

Prodigy Annotation Instructions

Requirements

These instructions assume you have Python, Prodigy, and the `en_core_web_sm` spaCy model installed and you are using a Unix machine (e.g., Mac or Linux). See 'Getting Help' below if you need assistance.

Data Sets

Please find the data set you have been allocated in the table below, and then download it (as a `.csv` file) to your local machine from the 'dataset' channel on Slack:

Name	Data Set
Mike Page	annotator_1.csv
Andrada Pumnea	annotator_2.csv
Teresa	annotator_3.csv
Jude	annotator_4.csv
Sophie	annotator_5.csv

The data sets each contains 1000 Tweets with their associated meta-data (e.g., no. of times retweeted, favorite count, etc.). The actual Tweets are stored in the `text` column, and this is what Prodigy will read in for you to annotate.

NB: If you would like to annotate more data, please send a message to Mike Page on Slack and another data set will be allocated to you.

Set Up Prodigy

Before you can load your data set for annotation, first you need to create a 'dataset' in Prodigy. A 'dataset' in this case means a local database instance where Prodigy will store your annotations. By default it uses SQLite meaning you don't have to worry about setting up or installing another database (such as MySQL). To create a 'dataset' in Prodigy, open a new terminal window and type the following command, replacing {Me} with your first name (no spaces):

```
prodigy dataset optout_proto "Opt Out Prototype Data" --author {Me}
```

Note: you only need to run the above command once, the first time you are setting up Prodigy for your allocated data set.

Load Data

To load in data into the Prodigy, type the following command, replacing {path_to_your_data} with the path to the `.csv` file you have been allocated:

```
prodigy textcat.teach optout_proto en_core_web_sm {path_to_your_data}
--label MISOGYNY --exclude optout_proto
```

After running the command a Prodigy server will start running. Open the front-end by clicking on the link in the terminal or opening <http://localhost:8080> in a browser.

NB: The `--exclude` tag in the command above means that you can save your progress at any point, and carry on the annotations from where you left off without having to start all over again. The command states that you want to *exclude* any Tweets you annotated previously. Therefore, whether it is the first time you are annotating Tweets in your allocated data set, or you want to pick up from where you left off, just type the same load command above.

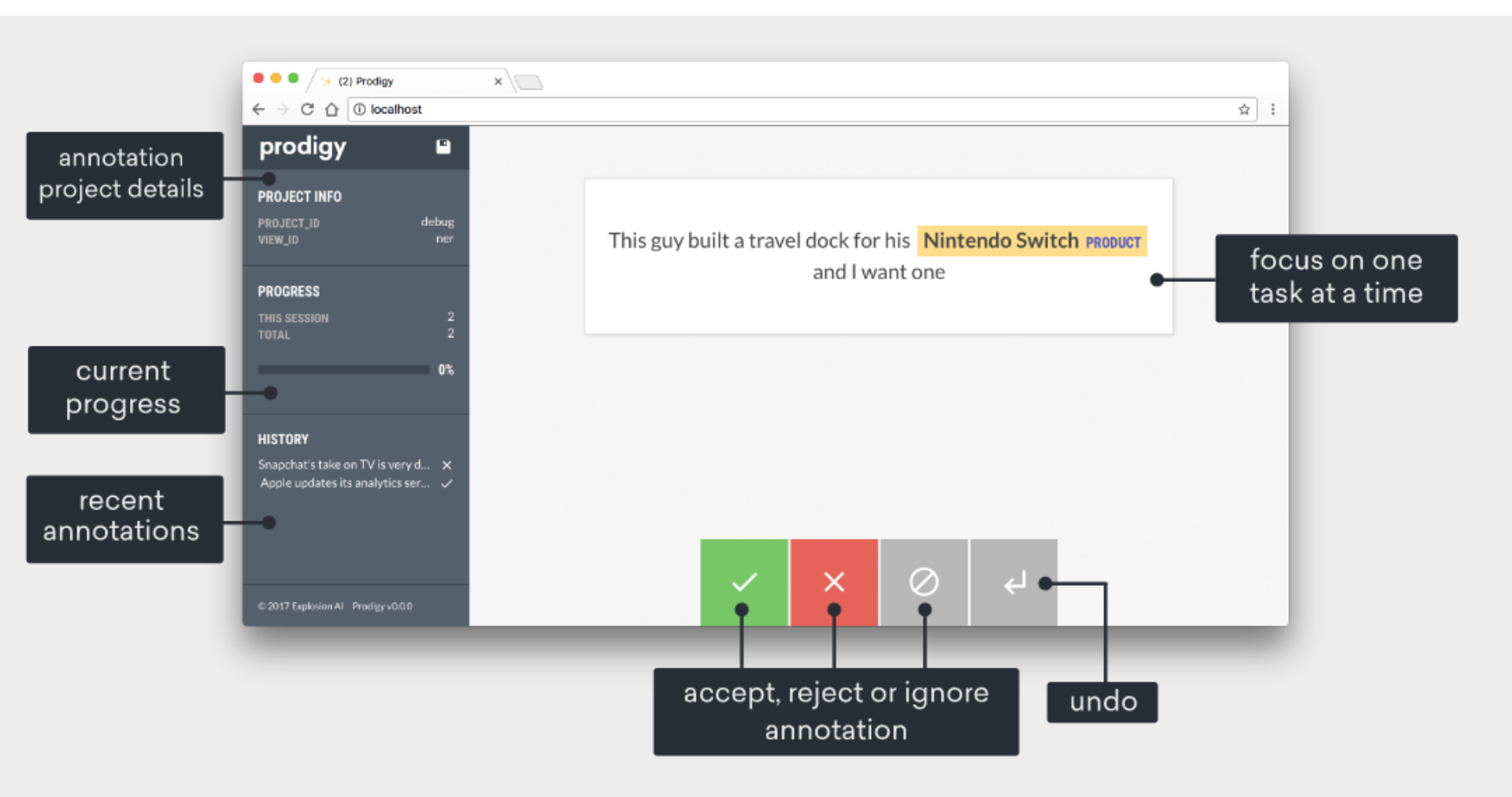
Annotate Data

To annotate the data, you need to either need to `accept`, `reject`, or `ignore` each Tweet as it is presented to you on the screen. These actions have the following implications:

- `accept`: You accept the Tweet is misogynistic
- `reject`: You reject the Tweet as being misogynistic (i.e., it is not misogynistic)
- `ignore`: You aren't sure if the Tweet is misogynistic, so ignore it.

Another way to think of this is prodigy assumes all Tweets are Misogynistic, and you either accept or reject (or ignore if you aren't sure) this claim on a Tweet by Tweet basis.

To `accept`, `reject`, or `ignore` each Tweet, click the appropriate button (see the diagram below).



You can also undo an annotation at any point by pressing the 'undo' button. Your progress (i.e., how many Tweets you have annotated) can be seen in the progress bar on the left of the screen.

Annotation Criteria

In order to decide if a tweet is misogynous, it should include a word or element, which could be classified under the following Misogynistic categories (cfr. Saha et al. 2018, Anzovino et al. 2018):

- Discredit:** slurring over, insulting, slandering or badmouthing with no other larger intention. For example:
I've yet to come across a nice girl. They all end up being bitches in the end #WomenSuck
- Stereotype:** widely held but fixed and oversimplified image or idea of a woman. For example:
I don't know why women wear watches, there's a perfectly good clock on the stove. #WomenSuck
- Objectification:** description of women's physical appeal and/or comparisons to narrow standards. For example:
You're ugly. Caking on makeup can't fix ugly. It just makes it worse!
- Sexual harassment:** describing actions as sexual advances, requests for sexual favors. Harassment of a sexual nature. For example:
Women are equal and deserve respect. Just kidding, they should suck my dick.
- Threat of violence:** intent to physically assert power over women, or to intimidate and silence women through threats of violence. For example:
Domestic abuse is never okay... Unless your wife is a bitch #WomenSuck
- Dominance:** asserting the superiority of men over women to highlight gender inequality, in order to preserve male control, protect male interests and to exclude women from conversation. For example:
We better not ever have a woman president @WomenSuckk
- Derailing:** justifying woman abuse, rejecting male responsibility; an attempt to disrupt the conversation in order to redirect women's conversations on something more comfortable for men. For example:
@yesllwomen wearing a tiny skirt is "asking for it". You're teasing a (hard- working, taxes paying) dog with a bone. That's cruel. #YesAllMen

Save Data

While annotations should be auto-saved in batches of 10 by default (i.e., every 10 annotations the data is automatically saved), It is *imperative* you manually save the annotations you have made before you end any session to guarantee they are saved (plus to account for any annotations under a multiple of 10 which won't be auto-saved). To do this, click the save button in the top left corner of the screen (see the diagram above) or press `CTRL + S` or `COMMAND + S`.

Export Data

Once you have annotated (and saved) all 1000 Tweets in your data set. You can export them with their annotation labels by typing the following command, replacing {x} with the same number in your `.csv` file you were allocated:

```
prodigy db-out optout_proto > annotator_{x}.jsonl
```

For example, Mike Page would run the following command:

```
prodigy db-out optout_proto > annotator_1.jsonl
```

Once saved, please upload the `.jsonl` file to the 'dataset' thread on Slack.

Getting Help

If you get stuck or need more help, please message Mike Page or Andrada Pumnea on Slack, or post a message on the 'dataset' thread.