

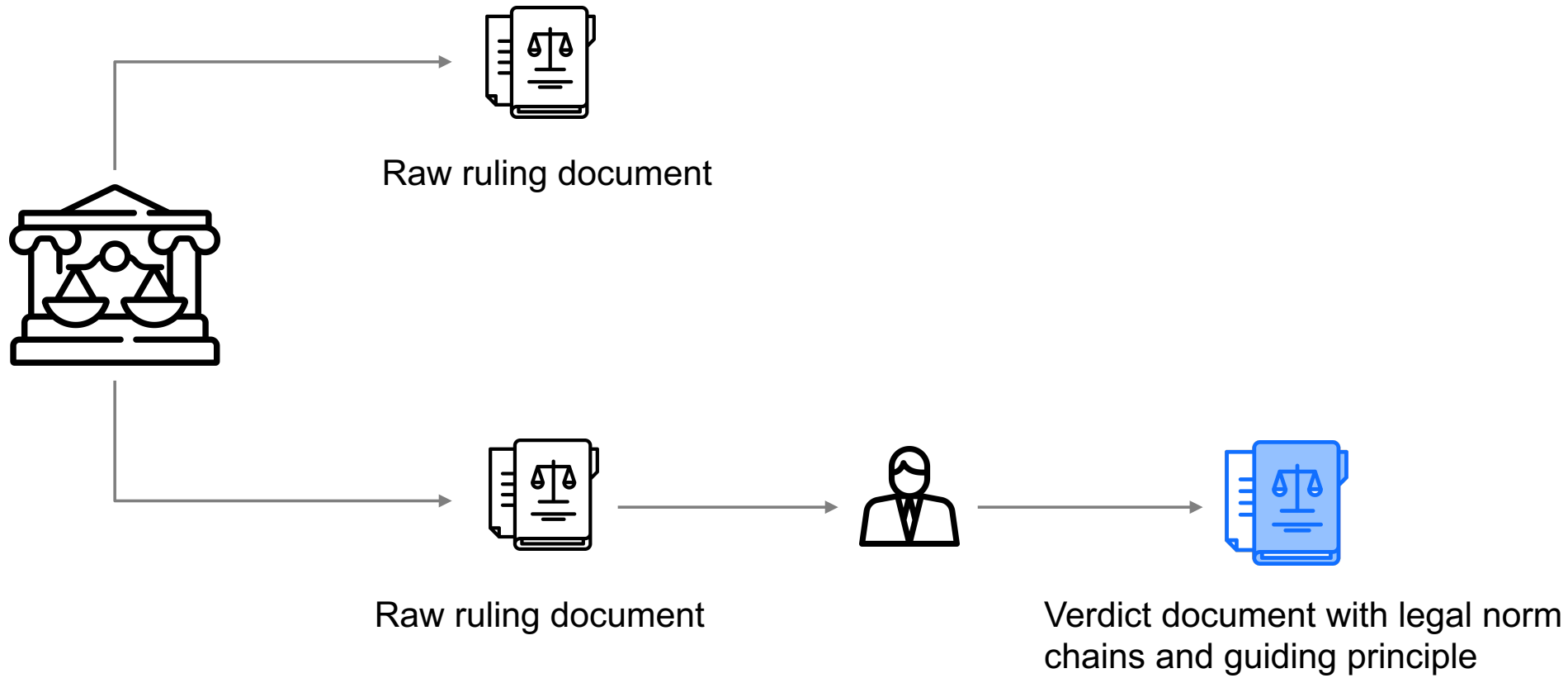
Automatic Norm Chain Generation for German Legal Verdicts

Jieyi Zhang, 20th July 2020, Master Thesis Final Presentation

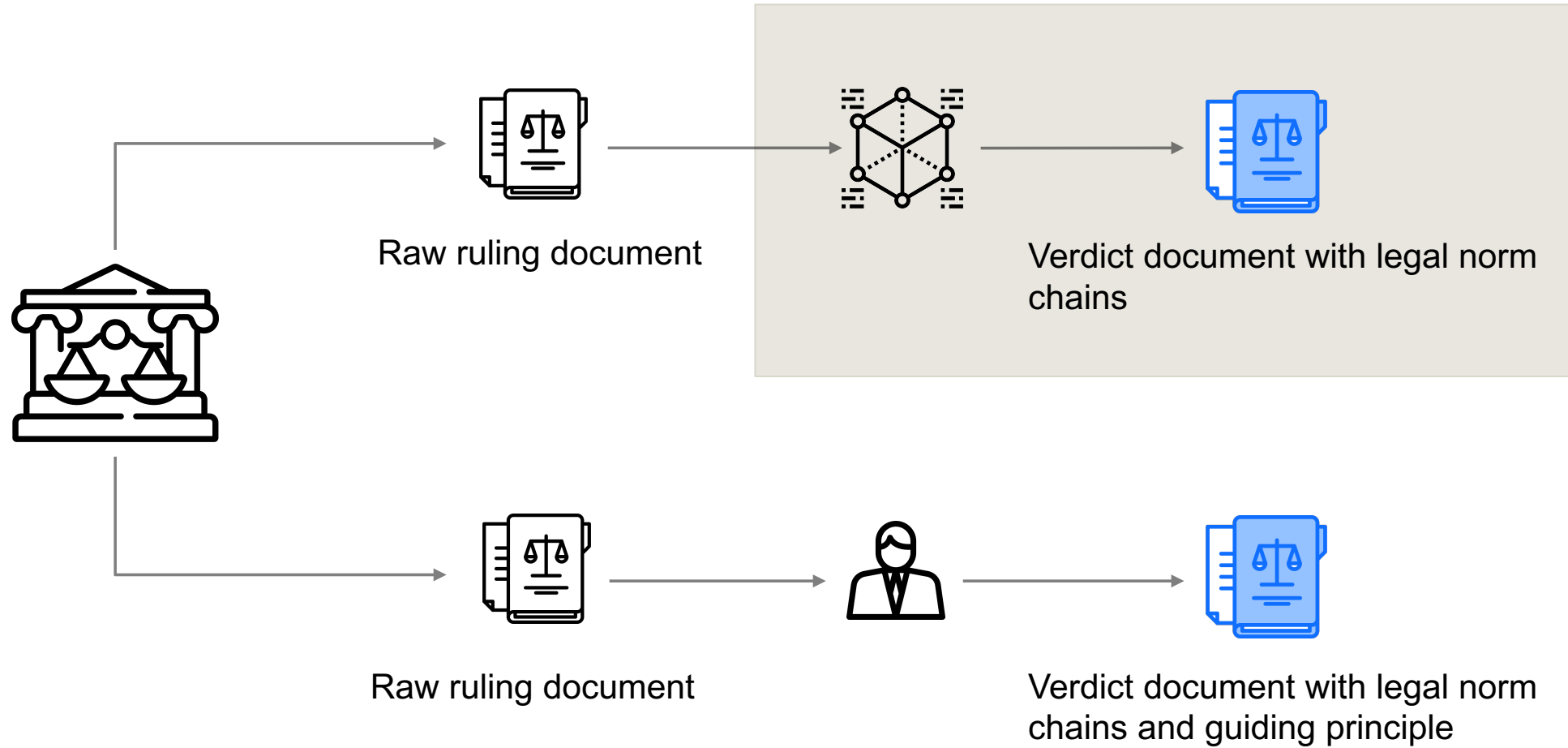
Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook

- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook



Motivation and problem



- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook

Research questions and approaches

- Research questions

RQ 1. How are the norm chains created by judges/legal authors?

RQ 2. How to technically generate the norm chains for each verdict document?

RQ 3. Can norm chains be generated just by the content in the respective verdict document?

Do we need extra dataset?

- Research approaches

- Literature research

- Interview with legal domain experts

- Statistical analysis of given dataset

- Mathematical modeling and experiments to predict the norm chains

- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook



BUNDESGERICHTSHOF

IM NAMEN DES VOLKES

URTEIL

VIII ZR 115/18

Verkündet am: 20. Februar 2019
Ernel,
Justizangestellte
als Urkundsbeamtin
der Geschäftsstelle

Nachschlagewerk: ja
BGHZ: nein
BGHR: ja

in dem Rechtsstreit

BGB § 651 Satz 1 aF, § 433 Abs. 2, § 316, § 315 Abs. 3 Satz 1, § 151, § 157 D,
§ 812 Abs. 1 Satz 2 Alt. 1

a) Stellt ein Krankenhaus in seiner hauseigenen Apotheke patientenindividuell Zytostatika für eine ambulante Behandlung des Patienten in seiner Klinik her, kommt regelmäßig (stillschweigend) eine Bruttopreisabrede zustande, bei der der darin

- 56,606 .xml files of German verdict documents (2010-2016)
 - Among them, 32,893.xml files (58.11 %) have the legal norm chains while the others don't.
- Norm chains (NORMENKETTE) and Norms (NORM) are the labels for each case
- In the documents, the sections highly relevant to our task include:
 - Guiding principle (LEITSATZ)
 - Fact of case (TATBESTAND)
 - Reasons (GRUENDE)
- Within the sections, Referral (VERWEIS-GS) are the tags that annotate most of the legal norms
- Additionally, we use regular expression to extract the missing norms in plain text and check if the norms are correctly annotated.

- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook

1. Rule-based algorithm

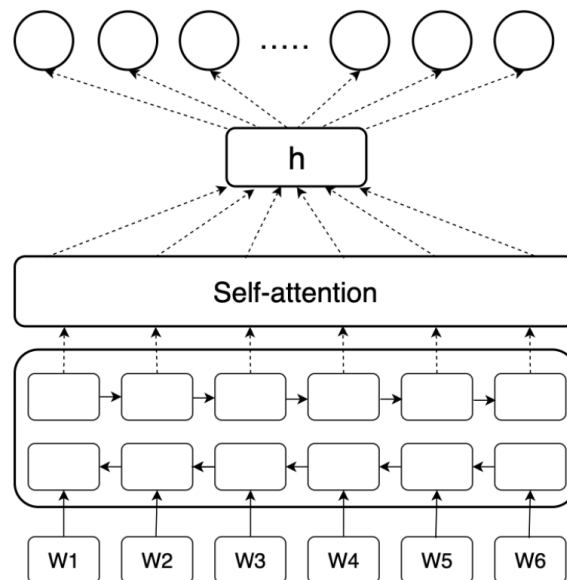
- Extract legal norms from the raw ruling text
- Assigning scores to each norm based on their frequencies and positions
- Pick the candidate norms with scores above a specific threshold

2. Text classification with machine learning models

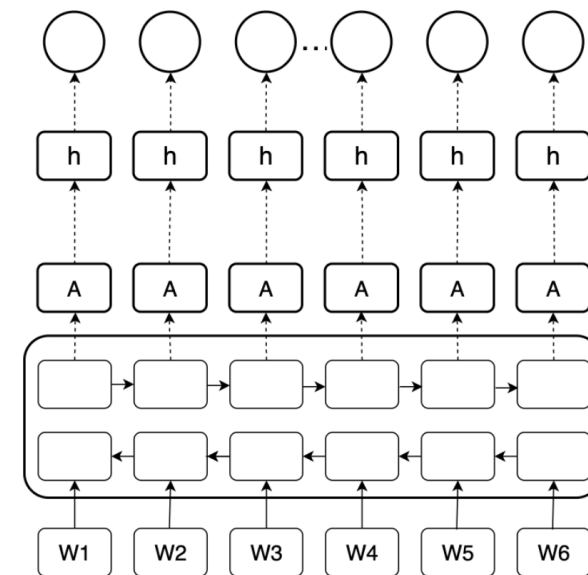
- Approach 1
 - Numeric vector representation of norms as input for each document
 - Entries are the frequencies of the norms
 - Train different classification models and fine-tuned the best performer
- Approach 2
 - Numeric vector representation of norms as input for each document
 - Each index of the vector represent the norm and its position
 - Entries are the frequencies of the norms in a specific section
 - Train different classification models and fine-tuned the best performer

3. Text classification with BIGRU-ATT and BIGRU-LWAN neural network models

- Preprocess input text
- Embed input text with GloVe as the input data
- Train the neural networks for large-scale text classification problem



Bidirectional GRU with self attention (BIGRU-ATT)



Bidirectional GRU with label-wise attention (BIGRU-LWAN)

4. Text summarization with BERT-Transformer model

- Preprocess input text
 - Clean the text data
 - Select fact of case and reasoning sections as input text
 - For applying BERT encoder, truncate and post-pad the input text to the fixed length of 512
- Encoding part of the model is the pre-trained BERT encoder for German language
- The hidden state vectors are then fed into the Transformer decoder part
- Train the whole model to predict the norm chains as summarization

- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook

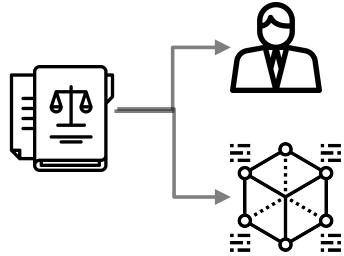
- Research methods: Literature research and Interview
 - 3 Interviews on Zoom
 - Interviewees are experienced lawyer in international law firm and editors in German legal publisher
 - Interview questions:
 - Manual norm chain generation process
 - Prediction results validation
 - Extra input dataset to improve prediction performance
 - Application
- Answer to RQ 1: Manual legal norm chain generation
 - Based on the analysis from the lawyers and legal authors
 - Select most important and relevant norms according to the ruling document and their domain knowledge
 - The sequence of norms doesn't follow a strict rule: logical / alphabetical / numerical... and they don't consider the sequence as a very important attribute in the chains

- Research methods: mathematical modeling and experiments

Model	Input	Code			Code + ParaNr.			Code + ParaNr. + SecNr.		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Rule-based	Text	0.5638	0.7440	0.6415	0.4732	0.6371	0.5431	0.4453	0.5958	0.5097
ML(MLP-2000)	NormVec	0.5834	0.7082	0.6397	0.5668	0.6486	0.6049	-	-	-
ML(MLP-1000)	NormVec w. Position	0.6905	0.9427	0.7971	0.4100	0.4732	0.4393	-	-	-
BIGRU-ATT *	Text-GloVe	0.7066	0.7192	0.7128	0.1655	0.2057	0.1834	-	-	-
BIGRU-LWAN *	Text-GloVe	-	-	-	0.0855	0.1070	0.1128	-	-	-
BERT-Transformer	Text	0.5247	0.4781	0.5003	0.3526	0.3845	0.3679	0.3026	0.3521	0.3255

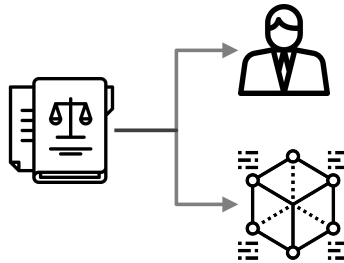
* The BIGRU-ATT model and BIGRU-LWAN model are trained with data including 'LEITSATZ' section

- Prediction demo: BERT-Transformer summarization model



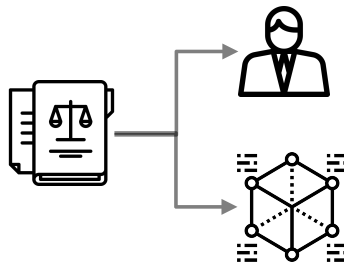
BGB 611, GG Art. 3

BGB 143, GG Art. 3



AO 173 Abs. 1 Nr. 2, AO 175 Abs. 1 S. 1 Nr. 2, EStG 3b

AO 173 Abs. 1 Nr. 1, AO 173 Abs 1, EStG 10 Abs. 1 Nr. 2



BetrVG 102 Abs. 1, BetrVG 102 Abs. 3, KSchG 2

KSchG 1 Abs 2, BetrVG 102 Abs. 2, BetrVG

- Rule-based algorithm
 - Pros:
 - Strong baseline model on all granularity levels
 - Easy to implement and explainable
 - Cons:
 - Cannot generate norms that doesn't exist in current input

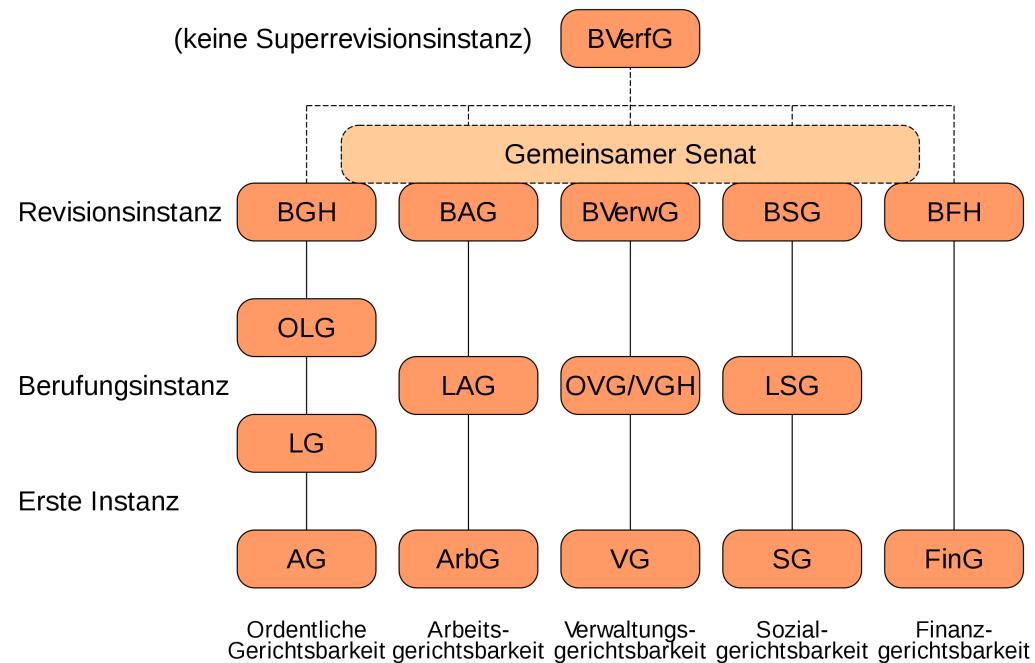
- Text classification with machine learning models
 - Pros:
 - Outperform other methods on more abstractive prediction level
 - Cons:
 - Stiff method, only predict to pre-defined specific format
 - Poor ability to predict more detailed information, e.g. section number

- Text classification with neural network models
 - Pros:
 - No norm extraction step in data processing
 - Cons:
 - Higher complexity without better performance
 - Stiffness in prediction
 - Poor performance in predicting more detailed clauses

- Text summarization with BERT-Transformer model
 - Pros:
 - Can generate 'new' clauses in prediction
 - Can generate more detailed norms and the norm sequences in the prediction
 - More flexibility in prediction
 - Cons:
 - Higher complexity
 - Prediction not accurate enough

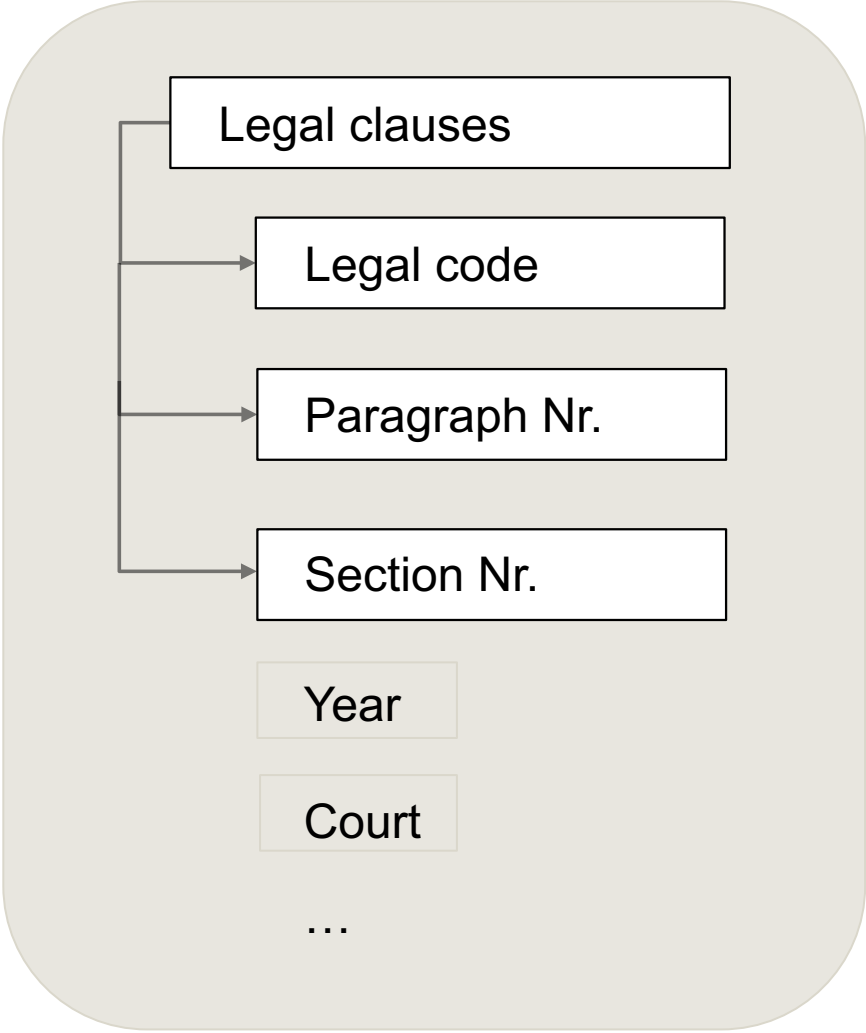
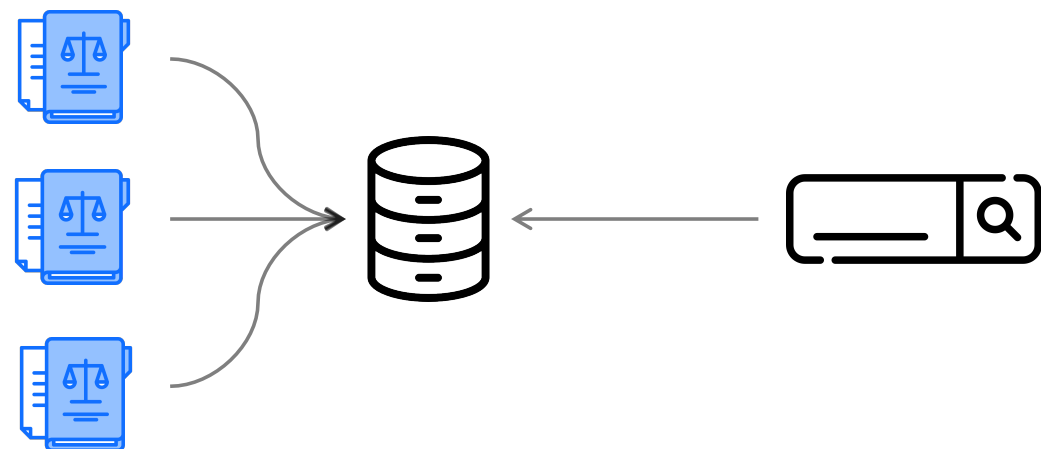
- Answer to RQ 2: Use models and algorithms to generate the legal norm chains
 - On the legal code and paragraph number level: MLP model with norm vectors as input
 - On the section number level: rule-based algorithm or BERT-Transformer neural network
 - More flexibility in the prediction: BERT-Transformer neural network

- Research method: Interview
- Answer to RQ 3: Extra dataset to generate better predictions
 - According to the interview results, additional dataset of legal codes and courts might be helpful
 - Courts information can indicate the type of cases
 - Law code complexity and length sometimes determine the granularity of the norms while being cited



- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential application
- Conclusion and outlook

Potential application



- Motivation and problem
- Research questions and approaches
- Dataset
- Algorithms and models
- Results
- Potential applications
- Conclusion and outlook

Summary

- Design and implement of automatic legal norm chain generation models for German legal verdicts
 - Rule-based model
 - Multi-label classification
 - MLP model
 - BIGRU with different attention mechanism
 - BERT-Transformer text summarization model
- Potential applications for the users in legal domain

Future work

- Adding extra input data to generate precise legal norm chain prediction
- Integration of this automation tool for law firms and legal publishers

- Recht ex machina: Formalisierung des Rechts im Internet der Dienste Oliver Raabe, Richard Wacker, Daniel Oberle, Christian Baumann (2012)
- Distributed Representations of Sentences and Documents, Quoc Le, Tomas Mikolov (2014)
- Extreme Multi-Label Legal Text Classification: A case study in EU Legislation, I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos (2019)
- KEA: Practical Automated Keyphrase Extraction, I. Witten, G. Paynter, E. Frank, C. Gutwin and C. Nevill-Manning (2005)
- Attention is all you need, A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser (2017)
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, J. Devlin, M. W. Chang, K. Lee, K. Toutanova (2018)
- Distributed Representations of Words and Phrases and their Compositionality, T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean (2013)
- GloVe: Global Vectors for Word Representation, J. Pennington, R. Socher, C. D. Manning (2014)
- Sequence to sequence learning with neural networks, I. Sutskever, O. Vinyals, Q. V. Le (2014)
- Neural Machine Translation by Jointly Learning to Align and Translate, D. Bahdanau, K. Cho, Y. Bengio (2014)
- Effective Approaches to Attention-based Neural Machine Translation, M. Luong, H. Pham, C. D. Manning (2015)
- Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification, R. Sathya, A. Abraham (2013)
- Natural Language Processing, E. D. Liddy (2001)
- A Survey on Methods of Abstractive Text Summarization, N. R. Kasture, N. Yargal, N. N. Singh (2014)



BSc.

Jieyi Zhang

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289. 17132
Fax +49.89.289.17136

matthes@in.tum.de
www.matthes.in.tum.de



Backup slides