

SLATE: A Super-Lightweight Annotation Tool for Experts

Jonathan K. Kummerfeld

Computer Science & Engineering

University of Michigan

Ann Arbor, MI, USA

jkummerf@umich.edu

Abstract

Many annotation tools have been developed, covering a wide variety of tasks and providing features like user management, pre-processing, and automatic labeling. However, all of these tools use Graphical User Interfaces, and often require substantial effort to install and configure. This paper presents a new annotation tool that is designed to fill the niche of a lightweight interface for users with a terminal-based workflow. SLATE supports annotation at different scales (spans of characters, tokens, and lines, or a document) and of different types (free text, labels, and links), with easily customisable keybindings, and unicode support. In a user study comparing with other tools it was consistently the easiest to install and use. SLATE fills a need not met by existing systems, and has already been used to annotate two corpora, one of which involved over 250 hours of annotation effort.

1 Introduction

Specialised text annotation software improves efficiency and consistency by constraining user actions and providing an effective interface. While current annotation tools vary in the types of annotation supported and other features, they are all built with direct manipulation via a Graphical User Interface (GUI). This approach has the advantage that it is easy for users who are not computer experts, but also shapes the design of tools to become large, complex pieces of software that are time-consuming to set up and difficult to modify.

We present a lightweight alternative that is not intended to cover all use-cases, but rather fills a specific niche: annotation in a terminal-based workflow. This goal guided the design to differ from prior systems in several ways. First, we use a text-based interface that uses almost the entire screen to display documents. This focuses attention on the data and means the interface can easily scale to assist vision-impaired annotators. Second,

we minimise the time cost of installation by implementing the entire system in Python using built-in libraries. Third, we follow the Unix Tools Philosophy (Raymond, 2003) to write programs that do one thing well, with flat text formats. In our case, (1) the tool only does annotation, not tokenisation, automatic labeling, file management, etc, which are covered by other tools, and (2) data is stored in a format that both people and command line tools like `grep` can easily read.

SLATE supports annotation of items that are continuous spans of either characters, tokens, lines, or documents. For all item types there are three forms of annotation: labeling items with categories, writing free-text labels, or linking pairs of items. Category labels are easily customisable, with no limit on the total number and the option to display a legend for reference. All keybindings are customisable, and additional commands can be defined with relatively little code. There is also an adjudication mode in which disagreements are displayed and resolved.

To compare with other tools we conducted a user study in which participants installed tools and completed a verb tagging task in a 623 word document. When using SLATE, participants finished the task in 13 minutes on average, with more than half spending 3 minutes or less on setup.

Two research projects have used SLATE for annotation: token-level classification of 25,624 tokens of cybercriminal web forum data (Portnoff et al., 2017), and line-level linking of 77,563 messages of chat data (Kummerfeld et al., 2019).

2 System Description

Rather than specifying a task, such as named entity recognition, SLATE is designed to flexibly support any annotation that can be formulated as one of three annotation types: applying categorical labels, writing free text, or linking portions of text. These can be applied to items that are single docu-

```

LSB- H -RSB- Moneybookers -LSB- W -RSB- Paypal

I got like $ 7 Moneybookers .
I need it exchanged to PP on 1:1 rate .
Pm me .

-----
file 1 / 1
want SPACE+a | have SPACE+s | rate SPACE+d
example_label SPACE+q | another_example SPACE+w
Current: have

```

```

[19:58] <wafflejock> corba, well depends on where on the local disk you're
trying to save, if it's your home folder should be fine
[19:58] <corba> MMM, i tried with NTFS and ext4
[19:58] <wafflejock> corba, someone here suggests thunar-shares-plugin but
from 2010 https://ubuntuforums.org/showthread.php?t=1582665
[19:59] <MMM> oh wait... heard you wrong. You are booted to xubuntu and
are trying to get to windows right? how did you set up samba?
[19:59] <corba> wafflejock, nope, its a separate disk
[19:59] <corba> wafflejock, ill look into that thanks
[19:59] <MMM> did you write your share into /etc/samba/smb.conf?
[19:59] <corba> MMM, i didn't write the share, shouldnt i only have to do
that if i was trying to share from xubuntu?
[20:01] <corba> i also tried with gigolo
[20:01] <MMM> oh jeeze: I keep crossing what you are saying iun my head. I
got it mixed up again and was right the first time. I dont know how to get
Windows from xubuntu... only xubuntu from windows. sorry
[20:02] <corba> no prob :)
=== nicomach2s is now known as nicomachus
[20:04] <corba> ive done it a thousand times in cinnamon but this is an
old computer and i have to use xfce or lxde, i dont know if i can run
mate...
[20:04] <wafflejock> ah yeah the shares plugin appears to be for adding
shared folders, might be useful but not for this issue
[20:05] <corba> not only that, im having problems downloading it form the

```

(a) Annotation of a cybercriminal forum post. The underlined token is the one currently selected. The mapping from colours to labels and the keys to apply them are indicated by the legend at the bottom. Two extra labels were added for this picture, to show how the legend will wrap as needed. The information about progress, the legend, and the current item can be hidden if desired.

(b) Adjudication of a linking task annotating reply-to relations on IRC. The current message in green, an antecedent that is agreed on is dark blue, one that is not agreed on is light blue, other messages with disagreements are red, and the underline indicates a message that is being considered for linking to the green message (incorrectly in this case).

Figure 1: Screenshots of terminal windows with SLATE running. We have intentionally used two different font sizes to show how the user interface can scale, which is helpful for visually impaired users.

ments, lines, tokens, or characters, and continuous spans of lines, tokens, and characters. For example, from this perspective, NER is a task with categorical labels over continuous token spans.

The tool also supports adjudication of annotation disagreements. Multiple sets of annotations are read in and compared to determine disagreements to be displayed to the user, which can then be resolved, producing a new annotation file.

In the process of describing the tool, we will refer to two datasets that it was used to annotate:¹

(1) Portnoff et al. (2017) studied cybercriminal web forums. Experts in computer security and NLP collaborated to label posts in which the user is trying to exchange money from one form into another. For each post, we labeled tokens that expressed what was being offered, what was being requested, and the rate. Three annotators labeled a set of 600 posts containing 25,624 tokens. 200 of the posts were triple annotated and disagreements were adjudicated using the tool.

(2) Kummerfeld et al. (2019) developed a new conversation disentanglement dataset an order of magnitude larger than all previously released datasets combined. 77,563 messages were annotated with links indicating the message(s) they were a response to. For 11,100 messages, multiple annotations were collected and adjudicated using the tool. Annotations took 7 to 11 seconds per

message depending on the complexity of the discussion, and adjudication took 5 seconds per message (lower because many messages did not have a disagreement, but not extremely low because those that did were the difficult cases). Overall, annotators spent approximately 240 hours on annotation and 15 hours on adjudication.

Display The interface is text-based and contained entirely within a terminal window. By default, the entire terminal area displays the text being annotated, as shown in Figure 1b. There are also options to use space to display information about the current annotations, as shown in Figure 1a.

Colour is used to indicate annotations. For categorical labels a different colour is used for each label, and a special colour is used to indicate when multiple labels have been assigned to the same item. For linking, one colour is used to indicate the item being linked, another is used to indicate what items are currently linked to it, and a third is used to indicate any item that has a link.

The most frequently needed visual customisation is changing the colours used to indicate labels when assigning categories. These colours are specified in a simple text file. More substantial changes, such as a different colour to indicate disagreements, requires changing the code, but colours for each situation are defined in one place with intuitive names.

Input All interaction occurs via the keyboard. Commands are designed to be intuitive, e.g. arrow

¹ No formal user satisfaction surveys were conducted, but the tool received positive feedback on both projects.

System	Annotation Types	Adjudication	External Dependencies	Programming Language	User Interface
SLATE	Classify, Link	Yes	-	Python	Terminal
brat (Stenetorp et al., 2012)	Classify, Link	Yes	apache	Python, Javascript	GUI
GATE (Bontcheva et al., 2013)	Classify, Link	Yes	-	Java	GUI
YEDDA (Yang et al., 2018)	Classify	-	-	Python	GUI
ANALEC (Landragin et al., 2012)	Classify, Link	-	-	Java	GUI
Anafora (Chen and Styler, 2013)	Classify, Link	Yes	-	Python	GUI
CAT (Lenzi et al., 2012)	Classify	-	apache	Java	GUI
Chooser (Koeva et al., 2008)	Classify, Link	-	-	C++, Python, Perl	GUI
CorA (Bollmann et al., 2014)	Classify	-	-	PHP, JavaScript	GUI
Djangology (Apostolova et al., 2010)	Classify	Yes	Django	Python	GUI
eHost (South et al., 2012)	Classify, Link	Yes	-	Java	GUI
Glozz (Widlöcher and Mathet, 2012)	Classify, Link	Yes	-	Java	GUI
GraphAnno (Gast et al., 2015)	Classify, Link	-	-	Ruby	GUI
Inforex (Marcinczuk et al., 2012)	Classify, Link	-	-	JavaScript	GUI
Knowtator (Ogren, 2006)	Classify, Link	Yes	Protégé	Java	GUI
MAE and MAI (Stubbs, 2011)	Classify, Link	Yes	-	Java	GUI
MMAX2 (Müller and Strube, 2006)	Classify, Link	-	-	Java	GUI
PALinkA (Orasan, 2003)	Classify, Link	-	-	Java	GUI
SAWT (Samih et al., 2016)	Classify	-	-	Python, PHP	GUI
SYNC3 (Petasis, 2012)	Classify	Yes	Ellogon	C	GUI
Stanford (Lee, 2004)	Classify	-	-	Java	GUI
UAM (O'Donnell, 2008)	Classify	Yes	-	Java	GUI
WAT-SL (Kiesel et al., 2017)	Classify	Yes	apache	Java	GUI
WebAnno (Yimam et al., 2013)	Classify, Link	Yes	-	Java	GUI
WordFreak (Morton and LaCivita, 2003)	Classify, Link	Yes	-	Java	GUI

Table 1: A comparison of annotation tools in terms of properties of interest (in some cases tools support extra types of annotation, e.g. syntax, that we do not consider here).

keys change which item is selected. Keybindings can be modified to suit user preferences and multi-key commands can be defined, like the labels in Figure 1a (e.g. `SPACE+a`), providing flexibility and an unlimited set of combinations.

Basic commands cover selecting a span, assigning an annotation, and undoing annotations. Exact string search is included, which was used in the disentanglement project to search for previous messages by a given user. Typing a number before a command will apply the command that many times. A range of commands exist to toggle properties of the view, including line numbering, the legend, showing the label for the selected item, and progress through a set of files.

Data Items in annotations are represented by tuples of integers and labels are represented as strings. Tokens are defined by splitting on whitespace, and characters are counted with separate numbering for each token. This makes the annotations easy to read in and interpret.

Externally, annotations are saved in stand-off files, with one annotation per line. This makes them easy to process with command line tools, e.g. using `wc` to count the number of annotations.

Extension The system is written with a modular design intended to be easily modifiable. For exam-

ple, at one point in the disentanglement project we wanted to go back and check messages that started conversations. This involved adding 21 lines of code, extending the existing search commands to jump to the next line that started a conversation.

Internal Architecture The system is written in 2,300 lines of Python, with extensive use of the standard built-in `curses` library for input and display handling. Setting up the tool involves downloading and unzipping the code. The tool can then be run from anywhere, does not require administrator privileges, and will only make modifications to the local directory it is being run from.

Dependencies On macOS and all Linux distributions we are aware of, there is nothing else to install: Python and the relevant libraries are already installed. On Windows, Python (either 2 or 3) and the `curses` library need to be installed.

3 Related Work

Table 1 presents a comparison of a range of annotation tools.² Note that all prior work has focused on graphical user interfaces. The top section shows the tools we consider in our user study, cho-

² Non-academic tools also focus on GUIs, including open source tools (e.g. `DocAnno`, `Flat`), and commercial tools (e.g. `TagTog`, `Prodigy`, `MAT`, `LightTag`, `DataTurks`).

sen because they are widely used (brat and GATE) or have a similar design motivation (YEDDA).

brat (Stenetorp et al., 2012), is a synchronous web-based system with a central server that users connect to using a web browser. The tool supports annotation of spans of text and relations between them, either binary relations, equivalence classes, or n-ary associations. It provides a search mechanism, automatic validation, tagging with external tools, multi-lingual processing, and a comparison mode that places two annotations side-by-side. The tool has been used to annotate a range of datasets, particularly for information extraction from biomedical documents.

GATE (Bontcheva et al., 2013), is both a desktop application and a web-based system. It is designed to support the entire lifecycle of a project, including data preparation, schema creation, annotation, adjudication, data storage and use. To achieve this, it contains a wide array of components, covering various annotation types and tools to define workflows that determine the stages of annotation of a document. The system has been used in a range of projects in both academic and commercial settings, including with users who did not have a computer science background.

YEDDA (Yang et al., 2018), is a desktop application that supports annotation of character spans with up to 8 categories using a GUI. It is designed to be lightweight, with no external dependencies, though it only supports Python 2. It has a built-in recommendation system that automatically proposes labels based on annotations so far, which Yang et al. (2018) found decreased named entity recognition annotation time by 16%. For input, the tool supports both selection with the mouse and specifying a sequence of character-level annotations with the keyboard, e.g. 2c3b to assign the label *c* to the next two characters and *b* to the three after that. The tool also has an analysis component that produces a LaTeX document comparing a pair of annotations of a document, or F-scores for all pairs of annotations. It does not support annotation of links or adjudication of disagreements.

4 User Study

We conducted a user study³ to investigate the effort required to install and use four tools. For each tool, participants had 25 minutes to install the tool and use it to identify verbs in a 623 word news

³ Approved by the Michigan IRB, ID: HUM00155689.

Tool	Time (minutes)	
	Ubuntu	macOS
SLATE	10	16
YEDDA	14	14
GATE	21	22
brat	-	-

Table 2: Average time for users to set up the tool and identify verbs in a 623 word news article. Only one participant managed to install and use brat, taking 18 minutes on Ubuntu. The differences between GATE and either SLATE or YEDDA are significant at the 0.01 level according to a t-test.

article.⁴ We measured how long they took to install the tool and to finish the task. After each tool, participants completed a survey asking about their experience installing and using the tool.

We set up the study to simulate a real usage scenario as closely as possible. Participants were computer science graduate students and all but one used their own computer. Four participants used macOS and four used Ubuntu. Every participant used all four tools. To reduce bias due to fatigue (which could make later tools appear slower) and familiarity with the news article (which could make later tools appear faster), the order of tools varied so that each one occurred 1st, 2nd, 3rd, and 4th once for each operating system.

Table 2 presents the time required to install each tool and complete the first annotation task. We combined these two because some participants read usage instructions while installing the tool, while others went back and forth during initial annotations. SLATE and YEDDA are comparable in effort, which fits with their common design as simple tools with minimal dependencies. Participants had great difficulty with brat, with only two managing to install it, one just as their time finished.

We also had an additional participant try using the tools in Windows. They were able to install and use SLATE and GATE, but did not complete the task on either. GATE had the easiest set up process, with a provided installer. SLATE required the curses package to be installed. They experienced difficulty with YEDDA because it is not Python 3 compatible, and rewriting the code to make it compatible led to other issues. brat only supports Windows via the use of a virtual machine.

⁴ We also considered having participants do a linking task, specifying the antecedents of pronouns, but found there was not enough time (also, linking is not supported by YEDDA).

4.1 Feedback Summary

brat Participants got stuck on various issues during installation and generally commented that they wanted more information in the documentation. The two most common issues were being unable to get the system to communicate with apache, and being unable to load files to annotate.

GATE Participants often had to try more than one installation method. Six participants commented on lag or missed clicks during annotation, and six commented that the documentation was not helpful for the process of installing the tool and understanding the user interface.

YEDDA All participants found installation easy, though three commented that it required familiarity with git, and one recommended adding it to the python package index. Every participant commented that the recommendation mode produced many false positives, making it inconvenient. Four commented that the way labels appear during annotation shifted the text in a way that made them lose their place or made it hard to read. All Ubuntu participants commented that the system scrolled to the top of the document after each annotation.

SLATE All participants found installation easy, though one recommended adding it to the python package index. One participant commented that they wished they could use their cursor as well, one suggested adding an explicit indication that saving was successful, one found the selected span hard to see, and one was surprised that pressing the down arrow selected the same position on the next line in terms of tokens, rather than visually.

Each of these tools has strengths and weaknesses, and this study was designed to test the specific gap our tool was designed to address. For projects with annotators who are not computer experts, we expect GUIs, and in particular web interfaces like the one provided by brat, will remain dominant. However, the results demonstrate that SLATE effectively fills the gap of an easy to use and efficient terminal-based tool.

5 Future Development

There are many directions in which we hope to develop SLATE further, by collaborating with other researchers and the open-source community. The system is designed to be simple to modify and extend, with some possible next steps including:

Multi-lingual Support The tool can already display a range of scripts given Python's unicode support, but to properly support annotation would require modifications. For example, changing the direction of script writing would involve adjusting the visual display module, while adding support for selecting parts of glyphs would require modifying the data representation.

Accessibility For users who require large fonts (or prefer small ones) the tool is extremely flexible and convenient. However, the extensive use of colour could be a problem for colour-blind users. The colours can be modified by editing a simple configuration file, or colour could be avoided entirely by modifying the display code to display in-line labels.

Data Formats Support for additional data formats, such as CoNLL-U, and [ISO 24612:2012](#).

Mouse Input While the motivation for this tool was to make a text-based interface controlled by the keyboard, the curses library does process mouse inputs, making it possible to add support for the mouse in the future.

6 Conclusion

A wide range of effective annotation tools already exist, but they all use a GUI and many involve substantial effort to set up. SLATE is designed for terminal-users who want a fast, easy to install, and flexible annotation tool. It supports a range of annotation types, and adjudication of disagreements in annotations. The code is publicly available⁵ under a permissive open-source license. The tool has already been used in two research projects, including one that involved over 250 hours of annotation.

Acknowledgements

Thank you to Will Radford and the anonymous reviewers for helpful suggestions. Also thanks to the study participants, and to the annotators who have used the tool. This material is based in part on work supported by IBM as part of the Sapphire Project at the University of Michigan, and by ONR under MURI grant N000140911081. Any opinions, findings, conclusions or recommendations expressed above are those of the author and do not necessarily reflect the views of the sponsors.

⁵<http://jkk.name/slate/>

References

- Emilia Apostolova, Sean Neilan, Gary An, Noriko Tomuro, and Steven Lytinen. 2010. *Djangology: A light-weight web-based tool for distributed collaborative text annotation*. In *Proceedings of LREC*.
- Marcel Bollmann, Florian Petran, Stefanie Dipper, and Julia Krasselt. 2014. *Cora: A web-based annotation tool for historical and other non-standard language data*. In *Proceedings of the LaTeCH Workshop*.
- 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.
- Wei-Te Chen and Will Styler. 2013. *Anafora: A web-based general purpose annotation tool*. In *Proceedings of NAACL, Demos*.
- Volker Gast, Lennart Bierkandt, and Christoph Rzym-ski. 2015. *Creating and retrieving tense and aspect annotations with graphanno, a lightweight tool for multi-level annotation*. In *Proceedings of the ISA Workshop*.
- ISO 24612:2012. 2012. *Language resource management – Linguistic annotation framework (LAF)*. Standard, International Organization for Standardization.
- Johannes Kiesel, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2017. *Wat-sl: A customizable web annotation tool for segment labeling*. In *Proceedings of EACL, Demos*.
- Svetla Koeva, Borislav Rizov, and Svetlozara Leseva. 2008. *Chooser: a multi-task annotation tool*. In *Proceedings of LREC*.
- Jonathan K. Kummerfeld, Sai R. Gouravajhala, Joseph Peper, Vignesh Athreya, Chulaka Gunasekara, Jatin Ganhotra, Siva Sankalp Patel, Lazaros Polymenakos, and Walter S. Lasecki. 2019. *A large-scale corpus for conversation disentanglement*. In *Proceedings of ACL*.
- Frederic Landragin, Thierry Poibeau, and Bernard Victorri. 2012. *Analec: a new tool for the dynamic annotation of textual data*. In *Proceedings of LREC*.
- Miler Lee. 2004. *Annotator / stanford manual annotation tool*. <http://nlp.stanford.edu/software/stanford-manual-annotation-tool-2004-05-16.tar.gz>. Stanford University.
- Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. 2012. *Cat: the celct annotation tool*. In *Proceedings of LREC*.
- Michal Marcinczuk, Jan Kocon, and Bartosz Broda. 2012. *Inforex – a web-based tool for text corpus management and semantic annotation*. In *Proceedings of LREC*.
- Thomas Morton and Jeremy LaCivita. 2003. *Wordfreak: An open tool for linguistic annotation*. In *Proceedings of NAACL, Demos*.
- Christoph Müller and Michael Strube. 2006. *Multi-level annotation of linguistic data with MMAX2*. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Peter Lang.
- Mick O’Donnell. 2008. *Demonstration of the UAM CorpusTool for text and image annotation*. In *Proceedings of ACL, Demos*.
- Philip V. Ogren. 2006. *Knowtator: A protégé plug-in for annotated corpus construction*. In *Proceedings of NAACL, Demos*.
- Constantin Orasan. 2003. *PALinkA: A highly customizable tool for discourse annotation*. In *Proceedings of SIGdial*.
- Georgios Petasis. 2012. *The sync3 collaborative annotation tool*. In *Proceedings of LREC*.
- Rebecca S. Portnoff, Sadia Afroz, Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. 2017. *Tools for automated analysis of cybercriminal markets*. In *Proceedings of WWW*.
- Eric S. Raymond. 2003. *Basics of the Unix Philosophy*. Addison-Wesley Professional.
- Younes Samih, Wolfgang Maier, and Laura Kallmeyer. 2016. *Sawt: Sequence annotation web tool*. In *Proceedings of the Second Workshop on Computational Approaches to Code Switching*.
- Brett South, Shuying Shen, Jianwei Leng, Tyler Forbush, Scott DuVall, and Wendy Chapman. 2012. *A prototype tool set to support machine-assisted annotation*. In *Proceedings of BioNLP*.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. *brat: a web-based tool for nlp-assisted text annotation*. In *Proceedings of EACL, Demos*.
- Amber Stubbs. 2011. *Mae and mai: Lightweight annotation and adjudication tools*. In *Proceedings of LAW*.
- Antoine Widlöcher and Yann Mathet. 2012. *The glozz platform: A corpus annotation and mining tool*. In *Proceedings of the ACM Symposium on Document Engineering*.
- Jie Yang, Yue Zhang, Linwei Li, and Xingxuan Li. 2018. *Yedda: A lightweight collaborative text span annotation tool*. In *Proceedings of ACL, Demos*.
- Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. *Webanno: A flexible, web-based and visually supported system for distributed annotations*. In *Proceedings of ACL, Demos*.