Annotation Guidelines for Time Expression Normalization in MeDESTO

Written by Natalia Viani and Sumithra Velupillai

Introduction

In this annotation task, you will normalize the time expressions (TIMEXes) that have been found inside the text. *Normalizing* a time expression means assigning a standardized value to it. For example, the date "1st December 2001" can be normalized as "2001-12-01". This step is necessary to refer each time expression to a specific point in time.

In general, TIMEX normalization follows the rules provided in the TimeML guidelines [1]. Each TIMEX type is normalized 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, on-going --> 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" normalized to "1989-07-06T17:00"
    b. "06/07/89" normalized to "1989-07-06"
- {Every day} at {1 o-clock} -->
    a. "1 o-clock normalized" to "XXXX-XX-XXT13:00"

Special cases: TIMEXes 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 2015, 7AM: the patient slept through {the night}" --> 2015-01-04TNI

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)

4) Frequency: normalized 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
- {fortnightly} --> P2W
- {ON} --> PT1NI
- {MANE} or {OM} --> PT1MO
- {TARDE} --> PT1EV

In this annotation task, we also consider the following expressions:

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 normalized, e.g. "yesterday", "that year", "two days before", …

Relative expressions can refer to:
* DCT from structured field
* Last referenced date (if this date cannot be easily resolved, use a generic normalized value, e.g., "PAST_REF")

Examples

- {three days back} = last referenced date - 3D
- {at that time} = last referenced date (might not be clearly written inside the text)

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 - 3 X (with X granularity)
- {a few days ago} (with DCT = “2010-05-06”) --> “2010-05-03”
- {many X ago} = DCT - 10 X (with X granularity)
- {many years ago} (with DCT = “2010-05-06”) --> “2000-05-06”
- {a few weeks ago} = DCT – 20
- {two to three months ago} (like three months ago) = DCT - 3 months

References

[1] Pustejovsky J, Castano JM, Ingria R, Sauri R, Gaizauskas RJ, Setzer A, et al. TimeML: Robust specification of event and temporal expressions in text. New directions in question answering. 2003;3:28–34.