Visualizing and Understanding the Errors of Information Extraction Results

Huan He, PhD, Sunyang Fu, PhD, Liwei Wang, MD, PhD, Andrew Wen, MS,
Sijia Liu, PhD, Sungrim Moon, PhD, Kurt Miller, MS, and Hongfang Liu, PhD
Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA

Introduction. Information extraction (IE) is a critical step in facilitating the use of healthcare data in clinical decision support and translational research by leveraging natural language processing (NLP) techniques. To understand and correct errors from IE systems or algorithms, it is essential to conduct an error analysis on the IE results by computing aggregated metrics (e.g., error rate by types, labels, etc.). However, as manually labeling errors is time-consuming and IE errors may be complex, it's challenging to make an in-depth error analysis to gain insights. To address this challenge, we created a visual error analysis tool using data visualization techniques to explore IE errors.

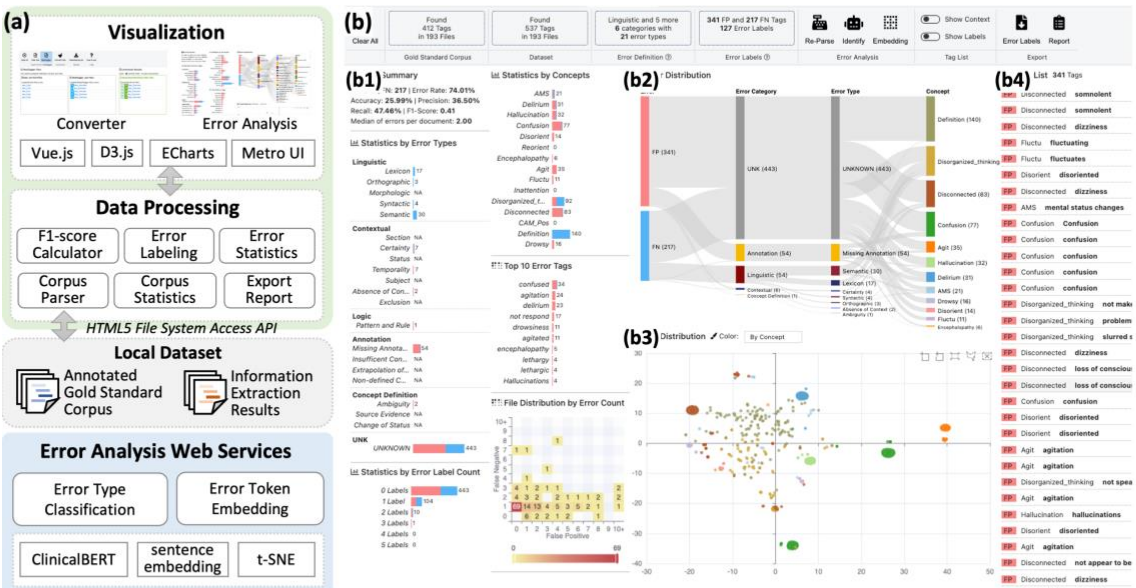

Figure 1. (a) The architecture of our proposed tool. (b) The screenshot of our tool showing (b1) the basic statistics of errors of multiple aspects; (b2) the error distributions between different types; (b3) the visualization of error embeddings; and (b4) the list of error tokens with contextual information.

System Design. We designed the tool based on a serverless text annotation tool, MedTator[2], to reuse the core modules of corpus processing and file system API. As shown in Fig. 1a, the tool consists of three components: the visualization module, the data processing module, and an optional error analysis web service module.

The visualization module integrated interactive charts and tables to show the statistical results of the errors. As shown in Fig. 1b, users can explore the error distribution from multiple aspects, such as error types, documents, and IE concepts. The detailed list of error tokens can be shown when clicking on any visual element in other charts. In addition, users can manually label errors by pre-defined error definitions, and the charts will be updated instantly. The data processing module implements evaluation functions for error analysis, such as F1-score calculation, error labeling, and error statistics, which can help to explore the errors between the gold standard and IE results. The error analysis web service module is designed to facilitate the error analysis by leveraging existing NLP models and algorithms, including (a) error type classification service that can classify the given IE error result based on pre-defined rules and error definitions to reduce manual efforts; (b) error token embedding service that can generate the dimension-reduced embeddings of given tokens to visualize the errors based on ClinicalBERT and t-SNE. The source code and online demo are available at [https://github.com/OHNLP/MedTator.](https://github.com/OHNLP/MedTator)

Discussion and Future Work. We tested our tool on sample datasets generated by our IE system. The preliminary results show that the visualization features can enable an easier process of error analysis. As the next step, we plan to improve the performance of the web service and conduct case studies to evaluate and enhance the visual design.

References

1. Wu T, Ribeiro MT, Heer J, Weld D. Errudite: Scalable, Reproducible, and Testable Error Analysis. In: Proceedings of the 57th Annual Meeting of the ACL; 2019 [cited 2022 Sep 9]. p. 747–63.
2. He H, Fu S, Wang L, Liu S, Wen A, Liu H. MedTator: a serverless annotation tool for corpus development. Bioinformatics, Vol. 38, Issue 6, 15 March 2022, Pages 1776–1778