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.

Annotation	Export 🙌 IAA							
COVID_VAX_AE 13 Entity Tags 1 Link Tags	A_Chill_07.txt_hh.xml 880 chars 14 tags	Save Save as	Document	 Rich Text Simple 	Simple Smart O Off Accept All	Show Links	Sample About	
Schema DTD File (.dtd)	Annotation File (.xml)	Save	Display Mode	Mark Mode	Hint Mode	Link Marks	Help	
Alter: 20 files	AN B During th	ne next 60 hou	rs I had the fo	ollowing reacti	ons, though not all	1 at the same t	ime: 02/19/21: 10:00PM No Nausea and extreme H0 headache ,	
A_ALarge_01.txt_hh.xml	16 🗢							
A_Chill_07.txt_hh.xml	14 O 11:00 PM	- 2 hours of	uncontrollable	shivering and	headache 2/20/21	1: 2:AM Extrem	e 🚹 headache (like Migraine), 🚻 Nausea, 🏼 PAO Body aching for the	
A_Chill_47.txt_hh.xml	5 o remainder	of the night						
A_Dizziness_17.txt_hh.xml	24 🗢		TREATMENT: T	0				
A_Fatigue_01.txt_hh.xml	5 O 2 Tried MO	Advil , then	M tylenol -	didn't tough 🔠	headache.		2 PYREXIA	
A_Headache_02.txt_hh.xml	2 O 3 7:00AM -	headache,	PA1 Body Aches	, po Fatique ,	co chills , mild	DIO diarrhea an	nd 👩 dehydration thr 3 CHILL y.	
A_Myalgia_09.txt_hh.xml	2 🗢						COUGH	
A_Nausea_11.txt_hh.xml	3 O 4 Slept, f:	inally that ni	ght.				5 FATIGUE	
A_Pain_11.txt_hh.xml	3 O 5 Advil	helped the he	ad and PA Body	Aches finally			🗣 q PAIN	
A_Pyrexia_10.txt_hh.xml	3 0						🔶 🖬 HEADACHE	
B_ALarge_01.txt_mm.xml	23 0 6 2/21/21 -	- Left with bo	dy 🗜 Fatigue,	mild DI diarr	hea, 🔟 dehydration	1.	🗣 🕘 DIZZINESS	
B_Chill_07.txt_mm.xml	14 O 7 Night swe	eats for most	of that night.	b.	1		🗣 r SORENESS	
B_Chill_47.txt_mm.xml	11 💿				1		🗣 t NAUSEA	
B_Dizziness_17.txt_mm.xml	13 0 8 2/22/21 -	 energy back, 	mild Di diarr	chea left.			🗣 a DIARRHEA	
B_Fatigue_01.txt_mm.xml	4 0 9 2/23/21 -	9 2/23/21 - good, except left with primid diarrhea Leary of taking anything for the pi diarrhea, as I t 🗣 🛛 OTHER AE 🤉.						
B_Headache_02.txt_mm.xml	2 0							
B_Myalgia_09.txt_mm.xml	3 O Through f	this whole ord	eal, I M2 consu	umed tons of wa	ter , almost drink:	ing nonstop due	to the thirst.	
🖿 All Tags	14 💊 CHILL	1 C0	392~398 c	hills	comment			
	• D							
2 PYREXIA	1 SOTHER_AE	iii 00	418~429 d	lehydration	comment			
3 CHILL	1							
4 COUGH	0 SMEDICATION	1 M2	814~836 cc	insumed tons of water	comment			
5 FATIGUE	0							
🗣 g PAIN	2 🔖 DIARRHEA	10 DI0	405~413 d	liarrhea	comment			
💛 🖬 HEADACHE	2							
e DIZZINESS	0 🗣 DIARRHEA	i DI1	703~716 п	nild diarrhea	comment			
SORENESS								

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.