Towards Classifying Parts of German Legal Writing Styles in German Legal Judgments

Stefanie Urchs
Chair of Data Science
University of Passau
Passau, Germany
Stefanie.Urchs@uni-passau.de

Jelena Mitrović
Chair of Data Science
University of Passau
Passau, Germany
Jelena.Mitrovic@uni-passau.de

Michael Granitzer
Chair of Data Science
University of Passau
Passau, Germany
Michael.Granitzer@uni-passau.de

***Abstract*—The main tool of a lawyer is their language. Legal prose is bound by writing styles, especially in Germany. These styles ensure that, i.a. judgments are written in a structured and comprehensive way. The writing style used for German judgements is called *Urteilsstil* and consist of several subcomponents. These subcomponents should be classifiable with the help of argumentation mining techniques. However, this classification is currently not possible because an annotated corpus, that considers such special structure of German legal text, is not available. This paper explores possibilities for classifying two subcomponents of the *Urteilsstil* by utilising argumentation mining. Furthermore, the creation of a new corpus for legal classification is proposed.**

Keywords—classification, German legal judgment, argumentation mining, legal writing style, Urteilsstil

I. Introduction

Germany is a country of poets and thinkers; legal professionals take this to heart, as their language is their main professional tool. However, official legal texts such as court decisions cannot be written as pure prose, the writing must be regulated. One form of legal text regulation are the writing styles used in legal texts. One of them is the so called *Urteilsstil* (judgement style), the second main style is *Gutachtenstil* (appraisal style). The first style consists of roughly three components: conclusion, definition and subsumption [1]. This style is used in the explanation of the decision in German court decisions. Classification of these components is no simple task because they could also be nested. Furthermore, these writing styles are more of a guideline than a hard regulation. Passages deemed too trivial are not written down and sometimes subcomponents are combined into one. These problems are comparable to problems that are tackled in the field of argumentation mining. First, arguments are detected in free text. Second, the argumentative propositions within the arguments have to be separated. In this paper, we discuss approaches to classify the definitions and subsumptions contained in German legal judgments by employing argumentation mining. However, to be able to classify text an annotated corpus is necessary. We propose the creation of a new corpus of German court decisions which makes the classification of components of *Urteilsstil* possible.

The remainder of the paper is structured as follows: chapter II discusses the related work. In chapter III the creation of a new corpus for classifying parts of legal writing styles is proposed and chapter IV specifies how classification of definition and subsumption could be conducted. The last section concludes the paper and gives an outlook on future work.

II. Related Work

In this section we investigate other research on text classification in the legal realm, followed by a short introduction to argumentation mining. In the end we discuss the German legal writing style *Urteilsstil*.

A. Text Classification in Legal Documents

Text classification in legal documents is pursued internationally and on many types of corpora. Mostly decisions from various courts and countries are used, predominantly using case law [2 - 5]. Others classify financial regulations with various methods [6 - 8], look at contracts [8; 9] or general legislative texts [10 - 12]. In Germany, researchers concentrate on the classification of norms [13; 14], tax laws [15] and contracts [16].

B. Argumentation Mining

Argumentation mining is a relatively young field situated between natural language processing, information retrieval and argumentation theory. The main aim is to detect arguments in free text. Other goals of argumentation mining cover the detection of structure(s) within the detected argumentation and the detection of the interaction between arguments [17].

The legal domain with its structured texts is an ideal basis for argumentation mining. In fact, one of the first works in the field (from 2007) was performed on legal texts [18]. Argumentation mining on German text is a developing research field. In some works, the argumentative structure in German texts is analysed [19; 20]. Others classify sentences depending on their argumentative content [21]. German texts are also integrated into cross-lingual argumentation mining [22]. The research in argumentation mining in German legal texts is sparser. To the best of our knowledge only one related project for developing a software prototype for automatic detection, analysis as well as recommendation of argumentation structures in court decisions (ARGUMENTUM [23]) exists at the moment. Furthermore, the International Research Group Computer Assisted Legal Linguistics developed a German Reference Corpus (JuReko) that contains "*all statutes of national law (legislation, recorded at one time); decisions and opinions of all federal courts and of a selection of courts at different instances (case law); commentaries, legal papers and articles of academic legal discourse, published in the most important and high*

ranked law journals"[1]. However, the corpus is not publicly available.

C. *Legal Writing Style* Urteilsstil

The *Urteilsstil* begins with the conclusion and proceeds with the reasoning. The most basic version of the *Urteilsstil* consists of three stages: the concrete legal consequences followed by the abstract legal facts and consequences (i.e. exact wording of the law) ending in the concrete facts. Between the abstract legal facts and consequences and the concrete facts a *Feststellungssatz* (determination sentence, the result of the subsumption) states if the legal requirements are fulfilled or not [1].

The determination sentence and the concrete facts together form the subsumption. This leads to the following structure as shown in Fig. 1:

Conclusion	Overall Result	The weather is nice today
Definition	Abstract Legal Facts and Consequences	The weather is always nice, if the sun is shining.
Subsumption	Determination Sentence Concrete Facts	The conditions are met. The sun is shining.

Fig. 1 - Example of "Urteilsstil" with named subcomponents [1]

In practice this basic schema is mostly extended. If norms refer to other regulations, these must first be defined before a subsumption can be made.

III. Corpus

In a supervised machine learning classification scenario, which would be the best suited for German legal writing styles, we need to have an annotated corpus. However, publicly available annotated corpora of German legal texts are sparse, and the proposed classification task is very specialised. No corpus for the classification of components of the *Urteilsstil* exists at the moment.

A good basis for such a corpus are the publicly available court decisions of the Bavarian state that are published on the website www.gesetze-bayern.de. We crawled over 30,000 court decisions from the website, out of which 11,477 are judgments. We randomly select 200 legal judgments for manual annotation. The website also provides a lot of metadata that could be useful for other tasks beside the described classification task. Therefore, two corpora will be published. One consists of over 30,000 court decisions, enriched with the metadata from the website. This metadata contains:

- Name of the court
- Type of the court decision
- Date of the court decision
- Reference number
- Title of the court decision
- Norm chains
- Guiding principles (if applicable)
- Keywords (if applicable)
- Previous court (if applicable)
- Further remarks (if applicable)
- Place of reference

Furthermore, the *Tenor* (purport), *Tatbestand* (facts) and *Entscheidungsgründe* (reasons for the decision) are clearly separated. Only judgments always contain a *Tatbestand*. Additionally, the text paragraphs are separated and numbered in the HTML code. Therefore, these implicit metadata can be easily retained and mapped into a JSON file. Fig 2 shows the resulting JSON format.

For the second corpus 200 judgments out of the 11,477 crawled ones are randomly chosen and annotated by a law expert possessing a first legal state exam. The expert gets an annotation guide that clarifies how *Urteilsstil* is defined for the project. (see definition in II.C.) Further instructions include: annotation of full sentences only; annotations should only be made if the annotator is 100% sure about the decision. Practice has shown that when using experts for annotation tasks it is sufficient to have each label annotated by only one annotator [24]. The annotated corpus will be saved in the JSON format shown in Fig. 3.

IV. Classification

The corpus proposed in Section III is then used to train a model to detect subsumptions and definitions in German judgments. In 2009 Mochales Palau and Moens [25] used argumentation mining to detect argumentation in legal text of the European Court of Human Rights. In the argumentative text the parts of an argument (premise and claim) are classified.

The procedure used in [25] lays the foundation for the classification of subsumption and definition. Firstly, a binary classification decides if the text is worth considering for further classification (detection of argumentative text in free text). All text that is deemed interesting is then classified in subsumption, classification or other (classification of argumentative proposition).

```
{   "meta": {
        "meta_title": "",
        "court": "",
        "decision_style": "",
        "date": "",
        "file_number":"",
        "title": "",
        "norm_chains": "",
        "decision_guidelines": "",
        "keywords": "",
        "lower_court": "",
        "additional_information": "",
        "decision_reference": ""
    },
    "decision_text":{
        "tenor": ["",""],
        "legal_facts": ["",""],
        "decision_reasons": ["",""]
    }
}
```

Fig 2: JSON format of court decision corpus

[1] https://cal2.eu/core-projects-and-associated-projects/german-legal-reference-corpus-jureko-en accessed on 2020-02-14

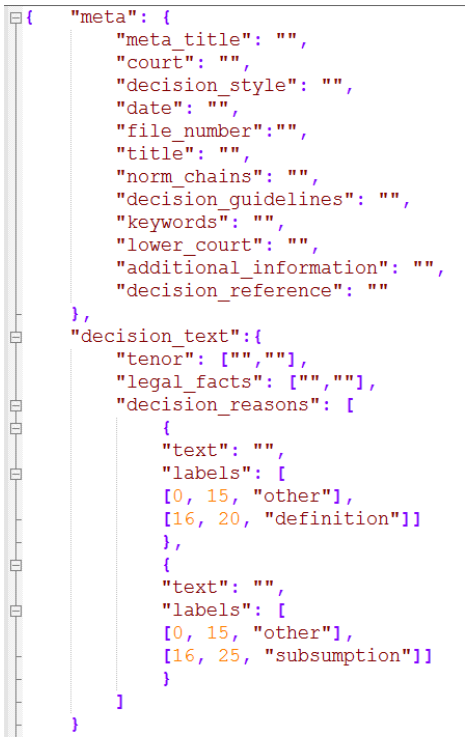

Fig. 3: JSON format of annotated judgment corpus

A. *Preprocessing*

In [25] no preprocessing steps are described. We also refrain from preprocessing, since it is not certain whether this will lead to the loss of important classification information. For example, the tense of verbs might be important. Additionally, words that are considered stop-words in non-legal text might contain valuable information for the classification task at hand.

B. *Features*

As stated above we need to handle two classification tasks. Following the approach in [25] we build our features on sentence level.

1) *Detection of* Urteilsstil

The classification of text that is written in the last two subcomponents of *Urteilsstil* can be compared to the detection of argumentative text in a document. Mochales Palau and Moens [25] suggest for the detection of argumentative text in free text the following features:

- Unigrams
- Bigrams
- Trigrams
- Adverbs
- Verbs
- Modal auxiliary
- Word couples
- Text statistics
- Punctuation sequence
- Key words
- Parse features

In their paper words are used as the smallest unit. We will instead go for tokens. In contrast to words, tokens include special characters like "§". We assume that the paragraph sign carries greate significance when looking for definitions.

Adverbs and verbs are detected with a standard Part of Speech (POS) tagger. We will use the TreeTagger [2] as provided by the Ludwig-Maximilians-Universität München.

In contrast to word couples, as combinations of every two words in the sentence independent of their position, we use the combination of token couples for our classification.

The text statistics consist of the length of a sentence, the average length of words in a sentence and the amount of punctuation marks in a sentence.

Punctuation was already mentioned in the text statistics feature. Still, the sequence of punctuation marks in a sentence is also important. Punctuation marks that are repeating (occur more than once in a row) are regarded as one pattern and marked with a + sign. E.g. two or more periods result in ".+".

We exclude the Key word list feature. No official lists of words to use in *Urteilsstil* are available. It was not possible to generate such a key word list from unofficial resources like web-blogs or forums. All hints given in these resources were too general and sometimes even contradictory.

We are also excluding the parse features. These look at the parse tree of every sentence and take the depth of the tree and the number of subclauses into account.

2) *Seperation of Definition and Subsumption*

Deciding whether a sentence belongs to a definition, subsumption or is something else could be a task comparable to deciding if a sentence belongs to a premise or conclusion. The features suggested by Mochales Palau and Moens [25] are mostly tailored to argumentation in legal cases, rather than the type of legal text that we are dealing with. However, we can still borrow the following features from their approach:

- Absolute sentence location in document
- Sentence length
- Information of first classification

The absolute sentence location in the documents needs to be adapted to the location in relation to the *Entscheidungsgründe* part of the judgment, because the *Tatbestand* and Tenor parts differ greatly in their length through the judgements. Furthermore, the sentence location could be computed relative to the text paragraph as provided by the website.

C. *Classifier*

For both kinds of classification different classification algorithms should be considered. The ones used in the base paper are introduced below.

1) *Detection of* Urteilsstil

Mochales Palau and Moens [25] used two statistical classifiers for their binary classification problem: a maximum entropy model and a multinomial naïve Bayes classifier. They also used a support vector machine (SVM) classifier but did not receive satisfying results. Therefore, we are omitting this classifier.

[2] https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ accessed on 2020-01-22

2) Seperation of Definition and Subsumption

To classify argumentative propositions (which corresponds to our classification into definition, subsumption and other) Mochales Palau and Moens [25] rely once again on a statistical classifier. The SVM that was not well suited for the binary classification of argumentative or not can be used for classifying if a proposition is a premise or a claim. The third class (other) is already detected by the first classifier. Therefore, we can regard this classification problem as a binary one.

D. Extended Classification

The paper of Mochales Palau and Moens [25] was written over ten years ago. Computing power and algorithms have greatly improved since then. Taking this into consideration we can use all proposed features in a multiclass classification setting. Furthermore, the combination of different features can be explored.

V. Conclusion and Future Work

We describe the formation of a new corpus for classifying German legal writing styles in this paper. Additionally, we will publish a second corpus containing all court decisions crawled from the website www.gesetze-bayern.de. This corpus will not be annotated, but it will contain all metadata described in section III.

We explore possibilities to classify the definitions and subsumptions of the *Urteilsstil* in the annotated corpus. We do not limit ourselves to our base paper and introduce different possibilities for classification. The proposed classification will be the content of a master thesis, that will be submitted in April of 2020.

In future work, the detected definitions can be ontologically modelled for argumentation mining, using the procedure of [26]. Furthermore, the German legal terminology can be modelled into the SUMO ontology [27]. Additionally, the trained *Urteilsstil* model can be used on documents written in *Gutachtenstil*, as they share the components Definition and Subsumption.

References

[1] Dr. Jörg Danger. „Urteil und Urteilsstil in der zivilrechtlichen Assessorklausur: eine praktische Hilfestellung". In: Juristische Arbeitsblätter 07 (2005)

[2] Kevin D Ashley and Stefanie Brüuninghaus. "Automatically classifying case texts and predicting outcomes". In: Artificial Intelligence and Law 17.2 (2009), pp. 125-165

[3] Paul Thompson. "Automatic categorization of case law". In: Proceedings of the 8th international conference on Artificial intelligence and law. ACM. 2001, pp. 70-77

[4] Nikolaos Aletras et al. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective". In: PeerJ Computer Science 2 (2016)

[5] Teresa Gonçalves and Paulo Quaresma. "Is linguistic information relevant for the classification of legal texts?" In: Proceedings of the 10th international conference on Artificial intelligence and law. ACM. 2005, pp. 168-176

[6] Alan Buabuchachart et al. "Classification of Regulatory Paragraphs by Discourse Structure, Reference Structure, and Regulation Type." 2013

[7] Kartik Asooja et al. "Semantic annotation of Finance regulatory text using multilabel classification". In: LeDA-SWAn (2015)

[8] James O' Neill et al. "Classifying sentential modality in legal language: a use case in financial regulations, acts and directives". In: Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law. ACM. 2017, pp. 159-168Michael Curtotti and Eric Mccreath. "Corpus based classification of text in Australian contracts". In: Proceedings of the Australasian Language Technology Association Workshop. 2010

[9] Kishore Varma Indukuri and P Radha Krishna. "Mining e-contract documents to classify clauses". In: Proceedings of the Third Annual ACM Bangalore Conference. ACM. 2010, p. 7

[10] Enrico Francesconi and Andrea Passerini. "Automatic classification of provisions in legislative texts". In: Artificial Intelligence and Law 15.1 (2007), pp. 1-17

[11] Emile de Maat, Kai Krabben, Radboud Winkels, et al. "Machine Learning versus Knowledge Based Classification of Legal Texts." In: JURIX. 2010, pp. 87-96

[12] Rob Opsomer et al. "Exploiting Properties of Legislative Texts to Improve Classification Accuracy". In: Legal Knowledge and Information Systems: JURIX 2009, the Twenty-second Annual Conference. Vol. 205. IOS Press. 2009, p. 136

[13] Bernhard Waltl et al. "Classifying Legal Norms with Active Machine Learning." In: JURIX. 2017, pp. 11-20

[14] Bernhard Waltl et al. "Semantic types of legal norms in German laws: classification and analysis using local linear explanations". In: Artificial Intelligence and Law 27.1

[15] Bernhard Waltl et al. "Predicting the outcome of appeal decisions in germany's tax law". In: International Conference on Electronic Participation. Springer. 2017, pp. 89-99

[16] Ingo Glaser, Elena Scepankova, and Florian Matthes. "Classifying Semantic Types of Legal Sentences: Portability of Machine Learning Models." In: JURIX. 2018, pp. 61-70

[17] Marie-Francine Moens. "Argumentation mining: Where are we now, where do we want to be and how do we get there?" In: Post-proceedings of the forum for information retrieval evaluation (FIRE 2013). Vol. 4. ACM; New York. 2013

[18] Marie-Francine Moens et al. "Automatic detection of arguments in legal texts". In: Proceedings of the 11th international conference on Artificial intelligence and law. ACM. 2007, pp. 225-230

[19] Andreas Peldszus. "Towards segment-based recognition of argumentation structure in short texts". In: Proceedings of the First Workshop on Argumentation Mining. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 88-97. doi: 10 . 3115 / v1 / W14 - 2112.

[20] Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. "Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications". In: Proceedings of the 2nd Workshop on Argumentation Mining. Denver, CO: Association for Computational Linguistics, June 2015, pp. 1-11. doi: 10.3115/v1/W15-0501.

[21] Matthias Liebeck, Katharina Esau, and Stefan Conrad. "What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld". In: Proceedings of the Third Workshop on Argument Mining (ArgMining2016). 2016, pp. 144-153

[22] Steffen Eger et al. "Cross-lingual Argumentation Mining: Machine Translation (and a bit of Projection) is All You Need!" In: (July 24, 2018). arXiv: http://arxiv.org/abs/1807.08998v1 [cs.CL]

[23] Constantin Houy et al. "Towards automated identification and analysis of argumentation structures in the decision corpus of the German Federal Constitutional Court". In: 2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST). IEEE. 2013, pp. 72-77

[24] Lars Wissler, Mohammed Almashraee, Dagmar Monett Díaz, and Adrian Paschke. "The gold standard in corpus annotation". In IEEE GSC, 2014.

[25] Raquel Mochales Palau and Marie-Francine Moens. "Argumentation mining: the detection, classification and structure of arguments in text". In: Proceedings of the 12th international conference on artificial intelligence and law. ACM. 2009, pp. 98-107

[26] Mitrović, J., O'Reilly, C., Mladenović, M., & Handschuh, S. (2017). Ontological representations of rhetorical figures for argument mining. Argument & Computation, 8(3), 267-287.

[27] Jelena Mitrović, Adam Pease, Michael Granitzer. "Modeling Legal Terminology in SUMO". In: Proceedings of TOTh 2019, Terminology and Ontology: Theories and Application