

Annotation Guideline

Reference

Leeuwenberg, A., & Moens, M. F. (2020). Towards Extracting Absolute Event Timelines from English Clinical Reports. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

This guideline

This guideline describes the annotation procedure for the temporal annotation scheme used in the referred article. The scheme's aim is to capture the temporal information in text that describes real world events: capturing the information needed to position all events mentioned in the text as precisely as possible on a calendar timeline. This guideline describes the scheme and how to use the annotation tool.

The Project

In this project we want to build an automatic computer model that can automatically read a medical record, and produce a timeline of the events that are described in the textual report. We will build such a model by using machine learning. To learn the model how to produce good timelines from the text, it needs a lot of examples. These examples of correct timelines is what you will be annotating.

The Annotation Task

As annotator, your task is to provide the most precise timeline of the events described in the medical report. You will soon notice that not for every event in the text the exact time is given in the text. This is not a problem, the scheme takes into account uncertainty, and you can indicate how certain you are about the timing of the events.

Start, duration, and end

To position each event on the timeline, we ask you to annotate two out of three types of information:

- The **start** time of the event,
- the **duration** of the event,
- and the **end** time of the event.

You only need to provide two out of these three options, as the third can always be automatically derived (this is done automatically by the tool). When annotating, you have to make sure your annotations are as precise as possible, so we ask you to annotate the information you are most certain about first! For example, in the following sentence:

*Pt had a 4-hour **surgery**, starting at 12:00.*

It would be best to annotate the start, and the duration (and not the end), as there is explicit information about the start and duration in the text.

Uncertainty about event timings

Because very often the exact timing of events is not given, each annotation is split into two parts

1. Annotating the most likely event timing
2. Annotating the temporal bounds, indicating what timings could also have been possible

The temporal bounds allow you to indicate uncertainty. For example, in the sentence:

*The patient was in a **coma** for several hours but woke up at 5 pm.*

There is explicit information about the duration, and the end of the coma. Although the end of the coma is given explicitly (5 pm). Probably the patient did not wake up exactly at 5 pm. So in this case, you would annotate the most likely end time as exactly 5 pm. However, it could have been somewhere between 4:50 pm and 5:10 pm. The duration of the coma is not given precisely. In this case you will have to annotate what you believe happened. So for example, most likely the duration was 3 hours. But it was at least 2 hours, and at most 5 hours. As you were not there, you don't know exactly what happened. You only have the report and your background knowledge. We assume that the amount of temporal information given in the text, or that can be interpreted by an informed reader (as yourself) is sufficient to construct a useful timeline.

Actuality of Events

In some cases it is not clear if an event happened at all. Or it may even be very clear that the event did not happen at all. However, in this annotation we do not include how likely it is that the event happened, but only when it (would have) happened. For example in the following sentence

*The patient will undergo an exercise tolerance test with viability evaluation to determine whether he would benefit from **revascularization**.*

It is not known if the patient will ever undergo revascularization. Nevertheless we can annotate temporal information on this revascularization for the case it would happen. The temporal bounds will just be quite wide, as there is much uncertainty. But surely the revascularization will not happen in the past. As another example

*She has not complained of any **fever**, **shaking**, or **chills**.*

Even though the fever, shaking and chills did not happen (probably). We can still attribute temporal information to the absence of these symptoms. As the doctor most likely asked the patient fairly recently in the past. A simple way to check, is there any temporal information present here, is by doing a thought experiment and try to imagine the event happening in the year 1629. If this seems unlikely, then you probably have some temporal intuition for when this event did take place, which you could annotate, and could be useful for the model to know about. Try to choose bounds of which you believe that there is at least a 90% chance that it happened within those bounds. Better safe than sorry.

The annotation process in short:

What is the most certain unannotated event?

What are the two most certain components from: start, duration, and end?

What is the most likely temporal value for each component?

What are the temporal bounds that fit at least 90% of the cases of each component?

12-hours or 24 hours?

The tool uses a 12 hour system (using AM and PM) to signify times of the day. If you are not completely familiar with this notation, please check the table on the right (especially the 12 AM, 12 PM distinction).

0:00 = 12:00 AM	12:00 = 12:00 PM
1:00 = 1:00 AM	13:00 = 1:00 PM
2:00 = 2:00 AM	14:00 = 2:00 PM
3:00 = 3:00 AM	15:00 = 3:00 PM
4:00 = 4:00 AM	16:00 = 4:00 PM
5:00 = 5:00 AM	17:00 = 5:00 PM
6:00 = 6:00 AM	18:00 = 6:00 PM
7:00 = 7:00 AM	19:00 = 7:00 PM
8:00 = 8:00 AM	20:00 = 8:00 PM
9:00 = 9:00 AM	21:00 = 9:00 PM
10:00 = 10:00 AM	22:00 = 10:00 PM
11:00 = 11:00 AM	23:00 = 11:00 PM

Keyboard shortcuts

You can copy timeline components from already annotated events to prevent redundant annotation work. We highly recommend you use this to save time. However, be careful to verify your copied values, as it also may introduce sloppiness. Sometimes event timings may seem similar at first glance, but appear to differ when taking a closer look. The keyboard shortcuts are:

Keypress S	starts at the same time as
Keypress D	has the same duration as
Keypress E	ends at the same time as
Keypress S and then E	starts at the end of
Keypress E and then S	ends at the start of
Spacebar	happens exactly at the same time as

Color Codes in the Tool

The events in the text are colored. The meanings of the colors are:

Orange: The event is **not annotated**.

Light blue: This is the **selected** event by the user for annotation.

Green: This event is **annotated**.

Purple/Pink: The event is annotated with a start/duration/end bound equal to the most likely timeline (purple). Or a duration bound lies very close to the most likely duration (pink). Both purple and pink show that you seem quite certain about the event's start/duration/end. This can happen if a lot of temporal information on the event is given **explicitly** in the text, but is generally quite rare. So, as a rule of thumb; if an event is purple/pink, there should be **explicit** temporal cues in the text about this event's timing.

Red: This event is annotated with bounds that do not include the most likely situation (i.e. this is a **wrong annotation!**).

So, generally in terms of color, when opening a new document, all events start off as orange, and after annotation most events should be green, and a few events could be purple or pink.