

Analyse von Rechtstexten mit Methoden der „Legal Analytics“: das Projekt „ManyLaws“

Günther Schefbeck

28. Deutscher EDV-Gerichtstag
Saarbrücken, 19. September 2019

Die Stimme des Propheten



“In a few years, lawyers will rely more and more on computers to perform many tasks for them. They will not rely on computers simply to do their bookkeeping, filing or other clerical tasks. They will also use them in their research and in the analysis and prediction of judicial decisions. In the latter tasks, they will make use of modern logic and the mathematical theory of probability, at least indirectly.”

Die Stimme des Propheten



“In a few years, lawyers will rely more and more on computers to perform many tasks for them. They will not rely on computers simply to do their bookkeeping, filing or other clerical tasks. They will also use them in their research and in the analysis and prediction of judicial decisions. In the latter tasks, they will make use of modern logic and the mathematical theory of probability, at least indirectly.”

Reed C. Lawlor, What computers can do: analysis and prediction of judicial decisions, American Bar Association Journal 49 (1963), 337-344

„Big Data Analytics“



- “Big Data”: Massendaten (“3 v”: “volume, velocity, variety”), nicht in relationalen Datenbanken gehalten, parallel verarbeitet, Herkunft z.B. aus Onlineverhalten, Suchmaschinennutzung, “Social Media”, elektronischer Kommunikation, Nutzung vernetzter Geräte, OGD
- “Big Data Analytics”: Einsatz lernfähiger elektronischer Werkzeuge zur Mustererkennung in Massendaten und Ableitung von Schlussfolgerungen
- “Predictive Analytics”: Ableitung von Prognosen aus Analyse von Massendaten
- Ermöglicht durch signifikantes Wachstum der verfügbaren Datenmengen und der Rechenleistung (Parallelisierung)
- Rascher Technologiewandel: z.B. Ablösung von MapReduce durch Spark
- Anwendungsbereiche: z.B. Produkt- und Servicemarketing, Kriminalitäts- und Terrorismusbekämpfung, Epidemievorhersage, Wahlbeeinflussung, Bonitätsprüfung, “social scoring”
- Erkenntnistheoretische Implikation: Korrelation statt Kausalität

„Legal Analytics“



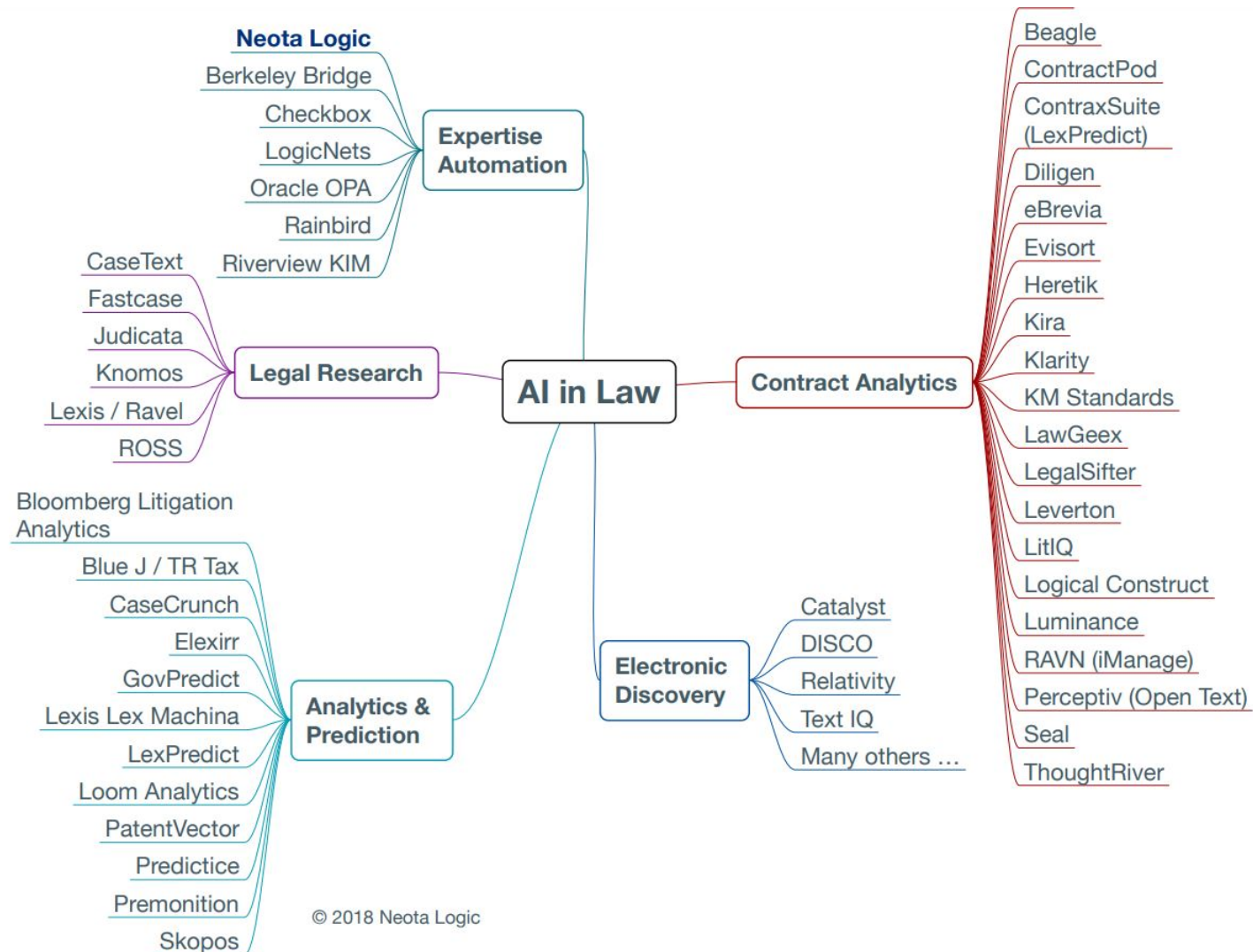
- “Legal Analytics”: Anwendung von “Big Data Analytics”- Methoden auf juristische Massendaten
- Begriff seit 2013 von LexisNexis-Dienst “LexMachina” verwendet
- Nachteil: Datenmengen kleiner als z.B. in kommerziellen oder medizinischen Anwendungen
- Vorteil: Daten oft besser strukturiert als herkömmliche Massendaten, z.T. sogar semantisch annotiert (z.B. Beschlagwortung nach EUROVOC, Auszeichnung mittels LegalRuleML)
- Anwendungsbereiche: z.B. juristische Recherche (implizite Verweisungen, Präzedenzfälle), Erkennung von Entscheidungsmustern (Behörden, Gerichte, Richter) und Vorhersage von Entscheidungen, Unterstützung in Abfassung juristischer Schriftsätze, Klassifikation, Evaluierung oder Generierung rechtlicher Dokumente (z.B. Verträge)

Beispiele



- Akademische Studien wie z.B. *Aletras et al., Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective, PeerJ Computer Science 2 (2016)*: Analyse von EGMR-Fällen mit NLP-Methoden (Bildung von N-grams und Wortclustern sowie Klassifizierung, Training eines "machine learning"-Algorithmus), erzielte Vorhersagewahrscheinlichkeit 79 %)
- Kommerzielle Anbieter, bisher vorzugsweise auf dem U.S. Markt, wie z.B. Ravel Law (Start-up, 2017 von RELX übernommen): „Case Analytics“, „Court Analytics“, „Judge Analytics“ („Motion Analytics“), „Firm Analytics“: Endbenutzerwerkzeuge zur Unterstützung in Recherche, Schriftsaterstellung und Entscheidungsvorhersage sowie Auswahl rechtsfreundlicher Hilfe
- Nationale Projekte wie z.B. LexDatafication (CIRSFID, Universität Bologna) zur Aufbereitung italienischer Rechtsdaten für semantische Analyse durch Migration nach AKN und semantische Annotation
- Europäische Projekte wie z.B. Lynx: Entwicklung von Diensten zur Unterstützung von Unternehmen in der Beachtung europäischer und nationaler Rechtsvorschriften

„Legal Analytics“: Dienstleister (Auswahl)



Markteffekt von „Legal Analytics“



Market Problem	AI Solution	Lawyer Impact	Examples
The volume costs of lawyer effort	Machine-learning programs perform lawyer tasks faster and therefore much less expensively.	Law firms can no longer sell lawyer hours spent on these tasks, reducing inventory.	E-discovery, contract analysis, legal research
The scarcity costs of lawyer expertise	Expert systems tap into legal knowledge databases to answer questions and forecast outcomes.	Lawyers' expertise becomes less exclusive, increasing supply and lowering price.	Expert applications, outcome prediction

(Jordan Furlong, Law21, 2018)

„Legal Analytics“: Konzepte und Methoden



- Rechtsinformation üblicherweise repräsentiert in Text (generelle und individuelle Normen, begleitende Information wie erläuternde Bemerkungen oder Kommentare)
- “Natural Language Processing” (NLP), z.B.
 - “N-grams”: Fragmentierung von Text in zusammenhängende Sequenzen von (n) Elementen (z.B. Wörtern)
 - “Bag of Words”: Repräsentation eines Dokuments als Multimenge von Termen (Wörtern oder “N-grams”) unter Vernachlässigung der sequentiellen Ordnung der Terme im Dokument
 - “Word (and/or phrase) embeddings”: Repräsentation von Worten/Phrasen als Vektoren in einem kontinuierlichen Vektorraum zur Ermöglichung der Berechnung von Distanzen/Ähnlichkeiten
 - “Topics”: statistisches Modell zur Entdeckung semantischer Strukturen in Texten durch “clustering” semantisch ähnlicher Wörter/“N-grams”
- Maschinelles Lernen (ML)
 - Unüberwacht: explorative, ergebnisoffene Mustererkennung durch “clustering” von Datenobjekten in iterativen Annäherungsverfahren
 - Überwacht: Training des Algorithmus anhand validierter Daten zur Ableitung eines Klassifikationsmodells



ManyLaws

Legal Text Mining unter Verwendung von
“Big Data”-Verarbeitungsinfrastrukturen

Projektparameter



- Dauer: 2 Jahre (Verlängerung um 6 Monate geplant)
- Start: 1. Oktober 2018
- Gefördert durch die CEF (Connecting Europe Facility)
- Partner:
 - Universität der Ägäis (Koordination)
 - Donau-Universität Krems
 - Intrasoft International
 - Griechisches Parlament
 - Österreichisches Parlament

Ziele



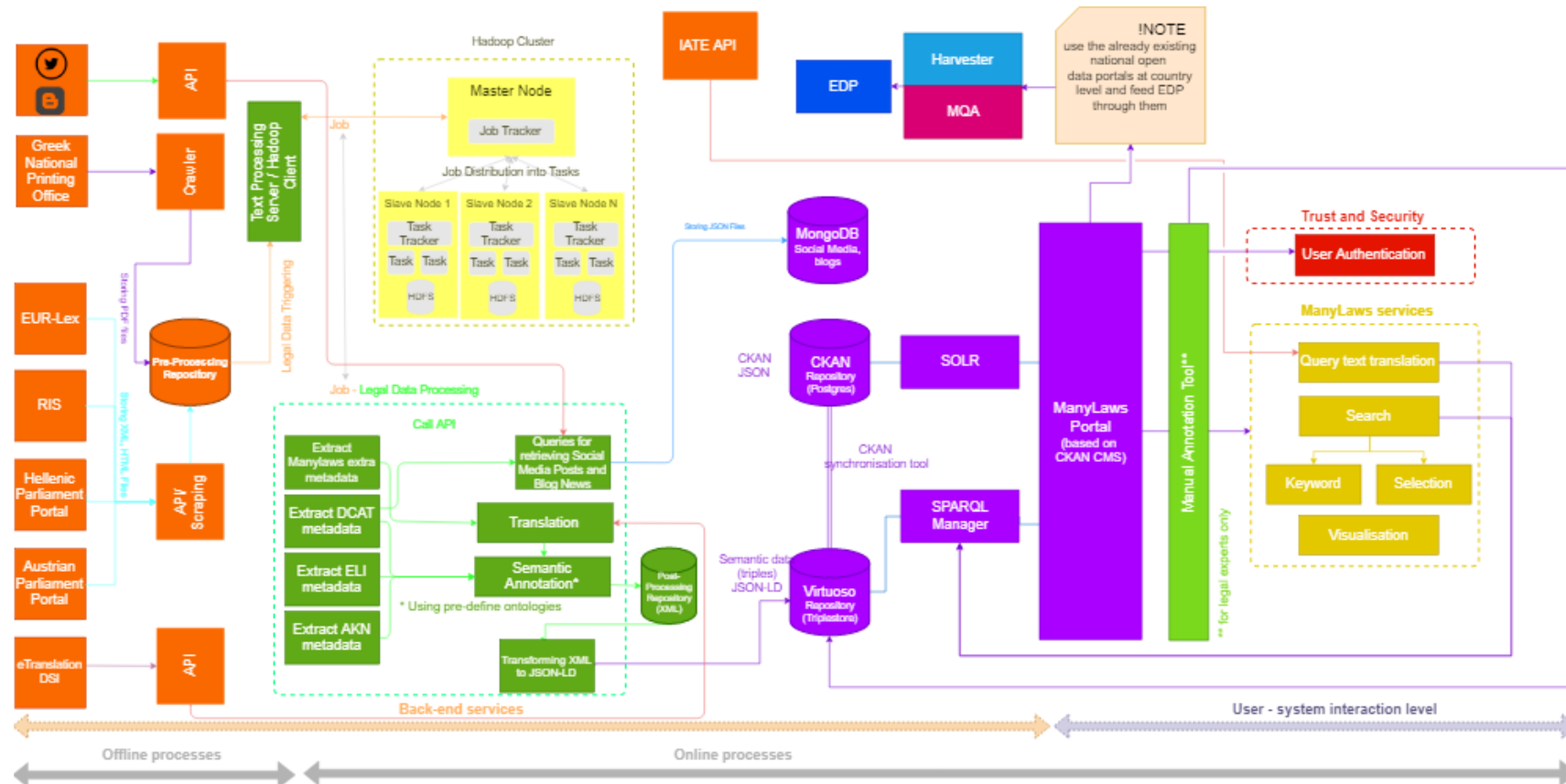
- Dienste für Bürger, Unternehmen, Verwaltung, Rechtsetzung
- Zugang zu öffentlicher Rechtsinformation unter Nutzung des EDP
- Unterstützung von Suchvorgängen durch mittels ML-basierter Mustererkennung angereicherte Metadaten
- Überwindung von Sprachbarrieren in Suchvorgang und Ergebnisdarstellung
- Verbesserung der Effizienz von Rechtsetzungsprozessen

Ressourcen und Werkzeuge



- Ressourcen:
 - Offene Rechtsdaten, eingebettet in nationale und europäische OGD-Strukturen (z.B. EU ODP, OGD Portal Austria, OGD Portal Greece)
 - Subsidiär proprietäre Rechtsdaten (z.B. Nomos-Datenbank)
- Architektur basiert auf CKAN als Quasi-Standardportalsoftware für das Teilen offener Daten:
 - CKAN-Extensionen (z.B. Harvester, DCAT, Extrafields/Scheming, SPARQL Interface, Spacial/Geoview, CMS, Visualisation libraries)
 - Virtuoso Triplestore
 - Mongo DB (optional für „Opinion Mining“)
 - IATE API und CEF DSI eTranslation für maschinelle Übersetzungen
- Metadatenschema nutzt drei vordefinierte Vokabularien:
 - DCAT-AP (interoperabel mit Dublin Core, EUROVOC etc.)
 - ELI
 - Akoma Ntoso

Systemarchitektur



Dienste (Auswahl)



- Sprachübergreifende Suche in europäischen und nationalen Rechtsvorschriften und legislativen Dokumenten
 - Maschinelle Transformation der Suchbegriffe
 - Maschinelle Arbeitsübersetzungen der Suchergebnisse
- Erweiterung auf Entscheidungen für Folgeprojekt geplant
- Erkennung vertikaler und horizontaler Interdependenzen und Äquivalenzen, z.B.
 - Umsetzung europäischer Rechtsakte in nationales Recht (Vollständigkeit, „goldplating“)
 - Vergleich der Umsetzung in verschiedenen nationalen Jurisdiktionen
 - Vergleich domänenspezifischer Regelungen in verschiedenen nationalen Jurisdiktionen
- Erkennung expliziter und impliziter Referenzen
- Analyse der Auswirkungen von Rechtsaktsentwürfen auf Rechtsordnung
- Visualisierungen, z.B.
 - Interdependenzen zwischen Rechtsvorschriften
 - Verlauf von Rechtsetzungsprozessen
 - Änderungsdynamik von Rechtsvorschriften
- Entscheidungsunterstützungssystem für Rechtsetzungsprozess (Option: Berücksichtigung der Ergebnisse von „Opinion Mining“/“Sentiment Analysis“)

Exemplarisches Nutzungsszenario

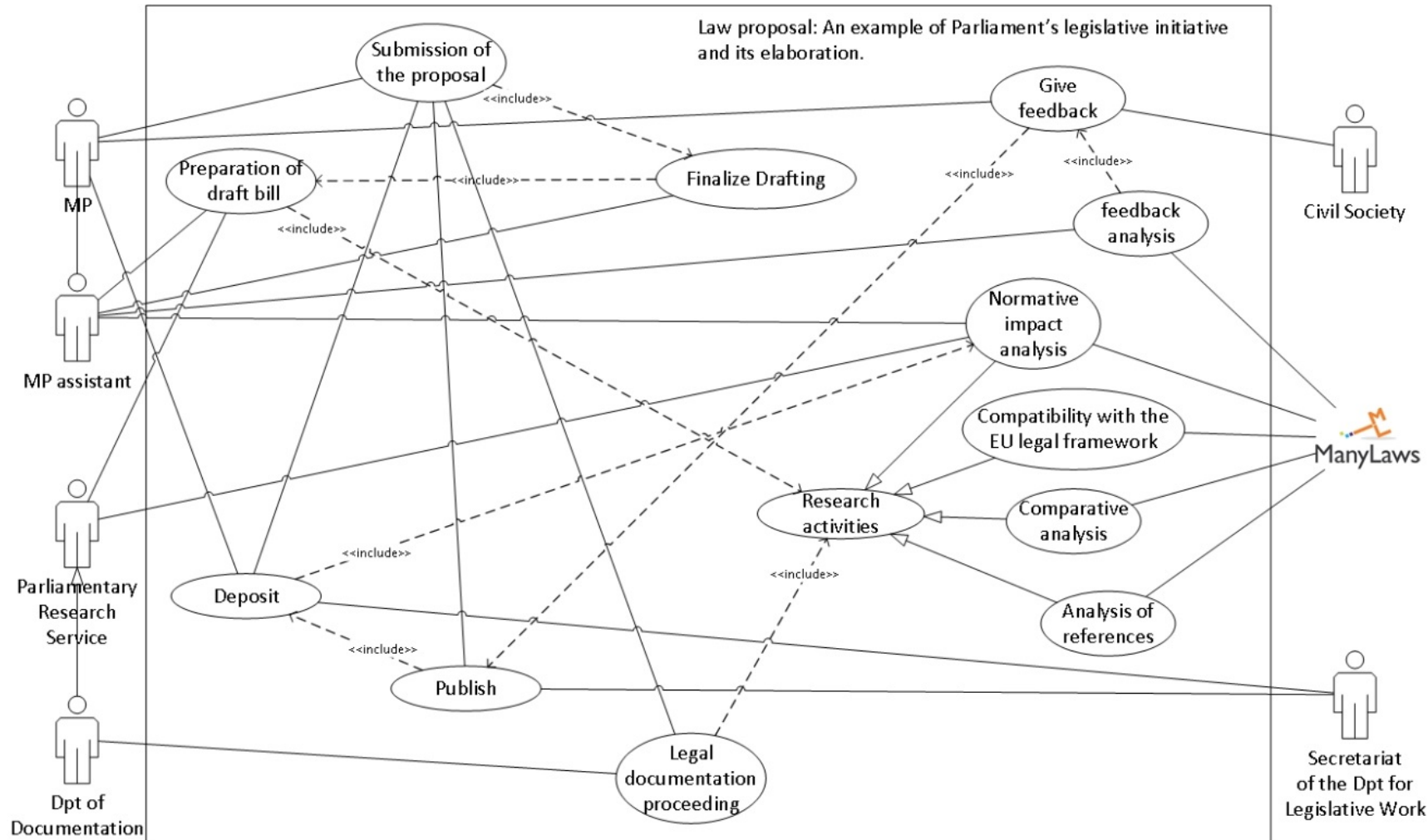


Diagramm:
Nektarios
Kyriakou

Danke für Ihre Aufmerksamkeit!

Weitere Information:

guenther.schefbeck@parlament.gv.at