Collaborative System Synergizing Human Expertise and Large-scale Language Models for Legal Knowledge Graph Construction*

Ryoma Kondo[1,2,*], Tomohiro Watanabe[1], Takahiro Yoshida[2], Kazuyuki Yamasawa[3] and Ryohei Hisano[1,2,*]

[1]Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[2]The Canon Institute for Global Studies, ShinMarunouchi Building 5-1 Marunouchi 1-chome, Chiyoda-ku, Tokyo 100-6511 Japan

[3]TKC Corporation, Karusaka MN Building, 2-1 Agemachi, Shinjuku-ku, Tokyo 162-8585, Japan

Abstract

Court judgments are rich with information that details the combination of facts and legal norms to arrive at judicial decisions. This intricate process can be conceptualized as a hierarchical tree structure, wherein facts are aggregated in a bottom-up manner to highlight pivotal facts. These critical facts are subsequently linked to legal norms, facilitating the derivation of specific decisions. Despite the intrinsic value of this information, there is no legal knowledge graph(KG) that can represent this, and even if there were, the task of text-mining or text-annotation of such legal structures from court judgments presents considerable challenges. We show a legal ontology that links facts and norms and a new collaborative framework that harnesses the capabilities of both LLMs and legal experts for extracting these intricate relationships. Our approach is underpinned by a carefully designed LLM, tailored through prompt engineering, that can capture such structures. The results from this are then refined by legal experts through a graphic user interface, also providing us with an overall score of the annotation. We believe that a synergistic integration of machine efficiency and human expertise will lead to the improvement of future legal KGs and legal search engines.

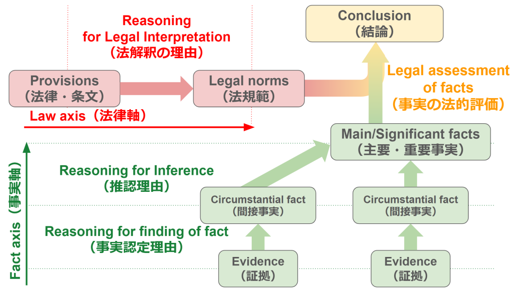
Keywords

Semantic web, legal knowledge graph, artificial intelligence and law, court judgments, large-scale language models, annotation tools

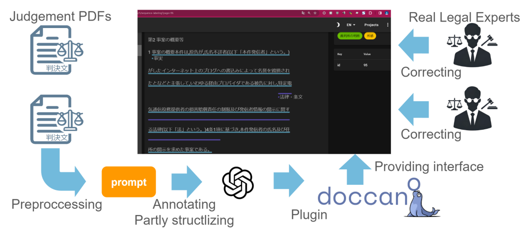
1. Introduction

Court judgments at the district level offer a wealth of information that merges factual evidence and legal principles to derive judicial decisions[1]. This process is consistent across both civil and criminal cases, where it invariably begins with the collection of evidence. As this evidence accumulates, it reveals certain detailed facts, which are subsequently leveraged to ascertain the fundamental facts of the case. These fundamental facts are pivotal, as they are the elements that the court can directly relate to applicable laws. Thus, these critical facts are subsequently

Posts, Demos, and Industry Tracks at ISWC 2024, November 13–15, 2024, Baltimore, USA
*Corresponding author.
kondor@g.ecc.u-tokyo.ac.jp (R. Kondo); hisanor@g.ecc.u-tokyo.ac.jp (R. Hisano)



(a) Logical structure in court judgments



(b) Collaborative system between large-scale language models and legal experts

Figure 1: Illustrations of the logical structure in court judgments and the collaborative system

linked to legal norms, facilitating the derivation of specific decisions. A diagram illustrating this process is provided in Figure 1a.

Given the significance of structured analysis, the task of extracting information from court judgments presents considerable challenges. Prior research in large-scale language models (LLMs), including projects like LegalBench[2] also emphasizes creating annotations for legal reasoning tasks. However, these efforts predominantly focus on straightforward labeling and fall short in structuring information through the precise definition of ontologies. Existing works on law and knowledge graphs(KGs) are limited to organizing legal citations in the judgment text[3], as well as linking text and law by citation relations[4]. Common ontologies like *schema.org* and *DBpedia* lack the resolution to represent phrase-wise entities and links in judgment texts. Although section-wise parsing is shown for organizing knowledge in academic papers[5], there is still no application to judgments or in a phrase-wise manner. Existing legal ontologies/KGs focus on organizing documents and the laws themselves[6, 7], but they do not achieve the linkage between factual phrases and legal phrases in actual judgments.

The development of KGs based on the legal ontologies is crucial. This addresses a critical gap in large-scale language models (LLMs), which often overlook relevant laws for specific cases and struggle with legal reasoning and analogies [8]-a gap that could be bridged with a well-structured KG[9]. This limitation is not new and echoes issues identified in earlier studies employing deep learning technologies in legal contexts[10]. Conversely, while human legal experts can realize such structured information, their involvement comes with significant costs.

Hence, we present a collaborative system synergizing human legal experts and LLMs to construct legal KGs. We start by defining an ontology that encapsulates the hierarchical reasoning structure (Figure 1a), blending factual aspects and legal principles. Using this ontology, alongside carefully crafted prompt engineering, we demonstrate how LLMs can effectively extract such information from written court judgments. Subsequently, our system displays these LLM-generated annotations through a graphical user interface, enabling legal experts to review and refine them. This process also allows for a direct comparison between the two annotations, facilitating a quantitative analysis of their agreement (Figure 1b).

2. Legal Ontology and Collaboration Between LLMs and Experts

We begin by establishing an ontology that reflects the hierarchical reasoning structure of court judgments (Figure 1a). This framework organizes arguments from plaintiffs, defendants, and courts. A court judgment includes attributes such as legal provisions, facts, legal norms, interpretations, decisions, and impressions. Legal provisions are specific references like Article 709 of the Civil Code on compensation for damages. Facts are court-recognized realities, while legal norms are principles derived from provisions. Interpretations apply provisions or norms to the facts. The judgment includes the court's conclusion, rationale, and legal assessment of the facts, with impressions covering opinions not directly tied to the logical structure. Using this ontology, we build a legal KG by extracting text segments corresponding to these attributes, similar to named entity recognition[12].

In legal applications, LLMs have shown that they can pass a bar exam in the US. However, they often miss important legal details, limiting their use mainly to drafting documents[8]. Their effectiveness also varies by legal field and task[11]. It is still unclear how well these models understand Japanese court judgments. Therefore, it is important to design a system that allows legal experts to guide these models to build better legal KGs.

Figure 1b depicts a schematic of our proposed collaborative system between LLMs and legal experts. Judgment documents are converted to text, segmented, and processed through the language model in smaller sections due to input character limits. The outputs are converted to JSON, and pattern matching identifies text segments, which are then linked to Doccano[1] for legal experts to review and correct.

Prompt engineering represents a crucial practice for using LLMs, notably in the context of expansive models such as GPT-4. This technique entails the strategic crafting of input prompts to effectively guide these models toward generating accurate and relevant outputs. Its significance is amplified in complex tasks requiring nuanced interpretation, particularly in specialized fields like legal analysis.

For legal experts, correcting false positives is easier than identifying important but undiscovered elements in the text. Therefore, we designed a prompt specifically for analyzing segments of civil court judgments, emphasizing the minimization of omissions over the reduction of false positives. The prompt directs GPT-4 to output sections corresponding to predefined labels: legal provisions, facts, legal norms, interpretation, judicial decisions, and impressions. These labels were selected to capture key elements of legal texts. The prompt ensures that GPT-4 outputs these elements without altering the original text or creating new content, which is essential for preventing hallucinations and maintaining fidelity to the source material. Full prompt engineering details are available in our repository[2].

3. Evaluation

To evaluate our system, we analyzed the modification history by legal experts on 11 representative civil court judgments from the district level, selected from 100 judgments provided

[1]https://github.com/doccano/doccano
[2]https://github.com/rkondo3/Collaborative_System_for_Legal_Knowledge_Graph_Construction

by the Supreme Court of Japan[3]. These texts cover key legal themes like trademark infringement, copyright law, defamation, labor-related injuries, land boundary disputes, and bona fide acquisition in negotiable instruments.

Our evaluation compared GPT-4 with legal experts on extracting relevant portions from court judgments, similar to named entity recognition (NER). We focused on phrase-level comparison rather than character-level accuracy, as it better captures meaningful differences in legal texts. Character-level accuracy can overly penalize minor discrepancies, while missing key points is more critical for understanding legal documents.

Table 1 demonstrates that GPT-4 can assist in labeling texts from Japanese civil district court cases, streamlining the annotation process for legal experts and reducing the time required compared to manual methods. The recall (Rec) score is 0.80, meaning that the LLM is good at finding the right points that legal experts consider important. The precision rate (Pre) is 0.73, showing that it is quite accurate, but the LLM is better at finding cases than always being right about them. Because missing important points in legal work can be a bigger problem, we believe that our prompt design, which allows the recall to be higher, is effective. Moreover, our collaborative system synergizing human legal experts and LLMs helps to capture essential details in legal documents, saving both time and effort to create annotations.

In line with our prompt design, the results show that the recall exceeds the precision across legal provisions, facts, legal norms, and legal interpretations-the four foundational and critical elements this study examined for connecting laws and facts. We also examined inaccuracies to assess GPT-4's characteristics and its potential within collaborative frameworks. Further details can be found in our repository.

Table 1
Total evaluation scores

Element	TP	FP	FN	TN	Pre	Rec
Provisions	52	8	6	64	0.87	**0.89**
Legal norms	10	18	4	29	0.36	**0.71**
Legal interpretation	7	10	0	16	0.41	**1.0**
Facts	152	43	22	195	0.77	**0.87**
Judicial decisions	51	22	38	99	**0.69**	0.57
Total	272	101	70	403	0.73	**0.80**

4. Conclusion

We have developed legal ontologies that encapsulate the hierarchical reasoning structure connecting the factual and legal aspects of a case. To efficiently construct this framework, we developed a collaborative model involving both LLMs and legal experts. Our evaluation indicated that LLMs serve as competent annotators, and this can lower the costs associated with annotations by experts[4].

[3]https://www.courts.go.jp/app/hanrei_jp/search1
[4]We thank TKC Corporation for helpful discussions concerning this work. R.K. is funded by JST, ACT-X Grant Number JPMJAX23CA, Japan. R.H. is funded by the JST, FOREST Program JPMJFR216Q, Japan, the UTEC-UTokyo

References

[1] S. Ota, "Creation of Law and Emergence of Social Norms," *Journal of the Japanese Society for Artificial Intelligence*, vol. 18, no. 4, pp. 408–415, in Japanese, Jul. 2003. urlhttps://-doi.org/10.11517/jjsai.18.4_408

[2] N. Guha, et al., "Legalbench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models," *arXiv preprint*, Aug. 2023. urlhttps://arxiv.org/abs/2308.11462

[3] N. Sakhaee, et al., "Information extraction framework to build legislation network," *Artificial Intelligence and Law*, Springer, vol. 29, pp. 35–58, Mar. 2021. urlhttps://-doi.org/10.1007/s10506-020-09263-3

[4] A. Sadeghian, et al., "Automatic semantic edge labeling over legal citation graphs," *Artificial Intelligence and Law*, Springer, vol. 26, pp. 127–144, Mar. 2018. urlhttps://-doi.org/10.1007/s10506-018-9217-1

[5] B. Zhang, et al., "ASKG: An Approach to Enrich Scholarly Knowledge Graphs through Paper Decomposition with Deep Learning," *ISWC (Posters/Demos/Industry)*, Nov. 2023.

[6] L.A. Avgerinos, et al., "Legal informatics from the aspect of interoperability: A review of systems, tools and ontologies," *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance.*, Association for Computing Machinery, pp. 731–-737, Oct. 2020. urlhttps://doi.org/10.1145/3428502.3428611

[7] G. Cristine, et al., "Legal powers, subjections, disabilities, and immunities: Ontological analysis and modeling patterns," *Data & Knowledge Engineering*, Elsevier, vol. 148, pp. 102219, Sep. 2023. urlhttps://doi.org/10.1016/j.datak.2023.102219

[8] A. Trozze, et al., "Large Language Models in Cryptocurrency Securities Cases: Can Chatgpt Replace Lawyers?" *arXiv preprint*, Nov. 2023. https://arxiv.org/abs/2308.06032

[9] M. Michalis, et al., "Real-Time Validation of ChatGPT Facts using RDF Knowledge Graphs," *ISWC (Posters/Demos/Industry)*, Nov. 2023.

[10] R. Kondo, et al., "Masked Prediction and Interdependence Network of the Law Using Data from Large-scale Japanese Court Judgments," *Artificial Intelligence and Law*, Springer, vol. 31, No. 4, pp. 739–771, Oct. 2022. https://doi.org/10.1007/s10506-022-09336-5

[11] D. Katz, et al., "GPT-4 Passes the Bar Exam," *SSRN preprint*, Apr. 2023. http://dx.doi.org/10.2139/ssrn.4389233

[12] L. Jing, et al., "A Survey on Deep Learning for Named Entity Recognition," *IEEE Transactions on Knowledge and Data Engineering*, IEEE, Jan. 2020. https://doi.org/10.1109/TKDE.2020.2981314

FSI Research Grant Program, and JSPS KAKENHI Grant Number JP24K03043.