

**Legal Knowledge Graphs –
Extraktion dogmatischen Wissens aus
Gesetzeskommentaren mit der Hilfe von
Large Language Models (LLMs)**

Exposé zur Bachelorarbeit

Elias Gürlich

Erstgutachter:
Prof. Dr. Ulf Leser

Zweitgutachter:
tba

INHALTSÜBERSICHT

A. Einführung	1
B. Hintergrund	1
I. Juristische Methodenlehre	1
1. Deklarativer Charakter von Rechtsnormen	1
2. Gesetzesauslegung und Rechtsfortbildung	2
3. Dogmatik als geronnene Methodik	2
II. Knowledge Graphs	2
III. Large Language Models (LLMs)	2
C. Forschungsfrage	3
D. Verwandte Arbeiten	3
E. Vorgehen	4
F. Evaluation	4

A. EINFÜHRUNG

Ein Rechtsstaat monopolisiert die Ausübung von Gewalt und gewährt die Durchsetzung subjektiver Rechte durch seine Institutionen.¹ Übersteigen die Kosten deren Inanspruchnahme jedoch den erwarteten Nutzen, entwickeln Rechtssuchende ein „rationales Desinteresse“. Die Rechtswirklichkeit bleibt dann insoweit hinter dem Anspruch des Einzelnen auf staatliche Justizgewähr zurück.²

Um die Kosten der Rechtsverfolgung zu senken, werden in der juristischen Literatur verschiedene Ansätze diskutiert, um die Rechtsanwendung ganz oder teilweise zu automatisieren.³ Eine besondere Herausforderung unter rechtsstaatlichen Gesichtspunkten ist dabei die Nachvollziehbarkeit der Lösung.⁴ Wie kann das materiellrechtliche Wissen, das der Rechtsanwender zur Lösung eines Falls benötigt, in eine formale Repräsentation als „Backbone“ für die automatisierte Begründung einer rechtlichen Entscheidung überführt werden?

In dieser Arbeit soll die Modellierung rechtlichen Wissens in der Form eines sog. Knowledge Graphs untersucht werden. Es soll ermittelt werden, wie akkurat die in Gesetzeskommentaren in natürlicher Sprache beschriebenen dogmatischen Strukturen mit der Hilfe generativer Sprachmodelle, sog. Large Language Models (LLMs), automatisch in solche formalen Repräsentationen überführt werden können.

B. HINTERGRUND

I. Juristische Methodenlehre

Im Folgenden wird das methodische Vorgehen skizziert, das die demokratische Legitimation (Art. 20 Abs. 1 und 2 GG), die Vorhersehbarkeit für die Rechtsunterworfenen und die Gesetzesbindung des Richters (Art. 20 Abs. 3 GG) bei der Rechtsanwendung gewährleistet.⁵

1. Deklarativer Charakter von Rechtsnormen

Rechtliche Normen setzen sich in aller Regel aus einem Tatbestand und einer Rechtsfolge zusammen. Sind die einzelnen Merkmale des Tatbestands einer Norm für einen Sachverhalt gegeben, so folgt hieraus, dass auch die Rechtsfolge in diesem Fall wahr sein muss.⁶ Hierbei kann sich das Vorliegen eines Tatbestandsmerkmals daraus ergeben, dass dies wiederum als Rechtsfolge einer anderer Norm folgt⁷ oder daraus, dass sich das Merkmal im Wege der Subsumtion zu einer Tatsache konkretisieren lässt, die im gegebenen Sachverhalt enthalten ist.⁸

¹Rosenberg/Schwab/Gottwald, § 1 Rn. 12.

²Hähnchen/Schrader/Weiler/Wischmeyer, JuS 2020, 625 (633).

³Für einen Überblick über verschiedene Gestaltungsoptionen: Rühl, JZ 2020, 809 (814 ff.).

⁴Hierzu grundlegend: Nink, 122 ff., 334 ff.

⁵Rüthers/Fischer/Birk, § 20 Rn. 649, 654.

⁶Ebd., § 4 Rn. 115.

⁷Ebd., § 4 Rn. 131 ff.

⁸Ebd., § 21 Rn. 677.

2. Gesetzesauslegung und Rechtsfortbildung

Jedoch muss die (nicht eindeutige) normative Struktur des Gesetzes zunächst durch Auslegung am konkreten Fall ausgehend von und in den Grenzen des Gesetzeswortlauts anhand von Systematik, Telos (Sinn und Zweck) sowie Norm- und Gesetzgebungshistorie erschlossen werden.⁹ Dieser Arbeitsschritt entspricht der Auslegung des Gesetzes. Bleibt ein Sachverhalt unregelt, der einem Lebensbereich zufällt, das durch ein Gesetz abschließend geregelt werden soll,¹⁰ und ist diese Regelungslücke „planwidrig“, muss diese vom Richter durch Fortbildung des Rechts geschlossen werden.¹¹ Dies geschieht etwa durch Analogieschlüsse¹² oder durch teleologische Reduktion.¹³

3. Dogmatik als geronnene Methodik

Durch die systematisch-wissenschaftliche Durchdringung des geltenden Rechts (Dogmatik)¹⁴ auf die oben beschriebene Weise tritt eine Stabilisierung der rechtlichen Kategorien ein, was eine zügigere und einheitlichere Rechtsanwendung erlaubt.¹⁵ Einen wesentlichen Teil des Schrifttums, das dieses dogmatische Wissen zusammenträgt und anreichert, machen Gesetzeskommentare aus.

II. Knowledge Graphs

Der Begriff des Knowledge Graphs bzw. des „semantisches Netzes“ wird in der Literatur nicht einheitlich verwendet.¹⁶ Einigkeit besteht darüber, dass ein Knowledge Graph über Entitäten als Knoten und Beziehungen zwischen den Entitäten (z.B. „A ist Mitarbeiter bei B“) als Kanten verfügt. Durch die Erweiterung mittels sog. „description logics“ (DL) lassen sich außerdem Integritätsbedingungen auf Attributen, Kardinalitäten und Typen von Beziehungen oder Aussagen über Klassen im Allgemeinen darstellen.¹⁷

III. Large Language Models (LLMs)

Unter Large Language Models (LLMs) versteht man eine Klasse von Sprachmodellen¹⁸, die eine Sequenz vektorisierter Tokens als Eingabe nehmen und diese auf eine Wahrscheinlichkeitsverteilung über das nächste Token abbilden.¹⁹ Durch die gezielte Gestaltung von Anfragen an LLMs (Prompt Engineering) eröffnet sich für die Modelle ein vielseitiger Anwendungsbereich. Hierbei spricht man von Zero-Shot

⁹So auch BVerfGE 11, 126 (Rn. 18).

¹⁰Larenz/Canaris, S. 187.

¹¹Für eine differenzierte Betrachtung der Regelungslücken: Zippelius/Würtenberger, 53 ff.

¹²Ebd., 56 ff.

¹³Rüthers/Fischer/Birk, § 23 Rn. 902 ff.

¹⁴Vgl. ebd., § 23 Rn. 321.

¹⁵Ebd., § 23 Rn. 322.

¹⁶Ein schneller Überblick findet sich bei: Ehrlinger/Wöß, S. 2.

¹⁷Leser/Naumann, S. 282.

¹⁸Zum charakteristischen Attention-Mechanismus: Vaswani/Shazeer/Parmar/Uszkoreit/Jones/Gomez/Kaiser/Polosukhin, 3 ff.

¹⁹Für GPT als prototypisches LLM: Radford/Narasimhan, S. 3.

Prompting, wenn die Anweisung sich auf Aufgaben bezieht, für die dem Modell keine Beispiele während des Trainings vorgegeben wurden. Werden mit der Anweisung einige Beispiele überliefert, so spricht man von Few-Shot Prompting. Unter Fine-Tuning versteht man die Anpassung der Modellparameter zur Optimierung für ein spezifisches Problem.²⁰

Bei der Inklusion externen Wissens als Kontext in eine Anfrage²¹ stellt sich das Problem, dass LLMs in Bezug auf faktisches Wissen zur „Halluzination“ neigen, was insbesondere für die Nachvollziehbarkeit bei rechtlichen Anwendungen problematisch ist.²²

C. FORSCHUNGSFRAGE

Diese Arbeit setzt sich mit der Frage auseinander, mit welcher Qualität sich eine vorab eingegrenzte Menge von ungesehenen Kommentierungen unterschiedlicher Normen des selben Gesetzeskommentares mit der Hilfe von LLMs in einen möglichst vollständigen und vordefinierten Integritätsbedingungen genügenden Knowledge Graph überführen lassen.

Die Integritätsbedingungen sind hierbei so zu wählen, dass – ausgehend von dem Knoten, der die kommentierte Norm widerspiegelt – nur Kanten der Art „ist Voraussetzung von“ und Tatbestandsmerkmale als Entitäten zulässig sind. In Abhängigkeit von den zeitlichen Möglichkeiten werden optional noch Kanten der Art „ist Alternative von“ und „ist Rechtsfolge von“ untersucht.

D. VERWANDTE ARBEITEN

Die Literatur zur formalen Repräsentation von rechtlichem Wissen ist im deutschsprachigen Raum überschaubar.²³ Die Rechtsinformatik der Achtzigerjahre hat versucht, rechtliches Wissen als logische Programme zu formalisieren.²⁴ In Abgrenzung dazu sollen in dieser Arbeit keine Regeln extrahiert werden, aufgrund derer neues Wissen interferiert werden könnte. Ein Knowledge Graph kann aber als Vokabular für solche Regeln dienen.

Unter dem Stichwort der „Juristischen Netzwerkforschung“ werden quantitative und graphentheoretische Ansätze in der Rechtswissenschaft diskutiert. Diese betreffen jedoch zumeist als Erkenntnisperspektive nicht wie hier die Rechtsdogmatik, sondern untersuchen vorrangig die Metadaten juristischer Informationsquellen (z.B. Konstruktion von Zitiergraphen).²⁵

Es existieren auch Arbeiten, die die prinzipielle Eignung von LLMs zur Konstruk-

²⁰Eine genauere Abgrenzung der drei Ansätze m.w.N. bei: *Brown/Mann/Ryder*, S. 3.

²¹Ein Beispiel findet sich bei *Guo/Cheng/Wang/Li/Liu*, S. 2.

²²*Ji/Lee/Frieske/Yu/Su/Xu/Ishii/Bang/Dai/Madotto/Fung*, S. 3.

²³Hierzu anschaulich: *Zippelius/Würtenberger*, 89 ff. m.w.N.

²⁴*Grundmann*, in: in *Erdmann / Fiedler / Haft / Traummüller*, Expertensysteme, 97 (98): „Prolog ist eine in ihrer Struktur der juristischen Denkweise besonders nahekommende Programmiersprache“.

²⁵Grundlegend hierzu: *Coupette*, 61 ff.

tion von Knowledge Graphs nahelegen.²⁶ Die dabei untersuchten Fragestellungen sind allgemeiner gefasst als die in dieser Arbeit untersuchte (s.o., C.), denn sie sind überwiegend nicht domänenspezifisch und beschäftigen sich nicht mit der deutschen Rechtsdogmatik.

E. VORGEHEN

Zunächst sind Integritätsbedingungen als DL zu definieren, die die Struktur des zu modellierenden Wissens (s.o., C.) widerspiegeln. Es werden dann zwei ausgewählte Pipeline-Architekturen für die Extraktion der Graphen getestet. Bei jedem Ansatz ist mit Few-Shot Learning und Fine-Tuning zu experimentieren.

Der erste „naive“ Ansatz besteht darin, die Kommentierungen zunächst in Abschnitte gleicher Länge zu segmentieren und diese jeweils an ein LLM zu übergeben, um sie unmittelbar in Knoten und Kanten zu überführen. Die Vereinigung dieser Teilgraphen ergibt dann den finalen Knowledge Graph.

Der zweite Ansatz setzt eine Question-Answering-Pipeline voraus, die mittels eines Retrievers (BM25, TF-IDF, etc.) geeignete Dokumentenabschnitte sucht und hieraus mittels eines LLM-Readers Antworten synthetisiert. Diese Pipeline wird verwendet, um ausgehend vom Knoten, der die kommentierte Norm repräsentiert, weitere Kanten sukzessive durch vorgefertigte Template-Fragen zu extrahieren. Dies wird rekursiv für die neu entstandenen Knoten fortgeführt.

Optional wird noch ein dritter Ansatz untersucht, der das LLM als strategisch agierenden „Agenten“ verwendet, welcher lediglich den Retriever als „Werkzeug“ erhält und damit selbständig im Kommentartext suchen kann.²⁷

F. EVALUATION

Zur Bewertung der verschiedenen Ansätze werden einige Kommentierungen manuell in Knowledge Graphs überführt, um Testdaten zu erhalten. Es wird dann für jeden Ansatz der Recall, also der Anteil der Knoten und Kanten in den Testdaten, die auch durch die LLMs extrahiert wurden, berechnet.²⁸

Ferner wird die Precision als Anteil der Knoten und Kanten in den durch die LLMs extrahierten Graphen, die auch in den Testdaten zu finden sind, berechnet. Für darüber hinausgehende Kanten, die extrahiert wurden, ist zu evaluieren, ob diese sinnvolle Erweiterungen der Testdaten oder bloße Halluzinationen darstellen.

²⁶Eine Zusammenfassung m.w.N. findet sich bei: *Carta/Giuliani/Piano/Podda/Pompianu/Tiddia*, S. 3.

²⁷Wie LLMs andere Modelle durch geschicktes Prompting planvoll als „Werkzeuge“ einsetzen können, zeigen: *Shen/Song/Tan/Li/Lu/Zhuang*, 2 f.

²⁸Grundlegend zur Abgrenzung von Recall und Precision: *Powers*, S. 2.

LITERATURVERZEICHNIS

- Brown, Tom / Mann, Benjamin / Ryder, Nick*: Language Models are Few-Shot Learners, 2020.
- Carta, Salvatore / Giuliani, Alessandro / Piano, Leonardo / Podda, Alessandro Sebastian / Pompianu, Livio / Tiddia, Sandro Gabriele*: Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction, 2023.
- Coupette, Corinna*: Juristische Netzwerkforschung, 1. Aufl., Tübingen 2019.
- Ehrlinger, Lisa / Wöß, Wolfram*: Towards a Definition of Knowledge Graphs, 2016.
- Grundmann, Stefan*: Juristische Expertensysteme – Brücke von (Rechts)informatik zu Rechtstheorie, in: *Erdmann, Ulrich / Fiedler, Herbert / Haft, Fritjof / Traunmüller, Roland* (Hrsg.), Computergestützte juristische Expertensysteme (Neue Methoden im Recht Bd. 1), Tübingen 1986, S. 97 (zitiert: *Grundmann* in *Erdmann / Fiedler / Haft / Traunmüller, Expertensysteme*).
- Guo, Zhicheng / Cheng, Sijie / Wang, Yile / Li, Peng / Liu, Yang*: Prompt-Guided Retrieval Augmentation for Non-Knowledge-Intensive Tasks, 2023.
- Hähnchen, Susanne / Schrader, Paul T. / Weiler, Frank / Wischmeyer, Jun-Thomas*: Legal Tech, JuS 2020, S. 625.
- Ji, Ziwei / Lee, Nayeon / Frieske, Rita / Yu, Tiezheng / Su, Dan / Xu, Yan / Ishii, Etsuko / Bang, Yejin / Dai, Wenliang / Madotto, Andrea / Fung, Pascale*: Survey of Hallucination in Natural Language Generation, 2023.
- Larenz, Karl / Canaris, Claus-Wilhelm*: Methodenlehre der Rechtswissenschaft, Berlin, Heidelberg 1995.
- Leser, Ulf / Naumann, Felix*: Informationsintegration, Heidelberg 2006.
- Nink, David*: Justiz und Algorithmen: Über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung, Berlin 2021.
- Powers, David M. W.*: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, 2020.
- Radford, Alec / Narasimhan, Karthik*: Improving Language Understanding by Generative Pre-Training, 2018.
- Rosenberg, Leo / Schwab, Karl-Heinz / Gottwald, Peter*: Zivilprozessrecht, 18. Aufl., München 2018.
- Rühl, Giesela*: Digitale Justiz, oder: Zivilverfahren für das 21. Jahrhundert, JZ 2020, S. 809.
- Rüthers, Bernd / Fischer, Christian / Birk, Axel*: Rechtstheorie und juristische Methodenlehre, 11. Aufl., München 2020.
- Shen, Yongliang / Song, Kaitao / Tan, Xu / Li, Dongsheng / Lu, Weiming / Zhuang, Yueting*: HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face, 2023.
- Vaswani, Ashish / Shazeer, Noam / Parmar, Niki / Uszkoreit, Jakob / Jones, Llion / Gomez, Aidan N. / Kaiser, Lukasz / Polosukhin, Illia*: Attention Is All You Need, 2017.

Zippelius, Reinhold / Würtenberger, Thomas: Juristische Methodenlehre, Bd. 93
(JuS-Schriftenreihe Studium), München 2021.