KORPUSLINGUISTIK IN DER RECHTSWISSENSCHAFT

Eine webbasierte Analyseplattform

für EuGH-Entscheidungen

Bettina Mielke, OLG Nürnberg / Universität Regenburg

Christian Wolff, Lehrstuhl für Medieninformatik, Universität Regensburg

Übersicht

- Einführung
  - Korpuslinguistische Arbeitsgruppen in der Rechtsinformatik
  - Was leisten korpuslinguistische Verfahren?
  - Aktuelle Analysebeispiele
- Designziele für eine Analyseplattform
  - Technisch
  - Inhaltlich / juristisch
- Technische Konzeption und Umsetzung
  - Korpuslinguistische Komponenten
  - Benutzerschnittstelle und Visualisierung
- Ausblick

Aktuelle korpuslinguistische Arbeiten im Rechtswesen – Arbeitsgruppen in Deutschland (Auswahl)

- Erlangen: Adrian / Evert – Anonymisierung von Gerichtsurteilen
- Berlin: Leibniz Linguistic Research into Constitutional Law – Analyse von Urteilen des BVerfG: umfassend annotiertes Korpus mit Entscheidungen des BVerfG
- Heidelberg: Hamann / Vogel – juristisches Referenzkorpus (JuReKo) des deutschsprachigen Rechts, das Entscheidungstexte, juristische Aufsatzliteratur und Normtexte von 1980 bis 2015 enthält
- Regensburg: Mielke / Wolff seit 2004 – Text Mining, Vergleich der deutsch-österreichischen (und „europäischen“) Rechtssprache, Aufbau Analyseinfrastruktur

Was leisten korpuslinguistische Verfahren?

- Einfache quantitative Textmetriken, Häufigkeiten (Wörter, Sätze) – Statistik
- Part-of-Speech-Analysen, Syntaxanalysen – sprachliches Wissen
- Inhaltsanalyse (Textklassifikation, Textzusammenfassung, Textvergleich, Information Retrieval)
- In den digitalen Geisteswissenschaften zuletzt zunehmend gefragt:
  - Stilometrie
  - Zuordnung von Autorenschaft
  - Emotion / Sentiment / Wertungen
  - Lesbarkeitsanalysen
  - Solche Analysen auch im juristischen Kontext relevant

Aktuelle Analysebeispiele aus den USA / zur Arbeit des Supreme Court

- Umgang mit Lexika und lexikalischer Bedeutung (Mouritsen 2010)
- Autorenschaft anhand des Schreibstils am Supreme Court (Rosenthal & Yoon 2011)
- Nachweis veränderter Aufgabenverteilung am Supreme Court durch die Untersuchung stilistischer Unterschiede in den Entscheidungen (Carlson et al. 2016)
- Längsschnittstudie (50 Jahre) zur Entwicklung der Urteile des Supreme Court (Livermore et al. 2017)
- Analyse einzelner Richter, hier: Neil Gorsuch (Varsava 2018)

Aktuelle Analysebeispiele aus dem deutschsprachigen Bereich – Überblick

- Wandel des Staatsbegriffs in einem Schweizer Textkorpus (Abegg & Bubenhofer 2016)
- Analyse der Grundrechtsverletzungen in erfolgreichen Verfassungsbeschwerden (Wendel 2020)
- Begriffsanalyse zu „geschäftsmäßig“ in der juristischen Fachsprache und im allgemeinen Sprachgebrauch durch Vergleich DeReKo/JuReKo (Vogel 2019)
- Untersuchung der Varietäten der deutschen Rechtssprache (Mielke & Wolff 2013, 2016, Berteloot et al. 2018, Auer et al. 2019)

Designziele für eine Analyseplattform

- Technisch:
  - Nutzung aktueller NLP-Frameworks und -Bibliotheken

- Inhaltlich / juristisch
  - Aufbau eines Textkorpus aus frei zugänglichen juristischen Texten
  - Analyse von EuGH-Entscheidungen

Datengrundlage EuGH-Urteile

- Import über das API von EUR-LEX, Nutzung der dort verfügbaren Webservices
- Volltext und Metadaten verfügbar
- 24 Sprachversionen
- Zusätzliche Metadaten
- Insg. ca 6.000 Dokumente – pro Sprachversion
- Textumfang ca. 1,2 GB

Database Name	Storage Size	Collections	Indexes
judgment_corpus	1.2GB	24	25

Collection Name	Documents	Avg. Document Size	Total Document Size
judgments_bg	5,555	6.6 KB	35.8 MB
judgments_cs	5,555	7.7 KB	41.8 MB
judgments_da	5,555	42.2 KB	229.0 MB
judgments_de	5,555	44.4 KB	240.8 MB
judgments_el	5,555	77.6 KB	420.9 MB
judgments_en	5,554	44.0 KB	238.9 MB

Fischer et al. 2020b, 9

```
_id: ObjectId("5f6cad54f7fb5c7af6db16b5")
reference: "d1c1196f-2248-4b88-93e4-6c4cf337073e_en"
title: "Judgment of the Court of 16 July 1956. # Fédération Charbonnière de Be..."
text: "Avis juridique important | 61955J0008 Judgment of the Court of 16 July..."
keywords: "++++ 1 . PROCEDURE - APPLICATION FOR ANNULMENT - DECISIONS OF THE HIGH..."
parties: "IN CASE 8/55 FEDERATION CHARBONNIERE DE BELGIQUE, REPRESENTED BY LOUIS..."
subject: "APPLICATION FOR THE ANNULMENT OF DECISION NO 22/55 OF THE HIGH AUTHORI..."
endorsements: "UPON READING THE PLEADINGS; UPON HEARING THE PARTIES; UPON HEARING THE..."
grounds: "P . 255 A - THE ADMISSIBILITY OF THE APPLICATION THE APPLICATION SEEKS..."
decisions_on_costs: null
operative_part: "THE COURT HEREBY : 1 . DECLARES THAT THE APPLICATION IS ADMISSIBLE; 2 ..."
celex: "61955CJ0008"
> author: Object
> subject_matter: Object
> case_law_directory: Object
ecli: "ECLI:EU:C:1956:7"
date: 1956-07-16T00:00:00.000+00:00
case_affecting: null
> applicant: Object
> defendant: Object
affected_by_case: null
> procedure_type: Object
```


Übernahme des EUR-LEX-Datenmodells in MongoDB

Fischer et al. 2020b, 11

Technische Konzeption und Umsetzung

- Dokumentakquise, Steuerung: Python
- Persistenzschicht; MongoDB (NOSQL-Datenbank)
- Redis / Celery zur Unterstützung der Datenverarbeitung
- Abfragemodell auf der Basis von JSON
- Projektmanagement: Trello (Scrum), GitHub

Verarbeitung der Urteile mit NLP-Komponenten

- Nutzung aktueller NLP-Frameworks
  - spaCy
  - textaCy
  - Stanza
- Nutzung vortrainierter Sprachmodelle:
  - Englisch-Blackstone-Modell des *Incorporated Council of Law Reporting for England and Wales*
  - Deutsch: SpaCy-Modell auf der Basis bekannter (nicht-juristischer) Korpora

BLACKSTONE

Open source natural language processing for Legal Texts

Blackstone is a spaCy model and library for processing long-form, unstructured legal text. Blackstone is an experimental research project from the Incorporated Council of Law Reporting for England and Wales' research lab, ICLR&D.

https://research.iclr.co.uk/blackstone

Textanalyse-Komponenten I

- Vorbereitung der Dokumente / Preprocessing
- Normalisierung der Materialien
- Einteilung der Texte in einzelne Token
- Part-of-Speech-Tagging (POS Tagging)
- Erkennung von Eigennamen (named entity recognition)
- linguistische Verarbeitungsschritte
- Grundformreduktion / Lemmatisierung
- Erkennung auffälliger N-Gramme / Kollokationen

Textanalyse-Komponenten II

- Schlagwortextraktion (PositionRank-Verfahren)
- Berechnung einer paarweisen Dokumentähnlichkeit (Word Embeddings-Verfahren)
- Metriken zur Lesbarkeit (Flesch-Reading-Ease, aus TextaCy)
- Sentiment-Analyse

Abfragebeispiel: Bigramme / Lesbarkeit

```
{
    "language": "en",
    "corpus": [
        {
            "column": "date",
            "start date": "1958-07-17",
            "end date": "1959-07-17"
        },
    "analysis": [
        {
            "type": "n-grams",
            "n": 2,
            "limit": 10
        },
        {
            "type": "readability"
        }
    ]
}
```

```
{
    "n-grams": [
        [
            "high authority",
            106
        ],
        [
            "ecsc treaty",
            30
        ],
        [
            "decisions nos",
            28
        ],
        [
            "common market",
            20
        ],
        [
            "national law",
            17
        ],
    ],
    "readability": 31.067219480746996
}
```


Fischer et al. 2020b, 12

Was leisten die trainierten Modelle? Beispiel Blackstone (https://research.iclr.co.uk/blackstone)

CASENAME	Case names	Smith v Jones, In re Jones, In Jones' case
CITATION	Citations (unique identifiers for reported and unreported cases)	(2002) 2 Cr App R 123
INSTRUMENT	Written legal instruments	Theft Act 1968, European Convention on Human Rights, CPR
PROVISION	Unit within a written legal instrument	section 1, art 2(3)
COURT	Court or tribunal	Court of Appeal, Upper Tribunal
JUDGE	References to judges	Eady J, Lord Bingham of Cornhill

Was leisten die trainierten Modelle? Beispiel Blackstone (https://research.iclr.co.uk/blackstone)

```
nlp = spacy.load("en_blackstone_proto")

text = """ 31 As we shall explain in more detail in examining the submission of the Secretary of State
(see paras 77 and following), it is the Secretary of State's case that nothing has been done by
Parliament in the European Communities Act 1972 or any other statute to remove the prerogative power of
the Crown, in the conduct of the international relations of the UK, to take steps to remove the UK from
the EU by giving notice under article 50EU for the UK to withdraw from the EU Treaty and other relevant
EU Treaties. The Secretary of State relies in particular on Attorney General v De Keyser's Royal Hotel
Ltd [1920] AC 508 and R v Secretary of State for Foreign and Commonwealth Affairs, Ex p Rees-Mogg
[1994] QB 552; he contends that the Crown's prerogative power to cause the UK to withdraw from the EU
by giving notice under article 50EU could c
words to that effect, alternatively by legi
Secretary of State contends that neither th
to have abrogated this aspect of the Crown'
implication.
"""
```

```
# Iterate through the entities identified by the model
for ent in doc.ents:
    print(ent.text, ent.label_)

>>> European Communities Act 1972 INSTRUMENT
>>> article 50EU PROVISION
>>> EU Treaty INSTRUMENT
>>> Attorney General v De Keyser's Royal Hotel Ltd CASENAME
>>> [1920] AC 508 CITATION
>>> R v Secretary of State for Foreign and Commonwealth Affairs, Ex p Rees-Mogg CASENAME
>>> [1994] QB 552 CITATION
>>> article 50EU PROVISION
```

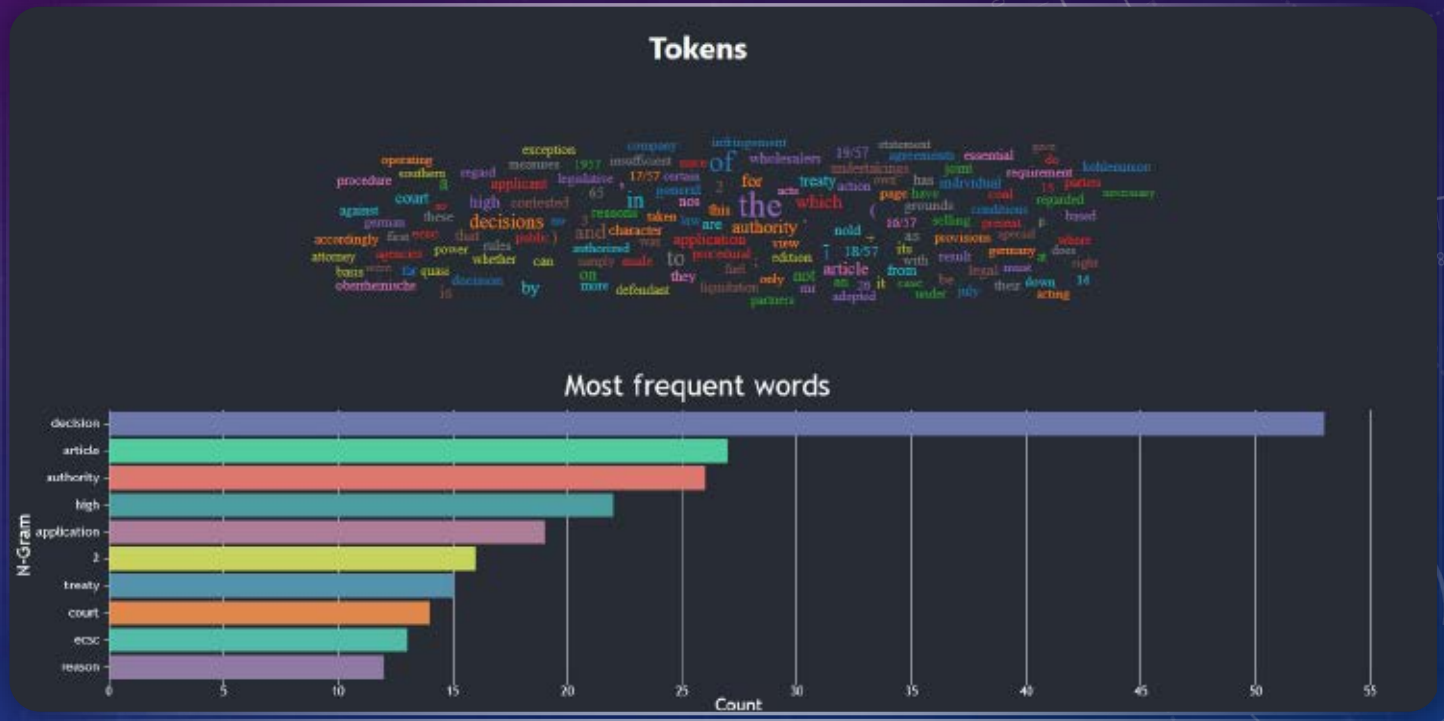
Benutzerschnittstelle und Visualisierung

- Realisierung als Web-App
- Implementierung mit JavaScript / React.js
- Abfrage von Metadaten
- Frei wählbare Analyseschritte

Fischer et al. 2020b, 17

Benutzerschnittstelle / Visualisierung

- Unterschiedliche Visualisierungsformate je nach Abfragetyp, z. B. Balkendiagramme und Wortwolken



Fischer et al. 2020b, 18

Danksagung

- Entwicklung der Workbench: Praxisseminar im Master Medieninformatik, SS 2020
  - Thomas Fischer
  - Philipp Hartl
  - Andreas Hilzenthaler
  - Lukas Jackermeier
- Fachliche Beratung: Thomas Schmidt, M. Sc., wissenschaftlicher Mitarbeiter am Lehrstuhl für Medieninformatik

Fazit/Ausblick

- Erster lauffähiger Prototyp
- Der eigentliche Analyse-/Forschungsteil kommt noch – IRIS 2022
- Weiterführung bestehender Kooperationen (Pascale Berteloot)
- Ggf. auch Nutzung im Kontext des LL. M. Legal Tech
- Kooperation mit einer neuen Professur an der Universität Regenburg zur Digitalisierung im Recht (Digital Law) geplant

Literatur

ABEGG, ANDREAS/BUBENHOFER, NOAH, Empirische Linguistik im Recht: Am Beispiel des Wandels des Staatsverständnisses im Sicherheitsrecht, öffentlichen Wirtschaftsrecht und Sozialrecht der Schweiz, Ancilla Iuris, 2016, S. 1-41.

CARLSON, KEITH/LIVERMORE, MICHAEL A/ROCKMORE, DANIEL, A quantitative analysis of writing style on the US Supreme Court, Wash. UL Rev., 2015, 93, S. 1461.

DEWILDE, BURTON, textacy Documentation, 2020, S.

EDER, MACIEJ, Rolling stylometry, Digital Scholarship in the Humanities, 2016, 31, S. 457-69.

FISCHER, THOMAS/HARTL, PHILIPP /HILZENTHALER, ANDREAS/JACKERMEIER, LUKAS (2020a), Aufbau und Analyse eines Korpus mit Entscheidungen des europäischen Gerichtshofs. Dokumentation, Lehrstuhl für Medieninformatik. Regensburg: Universität Regensburg.

FISCHER, THOMAS/HARTL, PHILIPP /HILZENTHALER, ANDREAS/JACKERMEIER, LUKAS (2020), Aufbau und Analyse eines Korpus mit Entscheidungen des europäischen Gerichtshofs. Präsentationsmaterial , Lehrstuhl für Medieninformatik. Regensburg: Universität Regensburg.

FLORESCU, CORINA/CARAGEA, CORNELIA (2017), Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1105-15.*

Literatur

LIVERMORE, MICHAEL A/RIDDELL, ALLEN B/ROCKMORE, DANIEL N, The Supreme Court and the judicial genre, Ariz. L. Rev., 2017, 59, S. 837.

MOURITSEN, STEPHEN C, The dictionary is not a fortress: Definitional fallacies and a corpus-based approach to plain meaning, Brigham Young University Law Review 2010, S. 1915-80.

PIOTROWSKI, MICHAEL (2013), 'Computerlinguistik und Digital Humanities', DHd - Digital Humanities im deutschsprachigen Raum. https://dhd-blog.org/?p=2532.

VARSAVA, NINA, Elements of Judicial Style: A Quantitative Guide to Neil Gorsuch's Opinion Writing, NYUL Rev. Online, 2018, 93, S. 75-112.

VOGEL, FRIEDEMANN /BÄUMER, BENJAMIN/DEUS, FABIAN /RÜDIGER, JAN OLIVER/TRIPPS, FELIX, Die Bedeutung des Adjektivs geschäftsmäßig im juristischen Fach- und massenmedialen Gemeinsprachgebrauch, LeGes, 2019, 30, S.

VOGEL, FRIEDEMANN/HAMANN, HANJO/GAUER, ISABELLE, Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies, Law & Social Inquiry, 2018, 43, S. 1340-63.