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

Introduction: Building a high-quality annotation dataset requires considerable time and expertise efforts to go through the whole annotation process with annotation tools[1]. 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 (Live demo and source code are available at: https://github.com/OHNLP/MedTator).

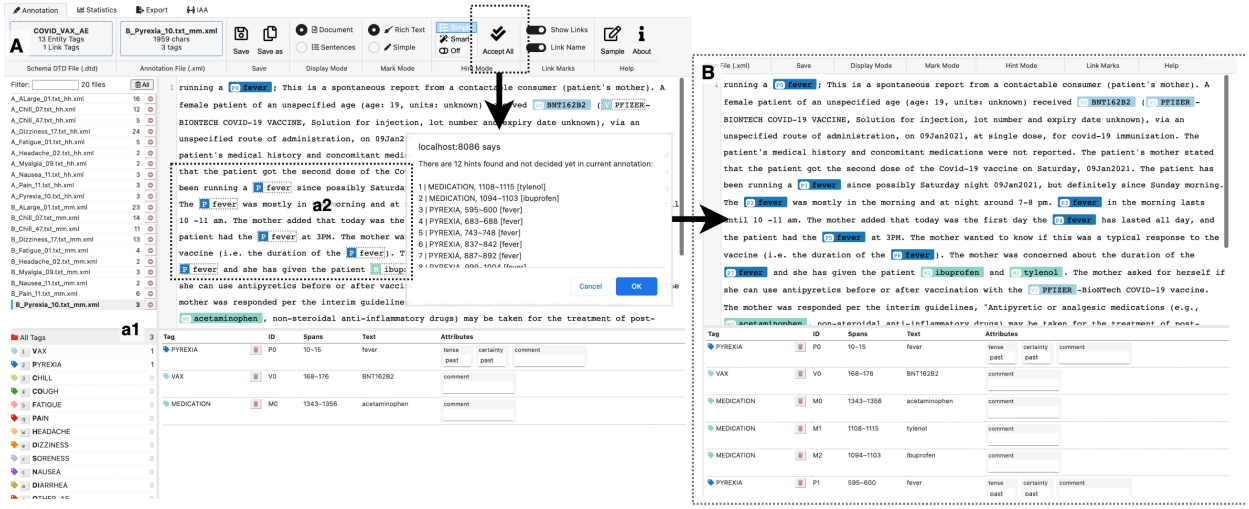

Figure 1. The screenshot of MedTator, including (A) the annotation editor showing (a1) real-time statistics on annotated tags and (a2) annotation hints. (B) Accept all hints as tags to reduce repetitive workload.

Design Principles: Since there have been several comprehensive annotation tools used in our past projects, we interviewed our domain experts who have extensive annotation experiences to get their comments on existing tools and annotation workflow. Based on their comments and task requirements, we summarized the design principles to guide our development, including: 1) *reducing the cognitive overhead*. To enable users to focus on the actual annotation tasks rather than other distracting tasks, we only selected those core features related to corpus annotation to be included in MedTator. And we customized visual designs to show useful information for the annotation task; all necessary functions are organized on single page with the same design language; the statistical results of the annotated tags are updated in real time to help users track annotation progress (Fig. 1(a1)); 2) *low physical effort*. To minimize the number of steps in each task, we studied the existing workflow and optimized the interactive process for each task. For example, to start a new annotation task from beginning, users only need to drag and drop the schema file and raw text files without any further operation; the annotation hints are optionally shown based on the completed annotations to reduce repetitive searching (Fig. 1(a2)); and adding all annotation hints as new tags only requires one click to confirm the suggestions (Fig. 1(B)); 3) *flexibility in use*. To address users' requirements for different workspace preferences, we use this principle to guide the tool design to balance the needs of tool features and the interface complexity. For example, we designed a variety of mode switches and organized them in a unified display layout for changing the visual effects.

Future work: Although our tool is still in development stage, we demonstrated our visual and interactive designs to domain experts and got positive feedbacks. We are going to improve its flexibility to meet the needs for downstream tasks and conduct a formal evaluation to improve the usability.

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.