

# Vertragsanalyse mittels Regel- und Machine Learning basierter Technologien

Ingo Glaser, 21.09.2017, Sebis Day

Lehrstuhl für Software Engineering betrieblicher Informationssysteme (sebis)  
Fakultät für Informatik  
Technische Universität München  
[www.matthes.in.tum.de](http://www.matthes.in.tum.de)

## Motivation

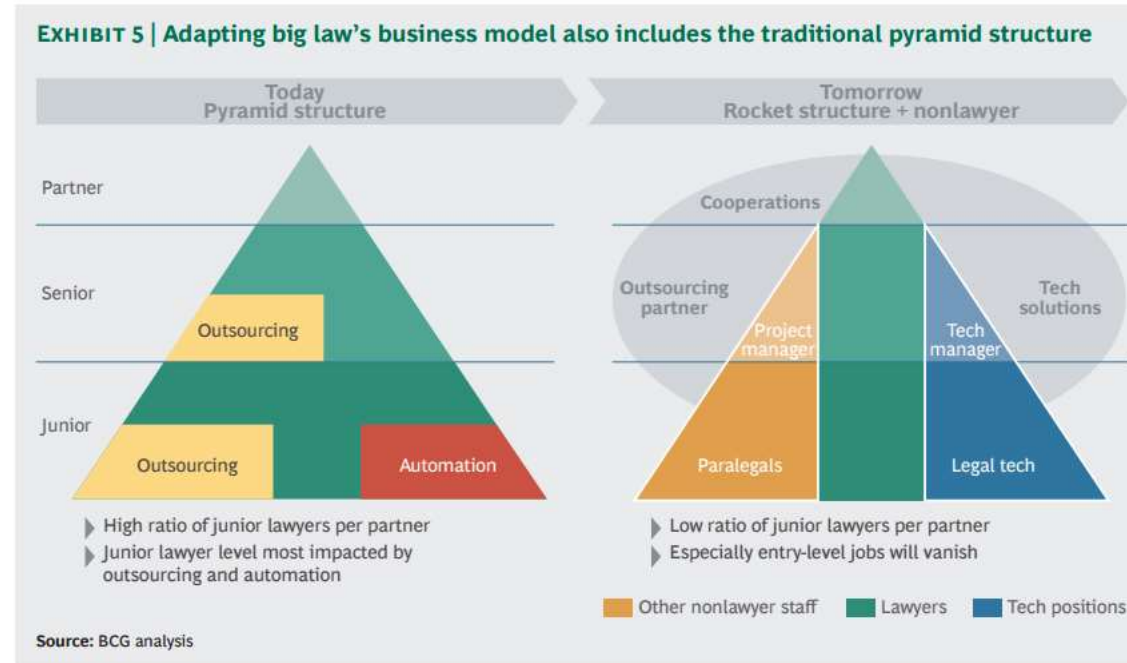
## Textgrundlage

## Themengebiete

- Klauselmatching
- Satzklassifikation
- Named Entity Recognition und Linking

## Fazit & Ausblick

### BCG Report: January 2016



*Powerful forces are reshaping the corporate legal-advice market, including price pressure from clients and **the digitization of legal data**. Legal-technology solutions can help law firms succeed in this new era, **but only if they understand the tech landscape** and redefine key elements of their business model.*

*- BCG Analysis, 2016*

*Derzeit haben wir folgende offene Stellen in FESTANSTELLUNG*

- ➔ **Anwendungsentwickler C#/.NET (m/w)**
- ➔ **Developer/Senior Developer/Architect für .Net & SharePoint (m/w)**
- ➔ **Consultant Microsoft Enterprise Systeme (m/w)**
- ➔ **Legal Data Scientist (m/w)**
- ➔ **Personalreferent (m/w)**

Quelle: Sinc GmbH  
Stand: 29. Mai, 2016

## **Tätigkeiten (Arbeitsabläufe, Prozesse bzw. Workflows) von Rechtsexperten (Wissenschaftler und Praktiker) sind ...**

- ... Zeit-intensiv
- ... Wissens-intensiv
- ... Daten-intensiv

## **Datenanalyse wird immer attraktiver, weil ...**

- ... Rechenzeit und Speicher immer billiger wird und
- ... Algorithmen immer präziser werden und
- ... Algorithmen immer schneller werden.

## **Größtmögliche Präzision im Rechtsbereich nur durch Anpassung an Domäne zu erreichen**

- Deutschsprachige Texte des Rechtwesens und der Justiz
- Datenmodell adaptieren
- Training von Algorithmen

Motivation

Textgrundlage

Themengebiete

- Klauselmatching
- Satzklassifikation
- Named Entity Recognition und Linking

Fazit & Ausblick

## Verträge

- Mietverträge
  - FuE – Verträge
  - Arbeitsverträge
  - Kaufverträge
  - Gesellschaftsverträge
  - ...
- 

## Alternativ

- Urteile (zB BGH, BVerfGH, etc.)
  - (Bundes-) Gesetze
    - Deutschland, Österreich und Schweiz
  - Urteile zu best. Domänen (zB Mietrecht)
- 

## Wahl

- Mietrecht
- Verträge
- Urteile bezüglich des Mietrechts

Motivation

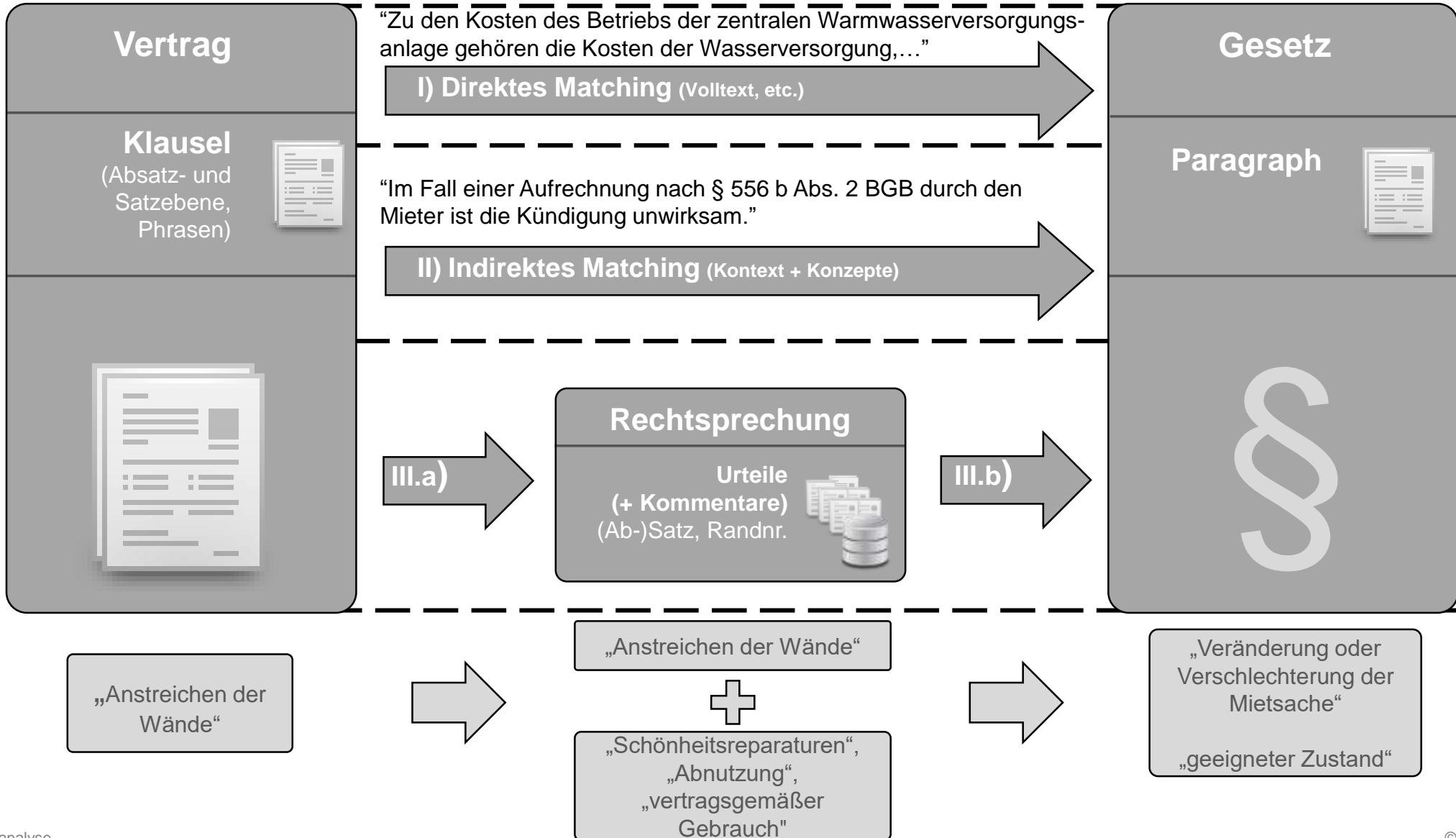
Textgrundlage

Themengebiete

- **Klauselmatching**
- Satzklassifikation
- Named Entity Recognition und Linking

Fazit & Ausblick





### Beispiel

- Mieterhöhung → Erhöhung der Mietsache
- Mietgebrauch → Gebrauch der Mietsache
  - Vertrag → BGH Urteil (VIII ZR 295/15)

### Ansätze

- MLT (TF-IDF)
- LSH

Motivation

Textgrundlage

Themengebiete

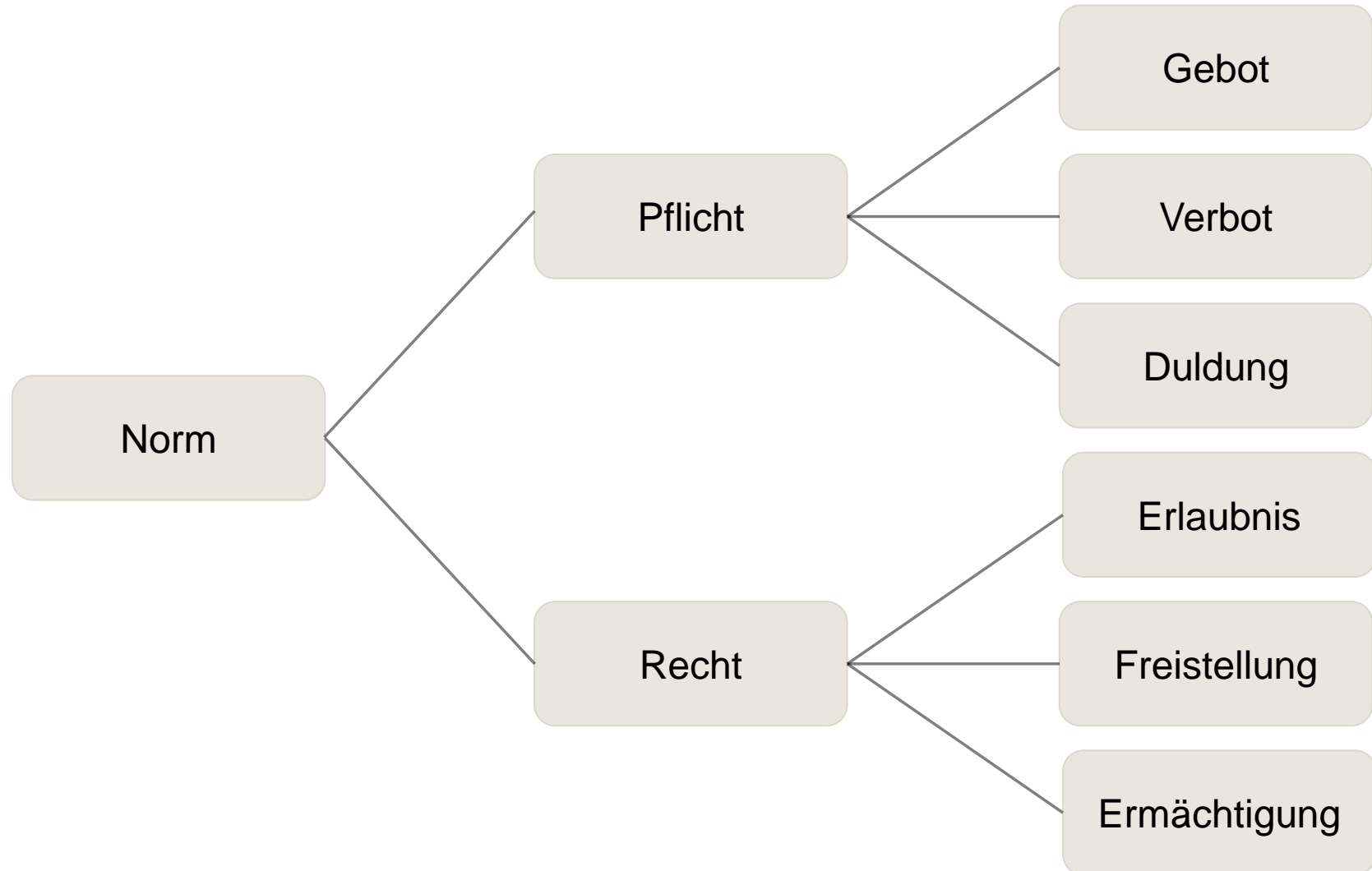
- Klauselmatching
- **Satzklassifikation**
- Named Entity Recognition und Linking

Fazit & Ausblick

## Taxonomie von Vertragsklauseln

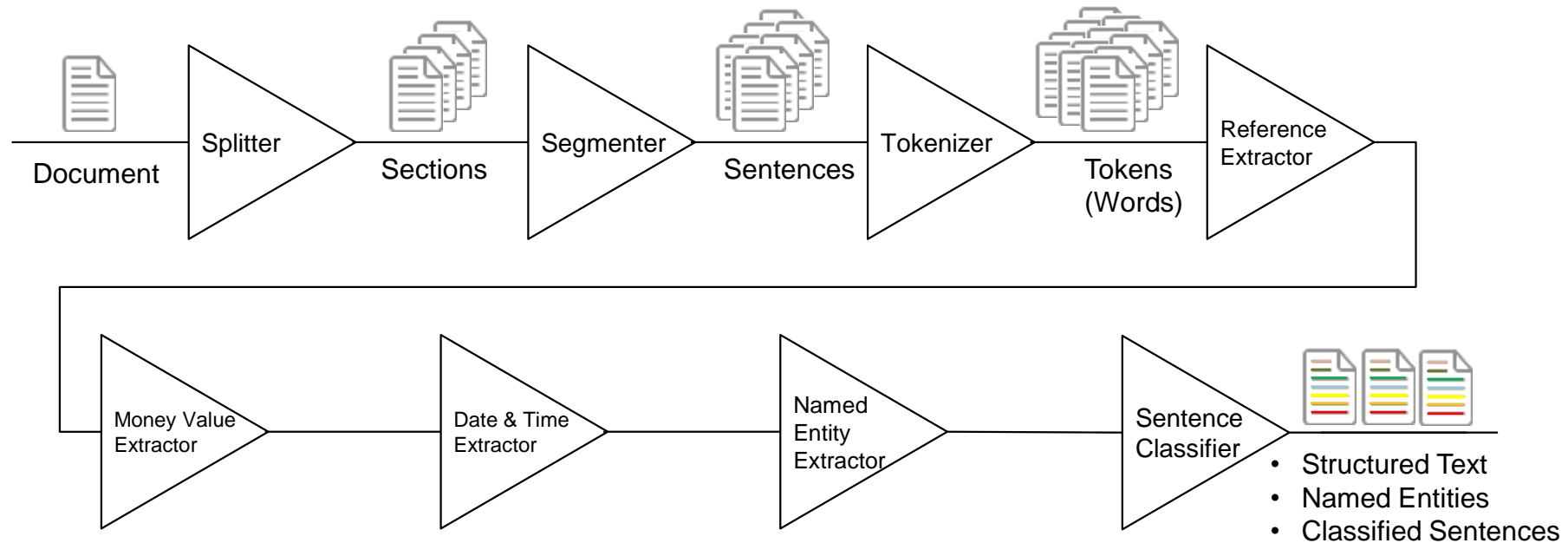
- Vertrags- und Domänenabhängig (!)
- Deontische Konzepte (Norm-Funktion)
  - Verbote, Gebote / Pflichten, Sollensanordnungen, Appelle, Rechte, Anspruchsnormen, etc.
- Inhaltliche Konzepte (Norm-Inhalt)
  - Haftung, Urheberrechtsbestimmungen, Zahlungsmodalität, Kündigungsklausel, Konkurrenzverbote, etc.
  - Überschriften und Nicht-disjunkt
- Meta-regeln
  - Gerichtsstand, Salvatorische Klausel, Verweisungen, etc.

## Analogie: Taxonomie von Normen bei Gesetzen



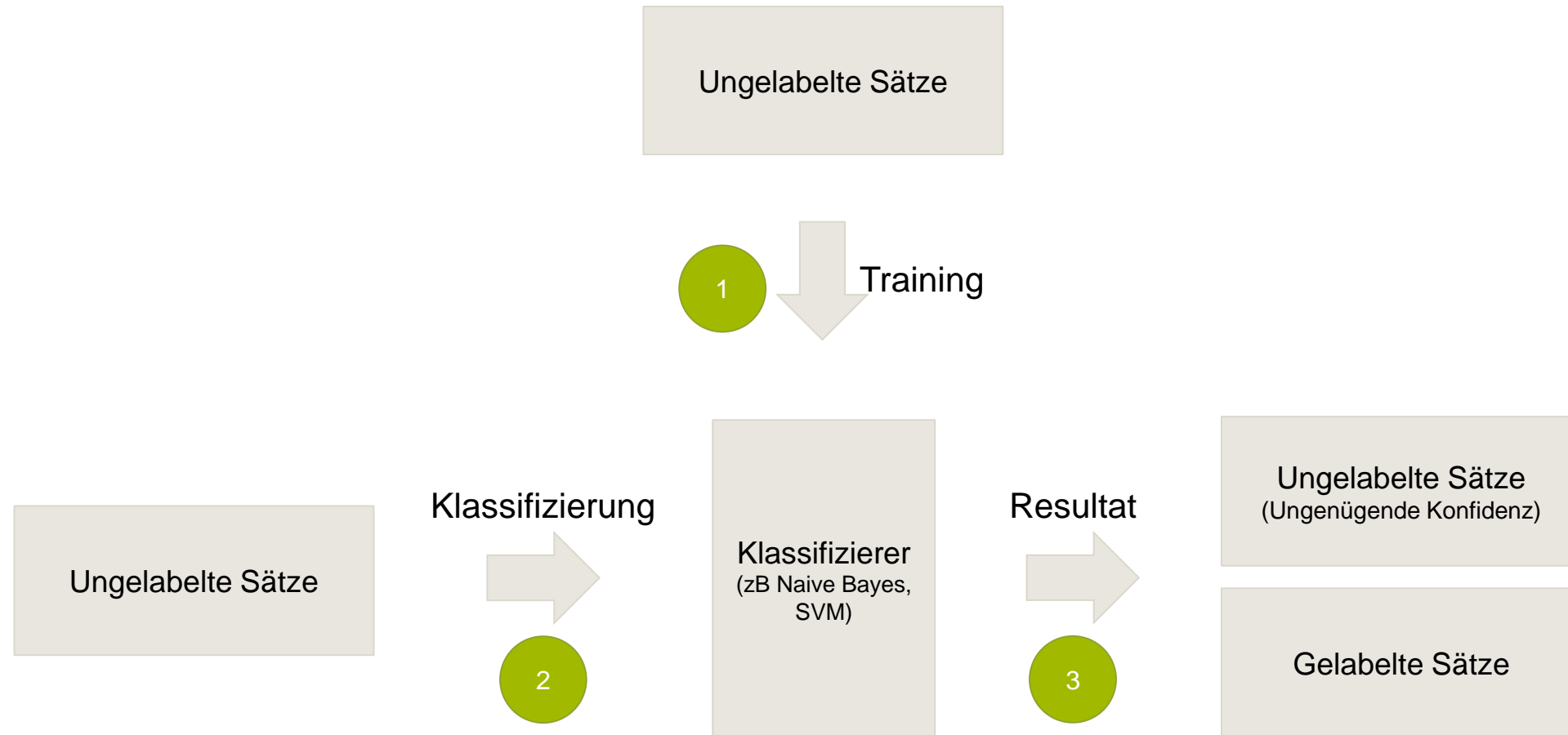
### Apache UIMA (Unstructured Information Management Architecture)

- Pipes & Filter Architektur
- Thread-sicher
- Apache RUTA um komplexe Pattern zu definieren



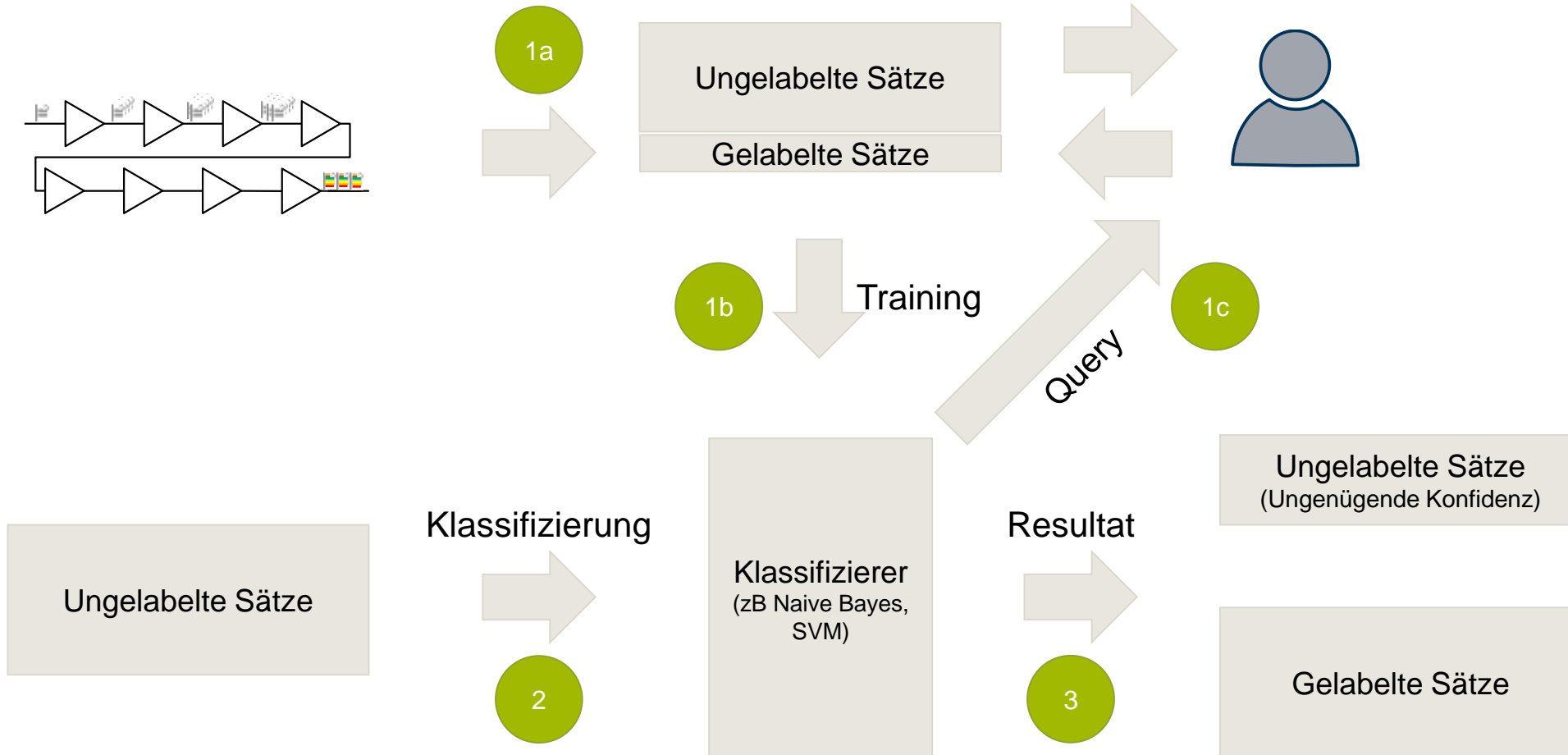
### LexML

- Machine Learning as a Service (MLaaS)
- Apache Spark



### Lexia + LexML

- Machine Learning as a Service (MLaaS)
- Apache Spark + Apache UIMA





### Größenordnung

- 378 Sätze zum Training (75%)
- 126 Sätze zum Testen (25%)
- BGB §535 - §595 (Mietrecht)

### Evaluation

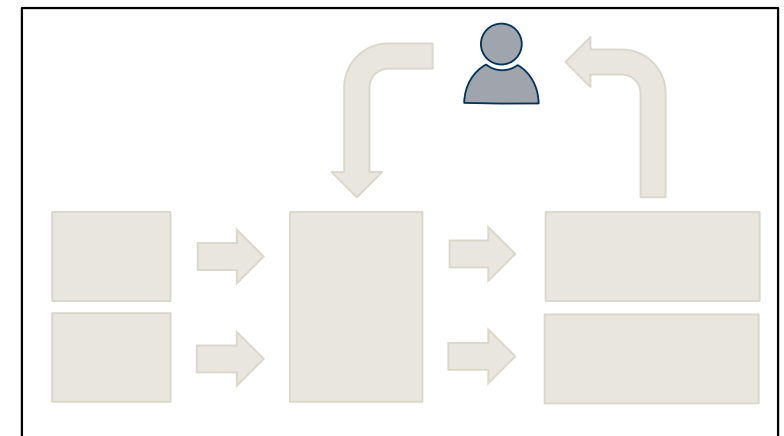
- Precision/Recall/Accuracy/F1
- Precision, Recall und F1 ca. 5-10% höher als SML
- Accuracy ca. 7% höher als SML

### Parameterstudien

- Wieviel bringt "Active" vgl. supervised ML?

### Lernverhalten

- ROC (Receiver Operator Characteristics)



Motivation

Textgrundlage

Themengebiete

- Klauselmatching
- Satzklassifikation
- **Named Entity Recognition und Linking**

Fazit & Ausblick

Geldwerte  
Datumswerte

Regelbasierte Verfahren  
(zB Apache Ruta, RegEx)

Verweisungen

Regelbasierte Verfahren  
(zB Apache Ruta, RegEx)

Named Entities:

- Personen
- Unternehmen
- Ort

Statistische (Machine Learning) Verfahren  
(zB GermaNer, Stanford NER)

Knowledge Bases  
(zB DBPedia, OpenCalais)

Template-basierte NER

### Geldwerte

- Absolut: 1.234 Euro; 1.234 Euro;
- Relativ: „50 % der Miete“;

### Datumswerte

- Absolut
  - „15. September bis 15. Mai“
- Relativ
  - „12 Monate nach Ende des Abrechnungszeitraums“
  - „4 Wochen nach XX“
  - „3 Monate vor Beginn der Bauarbeiten“

#### § 5 Versorgung mit Heizung und Warmwasser

1. Der Vermieter muss die Sammelheizung, soweit es die Witterung erfordert, mindestens aber in der Zeit vom 15. September bis 15. Mai in Betrieb halten. Eine Temperatur von mindestens 20°C bis 22°C zwischen 6.00 und 24.00 Uhr in den beheizbaren Räumen ist zu erreichen. In der übrigen Nachtzeit sind 18°C ausreichend.

6. Der Vermieter kann eine Nachzahlung auf die Heiz- und Betriebskosten nur verlangen, sofern er spätestens 12 Monate nach Ende des Abrechnungszeitraumes dem Mieter durch schriftliche Abrechnung nachweist, dass die Vorauszahlungen auf die Betriebskosten nicht ausgereicht haben. Ergibt sich ein Guthaben aus der Abrechnung für den Mieter, wird dies unverzüglich ausgezahlt. Eine Aufrechnung mit bestrittenen oder nicht rechtskräftig festgestellten Forderungen darf der Vermieter nicht vornehmen. Einwendungen des Mieters gegen die Abrechnung müssen dem Vermieter spätestens 12 Monate nach Zugang der Abrechnung mitgeteilt werden.
7. Nachforderungen des Vermieters werden 4 Wochen nach Zugang der ordnungsgemäßen Abrechnung fällig. Der Vermieter gewährt dem Mieter Einsicht in die Berechnungsunterlagen. Gegen Erstattung angemessener Kopier- und Portokosten kann der Mieter verlangen, dass ihm Kopien der Berechnungsunterlagen zugesandt werden.

### Referenzen

- „Teilkündigung und Verwertungskündigung §§ 573, 573a, 573b BGB“

### Personen

- „Herrn Martin Rollinger“
- „Prof. Dr. Florian Matthes“

### Unternehmen / Institutionen

- „Technische Universität München“
- „SINC GmbH“

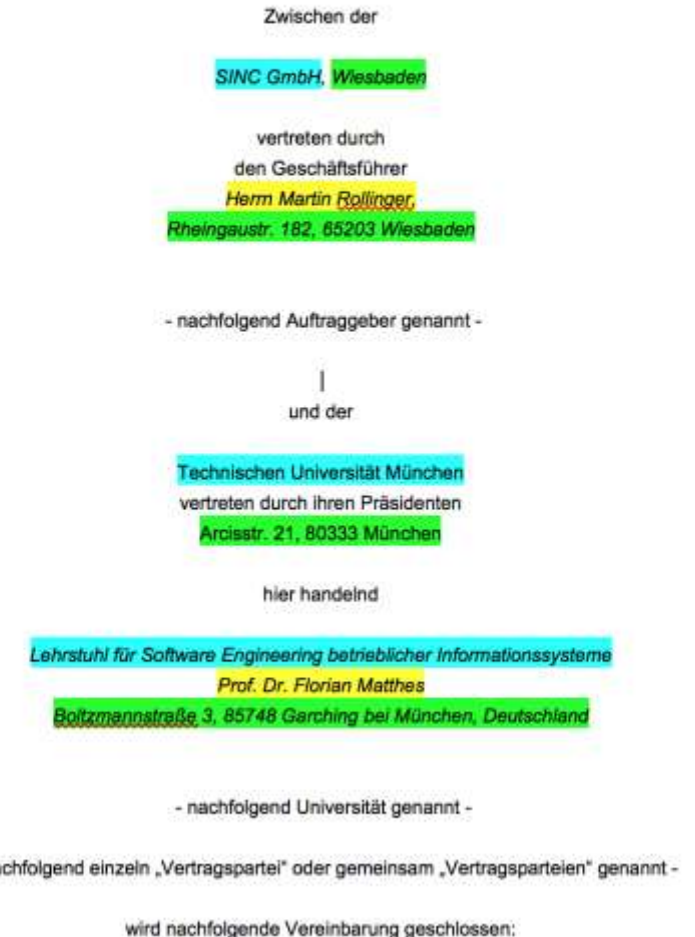
### Orte / Geographische Information

- „Rheingastr. 182, 65203 Wiesbaden“
- „Boltzmannstraße 3  
85748 Garching bei München, Deutschland“

### Rollen

- “Auftraggeber“ / „Auftragnehmer“

### Forschungs- und Entwicklungsvertrag



# Named Entity Recognition

Ein template-basierter Ansatz

**Mietvertrag für Wohnräume**

Zwischen XXXXX XXXXXXXX Vermieter

XXXXXXXXXX XX XXXXX XXXXXXXXX Telefon XXX XXXXXXXXX  
(Anschrift)

XXXXX@XXXX.XX Mobil XXXX XXXXXXXXXX  
(E-Mail-Adresse)

XXXXXXXXXXXXXXXXXXXX  
(Bevollmächtigter)

XXXXXXXX XX XXXXX XXXXXXXXX Telefon XXX XXXXXXXXX  
(Anschrift)

XXXXXXXXXXXXXXXXXXXX Mobil  
(E-Mail-Adresse)

und XXXXXXXXXXXXXXXXXXXXXXX Mieter  
(Vor- und Zuname, Geburtsdatum, Nummer des Personalausweises/Personalausweis)

XXXXXXXX XX XXXXX XXXXXXXXX Telefon XXX XXXXXXXXX  
(Anschrift)

max@mieter.de Mobil  
(E-Mail-Adresse)

XXXXXXXXXXXXXXXXXXXX  
(Kontostellung des Mieters, Name und Ort der Bank, Name des Kontoinhabers)

BLZ Kfs.-Nr.  
IBAN DEXXXXXXXXXXXXXXXXXXXX BIC XXXXXXXXXXXXXXX



**Mietvertrag für Wohnraum TEMPLATE**

**Präambel**

**§ 1 Allgemeines**

Zwischen --Vermieter--

--Vermieter.Anschrift-- Telefon --Vermieter.Telefonnummer--

--Vermieter.EMailAdresse-- Mobil --Vermieter.Mobilnummer--

--Bevollmächtigter--

--Bevollmächtigter.Anschrift-- Telefon --Bevollmächtigter.Telefonnummer--

--Bevollmächtigter.EMailAdresse-- Telefon --Bevollmächtigter.Mobilnummer--

und --Mieter--

--Mieter.Anschrift--

Telefon --Mieter.Telefonnummer--

--Mieter.EMailAdresse-- Mobil --Mieter.Mobilnummer--

--Mieter.Bankverbindung--



**Mietvertrag für Wohnraum**

**§ 1 Allgemeines**

Zwischen --Vermieter-- Vermieter.Anschrift-- Telefon --Vermieter.Telefonnummer--  
Vermieter.EMailAdresse-- Mobil --Vermieter.Mobilnummer-- Bevollmächtigter--  
Bevollmächtigter.Anschrift-- Telefon --Bevollmächtigter.Telefonnummer--  
Bevollmächtigter.EMailAdresse-- Telefon --Bevollmächtigter.Mobilnummer-- und  
Mieter-- Mieter.Anschrift-- Telefon --Mieter.Telefonnummer-- Mieter.EMailAdresse--  
Mobil-- Mieter.Mobilnummer-- Mieter.Bankverbindung-- Elena Seppachova

Boltzmannstraße 3 Telefon 089 289 17100

elena.seppachova@tum.de Mobil 0176 12345678

nicht vorhanden

nicht vorhanden Telefon nicht vorhanden

nicht vorhanden Telefon nicht vorhanden

und Bernhard Walt

Boltzmannstraße 3, 85748 Garching bei München

Telefon 089 289 17128

hawaii@mieter.de Mobil 0176 12345678

Musterfirma GmbH, Geschäftsführer: Deutsche Postbank AG, IBAN: 2512 0510 0430 0000 0000 00, BIC: 251205

1. Überführung Mustervertrag in Template

2. Analyse über „Platzhalter“

- --Vermieter--
- --Vermieter.Anschrift--
- --Mieter--
- --Mieter.Mobilnummer--

3. Auflösen der Platzhalter

# Named Entity Recognition

Ein template-basierter Ansatz: Beispiel

## Geschäftsraum-Mietvertrages

**§ 1 Parteien**

Zwischen  
BERNHARD WALTL  
und  
SINC GmbH  
wird folgender Geschäftsraummietvertrag geschlossen:

**§ 2 Mieträume**

1. Vermietet werden im Haus Rheingaustraße 182, 65203 Wiesbaden folgende Räume:

- a. Erdgeschoss: 1
- b. 1. Etage: 3
- c. Keller: 1
- d. Dachboden: 1

Die Mietfläche beträgt 75 qm.

2. Für die oben genannten Räume erhält der Mieter folgende Schlüssel:

2 x Haustüre, 1 x Keller, 1 x Dachboden

3. Schäden an diesen Räumen sind dem Vermieter unverzüglich anzuzeigen.

4. Der Mieter ist verpflichtet, eine Glasversicherung für sämtliche Fenster-, Schaufenster- und Türscheiben der Mieträume (oder sonstige Versicherungen nach Vereinbarung, wobei eine Doppelversicherung durch Mieter und Vermieter vermieden werden sollte) in entsprechender Höhe auf seinen Kosten abzuschließen und den

## Geschäftsraum-Mietvertrages TEMPLATE

**§ 1 Parteien**

Zwischen  
~~BERNHARD WALTL und SINC GmbH~~ **--Vermieter--**  
**und**  
**--Mieter--**  
wird folgender Geschäftsraummietvertrag geschlossen:

**§ 2 Mieträume**

1. Vermietet werden im Haus ~~Rheingaustraße 182, 65203 Wiesbaden~~ folgende Räume: **--Adresse--**

~~Erdgeschoss:1 1. Etage:3 Keller:1 Dachboden:1~~ **folgende Räume:**

- a. ~~Erdgeschoss:~~ **--Mietobjekt.Erdgeschoss--**
- b. ~~1. Etage:~~ **--Mietobjekt.1Etage--**
- c. ~~Keller:~~ **--Mietobjekt.Keller--**
- d. ~~Dachboden:~~ **--Mietobjekt.Dachboden--**

Die Mietfläche beträgt ~~75~~ **--Mietobjekt.Größe--** qm.

2. Für die oben genannten Räume erhält der Mieter folgende Schlüssel:

~~2 x Haustüre, 1 x Keller, 1 x Dachboden~~ **--Mietobjekt.Schlüssel--**

3. Schäden an diesen Räumen sind dem Vermieter unverzüglich anzuzeigen.

4. Der Mieter ist verpflichtet, eine Glasversicherung für sämtliche Fenster-, Schaufenster- und Türscheiben der Mieträume (oder sonstige Versicherungen nach



# Named Entity Recognition

Ein template-basierter Ansatz: Negatives Beispiel

## Mietvertrag für Wohnraum TEMPLATE

**§ 1 Allgemeines**

Zwischen --Vermieter--  
--Vermieter.Anschrift-- Telefon --Vermieter.Telefonnummer--  
--Vermieter.EMailAdresse-- Mobil --Vermieter.Mobilnummer--  
--Bevollmächtigter--  
--Bevollmächtigter.Anschrift-- Telefon --Bevollmächtigter.Telefonnummer--  
--Bevollmächtigter.EMailAdresse-- Telefon --Bevollmächtigter.Mobilnummer--  
und --Mieter--  
--Mieter.Anschrift--  
Telefon --Mieter.Telefonnummer--  
--Mieter.EMailAdresse-- Mobil --Mieter.Mobilnummer--  
--Mieter.Bankverbindung--

## Mietvertrag für Wohnraum

**§ 1 Allgemeines**

Zwischen ~~Vermieter~~ ~~Vermieter.Anschrift~~ ~~Telefon~~ ~~Vermieter.Telefonnummer~~ ~~Vermieter.EMailAdresse~~ ~~Mobil~~ ~~Vermieter.Mobilnummer~~ ~~Bevollmächtigter~~ ~~Bevollmächtigter.Anschrift~~ ~~Telefon~~ ~~Bevollmächtigter.Telefonnummer~~ ~~Bevollmächtigter.EMailAdresse~~ ~~Telefon~~ ~~Bevollmächtigter.Mobilnummer~~ und ~~Mieter~~ ~~Mieter.Anschrift~~ ~~Telefon~~ ~~Mieter.Telefonnummer~~ ~~Mieter.EMailAdresse~~ ~~Mobil~~ ~~Mieter.Mobilnummer~~ ~~Mieter.Bankverbindung~~ Elena Scepankova

Boltzmannstraße 3 Telefon 089 289 17100

elena.scepankova@tum.de Mobil 0176 12345678

nicht vorhanden

nicht vorhanden Telefon nicht vorhanden

nicht vorhanden Telefon nicht vorhanden

und Bernhard Waltl

Boltzmannstraße 3, 85748 Garching bei München

Telefon 089 289 17124

b.waltl@tum.de Mobil 0176 12345678

Mustermann GmbH, Kreditinstitut: Deutsche Postbank AG, IBAN: DE18 3601 0043 9999 9999 99, BIC: PBNKDEFF

## Problem

- Auflösen der Platzhalter bzw. Greedy-ness der Vergleichsoperation
- zB bei zwei Platzhalter unmittelbar hintereinander

## Nächste Schritte

- Aktuelle Implementierung basierend auf Myers' Diff Algorithmus
- Integration in eine Pipeline



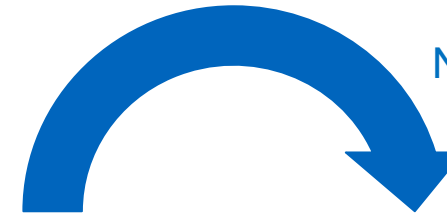
### Auflösen der Rolle einer Named Entity

- zB Arbeitsvertrag
  - Technische Universität München  Institution  Arbeitgeber
  - Bernhard Waltl  Person  Arbeitnehmer

NE Recognition



NE Linking



Text	Named Entity	Role
Technische Universität München	Institution	Arbeitgeber
Bernhard Waltl	Person	Arbeitnehmer
3 Monate	Zeitdauer	Kündigungsfrist
30. April 2018	Zeitangabe	Vertragsende
...	...	...

### Domänenmodell ist erforderlich

- Typen mit Attribute (evtl. Relationen)
- Kontext (Satz, Klausel, Document)

### 1. Zuweisung von NE durch Regeln

- Apache Ruta
- RegEx
- “Bootstrap“ von ML Verfahren (→ labelled data set)

### 2. Zuweisung von NE durch Heuristische Verfahren

- Active Machine Learning
- Apache Spark
  - Naive Bayes

Arbeitsvertrag z.B. im SocioCortex	
Arbeitgeber	<i>Institution</i>
Arbeitnehmer	<i>Person</i>
Kündigungsfrist	<i>Zeitdauer</i>
Vertragsende	<i>Zeitangabe</i>
Konkurrenzverbot	<i>Boolean</i>
...	...

Motivation

Textgrundlage

Themengebiete

- Klauselmatching
- Satzklassifikation
- Named Entity Recognition und Linking

Fazit & Ausblick

## Conclusion

- Klauselmatching mit **LSH** performt mit **niedrigem Recall**, aber **hoher Precision**
- Klauselmatching mit **TF-IDF** performt mit **hohem Recall**, aber **niedriger Precision**
- **AML ist SML überlegen** im Bereich der Satz- und Normklassifikation in rechtlichen Dokumenten
- **NER** ist im **Rechtsbereich** bereits **sehr gut anwendbar**
- Es existiert ein **Mangel an GOLD-Standards und Ontologien** im Bereich der Rechtstexte um NEL durchzuführen

## Outlook

- Definition von **angemessenen Anwendungsfällen**
- Wie kann der **Vertragserstellungsprozess verbessert** werden?
  - Verwendung von Klauselmatching
  - Recommender-Systeme
  - Vermeidung von unwirksamen Klauseln
  - Konsistente Formulierungen
- Linking von NEs zu **semantischen Modellen**



B.Sc.

**Ingo Glaser**

Studentische Hilfskraft

Technische Universität München  
Fakultät für Informatik  
Lehrstuhl für Software Engineering  
betrieblicher Informationssysteme

Boltzmannstraße 3  
85748 Garching bei München

Tel +49.89.289.17138

Fax +49.89.289.17136

ingo.glaser@tum.de  
[www.matthes.in.tum.de](http://www.matthes.in.tum.de)