

Annotation Guideline

This is the annotation guideline for the data sets for “Claim Extraction and Law Matching for COVID-19-related Legislation”. The annotation guideline starts with defining a claim (Claim definition) and describing the different kinds of claims that can be commonly found in newspaper articles (A taxonomy of claims). It then describes the rules that are to be followed while annotating (Rulebook).

While reading the annotation guideline, one thing should be kept in mind: The annotations are for both the claim extraction and the law matching task. However, there are many cases where a claim is not suitable for the law matching task, but is a perfectly valid claim for the claim extraction task.¹ This guideline tries to strike a balance so that the data can be used to make two separate datasets. For this is utilized labels (see “Label” section).

Claim definition

A claim can be defined as a conclusion whose merit must be established.² In regard to this task, we will choose a more narrow definition. Levy. et. all (2014) suggest the concept of a “Context Dependent Claim”: a general, concise statement that directly supports or contests the given topic, which is a short phrase that frames the discussion.

The context in our case is COVID-19 related legislation in Berlin. Thus, in our case, a claim is a statement that can be verified by looking at COVID-19 related rules in Berlin.³

One question that we will try to answer in this work is whether that is a sensible definition, especially for the claim extraction task. Because that means that the model did not only learn to recognize claims, but also to understand whether it is related to the topic (it could do that by looking at context information, like preceding text or titles).

General claim annotation style

1. Only laws that are valid in Berlin are annotated
2. We always try to reference the laws as precisely as possible. The style for annotation laws is the following:
§ 18 (1) S. 1, 3 InfSchMV
=> § *section (passage)* S. *sentence numbers*, where passage and sentences are only present if applicable⁴, and sentence numbers are comma separated.

¹ One example would be “Ab Mittwoch fallen die Regeln zum Maskentragen weg”. This kind of temporal claim (where a law is omitted) falls out of the scope of the law matching task, but not the of the claim extraction task.

² Levy et al.1489, Toulmin, 1958.

³ That does not only include legislation, but also other “rules”, like political agreements. See “Edge cases” -> “Non-law reasons that verify a claim”

⁴ Meaning that a section is subdivided by passages and sentences.

§ 2 (1) InfSchMV i.V.m. § 2 (3) Nr. 10 InfSchMV

=> [reference] i.V.m. [reference], where reference is in the style above. If more than one passage needs to be referenced, use this style.

Some more examples:

- § 9 (2) InfSchMV i.V.m. § 9 (3) Nr. 4, 5, 6 InfSchMV i.V.m. § 9 (10) InfSchMV
 - § 14 (1) SARS-CoV-2-EindmaßnV i.V.m. § 14 (2) SARS-CoV-2-EindmaßnV i.V.m. § 14 (3) j), k) SARS-CoV-2-EindmaßnV
3. The date for a reference is not when the claim was made or the newspaper written, but a date when the referenced law was valid (see notes for time-dependent claims)

A taxonomy of claims

In this section we will describe different structures of claims that appear in the newspaper articles, and how to annotate them.

These categories are overlapping, a claim can be in several of these descriptions.

Self-contained claim

Definition: A *self-contained claim* is a claim that contains enough information in its text such that its veracity can be determined by a (reasonably informed) human without requiring more context.

Example: “Wenn bei Veranstaltungen in Innenräumen mehr als fünf Personen zusammenkommen, darf das ebenfalls nur unter Nachweis eines negativen Corona-Tests passieren.”

Annotation style: Mark the whole claim. As text, add all laws (see Annotation Laws), and add a date (in “dd.mm.yy” format) when the claim was valid.

Usage: Self-contained claims are used for both CE and LM tasks.

Notes: The question on how much context is needed for a claim about a COVID-19 law basically depends on the prerequisites for facts (Tatbestandsvoraussetzungen) of the laws. So for the claim extraction task it is unreasonable to expect the model to decide whether a claim is self-contained or not.

Distributed claim

Definition: A *distributed claim* is a claim such that several, non-connected parts of a text have to be part of the claim in order for it to be self-contained.

Example: “Die bisher wegen der Corona-Pandemie geltenden Abstandsregeln in Gaststätten werden in Berlin gelockert. [Some text that is not part of the claim] Der Mindestabstand von 1,5 Metern zwischen den Tischen muss allerdings weiter eingehalten werden.”

Annotation style: If any part is self-contained, mark it as a claim, otherwise with the “no-context” label.

Usage: The parts with the “no-context” label will be used for the CE task, but not for LM.

Notes: The “no-context” label can be used for many sentences that are clearly a claim, but at the same time also not really suitable for LM.

Atomic claim

Definition: An atomic claim is a claim such that it is not possible to split the claim in two different claims without losing essential information.

Example: “In Friseursalons ist ein tagesaktuelles negatives Testergebnis Zugangsvoraussetzung.”

Counterexample: “In Friseursalons, Kosmetikstudios und bei anderen körpernahen Dienstleistungen ist ein tagesaktuelles negatives Testergebnis Zugangsvoraussetzung.”

Annotation style: Just like a claim.

Usage: Both CE and LM.

Notes: This definition is still a bit hairy, and not entirely precise. But the basic idea is that it probably is easier for a model to find the referenced laws if it looks at an atomic claim.

Compound claim

Definition: A claim that is not atomic.

Example: “In Friseursalons, Kosmetikstudios und bei anderen körpernahen Dienstleistungen ist ein tagesaktuelles negatives Testergebnis Zugangsvoraussetzung.”

Annotation style: Either mark one atomic claim, and the rest with the “no-context” label (not always possible, see the example). Or mark the whole compound claim as one claim, and annotate all relevant laws, joined with “i.V.m.”, e.g. “§ 14 (1) SARS-CoV-2-EindmaßnV i.V.m. § 14 (2) SARS-CoV-2-EindmaßnV i.V.m § 14 (3) Nr. j) SARS-CoV-2-EindmaßnV”.

Usage: Bot CE and LM.

Notes: Those are quite common, so the LM model will need to be able to deal with those.

No law-claims

Definition: A claim that can’t be justified by the text of a existing law (but e.g. with the absence of a law, or analogies)

Examples:

- “Für die Arbeitnehmerinnen und Arbeitnehmer besteht keine Testpflicht.”

- “Einkaufen soll dann mit vorher gebuchten Terminen möglich sein, solange die 7-Tage-Inzidenz nicht über 100 steigt.” (That claim was made on 04.03.21. On this time, there was only a political agreement about it, not a formalized law)
- "Religionsausübung sowie Demonstrationen bleiben davon unberührt" (edge case)
- “Zuvor war nur eine haushaltsfremde Person erlaubt.” (here the fact that a law was omitted is needed to verify the claim).

Annotation style: Mark with “**no-law**” label.

Usage: Only CE.

Notes: It is not reasonable to expect that the CE model could tell that a sentence is a “no-law” claim, since it is basically the responsibility for the LM model to find such a law.

Time dependent-claims

Definition: A claim that can only be verified by taking a time component into consideration.

Example: “Eine Ausgangssperre gibt es in der Hauptstadt noch nicht.” (also a non-law claim)

Annotation style: Mark with “**time-context**” label

Usage: Only CE. The LM model will not be able to deal with time contexts.

Notes: Often it is possible to annotate a time dependent-claim in a way that makes it time independent. E.g. “Ab nächsten Sonntag gilt folgendes: Masken müssen ...”. Here you can just annotate everything that follows after the colon. Make sure you remember to annotate the date correctly, i.e. not with the date of the newspaper article, *but with a date when the law was actually valid* (in this case: Sunday).

Citation

Definition: A claim that repeats verbatim the law.

Example: “Bei der Öffnung sind geeignete Vorkehrung zur Hygiene, zur Steuerung des Zutritts und zur Vermeidung von Warteschlangen zu treffen.”

Annotation style: Mark with “citation” label.

Usage: Only CE. Not super interesting for LM.

Edge cases and general notes

Only laws directly concerned with the claim should be annotated

“Museumsbesuch nur mit medizinischer Maske”.

⇒ § 4 (1) Nr. 14 InfSchMV (Tragen medizinischer Maske in kulturellen Einrichtungen).

Here, it is not totally clear whether we should also annotate the fact that museums can be opened in the first place (§ 20 (2) S. 1 InfSchMV). One could argue that this fact is implicitly contained, since in a pandemic it can be expected that all museums are closed. However, this should not be annotated: The claim itself speaks about the rules which apply to visiting museums, and not whether that is possible (that is merely a prerequisite).

For claims referencing the exception of a rule-exception law, both should be annotated

“Gartencenter, Blumenläden und Babyfachmärkte dürfen demnach ohne Terminvereinbarungen für Kunden öffnen.” § 15 (1) S. 1, 3 InfSchMV

=> In this case, both S. 1 (for visiting a shop, an appointment is required) and S. 3 (for those named, S. 1 does not apply) should be annotated. In this case, it is a classic rule-exception structure, and the exception only makes sense in context of the rule (it also explicitly references the rule).

That is different from the case mentioned before, where we can separate the question of “can I visit a museum?” and “when I visit a museum, what rules do apply?” (it is, however, a fine line).

“Prostitution zählt nicht zu den körpernahen Dienstleistungen und ist von den Lockerungen zunächst ausgenommen.” => § 18 (2) InfSchMV i.V.m. § 18 (5) InfSchMV

Prefer composite claims over distributed claims

“Auch körpernahe Dienstleistungen wie Massagen sind dann wieder erlaubt, allerdings müssen sich die Mitarbeiter mindestens einmal pro Woche einem Schnelltest unterziehen. Die Kunden und Kundinnen brauchen einen tagesaktuellen Schnelltest.” § 18 (2) Nr. 1, 3 InfSchMV i.V.m. § 18 (3) InfSchMV, 07.03.21

=> There are two ways on how this could be annotated:

1. Both sentences annotated separately: Then the first sentence would be one claim⁵, and the second one a Distributed Claim (since the second sentence requires the context of the first one).
2. Both sentences annotated together: Then this is a Composite Claim with two different, but connected, claims.

Since our hypothesis is that our law matching model can handle composite claims, but not distributed (context-lacking) claims, we should make this one claim.

⁵ One could argue that the first sentence is a Composite Claim too, saying “körpernahe Dienstleistungen sind erlaubt” and “Mitarbeiter müssen wöchentlich ein Schnelltest machen”. However, the “allerdings” makes it one claim.