

MedTator: A Lightweight Interactive Multi-Document Annotation Tool

Huan He, PhD, Sunyang Fu, PhD, Liwei Wang MD, PhD, Andrew Wen MS,
Sijia Liu, PhD, and Hongfang Liu, PhD

Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA

Introduction: Natural language processing (NLP) techniques have been widely applied in healthcare practice and research¹. While high-quality annotated datasets are usually indispensable to develop NLP models and evaluate their performances, building such datasets is a time-consuming and labor-intensive process. There have been several comprehensive annotation tools developed in the past to facilitate the annotation process, but a lightweight and easy-to-use tool for annotating clinical narrative documents is still demanded. Therefore, we developed MedTator, a serverless web application providing an intuitive and interactive user interface for multi-document annotation.

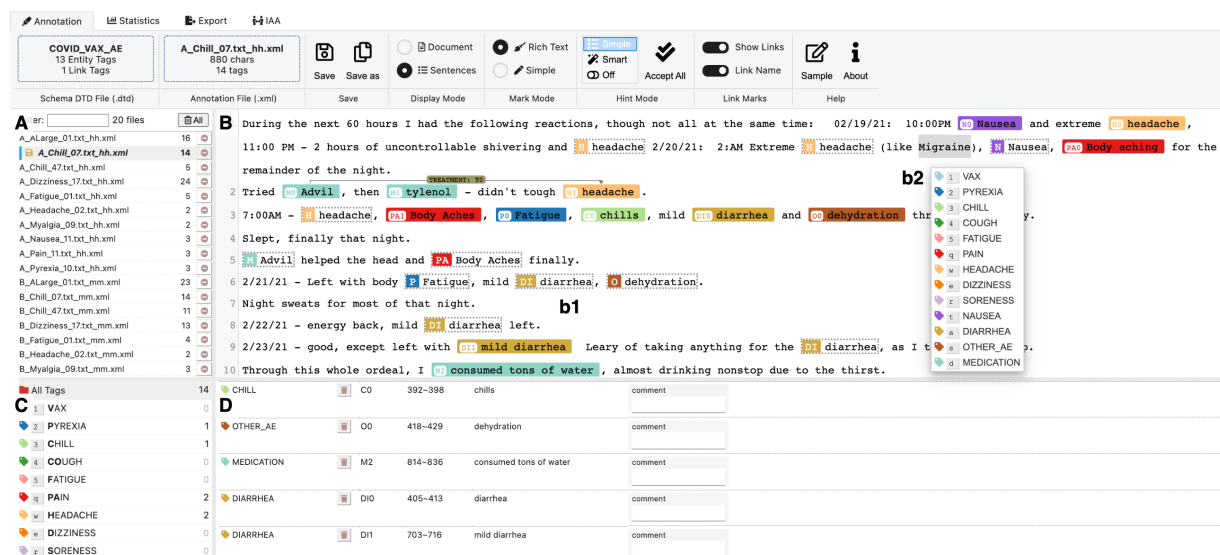


Figure 1. The screenshot of MedTator, including (A) document list view showing the summary of documents; (B) tag editor view showing visualized tags, (b1) annotation hints, and (b2) annotation menu for highlighted text; (C) concept list view showing all entity and relation concepts with the count of each concept annotated in a document; and (D) tag list view showing the detailed attributes of the annotated tags in the current document.

Tool Features: We have implemented the following features: (1) Serverless architecture. Due to the security and privacy concerns of clinical data, it is always required that data cannot leave specified computers or networks. Therefore, we adopted a serverless design which ensures all annotation tasks can be handled within local web browser. Moreover, unlike other web-based annotation tools, MedTator doesn't require any server at all. Users don't need to install any runtime environment on server or client. (2) A web-based user interface. To improve users' annotation experience, MedTator offers a series of high usability features to minimize the number of actions required to annotation multiple documents based on HTML5 and JavaScript techniques. As shown in Fig.1, we designed four coordinated views for several text annotation tasks, such as entity annotation, relation annotation, and document level annotation. (3) Machine-aided annotation support. We developed several functions to reduce user's workload during annotating. For example, MedTator can optionally show hints on potential words of interest based on the completed annotation; the statistics of the annotated data are updated in real-time to help the user track annotation progress. The source code of MedTator is available on GitHub with a live demo (<https://ohnlp.github.io/MedTator>).

Future work: MedTator is still under development. We plan to add more features to improve the annotation efficiency and support more output formats to export the annotated data for more downstream tasks such as building rule-based or machine learning based NLP systems.

References

1. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. Briefings in Bioinformatics. 2021 Jan 1;22(1):146–63.