Annotation guidelines for time expressions in MeDESTO

Based on the guidelines developed for the [2012 i2b2 challenge](#)[1] and the [THYME corpus](#)[2].
Written by Sumithra Velupillai and Natalia Viani.

Introduction

This document contains the annotation guidelines for annotating time expressions in clinical text. This annotation task is developed within the framework of the MeDESTO project (Measuring Duration of untreated psychosis by Extraction of Symptom and Treatment Onset from mental health records using language technology). The ultimate goal of this project is to be able to automatically identify the onset of symptoms related to psychosis and schizophrenia based on the information that is documented in mental health records.

The text to be annotated comes from de-identified documents from the Clinical Record Interactive Search (CRIS) database[3], from a cohort of patients who have been diagnosed with schizophrenia. Each patient has one document, this document is the longest document found in the database within the first 3 months of referral to SLaM services (the South London and Maudsley National Health Service Foundation Trust). This is a choice made to minimize the number of documents that have to be read in order to find the relevant information about a patient. The longest document is likely to contain relevant information, but this might not always be the case.

The goal of the annotation is to mark up temporal information present in clinical text in order to enable reasoning and queries over the timeline of clinically relevant events for each patient. The first step to reach this goal is to identify expressions that indicate time information.

Temporal expressions include all expressions related to time, such as dates, times, frequencies, and durations.

Time expressions

Temporal expressions are phrases that contain time information. The types of temporal expressions that we need to mark include dates (like "June 9th" or "May 2006"), times ("5PM"), durations ("for 15 years", "since last year"), frequencies ("once daily"), and age-related information ("at the age of 20", "in high school").

You will need to identify all the time expressions in a clinical record. As an important remark, you should not annotate those time expressions that are <u>not related to the patient</u>. For example, in a sentence containing a reference to the "*Best Practice Guidance: Specifications for Adult Medium Secure Services (DH, 2007)*", you should not annotate the date "2007".

Note that due to the nature of clinical notes, the texts that we will be annotating will often contain typos, grammatical errors, sentence fragments, ad-hoc abbreviations and many other deviations from the standard written English, which you will need to take into account during the annotation.

1. Date

The majority of the time expressions will be dates. These can be calendar dates (such as {January 4}) and other verbal expressions which can be mapped to calendar dates either concretely (such as {next Friday}, {Yesterday}, or {at this time}), or in a more fuzzy sense ({lately}, {in the past}).

- {At this time}, we see no reason to discontinue the treatment.

- The form was signed and scanned on {December 29, 2009}.

- I would probably restart her furosemide {tomorrow}.

- She came in {a couple months ago}.

DATE is also used for very generic sorts of time expressions, where you may not be able to point at a specific day, week, or month on a calendar, but can still gesture at the overall timeline. So, for instance, expressions like "in the past", "lately", "in the future", or even "previously", would be marked as DATE:

- She has experienced heavy bleeding {in the past}.

- She complains that she's felt tired {lately}.

- {Recently}, she has had several episodes of syncope.

Other examples for DATE

- {now}, {at present}

- {in this period of time}

- {over the weekend}

- {today}, {yesterday}, {Monday}

- {in three months}

- {during the summer holiday}

- {shortly after the school holidays started} (similar to "summer holidays")

- Date of birth: {05/10/1990} (annotate only "05/10/1990")

- **Note:** dates as part of nurse notes etc. should be annotated as dates.

- **Note:** {at the time of consultation}, {the day} in "{the day} of his admission" (these are the only cases in which we actually annotate an expression containing an "event")

2. Time

TIME is used for specific time points within a day, for instance, {3:00PM} or {23:45}, and once again can be relative:

- The patient's MRI is scheduled for {5:30pm}.

- Following the patient's latest seizure, {20 minutes ago}, we are re-evaluating her medications.

- Patient did well {overnight}.

- Surgery will need to be completed by {2:45} to have the biopsy to the lab sooner.

Put differently, temporal expressions which give minute-by-minute or hour-by-hour detail are marked as TIME. Day-by-day (or larger) detail is marked with DATE.

Other examples for TIME

- {throughout the night}, {overnight}

- {in the morning}, {at night}

3. Duration

Sometimes, you'll be given a single temporal expression interpreted as reflecting a span of time, rather than a point. These are things like "for 24 hours" or "for 2 years", and are marked with the class DURATION.

- The patient continuously experienced nausea {for nearly two weeks}.

- {For the next 12 hours}, we will lower the patient's morphine drip and then we will re-evaluate his pain.

- {Since August}, she has not had any episodes.

- {During the last 12 months}, she's been nauseous frequently.

- {Since February 18th}, he hasn't seen his doctor.

- He has been doing this {for five years}.

Remember again that more abstract temporal expressions ("lately", "in the past", "in the future"), although representing loosely defined spans of time, are considered DATE rather than DURATION.

Finally, two dates can be used to construct a duration:

- {From May 1st to the 3rd}, she will refrain from eating solid food.

Other examples for DURATION

- {for a period of time}

- {until a year ago}

- {prognosis is 6 to 9 months}

- {from 2pm to 5pm}

- {over the past three days},{over the past 48 hours},{for the past 4 weeks},{in the last two weeks}, {over the last few years},{for the last seven months}, {in the last two days}, {over the last weeks/months}, {for the last two weeks}

- {in the preceding nights}

4. Frequency

The frequency of symptoms or medication usage is important information, and these are annotated with type "FREQUENCY".

- The patient was discharged to home on Amiodarone 200 mg {daily}.

- The patient was started on Propafenone 300 mg {t.i.d.}.

Other examples for type FREQUENCY:

- {nocte}, {OD}

- Complied with his {night} prescribed medication

- The patient needs to remain on a {15 minute} observation (this expression actually indicates that it is important to check on the patient every 15 minutes)

- {Three nights in a row}

5. Age-related

Clinical records often include time expressions that are related to the age of the patient. Some expressions are explicitly related to an age ({at the age of 15}), while others are defined in a more contextual way ({in high school}, {when he was a child}).

- {After his birth}

- {whilst at University}

- {During his first year} (first year of University)

- smoking {since his teens} until a year ago

- {since she was 8 years old}

- **Note:** for Age-related only, you can remove the patient focus

Additional refinements to guidelines

- If you find two time expressions with different types written one after the other (e.g. "{Friday} {at 5pm}") you should create two different annotations. For expressions like "Wed 8th May", you can create only one annotation (date).

  E.g. "for a month in 2012": {for a month} (duration) in {2012} (date)

- To help disambiguate between different time expression types, you should include prepositions in the marked expressions (e.g., "in two days" is a Date, while "for two days" is a Duration).

- "On admission" should be considered an event and **not be annotated**

- "before" "after" should **not be annotated**

- When as in "when he's had a meal" should **not be annotated**, but "when he was 10 years old" **should**

- "then", "later" should **not be annotated**

- "on a few occasions" should **not be annotated**

- "at times" should **not be annotated**

- "often", "rarely" should **not be annotated** – for frequencies, only annotate if they can be pinpointed somehow to a time unit (e.g. day, week etc.)

- **As a general guiding point: think of whether or not the expression could be placed on a timeline. If it can be represented with a point, it would be probably a date. If it is better represented with a span of time, it would be probably a duration.**

References

1. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association. 2013;20(5):806–813.

2. Styler IV WF, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, et al. Temporal annotation in the clinical domain. Transactions of the Association for Computational Linguistics. 2014;2:143.

3. Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. BMJ Open. 2016;6(3).