4. REPRESENTATION OF EVENTS AND TIMES IN NATURAL LANGUAGE TEXT

A formal language to represent events and temporal relations in natural language text facilitates reasoning to solve many tasks in AI. The annotation schemes such as TIDES_TIMEX2 and TIMEML (Temporal markup Language) are popularly used to represent various features of events and times available in the natural language text. TIMEML (James Pustejovsky et al., 2003) is a rich markup language that can not only capture temporal characteristics of events available in natural language text documents but also facilitates temporal linking among them.

4.1 TIMEML

Temporal mark-up language TIMEML consists of a collection of tags inserted into a text, uses the representational principles of XML markup to annotate the analysis of core elements (i.e. events and temporal expressions) in a temporal framework. There are four major tags specified in TIMEML namely <EVENT>, <TIMEX3>, <SIGNAL>, and <LINK> to represent events, times and their temporal relation that are available in the natural language text.
4.1.1 EVENT Representation

<EVENT> tag in TIMEML is used to represent different class of events that occur in a natural language text. The specification of EVENT tag is as follows (The content in {} gives the attribute definition).

<attributes::= eid, class >

\[
\begin{align*}
\text{eid} & ::= \text{ID} \quad \{\text{eid} ::= \text{EventID where EventID} ::= e<\text{integer}>\} \\
\text{class} & ::= \text{’OCCURRENCE’} \mid \text{’PERCEPTION’} \mid \text{’REPORTING’} \mid \\
& \mid \text{’ASPECTUAL’} \mid \text{’STATE’} \mid \text{’I\_STATE’} \mid \text{’I\_ACTION’}
\end{align*}
\]

The set of attributes assigned to an event comprises of \textit{eid}, and \textit{class}. \textit{Event ID number (eid)} uniquely identifies an event in the text. TIMEML captures several different types of events that are stored in a \textit{class} attribute. A TIMEML event will fit into one of these categories such as \textit{OCCURRENCE} (describe something that happens or occurs in the world e.g. erupted), \textit{PERCEPTION} (includes verbs as events which express physical perception of other event, e.g. listen), \textit{REPORTING} (: events that describe action of a person or organization e.g. say), \textit{ASPECTUAL} (events focus on different facets of event history for e.g. stop, finish), \textit{STATE}: (describes circumstances in which something obtains or holds true e.g., injuries), \textit{I\_STATE} (includes states that refer to alternative or possible worlds, for e.g., believe), \textit{I\_ACTION}: (describes an action or situation, e.g., attempt).
For example

(1) Mr core called on Israel to continue taking risks for peace

   Mr Core
   <EVENT eid="e1" class="OCCURRENCE">
      called
   </EVENT>
   on Israel to
   <EVENT eid="e2" class="ASPECTUAL">
      continue
   </EVENT>
   taking risks for peace.

(2) Nobody believe this anymore

   Nobody
   <EVENT eid="e1" class="I_STATE">
      believe
   </EVENT>
   this anymore

(3) The entire world will see the images of pope in Cuba

   The entire world will
   <EVENT eid="e1" class="PERCEPTION">
      see
   </EVENT>
   the images of pope in Cuba.

Representation of events which belongs to different classes is shown in the examples (1), (2) and (3). For instance event “continue” in (1) belongs to the class of Aspectual events while event “called” belongs to
an Occurrence event. The representation of other event classes are shown in (2) and (3).

4.1.2 TIME Representation

An important first step towards automatic extraction and representation of temporal information is to analyze and identify different time expressions in natural language text. These temporal expressions denote fully specified point in time (explicit) and underspecified point in time (implicit). Explicit temporal information among events is detected by identification of temporal expressions that refer to calendar dates, times of day, durations etc, which refer to some area on timeline. <TIMEX3> tag in TIMEML is used to mark these explicit temporal expressions such as (Monday, yesterday, currently, while, today etc). The specification of TIMEX3 is defined as follows.

attributes ::= tid type

[functionInDocument][beginPoint][endPoint][quant][freq]

temporalFunction][value | valueFromFunction][mod] [anchorTimeID | anchorEventID]

tid ::= ID {tid ::= TimeID where TimeID ::= t<integer>}

type ::= 'DATE'| 'TIME'| 'DURATION'| 'SET'

functionInDocument ::= 'CREATION_TIME'| 'EXPIRATION_TIME'

| MODIFICATION_TIME'| 'PUBLICATION_TI
An attribute \textit{tid} identifies temporal expression uniquely. The attribute \textit{type} indicate the type of temporal expression retrieved such as DATE (e.g., \textit{September 13 2009}), TIME (e.g., \textit{five to eight}), DURATION (e.g., \textit{2 months}), SET (e.g., \textit{twice a month}). \textit{Value} and \textit{mod} are equivalent to VAL and MOD attributes of TIMEX2. \textit{BeginPoint} and \textit{endpoint}
attributes are used when duration is anchored by temporal expression. *Quant* and *freq* attributes are used when a temporal expression is of type SET. *Temporal function* is a binary attribute which expresses the information that the value of the temporal expression needs to be determined via evaluation of temporal function. *AnchorTimeID* is used to provide reference point to another time expression which have a functional interpretation, e.g “last week”. *AnchorEventID* is similar in function, and is used to point to an EVENT to serve as a reference point. The attribute *functionInDocument* in TIMEX3 provide a temporal anchor for other temporal expressions in the document which includes various values of time.

*For example*

(4) *Savin Corp, reported a third-quarter net loss of $35.2 million*

*Savin Corp*

```
<SPEC id="e1" class="REPORTING">
  reported
</SPEC>
```

*third-quarter*

```
<TIMEX3 tid="t1" TYPE="DATE">
  a
</TIMEX3>
```

*net*

```
<EVENT eid="e2" class="OCCURRENCE">
  loss
</EVENT>
```

*of $35.2 million*
(5) No more than 60 days

\[ \text{TIMEX3 tid="t1" type="DURATION" value="P60D" mod="EQUAL OR LESS"} \]

**No more than 60 days**

</TIMEX3>

In the example (4), (5) TIMEX3 tag identifies the type of temporal information with the help of attribute “type”. The temporal modifier that cannot be expressed either by proper value or via links or temporal functions is given by mod attribute. Period of 60 days is indicated “P60D” by the attribute value in <TIMEX3> tag.

### 4.1.3 Temporal Relations Representation

<TLink> tag in TIMEML provides a temporal relationship among the events in natural language text. The possible relations include the end points of the events which are represented using one of the 13 relations proposed in Allen (Allen 1983). Allen proposed an algebraic framework called Interval Algebra (IA) for qualitative reasoning with time intervals. The 13 atomic relations are: { before(p), meet(m), overlap (o), starts(s), finishes(f), during(d), equal(e), during-by(D), Overlapped-by(O), Started-by(S), Finished-by(F), Met-by(M), after(P))

<TLINK> has a fixed collection of 14 relation-types, each of which have a fixed interpretation. Among these, the 6 inverses relations are redundant. SIMULTANEOUS and IDENTITY is collapsed, since
IDENTITY is a subtype of SIMULTANEOUS. Similarly DURING and IS_INCLUDED are collapsed, since DURING is a subtype of IS_INCLUDED. IBEFORE (immediately before) corresponds to MEETS in Allen’s Interval calculus. Allen’s OVERLAP relation is not represented in TimeML. By using all these considerations TLINK relations collapse to a disjunctive classification of 6 temporal relations. TReIs= SIMULTANEOUS, IBEFORE, BEFORE, BEGINS, ENDS, INCLUDES). These 6 relations and their inverses map one-to-one to 12 of Allen’s 13 basic relations.

The specification of TLINK is as follows.

```
attributes ::= [lid] [origin] (eventInstanceID | timeID) [signalID] (relatedToEventInstance | relatedToTime) relType
```

![Figure 4.1 Allen’s Interval Relations](image-url)
\textbf{lid} ::= \text{ID} \{\text{lid} ::= \text{LinkID} \text{ where LinkID} ::= 1<\text{integer}>\}

\textbf{origin} ::= \text{CDATA}

\textbf{eventInstanceID} ::= \text{IDREF} \{\text{eventInstanceID} ::= \text{EventInstanceID}\}

\textbf{timeID} ::= \text{IDREF} \{\text{timeID} ::= \text{TimeID}\}

\textbf{signalID} ::= \text{IDREF} \{\text{signalID} ::= \text{SignalID}\}

\textbf{relatedToEventInstance} ::= \text{IDREF} \{\text{relatedToEventInstance} ::= \text{EventInstanceID}\}

\textbf{relatedToTime} ::= \text{IDREF} \{\text{relatedToTime} ::= \text{TimeID}\}

\textbf{relType} ::= 'BEFORE' | 'AFTER' | 'INCLUDES' | 'IS_INCLUDED' | 'DURING' | 'SIMULTANEOUS' | 'IAFTER' | 'IBEFORE' | 'IDENTITY' | 'BEGINS' | 'ENDS' | 'BEGUN_BY' | 'ENDED_BY'

The \textit{eventInstanceID} or the \textit{timeID} involved in temporal link is a obligatory attribute (one or the other of these needs to be present). \textit{Signal ID} is provided with the existence of signal in the text. \textit{relatedToEventInstance} or \textit{relatedToTime} is a ID of the entity that is being related to the event instance with \textit{ID=EventInstanceID} or time expression \textit{with ID=TimeID}. \textit{Reltype} is a temporal relation holding between events. In sentence (6), Qualitative interval information among the events is denoted by <TLINK> tag.
For example

(6) A company reported a rise in shares.

A company

\[
\text{<EVENT} eid="e1" \text{class="REPORTING" >
reported
</EVENT>}
\]

a

\[
\text{<EVENT} eid="e2" \text{class="OCCURRENCE" >
rise
</EVENT>}
\]

in shares

\[
\text{<TLINK} lid="l1" \text{relatedToEventInstance="ei2"}
\text{relType="BEFORE" eventInstanceID="ei1" />}
\]

4.2 Extension of TIMEML tags

The qualitative duration of events plays a crucial role to identify the overlap of event occurrences in the text. TIMEML tags can represent durations of events that are mentioned explicitly. None of the tags provides a mechanism to represent implicit duration of events. Therefore a novel method is proposed to represent implicit duration of events by extending \text{<EVENT>} and \text{<TLINK>} tags of TIMEML.

4.2.1 Extension of \text{<EVENT>} Tag

Extension of \text{<Event>} tag is contributed by introducing an additional attribute \textit{probabilistic duration (pdur)} in the event tag which represents implicit events duration. This attribute denote the estimated probability
duration of events. The duration probability calculated based on event occurrences for different durations at different points of time in corpus. The duration for which the event occurs for maximum number of times in the corpus is estimated to be the probable duration of event. This duration is tagged to “pdur” denoting it as implicit event duration. The specification of extended <EVENT> tag is as follows (The content in {} gives the attribute definition).

```
<attributes::= eid, class, pdur >

eid ::= ID {eid ::= EventID where EventID ::= e<integer>}

class ::= 'OCCURRENCE' | 'PERCEPTION' | 'REPORTING' | 'ASPECTUAL' | 'STATE' | 'I_STATE' | 'I_ACTION'

pdur::= event duration (probabilistic duration of events)
```

For example

(7) Revenue of Infosys shares rose by 33%.

Revenue of Infosys share

```
<Event eid="e1" class="OCCURRENCE" pdur="P3M"> rose
</EVENT>
```

by 33%

On applying extended <EVENT> tag to markup events in a given document, “rose” is marked as an occurrence of event with implicit
duration of events tagged automatically. In (7), the estimated probabilistic durations for “rose” event is extracted as “period of three months” represented as “P3M” tagged to “pdur” attribute of extended <EVENT> tag.

### 4.2.2 Extension of <TLINK> tag

In order to represent qualitative end-point and duration information an elegant framework INDU (Arun K. Pujari et.al. 1999) is proposed. INDU is a framework evolved by combining Allen’s Interval Algebra (IA)[James Allan 1983] with point algebra. The basic relations of interval algebra denote relative positions satisfied by two intervals wherein point algebra satisfies relative duration of two intervals. INDU has 25 basic relations that sufficiently represent qualitative information about interval and durations that relate any pair of events. These relations represent relative positions of interval duration information on a timeline. INDU is a framework which integrates both interval and duration information. The relations are denoted as

\[ E = \{ p^\prec, p^\succ, p^=', P^\prec, P^\succ, P^=', p^\prec, p^\succ, m^\prec, m^\succ, m^=', M^\prec, M^\succ, M^=', o^\prec, o^\succ, o^=', o^\prec, o^\succ, O^\prec, O^\succ, O^=', S^\prec, S^\succ, S^=', d^\prec, D^\prec, d^\succ, D^\succ, f^\prec, F^\prec, e^\} \]

<table>
<thead>
<tr>
<th>Basic Interval Relations</th>
<th>Converse relations</th>
<th>Schematic representation</th>
<th>Endpoint relations and relative durations of events</th>
</tr>
</thead>
<tbody>
<tr>
<td>p^\prec</td>
<td>P^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d&lt;Y^d</td>
</tr>
<tr>
<td>p^\succ</td>
<td>P^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>p^'</td>
<td>P'</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>p'^</td>
<td>P'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>m^\prec</td>
<td>M^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>m^\succ</td>
<td>M^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>m'^</td>
<td>M'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>m'^</td>
<td>M'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>M^\prec</td>
<td>o^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>M^\succ</td>
<td>o^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>M'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>M'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\prec</td>
<td>o^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\succ</td>
<td>o^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\prec</td>
<td>o^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\succ</td>
<td>o^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\prec</td>
<td>o^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\succ</td>
<td>o^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\prec</td>
<td>o^\prec</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o^\succ</td>
<td>o^\succ</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>o'^</td>
<td>o'^</td>
<td>xxx</td>
<td>X^b&lt;Y^b, X^b&lt;Y^c, X^c&lt;Y^b, X^d=Y^d</td>
</tr>
<tr>
<td>p&gt;</td>
<td>P&lt;</td>
<td>xxxx</td>
<td>XXXXX</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>m&lt;</td>
<td>M&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>m&lt;</td>
<td>M&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>m&lt;</td>
<td>M&lt;</td>
<td>xxxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>o&lt;</td>
<td>O&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>o&lt;</td>
<td>O&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>o&gt;</td>
<td>O&lt;</td>
<td>xxxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>s&lt;</td>
<td>S&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>f&lt;</td>
<td>F&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>d&lt;</td>
<td>D&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
<tr>
<td>e&lt;</td>
<td>E&lt;</td>
<td>xxx</td>
<td>yyyy</td>
</tr>
</tbody>
</table>

Xb, Xc, Xd denote begin point, end point and duration respectively for any interval X. The Table 4.1 gives the schematic of end point relations and the event durations.

**Table 4.1: 25 basic INDU Relations**

An extension of `<TLINK>` tag is contributed by incorporating a new attribute *relative duration* (rdur) in `<TLINK>` tag. The extended attribute represent the relative duration among the events based on estimated probable durations tagged to respective events. A binary relation `<'LESS' | 'EQUAL' | 'GREATER' > is tagged among events by comparing “pdur” attribute of extended `<EVENT>` tag. The specification of extended `<TLINK>` tag is represented as follows (The content in { } gives the attribute definition).
attributes ::= [lid] [origin] (eventInstanceID | timeID) [signalID]

(relatedToEventInstance | relatedToTime) relType rdur

lid ::= ID {lid ::= LinkID where LinkID ::= l<integer>}

origin ::= CDATA

eventInstanceID ::= IDREF {eventInstanceID ::= EventInstanceID}

timeID ::= IDREF {timeID ::= TimeID}

signalID ::= IDREF {signalID ::= SignalID}

relatedToEventInstance ::= IDREF {relatedToEventInstance ::= EventInstanceID}

relatedToTime ::= IDREF {relatedToTime ::= TimeID}

relType ::= 'BEFORE' | 'AFTER' | 'INCLUDES' | 'IS_INCLUDED' |

DURING | 'SIMULTANEOUS' | 'IAFTER' | 'IBEFORE' |

'IDENTITY' | 'BEGINS' | 'ENDS' | 'BEGUN_BY' |

'ENDED_BY'

rdur ::= 'LESS' | 'EQUAL' | 'GREATER'

For example

(8) A company reported a rise in shares

A company reported a rise in shares

A company

<EVENT eid="e1" class="REPORTING" pdur=P1D>

reported
</EVENT>

a

<EVENT eid="e2" class="OCCURRENCE" pdur="P3M>

rise
</EVENT>
The extended <EVENT> tag represent the probability duration of events which are further compared to obtain the relative duration among the events. In (8) event “Reported” and “rise” occurred has been tagged with its estimated durations in the extended event tag. This information is used by extended <TLINK> tag to represent relative duration (i.e. GREATER) among the identified events.

4.3 **Summary**

This chapter describes TIMEML tags to represent events, times and temporal relations among the events. <EVENT> tag in TIMEML is used to tag different class of events while <TIMEX3> tags explicit temporal information. Extension of various tags in TIMEML, to represent implicit event durations is proposed. <EVENT> tag is extended by introducing a new attribute “pdur” which represent probabilistic duration of event that is referred as implicit duration of events.

<TLINK> tag in TIMEML represents temporal relations among the events. <TLINK> tag is extended to represent relative duration among the events. An additional attribute “rdur” is added to <TLINK> tag to represent relative duration (Less, Equal, Greater) based on the “pdur” values annotated in extended <EVENT> tag.