

GATE Teamware 2: An open-source tool for collaborative document classification annotation

David Wilby and Twin Karmakharm
Research Software Engineering Team

Ian Roberts, Xingyi Song and Kalina Bontcheva
Natural Language Processing Research Group
Department of Computer Science
The University of Sheffield

Abstract

We present GATE Teamware 2: an open-source web-based platform for managing teams of annotators working on document classification tasks. GATE Teamware 2 is an entirely re-engineered successor to GATE Teamware, using contemporary web frameworks. The software allows the management of teams of multiple annotators, project managers and administrators - including the management of annotators - across multiple projects. Projects can be configured to control and monitor the annotation statistics and have a highly flexible JSON-configurable annotation display which can include arbitrary HTML. Optionally, documents can be uploaded with pre-existing annotations and documents are served to annotators in a random order by default to reduce bias. Crucially, annotators can be trained on applying the annotation guidelines correctly and then screened for quality assurance purposes, prior to being cleared for independent annotation. GATE Teamware 2 can be self-deployed, including in container orchestration environments, or provided as private, hosted cloud instances. GATE Teamware 2 is an open-source software and can be downloaded from Github.¹ A demonstration video of the system has also been made available.²

1 Introduction

Machine learning models are an important element of modern natural language processing (Cunningham et al., 2013; Otter et al., 2021). These models need to be trained and evaluated on gold-standard human-annotated datasets. Depending on the purpose of the model, there are two types of annotations: *classification* and *chunking*. Chunking tasks (e.g. named entity recognition) require the annotators to identify the subset (or span) of the given text, while classification tasks (such as sentiment and

stance classification) require annotators to assign labels to the given text. GATE Teamware 2 focuses on classification annotation tasks specifically.

Depending on text length or specific machine learning tasks, the classification annotation could vary from word level to document level, or single-label to multi-label annotation. In addition, other information (e.g. confidence score for the annotations) may also be needed. This requires annotation tools to be configurable and flexible to adapt to different classification annotation tasks.

Furthermore, machine learning model performance is highly dependent on annotation quality and as such, large annotation tasks often hire multiple annotators to collaborate to remedy this. Personal biases may be unknowingly introduced into annotations, so being able to collaborate with numerous annotators is a way to reduce the effect of this (Hovy and Prabhumoye, 2021).

Here we introduce GATE Teamware 2 (Karmakharm et al., 2022), a free open-source software tool for the management of teams of human annotators engaged in annotating datasets for document level classification, referred to here simply as document classification tasks. GATE Teamware 2 provides a highly flexible annotation configuration mechanism. Annotators can participate in more than one project. In order to ensure high quality human annotations, GATE Teamware 2 also provides functionality for training and testing annotators on a subset of example documents, prior to qualifying them to perform the given annotation task independently. GATE Teamware 2 is the spiritual successor to the popular GATE Teamware (Bontcheva et al., 2013), but has been entirely re-implemented in modern web frameworks with a containerised deployment architecture, as well as extended with new functionalities.

Teamware 2 has a number of unique features for an application specifically targeted at creating datasets for document classification tasks, includ-

¹<https://github.com/GATENLP/gate-teamware>

²<https://youtu.be/KoXkuhc4fmM>

ing:

- defined user roles and permissions,
- providing automatically ‘pre-annotated’ data, which the human annotators correct, rather than annotate from scratch, and
- optional training and testing qualification stages for annotators.

As a demonstration of its utility, Teamware 2 has already been used successfully for the annotation of several datasets consisting of tens of thousands of documents by numerous annotators per project on research topics such as COVID vaccine hesitancy on Twitter, stance detection in tweets (Figure 1) and emotion recognition.

2 Related Work

There is an abundance of software tools facilitating document annotation serving a variety of concerns in this area. [Neves and Ševa \(2020\)](#) collated a searchable directory of over 90 such tools and their features. In this section, the focus will be on the review of tools which can perform document classification while also offering collaborative features that allow multiple users to annotate the same set of documents.

According to [Neves and Ševa \(2020\)](#) and additional web searches, four annotation tools are available, which can:

- facilitate collaborative corpus annotation in document classification NLP tasks;
- permit management of multiple users;
- are actively maintained;
- and are available for general use.

These tools are [doccano \(Nakayama et al., 2018\)](#), [tagtog \(Cejuela et al., 2014\)](#), [Universal Data Tool \(Ibarluzea et al., 2022\)](#), and [Label Studio \(Tkachenko et al., 2020-2022\)](#).

These four tools all allow self-hosting of the software, enable multiple annotators to annotate documents through a web interface on a single hosted instance, and support a wide variety of annotation tasks out of the box in addition to document classification. However, they vary widely in availability of collaborative features and handling of document classification tasks.

The screenshot shows a web-based annotation interface. At the top, there is a text input field with a placeholder: "@mention this is where the tweet will appear." Below this is a section titled "Hostility categorisation" with a sub-instruction: "Please select the hostility type". It contains two radio buttons: "Hostile" and "Not Hostile". The next section is "Hostility Label Confidence" with the instruction: "Please select your confidence". It features five radio buttons labeled 1 through 5, with the button for '3' selected. Below this is a section titled "Comment on what made you uncertain" with a note: "NOTE: Please fill this if you selected confidence 3 or below". It contains a text input field. The following section is "Identity Characteristic Selection" with the instruction: "Please select the target identity characteristic(s)". It has four checkboxes: "Religion", "Gender", "Race/Ethnicity", and "None of the above". The next section is "Identity Characteristic Confidence" with five radio buttons labeled 1 through 5, with the button for '3' selected. Below this is another "Comment on what made you uncertain" section with the same note and a text input field. At the bottom, there are two buttons: "Submit" (green) and "Clear" (yellow).

Figure 1: An example of Teamware 2’s annotation interface, customised for the task of hostile tweet response classification. Actual tweet is not shown due to its sensitive nature. The application is able to capture multiple groups of labels in a single annotation using standard HTML form components (radio buttons, checkboxes, drop-down selectors and text boxes). Documents can be fully styled through HTML and CSS to suit various tasks.

In terms of suitability for document classification, while tagtog (Cejuela et al., 2014) can perform document classification as it allows document-wide labels to be captured, its focus is on sequence labelling tasks, and this is reflected in the user interface where these document-wide labels are only located in the sidebar, making it unsuitable for use especially when document classification is the only annotation task. Both Universal Data Tool (Ibarluzea et al., 2022) and doccano (Nakayama et al., 2018) have recipes for performing document classification, but both are only able to capture a single category; expect a fixed document layout; and have no options to customise the presentation of the text to be annotated by the user.

Label Studio (Tkachenko et al., 2020-2022) is perhaps the most similar tool to GATE Teamware 2 in that it provides a fine-grained way to customise the number of labels, their layout, and how the text to be annotated can be formatted and presented to the annotator.

In terms of collaborative features, Universal Data Tool (Ibarluzea et al., 2022), unlike the other tools, does not have user accounts and collaborators can join an annotation project through a generated link which would make the tool more suitable for smaller and shorter annotation projects. The free version of Label Studio (Tkachenko et al., 2020-2022) does not allow multiple users with clearly-defined user roles – all users are able to add and modify documents and annotations freely, though this may be different under a paid subscription. Finally, tagtog (Cejuela et al., 2014) and doccano (Nakayama et al., 2018) are the only two applications that have clear definition of user roles and limit annotators to performing annotation tasks, preventing them from modifying annotation project settings.

With these four tools, the expectation is for annotators to annotate every document in the project. With larger annotation projects that can contain thousands of documents, in order to limit the number of documents that have to be annotated by each individual annotator, it becomes necessary to split the corpus into multiple projects, recruiting different sets of annotators to each one. GATE Teamware 2 solves this problem by allowing limits to be placed on the number of annotations a document requires and the maximum proportion of the corpus a single annotator can annotate. This means annotation tasks can instead be automati-

cally distributed to annotators within the project whilst avoiding the problem of annotators having too many annotation tasks.

Relation to GATE Teamware

As stated above, GATE Teamware 2 is intended as a successor to GATE Teamware version 1 (Bontcheva et al., 2013). Many of the design concepts in GATE Teamware 2 originate in the original GATE Teamware, for example the management of teams of human annotators, their assignment to annotation projects, monitoring of the average time-per-document for each annotator, and the option for the project manager to provide “pre-annotations” for the annotators to correct, rather than having them annotate every document from scratch. However, the original GATE Teamware was designed around the annotation of *spans* of characters within text and is poorly suited to more general document classification tasks of the sort required for modern ML models. In GATE Teamware, annotation schemas can be provided to control the sets of labels and features that annotators may assign, but every label must be tied to a specific span within a longer document, which increases the cognitive load on annotators.

There are other desirable features missing from GATE Teamware, notably:

- Annotator training and validation – GATE Teamware included tools to facilitate resolution of disagreements among annotators or between a single annotator and a gold standard annotation set, but there is no real-time feedback to the annotators on how well they are following the annotation guidelines and no way to automatically check that annotators have reached a certain quality threshold before starting on a real project.
- Limiting the number of documents each annotator can consume – GATE Teamware ensured that each document is annotated by a set number of *annotators*, but cannot ensure that each annotator sees the same number of *documents*. An annotator will continue to be assigned documents until either all documents have been fully annotated, or they have been presented with every available document. In many cases, it is desirable to better balance the documents across annotators.

In addition, GATE Teamware was built as a Java

web application using a complex stack of software libraries, while the annotator interface was derived from the annotation editor of GATE Developer, a desktop Java Swing application. The interface was delivered to annotators as a Java Web Start application³, but this technology ceased to be supported by Oracle with the release of Java 11 in 2018, and is very difficult to operate in modern browsers. One of the key motivations for starting from scratch when implementing GATE Teamware 2 was to be able to use a more modern and performant development framework and purely browser-based user interface, vastly simplifying both server-side deployment and annotator on-boarding.

3 GATE Teamware 2 Overview

GATE Teamware 2 provides a database-backed web application for managing groups of annotators and providing document classification annotations. It is written in the modern web frameworks [Django](#) and [Vue.js](#) with a [PostgreSQL](#) database.

A Teamware 2 instance is set up by an administrator for a set group of users who can act as project managers and/or annotators. Project managers can create projects, which represent annotation tasks and are configured for the annotation of a single corpus of documents. Annotators can then be added to a project to create annotations according to the guidelines and configuration for that project.

4 User Roles

There are 3 user roles in Teamware 2:

Annotator: Can only annotate via the annotation view, when assigned to a project with annotation tasks available to them. Once a task is complete, they must contact a manager to be added to any new projects. This is the default role assigned to newly registered users. See [section 6](#).

Manager: Can annotate as the annotator above. Also has the ability to create new projects and edit all existing projects on the instance. See [section 5](#).

Admin: As roles above; also has total responsibility over an entire Teamware 2 instance, including the ability to promote annotator users to manager, and manage all users.

5 Annotation Projects

Annotation in GATE Teamware 2 is organised into ‘Projects’, each of which has a manager, who owns

³<https://docs.oracle.com/javase/8/docs/technotes/guides/javaws/>

the project.

Managers create projects simply through a web interface which populates the configuration form with default options. Alternatively, an existing project can be ‘cloned’ to create a new project with no documents or annotators, but an identical configuration.

Projects are configured to include a description and annotator guidelines, both of which can contain text formatted as markdown or HTML and can be previewed by the project manager. Annotation parameters are set at this stage to include: the total number of annotations required per document; the maximum proportion of documents in the corpus that any one user can annotate; and how long an annotator assigned a document has to annotate it before it is returned to the pool of documents. In the interface, each of these options has concise inline help text to remove ambiguity and streamline the project setup process.

5.1 Configuring annotation display and collected labels

Configuring how annotations will be presented and the details of collected labels is performed with a highly flexible JSON format in which elements are defined as JSON objects (see [Figure 2a](#)). Detailed documentation and examples are provided. Annotations can include HTML, fields or columns from the document data (see [subsection 5.2](#)), widgets such as radio buttons, free text fields, checkboxes, and drop-down menus. The result of the configuration JSON is previewed alongside its input in the form (e.g. [Figure 2c](#)), allowing managers to experiment with the best display for their project.

5.2 Documents

Under the ‘Documents & Annotations’ tab of the project management page, a corpus can be uploaded for annotation via single or multiple file upload, or by dragging and dropping files. Documents can be provided as JSON or CSV, and users can set a global preference for how they view documents for either one of these formats. Once documents are uploaded to the project, they are shown in a searchable list along with annotation statistics including how many annotations for each document are complete, aborted, rejected, timed out, and pending.

```
[
  {
    "name": "htmldisplay",
    "text": "{{text}}",
    "type": "html"
  },
  {
    "name": "sentiment",
    "type": "radio",
    "title": "Sentiment",
    "options": {
      "neutral": "Neutral",
      "negative": "Negative",
      "positive": "Positive"
    },
    "description": "Please select a
      ↪ sentiment of the text above."
  }
]
```

(a) Example project configuration JSON.

```
{
  "text": "<p>Document text can be
    ↪ <strong>arbitrarily
    ↪ styled</strong> using
    ↪ HTML.</p><p>Second
    ↪ paragraph.</p>"
}
```

(b) Example document JSON containing styling HTML.

Document text can be **arbitrarily styled** using HTML.

Second paragraph.

Sentiment

Please select a sentiment of the text above.

Neutral Negative Positive

(c) Example annotation display produced by configuration in Figure 2a and document text field in Figure 2b.

Figure 2: Example simple project configuration and document text with the resulting annotation view.

5.3 Pre-Annotation

A project can be configured to look for a field in the document that contains pre-annotation information. At the annotation stage, the annotation display will be pre-filled if the configured pre-annotation field exists in the document and contains data in the same layout as the expected annotation output. An example annotation output can be obtained by filling out the annotation display preview in the project configuration page.

5.4 Managing a Project

Currently available annotators can be added from a pool of those available, using a list of annotators which is searchable by email and username. Once added to the project, user status can be monitored with respect to completion of training and testing stages (optional) and annotations. Annotators can then be marked as complete or rejected from a project, as well as made active again in a project, provided that they have not met their quota of annotations. These user management actions can also be performed in batch mode to save time.

The status of documents and their annotations can be monitored throughout the project, including icons showing the number of annotations and their status. Annotations can also be edited by project managers. Similarly, annotator status is shown in the ‘Annotators’ tab.

5.5 Training & Testing Annotators

Training and testing stages can be enabled for any project to ensure quality in annotations. Training mode allows project managers to supply interactive examples for annotators to complete, prior to participating in a project. Similarly, a testing stage can be enforced with a threshold of correct answers that an annotator must achieve before being allowed to annotate. For both options, a lightweight JSON format, similar to the main project documents, is used, with an additional field indicating the correct answers.

Depending on choice at the project configuration stage, annotators can be automatically advanced to the annotation stage once they have successfully scored over the pass threshold or, from the project management screen, be approved by a project manager or admin.

5.6 Collected Annotations

Annotations can be exported as CSV, JSON or JSONL format, of which the latter two can be either formatted in the same style as uploaded JSON documents or in the format used by GATE⁴. The project configuration (section 5 & Figure 2a) also defines how the annotations are exported.

By default, each exported annotation contains a unique numerical key associated to an annotator instead of the annotator's username. This is to ensure anonymisation when sharing the dataset, while at the same time allowing cross-referencing of annotators between exported datasets from the same Teamware 2 instance. An option is available to export with usernames instead if required.

6 Annotating with GATE Teamware 2

Once logged in, annotators are only able to view the annotation view, which presents them with the next annotation task assigned to them. Depending on the project configuration, there are several options made available; such as rejecting an annotation, or adding a training or test stage to the project. Once there are no more documents available, they are instructed to contact the project manager to be assigned to a new project (if applicable).

During an annotation task, annotators are able to go back to the previous annotation to make a correction. Annotations can also be changed at any stage by visiting the 'My Annotations' view which displays all of an annotator's annotations. A complete history of all annotations and changes made is kept for posterity.

7 Using GATE Teamware 2

GATE Teamware 2 is available under an open source [AGPL license](#) and as such is free to use and deploy. However, subject to arrangements, private hosted instances can be provided for NLP projects⁵ as a low-overhead way to set up an instance of the software. For hosting your own instance, GATE Teamware 2 has been designed so users can deploy via a container orchestration tool such as [Kubernetes](#) or [Docker Compose](#), for instance on a cloud service. Deployment options can be chosen by editing a single text file and without changing the

⁴<https://gate.ac.uk/userguide/sec:creole:gatejson>

⁵Please contact the team at <https://gate.ac.uk/g8/contact> to arrange.

source code, enabling easy annotation task management.

8 Software Quality Assurance

Quality and consistency of GATE Teamware 2's source code is ensured via extensive software testing using [pytest](#) for the back end unit tests, [jest](#) for the front end unit tests and [Cypress](#) for integration testing. Testing currently covers 85% of the codebase. The software is under active development and uses a continuous integration approach using [GitHub Actions](#) to run tests on all changes to the source code, therefore reducing breaking changes to functionality.

9 Conclusions & Future work

GATE Teamware 2 is a fast and simple to use platform for facilitating collaborative document annotation, and includes a number of important features such as pre-annotation upload, flexible annotation configuration and multiple user management.

The software is under active development and a number of valuable features will be coming to the software soon. For instance, in addition to document classification, sequence labelling is planned to be integrated in the future, allowing it to be used in an even greater range of NLP annotation tasks. Other planned future features include: calculation of inter-annotator agreement metrics, ability to define hierarchical/dependent annotations, exposing an annotation management API for integration with active learning algorithms, control over project visibility and ownership, and availability of the platform as a cloud service.

Limitations

The major current limitations of GATE Teamware 2 align closely with the future work above and so many of its limitations will be resolved in the short to medium term. Primarily, Teamware 2 is designed for document classification tasks and so at present cannot be used for chunk annotation. There is no facility yet for calculation of inter-annotator agreements within the application, though these can be simply calculated from the exported annotation data. Similarly, Teamware 2 does not yet have web-based annotation adjudication capabilities and annotation quality can only be evaluated based on exported data. Hierarchical annotations, *i.e.* nested options which depend on the initial annotation, are not yet available. As Teamware 2 is designed so

far, all projects are visible and editable by all other managers and admins on the instance, however this is something that may become an option to change in the future. In the mean time if such separation is required (e.g. to protect data from one project being seen by managers of another), a workaround is possible by having the different annotation projects run on separate GATE Teamware 2 deployments.

Ethics Statement

The development of GATE Teamware 2 as an annotation tool was in line with the ethical clearances of the respective research projects that provided the funding. Ethical approvals for each annotated dataset and the respective volunteer annotators recruited for that project may need to be sought in addition, in line with the ethical guidelines of the user's institution and any data protection and privacy laws that apply to the managing organisation and to the annotators if they are located in a different jurisdiction. For example, for the authors' own annotation projects, individual ethical approvals have been obtained via the standard University of Sheffield research ethics process, and it is envisaged that similar processes would be followed by users at other institutions.

A limited amount of personally identifiable information, namely an email address and username, is collected from each annotator who registers an account with a given installation of GATE Teamware 2. This is made clear to users in a privacy policy prior to registration, and provision is made to remove all personal identifiers from the user's account if they choose to withdraw from participating in annotation projects on that installation. The annotations they have performed so far are not necessarily deleted⁶, and will remain linked to the dormant account, so that it is still possible to determine whether disparate sets of annotations were created by the same individual, but the username and email address on their account profile are replaced by anonymous placeholders so the account is no longer linked to an identifiable person.

Acknowledgements

This work has been partially supported by a European Union – Horizon 2020 Program,

⁶Whether a user's pre-existing annotations must be deleted is a matter for the ethical approval of the particular project or projects they are working on, not something enforced at the platform level.

grant no. 824091 (RISIS 2), the European Union Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042("SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics"(http://www.sobigdata.eu)) and a UKRI grant no. EP/W011212/1.

References

- Kalina Bontcheva, Hamish Cunningham, Ian Roberts, Angus Roberts, Valentin Tablan, Niraj Aswani, and Genevieve Gorrell. 2013. *GATE teamware: a web-based, collaborative text annotation framework*. *Language Resources and Evaluation*, 47(4):1007–1029.
- Juan Miguel Cejuela, Peter McQuilton, Laura Ponting, Steven J. Marygold, Raymund Stefancsik, Gillian H. Millburn, Burkhard Rost, and the FlyBase Consortium. 2014. *tagtog: interactive and text-mining-assisted annotation of gene mentions in PLOS full-text articles*. *Database*, 2014. Bau033.
- Hamish Cunningham, Valentin Tablan, Angus Roberts, and Kalina Bontcheva. 2013. *Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics*. *PLOS Computational Biology*, 9(2):e1002854. Publisher: Public Library of Science.
- Dirk Hovy and Shrimai Prabhumoye. 2021. *Five sources of bias in natural language processing*. *Language and Linguistics Compass*, 15(8):e12432.
- Severin Ibarluzea et al. 2022. *Universal data tool*. <https://github.com/UniversalDataTool/universal-data-tool>.
- Twin Karmakharm, David Wilby, Ian Roberts, and Kalina Bontcheva. 2022. *GATE Teamware*. Open source software available from <https://github.com/GATENLP/gate-teamware>.
- Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. *doccano: Text annotation tool for human*. Software available from <https://github.com/doccano/doccano>.
- Mariana Neves and Jurica Ševa. 2020. *Annotation-saurus: A searchable directory of annotation tools*.
- Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. 2021. *A Survey of the Usages of Deep Learning for Natural Language Processing*. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):604–624. Conference Name: IEEE Transactions on Neural Networks and Learning Systems.
- Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. 2020-2022. *Label Studio: Data labeling software*. Open source software available from <https://github.com/heartexlabs/label-studio>.