

# Time and the Semantic Web

James Pustejovsky  
Computer Science Department  
Brandeis University  
Waltham, MA, 02254 USA  
jamesp@cs.brandeis.edu

## Abstract

*In this paper, we discuss the role that temporal information plays in natural language text, specifically in the context of enriching the semantics of web texts and web interactions. We present a language, TimeML, which attempts to capture the richness of temporal and event related information in language, while demonstrating how it can play an important part in the development of more robust semantic ontologies. Specifically, we propose to demonstrate how a TimeML markup of text is interpreted within the DAML-Time Ontology and time framework of Hobbs (2002).*

## 1. Introduction

In order for the semantic web to be adopted as an environment and platform for commerce, web delivery, and communication, it will require a robust specification of temporal and event information. For example, someone who does a Web search trying to find a place to buy a book needed before next Tuesday may or may not be able to use an online bookstore that promises delivery within five business days. Someone doing a genealogical search may want to specify that the birthdate of a person is between 15 and 45 years before a known marriage date. In response to this need, in connection with the DARPA-sponsored DAML program, an ontology of temporal concepts has been developed, DAML-Time, that covers the basic topological temporal relations on instants and intervals, measures of duration, and the clock and calendar (Hobbs, 2002). This will enable the statement of and reasoning about the temporal aspects of the contents and capabilities of web resources and of the needs of users.

But since most of the information on the Web is in natural language, there is little chance that it will ever be marked up for semantic retrieval if that has to be done by hand. Natural language programs will have to process the contents of web pages to produce annotations. Special semantic web

tag editors might facilitate some of the markup task, and it is true that remarkable progress has been made in the last decade in the use of statistical techniques for analyzing text. However, these techniques for the most part depend on having large amounts of annotated data, and annotations require an annotation scheme. It is for this reason that the ARDA-funded AQUAINT program has sponsored the development of an mark-up language TimeML for temporal information in texts (Pustejovsky et al, 2005). These annotations are most relevant to the Semantic Web enterprise if the annotation scheme meshes well with temporal ontologies used in the Semantic Web. The aim of this paper is to outline just such a mapping.

Reasoning about time is one of the most important aspects of commonsense reasoning. Linking a formal theory for time with an annotation scheme aimed at extracting rich temporal information from natural language text is significant for at least two reasons. It will allow us to use the multitude of temporal facts expressed in text as the ground propositions in a system for reasoning about temporal relations. It will also constitute a forcing function for developing the coverage of a temporal reasoning system, as we encounter phenomena not normally covered by such systems, such as complex descriptions of temporal aggregates.

The automatic recognition of temporal and event expressions in natural language text has recently become an area of intense research in computational linguistics and Artificial Intelligence. The importance of temporal awareness to question answering systems has become more obvious as current systems strive to move beyond keyword and simple named entity extraction. Named entity recognition (Chinchor et al, 1999) has moved the fields of information retrieval and information exploitation closer to access by content, by allowing some identification of names, locations, and products in texts. Beyond such metadata tags, however, there is only a limited ability at marking up text for real content. One major problem that has not been solved is the recognition of events and their temporal anchorings.

## 2 The TimeML Specification

Unlike most previous attempts at event and temporal specification, TimeML separates the representation of event and temporal expressions from the anchoring or ordering dependencies that may exist in a given text. There are four major data structures that are specified in TimeML (Ingria and Pustejovsky, 2002, Pustejovsky et al., 2002): **EVENT**, **TIMEX3**, **SIGNAL**, and **LINK**. These are described in some detail below. The features distinguishing TimeML from most previous attempts at event and time annotation are summarized below:

1. Extends the TIMEX2 annotation attributes;
2. Introduces **Temporal Functions** to allow intensionally specified expressions: *three years ago*, *last month*
3. Identifies signals determining interpretation of temporal expressions; *for*, *during*, *on*, *at*, *before*, *after*, *while*.
4. Identifies all classes of event expressions;
  - (a) Tensed verbs: *has left*, *was captured*, *will resign*;
  - (b) Stative adjectives: *sunken*, *stalled*, *on board*;
  - (c) Event nominals: *merger*, *Military Operation*;
5. Creates dependencies between events and times:
  - (a) Anchoring: *John left on Monday*.
  - (b) Orderings: *The party happened after midnight*.
  - (c) Embedding: *John said Mary left*.

TimeML considers “events” (and the corresponding tag <EVENT>) a cover term for situations that *happen* or *occur*. Events can be punctual or last for a period of time. We also consider as events those predicates describing *states* or *circumstances* in which something obtains or holds true. Not all stative predicates are marked up, however, as only those states which participate in an opposition structure in a given text are marked up. Events are generally expressed by means of tensed or untensed verbs, nominalizations, adjectives, predicative clauses, or prepositional phrases. The specification of **EVENT** is shown below:

```
attributes ::= eid class tense aspect
eid ::= ID
{eid ::= EventID
EventID ::= e<integer>}
class ::= 'OCCURRENCE' | 'PERCEPTION' | 'REPORTING'
        | 'ASPECTUAL' | 'STATE' | 'I_STATE' |
        | 'I_ACTION' | 'MODAL'
tense ::= 'PAST' | 'PRESENT' | 'FUTURE' | 'NONE'
aspect ::= 'PROGRESSIVE' | 'PERFECTIVE' |
          | 'PERFECTIVE_PROGRESSIVE' | 'NONE'
```

Examples of each of these event types are: Occurrence: *die*, *crash*, *build*, *sell*; State: *on board*, *kidnapped*, *love*; Reporting: *say*, *report*, *announce*; I-Action: *attempt*, *try*, *promise*, *offer*; I-State: *believe*, *intend*, *want*; Aspectual: *begin*, *finish*, *stop*, *continue*; Perception: *See*, *hear*, *watch*, *feel*.

The TIMEX3 tag is used to mark up explicit temporal expressions, such as times, dates, durations, etc. It is modelled on both Setzer’s (2001) TIMEX tag, as well as the TIDES (Ferro, et al. (2002)) TIMEX2 tag. There are types of TIMEX3 expressions:

- (a) Fully Specified Temporal Expressions, *June 11, 1989*, *Summer, 2002*;
- (b) Underspecified Temporal Expressions, *Monday*, *Next month*, *Last year*, *Two days ago*;
- (c) Durations, *Three months*, *Two years*.

The treatment of temporal functions in TimeML allows any time-value dependent algorithms to delay the computation of the actual (ISO) value of the expression. The following informal paraphrase of some examples illustrates this point, where DCT is the Document Creation Time of the article.

1. *last week* = (predecessor (week DCT)) : That is, we start with a temporal anchor, in this case, the DCT, coerce it to a week, than find the week preceding it.
2. *last Thursday* = (thursday (predecessor (week DCT))) : Similar to the preceding expression, except that we pick out the day named ‘thursday’ in the predecessor week.
3. *the week before last* = (predecessor (predecessor (week DCT))) : Also similar to the first expression, except that we go back two weeks.
4. *next week* = (successor (week DCT)) : The dual of the first expression: we start with the same coercion, but go forward instead of back.

A set of **LINK** tags encode the various relations that exist between the temporal elements of a document. There are three types of link tags.

- **TLINK**: A TLINK or Temporal Link represents the temporal relationship holding between events or between an event and a time, and establishes a link between the involved entities making explicit if they are: simultaneous, before, after, including, holds, beginning, and ending.
- **SLINK**: An SLINK or Subordination Link is used for contexts introducing relations between two events, or an event and a signal. SLINKs are of one of the following sorts: Modal, Factive, Counter-factive, Evidential, Negative evidential, and Negative.
- **ALINK**: An ALINK or Aspectual Link represents the relationship between an aspectual event and its argument event. Examples of the aspectual relations to be encoded are: initiation, culmination, termination, continuation.

To illustrate the function of these links, consider the sentence below: *John left 2 days before the attack*. A TLINK orders the two events mentioned in the sentence, with a magnitude denoted by the temporal expression.

```
<TLINK eventInstanceID="ei1" signalID="s1"
relatedToEvent="ei2"
relType="BEFORE" magnitude="t1"/>
```

For a modally subordinating predicate such as *want*, an SLINK is introduced, as shown below. *Bill wants to teach on Monday*.

```
Bill
<EVENT eid="e1" class="I_STATE" tense="PRESENT"
aspect="NONE"> wants </EVENT>
<MAKEINSTANCE eiid="ei1" eventID="e1"/>
<SLINK eventInstanceID="ei1" signalID="s1"
subordinatedEvent="e2" relType="MODAL"/>
<SIGNAL sid="s1"> to </SIGNAL>
<EVENT eid="e2" class="OCCURRENCE" tense="NONE"
aspect="NONE"> teach </EVENT>
<MAKEINSTANCE eiid="ei2" eventID="e2"/>
<SIGNAL sid="s2"> on </SIGNAL>
<TIMEX3 tid="t1" type="DATE" temporalFunction="true"
value="XXXX-WXX-1"> Monday </TIMEX3>
<TLINK eventInstanceID="ei2" relatedToTime="t1"
relType="IS_INCLUDED"/>
```

In a similar fashion, ALINKs are introduced by aspectual predicates. It should be pointed out that an ALINK captures aspectual phase information associated with an event, and is logically (temporally) distinct from the BEGIN relType in the TLINK above. For example, in a sentence *The boat began to sink when the torpedo hit it*, the torpedo hitting the boat is TLINKed to the sinking event through the relType BEGIN, while the beginning of the sinking is ALINKed through the phase relType INITIATE.

In order to provide an interpretation of the TimeML specification of event and temporal expressions described above, we will adopt the DAML-Time ontology for time as a model.

### 3 DAML-Time

DAML-Time is an ontology of temporal concepts, for describing the temporal content of Web pages and the temporal properties of Web services. Its development is being informed by temporal ontologies developed at a number of sites and is intended to capture the essential features of all of them and make them and their associated resources easily available to a large group of Web developers and users. Here I will adopt the the ontology as outlined in Hobbs and Pustejovsky (2003) and developed in Hobbs and Pan (2004). There it is argued that there are two subclasses of temporal-entity: *instant* and *interval*.

$$\begin{aligned} \text{instant}(t) &\supset \text{temporal-entity}(t) \\ \text{interval}(T) &\supset \text{temporal-entity}(T) \end{aligned}$$

*start-of* and *end-of* are functions from temporal entities to instants.

$$\begin{aligned} \text{temporal-entity}(T) &\supset \text{instant}(\text{start-of}(T)) \\ \text{temporal-entity}(T) &\supset \text{instant}(\text{end-of}(T)) \end{aligned}$$

*inside* is a relation between an instant and an interval.

$$\text{inside}(t, T) \supset \text{instant}(t) \wedge \text{interval}(T)$$

*interval-between* is a relation among a temporal entity and two instants.

$$\begin{aligned} \text{interval-between}(T, t_1, t_2) \\ \supset \text{temporal-entity}(T) \wedge \text{instant}(t_1) \\ \wedge \text{instant}(t_2) \end{aligned}$$

Given these definitions, a *before* relation on temporal entities is defined, which gives directionality to time. If temporal entity  $T_1$  is before temporal entity  $T_2$ , then the end of  $T_1$  is before the start of  $T_2$ . Thus, before can be considered to be basic to instants and derived for intervals.

$$\begin{aligned} \text{before}(T_1, T_2) \\ \equiv \text{before}(\text{end-of}(T_1), \text{start-of}(T_2)) \end{aligned}$$

The *before* relation is anti-symmetric and transitive, and the relation *after* is defined in terms of *before*.

$$\begin{aligned} \text{before}(T_1, T_2) &\supset \neg \text{before}(T_2, T_1) \\ \text{before}(T_1, T_2) \wedge \text{before}(T_2, T_3) \\ &\supset \text{before}(T_1, T_3) \end{aligned}$$

Next, Hobbs and Pustejovsky (2003) assumes that the relations between intervals defined in Allen's temporal interval calculus (Allen and Kautz, 1997) can be defined in a straightforward fashion in terms of *before* and identity on the start and end points. We illustrate this with the relations *int – meets* and *int – finishes*.

$$\begin{aligned} \text{interval}(T_1) \wedge \text{interval}(T_2) \\ \supset [\text{int-meets}(T_1, T_2) \\ \equiv \text{end-of}(T_1) = \text{start-of}(T_2) \\ \text{interval}(T_1) \wedge \text{interval}(T_2) \\ \supset [\text{int-finishes}(T_1, T_2) \\ \equiv \text{before}(\text{start-of}(T_2), \text{start-of}(T_1)) \\ \wedge \text{end-of}(T_1) = \text{end-of}(T_2)] \end{aligned}$$

### 4 Interpreting TimeML in DAML-Time

Following Hobbs and Pustejovsky (2003), the time ontology links to other objects in the domain through the predicates *at-time* and *during*. As in TimeML, the term “eventuality” will be used to cover events, states, processes, propositions, states of affairs, and anything else that can be located with respect to time. The possible natures of eventualities would be spelled out in the event ontologies. The

predicate *at-time* relates an eventuality to an instant, and is intended to say that the eventuality holds, obtains, or is taking place at that time:  $at-time(e, t) \supset instant(t)$ .

The predicate *during* relates an eventuality to an interval, and is intended to say that the eventuality holds, obtains, or is taking place during that interval.

$$during(e, T) \supset interval(T)$$

If an eventuality obtains during an interval, it obtains at every instant inside the interval:

$$during(e, T) \wedge inside(t, T) \supset at-time(e, t)$$

Whether a particular process is viewed as instantaneous or as occurring over an interval is a granularity decision that may vary according to the context of use, and is assumed to be provided by the event ontology.

The event ontology could extend temporal functions and predicates to apply to events in the obvious way, e.g.,

$$\begin{aligned} ev-start-of(e) &= t \\ &\equiv time-span-of(T, e) \wedge start-of(T) = t \end{aligned}$$

This would not be part of the time ontology, but would be consistent with it.

The intended interpretation of the TimeML specifications can now be provided in terms of the DAML-Time ontology. This allows one to move back and forth between the two representations as required for linguistic analysis and reasoning. The “tense” attribute of EVENT tags is expressed in terms of a *before* or equality relation, once the deictic anchor (“now”) is known.

The “value” in a TIMEX3 tag is a date, time, or duration in the ISO 8601 standard format, as extended by TIDES2. There is a straightforward mapping between *time-of* predication and date and time formats, and between *duration-of* predication and duration formats. The “XX” notation in the ISO standard corresponds to existentially quantified variables in the logic.

The “relType” values of the TLINK tag correspond roughly to the interval relations in Allen’s interval calculus and can be defined either in terms of these or in a similar manner. For example, an IAFTER relation can be defined as the interval during which one event occurs meeting the interval during which the other occurs. IDENTITY is not a temporal relation but is expressed as equality involving reified events.

The “relType” values of the ALINK tag can be defined similarly. An event that INITIATES another, for example, occurs at the start of the interval during which the other occurs. The ALINK tag conveys something more than just temporal information. As mentioned above, the beginning of a sinking of a boat is not the same event as the event of a torpedo hitting the boat, even if they are simultaneous. This distinction, however, is not part of a theory of time but

a theory of the structure of events. Similarly, the distinction between TERMINATES and CULMINATES is not a temporal one but something that would have to come from event theory.

The logical representation of the temporal information in the sentence *John left 2 days before the attack*. can now be read off the TimeML tags in the markup. Recall that the ordering between the events is encoded in a TLINK as:

```
<TLINK eventInstanceID="ei1" signalID="s1"
relatedToEvent="ei2" relType="BEFORE"
magnitude="t1"/>
```

Assume  $e_1$  is John’s leaving and  $e_2$  is the attack. If the TLINK relType is interpreted as *before*, and the magnitude attribute is associated with *duration*, then the DAML-Time representation of the temporal information from TimeML is as shown below.

$$\begin{aligned} at-time(e_1, t_1) \wedge at-time(e_2, t_2) \\ \wedge before(t_1, t_2) \quad \wedge \quad interval- \\ between(T, t_1, t_2) \\ \wedge duration(T, *Day*) = 2 \end{aligned}$$

This short note has presented TimeML as an interpreted markup language for use in the annotation of natural language text on the web. For details, see Pustejovsky et al (2005).

## References

- [1] Allen, J.F. (1984). Towards a general theory of action and time. *Artificial Intelligence* 23, pp. 123-154.
- [2] Allen, James F., and Henry A. Kautz. 1985. “A Model of Naive Temporal Reasoning”, *Formal Theories of the Commonsense World*, ed. by Jerry R. Hobbs and Robert C. Moore, Ablex Publishing Corp., pp. 251-268.
- [3] Allen, J.F. and G. Ferguson (1997). Actions and events in interval temporal logic. In Oliveiro Stock (ed.), *Spatial and Temporal Reasoning*, Kluwer Academic Publishers, pp. 205-245.
- [4] Ferro, Lisa, Inderjeet Mani, Beth Sundheim and George Wilson, 2001, TIDES Temporal Annotation Guidelines. Version 1.0.2 MITRE Technical Report, MTR 01W0000041.
- [5] Mani, Inderjeet and George Wilson. 2000, “Robust Temporal Processing of News”, *Proceedings of the ACL’2000 Conference*, 3-6 October 2000, Hong Kong.
- [6] Pustejovsky, James, Robert Gaizauskas, Roser Sauri, Andrea Setzer, Robert Ingria, 2002. “Annotation Guideline to TimeML 1.0”, available at <http://time2002.org>.