A Dataset of German Legal Reference Annotations

Harshil Darji
University of Passau
Germany
Harshil.Darji@uni-passau.de

Jelena Mitrović
University of Passau
Germany
Institute for AI R&D of Serbia
Serbia
Jelena.Mitrovic@uni-passau.de

Michael Granitzer
University of Passau
Germany
Michael.Granitzer@uni-passau.de

ABSTRACT

The field of legal Natural Language Processing faces a lot of challenges due to the unavailability of properly structured datasets. One such instance is the need for a dataset that not only separates different parts of legal references, such as an article or paragraph number but also provides information about what a particular legal reference dictates. Having access to such a dataset can provide easy access to researchers working on experiments such as context similarity between law texts and legal cases that refer to a particular law. In this paper, we present a dataset of 2944 legal references in German law that are manually annotated by law experts. This dataset has 21 properties for each law reference in the dataset, such as *Buch*, *Teil*, *Titel*, *Untertitel*, etc. It also provides the complete text of each law reference in the dataset, along with specific paragraph text mentioned in the law reference. Furthermore, using this dataset together with Open Legal Data, we perform a law reference prediction task to compare the performance between predicting full law reference and only the base law reference.

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**; *Neural networks*.

KEYWORDS

Legal Language Processing, NLP, Sentence Transformers, Open Legal Data, Law References, Law Reference Annotations

1 INTRODUCTION

In the legal field, the amount of data is growing more and more every day, while digitization faces the same challenges as every other field. Manually processing and analyzing this much data is truly a monotonous and time-consuming task. For this reason, legal NLP

and Legal Tech aim at applying state-of-the-art computer science techniques to improve and speed up this process. The use and significance of Artificial Intelligence in the legal domain for reasoning with cases and argumentation are expressed in [5]. Studies mentioned in this paper show that the use of computing methodologies for reasoning with legal cases goes as far back as 1987 [9]. However, one of the bottlenecks here is the lack of properly structured datasets on which new models or approaches can safely be trained and published. Lack of access to such properly structured datasets in the legal domain acts as a barrier for researchers working in Legal Tech [1, 14].

Datasets in the German legal domain that are currently available mainly focus on providing the entirety of legal cases, e.g. Open Legal Data[1], or named entities in those legal cases such as the Legal Entity Recognition dataset [6]. However, there are certain tasks such as comparing contextual similarities between a legal case and the law text this case refers to, or analyzing the importance of law in reference to other laws. For such tasks, it is necessary to have a dataset of legal references that not only provides researchers with full texts for each legal reference mentioned in a given legal case but also text from a specific paragraph, if required.

The legal references in the dataset we are presenting here are obtained from the Open Legal Dataset using a German BERT model [2] trained on the Legal Entity Recognition dataset [6]. These legal references are then manually annotated, including a link to their online version. We conducted experiments on law reference prediction tasks using Open Legal Data. Our goal was to predict both full law references and base law references from this dataset.

2 RELATED WORK

There are very few open and available legal datasets for German law, one of which is Open Legal Data, published by Ostendorff et al. [8] in 2020. This dataset contains metadata for more than 250,000 German laws and court decisions. The primary source for this dataset are courts and government data, as they have the highest level of reliability. Secondary sources are scrapping data directly from trusted websites such as Gesetze im Internet[2] and Rechtsprechung im Internet[3], and Linked Open Data.

Another important dataset was published by Leitner et al. [6] in 2019. This dataset was generated for Named Entity Recognition in German federal court decisions. It contains approximately 67,000 sentences and 54,000 manually annotated entities. These entities are characterized into 7 coarse-grained classes: *Person, Location, Organization, Legal norm, Case-by-case regulation, Court decision,*

---

[1] https://de.openlegaldata.io/
[2] https://www.gesetze-im-internet.de/
[3] https://www.rechtsprechung-im-internet.de/

and *Legal literature*. These are further sub-characterized into 19 fine-grained classes: *person, judge, lawyer, country, city, street, landscape, organization, company, institution, court, brand, law, ordinance, European legal norm, regulation, contract, court decision, and legal literature*. The dataset is publicly available in the CoNNL-2002 format [10] on GitHub[4].

Furthermore, Urchs et al.[12] published another important dataset in 2021. It is available in two parts, one consists of a total of 32,748 decisions from 131 German courts, while the second part consists of 200 randomly chosen judgments. The judgments in the second part of the dataset are annotated with the conclusion, definition, and subsumption annotations.

In [11], Urchs et al. also identified and addressed the lack of availability of German legal corpora that follows the *Urteilsstil* writing style for German judgments. As stated by the authors, this writing style "*begins with the conclusion and proceeds with the reasoning*". To create the corpus, the authors crawled over 30,000 court decisions from www.gesetze-bayern.de. Out of these court decisions, 11,477 are judgments. The authors then proceeded with 200 randomly selected judgments for manual annotations and published them as one part of the dataset. The second part consists of 30,000 court decisions with *Name of the court, Type of the court decision, Date of the court decision, Reference number, Title of the court decision, Norm chains, Place of reference* metadata.

Other than these publicly available datasets, there are a few cost-based databases such as juris[5], beck-online[6], LexisNexis[7], and Wolters Kluwer Online[8]. All these datasets contain a large number of case records, federal statutes, commentaries, and more.

To the best of our knowledge, there is no publicly available dataset for the German legal domain that provides researchers with full or partial texts of law references for a legal case.

3 DATASET

In this section, we explain the process of gathering data from Open Legal Data, extracting law references from this data using Transformers [13], and finally, annotating these law references.

3.1 Pre-processing Open Legal Data

In order to create a dataset of law reference annotations, we first need a list of law references to annotate. For this purpose, we gathered the case records from the Open Legal Data, preprocessed them, and created a new dataset of clean Open Legal Data, which is currently publicly available at zenodo[9]. This dataset is approximately 1.1 GB in size and contains 43337 rows and the following 12 features.

The next step is to extract law references from this dataset. Various ways to accomplish this task include pattern matching as law references usually start with the article sign ("§") [7]. However, we decided to take a slightly different approach using a transformer[13] model.

Table 1: Features in the clean Open Legal Data dataset with statistics.

Features	Total
id	43337
slug	43337
ecli	10831
date	43337
court	43337
jurisdiction	43337
level_of_appeal	43337
type	43337
tenor	36282
tatbestand	24243
gründe	27144
entscheidungsgründe	24038

3.2 German BERT for Legal NER

To extract law references from pre-processed Open Legal Data, we use a German BERT[10] model fine-tuned on the Legal Entity Recognition dataset mentioned in related works with a micro-F1 score of 99.49[4]. This fine-tuned German BERT is available on HuggingFace[11]. We then used this fine-tuned model on the *tenor, tatbestand, gründe,* and *entscheidungsgründe* parts of the clean Open Legal Data to extract law references.

Three legal experts, who have already passed their first state examinations in law, manually annotated 21 labels for the law references. To ensure that the annotation task was correctly set up, they performed a test annotation task on 523 law references. The level of agreement between the three annotators was then measured by Fleiss' kappa score, which has been calculated to be 0.87, indicating a strong inter-annotator agreement. Afterward, the annotators worked independently on annotating the remaining law references using all the physical and digital legal resources necessary for the tasks. These 21 annotation labels are as follows:

Gesetzbuch/Norm, Buch , Teil, Abschnitt, Titel, Untertitel, Kapitel, Artikel, Absatz, Buchstabe, Unterabsatz, Satz, Nummer, Buchstabe , Online Link Gesetzbuch, Online Link Exakt, Alternative Schreibweise 1, Alternative Schreibweise 2, law_title, full_text, absatz_text

Here, *Gesetzbuch/Norm* refers to the standard written description of a specific law, *Online Link Gesetzbuch* refers to the link to the base of the law, while *Online Link Exakt* refers to the link to the specific article of the law. *Alternative Schreibweise 1* and *Alternative Schreibweise 2* show the normalized variant to write a law reference.

For example, given a law reference, **Art. 1 Abs. 1 GG**[12], *Gesetzbuch/Norm* is **Grundgesetz**, the base of the law is **GG**, and the specific article is **Art. 1**. *Alternative Schreibweise 1* and *Alternative Schreibweise 2* is **Art. 1 (1) GG** and **Art. 1 I GG** .

3.3 Gathering corresponding law texts

Next, we use the links available in *Online Link Exakt* to get the *law_title*, *full_text*, and *absatz_text*. However, we were not able to get these details for all the law references due to various reasons such as some laws not being in German or specified *Absatz* not being available in the linked text.

For the above example, the *law_title* is **Grundgesetz für die Bundesrepublik Deutschland Art 1**, the *full_text* is

> *(1) Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.*
> *(2) Das Deutsche Volk bekennt sich darum zu unverletzlichen und unveräußerlichen Menschenrechten als Grundlage jeder menschlichen Gemeinschaft, ...,*

and the *absatz_text* is

> *(1) Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.*

Also, some law references in the dataset simply refer to the base article. Therefore, there are also law references for which many features have absent values, as shown in Table 2.

Table 2: Number of absent values for each feature in our dataset, with the English translation of each feature.

Features	English	Absent
Referenz	Reference	0
Gesetzbuch /Norm	Code/Norm	0
Buch	Book	2264
Teil	Part	1943
Abschnitt	Section	674
Titel	Title	2484
Untertitel	Subtitle	2872
Kapitel	Chapter	2294
Artikel	Article	0
Absatz	Paragraph	971
Buchstabe	Letter	2933
Unterabsatz	Subparagraph	2926
Satz	Sentence	2129
Nummer	Number	2543
Buchstabe	Letter	2928
Online Link Gesetzbuch	Online link law book	0
Online Link Exakt	Online link Exact	0
Alternative Schreibweise 1	Alternate spelling 1	11
Alternative Schreibweise 2	Alternate spelling 2	11
law_title	Law title	0
full_text	Full text	1
absatz_text	Paragraph text	973

Figure 1 shows the total number of occurrences of law reference for the top-10 Gesetzbuch /Norm.
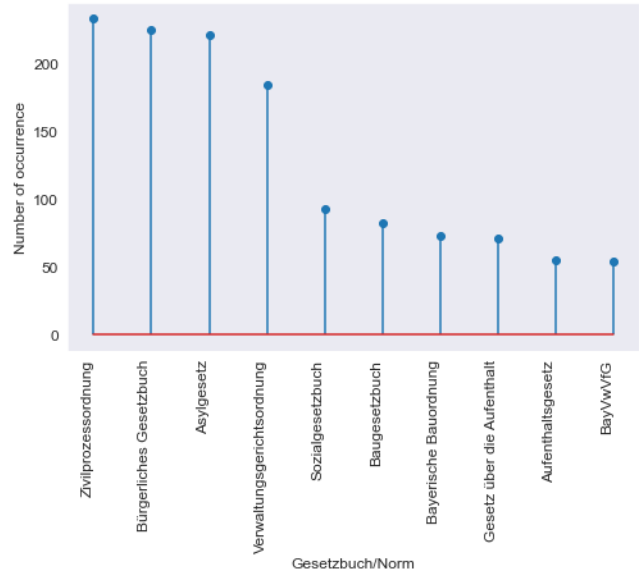

Figure 1: Number of law reference for top-10 Gesetzbuch/Norm

As we can see in Figure 1, the most used *Gesetzbuch /Norm* is *Zivilprozessordnung*, which translates to Code of Civil Procedure. This law is mainly used to regulate actions in civil courts, where one civilian of a country can sue another to resolve a dispute between them.

Figure 2 shows the total number of occurrences of law reference for the top-10 Gesetzbuch/Norm and article combination. The plot shows that the most used law reference is *Gesetz über die Aufenthalt, die Erwerbstätigkeit und die Integration von Ausländern im Bundesgebiet 1*, which translates to *Act on the Residence, Employment and Integration of Foreigners in Federal Territory 1*. This law is used to govern the residence and employment of nationals from a different country of origin. This high occurrence can be explained by a higher number of case records in the clean Open Legal Data from 2015 and 2016. 2015 was a year of migration crisis across Europe due to several civil wars in Lybia, Syria, and Iraq. Germany accepted a total of 1,091,894 and 321,361 refugees in 2015 and 2016, respectively[13].

4 EXPERIMENTS

In this section, we show how this dataset can be used in practice. In our previous research [3], we identified that case texts that refer to the same law are closely grouped in an embedding space. Therefore, it is safe to say that we can use these embeddings to predict law references, given case text. We also fine-tune the German BERT model to predict law references, given law texts. Here, we perform prediction based on the base law reference (e.g. *§ 97 ZPO.*) and the full law reference (e.g. *§ 97 Abs. 1 ZPO.*).

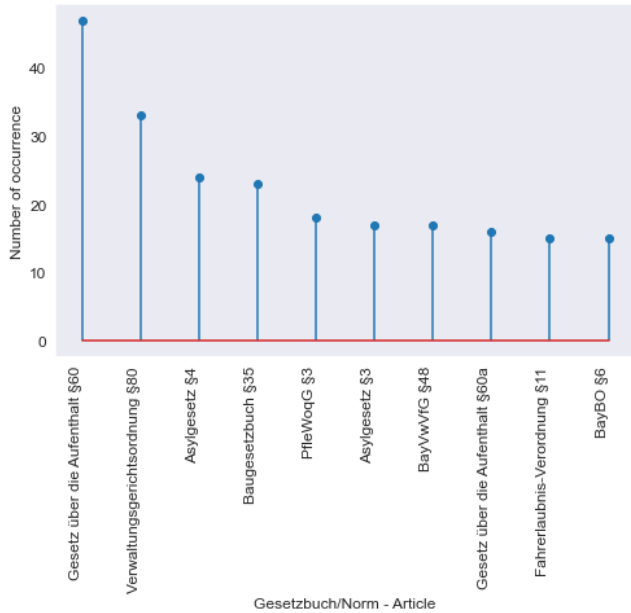[13] https://www.statista.com/statistics/911484/number-newly-registered-refugees-germany/

Figure 2: Number of law reference for top-10 Gesetzbuch/Norm-Article

4.1 Predicting base law references

Here, we experiment with predicting only the base law referred to in the case text, while ignoring the *Absatz* or *Satz* part of the law reference. For example, given two law references, *§ 86b Abs. 1 SGG* and *§ 86b Abs. 2 Satz 2 SGG*, we train the model to only predict only the base law, which in this case is *§ 86b SGG*.

In our dataset, over 50% of data refers to the top 20 base law references, which is approximately 7% of unique base law references available. Therefore, we train our model to predict these top 20 base law references, making this essentially a 20-class classification problem.

In order to make sure the underlying German BERT model is trained on an equal amount of data for each class label, we perform a stratified train-test split of the dataset. Figure 3 shows the performance of fine-tuning the German BERT model.

The bar plot in Figure 4 shows the performance for individual base law references.

4.2 Predicting full law references

Here, we fine-tune the German BERT model to predict full law references as they appear in the case text. Similar to predicting the base law references, we treat this task as a 20-class classification problem. This is because, similar to base law references, over 50% of the dataset refers to the top 20 law references available in our dataset. Here as well, we perform a label-wise stratified train-test split over the dataset.

Figure 5 shows the performance of fine-tuning the German BERT model, while the bar plot in Figure 6 shows the performance for individual full law references.
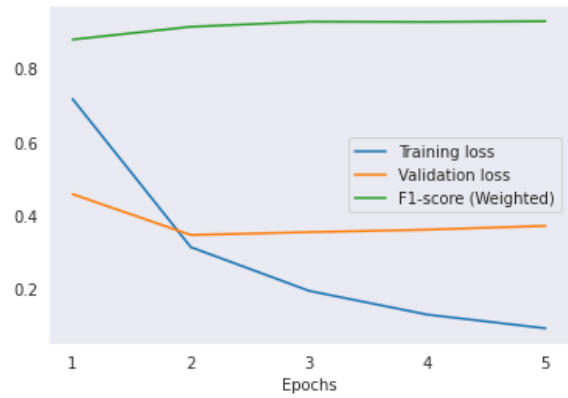

Figure 3: Performance of fine-tuning the German BERT model to predict base law references. The fine-tuned model provides a consistent performance of 0.9 weighted F1-score just after the second epoch.
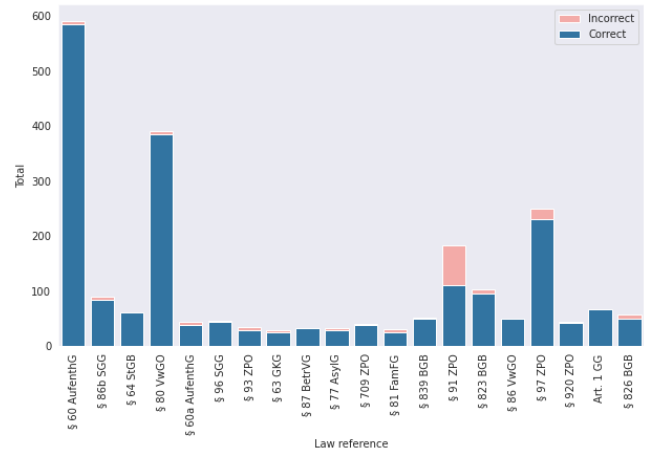

Figure 4: Bar plot visualizing the number of correctly predicted base law references compared to incorrect predictions.

Table 3 shows the validation performance for predicting the base law references and full law references.

Table 3: Accuracy and weighted performances for the base and full law reference prediction.

Metrics	Base law reference	Full law reference
Accuracy	0.92853	0.81336
Weighted-Precision	0.92994	0.80539
Weighted-Recall	0.92853	0.81336
Weighted-F1	0.92707	0.80565

5 CONCLUSION

In this paper, we present a dataset of annotations for legal references done by legal experts. This dataset will enable NLP researchers
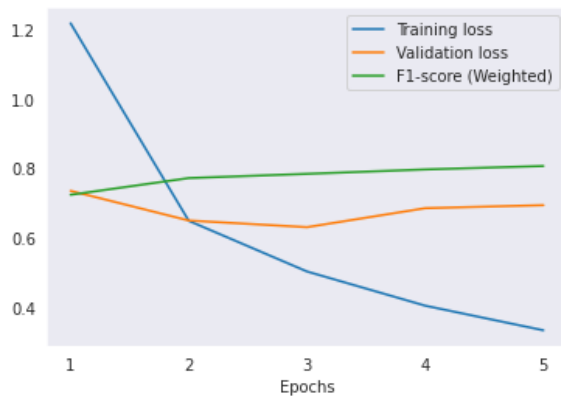
Figure 5: Performance of fine-tuning the German BERT model to predict full law references. After five epochs, the fine-tuned model provides an approximate performance of 0.8 weighted F1-score.
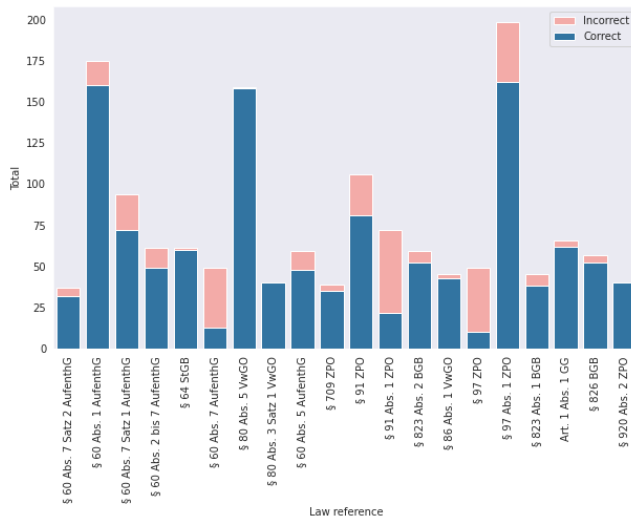

Figure 6: Bar plot visualizing the number of correctly predicted full law references compared to incorrect predictions.

working with legal data to avoid crawling and scrapping data from different sites to gather information about law references.

One of the ways to use this dataset is shown in section 4, where we perform a law reference prediction task, using full law reference and base law reference. Results from this experiment indicate that the task of predicting base law references gives a much better performance than the task of predicting full law references.

The overlapping behavior of semantically similar case text with similar base law references suggests that we can use sentence embeddings to perform the link prediction tasks. This is one of the future tasks that can be performed using this dataset - to identify instances of case texts very similar to queried case text with a specific law reference.

REFERENCES

[1] Anu Bradford, Adam S Chilton, Christopher Megaw, and Nathaniel Sokol. 2019. Competition law gone global: Introducing the comparative competition law and enforcement datasets. *Journal of Empirical Legal Studies* 16, 2 (2019), 411–443.
[2] Branden Chan, Stefan Schweter, and Timo Möller. 2020. German's next language model. *arXiv preprint arXiv:2010.10906* (2020).
[3] Harshil Darji, Jelena Mitrović, and Michael Granitzer. 2023. Exploring Semantic Similarity between German Legal Texts and Referred Laws. *Communications in Computer and Information Science* (2023).
[4] Harshil Darji, Jelena Mitrović, and Michael Granitzer. 2023. German BERT Model for Legal Named Entity Recognition. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART*. INSTICC, SciTePress, 723–728. https://doi.org/10.5220/0011749400003393
[5] Guido Governatori, Trevor Bench-Capon, Bart Verheij, Michał Araszkiewicz, Enrico Francesconi, and Matthias Grabmair. 2022. Thirty years of Artificial Intelligence and Law: the first decade. *Artificial Intelligence and Law* 30, 4 (2022), 481–519.
[6] Elena Leitner, Georg Rehm, and Julián Moreno-Schneider. 2020. A dataset of german legal documents for named entity recognition. *arXiv preprint arXiv:2003.13016* (2020).
[7] Tobias Milz, Michael Granitzer, and Jelena Mitrovic. 2021. Analysis of a German Legal Citation Network. In *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2021, Volume 1: KDIR, Online Streaming, October 25-27, 2021*, Rita Cucchiara, Ana L. N. Fred, and Joaquim Filipe (Eds.). SCITEPRESS, 147–154. https://doi.org/10.5220/0010650800003064
[8] Malte Ostendorff, Till Blume, and Saskia Ostendorff. 2020. Towards an open platform for legal information. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 385–388.
[9] E. L. Rissland and K. D. Ashley. 1987. A Case-Based System for Trade Secrets Law. In *Proceedings of the 1st International Conference on Artificial Intelligence and Law* (Boston, Massachusetts, USA) *(ICAIL '87)*. Association for Computing Machinery, New York, NY, USA, 60–66. https://doi.org/10.1145/41735.41743
[10] Erik F Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050* (2003).
[11] Stefanie Urchs, Jelena Mitrović, and Michael Granitzer. 2020. Towards classifying parts of german legal writing styles in german legal judgments. In *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*. IEEE, 451–454.
[12] Stefanie Urchs, Jelena Mitrović, and Michael Granitzer. 2021. Design and Implementation of German Legal Decision Corpora. 515–521. https://doi.org/10.5220/0010187305150521
[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
[14] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, et al. 2018. Cail2018: A large-scale legal dataset for judgment prediction. *arXiv preprint arXiv:1807.02478* (2018).