Temporal Information and Event Markup Language (TIE-ML)

Billy Dickson, Soyoung Kim, Damir Cavar, Ali Aljubailan
Indiana University Bloomington

ABSTRACT

TIE-ML is a markup strategy and annotation schema to improve the productivity and accuracy of temporal and event related annotation of corpora to facilitate machine learning based model training. For the annotation of events, and temporal sequencing and duration of events, it is significantly simpler and easier to use than much more sophisticated formalisms and approaches such as TimeML.

INTRODUCTION

Natural languages use various means to express events and place them in time. Tense, aspect, mood, and modality make up the foundations of this process, and each language utilizes a certain combination of these tools to indicate temporal information.

Developing corpora and data sets is essential for quantitative studies of distributional properties of temporal and event logic phenomena and expressions. It also allows us to develop machine learning based automatic annotation and processing of event sequencing and temporal aspect as for example duration.

Event Sequencing:

We observe that presentation order of events does not always match the temporal sequence.

(1) *Wash the veggies, chop them, and fry them.*
1 2 3
(2) *Before you fry the veggies, wash, and chop them.*
3 1 2

Events can also overlap and occupy simultaneous points in time:

(3) *John lived in Paris while Susan studied in Berlin.*
E: 1 2
T: 1 1

Tense (Reichenbach 1947)

Reichenbach theory introduces time variables: event time (E), reference time (R), and speaker time (S). Event time refers to the time of an event, reference time is the point of reference along the time axis that the event occurs with respect to, and speaker time refers to time of utterance. The relative sequence of R and S in Reichenbach's model reflects the tense categories *present* (S and R overlap), *past* (R precedes S), and *future* (S precedes R).

We can also use the ordering of these variables to determine assertion of fact.

(4) *Apple merged with Alphabet.*
(5) *Apple will merge with Alphabet.*

If E precedes S, the event could be asserted to be factual, while S preceding E implies that the event is a hypothetical projection into the future.

The situation changes if a predicate with a tense as in (4) is in the scope of another predicate and specific tense. While example (6) is equivalent to (4) with respect to the embedded predicate, the matrix predicate and tense in (7) render the embedded predicate hypothetical.

(6) *Reuters reported that Apple merged with Alphabet.*
(7) *Reuters will report that Apple merged with Alphabet.*

TimeML

EVENT, TIMEX3, SIGNAL, LINK

TIE-ML

- **EventID**

Events in the TIE-ML schema are individual predicates that are usually clauses. Each clause or independent predicate is given a numerical event identifier (eventid), shown in Figure 1, that serves both to mark relationships between events, as well as track the presentation order of events in text.

```
<s> <c eventid="1">
    Danny watched the movie
    </c>
    <c eventid="2">
    and ate popcorn
    </c>. </s>
<s> <c eventid="3">
    Josh brought the pizza
    </c>. </s>
```

- **Tense, Perfect, Progressive**

For each event, TIE-ML provides the possibility for the tense of the predicate, as well as the presence of perfect and progressive aspect to be explicitly annotated using the *tense* attribute, and Boolean *perfect* and *progressive* attributes as shown

```
<s> <c tense="PAST" perfect= "TRUE"
       progressive="TRUE">
    The patient had been experiencing
    stomach pain
</c>. </s>
```

- **Reference**

To be able to capture the concrete reference time for an event, we provide a designated attribute to capture concrete date or time point expressions that anchor R on the real time axis.

```
<s> <c reference="264 BC">
    The First Punic War broke out on the
    island of Sicily in 264 BC.
</c> </s>
```

- **Reichenbach E, R, S**

We assign positive and negative numbers to sequence time variables, as well as encode tense directly into the value where negative values equal the past, positive the present, and 0 the present.

```
<s> <c E="-1" R="-1" S="0">
    Danny watched the movie.
</c> </s>

<s> <c E="-2" R="-1" S="0">
    Josh had watched the movie.
</c> </s>
```

- **Timeslot**

To capture the presentation time and the relative timeslot association of events, the TIE-ML schema provides a timeslot attribute representing in its value the relative ordering of events along the time axis.

```
<s> <c eventid="1" timeslot="2">
    Prepare the vegetables </c>
    <c eventid="2" timeslot="1">
       after starting the oven
</c>. </s>
```

CROSS-LINGUISTIC DATA

Our aim is to develop cross linguistics data sets

German

```
<s> <c eventid="1" tense="FUT"
       reference="morgen"
       E="1" R="1" S="0">
    Morgen werde ich ein Buch lesen
    <!-- Tomorrow I will read a book -->
</c>. </s>
```

Arabic

```
<s> <c eventid="1" timeslot="2"
       tense="PAST" E="1" R="1" S="2">
التقى أحمد بصديقه        </c>
    <c eventid="2" timeslot="1"
       tense="PAST" E="1" R="1" S="2">
بعد أن زار والده الثلاثاء الماضي
<!-- Ahemd met with his friend after he
 visited his father last Tuesday  -->
</c>. </s>
```

Hebrew

```
<s> <c eventid="1" timeslot="2"
       tense="PAST" E="1" R="1" S="2">
נתן נפגש עם רבקיה          </c>
    <c eventid="2" timeslot="1"
       tense="PAST" E="1" R="1" S="2">
לאחר ששיים לעבוד אחר הצהריים
<!-- Nathan met with Rebecca after he
finished working this afternoon -->
</c>. </s>
```

FUTURE WORK

Expansion of the annotation scheme to include Lexical Semantics:

(8) John is reading Harry Potter.

In (8), the reading event can be interpreted as though John is in the process of reading Harry Potter but not currently reading at the speech time. Or it can be the case John is indeed reading at the speech time.

i.e. An event can be comprised of smaller iterative events.

This also relates to capturing and annotating sub events:

(9) Event 1: *The students went to the fair on Monday.*
(10) Event 2: *In the morning the students saw the animals.*
(11) Event 3: *In the afternoon they got cotton candy.*
(12) Event 4: *In the evening they ate an elephant ear..*

Mark typical event durations in order to identify abnormalities.

(13) *Geoff ate his dinner in 8 hours.*

CONCLUSIONS

TIE-ML is more efficient than TIME-ML.Annotations of corpora using TimeML can be mapped to TIE-ML with a loss, and TIE-ML annotations can be fully mapped to TimeML with certain under-specification.

https://github.com/dcavar/tieml

REFERENCES

J. Pustejovsky, et al. TimeML: Robust Specification of Event and Temporal Expressions in Text. New Directions in Question Answering, 2003, pp. 28-34.
D. Cavar. et al. Temporal Information and Event Markup Language: TIE-ML Markup Process and Schema Version 1.0. Proceedings of SEMAPRO 2021, Barcelona, Spain.
H. Reicehenbach. Elements of Symbolic Logic. The Free Press, New York, 1947.