Guidelines for Time Expression Annotation and Normalisation

Based on the TimeML annotation guidelines (www.timeml.org) and adapted for the MTSamples (www.mtsamples.com/) collection.

Written by Natalia Viani and Sumithra Velupillai in June 2018

Introduction

This document contains the annotation guidelines for annotating time expressions in clinical text. The text to be annotated comes from MTSamples, a collection of Medical Transcription Sample reports for multiple clinical specialties, created for educational purposes and for working transcriptionists.

The goal of the annotation task 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 (e.g., "in June 2012", "yesterday").

In this annotation task, you will identify and normalise time expressions inside the text. *Normalising* a time expression means assigning a standardized value to it. For example, the date "1st December 2001" can be normalised as "2001-12-01". This step is necessary to refer each time expression to a specific point in time.

Time expressions

Time expressions are phrases that contain time information. The *types* of expressions that you need to mark include dates ("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").

In general, time expression normalisation follows the rules provided in the TimeML guidelines. Each time expression type is normalised as follows:

1) Date: YYYY-MM-DD

Examples
- {1st January 2009} --> 2009-01-01
- {06/07/89} --> 1989-07-06
- {January 2002} --> 2002-01

Special cases

- In the past --> PAST_REF
- At present --> PRESENT_REF
- Currently, ongoing --> PRESENT_REF
- Recently --> PAST_REF
- In the future --> FUTURE_REF

2) Time: YYYY-MM-DDThh:mm

Examples

- On {06/07/89} at {5PM} -->
    a. "5PM" normalised to "1989-07-06T17:00"
    b. "06/07/89" normalised to "1989-07-06"
- {Every day} at {1 o-clock} -->
    a. "1 o-clock normalised" to "XXXX-XX-XXT13:00"

Special cases: expressions related to a time of the day, but not to a specific day of the year

- at night = XXXX-XX-XXTNI
- morning (MO), afternoon (AF), daytime (DT), weekend (WE), evening (EV)
- "05 Jan 2018, 7AM: the patient slept through {the night}" --> 2018-01-04TNI

3) Duration: P+T(optional)+number+letter,
where the letter indicates a temporal granularity (e.g., day, month)

Granularity values: years (Y), months (M), days (D), weeks (W), hours (H), minutes (M), seconds (S)

- For years, months, days, weeks: P+number+letter
- For hours, minutes, seconds: PT+number+letter

Examples

- {for three days} --> P3D
- {for 2 years} --> P2Y
- {For three minutes} --> PT3M
- {for two and a half years} --> P2.5Y

Special cases: durations with prepositions or ranges

- until around 2004 --> UNT2004
- since 2005 --> SIN2005
- APRIL-JUNE 2014 --> (2014-04,2014-06)
- 01 JAN – 06 JAN --> (XXXX-01-01, XXXX-01-06)

- Pregnancy week: e.g., P33W, P5W3D

4) Frequency: normalised like durations (an attribute would be required to take into account the frequency aspect, not considered for now)

Examples
- {daily} --> P1D
- {three times a day}--> P1D (we would then use a "freq" attribute to mark "3": not used for now)
- {fortnightly} --> P2W
- {ON} --> PT1NI
- {MANE} or {OM} --> PT1MO
- {TARDE} --> PT1EV

5) Age-related time expressions: P+number+letter *or* A+number+letter
- P: when the TIMEX refers to the age of the patient
- A: when the TIMEX refers to a point in time described as the age of the patient

Examples
- {25 years old} = P25Y
- {aged 21 years} = A21Y
- {at the age of 8} = A8Y
- {in year 3} … {in year 11} (referred to school) = A8Y … A16Y
- {since she was 8} --> SINA8Y
- {in his mid 50s} --> A55Y
- {in his late 50s} --> A58Y

Special cases for age-related
- {when he was quite young}, {during childhood} = CHILD_REF
- {throughout school} = SCHOOL_REF
- {during University} = UNI_REF
- {since his teens} = SINTEENS_REF
- {in later life} = ADULT_REF

6) Other special cases

6.1) Weeks/weekends
- 2014-W45
  "in the past week", where 45 is the number of the week in the year, in this case: November 3-9, 2014
- 2014-W45-WE

"over the weekend", where 45 is the number of the week in the year, in this case: November 3-9, 2014

6.2) Relative expressions are the TIMEXes that require other time expressions or the document creation time (DCT) to be correctly normalised, e.g. "yesterday", "that year", "two days before", …

Relative expressions can refer to:
* DCT from structured field (not available for MTSamples)
* Last referenced date (if this date cannot be easily resolved, use a generic normalised value, e.g., "PAST_REF")
* Domain-specific dates (e.g., admission, discharge, operation, hospitalization)

Examples
- {three days ago} = DCT - P3D
- {three days back} = last referenced date - P3D
- {at that time} = last referenced date (might not be clearly written inside the text)
- {yesterday} = DCT – P1D
- {today} = DCT
- {postoperative day #2} = OP + P2D
- {during his hospitalization} = HOSP_REF
- {discharge day} = DIS
- {admission day} = ADM

6.3) Uncertain expressions
Use the X value for durations:
- {for a few days} = PXD
- {for some time} = PXM (use the "month" granularity)

Use a specific value for relative expressions:
- few, a couple, several = "3"
- many, a lot = "10"
- a day or two later = two days later (always pick the second one)

Examples
- {a few X ago} = DCT - P3X (with X granularity)
- {a few days ago} --> DCT – P3D
- {many X ago} = DCT - 10 X (with X granularity)
- {many years ago} --> DCT – P10Y
- {two to three months ago} (like three months ago) = DCT – P3M