Annotation Guidlines for Explainability Annotations for Legal Judgment Prediction in Switzerland

Nina Baumgartner

1 Introduction

1.1 Annotation Goal

Recently Niklaus, Chalkidis, and Stürmer (2021) presented a dataset for legal Judgment prediction including 85k Swiss Federal Supreme Court decisions. Using Hierarchical BERT they achieved a Macro-F1 Score of approximately 68-70%. Nevertheless, the inner workings of such models are still mostly unknown and the results thus not interpretable. For this reason, this annotation task focuses on the explainability of these predictions. With your annotation you will give your insight as a legal expert and tag parts of the facts that support or oppose the judgment. These guidelines are based on the work of Reiter (2020), Leitner (2019 [Online]) and Pustejovsky and Stubbs (2012). They are a work-in-progress in collaboration with Lynn Grau, Angela Stefanelli and Thomas Lüthi.

1.2 Dataset

For this task a subset of the Swiss court ruling corpus presented by Niklaus et al. (2021) containing 108 cases was created. In most of these cases the model worked well and indicated the correct judgment form the facts given to it. The 108 cases are equally distributed among the three languages German, French and Italian. Each language set contains six cases over six years. With each year having two cases per legal area[1]: One with the verdict approved and one with the verdict dismissed.

1.3 Disclaimer

This document is a work-in-progress. If you have questions or find any errors in these instructions while doing the annotation please feel free to contact the maintainer. Please help with collecting examples to complete these guidelines.

2 Annotation Entities

Although you will only be annotating the fact section of a ruling, you will have access to the full document (via a link on Prodigy) and the judgment will be clearly indicated on the prodigy interface. You can and should use these other resources as an indicator on which part of the facts are of greatest importance. The annotation will focus on the sentences and sub-sentences. @ToDo define Sentences?

We focus on this linguistic entity because it is large enough to understand the context but still sufficiently versatile to accurately annotate. Thus, it is possible that a sentence contains two sub-sentences opposing each other, which would be consequently annotated with different labels. We hope that by choosing this scope it is possible to indicate what the different part of the sentences denote in the context of the judgment and to subsequently better explain the decisions of the model.

3 Annotation Categories

To annotate the sentences of each fact section you will be using two labels, *Supports judgment* and *Opposes verdict*. In addition, you will be given several options for dealing with problematic cases, which should help to improve the dataset, these guidelines and the annotations themselves.

---

[1]The chosen legal areas are categorized as penal law, social law and civil law

3.1 Supports Judgment

This label is used when a sentence or sub-sentence supports the judgment. Every sub-sentence that supports the judgment should be annotated. @ToDo Example and more explanation

3.2 Opposes Verdict

This label is used when a sentence or sub-sentence opposes the judgment. Every sub-sentence that opposes the judgment should be annotated. @ToDo Example and more explanation

3.3 Neutral

Every not-labeled sub-sentence is considered neutral. This is not a label per say but merely how the system interprets words or sentences which are not assigned one of the labels above. It is important for the analysis that even the neutral sentences are annotated which in our case means to omit them.

One example in German of an neutral expression which should not be tagged with a label is the word "Sachverhalt:". This word only indicates the beginning of the fact section and should be left out as a neutral part of the facts because it does not give us any further information on the explainability of the judgment.

3.4 Problematic Cases

Problematic cases can occur because this annotation task is an iterative process. For now, we differentiate between three possible types of such cases.

3.4.1 Rejected Cases

If a case is badly tokenized[2] or there is another formal error it should be rejected. Please state your reasoning in the comment window using the comment pattern below and reference the Reject or Ignore a Case section of this document for the details on how to properly reject a case.
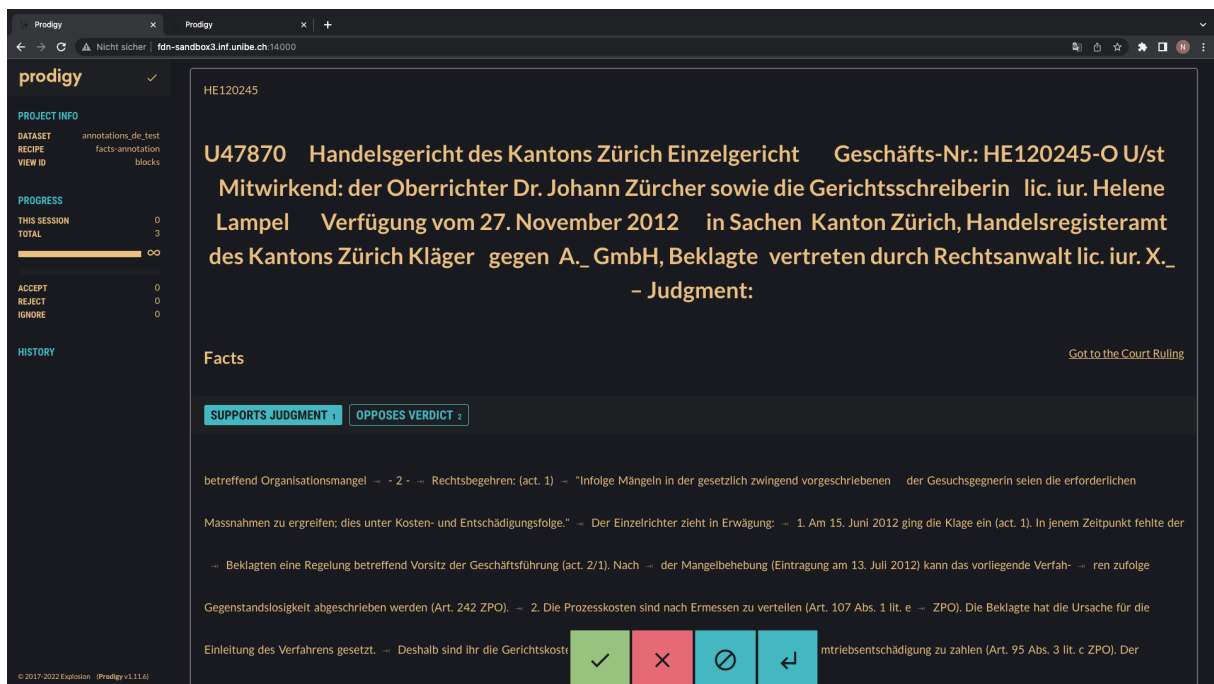

Figure 1: Example of a case containing a formal error which should be rejected. Here the title was parsed incorrectly, the judgment is missing and the facts are tokenized wrongly and incomplete.

---

[2]Tokenized means that the system did not properly separate the words.

3.4.2 Ignored Cases

If a case is to short or otherwise unfit for the annotation it should be ignored. To ignore it please state your reasoning in the comment section and follow the steps explained in the Implementation section of this document below.

3.4.3 Other Problematic Cases

There might be cases without formal errors where you have difficulties to annotate (neither reject nor ignore). In such cases, please annotate to the best of your ability and explain your reasoning in the comment section.

3.4.4 Comment Structure

Comment for rejecting and ignoring case

Number of case – Annotators name

- Why did you ignore/reject this case?

Comment for generally problematic case

Number of case – Annotators name

- Why is this case problematic and difficult to annotate?

- How did you decide on your annotation?

---

4 Implementation: How to Annotate the Dataset Using Prodigy

This section explains how to use the annotation tool Prodigy[3]. We built a custom recipe for this task which lets you annotate the facts section of a given court decision.

4.1 Access

The Prodigy instance can only be accessed via the University of Bern network. If you want to annotate from home you must use the VPN of the University of Bern[4]. If you are connected to the university network you can access Prodigy via one of the following URLs:

- Angela: http://fdn-sandbox3.inf.unibe.ch:11001/

- Lynn: http://fdn-sandbox3.inf.unibe.ch:11002/

- Thomas: http://fdn-sandbox3.inf.unibe.ch:11003/

- FR-Annotations: http://fdn-sandbox3.inf.unibe.ch:12000/

- IT-Annotations: http://fdn-sandbox3.inf.unibe.ch:13000/

Before you can start you will be asked to provide a *username* and a *password*, which will be give to you by the maintainer of the annotation process. After the login procedure you should now see an overview of the case and you can start with your annotation.

4.2 Annotate a Sub-Sentence

To label a phrase with a tag, highlight it with your cursor and choose the corresponding label. To delete a tag simply click on the tagged words again.

---

[3]https://prodi.gy/
[4]https://serviceportal.unibe.ch/sp?id=kb$_a$rticle$_v$iewsys$_k$b$_i$d $= 00cb11e51b005050134ddc6a9b4bcb49$
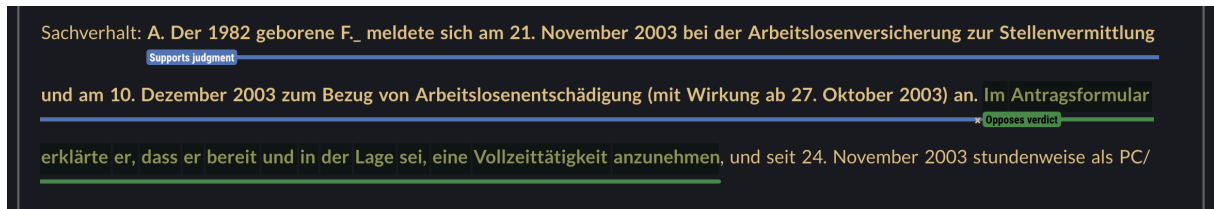
Figure 2: Screenshot of sentence labeling in prodigy.

If you are happy with your annotation you can accept it by clicking on the green check [1] and save it by pressing the save button in the left corner [2]. To see your progress you can look at the information displayed on the left [3]. If you want to access the original document you can click on the link in the right corner [4]. Please do not forget to safe [2] your progress.
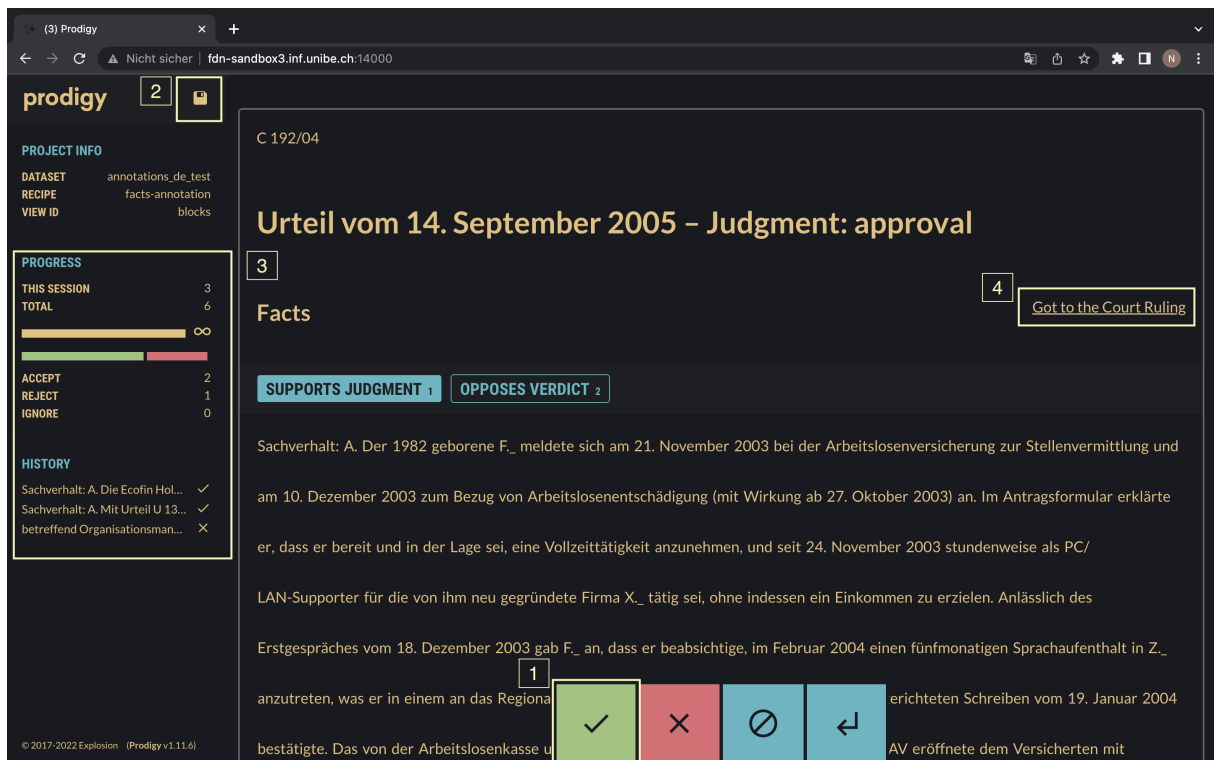

Figure 3: Screenshot of the case overview on prodigy

4.3 Reject or Ignore a Case

To reject a case state your reasoning in the comment section and press the red cross to reject it. To ignore it, press the blue button with the stop signal after commenting. Do not forget to save your progress.
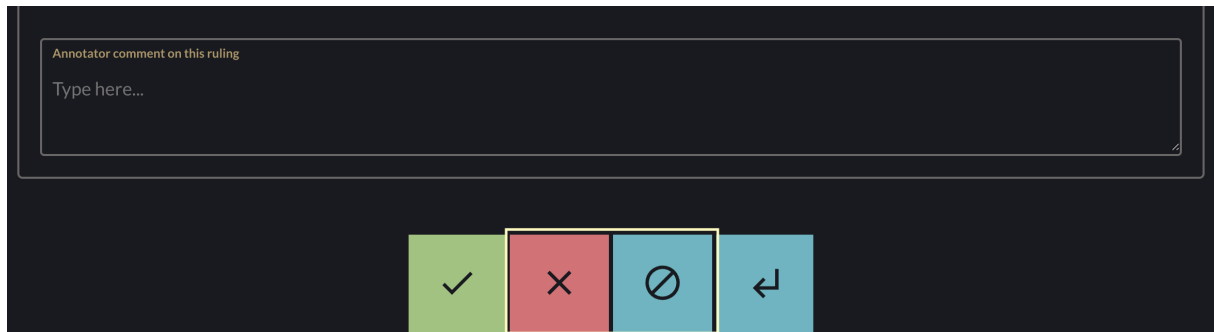

Figure 4: Reject and Ignore buttons

5 Change Log

This change log documents the progress of these guidelines. When adapting these guideline please also add a new entry to the changelog using the following structure

---

Change log

Date – Title of changes

- Which parts were changed in which iteration?

- Why was this part changed

References

Leitner, E. (2019 [Online], Feb). *Annotationsrichtlinien zum juristischen korpus ler.* GitHub. Retrieved from https://github.com/elenanereiss/Legal-Entity-Recognition/blob/master/docs/Annotationsrichtlinien.pdf

Niklaus, J., Chalkidis, I., & Stürmer, M. (2021). *Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark.* arXiv. Retrieved from https://arxiv.org/abs/2110.00806 DOI: 10.48550/ARXIV.2110.00806

Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning* (Nos. Bd. 9,S. 878). O'Reilly Media, Incorporated.

Reiter, N. (2020). Anleitung zur erstellung von annotationsrichtlinien. In N. Reiter, A. Pichler, & J. Kuhn (Eds.), *Reflektierte algorithmische textanalyse: Interdisziplinäre(s) arbeiten in der creta-werkstatt* (pp. 193–202). De Gruyter. Retrieved from https://doi.org/10.1515/9783110693973-009 DOI: doi:10.1515/9783110693973-009