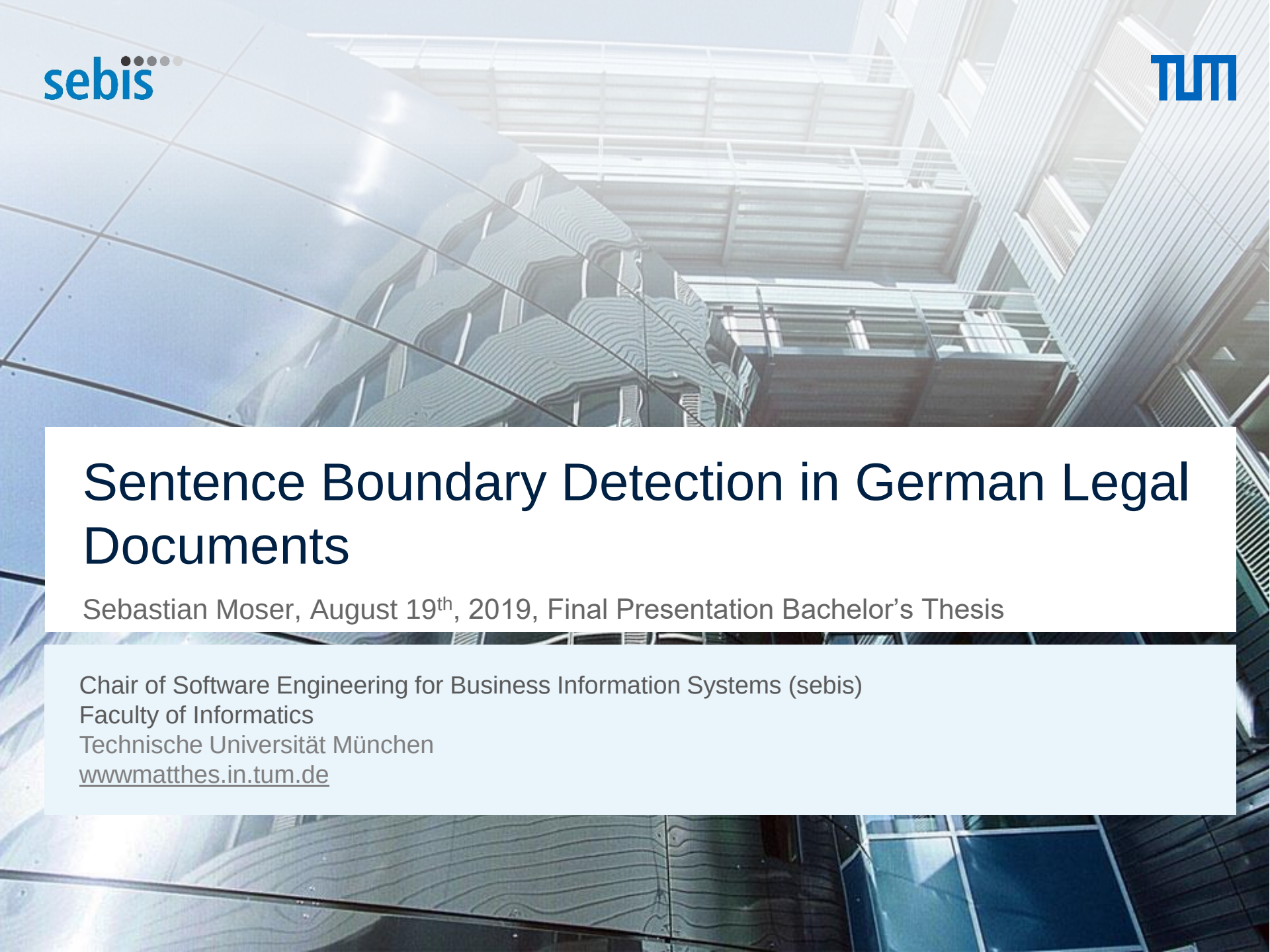


Sentence Boundary Detection in German Legal Documents

Sebastian Moser, August 19th, 2019, Final Presentation Bachelor's Thesis

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

Outline

Introduction
- Motivation
- Research Questions
- Sentence Boundaries in Legal Documents

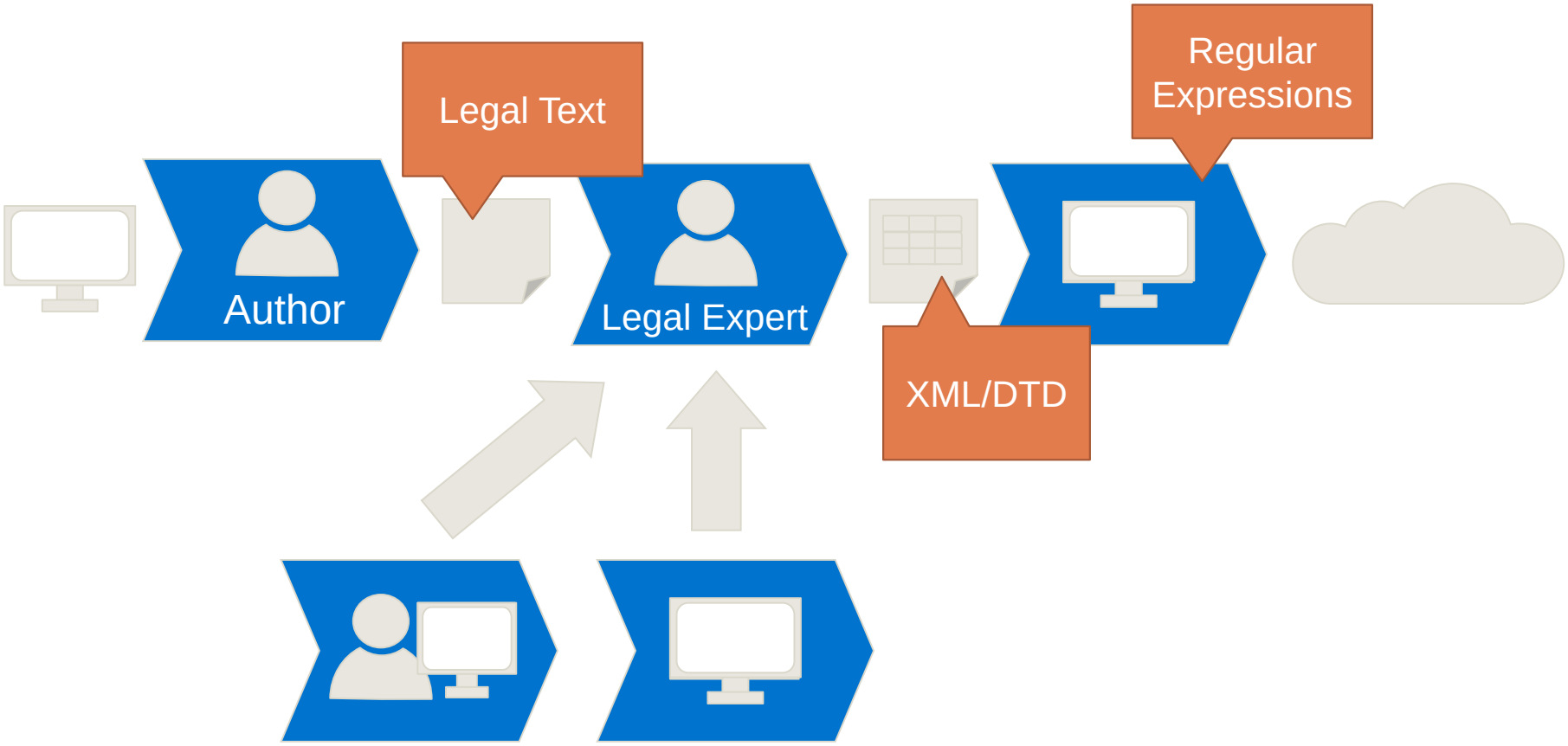
Dataset

SBD System
- Overview
- Existing Approaches
- Rule-Based
- CRF
- NN

Evaluation
- Legal Documents
- Wikipedia Articles
- XML Documents

Conclusion + Demo

Motivation - Publisher



Research Questions

What are sentences in the legal domain?

How should the document corpus be build?

Which methods are state-of-the-art solutions in other domains?

What are the best methods for SBD on German legal documents?

What are the functional/non-functional requirements of the SBD system?

How good are existing approaches on German legal documents?

Are different solutions required for different legal document types?

Research Questions

What are sentences in the legal domain?

How should the document corpus be build?

Which methods are state-of-the-art solutions in other domains?

What are the best methods for SBD on German legal documents?

What are the functional/non-functional requirements of the SBD system?

How good are existing approaches on German legal documents?

Are different solutions required for different legal document types?

Sentence Boundaries in Legal Documents

§ 556g Rechtsfolge; Auskunft über die Mieter[1]

(1) Eine zum Nachteil des Mieters von den Vorschriften dieses Unterkapitels abweichende Vereinbarung ist unwirksam.

(1a) Der Vermieter ist verpflichtet, dem Mieter vor dessen Abgabe der Vertragserklärung über Folgendes unaufgefordert Auskunft zu erteilen:

1. im Fall des § 556e Abs. 1 darüber, wie hoch die Vormiete ein Jahr vor Beendigung des Vormietverhältnisses war, […]

4. im Fall des § 556f Satz 2 darüber, dass es sich um die erste Vermietung nach umfassender Modernisierung handelt. […]

(4) Sämtliche Erklärungen nach den Absätzen 1a bis 3 bedürfen der Textform.

Fußnote

(+++ § 556g: Zur Anwendung vgl. §§ 557a, 557b +++)

(+++ § 556g: Zur Nichtanwendung vgl. § 35 BGBEG +++)

(+++ § 556g: Zur Anwendung vgl. Art. 229 § 49 Abs. 2 BGBEG +++)

Unterkapitel 2

Regelungen über die Miethöhe

§ 557 Mieterhöhungen nach Vereinbarung oder Gesetz

[…]

[1] BGB § 556g shortened, slightly changed

Sentence Boundaries: Abbreviations [1]

§ 556g Rechtsfolge; Auskunft über die Mieter[1]

(1) Eine zum Nachteil des Mieters von den Vorschriften dieses Unterkapitels abweichende Vereinbarung ist unwirksam.

(1a) Der Vermieter ist verpflichtet, dem Mieter vor dessen Abgabe der Vertragserklärung über Folgendes unaufgefordert Auskunft zu erteilen:

1. im Fall des § 556e Abs. 1 darüber, wie hoch die Vormiete ein Jahr vor Beendigung des Vormietverhältnisses war, […]

4. im Fall des § 556f Satz 2 darüber, dass es sich um die erste Vermietung nach umfassender Modernisierung handelt. […]

(4) Sämtliche Erklärungen nach den Absätzen 1a bis 3 bedürfen der Textform.

Fußnote

(+++ § 556g: Zur Anwendung vgl. §§ 557a, 557b +++)

(+++ § 556g: Zur Nichtanwendung vgl. § 35 BGBEG +++)

(+++ § 556g: Zur Anwendung vgl. Art. 229 § 49 Abs. 2 BGBEG +++)

Unterkapitel 2

Regelungen über die Miethöhe

§ 557 Mieterhöhungen nach Vereinbarung oder Gesetz

(1) Während des Mietverhältnisses können die Parteien eine Erhöhung der Miete vereinbaren.

[1] BGB § 556g shortened, slightly changed

Sentence Boundaries: Abbreviations [1]

§ 556g Rechtsfolge; Auskunft über die Mieter[1]

(1) Eine zum Nachteil des Mieters von den Vorschriften dieses Unterkapitels abweichende Vereinbarung ist unwirksam.

(1a) Der Vermieter ist verpflichtet, dem Mieter vor dessen Abgabe der Vertragserklärung über Folgendes unaufgefordert Auskunft zu erteilen:

1. im Fall des § 556e Abs. 1 darüber, wie hoch die Vormiete ein Jahr vor Beendigung des Vormietverhältnisses war, […]

4. im Fall des § 556f Satz 2 darüber, dass es sich um die erste Vermietung nach umfassender Modernisierung handelt. […]

(4) Sämtliche Erklärungen nach den Absätzen 1a bis 3 bedürfen der Textform.

Fußnote

(+++ § 556g: Zur Anwendung vgl. §§ 557a, 557b +++)

(+++ § 556g: Zur Nichtanwendung vgl. § 35 BGBEG +++)

(+++ § 556g: Zur Anwendung vgl. Art. 229 § 49 Abs. 2 BGBEG +++)

Unterkapitel 2

Regelungen über die Miethöhe

§ 557 Mieterhöhungen nach Vereinbarung oder Gesetz

(1) Während des Mietverhältnisses können die Parteien eine Erhöhung der Miete vereinbaren.

[1] BGB § 556g shortened, slightly changed

Sentence Boundaries: Mandatory Positions [2]

§ 556g Rechtsfolge; Auskunft über die Mieter[1]

(1) Eine zum Nachteil des Mieters von den Vorschriften dieses Unterkapitels abweichende Vereinbarung ist unwirksam.

(1a) Der Vermieter ist verpflichtet, dem Mieter vor dessen Abgabe der Vertragserklärung über Folgendes unaufgefordert Auskunft zu erteilen:

1. im Fall des § 556e Abs. 1 darüber, wie hoch die Vormiete ein Jahr vor Beendigung des Vormietverhältnisses war, […]

4. im Fall des § 556f Satz 2 darüber, dass es sich um die erste Vermietung nach umfassender Modernisierung handelt. […]

(4) Sämtliche Erklärungen nach den Absätzen 1a bis 3 bedürfen der Textform.

Fußnote

(+++ § 556g: Zur Anwendung vgl. §§ 557a, 557b +++)

(+++ § 556g: Zur Nichtanwendung vgl. § 35 BGBEG +++)

(+++ § 556g: Zur Anwendung vgl. Art. 229 § 49 Abs. 2 BGBEG +++)

Unterkapitel 2

Regelungen über die Miethöhe

§ 557 Mieterhöhungen nach Vereinbarung oder Gesetz

(1) Während des Mietverhältnisses können die Parteien eine Erhöhung der Miete vereinbaren.

[1] BGB § 556g shortened, slightly changed

Sentence Boundaries: Legal Domain [3][4]

§ 556g Rechtsfolge; Auskunft über die Mieter[1]

(1) Eine zum Nachteil des Mieters von den Vorschriften dieses Unterkapitels abweichende Vereinbarung ist unwirksam.

(1a) Der Vermieter ist verpflichtet, dem Mieter vor dessen Abgabe der Vertragserklärung über Folgendes unaufgefordert Auskunft zu erteilen:

1. im Fall des § 556e Abs. 1 darüber, wie hoch die Vormiete ein Jahr vor Beendigung des Vormietverhältnisses war, […]

4. im Fall des § 556f Satz 2 darüber, dass es sich um die erste Vermietung nach umfassender Modernisierung handelt. […]

(4) Sämtliche Erklärungen nach den Absätzen 1a bis 3 bedürfen der Textform.

Fußnote

(+++ § 556g: Zur Anwendung vgl. §§ 557a, 557b +++)

(+++ § 556g: Zur Nichtanwendung vgl. § 35 BGBEG +++)

(+++ § 556g: Zur Anwendung vgl. Art. 229 § 49 Abs. 2 BGBEG +++)

Unterkapitel 2

Regelungen über die Miethöhe

§ 557 Mieterhöhungen nach Vereinbarung oder Gesetz

(1) Während des Mietverhältnisses können die Parteien eine Erhöhung der Miete vereinbaren.

[1] BGB § 556g shortened, slightly changed

Outline

Introduction
- Motivation
- Research Questions
- Sentence Boundaries in Legal Documents

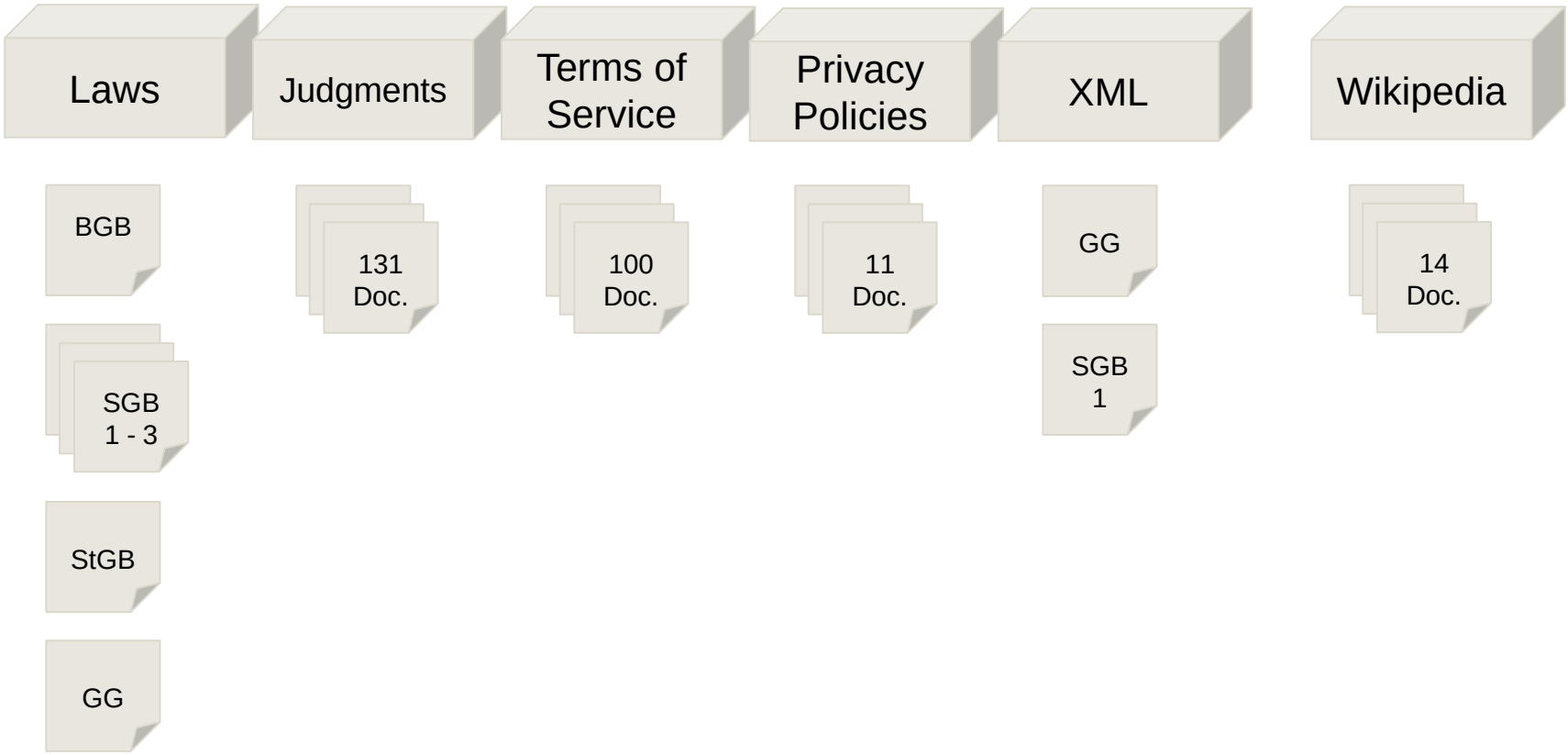
Dataset

SBD System
- Overview
- Existing Approaches
- Rule-Based
- CRF
- NN

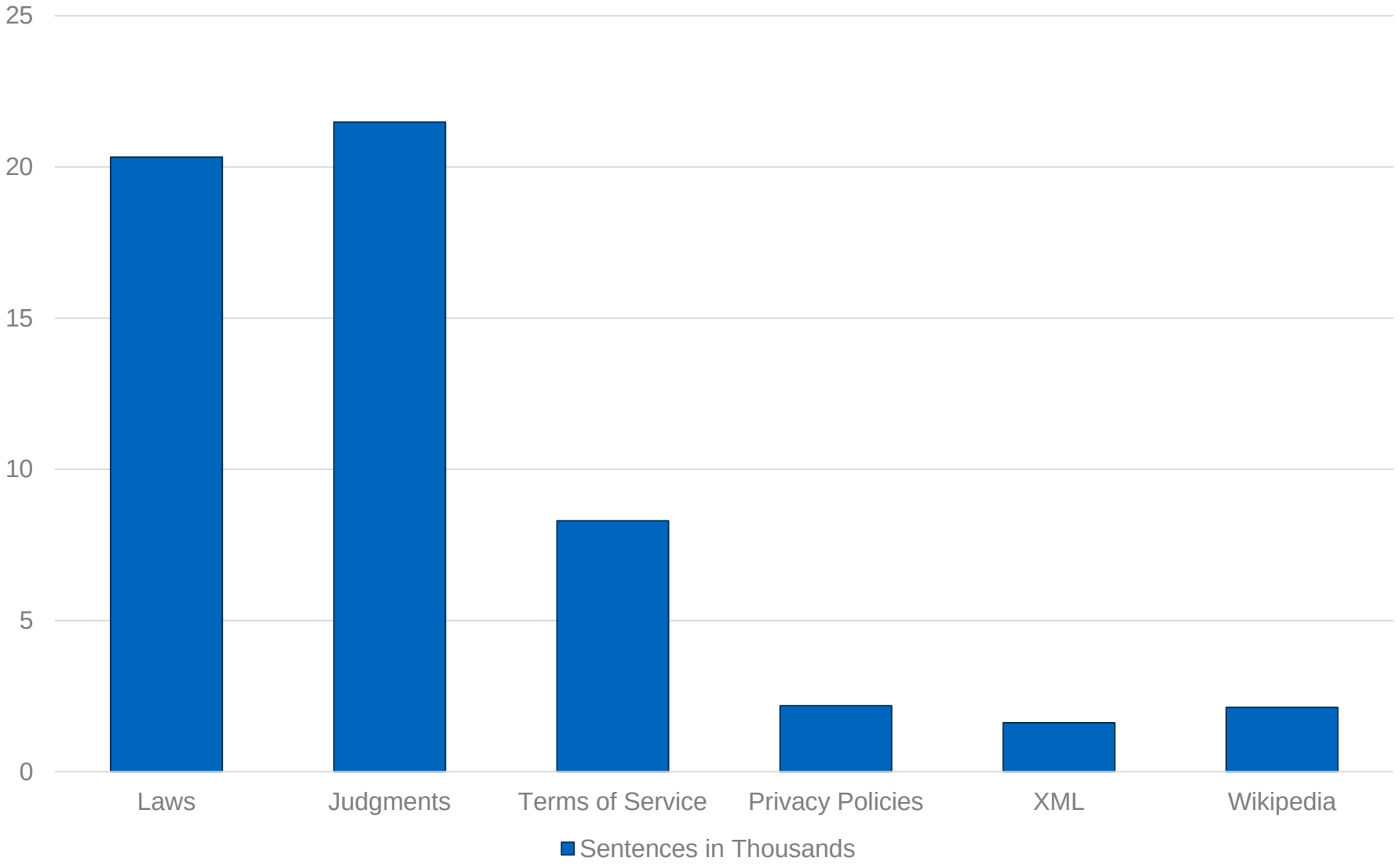
Evaluation
- Legal Documents
- Wikipedia Articles
- XML Documents

Conclusion + Demo

Dataset



Dataset



Outline

Introduction
- Motivation
- Research Questions
- Sentence Boundaries in Legal Documents

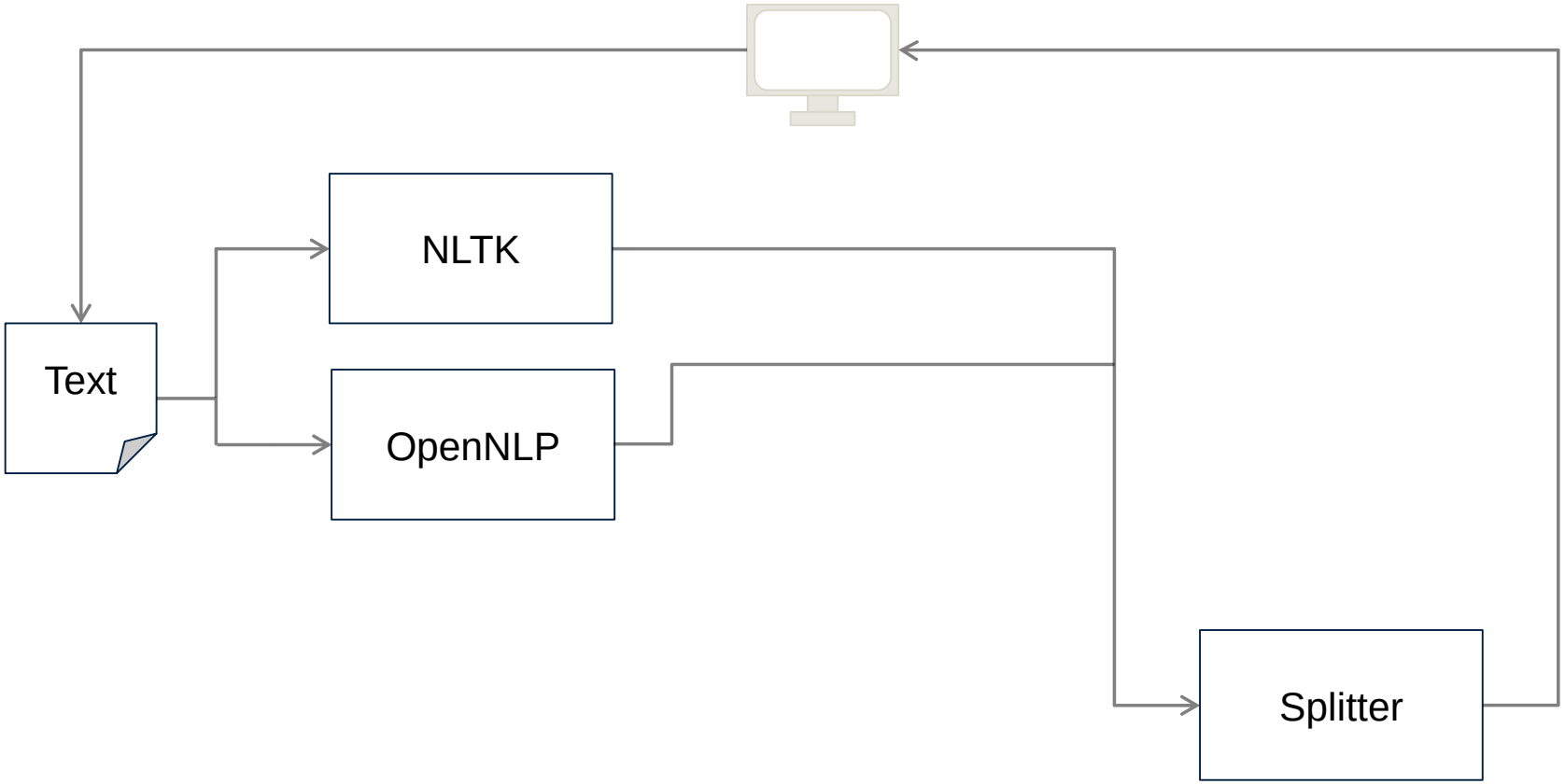
Dataset

SBD System
- Overview
- Existing Approaches
- Rule-Based
- Conditional Random Fields
- Recurrent Neural Network

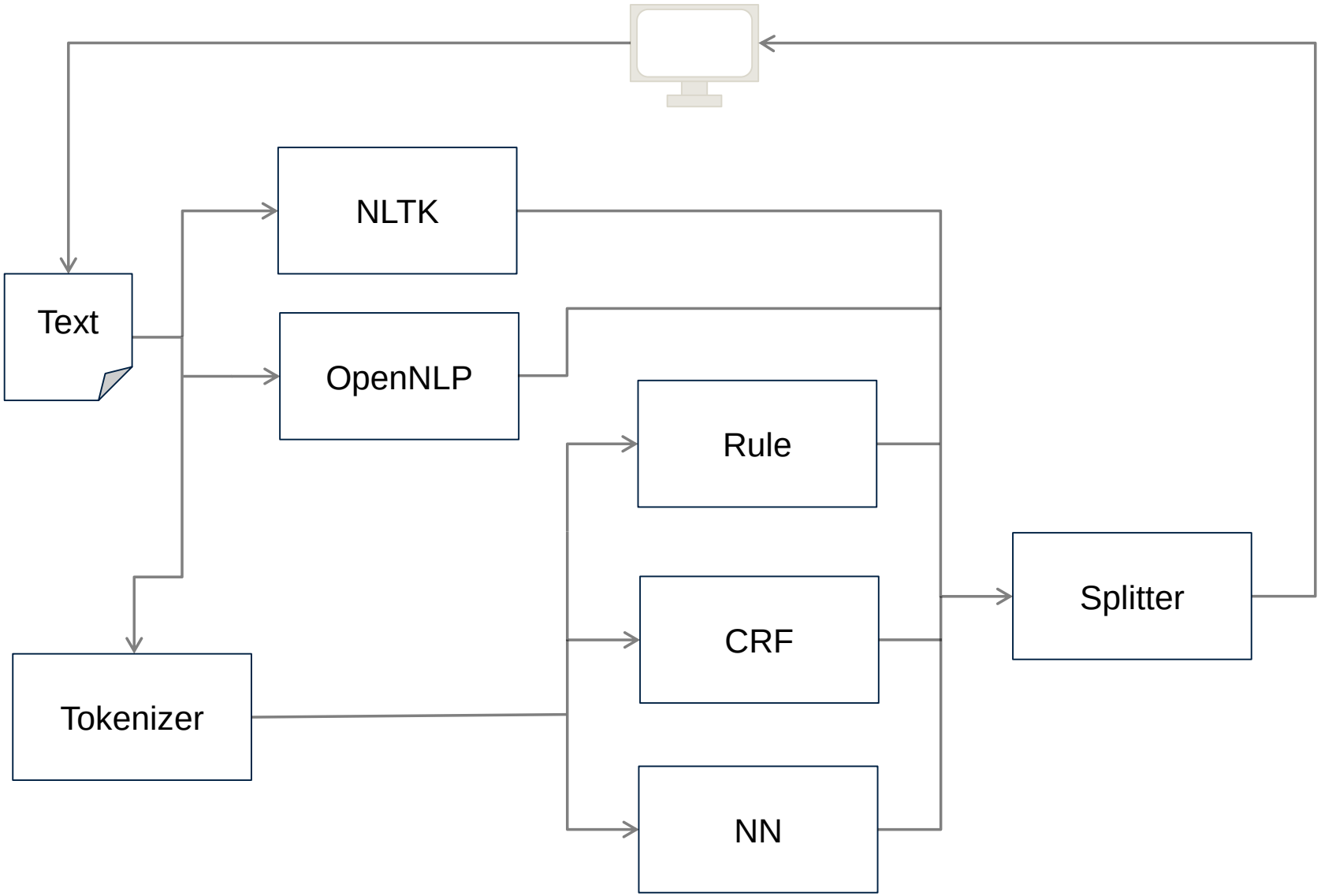
Evaluation
- Legal Documents
- Wikipedia Articles
- XML Documents

Conclusion + Demo

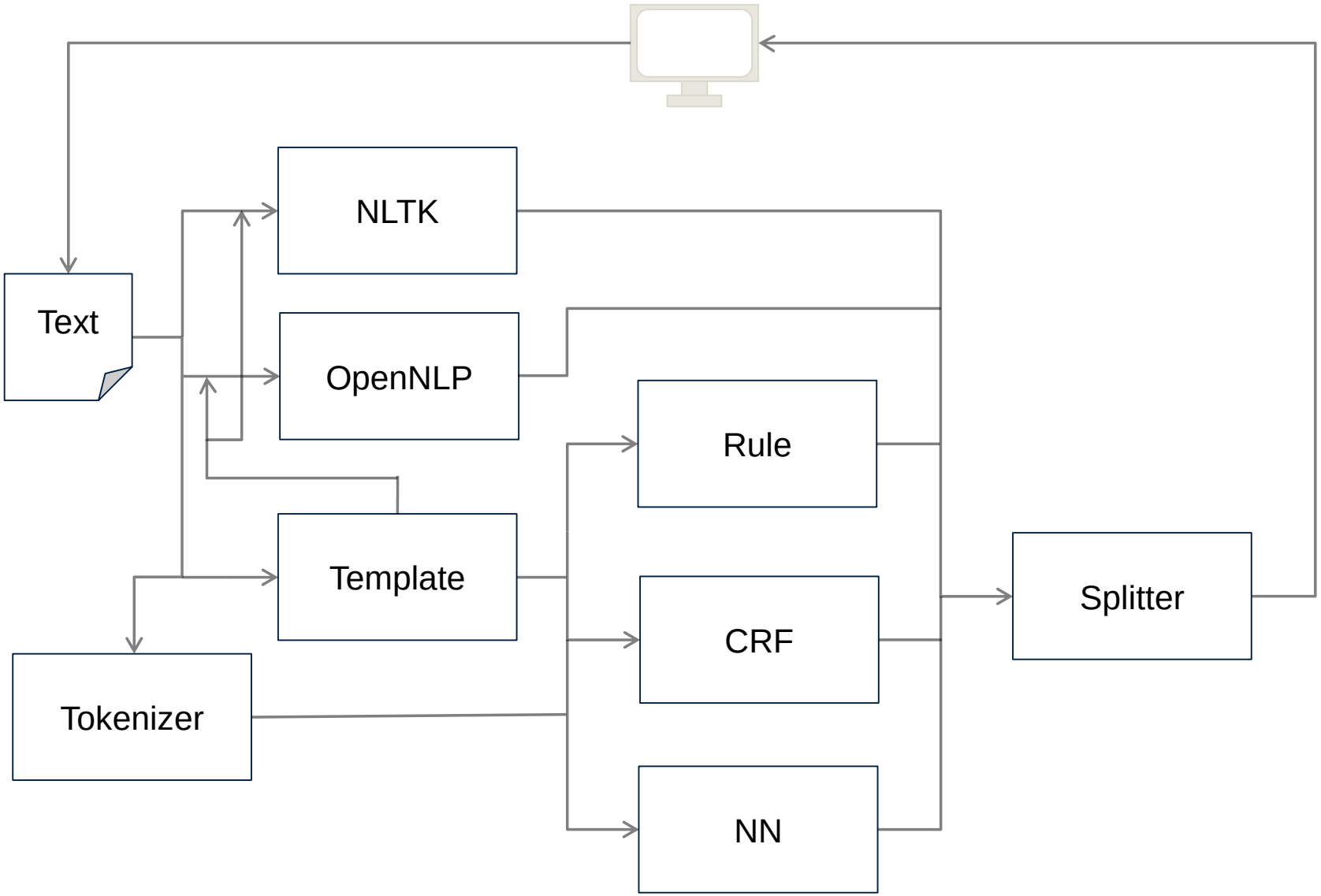
Overview



Overview



Overview



Existing Approaches

NLTK/Punkt [5]

- **Unsupervised**
- Abbreviation Disambiguation
- Hypothesis testing
- Python

OpenNLP [6]

- **Supervised**
- Statistical model with hardcoded features
- Java

Rule-Based

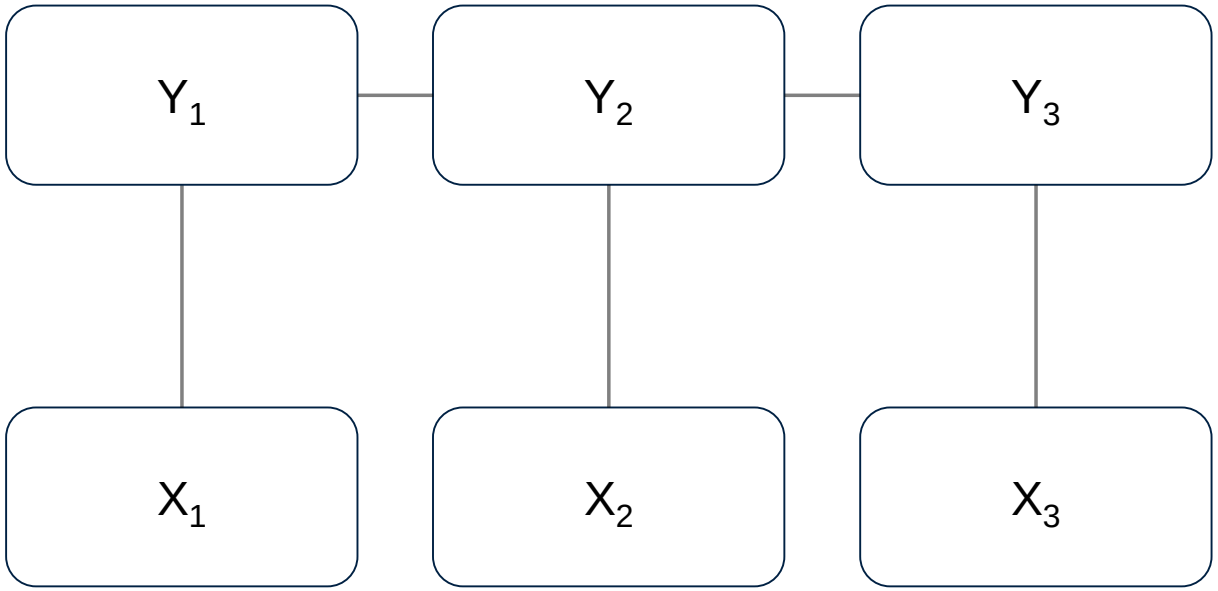
- Definition of **rules** for sentence boundaries (SB)
- Rules based on **context window** and regular expressions
- **Positive** rules → SB
- **Negative** rules → remove SB

Paragraph
[§|Upper, Number, AlphaNumeric, \n]
↔ § 81 Stiftungsgeschäft \n

Abbreviation
[Abs|Art|Rn|Urt|Buchst|bzw|…]

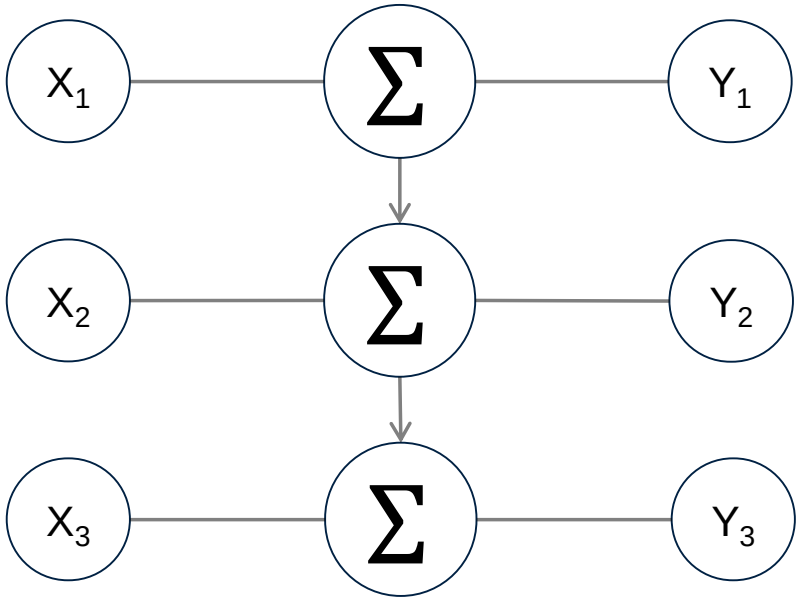
Conditional Random Fields

- Statistical model for **sequence modelling**
- Label probability inferred via predefined **features** for individual tokens
- Features used: *Special, Lowercase, Length, Signature, Lower, Upper, Number*
- Implementation with CRFSuite
- Dependencies between input/output sequence → **linear-chain CRF**



Recurrent Neural Network

- Recurrent Neural Networks **keep information** from previous processing steps
- Input: **Word2vec** word embeddings (pretrained + trained on corpus)
  + **Context** around token
- Implementation in PyTorch
- Combination of **bidirectional recurrent** (RNN, LSTM, GRU) and **linear processing units**



Outline

Introduction
- Motivation
- Research Questions
- Sentence Boundaries in Legal Documents

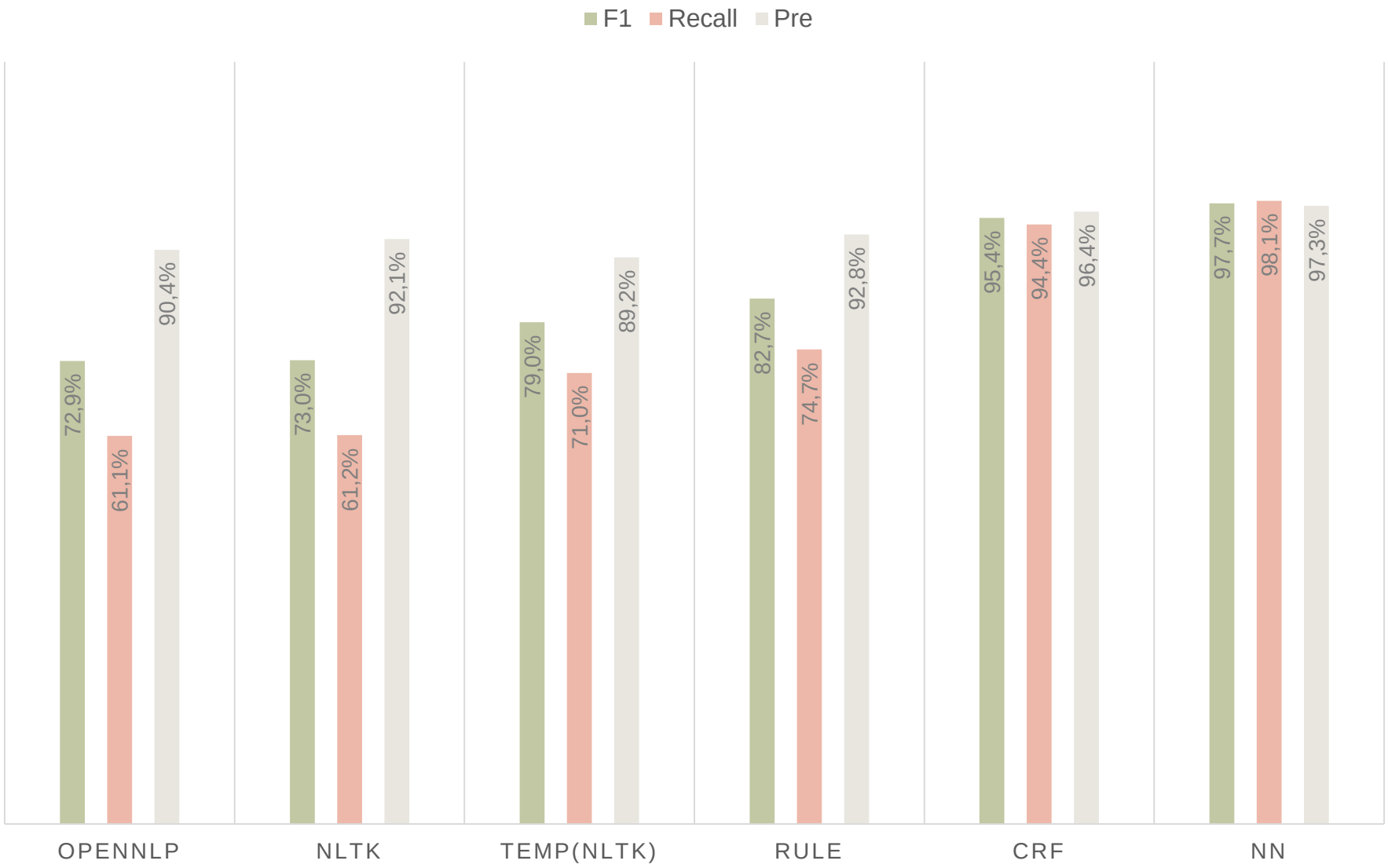
Dataset

SBD System
- Overview
- Existing Approaches
- Rule-Based
- Conditional Random Fields
- Recurrent Neural Network

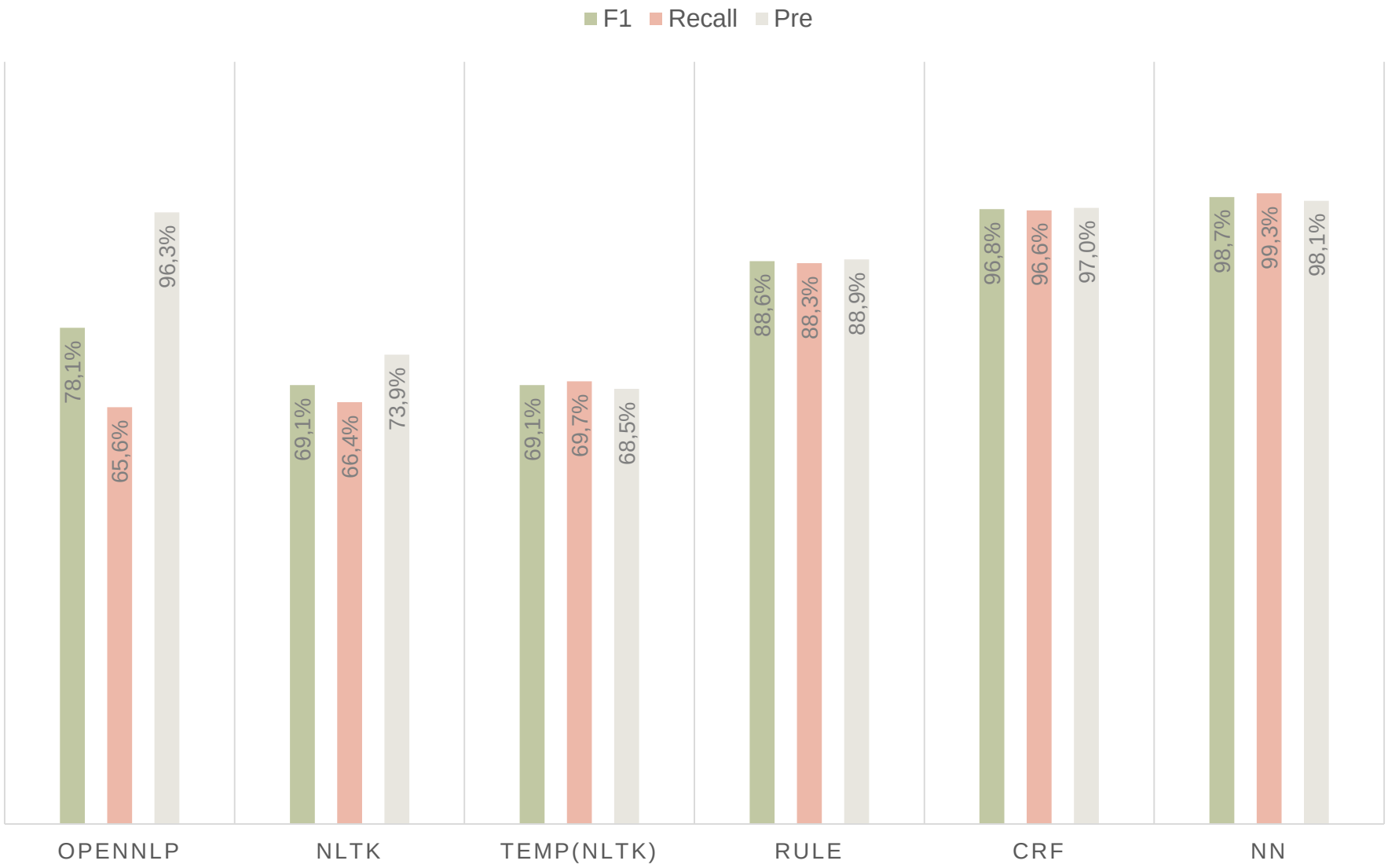
Evaluation
- Legal Documents
- Wikipedia Articles
- XML Documents

Conclusion + Demo

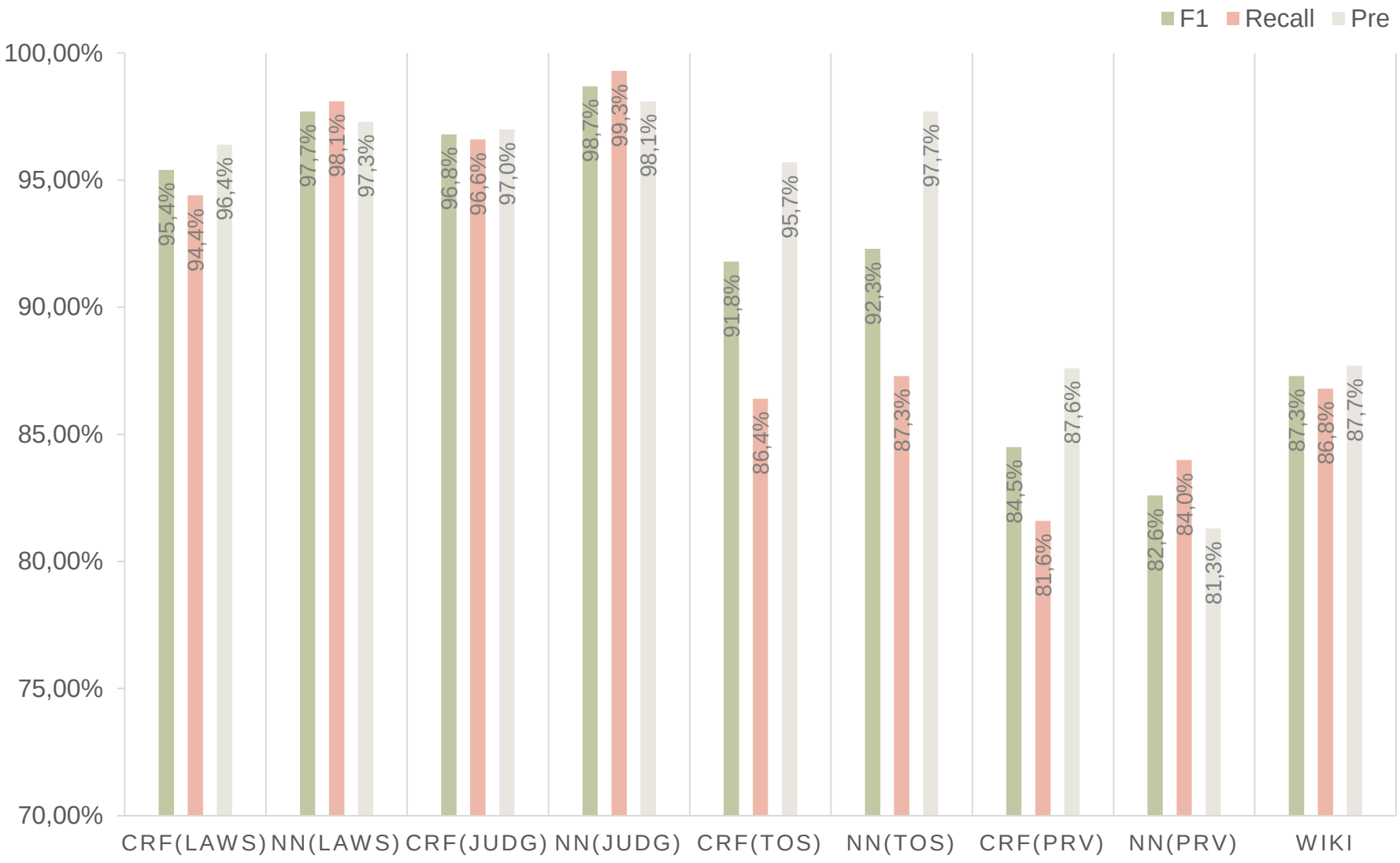
Legal Documents: Laws



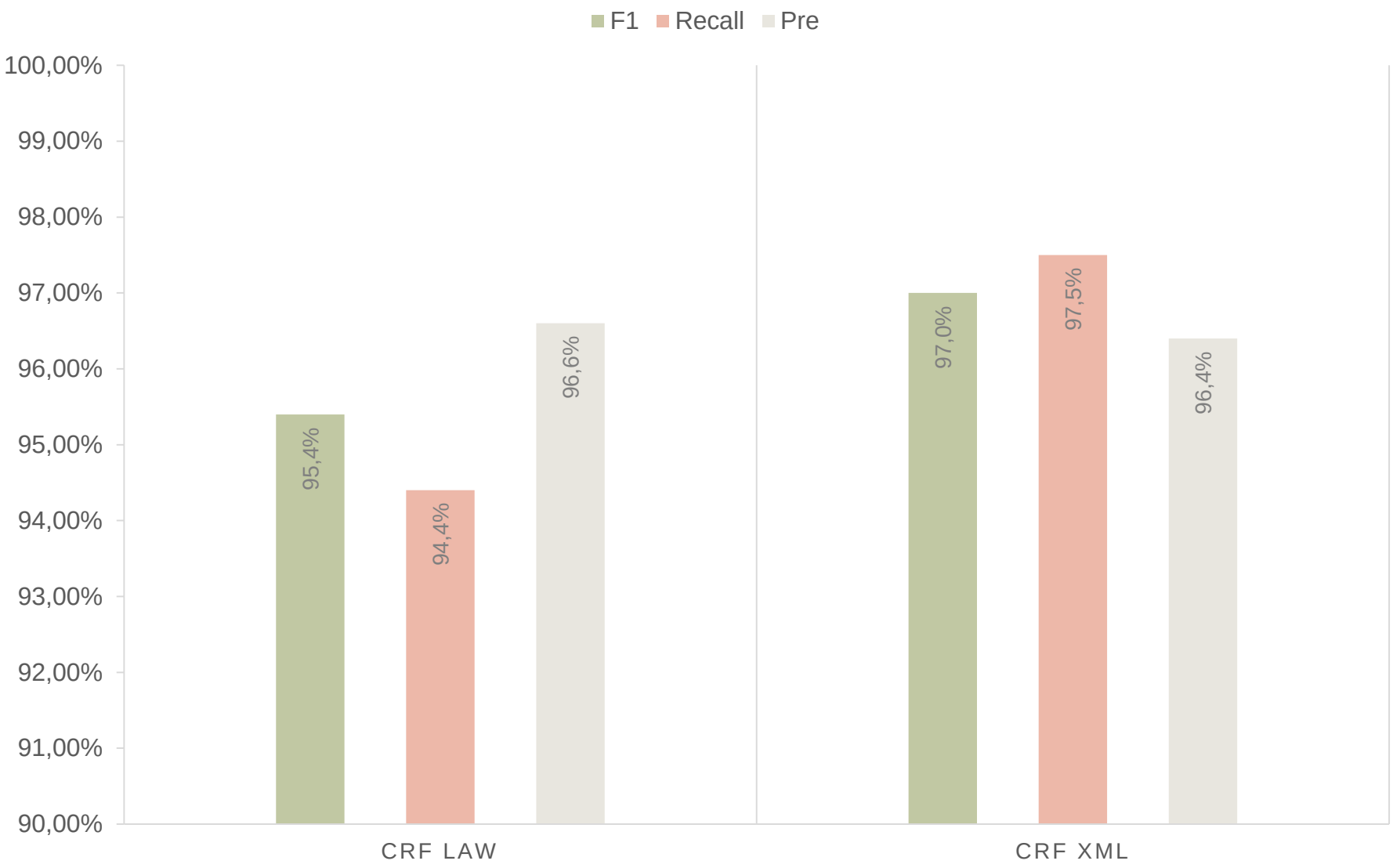
Legal Documents: Judgments



Legal Documents: Privacy Policies, Terms of Service



XML Documents



Outline

Introduction
- Motivation
- Research Questions
- Sentence Boundaries in Legal Documents

Dataset

SBD System
- Overview
- Existing Approaches
- Rule-Based
- Conditional Random Fields
- Recurrent Neural Network

Evaluation
- Legal Documents
- Wikipedia Articles
- XML Documents

Conclusion + Demo

Conclusion

Implementation of **tailored SBD system** for the German legal domain

Creation of **SBD dataset** for the German legal domain

Performance evaluation:

→ **Existing** solutions **not useful** for legal texts

→ **Highly specialized** methods needed

→ **state-of-the-art** results with recurrent neural networks

Legal documents are harder to process than normal text

Demonstration

References

[1] Reynar, J. C.; Ratnaparkhi, A.: *A Maximum Entropy Approach to Identifying Sentence Boundaries.* In *Proceedings of COLING 2012: Posters.* Pages 985-994. Mumbai, India. December 2012.

[2] Mikheev, A.: *Tagging Sentence Boundaries.* In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference.* NAACL 2000. Pages 264-271. Stroudsburg, PA, USA. 2000.

[3] de Maat, E.: *Making sense of legal texts.* PhD thesis. University of Amsterdam. 2012.

[4] Savelka, J.; Ashley, K. D.: *Sentence Boundary Detection in Adjudicatory Decisions in the United States. Traitement automatique des langues.* 58(February):21-45. 2017.

[5] Natural Language Toolkit. https://www.nltk.org. Last Access: August 16, 2019

[6] Apache OpenNLP. https://opennlp.apache.org/. Last Access: August 16, 2019

[7] Okazaki, N.: *CRFSuite: a fast implementation of Conditional Random Fields (CRFs).* 2007.

Sebastian Moser

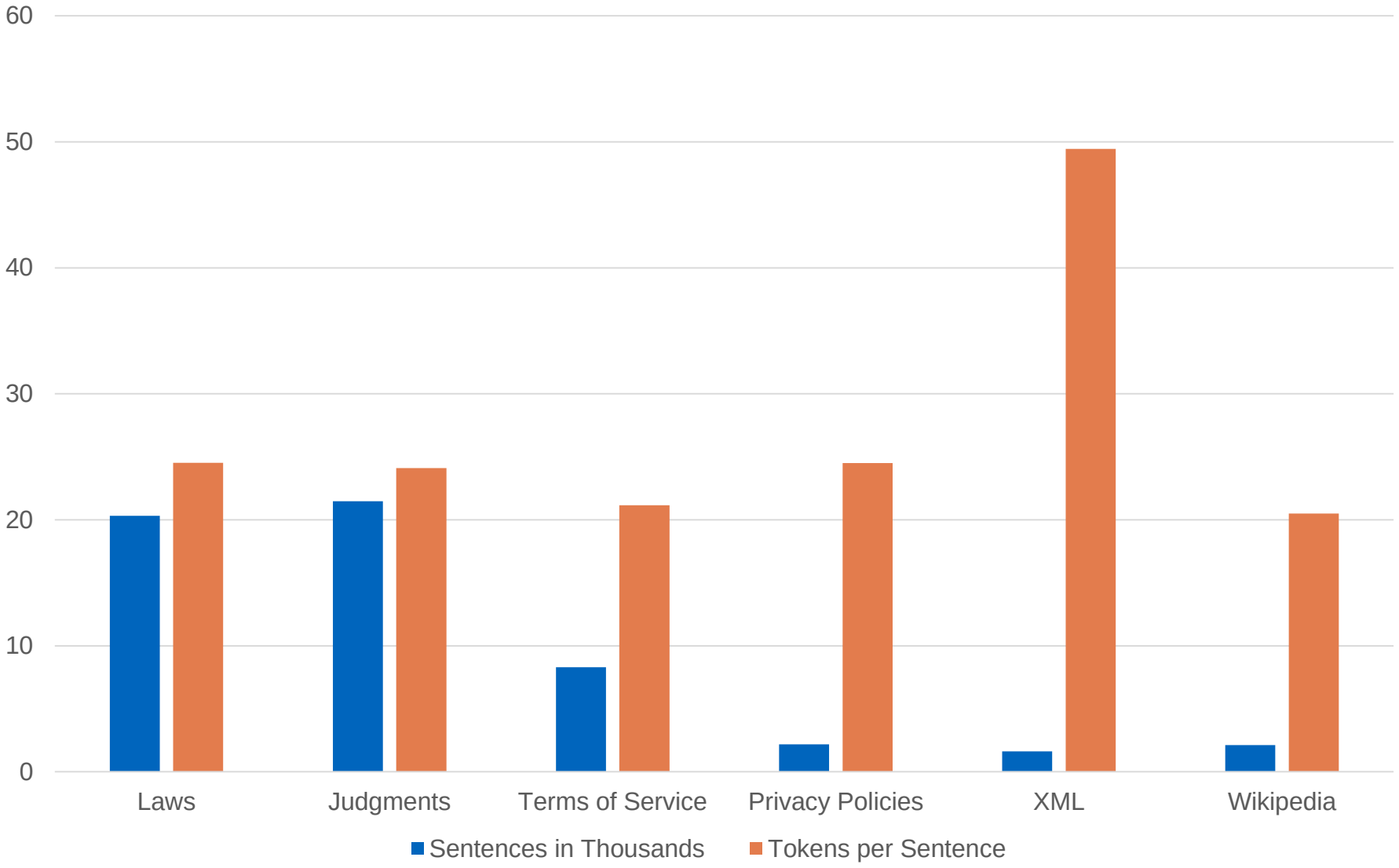
Technische Universität München
Faculty of Informatics
Chair of Software Engineering for
Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

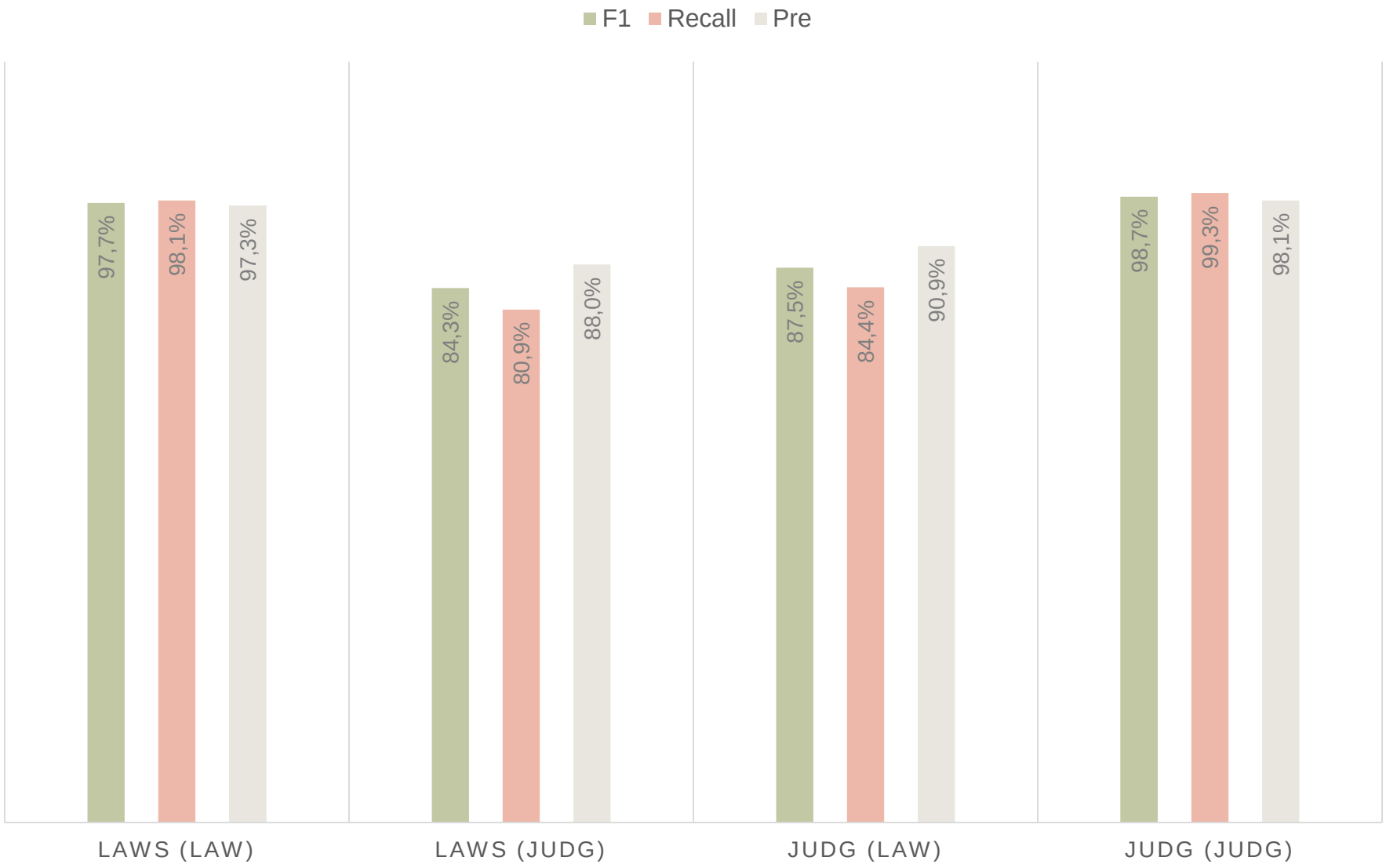
Tel +49.89.289.17132
Fax +49.89.289.17136

matthes@in.tum.de
wwwmatthes.in.tum.de

Dataset



Neural Network Performance on Legal Documents

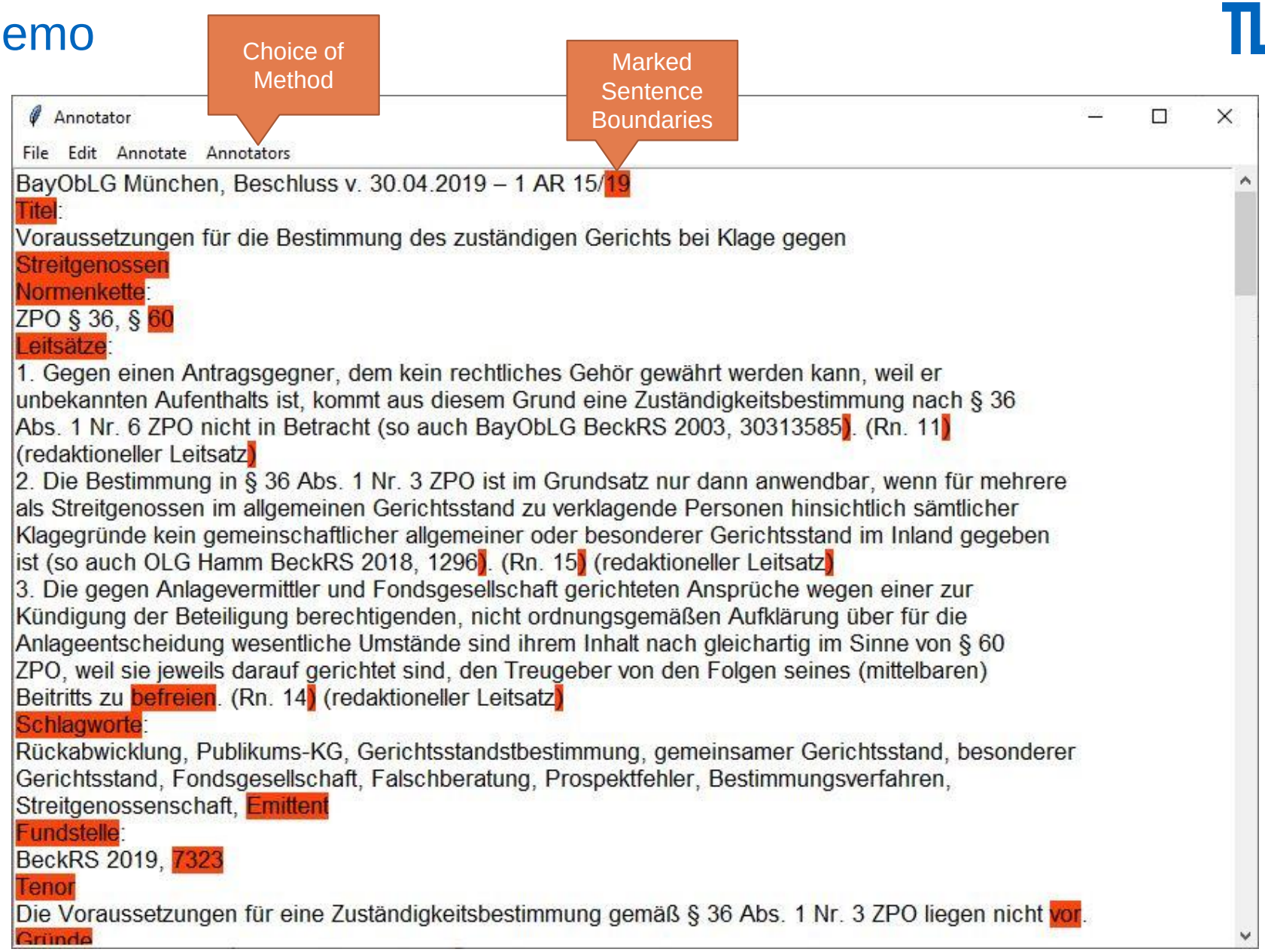


Template

- Idea: Automatic detection of structure
- Identify headlines, paragraphs,…
- Pre-processing method
- Algorithm:
  1. Find **numbered pattern**

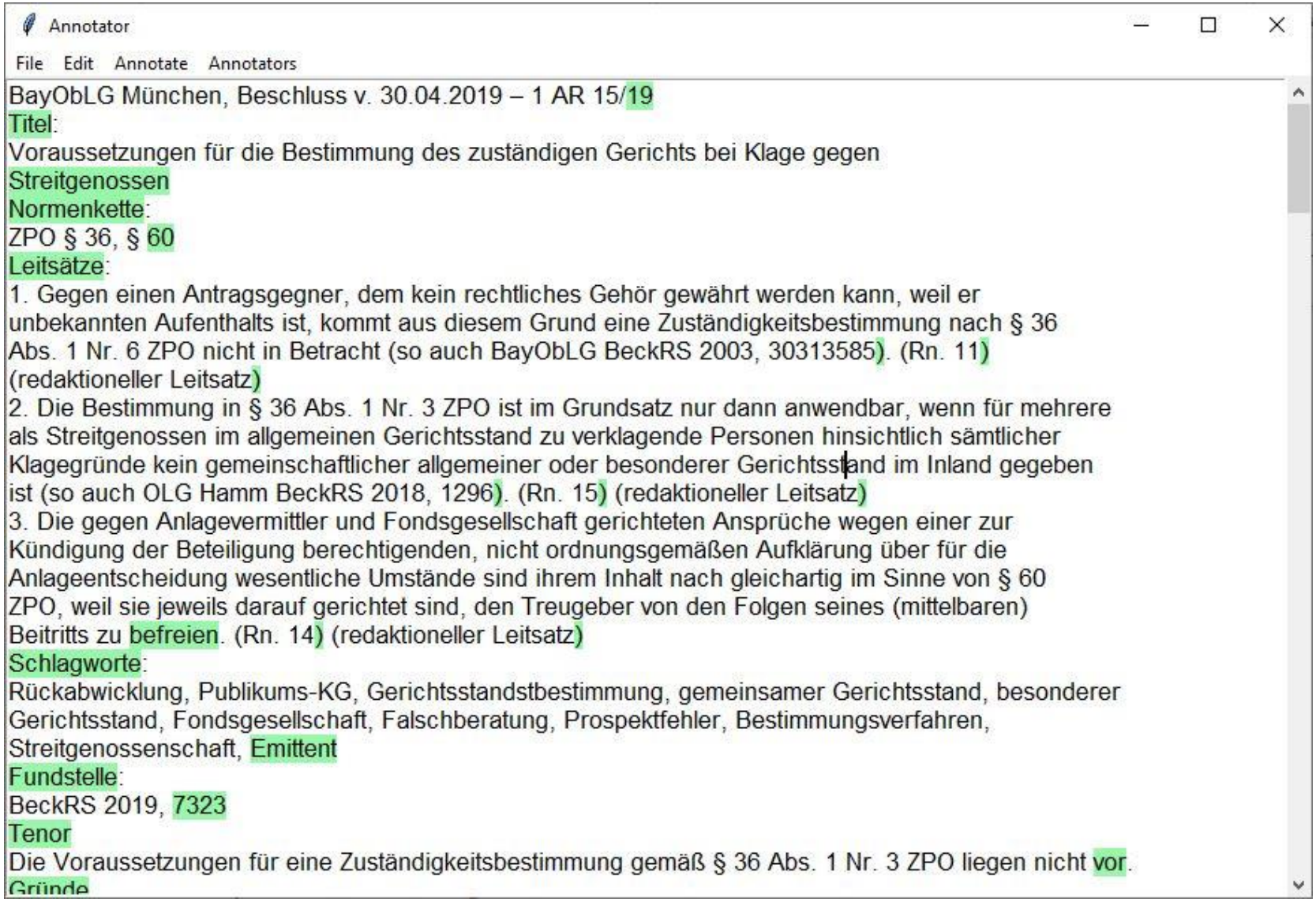
  → Construct regular expressions
  2. Segment into **coarser structure**
  3. Repeat
- Individual sentences determined by **other methods**

Demo


Choice of Method
Marked Sentence Boundaries
Annotator
File Edit Annotate Annotators
BayObLG München, Beschluss v. 30.04.2019 – 1 AR 15/19
Titel:
Voraussetzungen für die Bestimmung des zuständigen Gerichts bei Klage gegen Streitgenossen
Normenkette:
ZPO § 36, § 60
Leitsätze:
1. Gegen einen Antragsgegner, dem kein rechtliches Gehör gewährt werden kann, weil er unbekannten Aufenthalts ist, kommt aus diesem Grund eine Zuständigkeitsbestimmung nach § 36 Abs. 1 Nr. 6 ZPO nicht in Betracht (so auch BayObLG BeckRS 2003, 30313585). (Rn. 11) (redaktioneller Leitsatz)
2. Die Bestimmung in § 36 Abs. 1 Nr. 3 ZPO ist im Grundsatz nur dann anwendbar, wenn für mehrere als Streitgenossen im allgemeinen Gerichtsstand zu verklagende Personen hinsichtlich sämtlicher Klagegründe kein gemeinschaftlicher allgemeiner oder besonderer Gerichtsstand im Inland gegeben ist (so auch OLG Hamm BeckRS 2018, 1296). (Rn. 15) (redaktioneller Leitsatz)
3. Die gegen Anlagevermittler und Fondsgesellschaft gerichteten Ansprüche wegen einer zur Kündigung der Beteiligung berechtigenden, nicht ordnungsgemäßen Aufklärung über für die Anlageentscheidung wesentliche Umstände sind ihrem Inhalt nach gleichartig im Sinne von § 60 ZPO, weil sie jeweils darauf gerichtet sind, den Treugeber von den Folgen seines (mittelbaren) Beitritts zu befreien. (Rn. 14) (redaktioneller Leitsatz)
Schlagworte:
Rückabwicklung, Publikums-KG, Gerichtsstandstbestimmung, gemeinsamer Gerichtsstand, besonderer Gerichtsstand, Fondsgesellschaft, Falschberatung, Prospektfehler, Bestimmungsverfahren, Streitgenossenschaft, Emittent
Fundstelle:
BeckRS 2019, 7323
Tenor
Die Voraussetzungen für eine Zuständigkeitsbestimmung gemäß § 36 Abs. 1 Nr. 3 ZPO liegen nicht vor.
Gründe

Demo: Rule-based

Demo: NN


Annotator
File Edit Annotate Annotators
BayObLG München, Beschluss v. 30.04.2019 – 1 AR 15/19
Titel:
Voraussetzungen für die Bestimmung des zuständigen Gerichts bei Klage gegen Streitgenossen
Normenkette:
ZPO § 36, § 60
Leitsätze:
1. Gegen einen Antragsgegner, dem kein rechtliches Gehör gewährt werden kann, weil er unbekannten Aufenthalts ist, kommt aus diesem Grund eine Zuständigkeitsbestimmung nach § 36 Abs. 1 Nr. 6 ZPO nicht in Betracht (so auch BayObLG BeckRS 2003, 30313585). (Rn. 11) (redaktioneller Leitsatz)
2. Die Bestimmung in § 36 Abs. 1 Nr. 3 ZPO ist im Grundsatz nur dann anwendbar, wenn für mehrere als Streitgenossen im allgemeinen Gerichtsstand zu verklagende Personen hinsichtlich sämtlicher Klagegründe kein gemeinschaftlicher allgemeiner oder besonderer Gerichtsstand im Inland gegeben ist (so auch OLG Hamm BeckRS 2018, 1296). (Rn. 15) (redaktioneller Leitsatz)
3. Die gegen Anlagevermittler und Fondsgesellschaft gerichteten Ansprüche wegen einer zur Kündigung der Beteiligung berechtigenden, nicht ordnungsgemäßen Aufklärung über für die Anlageentscheidung wesentliche Umstände sind ihrem Inhalt nach gleichartig im Sinne von § 60 ZPO, weil sie jeweils darauf gerichtet sind, den Treugeber von den Folgen seines (mittelbaren) Beitritts zu befreien. (Rn. 14) (redaktioneller Leitsatz)
Schlagworte:
Rückabwicklung, Publikums-KG, Gerichtsstandstbestimmung, gemeinsamer Gerichtsstand, besonderer Gerichtsstand, Fondsgesellschaft, Falschberatung, Prospektfehler, Bestimmungsverfahren, Streitgenossenschaft, Emittent
Fundstelle:
BeckRS 2019, 7323
Tenor
Die Voraussetzungen für eine Zuständigkeitsbestimmung gemäß § 36 Abs. 1 Nr. 3 ZPO liegen nicht vor.
Gründe

Class Overview