Streamlining Temporal Information Extraction: Integrating Rule-based Methods into MedspaCy for Clinical Application

Mengke Hu, PHD[1,2], Alec B. Chapman, MS[1,2], Patrick R. Alba, MS[1,2], Jianlin Shi, MD, PHD[1,2]
[1]VA Salt Lake City Health Care System, UT; [2]University of Utah, Salt Lake City, UT, USA

Introduction

- It is crucial to identify the timing of a clinical event in Electronic Health Records (EHRs).
- MedspaCy [1] is a package for clinical information extraction but has no temporal component for extracting temporal information.
- Current temporal extraction efforts exist but are not compatible with spaCy-styled tokenizer.
- In this work, we build from existing work and introduce a rule-based temporal entity extraction package [2] for medspaCy, which can adapt to any spaCy-styled tokenizers, as well as handle additional types of implicit expressions.

Methods

- Our temporal entity extraction package was implemented in Python using medspaCy. It can be directly employed in medspaCy NLP pipeline as shown in Figure 1.
- Our package consists of **two stages**:
  - Concept Tagging: temporal related tokens are labeled.
  - Temporal NER: temporal expressions can be distinguished by constraining the tokens around. One example is shown in Figure 2.
- Each temporal expression will be identified and classified into 4 categories shown in Table 1.
- NER Performance validation on the following datasets:
  1. VA shelter dataset
  2. N2C2 2012 temporality challenge dataset

TIMEX3 Categories	Examples
DATE	• At **{this time}**, we se no reason to continue the treatment. • He was admitted on **{December 29, 2001}**.
TIM	• The patient's MRI is scheduled at **{4:45pm}**. • **{20 minutes ago}**, we re-evaluated her medications.
DURATION	• The patient had fever for **{5 days}**. • **{During last month}**, she's been nauseous frequently.
FREQUENCY	• Patient has checked to ER roughly **{3 times a month}**. • Mirtazapine tablet by mouth **{every bedtime}**.

Table 1: Examples for TIMEX3 categories

Results

- Our package is released [3]: https://github.com/VINCI-AppliedNLP/temporal_expression_extraction
- On shelter-related temporal dataset, our package identified 90% of the temporal expressions. On N2C2 2012 temporality challenge, our system achieved 81% in F1 score.

Future Work

- To identify frequency expressions such as "2 episodes of homelessness", as well as some vague date expressions such as "within 3.5 weeks"
- Implementing temporal relationship detection in our system will be our next goal to extend in medspaCy

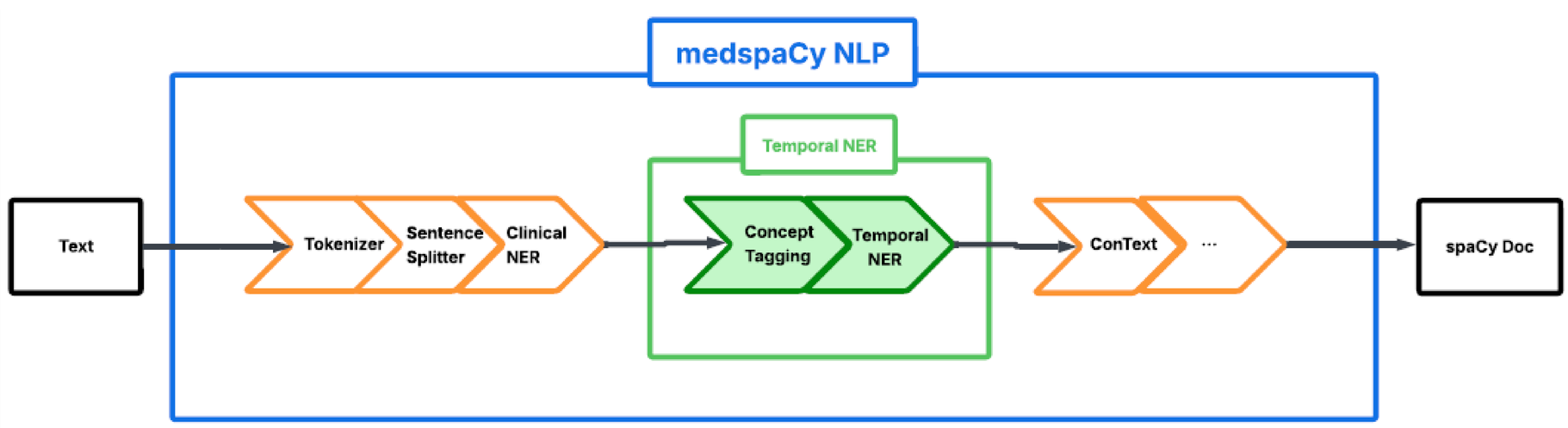

Figure 1: Temporal Component inserted in medspaCy pipeline

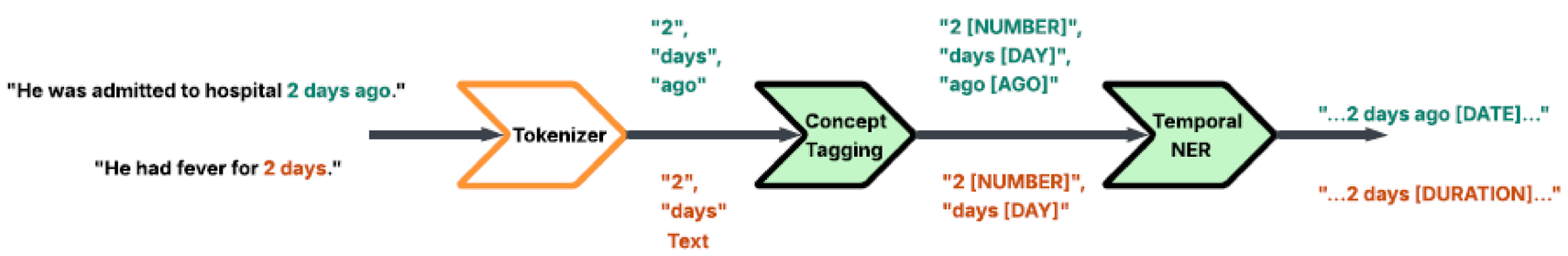

Figure 2: Two Ambiguous Examples in Distinct Stages of Temporal Component

Key References

- [1] H. Eyre, A. B. Chapman, K. S. Peterson, J. Shi, P. R. Alba, M. M. Jones, T. L. Box, S. L. DuVall and O. V. Patterson, "Launching into clinical space with medspaCy: a new clinical text processing toolkit in Python," *AMIA Annu Symp Proc,* pp. 438--447, 2021
- [2] M. Hu, "Temporal Entity Extraction," VINCI Applied NLP Team, 2025. [Online]. Available: https://github.com/VINCI-AppliedNLP/temporal_expression_extraction.
- [3] W. F. S. IV, S. Bethard, S. Finan, M. Palmer, S. Pradhan, P. C. d. Groen, B. Erickson, T. Miller, C. Lin, G. Savova and J. Pustejovsky, "Temporal Annotation in the Clinical Domain," *Trans Assoc Comput Linguist,* pp. 143--154, April 2014.
- [4] A. B. Chapman, A. Jones, A. T. Kelley, B. Jones, L. Gawron, A. E. Montgomery, T. Byrne, Y. Suo, J. Cook, W. Pettey, K. Peterson, M. Jones and R. Nelson, "ReHouSED: A novel measurement of Veteran housing stability using natural language processing," *Journal of Biomedical Informatics,* vol. 122, no. 1532-0464, p. 103903, 2021.