

## SOFTWARE

# MedTAG: A Portable and Customizable Annotation Tool for Biomedical Documents

Fabio Giachelle, Ornella Irrera and Gianmaria Silvello\*

\*Correspondence:

gianmaria.silvello@unipd.it  
Department of Information  
Engineering, University of Padua,  
Padua, Italy  
Full list of author information is  
available at the end of the article

## Abstract

**Background:** Semantic annotators and *Natural Language Processing* (NLP) methods for *Named Entity Recognition and Linking* (NER+L) require plenty of training and test data, especially in the biomedical domain. Despite the abundance of unstructured biomedical data, the lack of richly annotated biomedical datasets poses hindrances to the further development of NER+L algorithms for any effective secondary use. In addition, manual annotation of biomedical documents performed by physicians and experts is a costly and time-consuming task. To support, organize and speed up the annotation process, we introduce MedTAG, a collaborative biomedical annotation tool that is open-source, platform-independent, and free to use/distribute.

**Results:** We present the main features of MedTAG and how it has been employed in the histopathology domain by physicians and experts to annotate more than seven thousand clinical reports manually. We compare MedTAG with a set of well-established biomedical annotation tools, including BioQRator, exTag, MyMiner, and tagtog, comparing their pros and cons with those of MedTAG. We highlight that MedTAG is the only open-source tool provided with an open license and a straightforward installation procedure supporting cross-platform use.

**Conclusions:** MedTAG has been designed according to five requirements (i.e. available, distributable, installable, workable and schematic) defined in a recent extensive review of manual annotation tools. Moreover, MedTAG satisfies 20 over 22 criteria specified in the same study. Finally, we plan to introduce additional features, such as the integration with PubMed, to improve MedTAG.

**Keywords:** Biomedical annotation tools; Entity extraction; eHealth; Digital health

## Background

In the last decades, exascale volumes of biomedical data have been produced, where the vast majority is available as unstructured text [1]. Health-care professionals traditionally rely on free-text reporting for communicating patient information such as diagnosis and treatments. For instance, narrative clinical reports are usually conceived as free-text reports, which are human-readable but not machine-readable. This brings interoperability issues and limitations to effective secondary reuse of data, essential for medical decision making and support. In order to process the vast amount of unstructured biomedical data from clinical reports and *Electronic Health Records* (EHRs), *Information Extraction* (IE) algorithms and NLP techniques have been developed and are currently exploited.

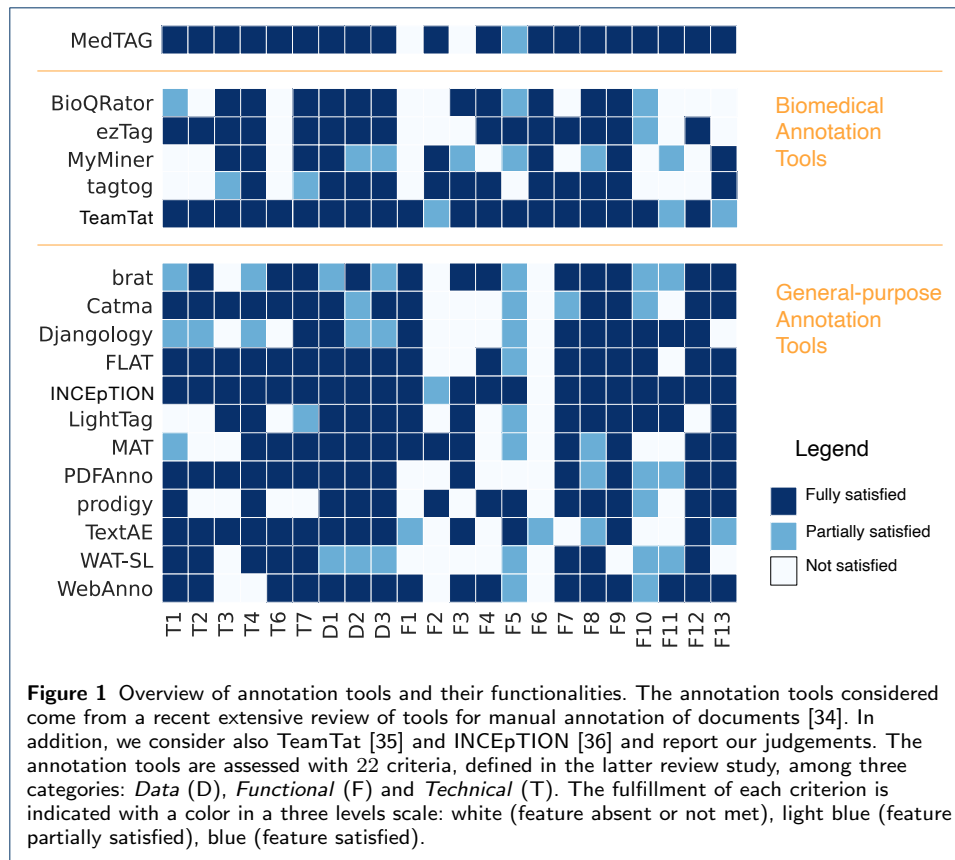
To this aim, significant efforts have been dedicated to applying *Named Entity Recognition and Linking* (NER+L) methods for entity extraction and semantic an-

notation [2–6]. Semantic annotation is the NLP task of identifying the type of an entity and uniquely linking it to a corresponding knowledge base entry [7]; it leverages both text-processing and *Machine Learning* (ML) techniques to tackle biomedical information extraction challenges such as terms and abbreviations disambiguation. Furthermore, semantic annotation tasks based on ML methods are often carried out in a supervised context where large-scale training and test annotated corpora are required. Moreover, even in an unsupervised context, NER+L models require annotated datasets for evaluation purposes. However, the lack of manually annotated biomedical datasets poses hindrances to the further development of NER+L systems. In addition, most of the training data available for the biomedical domain covers mainly common entity types (e.g., drugs, genes, and diseases) [8–11], thus the coverage of some biomedical sub-domains is limited. For these reasons, several attempts to create large annotated biomedical corpora have been conducted [12–19].

To achieve the high-quality standards required for the biomedical domain, the annotation process demands human-expert supervision. Nevertheless, manual annotation of large datasets is an expensive and time-consuming task requiring plenty of expert annotators with extensive experience in biomedical contents. To support, organize and speed up the annotation process, several annotation tools have been developed [20–33]. However the biomedical domain is particularly challenging, since biomedical texts contain mentions that are burdensome for semantic annotation, such as the abbreviations of genes and proteins. Moreover, the specificity of some biomedical sub-domains, such as histopathology, requires fine-grained annotation systems designed to be customizable according to physicians' and experts' needs.

In recent years, several biomedical annotation tools have been released [34, 37]. Motivation for the wide variety of biomedical annotation tools available could be the necessity of domain-specific functionalities that might be only partially supported or not by other well-established tools. Hence, some tools could be handier than others for a specific task of interest.

A recent extensive review of both general-purpose and biomedical annotation tools provides a detailed comparison of state-of-the-art annotation tools [34]. Some of the common limitations of the available tools are, for instance, the non-availability of the source code or the raised exceptions and failures during the installation process. In addition, even the most popular annotation tools present drawbacks such as a burdensome installation procedure or the lack of documentation. As an example, WebAnno [38] and brat [39] are popular general-purpose annotation tools with a comprehensive set of functionalities, but their installation process is rather complex for the not technology-savvy users. INCEpTION [36, 40] is a more recent general-purpose annotation tool from the authors of WebAnno [38], that mitigates this issue with a web service enabling the users to work online. Moreover, general-purpose annotation tools often do not fulfill the needs of biomedical experts; thus, domain-specific solutions are preferable for this field. Even though brat [39] has been used in several biomedical projects [41–45], it is designed for general-purpose annotation, thus it provides additional features that are not suited for physicians and experts of the biomedical domain. Since the annotation process is a time-consuming task, biomedical annotation tools should be designed to offer an intuitive streamlined interface that minimizes redundant features, fulfill domain-specific requirements and reduce the annotators workload.



For the in-depth analysis, we focus on the tools specifically designed for biomedical annotations: BioQRator [25], ezTag [26], TeamTat [35], MyMiner [27] and tagtog [28]. Additionally, we also consider two general-purpose annotation tools that are used by the biomedical community as well - i.e., brat [39] and INCEpTION [36].

In Figure 1, we can see a heat-map reporting on the functionalities of the current text annotation tools as analyzed by a very recent extensive survey [34]. The provided heat-map is to be used as a visual summary of the features provided by each annotation tool. <sup>[1]</sup> In particular, the heat-map considers a list of 15 annotation tools selected according to five major requirements: (i) **Available**: the executable and project source code should be available; (ii) **Web-based**: the tool should be provided as an online web application or as an installable application running in a web browser container; (iii) **Installable**: the installation process should last two hours at most; (iv) **Workable**: it should work for hands-on experiments; (v) **Schematic**: users should be able to configure the annotation schema at will. Hence, several biomedical annotation tools such as Argo [29], Egas [24], Marky [30], ODIN [31], Pubtator [32] and Textpresso [33] are not considered since they do not satisfy one or more of the previous five requirements.

Moreover, the selected annotation tools are compared according to a set of 22 criteria chosen among the original 26 criteria of the same study [34]. In particular,

<sup>[1]</sup>We report details about the main features considered in [34] to ease the comprehension of our analysis.

the criteria are grouped in four categories: (i) *Data*, (ii) *Functional*, (iii) *Publication* and (iv) *Technical*. We excluded the publication criteria (i.e., the four missing criteria) since we are interested in comparing the facilities and functionalities provided by the different tools and not on their coverage in scientific publications.

The *data* criteria are: (D1) format of the schema – whether it is configurable or uses standard formats (e.g. JSON, XML); (D2) input format for documents – whether the input documents are based on standard formats (e.g. JSON, XML) and (D3) output format for annotations – whether the annotations are based on standard formats (e.g. JSON, XML).

The *functionality* criteria are: (F1) support for overlapping mentions/annotations; (F2) support for document-level annotations – users can specify the labels that apply to the whole document (not just for a textual portion); (F3) support for relationship annotations; (F4) support for ontologies and terminologies (i.e. a procedure to import terminology resources is provided); (F5) Support for built-in predictions and active learning from pre-annotated documents; (F6) Integration with PubMed – users can annotate PubMed abstracts just providing a list of PubMed ids; (F7) Suitability for full texts (i.e., tool capable of displaying long text correctly, without compromising readability); (F8) Allowance for saving documents partially (i.e., holding annotations partially to later continue the annotation process); (F9) Ability to highlight parts of the text; (F10) Support for users and teams; (F11) Support for *Inter-Annotator Agreement* (IAA); (F12) Data privacy (i.e., can be used for private data); (F13) multilingual support (i.e., annotating multilingual documents, that might contain special characters).

The *technical* criteria are (T1) Date of the last version – whether the last version (or commit) has been released within the past five years; (T2) availability of the source code – whether the source code is available in version control platforms; (T3) online availability for use; (T4) easiness of installation – i.e., available online (no installation required) or easy and fast to install (up to half-hour time); (T6) license allowing modification and redistribution; (T7) free of charge. We excluded T5 (quality of documentation) from the technical criteria since we are interested in objective and assessable criteria.

Figure 1 shows that several tools lack one of the following functionalities: (i) document-level annotation; (ii) ontology and terminology resources support; (iii) support for multi-label annotation; and (iv) support for collaborative annotations with users and teams. Moreover, seven over the fifteen selected tools are provided through a license that limits modifications and redistribution.

To mitigate this, we introduce MedTAG, a customizable, collaborative, web-based annotation tool provided as a docker container to enable cross-platform support and quick and easy installation. MedTag provides a step-by-step schema configuration, by which the project/team leader can specify in detail which document parts or document fields can be annotated. We designed MedTAG according to the five primary annotation tools' requirements previously discussed. Besides, we determined the feature coverage provided by MedTAG concerning the former criteria. Figure 1 shows that MedTAG satisfies most of the criteria, having a feature coverage of 20 criteria over 22. The rest of the criteria currently not covered by MedTAG, such as the relationship annotations and active learning capabilities, are planned as future work.

## Implementation

MedTAG has been designed to be flexible and customizable, so that users can easily install it and configure the annotation schema at will. Hence, MedTAG is not limited to a specific (sub-)domain (e.g., histopathology), but it can be seamlessly used in different biomedical sub-domains. The key MedTAG functionalities are: (i) a web-based collaborative annotation platform with support for users and roles; (ii) a user-friendly interface with support for click-away mention annotation, mentions highlighting in different colors and automatic saving every time an action is performed; (iii) sorting of documents based on the lexicographic order or the “unannotated-first” policy; (iv) support for mobile devices; (v) download of annotations in several formats (i.e., BioC/JSON, BioC/XML, CSV, JSON); (vi) support for multi-label annotation; (vii) support for document-level annotations; (viii) multilingual support; (ix) support for ontologies/concepts to use for the annotation process; (x) support for IAA; (xi) integration with PubMed; (xii) support for automatic built-in predictions; (xiii) support for schema configuration, so that users can easily import data (i.e., documents, labels and concepts), as CSV files, and choose which document fields to annotate. In order to achieve automatic annotations and built-in predictions, we integrated the *Semantic Knowledge Extractor Tool* (SKET)<sup>[2]</sup> in MedTAG. Note that the support for built-in predictions is currently limited to three cancer use-cases (i.e., cervix, colon, and lung cancer). Nevertheless, we plan to extend the support for automatic built-in predictions also for other use-cases. General-purpose automatic annotation methods are of limited efficacy for the biomedical domain; nevertheless, the integration of SKET paves the road for the integration of other third-party libraries users may want to employ.

To exploit the concept linking functionality, MedTAG requires the admin user to specify, during the configuration phase, the CSV file containing all the concepts used for annotating the clinical reports. During the first configuration, the admin user is not defined yet, thus the configuration is handled by the *Test* user in *Test mode*, as described in the *Installation and customization* section. Figure 5.2 shows the configuration interface that allows the users to specify the CSV file for the concepts. Moreover, the users can choose whether to use the concepts of the ExaMode ontology<sup>[3]</sup> (necessary for the automatic annotation module using SKET) or a set of concepts from a different ontology. Then, the concepts provided in the CSV file populates the MedTAG database and are integrated in the drop-down menu available to the user to select the concepts. Every concept defined in the provided CSV is uniquely identified with a concept IRI. Thus, users could use concepts defined in different ontologies at the same time. Since the CSV file with the concepts for the annotation process is provided by the admin user, the coherence of the data (e.g., the same concept mapping to more than one IRI from different ontologies) should be checked and enforced by the admin herself. Nevertheless, in the case of the same entity mapping to different ontologies, MedTAG differentiates the concepts in the user interface based on the IRIs and other concept information such as use-cases and semantic areas. Thus, users have the means to disambiguate between potentially similar concepts.

---

<sup>[2]</sup><https://github.com/ExaNLP/sket/>

<sup>[3]</sup><http://examode.dei.unipd.it/ontology/>

MedTAG source code and the documentation are publicly available at this URL: <https://github.com/MedTAG/medtag-core>.

### Architecture

Figure 2 illustrates the MedTAG architecture, which consists of three logic layers (i.e., *Data*, *Business* and *Presentation* layer). The data layer concerns information and data management; it consists of two main relational databases realized with PostgreSQL, namely, the *MedTAG data* and the *Log data* databases. The former contains documents, entity concepts/labels, and the relations among them. The latter takes care of logging data such as user-provided information about issues with the documents to be annotated. The business layer controls the whole information flow as the information is displayed in the web interface and stored in the MedTAG database. It consists of two business units, the business logic, and the REST APIs end-point. The first one consists of Python routines and a controller that invokes the proper routine based on the received request. The second one is the back-end entry-point of MedTAG; it handles all the user requests from the web interface, invoking the business logic controller and returning its result to the front-end. The presentation layer provides the MedTAG front-end; it consists of a web interface to navigate the documents, annotate them and download the annotations in different formats (i.e., BioC/JSON, BioC/XML, CSV, and JSON).

Figure 2 shows the technologies adopted for each logic layer: (i) the front-end interface built with React.js<sup>[4]</sup>, HTML5 and CSS3; (ii) the back-end for web API and services built with the Python web framework Django<sup>[5]</sup>; (iii) the *MedTAG data* relational database implemented using PostgreSQL.

Due to the multitude of architecture components, manually installing and configuring each one would be cumbersome and error-prone. To mitigate this, we provide a fast and reliable installation by distributing MedTAG as a docker container.

### Installation and customization

Since MedTAG is provided as a Docker container, both *docker*<sup>[6]</sup> and *docker-compose*<sup>[7]</sup> are required. The detailed installation procedure is described at <https://github.com/MedTAG/medtag-core/tree/main#installation>. We can summarize the MedTAG installation in three steps:

- 1 Check the Docker daemon - i.e., `dockerd` - is up and running.
- 2 Download the `MedTAG_Dockerized`<sup>[8]</sup> folder from the `medtag-core`<sup>[9]</sup> repository, or clone it.
- 3 Open the `MedTAG_Dockerized` project folder and, on a new terminal session, type `docker-compose up`.

Once the installation process has been completed, MedTAG is available on your browser at `http://0.0.0.0:8000`. At this stage, users can access MedTAG only in *Test mode* – i.e., by using the pre-loaded documents. The pre-loaded documents

---

<sup>[4]</sup><https://reactjs.org/>

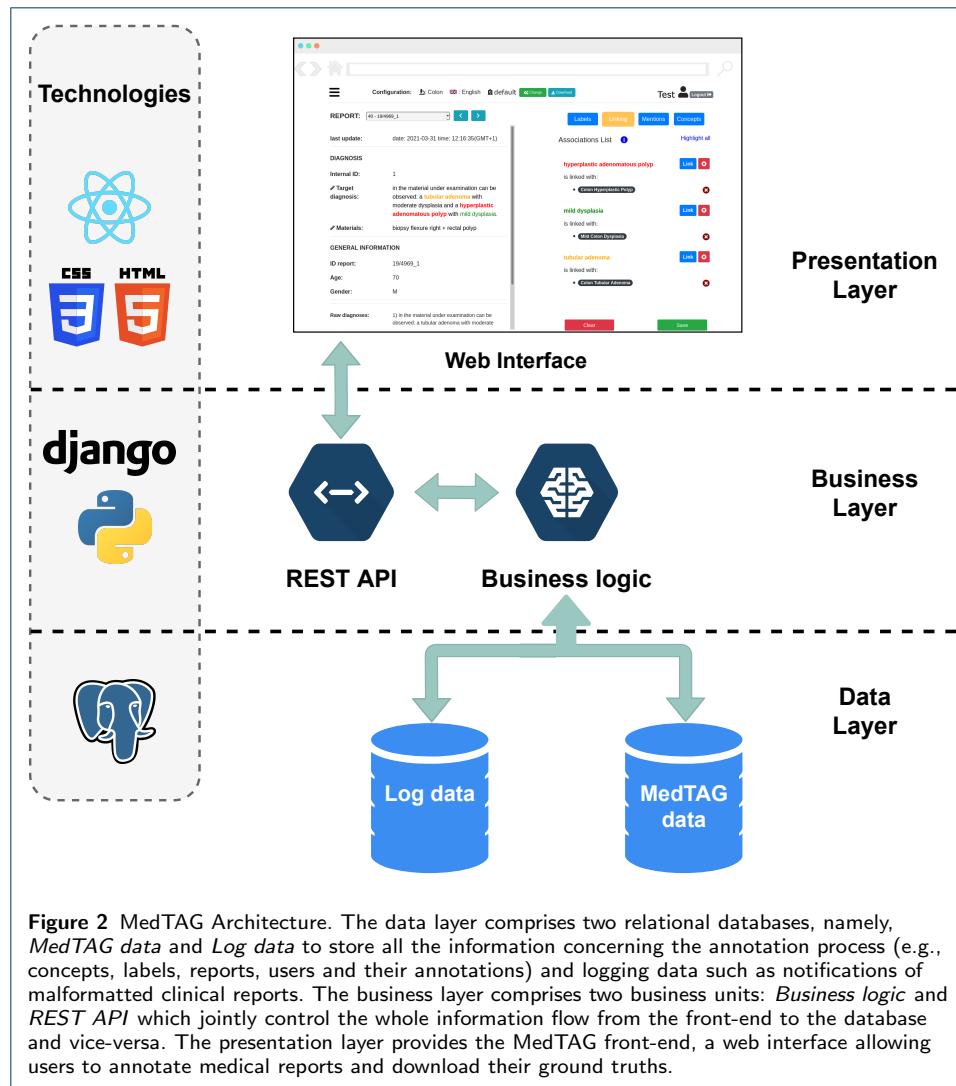
<sup>[5]</sup><https://www.djangoproject.com/>

<sup>[6]</sup><https://docs.docker.com/engine/reference/commandline/docker/>

<sup>[7]</sup><https://docs.docker.com/compose/>

<sup>[8]</sup>[https://github.com/MedTAG/medtag-core/tree/main/MedTAG\\_Dockerized](https://github.com/MedTAG/medtag-core/tree/main/MedTAG_Dockerized)

<sup>[9]</sup><https://github.com/MedTAG/medtag-core>



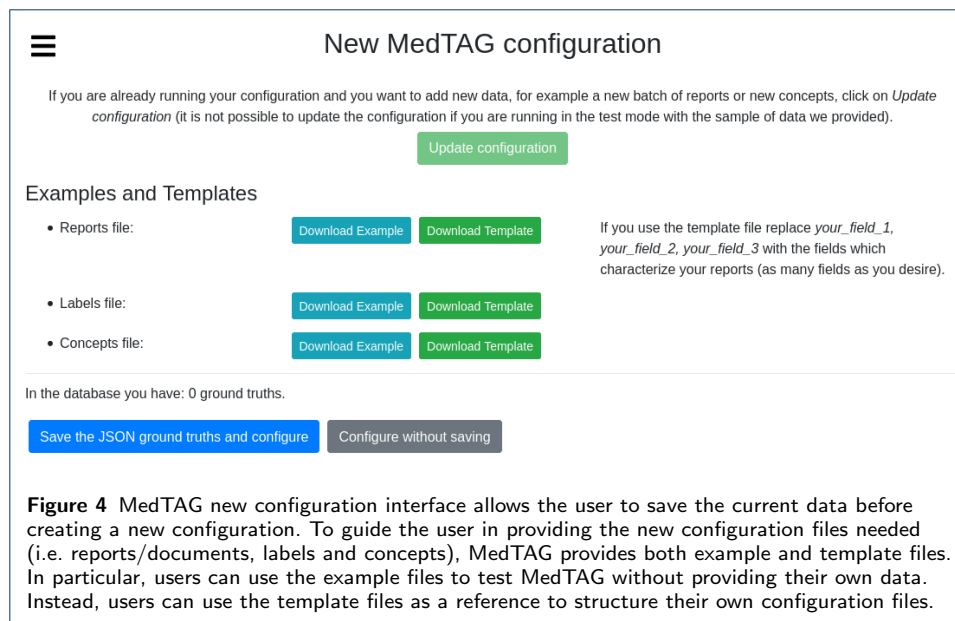
for the test mode are taken from the histopathology domain because we chose this domain as a use case for introducing and testing MedTAG functionalities.

Users can log into MedTAG and test it with the preloaded medical reports using *Test* as username and password.

To customize MedTAG, the users need to follow three steps: (i) open the menu and click on *Configure*, as shown in Figure 3; (ii) follow the instructions of the guided procedure – i.e., users are asked to provide both the admin user credentials and three CSV files: *concepts\_file*, *labels\_file* and *reports\_file*, as shown in Figure 4. The users are provided with CSV templates and with examples containing real data to speed-up the data preparation procedure; (iii) choose which document fields to display and annotate as shown in Figure 5; the *Check* button activates the file compliance procedures that will produce some state messages in different colors to inform the user about whether the CSV files provided are well formatted or not. Figure 5 shows the configuration interface that allows the users to specify whether to use the ExaMode concepts (indicated with number two) and labels (indicated with number three) or to upload a new set of concepts from different ontologies. The



**Figure 3** MedTAG sidebar provides the *Configure* option, indicated by the orange arrow, to set up a new custom configuration.



**Figure 4** MedTAG new configuration interface allows the user to save the current data before creating a new configuration. To guide the user in providing the new configuration files needed (i.e. reports/documents, labels and concepts), MedTAG provides both example and template files. In particular, users can use the example files to test MedTAG without providing their own data. Instead, users can use the template files as a reference to structure their own configuration files.

latter are necessary in case users want to take advantage of automatic annotation features. In addition, users can choose whether to annotate custom documents or PubMed abstracts and titles. In the first case, users are required to provide all the reports to annotate as a CSV file, that is, `reports_file`. Then, users can choose the report fields to annotate at will. In the second case, users have to specify a list of PubMed identifiers as a CSV file. Then, users can annotate both abstract and title of each PubMed article specified.

The detailed customization procedure is available at <https://github.com/MedTAG/medtag-core#customize-medtag>.



### Configure MedTAG with your data

#### PubMed IDs (Mandatory one between PubMed and reports files)

Insert here the .CSV file with the PubMed IDs of the articles you want to annotate. *Abstract and Title* sections will be annotable. Example

Nessun file selezionato

#### Reports (Mandatory one between PubMed and reports files)

Insert here the .CSV file with the reports to annotate. Example

clinical\_reports.csv

✗ Delete file

**Reports' fields** (It is mandatory to set at least one field to display. It is optional to set at least one field to annotate)

Below you can find all the fields which characterize your reports. For each key, you have to decide if you want to display, hide, or display and annotate the value corresponding to that key. Remember that to perform mentions annotation, you must select at least one field checking the button *Display and Annotate*

age	<input type="radio"/> Hide	<input checked="" type="radio"/> Display	<input type="radio"/> Display and Annotate
target_diagnosis	<input type="radio"/> Hide	<input type="radio"/> Display	<input checked="" type="radio"/> Display and Annotate
date_of_hospitalization	<input type="radio"/> Hide	<input checked="" type="radio"/> Display	<input type="radio"/> Display and Annotate
clinical_notes	<input type="radio"/> Hide	<input type="radio"/> Display	<input checked="" type="radio"/> Display and Annotate
clinical_studies	<input type="radio"/> Hide	<input type="radio"/> Display	<input checked="" type="radio"/> Display and Annotate
treatments	<input type="radio"/> Hide	<input type="radio"/> Display	<input checked="" type="radio"/> Display and Annotate

#### Files for CONCEPTS IDENTIFICATION AND LINKING (Optional)

If you want to perform concepts identification and linking you must provide the concepts.

You inserted reports for: **Colon**; EXAMODE ontology is available for these secases. **You can use the EXAMODE ontology concepts OR upload your own ones.** Get EXAMODE concepts

Insert here the .CSV file with the concepts. Example

Nessun file selezionato

#### Files for LABELS ANNOTATION (Optional)

If you want to perform labels annotation you must provide the labels related to the use cases you are interested in.

You inserted reports for: **Colon**; EXAMODE labels are available for these usecases. **You can insert the EXAMODE labels or upload your own ones.** Get EXAMODE labels

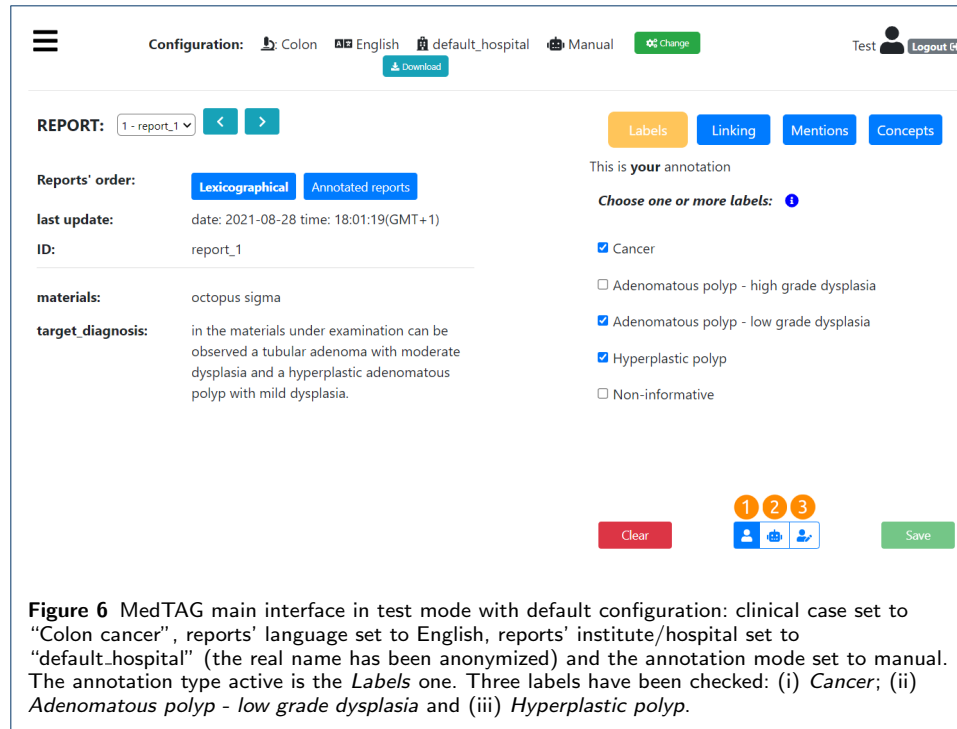
Insert here the .CSV file with the annotation labels. Example

Nessun file selezionato

**Figure 5** MedTAG main interface for data configuration. Users can provide their own CSV files for the reports/documents to annotate and the concepts and labels to use for the annotation process. Moreover, MedTAG detects automatically the document fields and allows users to specify which of them to annotate and/or display in the interface, as shown in the orange box (1).

### User interface and interaction

The MedTAG web interface has been developed based on the positive feedback received from physicians and experts in the digital pathology domain where an instance of MedTAG - i.e., ExaTAG - has been released. Figure 6 shows the main MedTAG web-interface for the annotation of medical documents or reports. On the top of the web page, there is the header section with the current MedTAG configuration: (i) the clinical case (e.g., Colon cancer); (ii) the report language (e.g., English); (iii) the hospital/institute which provided the report's dataset (e.g., "default.hospital" identifies the institute which provided the datasets of reports pre-



loaded in MedTAG in test mode) and (iv) the annotation mode (i.e., manual or automatic) used for the annotation process. In addition, the menu button (left-side) and the user section (right-side) are included in the header as well. It is worth noting that when the automatic annotation mode is active the users visualize the automatic annotations generated by the built-in annotation module. Any user edit concerning the automatic annotations is also replicated in the user profile, available for further edits in manual annotation mode. The user section shows the current username along with the *Logout* button. Below the header, the interface body is divided into two sections: the diagnostic report and the annotation section. The first one (left-side) shows the information regarding the textual document, that in the case of a medical report may contain the diagnosis and the patient’s information. Users can navigate between documents using either the keyboard arrows or the *next* and *previous* buttons. The annotation section (right side) shows the information concerning annotation labels, ontological concepts and the mentions identified in the selected document.

MedTag allows the users to use four different annotation types that can be activated alternatively by clicking on the corresponding buttons: (i) *Labels* is a form of document-level annotation where the reports are classified into predefined categories, (ii) *Mentions* where the user selects words in the text of the reports, (iii) *Linking* where the identified mentions are linked to ontological concepts, and (iv) *Concepts*, another form of document-level annotation, where the reports are annotated with ontological concepts not strictly tied to specific mentions.

In Figure 6 the *Labels* action is activated. We can notice three selected labels: “Cancer”, “Adenomatous polyp - low grade dysplasia” and “Hyperplastic polyp”. The labels describe properties or attributes that apply to the whole document, such

**Figure 7** MedTAG main interface in test mode with default configuration: clinical case set to “Colon cancer”, reports’ language set to English, reports’ institute/hospital set to “default.hospital” (the real name has been anonymized) and the annotation mode set to manual. The annotation type active is the *Linking* one. Three mentions have been identified and linked to the corresponding concepts: (i) *hyperplastic adenomatous polyp* is linked to *Colon Hyperplastic Polyp*; (ii) *mild dysplasia* is linked to *Mild Colon Dysplasia*; and (iii) *tubular adenoma* is linked to *Colon Tubular Adenoma*.

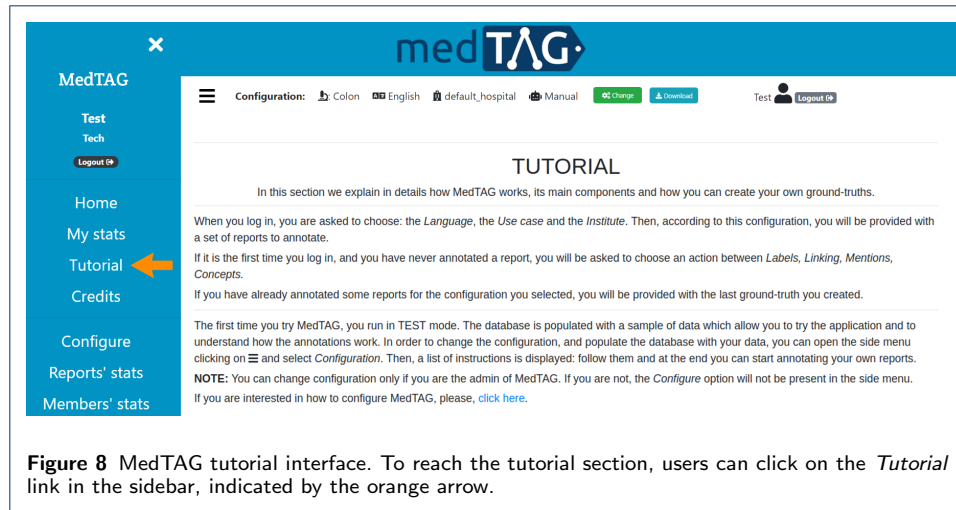
as the presence or the absence of cancer in the diagnosis of a clinical report. The set of labels used for the document-level annotation process, are provided by the user during the configuration phase, as previously discussed.

In Figure 7 the *Linking* action is activated. We can see three selected multi-word mentions in the text: “tubular adenoma”, “hyperplastic adenomatous polyp” and “mild dysplasia”. These mentions are linked to concepts taken from an histopathological ontology: (i) *hyperplastic adenomatous polyp* is linked to *Colon Hyperplastic Polyp*; (ii) *mild dysplasia* is linked to *Mild Colon Dysplasia*; and (iii) *tubular adenoma* is linked to *Colon Tubular Adenoma*.

The ontological concepts linked to the mentions can be selected via a drop-down menu (that, in turn, can be divided into semantic areas) or manually typed in a text field; in the latter case, the user is aided by auto-completion facilities.

To add a new mention, a user can click on any text token. After clicking on a text token, it gets highlighted with a new color, and the neighbor tokens turn highlighted as well, meaning that they could be selected as part of the current mention. All the mentions are highlighted with a different color in the document text and in the list of mentions for fast detection. The users can add, edit and delete the associations at will. Moreover, every time an action is performed, all the concerning information is asynchronously saved in the database; there is also manual saving via the *Save* button. Users can delete (after confirmation) all the annotations related to the current action button selected by clicking on the *Clear* button.

MedTAG enables the team members to collaborate during the annotation process. In particular, users can see anytime the annotations done by other team members for each clinical report by clicking on the button (3) of Figure 6. This feature is handy in case of annotation uncertainty (e.g., which concepts to associate to an



**Figure 8** MedTAG tutorial interface. To reach the tutorial section, users can click on the *Tutorial* link in the sidebar, indicated by the orange arrow.

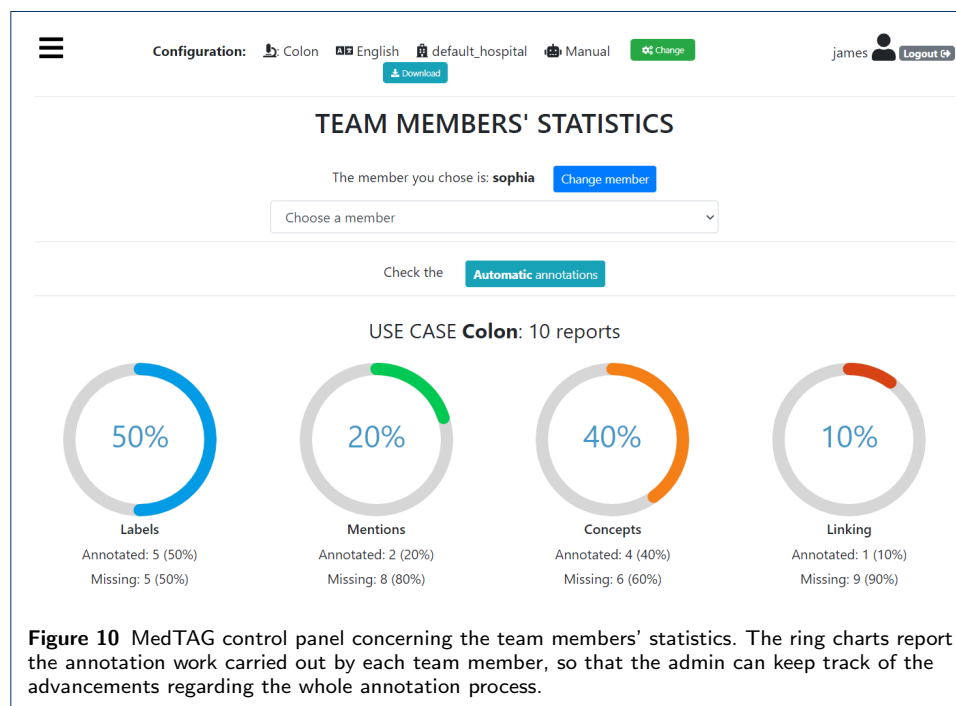
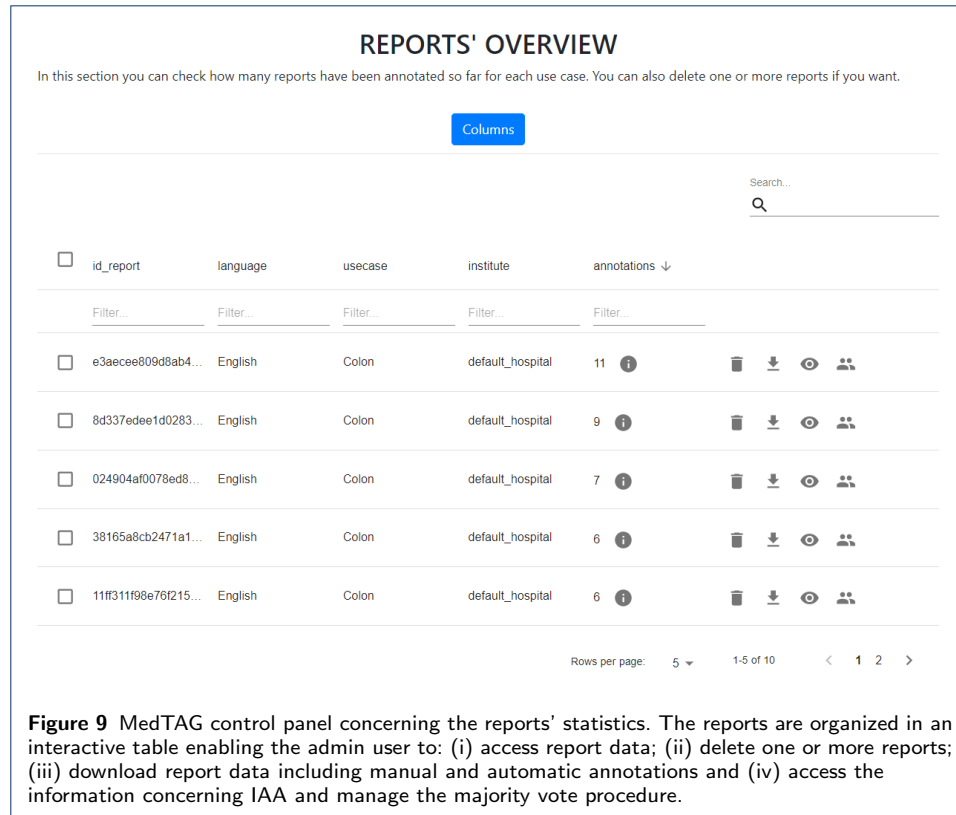
identified mention). To attain high-quality annotations, users can take advantage of the expertise and work other team members have previously done. In addition, users can visualize the automatic annotations made by the robot user - i.e., the automatic annotation module SKET - by clicking on button (2) of Figure 6. Moreover, the users can consult and edit the automatic annotations so that new edits are automatically copied in the user profiles for further modifications in manual annotation mode. Hence, users can take advantage of automated annotation facilities to reduce the annotation workload. Moreover, the admin can oversee the overall annotation process from the *Team members' statistics* section of the control panel. This section provides the admin user an overview of the annotation work carried out by each team member, providing information such as the number and the percentage of annotated reports for each use-case and annotation type. Hence, the admin can make decisions to coordinate the work of team members and keep track of the advancements in the annotation process.

Finally, users can download their annotations in different formats (i.e., BioC/JSON, BioC/XML, CSV and JSON), by clicking on the *Download* button.

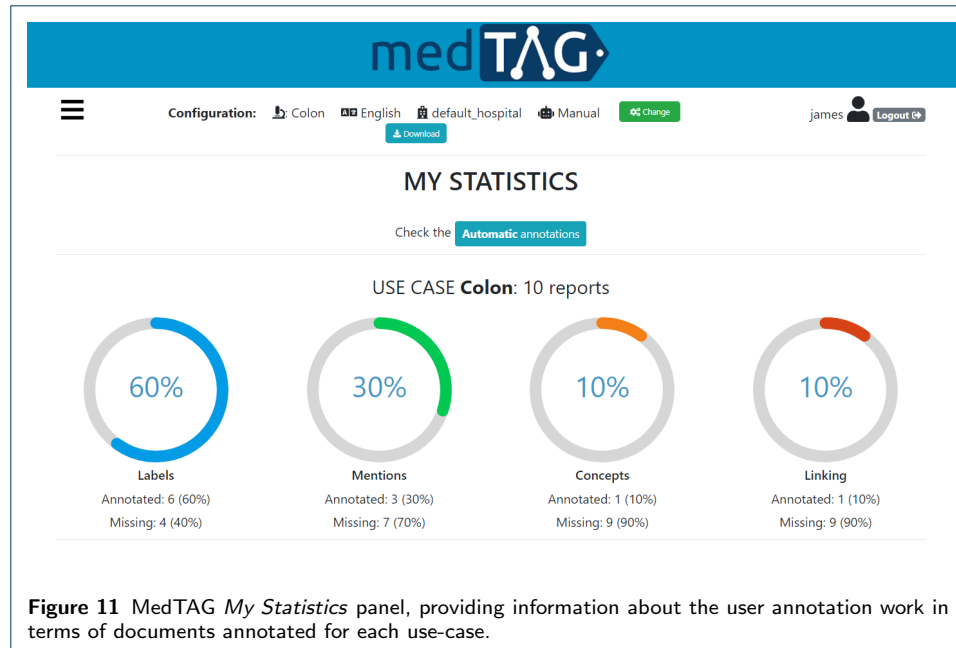
Overall, a detailed graphical tutorial is always available to the users to learn how to use MedTAG; the *Tutorial* link is provided in the sidebar, as shown in Figure 8.

### MedTAG control panel for statistics and Inter-Annotator Agreement (IAA)

MedTAG provides a unified interface that allows the admin user to access the annotation statistics (e.g., the number of users that annotated each report) and access the information concerning IAA for each report. It is worth noting that only the admin user can consult the statistics concerning the overall annotation process. Instead, other members can only access their statistics in the dedicated menu section *My statistics*. Figure 9 shows the control panel information organized in a dynamic table, where the admin can search, access, and filter the reports according to a selection of columns filters. Moreover, the admin can choose anytime which columns to show by clicking on the *Columns* button. The last column provides the following action buttons:



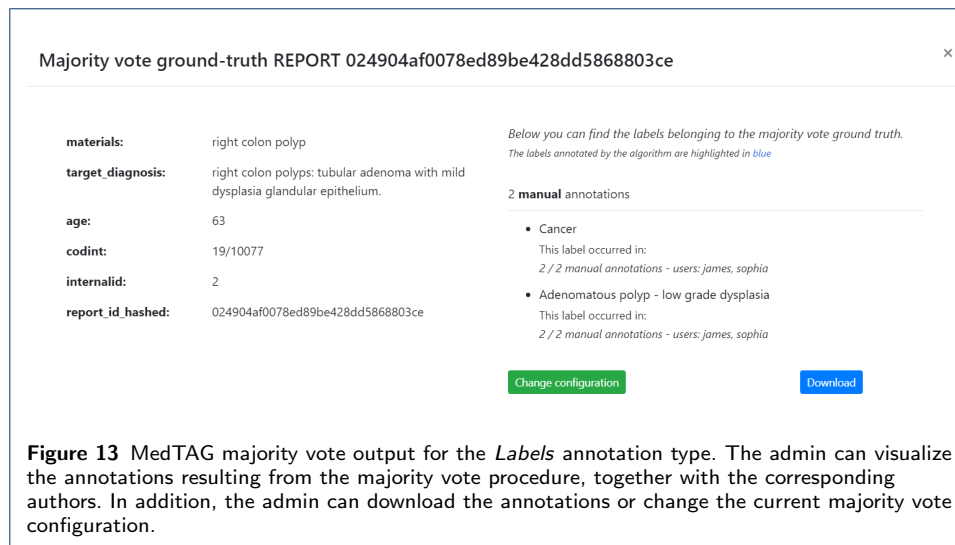
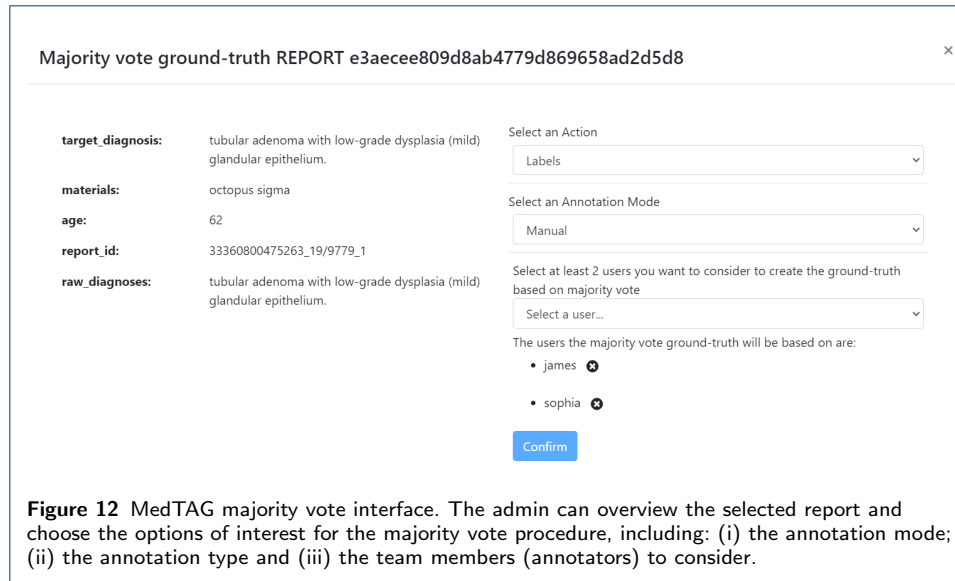
- **Delete:** enables the admin user to remove the corresponding reports.
- **Download:** allows the admin user to download either the original annotations or the ones resulting from the majority vote procedure. Moreover, the admin



user can also download the automatic annotations generated by the built-in prediction system. Several download options are provided, including the output file format, the annotation mode (i.e., manual or automatic) and type (i.e., *Labels*, *Concepts*, *Mentions* and *Linking*).

- **Inspect statistics:** allows the admin user to consult the report information as well as the statistics concerning the annotations of the selected report. The annotation statistics regards all the annotation types provided in MedTAG (i.e., *Labels*, *Concepts*, *Mentions* and *Linking*) and include the number of users that identified each label, mention or concept in the report. In addition to user annotations, the interface shows the automatic annotations highlighted in blue produced by the built-in prediction system.
- **IAA and majority vote:** allows the admin user to access the information concerning IAA for each report. Figure 12 shows the pop-up modal by which the admin can specify the options for the majority vote procedure. The admin can choose from a drop-down menu which team members (annotators) to consider, as well as the annotation mode and type. The procedure returns only the annotations that achieved more than fifty percent of agreement among the annotators considered. Then, the admin can download the annotations resulting from the majority vote procedure, as shown in Figure 13.

Figure 10 shows the *Team members' statistics* section of the control panel, which provides the information about the advancements in the annotation work for each team member. Access to this section is restricted to the admin user. The admin can overview the annotation work carried out for each use-case and annotation type using ring charts providing information about the number of annotated reports and the corresponding percentage out of the total. Moreover, Figure 11 shows that team members can keep track of their work by consulting the section *My statistics*, where other ring charts visually summarize the personal annotation statistics.



## Results and Discussion

MedTAG has been used to annotate diagnostic reports to produce both training and test annotated data. In particular, a specific instance of MedTAG (ExaTAG) has been used to generate more than seven thousand annotated reports and more than eight thousand annotations overall. ExaTAG<sup>[10]</sup> is an instance of MedTAG tailored for the histopathology domain. By connecting to ExaTAG, users can try MedTAG functionalities with real (anonymized) clinical reports from the digital pathology domain without downloading and installing it. ExaTAG has been customized to meet the needs of the physicians and experts concerning three cancer use-cases (i.e., cervix, colon, and lung cancer) within the ExaMode H2020 EU project<sup>[11]</sup>. Physicians and experts have used ExaTAG to annotate the diagnostic reports from two healthcare institutions, namely, the *Azienda Ospedaliera Cannizzaro, Catania*

<sup>[10]</sup><http://w3id.org/exatag/> access granted with username and password *Test* for reviewing purposes

<sup>[11]</sup><https://www.examode.eu/>

<b>Use-case</b> <b>Language</b>	Cervix cancer	Colon cancer	Lung cancer	<b>Total</b>
Dutch	-	889	-	889
English	2,361	-	-	2,361
Italian	1,828	239	2,005	4,072
<b>Total</b>	4,189	1,128	2,005	7,322

**Table 1** Number of diagnostic reports annotated per language and use-case.

<b>Use-case</b> <b>Annotation type</b>	Cervix cancer	Colon cancer	Lung cancer	<b>Total</b>
Labels	16,033	9,309	2,066	27,408
Concepts	12,936	11,932	2,336	27,204
Mentions	12,070	10,926	2,336	25,332
Linking	12,936	11,932	2,336	27,204
<b>Total</b>	53,975	44,099	9,074	107,148

**Table 2** Number of labels, concepts, mentions and links (mention - concept) automatically annotated per use-case.

(Italy) and the Radboud University Medical Center, Nijmegen, (The Netherlands). For the time being, ten annotators between physicians and experts have annotated thousands of medical reports in three languages (Dutch, English, and Italian). Table 1 reports some statistics about the manual annotation process conducted so far. Instead, Table 2 shows the number of automatic annotations done by SKET (i.e. the automatic annotation module) for each annotation type and use-case.

#### Biomedical annotation tools comparison

The biomedical annotation tools selected for the comparison, according to the five requirements presented above, are BioQRator [25], ezTag [26], MyMiner [27], tagtog [28] and TeamTat [35]. Moreover, we also consider brat [39] and INCEPTION [36] because they are used by the biomedical community in some settings. Figure 1 shows that several of the considered tools lack T6 (license allowance to modify and redistribute the tool) and F1 (support for overlapping mentions). Almost half of the tools (three out of seven) lacks T2 (availability of the source code), F2 (support for document-level annotation), F11 (support for IAA), F12 (data privacy) and F13 (multilingual support). In contrast, MedTAG satisfies: (T6) MedTAG is provided through the MIT license, permitting the use, modification and distribution of the tool free of charge; (T2) the source code of MedTAG is publicly available<sup>[12]</sup>; (F12) MedTAG enables the utilization of data on a local system without any sharing with external servers, thus ensuring data privacy; (F2) MedTAG supports two types of document-level annotations, namely, label and concept annotations. The label annotation feature allows the user to tag a document according to a customizable set of labels.

The concept annotation feature allows the users to mark a document as pertinent for one or more ontological concepts. Users can leverage the auto-complete feature to search for the relevant concepts to assign. Note that, as analyzed in [34], only a tiny

<sup>[12]</sup><https://github.com/MedTAG/medtag-core/>



minority of annotation tools on the market fully support document-level annotation. For instance, MyMiner supports document annotation, but due to limits in the customization process, the annotators must re-define the labels every time new documents are added to the system. Moreover, most of the other annotation tools allow the users to provide document-level annotations only using some workaround such as zero-width annotations and annotations of pre-defined placeholders placed at the beginning or at the end of the document to annotate. However, this practice is additional overhead that further complicates and slows down the annotation process; (F10) MedTAG supports users and roles.

MedTAG is distributed as a Docker container, thus it can seamlessly be deployed in a local environment or a remote cloud solution. Therefore, the administrator can choose whether to grant MedTAG access to annotators only within a local network or “worldwide”; (F13) MedTAG provides multilingual support. It allows the users to annotate the same document (same document identifier) in different languages.

When dealing with thousands of biomedical documents to annotate, time is crucial. Hence, web-based annotation tools provided with the modality of *Software as a service* (SaaS) are not necessarily the best solution in this context due to possible network delays. For instance, network delays might be experienced when uploading high volumes of data. A local installation can avoid network delays and operate better in the case of large corpora to be annotated. However, several annotation tools present difficulties about the installation process, such as lack of documentation or dependency issues, as stated in [34]. For instance, tagtog can be installed locally only in its commercial version, whereas ezTAG and TeamTat can be installed free of charge. Still, the procedure could be quite complex for the not technology-savvy; ezTag and TeamTat require the user to install and configure some frameworks and software packages manually (e.g., Ruby, Rails, and MySQL) as prerequisites. In contrast, MedTAG provides an easy installation procedure; it requires the user only to execute the `docker-compose up` command (provided that Docker is installed). The MedTAG installation procedure is available and thoroughly described online<sup>[13]</sup>.

Note that TeamTat provides high-level inter-annotator agreement statistics since the project manager can calculate the agreement among annotators. In contrast, MedTAG provides fine-grained statistics by allowing the users to access the information concerning IAA for each report and to download the annotations resulting from the majority vote procedure. For this reason, we consider the criterion (F11) partially satisfied by TeamTat (see Figure 1). TeamTat supports the annotation of documents compliant with the Unicode Standard, meaning that documents with special characters are visualized and annotated correctly. However, TeamTat does not provide additional facilities to manage, organize and search documents according to their languages (unless using a specific workaround such as creating language-specific document collections). In contrast, MedTAG allows the users to organize and filter documents according to their languages out-of-the-box; no additional configuration or effort is required. For this reason, we consider (F13) partially satisfied by TeamTat and entirely by MedTAG.

Several biomedical tools let the users upload biomedical documents by using tool-specific procedures and formats. For instance, BioQRator and ezTag only accept

---

<sup>[13]</sup><https://github.com/MedTAG/medtag-core/tree/main#installation>

medical documents in BioC format. Despite BioC being a well-established file format in the biomedical domain, adopting it as the only valid format poses hindrances to annotating biomedical documents in other formats. For instance, narrative clinical reports are usually available in an unstructured format, such as plain text. Thus, to use them in BioQRator and ezTag, they need to be converted in BioC format in advance. In contrast, MedTAG allows the users to provide the medical documents as customizable CSV files, letting the user decide and set up which fields to display and annotate. This feature turns out to be helpful, especially when dealing with high volumes of long biomedical documents, where changing data format is not always a feasible or reasonable operation for annotators.

For what concerns the general-purpose annotation tools - i.e., brat and INCEPTION - they are substantially different from MedTAG. For instance, brat [39] is a well-established web-based annotation tool specifically suited for entity and relationship annotations. It has been extensively used for the annotation of biomedical projects [41–45]. Brat is not available for online use; it requires to be installed locally in a UNIX-like environment. Hence, the procedure could be complex for not technology-savvy users, as stated in [34]. In contrast, MedTAG is provided as a portable and easy-to-run Docker container. Users can configure brat via plain-text schema configuration. Moreover, users can import raw documents and export the annotations in plain-text format. Conversely, MedTAG provides support for several file formats such as BioC/JSON and BioC/XML, which are standard formats for the annotations in the biomedical domain. In addition, MedTAG also provides several other features that brat currently does not support, as (T3) online availability; (F2) support for document-level annotation; and, (F6) integration with PubMed.

INCEPTION is another general-purpose tool used also by the biomedical community [46–49]. It is an open-source web-based annotation tool both available online and for local installation. For the local use, it requires Java, as described in the online documentation<sup>[14]</sup>. Figure 1 shows that INCEPTION covers most of the considered criteria (21 over 22). For instance, it provides active learning facilities to improve suggestions over time in a human-in-the-loop environment and a comprehensive set of features to adapt to different annotation scenarios. However, the INCEPTION interface provides several functionalities not specifically designed for the biomedical domain, which can be perceived as redundant by the biomedical community. Moreover, to achieve annotation flexibility, INCEPTION introduces additional levels of abstraction that increase the complexity of the annotation task, thus resulting potentially not within reach of not technologically-savvy users. For instance, document-level annotation is, at the time of writing, an experimental feature that needs to be explicitly enabled by manually editing a settings file. Moreover, to enable document-level annotations, the user must define a “Document metadata” annotation layer in the project settings. For such a reason, we judge the criterion (F2) as partially satisfied by INCEPTION (see Figure 1). In contrast, MedTAG provides document-level annotation facilities off-the-shelf since no additional configuration is required. In addition, MedTAG provides native PubMed integration facilities - i.e., users can annotate PubMed titles and abstracts – whereas INCEPTION employs a

---

<sup>[14]</sup><https://inception-project.github.io/documentation/>

third-party tool (i.e., PubAnnotation [50]) to retrieve the documents to annotate from PubMed Central, as stated in [51].

#### Quantitative comparison of biomedical annotation tools

To quantitatively assess MedTAG performance, we conducted several experiments designed to evaluate MedTAG concerning two annotation tasks: document-level annotation and mention identification. The first one concerns annotations that refer to the whole document, such as labels describing the overall document content (e.g., the “cancer” label may indicate whether a clinical report suggests a cancer condition). Instead, mention identification regards entity mentions identified in the textual content of a document. The annotation tools are compared regarding the number of actions and elapsed time required to complete the overall annotation process. To the best of our knowledge, this is the first available quantitative evaluation of biomedical annotation tools. The analysis we conducted considers a set of web-based biomedical annotation tools - i.e., ezTag, MedTAG, MyMiner, tagtog and TeamTat - evaluated on a sample of one hundred documents, randomly chosen from a real dataset concerning the digital pathology domain (i.e., clinical reports related to colon cancer). For the comparison, we consider only web-based publicly available tools since many biomedical annotation tools are not available for local installation or are not easy to install for not technologically savvy end-users. It is worth noting that our analysis does not focus on usability and *Human-Computer Interaction* (HCI) aspects (e.g., *User Experience* (UX) and interface look and feel) that may vary subjectively. Nevertheless, the latter are essential points that should be treated with specific user studies. In contrast, we focused on the annotation work regarding the number of actions (e.g., mouse clicks and keys pressed) and elapsed time to achieve the same annotations in different tools. To perform a fair comparison, we used automatic agents (web robots) designed to annotate using the same annotation speed - i.e., exact time to simulate a mouse click or a key pressed for each annotation tool. The automatic agents have been implemented using the Python Web automation library Selenium<sup>[15]</sup>. The source code of the automated agents used for the experiments is publicly available<sup>[16]</sup>. Since the automatic agents are generally way faster than any human annotator, we introduced a short delay (about two hundred milliseconds) between two consecutive actions, which is also necessary to avoid overloading the server with too many requests.

Table 3 and Table 4 show the experimental results in terms of the number of actions and elapsed time for annotating one hundred documents. The elapsed time for each tool was recorded forty times; the resulting mean value and standard deviation are reported in the tables. Table 3 shows the performance analysis concerning the document-level annotation task. For the latter task, we considered three tools - i.e., MedTAG, MyMiner, and tagtog - since ezTag does not support document-level annotation, whereas TeamTat provides different document-level annotation facilities. In particular, TeamTat allows us to annotate entities in different documents and then to create relationships between them; this is different from the functional

---

<sup>[15]</sup><https://www.selenium.dev/>

<sup>[16]</sup><https://github.com/MedTAG/medtag-core/tree/main/benchmark>

criterion (F2), indicating whether the users can specify labels at the document-level. For this reason, we consider the latter criterion only partially satisfied. The experiments concerning document-level annotation consist of assigning one label for each document to annotate. The labels, mentions, and documents used for testing are publicly available<sup>[17]</sup> for reproducibility purposes. Table 3 shows that MyMiner requires fewer actions than other tools to achieve the same annotations, whereas MedTAG turns out to be the fastest tool in terms of elapsed time. Nevertheless, the time difference between MyMiner and MedTAG is about ten seconds, which is negligible considering different server response times. According to Table 3, tagtog requires more actions and time than other tools to complete the annotation process. However, these results are motivated considering that tagtog is one of the most flexible annotation tools and allows to specify whether a document label is *true*, *false* or *unknown*. To this aim, tagtog allows users to choose the correct value from a drop-down menu for a document label. Thus, the users have to click on the drop-down menu two times: the first one to open the pop-up menu and the second for the value selection. In contrast, MyMiner and MedTAG require just one click on a checkbox, based on the assumption that a label may apply for a document or not (the *unknown* state is not allowed). Moreover, MyMiner requires fewer actions than MedTAG to complete the annotation process since it automatically moves on to the following document to annotate after the user selection. However, MyMiner does not allow to specify more labels for a document. In contrast, MedTAG goes beyond this limitation and allows to specify of several labels at the same time for each document. Thus, users can decide on their own when to move on to the next document to annotate.

Table 4 shows the performance analysis concerning the mention identification task. The experiments concerning mention identification consist of identifying entity mentions within the documents' textual content. To this aim, we used a set of pre-identified mentions for each of the documents considered. According to Table 4, the tools with the lowest number of actions required are ezTag and TeamTat, whereas MyMiner and MedTAG are the fastest tools in terms of elapsed time. TeamTat and ezTag achieved comparable performance since they are similar in terms of functionalities provided. The experimental results show that MyMiner is the fastest tool in terms of elapsed time. MyMiner provides a neat interface that requires low network resources and bandwidth to work, thus reducing loading time and making the annotation process faster. However, it lacks several functionalities such as (i) support for users and teams, (ii) availability for local installation, and (iii) data privacy (upload of the documents to annotate is required) that could be relevant for the needs of the biomedical community. In contrast, MedTAG is designed to be portable (i.e., local installation is available) and flexible; it provides annotation facilities, such as schema configuration, that allow users to customize the annotation experience. Moreover, MedTAG is faster than other tools, even if it requires more actions. A possible explanation could be the different mention annotation functionality. Indeed, most of the annotation tools allow identifying entity mentions within the text using drag-and-drop facilities. In contrast, MedTAG enables users to annotate mentions with a single click on each text token. The latter facility turns out

---

<sup>[17]</sup><https://github.com/MedTAG/medtag-core/tree/main/benchmark/datasets>

Tool	Number of actions	Elapsed time in seconds (mean)	Standard deviation in seconds
MedTAG	200	46.840	0.803
MyMiner	100	56.677	0.416
tagtog	400	205.740	5.471

**Table 3** Document-level annotation performance analysis in terms of number of actions (e.g. mouse clicks and keys pressed) and elapsed time required to complete the whole annotation process.

Tool	Number of actions	Elapsed time in seconds (mean)	Standard deviation in seconds
MedTAG	519	159.337	0.479
ezTag	307	260.340	0.576
MyMiner	414	114.390	1.507
tagtog	404	304.692	10.067
TeamTat	307	271.577	1.542

**Table 4** Mention-level annotation performance analysis in terms of number of actions (e.g. mouse clicks and keys pressed) and elapsed time required to complete the whole annotation process.

to be convenient in short mentions, whereas drag-and-drop is more suitable in the case of long ones.

To summarize, we quantitatively compared a set of web-based biomedical annotation tools on two tasks: document-level annotation (one label per document) and mention identification. We conducted several experiments to assess each annotation tool regarding the number of actions and elapsed time required to complete the overall annotation process. From the experimental results emerge that, depending on the task, some tools perform better than others. Despite the higher number of actions required to complete the annotation process, MedTAG turns out to be faster than other tools, especially for the document-level annotation task.

Finally, it is worth noting that the present study focuses on evaluating a set of biomedical annotation tools only on physical aspects such as the number of actions and elapsed time required to annotate all the documents considered. Hence, we do not consider several critical human-centric factors (e.g., UX and HCI) that should be investigated in dedicated usability studies.

## Conclusions

We presented MedTAG, a customizable, portable, collaborative, web-based biomedical annotation tool. We described an instance of MedTAG adopted in the histopathology domain, where MedTAG has been used by physicians to annotate more than seven thousand clinical reports in three languages (Dutch, English and Italian), from two health-care institutions. MedTAG is provided as a docker container to make it distributable, platform-independent and easy to install/deploy. We designed MedTAG according to the five requirements (i.e. available, distributable, installable, workable and schematic) defined in a recent extensive review of manual annotation tool [34]. Moreover, MedTAG satisfies 20 over 22 criteria defined in the same study.

The key points of strength of MedTAG are: (i) fast and easy installation because only one command is necessary to install it in less than 10 minutes on a current notebook; (ii) cross-platform support since MedTAG can be installed in every platform

supporting docker; (iii) a collaborative web-based platform supporting users and roles; (iv) broad data formats support including BioC/JSON, BioC/XML, CSV, and JSON; (v) support for schema configuration where the users provide the documents to annotate by using custom CSV files and can decide which fields to display and annotate.

### Limitations and Future Work

MedTAG, as the name suggests, is a customizable annotation tool for the biomedical domain; Thus, it is not intended for general-purpose annotations since the users could not exploit domain-specific features such as automatic annotation. It is worth noting that the automatic annotation is currently provided for three cancer use-cases (i.e., cervix, colon, and lung cancer). Nevertheless, we plan to extend the automatic annotation support for other use-cases according to the needs of the biomedical community. The integration of SKET as an automated annotation tool shows the flexibility of MedTAG and how annotation automation may work with MedTAG. Another limitation concerns the file format of the input documents since MedTAG currently supports only plain-text documents. We believe that PDF annotation would be particularly useful, especially when dealing with scientific paper annotation. Hence, we plan to include this feature in the future version of MedTAG. For the time being, MedTAG does not support both overlapping mentions (also known as multi-label annotations) and relationship annotations that are left as future work. Indeed, even if MedTAG allows assigning multiple concept labels to the same mention, it is currently impossible to annotate any sub-mention. Finally, it is worth noting that even if MedTAG is designed for the biomedical domain, it could also be used for general-purpose annotation as long as a suitable schema configuration is provided. As future work, we plan to enrich MedTAG by adding (i) the support for overlapping mentions; (ii) the support for relationship annotations; (iii) the support for active learning capabilities; (iv) the support for PDF annotation; (v) the automatic annotation support for other use-cases relevant for the biomedical community. Thereby, we aim to improve MedTAG according to the biomedical community's needs and foster further developments in this field.

### Availability and requirements

- Project name: MedTAG
- Project home page: <https://github.com/MedTAG/medtag-core>
- Operating system(s): Platform independent.
- Other requirements: Docker and docker-compose
- License: MIT License
- Any restrictions to use by non-academics: No

### Abbreviations

EHRs: Electronic Health Records; HCI: Human-Computer Interaction; IAA: Inter-Annotator Agreement; IE: Information Extraction; ML: Machine Learning; NCIT: National Cancer Institute Thesaurus; NER+L: Named Entity Recognition and Linking; NLP: Natural Language Processing; PMC: PubMed Central; SaaS: Software as a service; SKET: Semantic Knowledge Extractor Tool; UI: User Interface; UX: User Experience;

### Ethics approval and consent to participate

Not applicable

### Consent to publish

Not applicable

**Availability of data and materials**

The full source code of MedTAG is available at <https://github.com/MedTAG/medtag-core>. An instance of MedTAG (i.e. ExaTAG) is publicly available for testing it, using the following credentials: username and password *Test*.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the ExaMode Project, as a part of the European Union Horizon 2020 Program under Grant 825292. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Authors' contributions**

All the authors participated to the design of MedTAG. OI and FG co-developed MedTAG tool and wrote the documentation revised by GS. OI implemented the docker version of MedTAG. GS oversaw the implementation of the MedTAG tool. FG, OI and GS co-authored the first draft of the manuscript. All the authors discussed the manuscript at all stages. All the authors read and approved the final manuscript.

**Acknowledgements**

The authors wish to thank Stefano Marchesin for the work on SKET which has been integrated in this work.

**Author details**

Department of Information Engineering, University of Padua, Padua, Italy.

**References**

1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *Jama*. 2013;309(13):1351–1352.
2. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*. 2010;17(3):229–236.
3. Gorrell G, Song X, Roberts A. Bio-yodie: A named entity linking system for biomedical text. *arXiv preprint arXiv:181104860*. 2018;.
4. Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. *Journal of the American Medical Informatics Association*. 2018;25(5):530–537.
5. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*. 2010;17(5):507–513.
6. Kraljevic Z, Bean D, Mascio A, Roguski L, Folarin A, Roberts A, et al. MedCAT—Medical Concept Annotation Tool. *arXiv preprint arXiv:191210166*. 2019;.
7. Jovanović J, Bagheri E. Semantic annotation in biomedicine: The current landscape. *Journal of Biomedical Semantics*. 2017;8(1):1–18.
8. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, Ruch P, et al. Overview of BioCreative II gene normalization. *Genome biology*. 2008;9(2):1–19.
9. Li J, Sun Y, Johnson RJ, Sciaky D, Wei CH, Leaman R, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database*. 2016;2016.
10. Doğan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. *Journal of biomedical informatics*. 2014;47:1–10.
11. Krallinger M, Rabal O, Leitner F, Vazquez M, Salgado D, Lu Z, et al. The CHEMDNER corpus of chemicals and drugs and its annotation principles. *Journal of cheminformatics*. 2015;7(1):1–17.
12. Mohan S, Li D. Medmentions: a large biomedical corpus annotated with UMLS concepts. *arXiv preprint arXiv:190209476*. 2019;.
13. Nye B, Li JJ, Patel R, Yang Y, Marshall IJ, Nenkova A, et al. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In: *Proceedings of the conference. Association for Computational Linguistics. Meeting*. vol. 2018. NIH Public Access; 2018. p. 197.
14. Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I, et al. Building a semantically annotated corpus of clinical texts. *Journal of biomedical informatics*. 2009;42(5):950–966.
15. Campillos L, Deléger L, Grouin C, Hamon T, Ligozat AL, Névéal A. A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMS annotated Text corpus (MERLOT). *Language Resources and Evaluation*. 2018;52(2):571–601.
16. Ohta T, Pyysalo S, Tsujii J, Ananiadou S. Open-domain anatomical entity mention detection. In: *Proceedings of the workshop on detecting structure in scholarly discourse*; 2012. p. 27–36.
17. Kafkas S, Lewin I, Milward D, van Mulligen EM, Kors JA, Hahn U, et al. CALBC: Releasing the Final Corpora. In: *LREC*; 2012. p. 2923–2926.
18. Van Auken K, Schaeffer ML, McQuilton P, Laulederkind SJ, Li D, Wang SJ, et al. BC4GO: a full-text corpus for the BioCreative IV GO task. *Database*. 2014;2014.
19. Johnson AE, Pollard TJ, Shen L, Li-Wei HL, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Scientific data*. 2016;3(1):1–9.
20. Dobbie S, Strafford H, Pickrell WO, Fonferko-Shadrach B, Jones C, Akbari A, et al. Markup: A Web-Based Annotation Tool Powered by Active Learning. *Frontiers Digit Health*. 2021;3:598916. Available from: <https://doi.org/10.3389/fgdth.2021.598916>.
21. Searle T, Kraljevic Z, Bendayan R, Bean D, Dobson R. MedCATTrainer: A biomedical free text annotation interface with active learning and research use case specific customisation. *arXiv preprint arXiv:190707322*. 2019;.

22. Zvára K, Tomecková M, Peleška J, Svátek V, Zvárová J. Tool-supported interactive correction and semantic annotation of narrative clinical reports. *Methods of information in medicine*. 2017;56(03):217–229.
23. Bontcheva K, Cunningham H, Roberts I, Roberts A, Tablan V, Aswani N, et al. GATE Teamware: a web-based, collaborative text annotation framework. *Language Resources and Evaluation*. 2013;47(4):1007–1029.
24. Campos D, Lourenço J, Matos S, Oliveira JL. Egas: a collaborative and interactive document curation platform. *Database*. 2014;2014.
25. Kwon D, Kim S, Shin SY, Wilbur WJ. BioQRator: a web-based interactive biomedical literature curating system. In: *Proceedings of the Fourth BioCreative Challenge Evaluation Workshop*. vol. 1; 2013. p. 241–246.
26. Kwon D, Kim S, Wei CH, Leaman R, Lu Z. ezTag: tagging biomedical concepts via interactive learning. *Nucleic acids research*. 2018;46(W1):W523–W529.
27. Salgado D, Krallinger M, Depaule M, Drula E, Tendulkar AV, Leitner F, et al. MyMiner: a web application for computer-assisted biocuration and text annotation. *Bioinformatics*. 2012;28(17):2285–2287.
28. Cejuela JM, McQuilton P, Ponting L, Marygold SJ, Stefancsik R, Millburn GH, et al. tagtog: interactive and text-mining-assisted annotation of gene mentions in PLOS full-text articles. *Database*. 2014;2014.
29. Rak R, Rowley A, Black W, Ananiadou S. Argo: an integrative, interactive, text mining-based workbench supporting curation. *Database*. 2012;2012.
30. Pérez-Pérez M, Glez-Peña D, Fdez-Riverola F, Lourenço A. Marky: a tool supporting annotation consistency in multi-user and iterative document annotation projects. *Computer methods and programs in biomedicine*. 2015;118(2):242–251.
31. Rinaldi F, Clematide S, Schneider G, Romacker M, Vachon T. ODIN: an advanced interface for the curation of biomedical literature. *Nature Precedings*. 2010;p. 1–1.
32. Wei CH, Allot A, Leaman R, Lu Z. PubTator central: automated concept annotation for biomedical full text articles. *Nucleic acids research*. 2019;47(W1):W587–W593.
33. Müller HM, Van Auken KM, Li Y, Sternberg PW. Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. *BMC bioinformatics*. 2018;19(1):1–16.
34. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. *Briefings in bioinformatics*. 2021;22(1):146–163.
35. Islamaj R, Kwon D, Kim S, Lu Z. TeamTat: a collaborative text annotation tool. *Nucleic Acids Research*. 2020 05;48(W1):W5–W11.
36. Klie JC, Bugert M, Boullosa B, de Castilho RE, Gurevych I. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Association for Computational Linguistics; 2018. p. 5–9. Available from: <http://tubiblio.ulb.tu-darmstadt.de/106270/>.
37. Neves M, Leser U. A survey on annotation tools for the biomedical literature. *Briefings in bioinformatics*. 2014;15(2):327–340.
38. Yimam SM, Gurevych I, de Castilho RE, Biemann C. Webanno: A flexible, web-based and visually supported system for distributed annotations. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*; 2013. p. 1–6.
39. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. BRAT: a web-based tool for NLP-assisted text annotation. In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*; 2012. p. 102–107.
40. Klie J. INCEpTION: Interactive machine-assisted annotation. In: *Proc. of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems (DESIREs 2018)*. vol. 2167 of CEUR Workshop Proceedings. CEUR-WS.org; 2018. p. 105. Available from: <http://ceur-ws.org/Vol-2167/short8.pdf>.
41. Mihăilă C, Ohta T, Pyysalo S, Ananiadou S. BioCause: Annotating and analysing causality in the biomedical domain. *BMC bioinformatics*. 2013;14(1):1–18.
42. Zerva C, Batista-Navarro R, Day P, Ananiadou S. Using uncertainty to link and rank evidence from biomedical literature for model curation. *Bioinformatics*. 2017;33(23):3784–3792.
43. Kors JA, Clematide S, Akhondi SA, Van Mulligen EM, Rebholz-Schuhmann D. A multilingual gold-standard corpus for biomedical concept recognition: the Mantra GSC. *Journal of the American Medical Informatics Association*. 2015;22(5):948–956.
44. Campos D, Matos S, Oliveira JL. A modular framework for biomedical concept recognition. *BMC bioinformatics*. 2013;14(1):1–21.
45. Verspoor K, Jimeno Yepes A, Cavedon L, McIntosh T, Herten-Crabb A, Thomas Z, et al. Annotating the biomedical literature for the human variome. *Database*. 2013;2013.
46. Tutubalina E, Alimova I, Miftahutdinov Z, Sakhovskiy A, Malykh V, Nikolenko SI. The Russian Drug Reaction Corpus and neural models for drug reactions and effectiveness detection in user reviews. *Bioinform*. 2021;37(2):243–249. Available from: <https://doi.org/10.1093/bioinformatics/btaa675>.
47. Canales L, Menke S, Marchesseau S, D'Agostino A, del Rio-Bermudez C, Taberna M, et al. Assessing the Performance of Clinical Natural Language Processing Systems: Development of an Evaluation Methodology. *JMIR Med Inform*. 2021 Jul;9(7):e20492. Available from: <https://medinform.jmir.org/2021/7/e20492>.
48. Yi Y, Shen Z, Bompelli A, Yu F, Wang Y, Zhang R. Natural Language Processing Methods to Extract Lifestyle Exposures for Alzheimer's Disease from Clinical Notes. In: *2020 IEEE International Conference on Healthcare Informatics (ICHI)*; 2020. p. 1–2.
49. Schulz C, Meyer CM, Kiesewetter J, Sailer M, Bauer E, Fischer MR, et al. Analysis of Automatic Annotation Suggestions for Hard Discourse-Level Tasks in Expert Domains. In: *ACL*; 2019. .
50. Kim J, Wang Y. PubAnnotation - a persistent and sharable corpus and annotation repository. In: Cohen KB, Demner-Fushman D, Ananiadou S, Webber BL, Tsujii J, Pestian J, editors. *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, BioNLP@HLT-NAACL Montréal, Canada, June 8, 2012*. Association for Computational Linguistics; 2012. p. 202–205. Available from: <https://aclanthology.org/W12-2425/>.
51. de Castilho RE, Ide N, Kim JD, Klie JC, Suderman K. Towards cross-platform interoperability for



machine-assisted text annotation. *Genomics & Informatics*. 2019;17.