

# Computergestützte Textanalyse zur Unterstützung juristischer Arbeitsabläufe

Bernhard Wautl, 05.07.2017

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

- Introduction & Background
  
- Semantic Analysis of Legal Documents
  - a. Supporting **editorial** processes
  - b. Supporting **drafting** processes
  - c. Supporting **search** processes
  
- Technology for Semantic Analysis
  - a. Pipes & Filters – Architecture
  - b. Machine Learning Approach

- Research associate at the Technical University of Munich
  - Since 2014 at the chair „Software Engineering for Business Information Systems“
  - Software Engineering
    - Collaborative Information Systems
    - Databases
  - Legal Informatics
    - Legal Reasoning
    - Data and Text Analytics
- Lexalyze: Interdisciplinary Research Program
  - Technical University of Munich: Prof. Florian Matthes, et al.
  - Ludwig-Maximilians University: Prof. Hans-Christoph Grigoleit, et al.
  - [www.lexalyze.de](http://www.lexalyze.de)

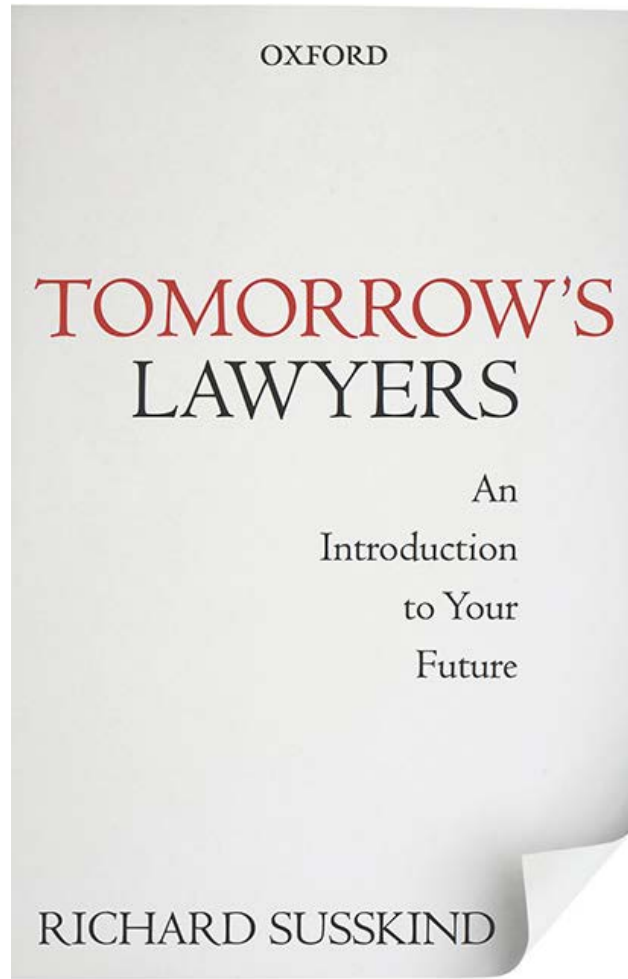
## Motivation

- Digitization throughout all domains of society
- Legal sciences and practices still lack of digitization

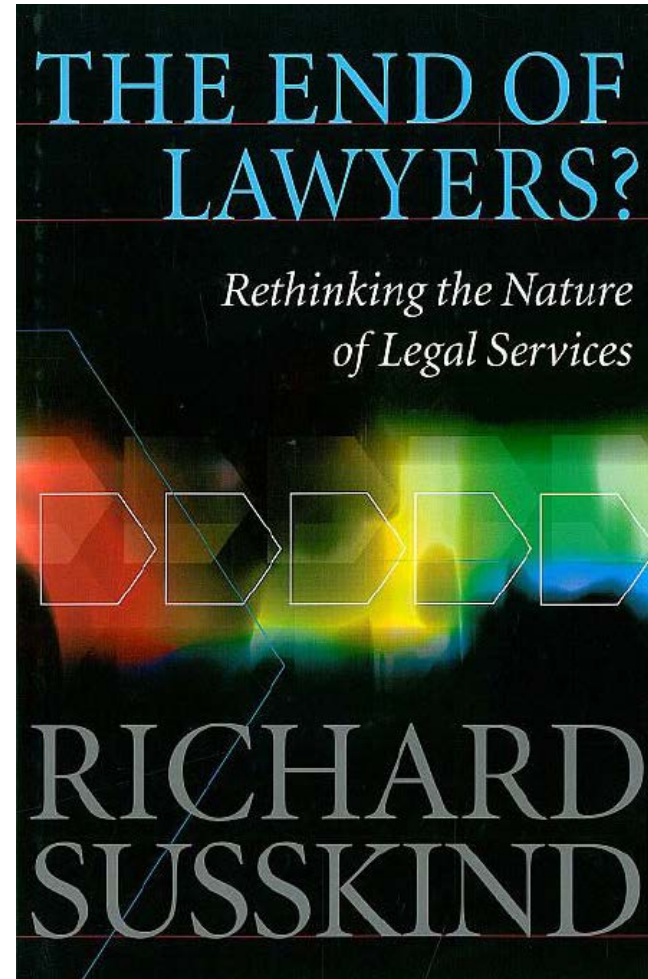
→ **Unexploited synergies**

## Approach

- **Usage of Informatics** for judicial problem solving
- Synergies through the development of **a common language**
- Development of **interdisciplinary competencies**

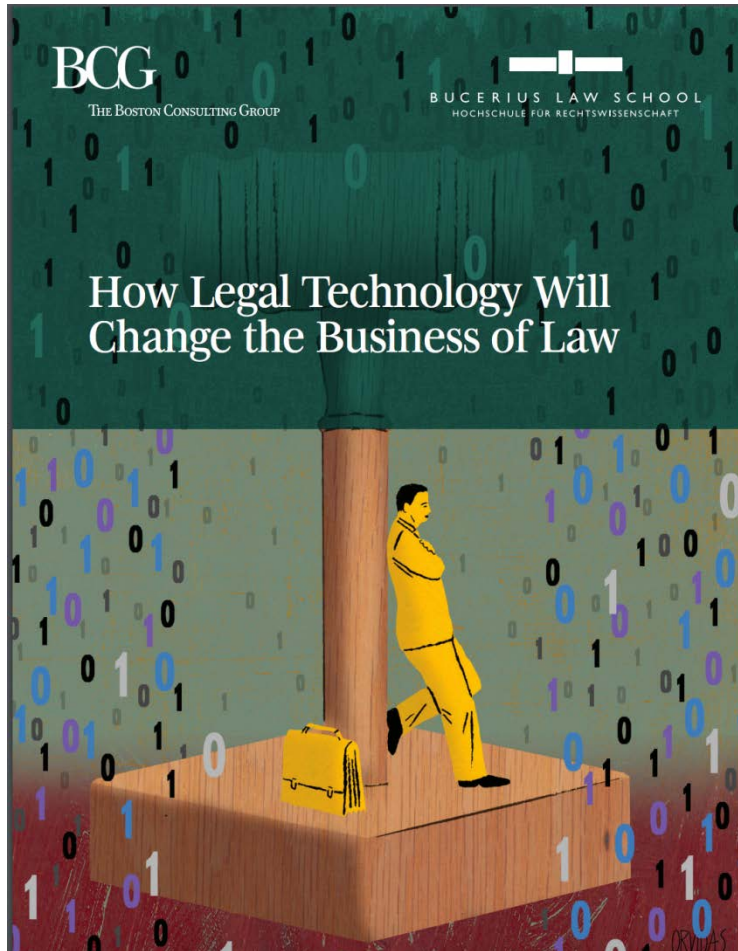


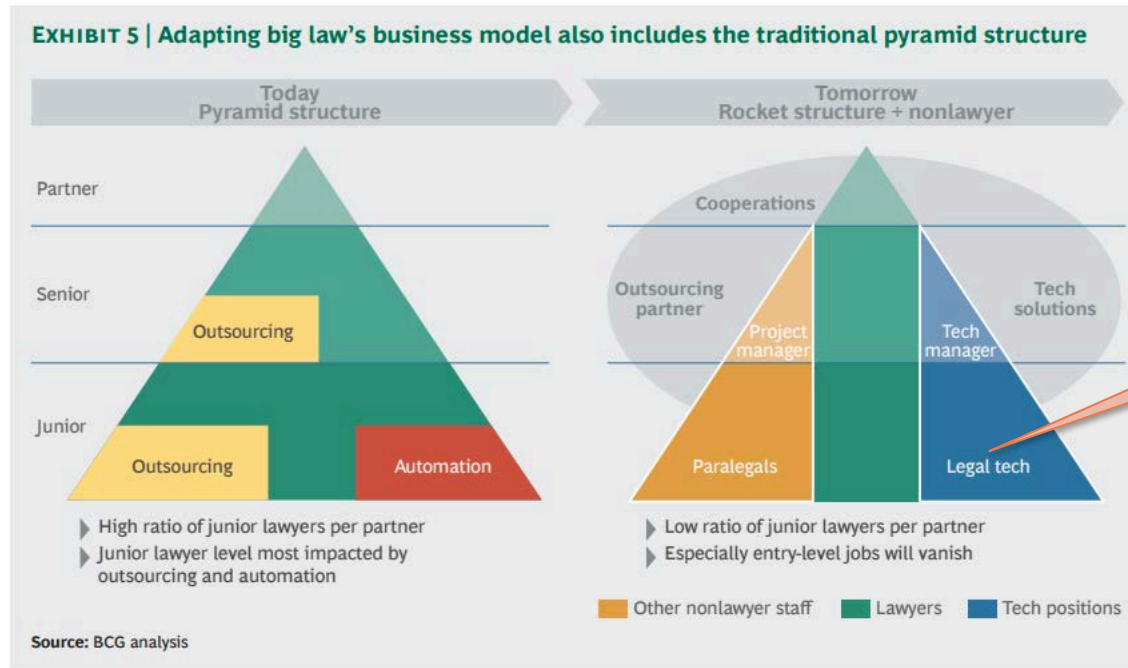
2013



2010







What will the role of LegalTech be?

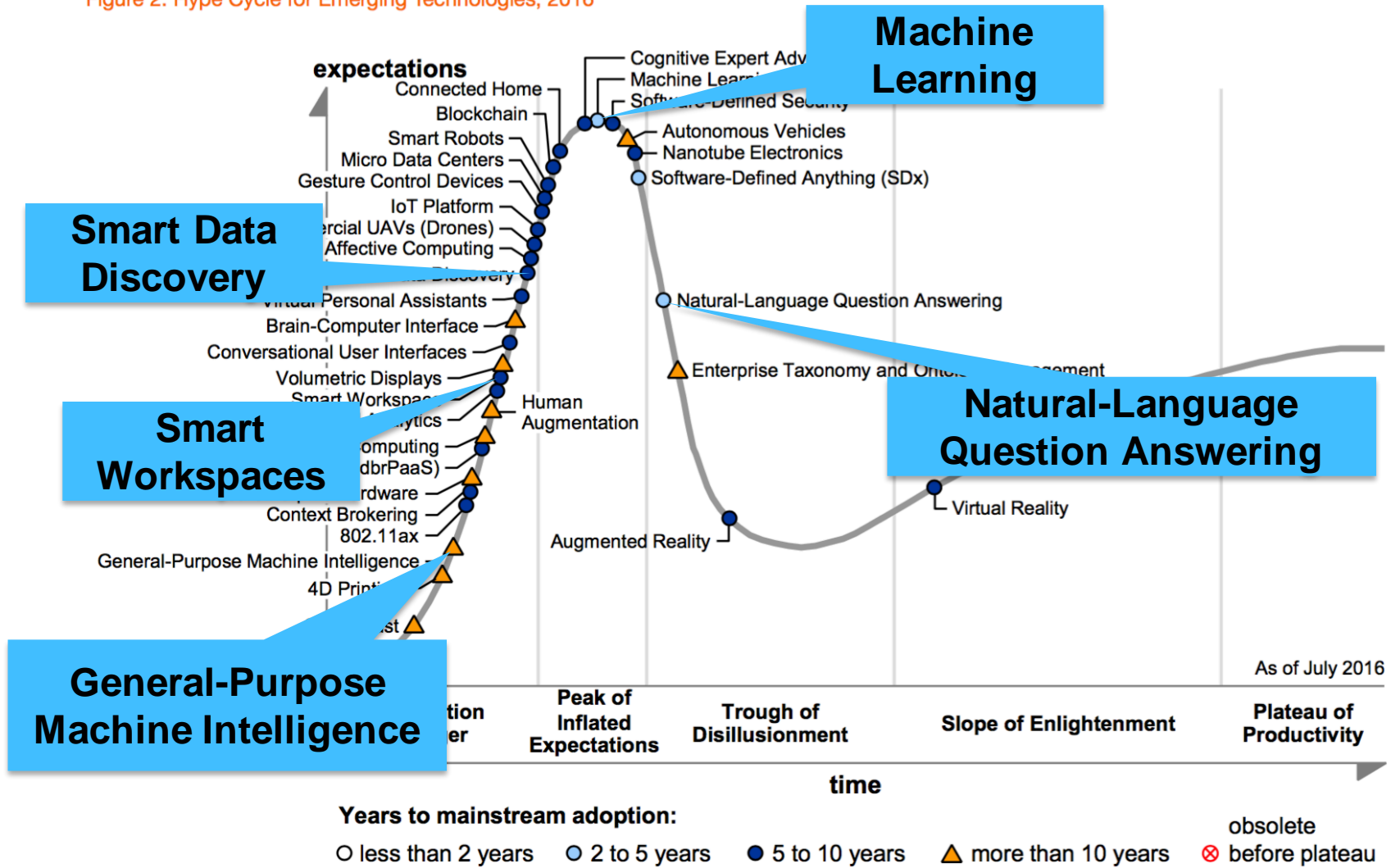
*Powerful forces are reshaping the corporate legal-advice market, including price pressure from clients and **the digitization of legal data**. Legal-technology solutions can help law firms succeed in this new era, **but only if they understand the tech landscape** and redefine key elements of their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

*- BCG Analysis, 2016*

# Motivation

## Gartner Hype Cycle July 2016

Figure 2. Hype Cycle for Emerging Technologies, 2016



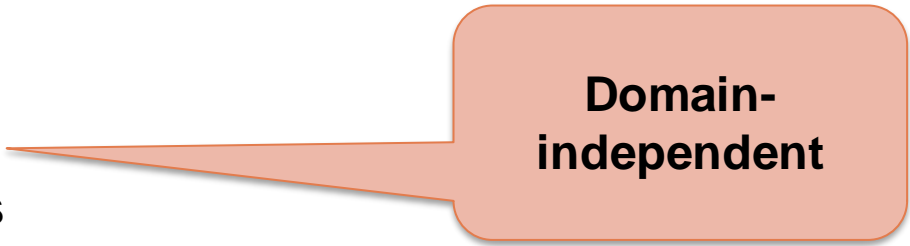
Source: Gartner (July 2016)



- Processes of Legal Experts (Scientists and Lawyers) are...
  - ... time-intensive
  - ... knowledge-intensive
  - ... data-intensive
- Legal Data Science is becoming more and more attractive, because...
  - ... process time and memory space are cheap
  - ... algorithms can process data fast and accurate
- In order to achieve highest accuracy, algorithms and data models need an adaption to the domain...
  - German legal texts (laws, judgments, contracts, etc.)
  - Data model tailored to judicial data

- Introduction & Background
- Semantic Analysis of Legal Documents
  - a. Supporting **editorial** processes
  - b. Supporting **drafting** processes
  - c. Supporting **search** processes
- Technology for Semantic Analysis
  - a. Pipes & Filters – Architecture
  - b. Machine Learning Approach

- Automated extraction of linguistic and semantic entities
  - **Basic Named Entities**
    - **Dates**
    - **Money values**
    - **Locations and places**
    - **Persons and institutions**
    - **References**
    - **Companies**
  - **Legal Entities**



**Domain-  
independent**

### SECTIONS

Open Close

### SEMANTICS

- Linguistic
- Legal Information 
  - AntragssatzIndicator
  - AntragssatzWithYear
  - Day
  - Jahre
  - Konnektor
  - Month
  - PreIndicator
  - SpecificDate
  - Streitjahr
  - StrictPostIndicator
  - StrictPreIndicator
  - Timespan
  - Year
- Comments

Enable comment mode

## Steuerliche Anerkennung von Verlusten...

2016/05/20

unberücksichtigt bleiben.

### Statement of facts

Die Beteiligten streiten im Anschluss an eine Außenprüfung über die steuerliche Anerkennung von Verlusten aus einer Finanzanlage sowie die Aktivierung vertraglicher Zinsansprüche.

Die Klägerin ist Obergesellschaft der Y-Gruppe. Sie hat ein Stammkapital von 6.000.000 DM. Gesellschafter sind Frau A und Herr B. Alleinvertretungsberechtigter Geschäftsführer ist ihr Vater C. Die Klägerin hat ein vom Kalenderjahr abweichendes Wirtschaftsjahr (1. Februar bis 31. Januar des Folgejahres). Zum 1. Oktober 2001 erwarb die Klägerin von ihren Gesellschaftern aus deren Privatvermögen Anteile an drei Aktiengesellschaften, welche Plantagen in Übersee betreiben bzw. deren Eigentümer sind, und zwar der E-AG, der F-AG und der G-AG.

Mit Vertrag vom 22. Februar 2002 veräußerte die Klägerin die zuvor erworbenen Plantagengesellschaften an die H-AG schuldenfrei zum Preis von 30.000.000 US-\$. Die Zahlung des Kaufpreises sollte gemäß einem beigefügten Ratenzahlungsplan über die Dauer von 20 Jahren gestreckt werden. Mit der Tilgung des Kaufpreises sollte ab dem 30.06.2008 begonnen werden. Im Vertrag ist eine Verzinsung des offenen Kaufpreises für die ersten 10 Jahre in Höhe von 5 % p.a. und für die Jahre 11 bis 20 in Höhe von 6 % p.a. festgelegt. Als Sicherheit für den Verkäufer wurde die Sicherungsübereignung der veräußerten Aktien vereinbart.

Zum 1. Mai 2002 verbuchte die Klägerin aus dem Geschäft mit der H-AG gemäß § 8 b KStG steuerfreie Veräußerungsgewinne ("Erträge aus Beteiligungen") in Höhe von 8.512.195,78 € und 8.226.000 € = 16.738.195,78 €. Im Rahmen einer früheren Außenprüfung für die Veranlagungszeiträume 1999 bis 2003 wurde der erklärte steuerfreie Veräußerungsgewinn im Umfang zusätzlicher Anschaffungskosten infolge einer verdeckten Einlage des Gesellschafters B um 14.914.000 € auf 1.824.195,78 € gekürzt.

In ihrer Bilanz zum 31. Januar 2003 wies die Klägerin wegen der Kaufpreisforderung aus dem Veräußerungsgeschäft mit der H-AG eine Fremdwährungsforderung in Höhe von 33.215.235 € (= 30 Mio. US-\$) aus. Mit Kauf- und Abtretungsvertrag vom 16. Juni 2003 übertrug die Klägerin 1/3 ihrer Forderung gegenüber der H-AG auf die C-Holding GmbH. Hierdurch wurden im Wirtschaftsjahr 2003/2004 Wechselkursverluste in Höhe von 2.676.849 € realisiert. Mit Vertrag vom 21. November 2003 wurde die C-Holding GmbH auf die Klägerin verschmolzen. Die Klägerin wurde hierdurch wieder vollständige Inhaberin der Forderung gegenüber der H-AG.

Umgehung der Regelung des § 6 EStG. Nach dieser Vorschrift ist eine Teilwertabschreibung nur unter engen Voraussetzungen möglich, die hier nicht vorlagen. Eine Teilwertabschreibung i.H.v. 25.189.235 € auf die

### HIGHLIGHT

Keyword

Next

Prev

Clear

### INFORMATION

#### QUANTIFICATION

► Linguistic KPIs

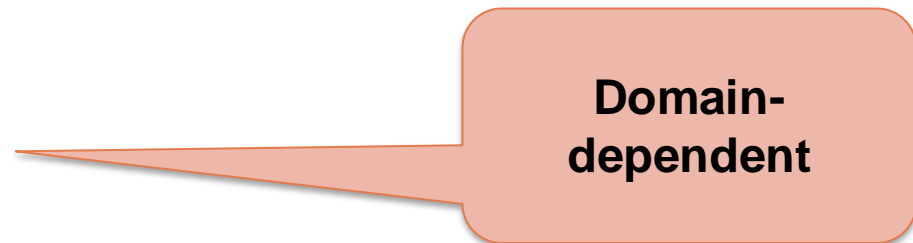
► Semantic KPIs

#### SEMANTIC LABELS

► 1999 bis 2003

► 2005

- Automated extraction of linguistic and semantic entities
  - Basic Named Entities
  - **Legal Entities**
    - **Legal definitions**
    - **Obligations**
    - **Permissions**
    - **Contracting Parties**
    - **Liabilities**
    - **“Change-of-control”**





### SECTIONS

Open Close

### SEMANTICS

- Linguistic Entities
- Named Entities
- Legal Entities

- Beweislast
- BeweislastAnordnung
- BeweislastUmkehr
- EinwendungRV
- Erlaubnis
- Erloeschen
- Fiktion
- Fortfuehrung
- FortfuehrungAusnahme
- FortfuehrungEinschraenkung
- FortfuehrungErweiterung
- FortfuehrungGleichstellung
- Konkurrenz
- LegalDefinition
- Pflichtzuweisung

## Gesetz über die Haftung für fehlerhafte Produkte 03.04.2013

ein eingeführtes Produkt, wenn sich bei diesem die in Absatz 2 genannte Person nicht feststellen läßt, selbst wenn der Name des Herstellers bekannt ist.))

### § 5 Mehrere Ersatzpflichtige

(Sind für denselben Schaden mehrere Hersteller nebeneinander zum Schadensersatz verpflichtet, so haften sie als Gesamtschuldner.) (Im Verhältnis der Ersatzpflichtigen zueinander hängt, soweit nichts anderes bestimmt ist, die Verpflichtung zum Ersatz sowie der Umfang des zu leistenden Ersatzes von den Umständen, insbesondere davon ab, inwieweit der Schaden vorwiegend von dem einen oder dem anderen Teil verursacht worden ist.) (Im übrigen gelten die §§ 421 bis 425 sowie § 426 Abs. 1 Satz 2 und Abs. 2 des Bürgerlichen Gesetzbuchs.)

### § 6 Haftungsminde rung

((1) Hat bei der Entstehung des Schadens ein Verschulden des Geschädigten mitgewirkt, so gilt § 254 des Bürgerlichen Gesetzbuchs.) (Im Falle der Sachbeschädigung steht das Verschulden desjenigen, der die tatsächliche Gewalt über die Sache ausübt, dem Verschulden des Geschädigten gleich.)

((2) Die Haftung des Herstellers wird nicht gemindert, wenn der Schaden durch einen Fehler des Produkts und zugleich durch die Handlung eines Dritten verursacht worden ist.) (§ 5 Satz 2 gilt entsprechend.)

### § 7 Umfang der Ersatzpflicht bei Tötung

((1) Im Falle der Tötung ist Ersatz der Kosten einer versuchten Heilung sowie des Vermögensnachteils zu leisten, den der Getötete dadurch erlitten hat, daß während der Krankheit seine Erwerbsfähigkeit aufgehoben oder gemindert war oder seine Bedürfnisse vermehrt waren.) (Der Ersatzpflichtige hat außerdem die Kosten der Beerdigung demjenigen zu ersetzen, der diese Kosten zu tragen hat.)

((2) Stand der Getötete zur Zeit der Verletzung zu einem Dritten in einem Verhältnis, aus dem er diesem gegenüber kraft Gesetzes unterhaltspflichtig war oder unterhaltspflichtig werden konnte, und ist dem Dritten infolge der Tötung das Recht auf Unterhalt entzogen, so hat der Ersatzpflichtige dem Dritten insoweit Schadensersatz zu leisten, als der Getötete während der mutmaßlichen Dauer seines Lebens zur Gewährung des Unterhalts verpflichtet gewesen wäre.) (Die Ersatzpflicht tritt auch ein, wenn der Dritte zur Zeit der Verletzung gezeugt, aber noch nicht geboren war.)

### § 8 Umfang der Ersatzpflicht bei Körperverletzung

### HIGHLIGHT

Keyword

Next Prev Clear

### INTERACTIVE SEARCH

Enable interactive mode

Comments

### QUANTIFICATION

### SEMANTIC LABELS

- (4) Für den Fehler, den Schaden und d...
- Ist streitig, ob die Ersatzpflicht ge...
- (3) Kann der Hersteller des Produkts ...
- Im Falle der Verletzung des Körpers o...
- Als Hersteller gilt auch jeder, der s...
- (2) Als Hersteller gilt ferner, wer e...
- (3) Kann der Hersteller des Produkts ...
- (2) Die Haftung des Herstellers wird ...
- (1) Der Anspruch nach § 1 erlischt ze...
- Entgegenstehende Vereinbarungen sind ...
- (4) Für den Fehler, den Schaden und d...

- **Collaborative drafting of documents**
- **Support during the creation of legal documents**
  - Laws and judgments
  - Contracts, regulations, administrative orders, etc.
- **Instant feedback**
  - **Readability:** Flesch-Reading-Ease, Vocabulary, etc.
  - **Basic entities:** Date, locations, persons, institutions, references, etc.
  - **Legal entities:** Obligations, permissions, legal definitions, etc.

# Drafting of Legal Documents

## Support of drafting processes

Live update on Recommendations

**HIGHLIGHT**

Keyword

Next Prev Clear

Inhalt

§1 Schönheitsreparaturen

### Mietvertrag

+ Group - Group

File Edit Insert View Format Table Tools

Formats B I

Schönheitsreparaturen umfassen das Tapezieren, Anstreichen der Wände und Decken, das Streichen der Heizkörper einschließlich Heizrohre, der Innentüren sowie der Fenster und Außentüren von innen, das Streichen der Fußleisten und der offen liegenden Ver- und Entsorgungsleitungen.

+ Group + Article - Article

**SETTINGS**

Save Cancel

**RECOMMENDATIONS**

For article: § 1 Schönheitsreparaturen  
Based on highlighted text: Anstreichen der Wände...

Nr	Document
+ 1	Wohnraummiete: Wirksamkeit der Schönheitsreparaturklausel mit der Verpflichtung des Mieters zum Streichen der Wohnungseingangstür und der Fenster
+ 2	Mietvertrag
+ 3	Formularmäßiger Wohnraummietvertrag: Inhaltskontrolle für ein Quotenabgeltungsklausel für bei Mietende noch nicht fällige Schönheitsreparaturen
+ 4	Wohnraummiete: Inhaltskontrolle einer Schönheitsreparaturklausel; Gesamtnwirksamkeit wegen unangemessener Benachteiligung des Mieters
+ 5	Formularmäßiger Wohnraummietvertrag: Unwirksamkeit eines

- I. **Smart Recommendations** → Missed some highly relevant information?
- II. **Missing clauses** → Have you missed a liability clause?
- III. **Suspicious clauses** → Clause "looks" very unusual?
- IV. **Clause Effectiveness** → Courts have decided on some clause?

- Introduction & Background
  
- Semantic Analysis of Legal Documents
  - a. Supporting **editorial** processes
  - b. Supporting **drafting** processes
  - c. Supporting **search** processes
  
- Technology for Semantic Analysis
  - a. Pipes & Filters – Architecture
  - b. Machine Learning Approach



# Reference Architecture

## Implementation Details

- Java web application
- ElasticSearch
- Apache UIMA
- Apache Ruta

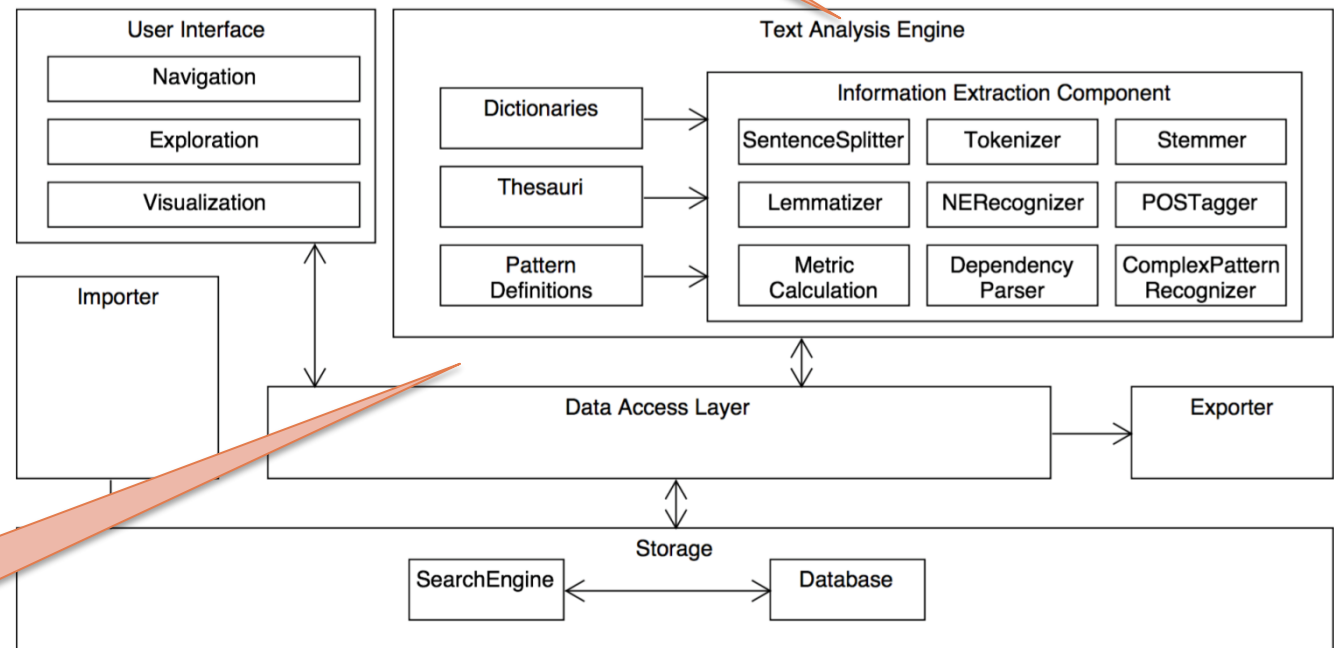


elastic



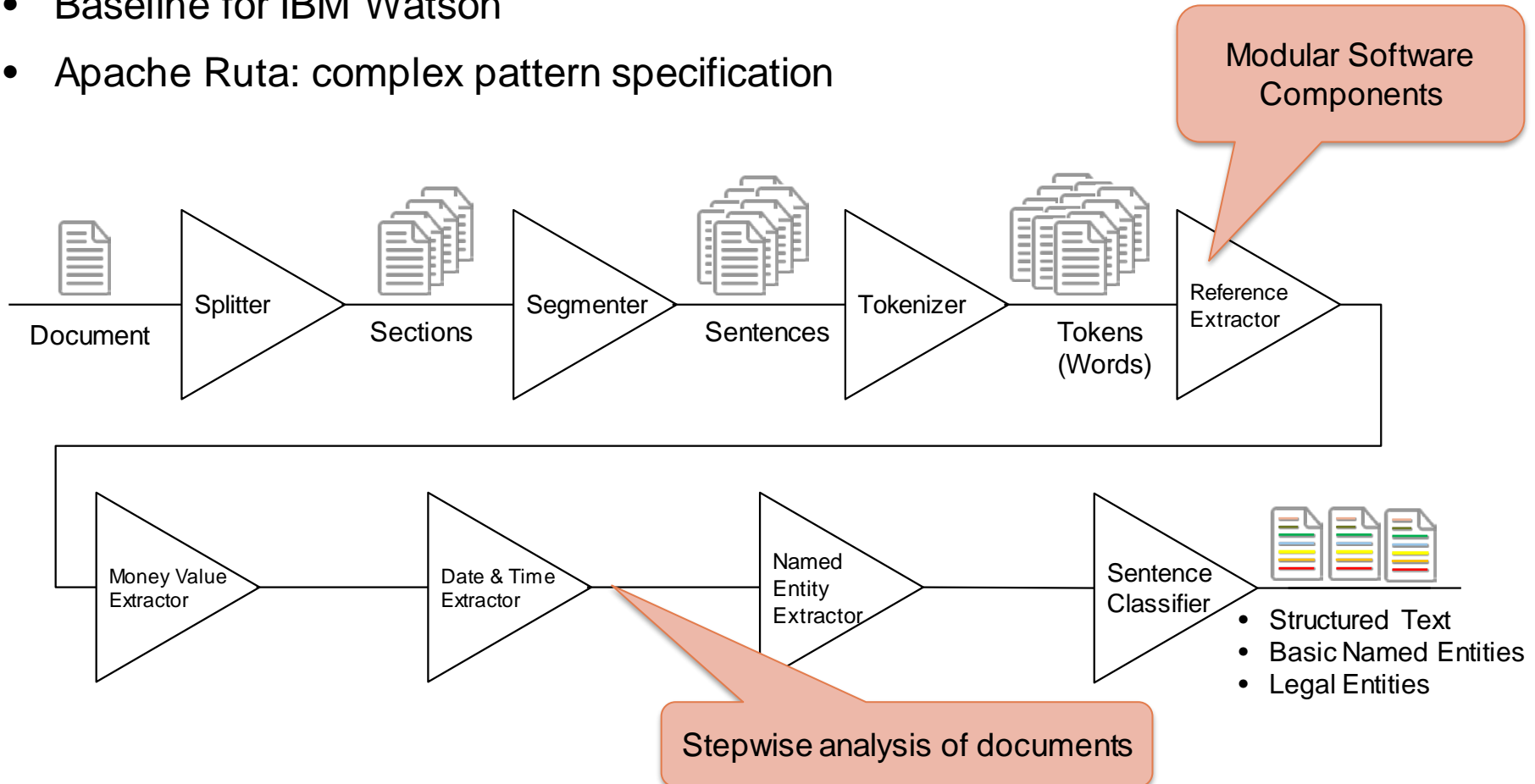
Extensibility

Modularity



## UIMA (Unstructured Information Management Architecture)

- Pipes & Filters architecture
- Baseline for IBM Watson
- Apache Ruta: complex pattern specification

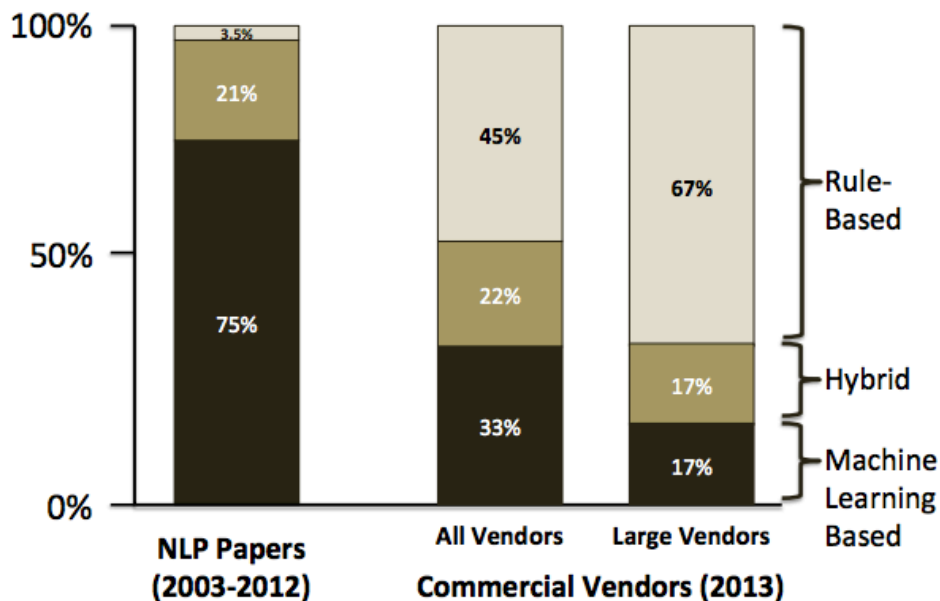


# Extraction of semantic information

## Entity extraction in practice

- Survey among 54 different entity extraction tool vendors in 2013

### Implementations of Entity Extraction



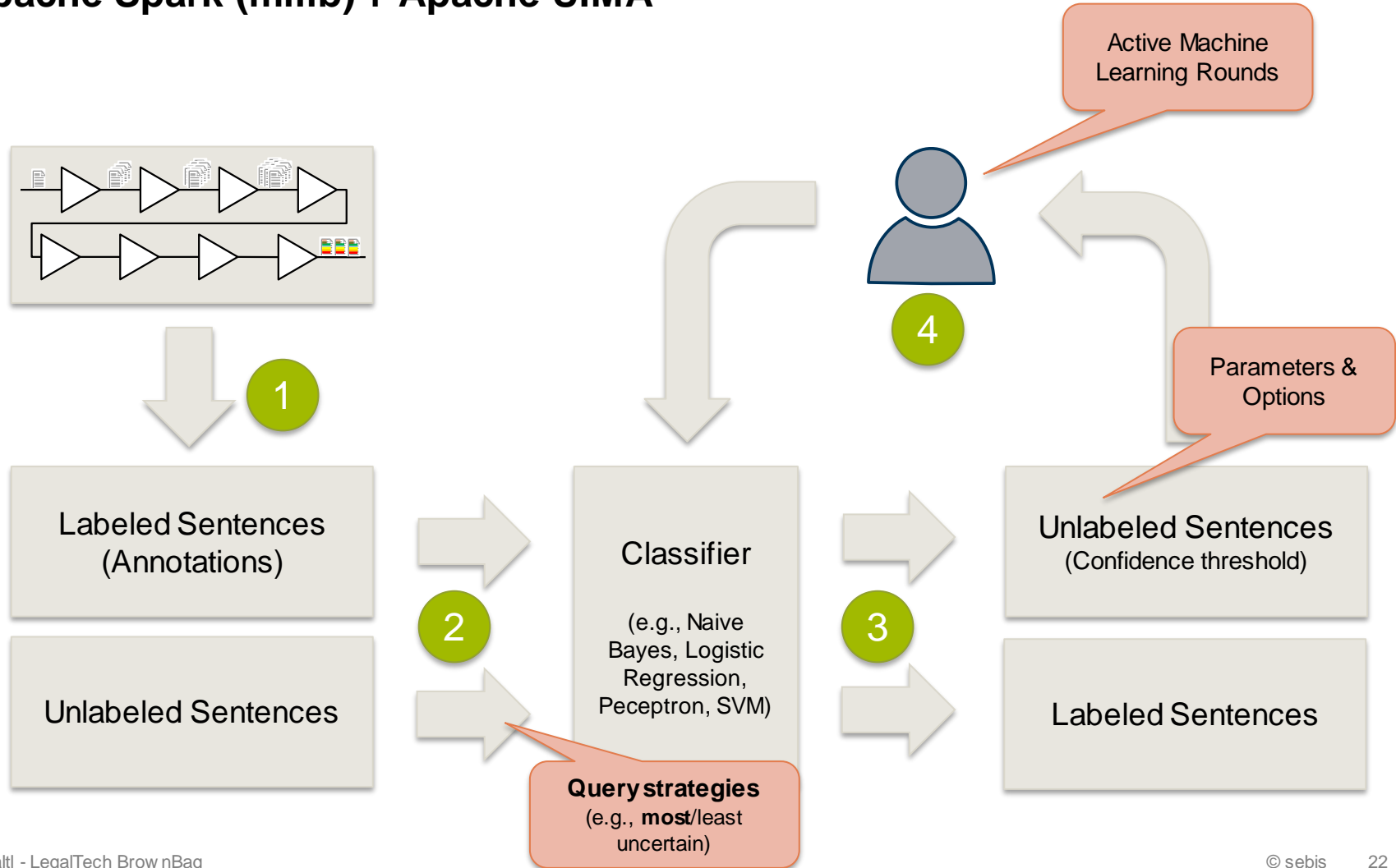
	Pros	Cons
Rule-based	<ul style="list-style-type: none"> <li>• Declarative</li> <li>• Easy to comprehend</li> <li>• Easy to maintain</li> <li>• Easy to incorporate domain knowledge</li> <li>• Easy to trace and fix the cause of errors</li> </ul>	<ul style="list-style-type: none"> <li>• Heuristic</li> <li>• Requires tedious manual labor</li> </ul>
ML-based	<ul style="list-style-type: none"> <li>• Trainable</li> <li>• Adaptable</li> <li>• Reduces manual effort</li> </ul>	<ul style="list-style-type: none"> <li>• Requires labeled data</li> <li>• Requires retraining for domain adaptation</li> <li>• Requires ML expertise to use or maintain</li> <li>• Opaque</li> </ul>

Chiticariu, Laura, Yunyao Li, and Frederick R. Reiss. "Rule-based information extraction is dead! long live rule-based information extraction systems!." *EMNLP*. No. October. 2013.

# Classification of Norms

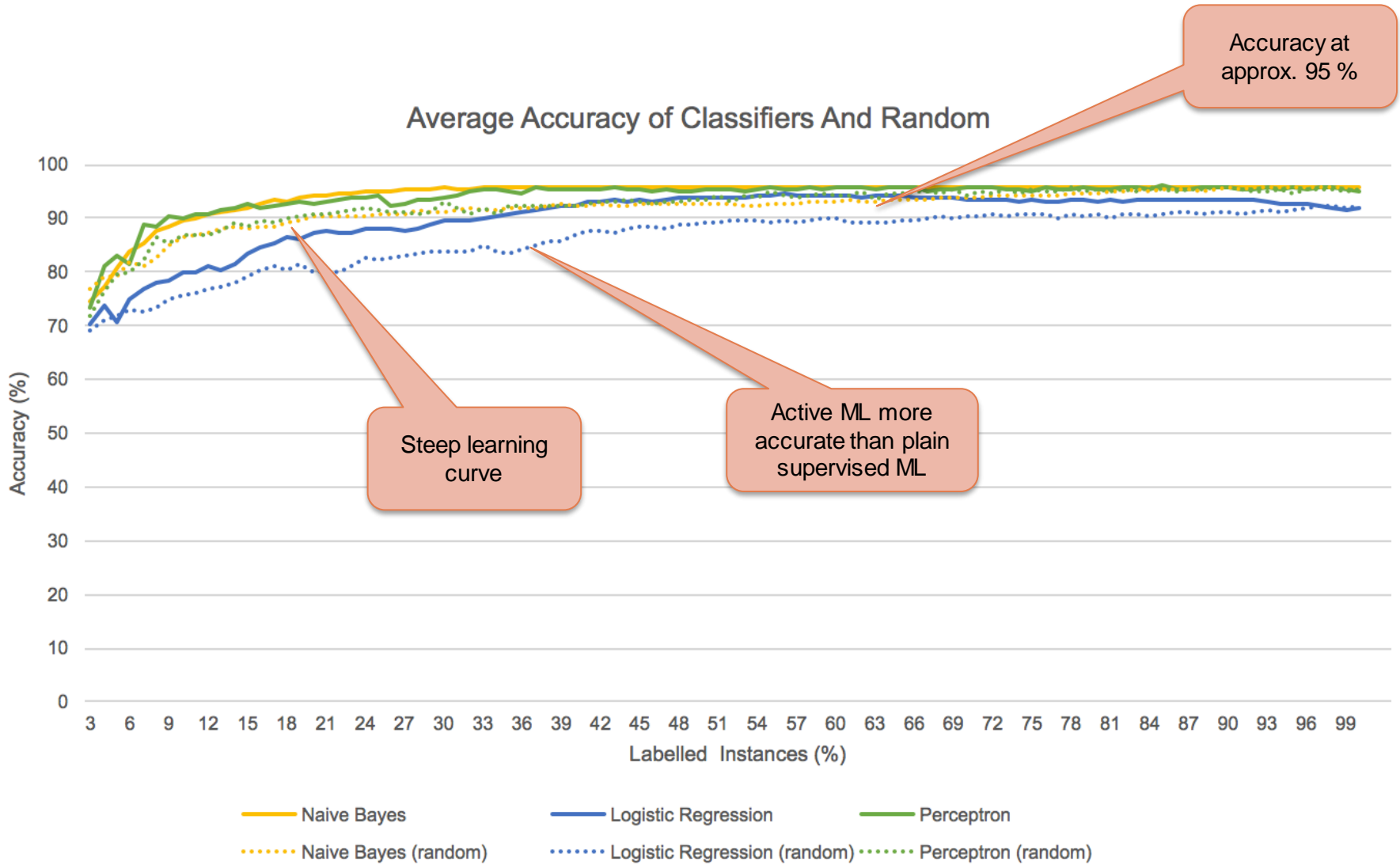
## Active Machine Learning: A Hybrid Approach

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



# Preliminary Results

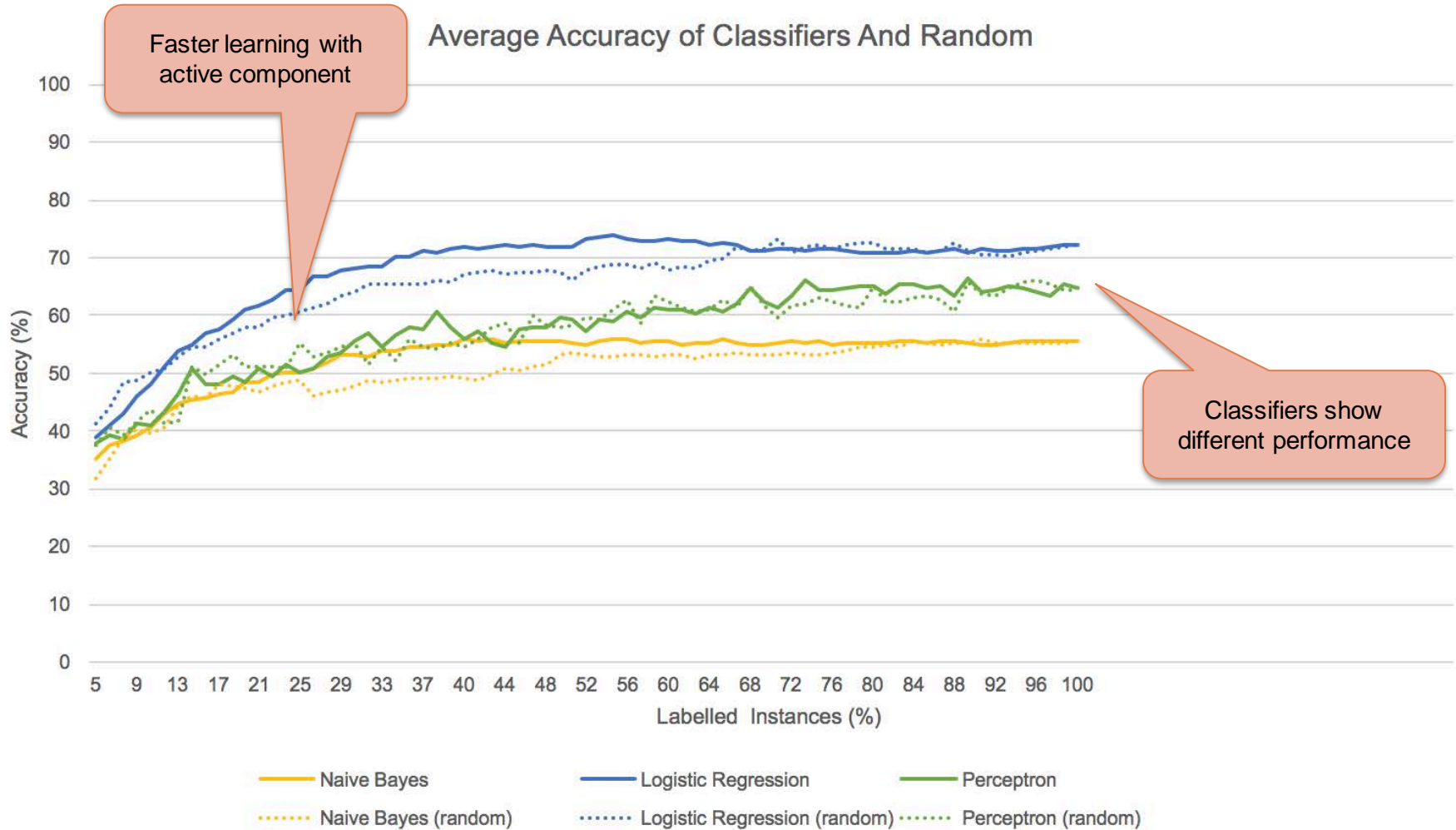
## Document Classification





# Preliminary Results

## Norm Classification



# Conclusion

- **Legal Informatics** is an vital research area with many different facets

- Automated annotation of textual data
- Extraction of relevant information



- Legal Tech is an interesting field but needs different methods and objectives.

- Interdisciplinary research is highly interesting ([www.lexalyze.de](http://www.lexalyze.de))

- Legal scientists and practitioners
- Linguists
- Software Engineers
- etc.

