Design Principles Towards Higher Efficiency in Multi-document Annotation

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

Department of AI and Informatics, Mayo Clinic, Rochester, MN, USA

* He.Huan@mayo.edu
† Liu.Hongfang@mayo.edu

Background

Building a high-quality annotation dataset requires considerable time and expertise efforts to go through the whole annotation process with annotation tools. Although existing text annotation tools may provide many features to meet different needs in annotation, limited attention is paid to how to improve the annotation efficiency and user experience.

In response, we propose applying human-computer interaction (HCI) design principles to guide the design of annotation tools. Moreover, based on the design principles and the feedback summarized from our experienced domain experts, we developed *MedTator*, a lightweight, open-source, web-based application for multi-document annotation.
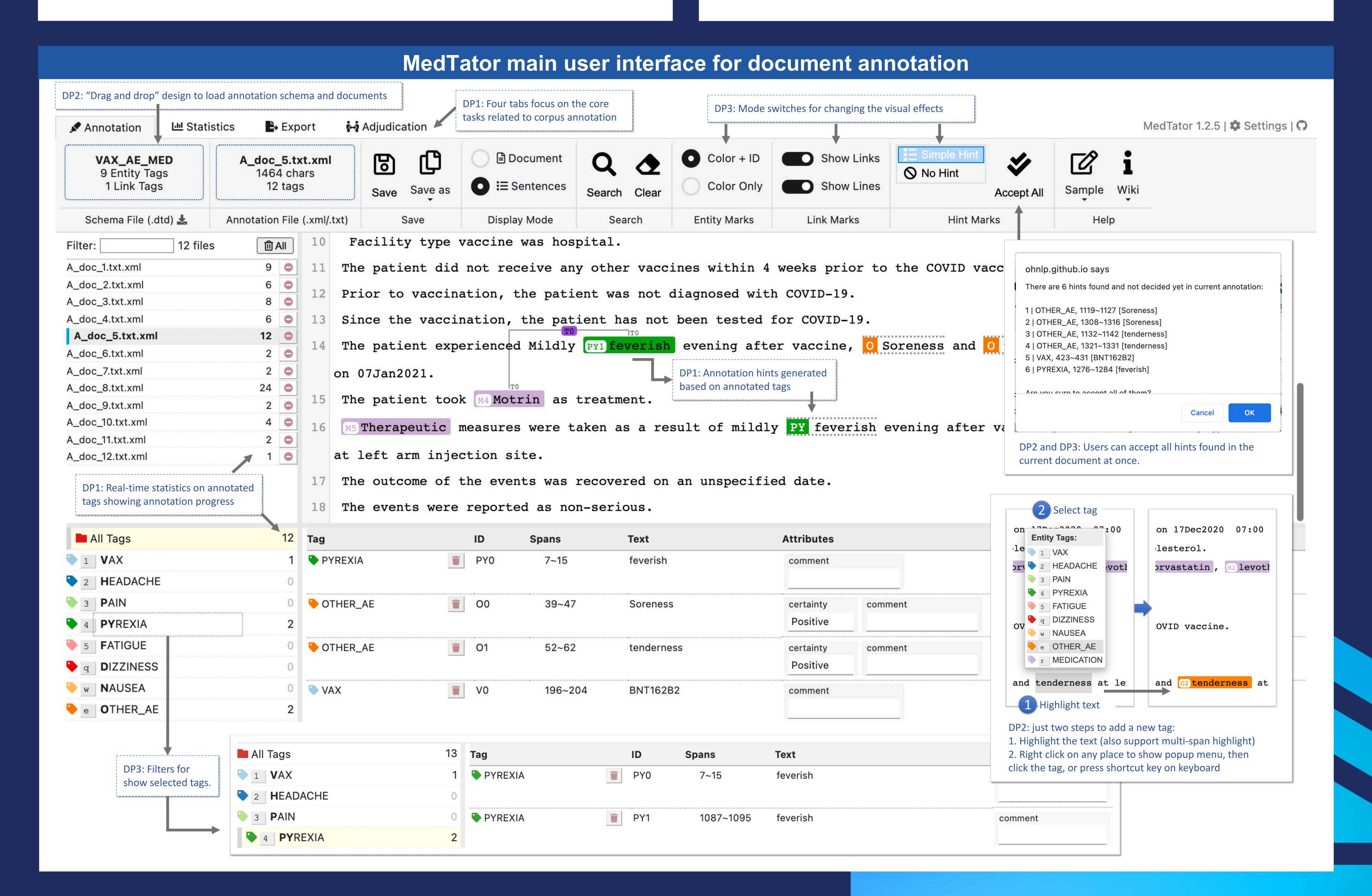
MedTator doesn't require any runtime installation. You can check our online live demo at: https://ohnlp.github.io/MedTator/ .

Source code are available at: https://github.com/OHNLP/MedTator

Human-computer interaction design principles

We interviewed our domain experts who have extensive annotation experiences to get their comments on existing tools and annotation workflow. Based on their comments, we summarized the design principles to guide our development, including:

1) **Reducing the cognitive overhead**: The design of existing tools may distract users from annotation and increase the cognitive load by other unrelated tasks. To enable users to focus on the actual annotation, we use visual designs to show more useful information.

2) **Low physical effort**: Users should be able to complete a task without any "unnecessary" operations. We studies existing workflow and tried to minimize the steps in each task.

3) **Flexibility in use**: When annotating documents, users may have different working preferences and requirements.

MedTator main user interface for document annotation

Lessons learned during the development of MedTator

Ease of installation: As an open-source software, a very important aspect is its installation. Most of existing tools usually require a runtime environment to be installed and configured correctly before starting corpus annotation. We propose using serverless design to address this issue, so that user don't need to install any runtime. All dependencies can be solved based on HTML5 techniques and users can use MedTator as a webpage or as a standalone program.

Documentation: While the development and testing of MedTator, our domain experts and annotators have tried to apply the prototypes and early versions in our projects for corpus annotation and adjudication. Then we found that a variety of documents are needed, such as tutorials for basic workflow, instructions for all necessary functions, and best practices for each task during annotation. It will be very helpful to provide comprehensive documents to users to cover different needs.

Conclusion

We identified some HCI design principles to guide the development of a lightweight annotation tool for corpus annotation. We have released several versions on our GitHub repository, but we are still developing new features and fixing bugs to improve its usability and flexibility.

More technical details can be found at the wiki page on its GitHub repo.