

# ISO-TimeML and its Applications

Kiyong Lee, Branimir Boguaraev, Harry Bunt, and James Pustejovsky

## Abstract

As is specified in ISO WD 24616-1 *Language Resource Management Semantic Annotation Framework -Part 1: Time and Events* (SemAf/Time), ISO-TimeML is an XML-based specification language for annotating temporal and event-related information in natural language texts. Texts or corpora with such annotation provide useful language resources for information extraction, summarization, question-answering or multilingual translation, especially related to time and events in texts such as news reports on the Web. The purpose of this presentation is to show how these applications may work in multilingual contexts, specifically with annotated English and Korean versions of news articles on the same topic that are published on the same date on the Internet.

Today's presentation focuses on specific ways of annotating temporal and event-related information in some small fragments of texts, while aiming at possible applications of such annotated texts. The extraction of information on temporal orderings of events that are introduced by three types of links, temporal T-Link, aspectual A-Link, and subordinating S-Link into texts, for instance, helps visualize all the related sub-events mentioned in a big event reported in a news article in a sequential format. Based on the date and time specification of ISO 8601 and its subsequent developments, Timex2 and Timex3 in TimeML, ISO-TimeML also provides annotated information which is necessary for WH-questions of three types: time and date (*when*), frequencies (*how often*), and duration (*how long*). Finally, applied to languages other than English, ISO-TimeML provides multilingual semantic resources for the summarization of information content in several different documents or texts and also for the translation of time and event-related information in those texts. It should be easily claimed that, since every well-formed sentence or nominal phrase is headed by a predicate that refers to an event and often with temporal modification, ISO-TimeML annotates most of the key temporally sensitive information in text.

The tags employed in ISO-TimeML are all intended to assist in the understanding of time so that questions and corpora can be modeled, leading to eventual question answering. To that end, TimeML used four different tag types. The TIMEX3 tag is used to capture all temporal expressions. The EVENT tag captures all temporal events. Functional words such as *at* and *from* are annotated with the SIGNAL tag. Finally, all relationships between the other tags are represented with the LINK tags: TLINK, SLINK, and ALINK.

Natural languages have many mechanisms for expressing the temporal properties of events. In this talk, we turn to considerations of how temporal information might be extracted from natural language texts. A temporal annotation language should also capture the kind of information that might be required by tools and algorithms for automatically annotating

temporal information. Such tools and algorithms include machine learning techniques which might use human-annotated documents as training data, as well as more rule-based techniques, which might exploit linguistic regularities to derive temporal relations. We envision a multi-step, layered, information extraction process in which distinct modules may be responsible for extracting different pieces of information and incrementally marking up a knowledge base so information can be retrieved from it. Because the temporal relations representing the anchoring and ordering of events are the ultimate goal of an extraction process, such a process requires a practical way to represent all the building blocks which might be used to determine