

Doccano Guide

General

Doccano is an annotation management system for text data. It can be used for developing datasets for classification, sentiment analysis, entity tagging, or translation tasks. We will use it for text classification purposes, so each entry should be labeled with tags.

First, take a look at the online demo for sentiment analysis (the same procedure can be followed for classification as well): <http://doccano.herokuapp.com/demo/text-classification>

In a few words, the annotator should assign a label (“positive” or “negative”) to each entry. The annotations are stored automatically and maintained for the next sign in.

For our needs (annotator)

First, visit our doccano installation at <http://pharm-doccano.lvrysis.com>. Sign in using your credentials as annotator (Figure 1).

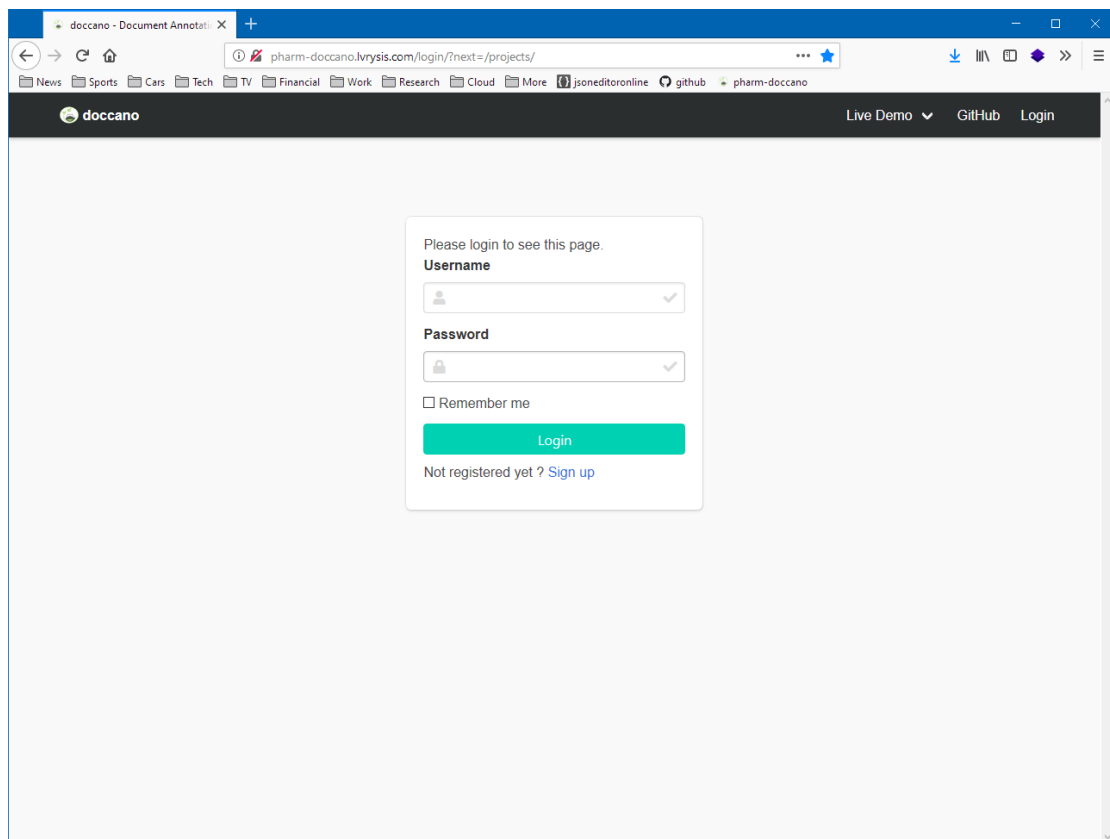


Figure 1 Login screen

Choose the project you have been asked to collaborate (Figure 2).

Next, read the text (for the current entry), choose either “Hate” or “No Hate” (if the text contains content against refugees or migrants or not), and mark it as “Positive”, “Neutral” or “Negative” (according to the sentiment of the text). It is mandatory to choose one “blue” and one “purple” label for each entry, as the annotation is performed for two individual classification schemes simultaneously. “Blue” labels are for hate speech detection, while “purple” labels are necessary for sentiment analysis. Annotated entries are highlighted with a checkmark icon. Finally, choose “next” to move on to the next entry.

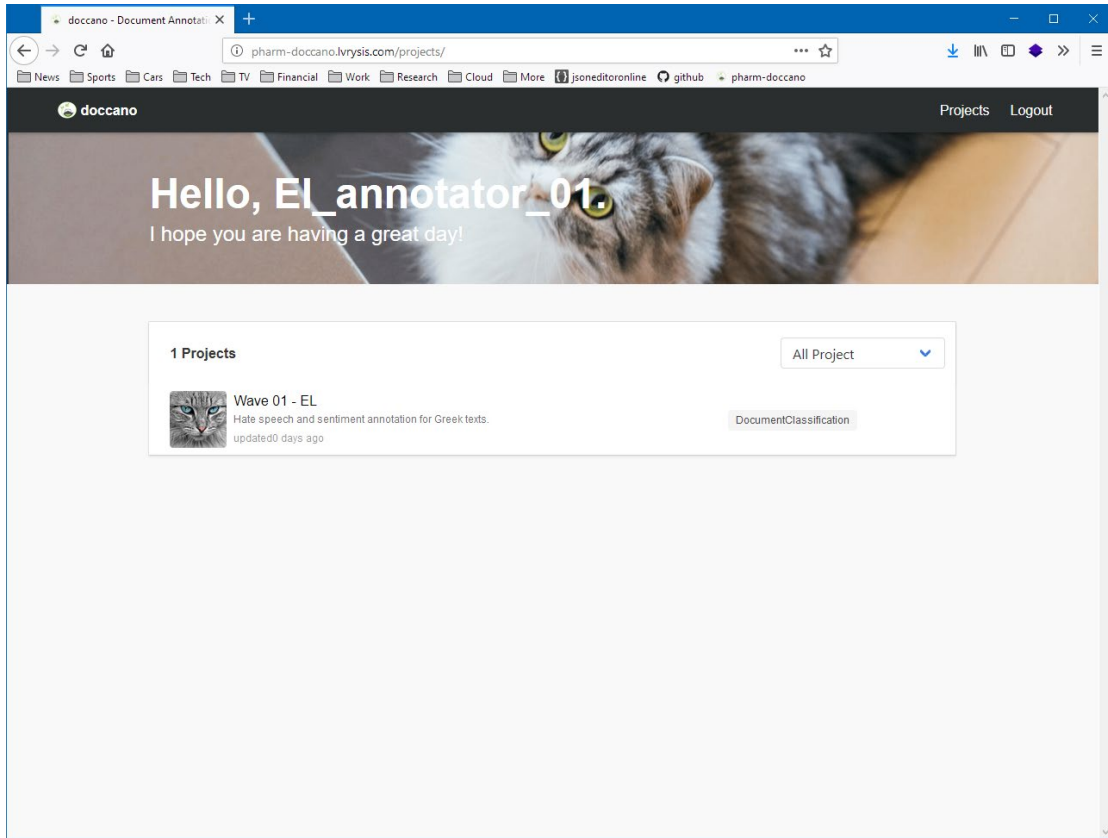


Figure 2 Project selection screen

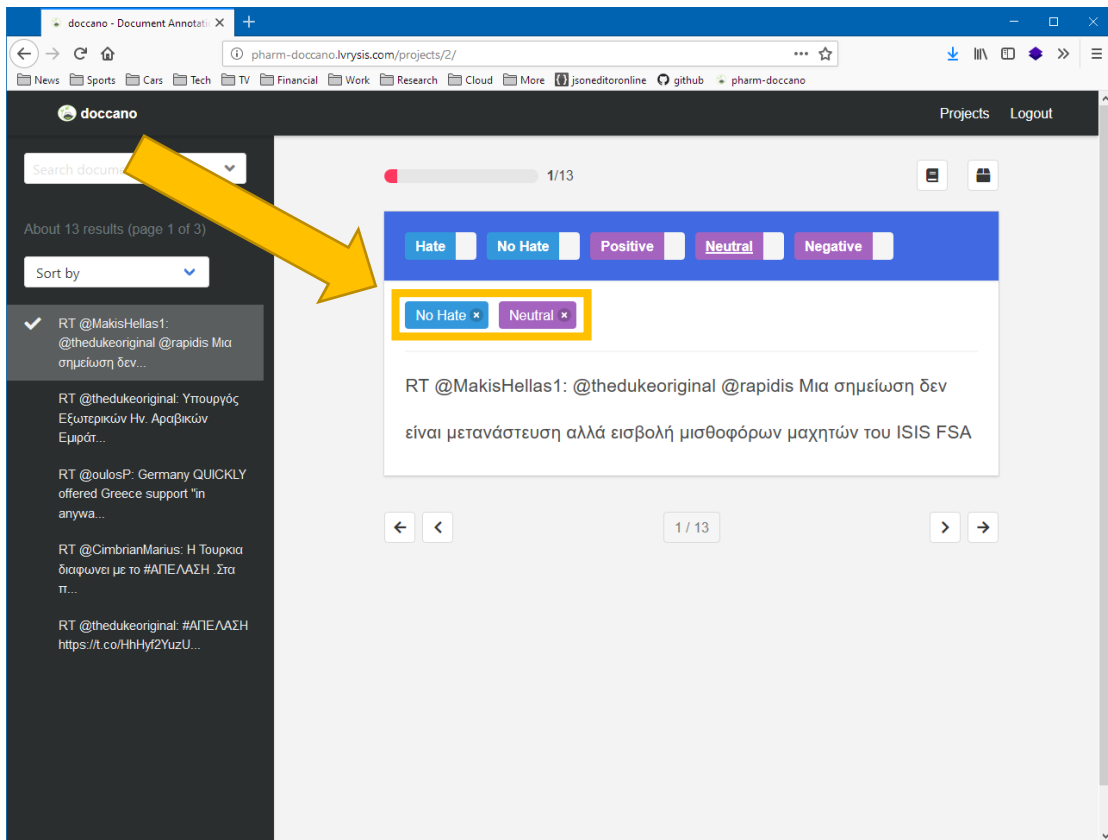


Figure 3 Annotation screen and labels

For our needs (approver)

Log in to doccano and select the appropriate project (follow the same steps as described previously). Approve an annotation using the button shown in Figure 4. Approved entries are shown with a thumbs up icon.

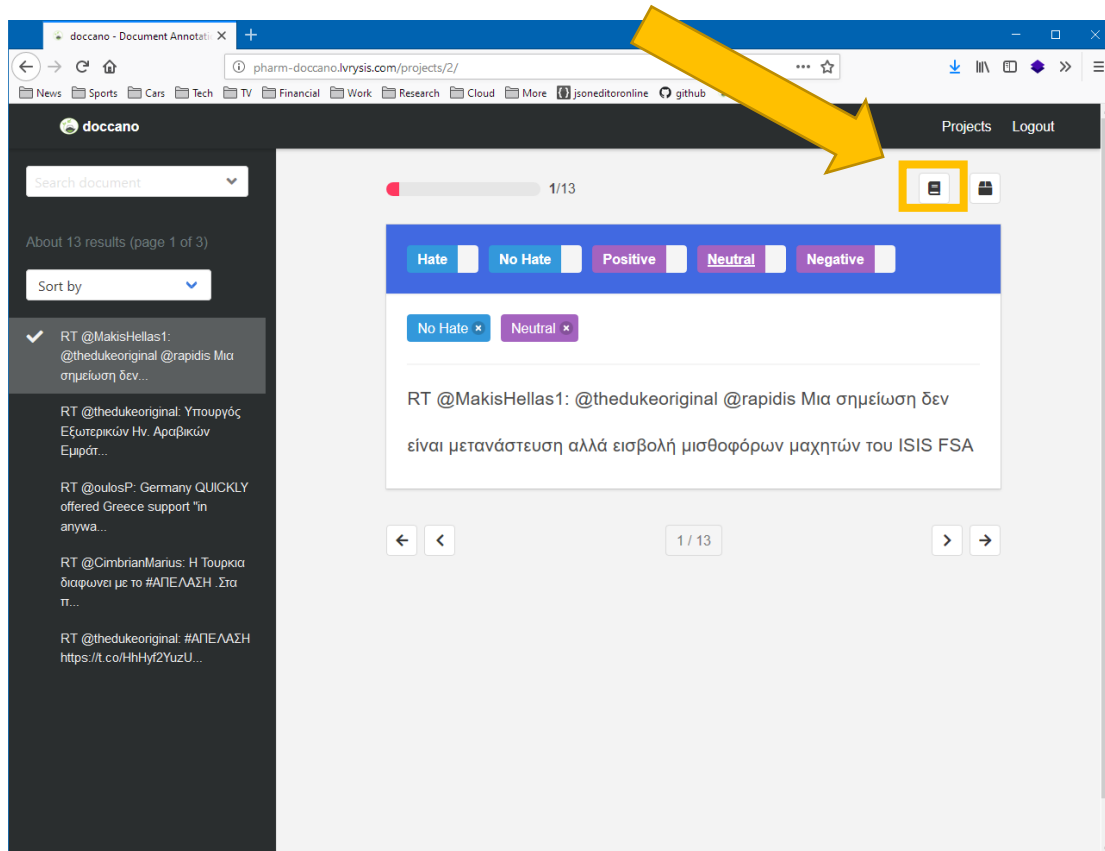


Figure 4 Annotation screen and approval