

MetaTron: Streamlining Collaborative Annotation for Biomedical Documents

Ornella Irrera¹, Stefano Marchesin¹ and Gianmaria Silvello¹

¹University of Padova, via Gradenigo 6/B, Padova, Italy

Abstract

Manual annotation of biomedical texts, such as electronic health records and medical reports, is crucial to creating reliable corpora for training automated methods for tasks like relation extraction and entity linking. To streamline this process, we introduce MetaTron, a web-based collaborative tool that supports mention-level and document-level annotations and automatic built-in predictions¹.

1. Motivation

The exponential growth of biomedical textual data such as Electronic Health Records (EHRs), medical reports, and clinical notes, has created a growing demand for novel methods to extract and organize machine-readable knowledge, enabling easier access and reuse of information for training and testing models and applications in the biomedical domain. A key challenge in this context is the creation of large annotated corpora, which can be used as reliable ground truths for developing robust models for tasks such as relation extraction, topic recognition, and entity linking. Manual annotation requires the effort of domain experts, which is resource- and time-demanding but often unavoidable.

To assist healthcare experts in the annotation process and promote the creation of annotated corpora, we introduce MetaTron, a free, open-source, web-based tool designed for collaborative annotation of biomedical data. It supports six annotation types and includes automatic built-in predictions. MetaTron offers a unique combination of features, including relation annotation, which is rarely found alongside capabilities like mention and entity annotation, collaborative annotation, and automated inter-assessor agreement methods. In addition, it is fully customizable to meet specific needs.

We presented MetaTron in [1] and here we present an updated version with some additional features, such as tag-based annotations and a revised user interface. Our goal is to engage with the community and discuss its recent updates and expansions.

2. MetaTron

MetaTron [1] is a web-based annotation tool designed to support human annotators in the annotation of biomedical documents. MetaTron is available online at <https://metatron.dei.unipd.it> and as Docker container <https://github.com/GDAMining/metatron/>. This ensures portability, scalability, and isolation, while enabling private document uploads, offline use, and feature customization. The annotation interface of MetaTron is shown in Figure 1. On the left side, a vertical toolbar allows annotators to quickly access the tool's features during the annotation process, facilitating easy selection of different annotation types and improving workflow efficiency. The main webpage displays the textual document annotators interact with. MetaTron supports six annotation types, categorized into mention-level and document-level annotations. Mention-level annotations focus on specific portions of the text and include mention annotations, entity tagging, entity linking, and relationship annotation. Document-level annotations, on the other hand, consider the entire document as a unit, independent of individual

¹MetaTron was initially introduced in [1] and is further described in this work.

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

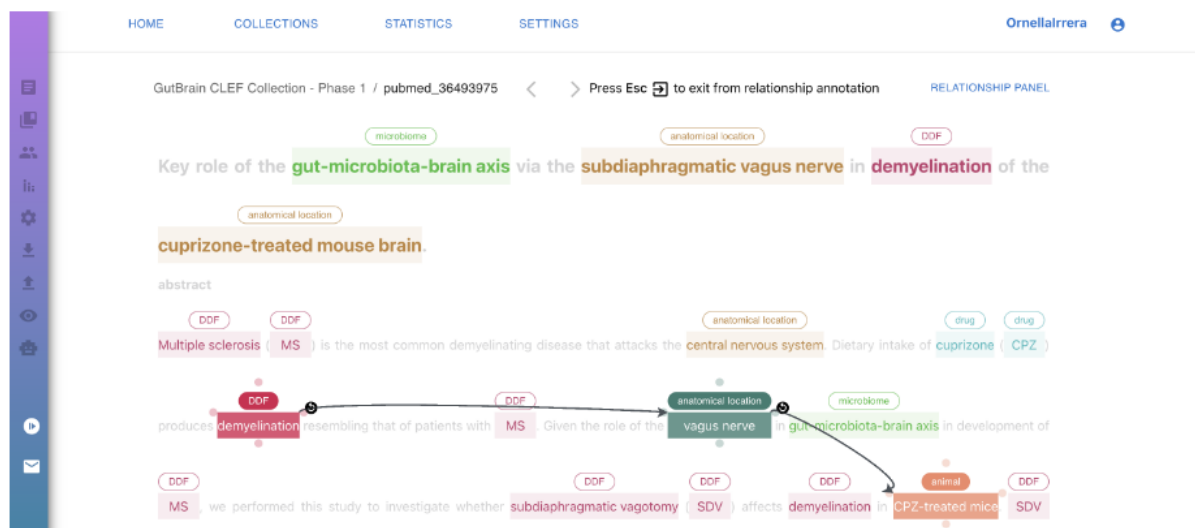
✉ ornella.irrera@unipd.it (O. Irrera); stefano.marchesin@unipd.it (S. Marchesin); gianmaria.silvello@unipd.it (G. Silvello)

🆔 0000-0003-2284-5699 (O. Irrera); 0000-0003-0362-5893 (S. Marchesin); 0000-0003-4970-4554 (G. Silvello)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Figure 1: The MetaTron annotation interface. The figure illustrates the annotation of a relationship consisting of a subject (red), a predicate (green), and an object (orange). Mentions are highlighted in various colors, with tags and linked concepts displayed above each mention.



words. These include label annotations and assertions in the form of subject-predicate-object RDF triples, where each entity can be linked to ontological concepts.

MetaTron also implements AutoTron, a framework designed to automatically annotate relationships and assertions. By offering an initial set of annotations that can be modified, AutoTron improves the user experience and accelerates the annotation process. Annotators can share collections to annotate among several annotators. MetaTron supports multiple annotation rounds, promoting continuous improvement and ensuring the creation of robust, reliable annotated corpora. It implements detailed document statistics to monitor annotators' progress and track the quality of their annotation. Examples are Cohen's and Fleiss' kappa statistics which help in keeping track of the consistency and reliability of annotations. To assess the tool's usability, understandability, we conducted a user study involving three human annotators. The user study found MetaTron to be intuitive and user-friendly. In recent months, MetaTron has been adopted by several research groups for annotating biomedical documents. A significant annotation initiative is currently underway, utilizing MetaTron as the primary tool for annotating PubMed abstracts with entities and relationships.

3. Final remarks

We outlined MetaTron, a biomedical manual annotation tool that stands out for its intuitive interface, collaborative features, and support for automatic annotations, making the annotation process more efficient and reliable. Its full customizability makes this tool highly versatile, targeting diverse needs in biomedical textual data annotation.

Acknowledgments

This work is supported by the HEREDITARY Project under Grant Agreement No GA 101137074.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] O. Irrera, S. Marchesin, G. Silvello, MetaTron: advancing biomedical annotation empowering relation annotation and collaboration, *BMC Bioinformatics* 25 (2024) 112.