002-2540-9991

Feng Gao obtained his Ph.D. degree in computer science from National University of Galway, Ireland in 2016 and has been working at Wuhan University of Science and Technology since 2016. His main research area is Knowledge Graph/Semantic Web. ORCID: 0000-0002-2396-1360

Fangfang Xu is currently a PhD candidate at the School of Computer Science and Technology, Wuhan University of Science and Technology. She is a lab master. Her research interest is Semantic Computing. ORCID: 0000-0003-1057-6444

Jinguang Gu obtained his Ph.D. degree from Wuhan University in 2005. He is currently a professor and doctoral supervisor at Wuhan University of Science and Technology. His main research areas are Distributed Computing and Knowledge Graph. He has published more than 100 papers and chaired more than 10 projects at provincial and ministerial levels. ORCID: 0000-0002-8823-8480