Annotation Guidelines 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 diachronic multilingual (German, French, Italian) dataset for Legal Judgment Prediction LJP including 85k Swiss Federal Supreme Court decisions. Using Hierarchical BERT, they achieved a Macro-F1 Score of up to 70%, considering penal law exclusively, they even achieved a score of up to 80%. To use Artificial Intelligence (AI) safely in high stakes domains such as law we need explanations on how these decisions are made. To investigate explainabilty in the legal area of AI we want to gather some human and model-generated explanation of decision from the SwissJudgmentPrediction (SJP) corpus.

This annotation task has the goal to gather the human part of the explanation. With your annotation you will give your insight as a legal expert and tag parts of the facts with specific labels. These guidelines should help you to identify the important parts of the facts and create consistent annotations. They are based on the work of Reiter (2020), Leitner, Rehm, and Moreno-Schneider (2019) and Pustejovsky and Stubbs (2012). They are a work-in-progress in collaboration with Lynn Grau, Angela Stefanelli and Thomas Lüthi.

1.2 Dataset

The SJP dataset is split into training, validation and testing set. For this annotation task a balanced subset of the SJP containing 108 cases taken from the test and validation set was created. The dataset is deemed balanced because the 108 cases are equally distributed among the three languages contained in the Swiss judicial system German, French and Italian. Each language set contains six cases over six years (2015 until 2020). With each year having two cases per legal area[1]: One with the verdict approved and one with the verdict dismissed. In addition, preference was given to cases where the model decided the correct judgment form the facts given to it, with some outliers in the French and Italian subset.

1.3 Disclaimer

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

2 The Annotation Cycle

To produce quality annotations and guidelines, which make the annotation task scalable and reproducible the annotations have to be done in cycles. Pustejovsky and Stubbs (2012) call this process the MAMA (Model-Annotate-Model-Annotate) cycle (see Figure 1 for details).

Using the annotation guidelines to identify the right parts of the text, multiple annotations by multiple individual annotators are done on the same input. Then these annotations are analyzed and the guidelines adapted accordingly to provide consistency in the annotations. Therefore, it is important that for this first few cycles the annotations are done individually. Later the gold standard annotation for this corpus emerge from this process. Pustejovsky and Stubbs (2012) describe gold standard annotations as the final version of the annotations, which uses the most up-to-date guidelines and has everything labeled

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

correctly. For this work these gold standard annotation will be done as a team. For the practical aspect of this process please reference section 5.1.1, 5.1.2, 5.1.3.

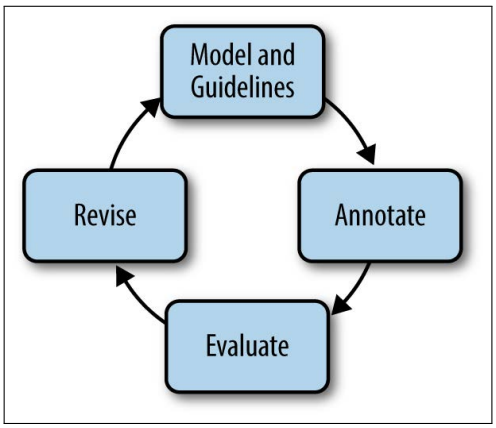


Figure 1: The inner workings of the MAMA cycle (Pustejovsky & Stubbs, 2012).

3 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.

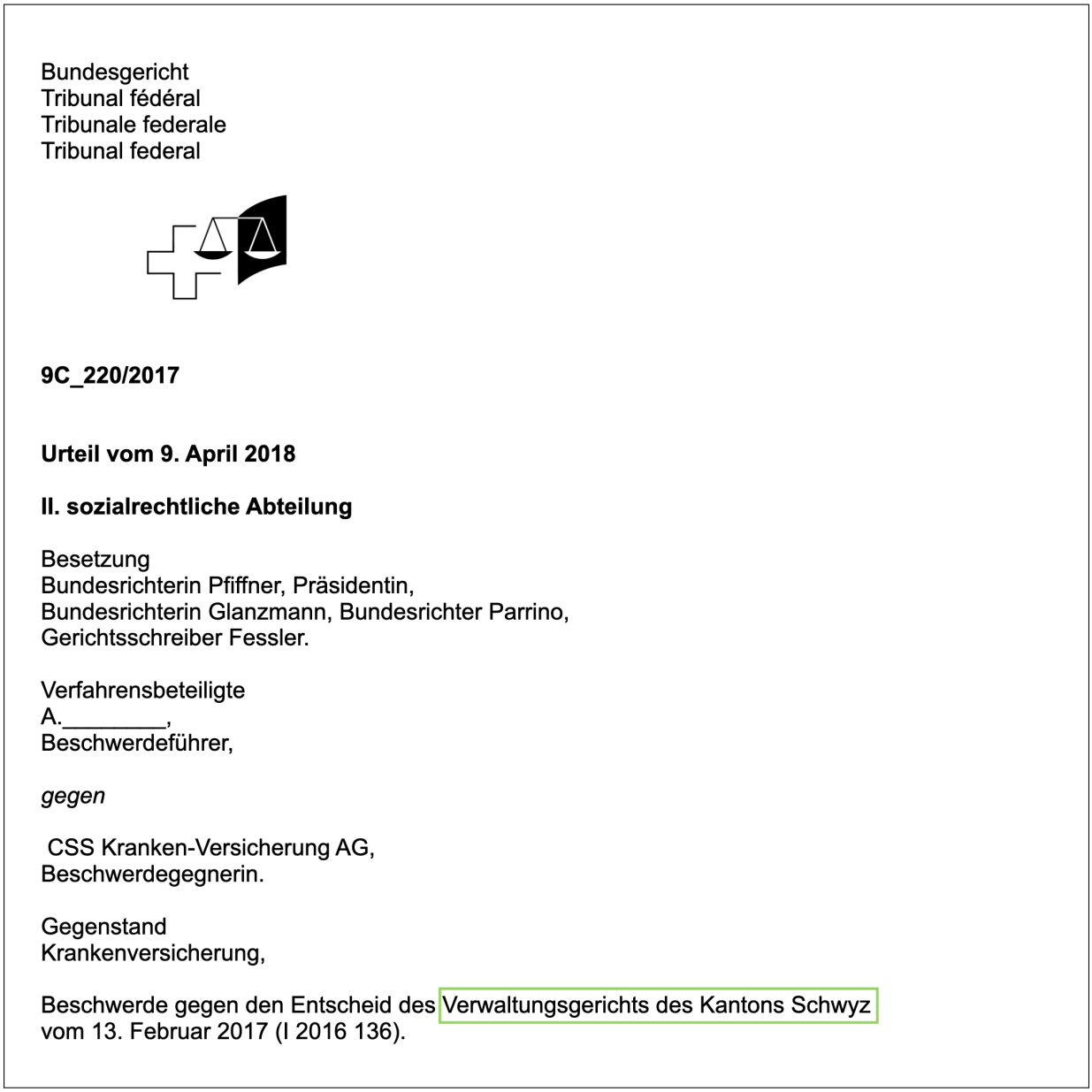
3.1 Sentences and Sub-Sentences

Wiegreffe and Marasovic (2021) identify three types of explanations in the Explainable Natural Language Processing (ExNLP) literature: highlights, free-text, and structured explanations. The explainability annotations for this task focus mainly on highlights with some addition of free-text explanations.

To add highlights you will label sentences or sub-sentences as supporting or opposing the judgment. For this task we define a sentence as a self-contained linguistic unit consisting of multiple words, terminated with a period, semicolon, colon, question mark or exclamation mark. A entire sentence is the largest entity to be annotated. A sentence can consist of multiple sub-sentences usually separated with a "and" or a comma. It is possible that a sentence contains two sub-sentences opposing each other, which should be consequently annotated with different labels. These sub-sentences are the smallest units that should be annotated. So single words or expressions should never be annotated. We hope that by choosing those units it is possible to indicate what the different parts of the sentences denote in the context of the judgment and to subsequently better explain the decisions of the model.

3.2 Lower Court

In addition to sentences you will also have to annotate the last lower court of each case. As seen in Figure 2 the Rubrum of the ruling indicates the last lower court. The last lower court is composed of the name of the court e.g. "Verwaltungsgericht" and the location "Kanton Luzern". Please annotate all instances of the lower court where it appears as a complete constellation. So for example if "Verwaltungsgericht des Kanton Luzern" appears multiple times in the facts please label it each time. Please Note that you should only annotate the lower court itself please do not label prepositions like "beim" or "zum" or verbs like "sprach" which are often found next to the lower court.


Bundesgericht
Tribunal fédéral
Tribunale federale
Tribunal federal

9C_220/2017

Urteil vom 9. April 2018

II. sozialrechtliche Abteilung

Besetzung
Bundesrichterin Pfiffner, Präsidentin,
Bundesrichterin Glanzmann, Bundesrichter Parrino,
Gerichtsschreiber Fessler.

Verfahrensbeteiligte
A._____,
Beschwerdeführer,

gegen

CSS Kranken-Versicherung AG,
Beschwerdegegnerin.

Gegenstand
Krankenversicherung,

Beschwerde gegen den Entscheid des Verwaltungsgerichts des Kantons Schwyz
vom 13. Februar 2017 (I 2016 136).

Figure 2: Screenshot of a Rubrum with the lower court highlighted Judgment (of the Federal Court) from September 8th 2017.

4 Annotation Categories

To annotate the sentences of each fact section you will be using two labels, *Supports judgment* and *Opposes verdict*. You should also highlight the lower court for each judgement. 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.

4.1 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.

4.2 Opposes Judgment

This label is used when a sentence or sub-sentence opposes the judgment. Every sub-sentence that opposes the judgment should be annotated.

4.3 Lower Court

This label is used to highlight the last lower court of the case. To label the last lower court highlight the name and the location of the court as one instance (see Figure 3).

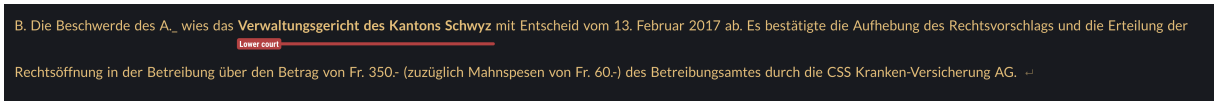


Figure 3: Example of a highlighted lower court in Prodigy.

4.4 Neutral

Every not-labeled sub-sentence is considered neutral. This is not a label per se 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.

Another example of a neutral part of the facts are the section indicators labeled with capital letters (e.g. A., B., A.a., A.b and so on). Note that witnesses, accused persons and other involved parties are also labeled with uppercase letters and should be annotated if part of a sentence (see 4 below as illustration).

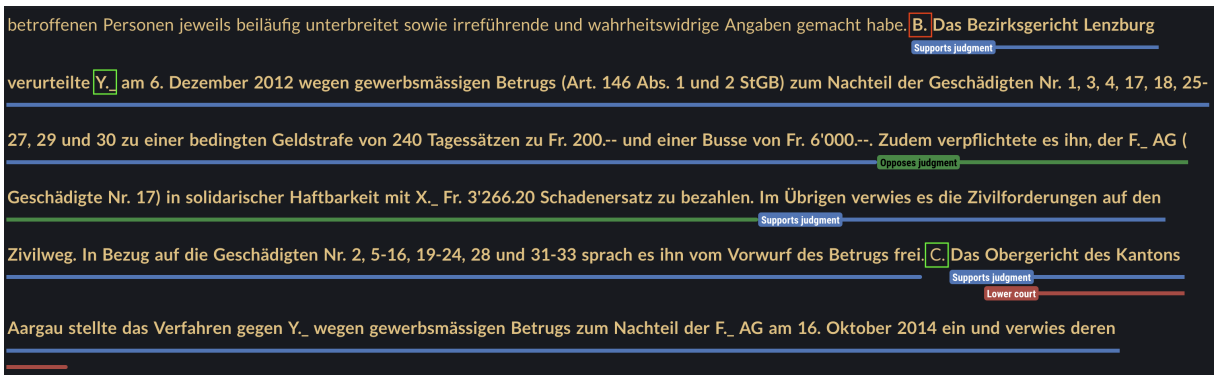


Figure 4: Example an annotation where the uppercase letters are first wrongly (marked red) and then correctly annotated (marked green).

4.5 Problematic Cases

Problematic cases can occur. For now, we differentiate between three possible types of such cases.

4.5.1 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 5 is an example of a case with formal errors.

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

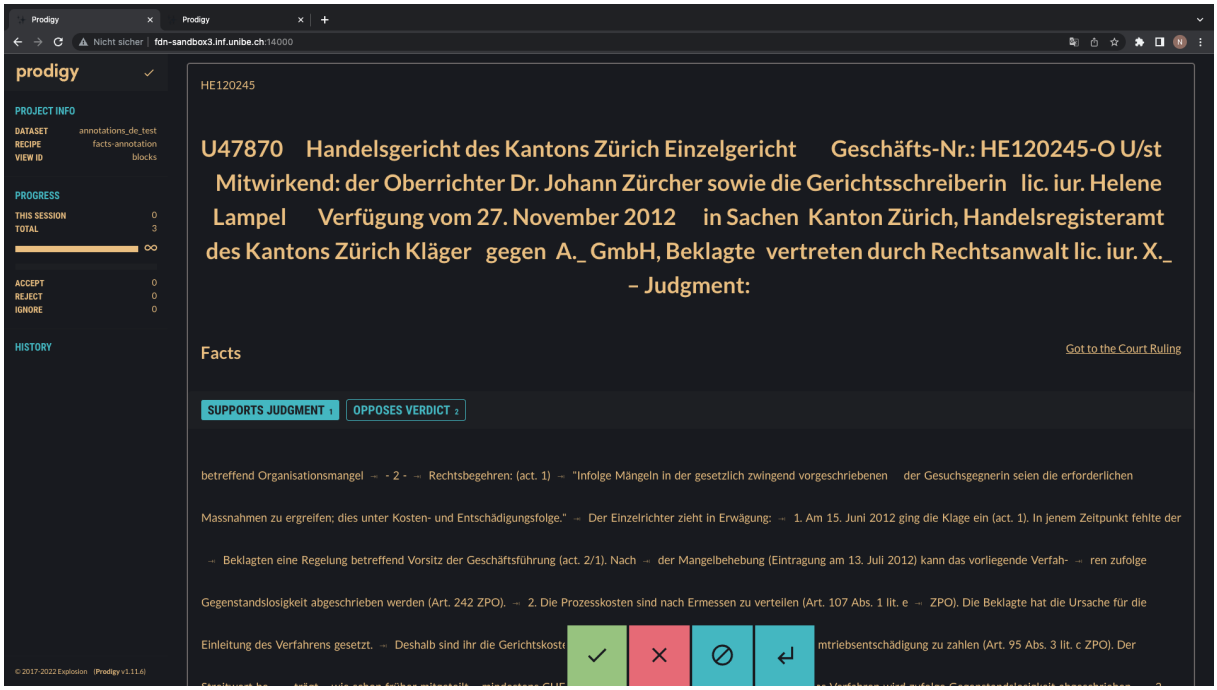


Figure 5: Example of a case containing a formal error which should be rejected. Here the title was parsed incorrectly, the judgment is missing and the facts are tokenized wrongly and incomplete.

4.5.2 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 Reject or Ignore a Case section of this document below.

An example of a case which was ignored from a annotator is the Judgment (of the Federal Court) from April 8th 2020 (please reference whole case online via link). The annotator which ignored this case explained his reasoning as follow in the free text explanation:

> "Before the Federal Court, only the question of party compensation was in dispute. The underlying facts, however, actually have nothing to do with the court's decision."

This argumentation can be supported with the following parts of the facts section from from April 8th 2020:

> "B. [...] Allerdings verpflichtete [das Verwaltungsgericht des Kantons Bern] die Suva-MV, A._____ eine Parteientschädigung in der Höhe von Fr. 3610.15 [...] zu bezahlen.
>
> C. Die Suva-MV erhebt Beschwerde in öffentlich-rechtlichen Angelegenheiten und beantragt sinngemäss, der angefochtene Entscheid sei bezüglich der Zusprache der Parteientschädigung aufzuheben. A._____ beantragt, auf die Beschwerde sei nicht einzutreten, eventuell sei sie abzuweisen."

4.5.3 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.

4.5.4 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?

5 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.

5.1 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 URLs in the following three sections. 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.

5.1.1 First cycle

The following links will be used for your pilot annotations (first iteration). If you completed the annotation on this dataset ignored and rejected cases will be replaced with other cases having the same legal area, year and judgment. This process is ongoing until we reach 36 accepted cases.

- German case annotations:
  - Angela: http://fdn-sandbox3.inf.unibe.ch:11000/?session=angela
  - Lynn: http://fdn-sandbox3.inf.unibe.ch:11000/?session=lynn
  - Thomas: http://fdn-sandbox3.inf.unibe.ch:11000/?session=thomas
- French case annotations: http://fdn-sandbox3.inf.unibe.ch:12000/
- Italian case annotations: http://fdn-sandbox3.inf.unibe.ch:13000/

Note that sessions can be added dynamically by adding the suffix /?session=SessionName to the url.

5.1.2 Further Cycles and Corrections

If you have completed all the pending annotations on the above URLs Prodigy will display a message saying no task available. This is your indicator to continue to this part of the annotations. Reference the Guideline for recent changes and adapt your annotations accordingly. You can repeat this process with a new session as often as you want (see session management example below).

- German case annotations:
  - Angela: http://fdn-sandbox3.inf.unibe.ch:11001/?session=angela
  - Lynn: http://fdn-sandbox3.inf.unibe.ch:11002/?session=lynn

[3] https://prodi.gy/

[4] https://serviceportal.unibe.ch/sp?id=kb$_article_viewsysparm_article = KB0010032$

  - Thomas: http://fdn-sandbox3.inf.unibe.ch:11003/?session=thomas

- French case annotations: http://fdn-sandbox3.inf.unibe.ch:12000/?session=lynn
- Italian case annotations: http://fdn-sandbox3.inf.unibe.ch:13000/?session=angela

If you need to do multiple corrections on the same case please add a number behind you link as seen below to distinguish between the sessions:

- Session 1: http://fdn-sandbox3.inf.unibe.ch:11001/?session=angela1
- Session 2: http://fdn-sandbox3.inf.unibe.ch:11001/?session=angela2

5.1.3 Final Gold Standard Annotations

After some iterations you and the other annotator will get together and decide on the final annotation using the below link.

- German gold standard annotations:
  - http://fdn-sandbox3.inf.unibe.ch:8080/?session=gold

5.2 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. As seen in Figure 6 the two label appear in two different colors. By hovering over an annotated section the delete toggle appears.

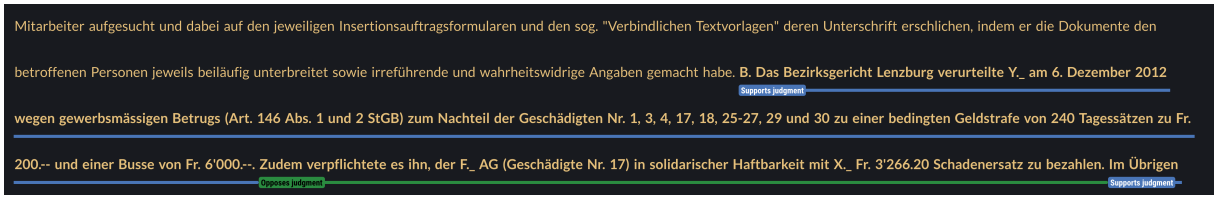


Figure 6: Screenshot of sentence labeling in prodigy.

If you are happy with your annotation you can accept it by clicking on the green check labeled with [1] in Figure 7 and save it by pressing the save button in the left corner referenced by the number [2]. To see your progress you can look at the information displayed on the left (see number [3] on Figure 7). If you want to access the original document you can click on the link in the right corner (see number [4]). Please do not forget to safe your progress using the save button [2].

If you want to skip a case, because you already annotated it. Please use the accept button [1] to get to the next case.

Figure 7: Screenshot of the case overview on prodigy

5.3 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 8 shows the interface of the comment section and the ignore and reject buttons.

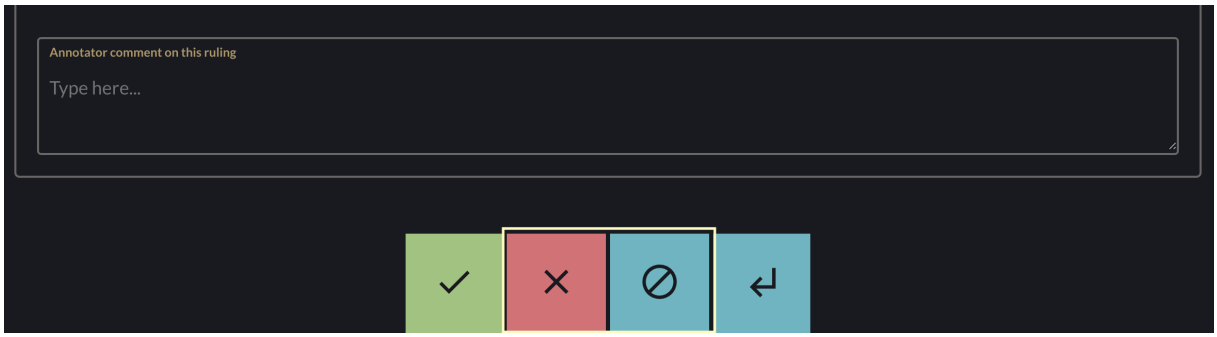


Figure 8: Reject and Ignore buttons

6 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

Template

Date – Title of changes

- Which parts were changed this iteration?
- Why was this part changed?

10.04.2022 – Formal changes after first feedback

- Which parts were changed this iteration?
  - Changed E. Leitner reference to the published article.
  - Corrected some spelling errors.
  - Integrated the figures into the text of the guidelines.
  - Changed label "opposes verdict" to "opposes judgment"
- Why was this part changed?

  With this first adaption of the guidelines we mainly worked on some formal errors to standardize the format and clarify the instruction (especially with integrating the figures into the text). The label was changed to make the annotation and their interpretation more consistent.

23.04.2022 – Changes to Prodigy setup and new label

- Which parts were changed this iteration?
  - Named multi-user sessions where added to the Prodigy setup, which changed the annotators URLs in this document.
  - The label lower court was added as a new annotation category and subsequently to the prodigy setup. Explanation how and when to use it were added to section 2 and 3.
  - Directions on how to skip already annotated cases were added. To section 4.2
- Why was this part changed?

  The named multi-user session was a pending part of the prodigy setup, which is now resolved. The URLs of the annotators had to be adapted accordingly. After a meeting with the lawyer and annotator Thomas Lüthi, we decided on adding the new label "lower court", to highlight it as a separate entity additionally to the existing two labels. Correct sessions where not yet implemented in the first setup of Prodigy used for some annotations, for this reason directions on how to skip a case a already annotated case were added.

12.05.2022 – Changes to Prodigy setup, introduction revision, explanation of the annotation cycle

- Which parts were changed in this iteration?
  - The Prodigy setup was extended to enable the iterative work on the annotations. Therefore, an explanation on when to use which link was added. In addition explanation on the annotation cycle itself where added.
  - Updated images because Prodigy Interface changed
  - After writing the proposal of the thesis corresponding to these guidelines the introduction was adapted accordingly.
  - Ignored Case example was added
- Why was this part changed

  Enabling the iterative process is an important step to provide quality annotations and guidelines. Therefore after the setup was implemented the guidelines had to be adapted. The images had to be updated because they where no longer up to date and to provide consistency in these guidelines. To give the annotator a better understanding of the task the introduction was updated with some input from the proposal. After analysing the currently done annotation a example of a ignored case could be added to these guidelines.

21.07.2022 – Language extensions, clarification of the instructions

- Which parts were changed in this iteration?
  - Extensions to French and Italian in the iterative annotation cycle in the implementation part of these guidelines.
  - Added some clarification to the lower court label which also specifies how often it should be annotated.
  - Added the section dividing capital letters as a new neutral element.
- Why was this part changed

  Enabling the iterative process is an important step to provide quality annotations and guidelines. Therefore after the setup was implemented in Italian and French the guidelines had to be adapted. After reviewing first results of the annotation done using these guidelines some clarification for more consistent annotation where added. The new neutral element and clarification in the lower court section will help to prevent distortion in the annotator agreement caused by minor shifts of the start of the label. The clarification that all lower court instances appearing in complete form should be annotated, was added so that the annotation where most similar to the models output (the model will extract all instances of lower court).

References

from April 8th 2020, C. . (2020). *Bgu 8c 641/2019.* Retrieved from https://www.bger.ch/ext/eurospider/live/de/php/aza/http/index.php?highlight_docid=aza%3A%2F%2F08-04-2020-8C_641-2019&lang=de&type=show_document&zoom=YES&

from September 8th 2017, C. . (2017). *Bgu 9c 424/2017.* Retrieved from https://www.bger.ch/ext/eurospider/live/de/php/aza/http/index.php?lang=de&type=highlight_simple_query&page=1&from_date=20.08.2017&to_date=08.09.2017&sort=relevance&insertion_date=&top_subcollection_aza=all&query_words=&rank=9&azaclir=aza&highlight_docid=aza%3A%2F%2F08-09-2017-9C_424-2017&number_of_ranks=456

Leitner, E., Rehm, G., & Moreno-Schneider, J. (2019, 9). Fine-grained Named Entity Recognition in Legal Documents. In M. Acosta, P. Cudré-Mauroux, M. Maleshkova, T. Pellegrini, H. Sack, & Y. Sure-Vetter (Eds.), *Semantic systems. the power of ai and knowledge graphs. proceedings of the 15th international conference (semantics 2019)* (pp. 272–287). Karlsruhe, Germany: Springer. (10/11 September 2019)

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

Wiegreffe, S., & Marasovic, A. (2021). Teach me to explain: A review of datasets for explainable NLP. *CoRR*, *abs/2102.12060*. Retrieved from https://arxiv.org/abs/2102.12060

List of Figures