LegalSearchLM: A Generative Legal Language Model for Legal Case Retrieval

Anonymous ACL submission

Abstract

Legal Case Retrieval (LCR), which retrieves relevant cases from a query case, is a fundamental task for legal professionals in legal research and decision-making. Previous studies have focused on lexical matching or embedding-based retrieval methods, which often fail to capture detailed legal factors from complex cases. In this paper, we introduce a benchmark and a novel retrieval approach: (1) LEGAR BENCH, the first Korean LCR benchmark covering the widest range of criminal case types, supporting two dataset versions based on different relevance criteria; (2) LegalSearchLM, a generative retrieval model that can generates key legal elements from query cases with complex legal conditions through entry point-aware identifier generation. Our experiments on LEGAR BENCH show that our LegalSearchLM outperforms the most powerful baseline by 17%, achieving state-of-the-art results. It also demonstrates remarkable out-of-domain performance across diverse criminal cases.

1 Introduction

Legal AI has increasingly gained attention from legal professionals (e.g., lawyers and judges) to raise the productivity of their work. Among various legal applications, *Legal Case Retrieval* (LCR) (Feng et al., 2024; Deng et al., 2024b; Su et al., 2024; Deng et al., 2024a; Li et al., 2023a; Xiao et al., 2021; Li et al., 2023b,c; Zhang et al., 2023), which identifies relevant precedents for a given case, plays a particularly crucial role in maintaining judicial fairness and supporting the decision-making process of legal experts.

While research on LCR has expanded in recent years, existing studies rely on lexical matching or embedding similarity search, often resulting in imprudent matches or failing to capture subtle distinctions from a legal perspective (Magesh et al., 2024). They do not consider document identifiers that meet legal elements required from specific criminal case. This underscores the urgent need for more advanced retrieval methods that can handle complex legal details. To address these limitations, we introduce a new retrieval approach.

First, we present LEGAR BENCH (**Lega**l Case **R**etrieval **bench**mark), the first Korean LCR benchmark that covers a broad range of crime types with relevance criteria rigorously defined by legal experts. LEGAR BENCH comprises two dataset versions tailored to different evaluation needs: (1) LEGAR BENCH$_{Standard}$, designed for a comprehensive assessment of most crime categories. It constructs relevant cases based on charge titles and statutory provisions using a top-down approach, consisting of 411 similar case groups across 33 crime categories. (2) LEGAR BENCH$_{Stricter}$, which imposes higher relevance requirements by ensuring identical factual details and legal disputes, aligning with expert-level expectations. It includes 160 similar case groups across 8 categories. Both datasets support flexible query selection from similar case groups, enabling scalable evaluation (See Section 2).

Second, we propose LegalSearchLM, a generative retrieval model designed to capture key legal factors from a query case as document identifiers through entry point-aware decoding. While content generation retrieval approach [1] has shown promise and practical (Kim et al., 2024; Li et al., 2023d,e; Bevilacqua et al., 2022), applying it to the legal domain presents two challenges: (1) Legal relevance is inherently tied to domain knowledge, making it difficult to obtain sufficient case pairs, especially in long-tail or underexplored areas. (2) Naive document identifiers often fail to probabilistically model relevance, leading to deviations during constrained decoding. To address the first challenge,

[1]This approach utilizes an autoregressive generation model for retrieval, where decoding is conditioned on a database to generate document content as document identifiers.

	Language	Crime types of query	Query case	Retrieval pool	Target case per query
COLIEE2024	English	-	400	1,734 (per query)	-
LeCaRD	Chinese	-	107	100 (per query)	10.33
LeCaRDv2	Chinese	-	800	55,192 (per query)	20.89
LEGAR BENCH$_{Standard}$	Korean	450	450*N	1,226,814	200
LEGAR BENCH$_{Stricter}$	Korean	160	160*N	169,230	14.69

Table 1: Comparison of the scale of the LCR Benchmark.

we introduce a self-supervised fine-tuning (SSFT) retrieval approach optimized for the LCR task, alleviating the need of expensive query-target case pairs required to train standard case retrievers. For the second, we propose a first token-aware generation method during constrained decoding, enabling the successful generation of subsequent identifiers from the database (See Section 3).

We evaluate LegalSearchLM on LEGAR BENCH, comparing it with competitive baselines, including sparse lexical matching and embedding similarity search from both general and legal domains. In Section 4.2, the results show that LegalSearchLM outperforms the top retrieval baseline by 17% in precision and matches the performance of a reranked model using passage-level multiple inference. Additionally, it demonstrates strong generalization on out-of-domain criminal types, with a 16% improvement over generative retrieval with naive identifiers on in-domain cases. (See Section 5).

In summary, our contributions are as follows.

- We introduce the first Korean LCR benchmark, **LEGAR BENCH**, which covers the widest range of criminal cases and is built with rigorously defined relevance criteria.
- We present a specialized legal expert retrieval model, **LegalSearchLM**, which generates optimal document identifiers to precisely match individual legal factors through entry token-aware generation.
- Our LegalSearchLM achieves *state-of-the-art performance* on LEGAL BENCH, capturing legal semantic nuances and demonstrating remarkable *generalization ability*.

2 LEGAR BENCH

LEGAR BENCH features the most comprehensive set of query and target cases (See Table 1), with rigorously defined relevance criteria for lawyers involved in the construction process. It offers two dataset versions based on different evaluation needs. In this section, we share relevance criteria and the construction process of LEGAR BENCH$_{Standard}$ and LEGAR BENCH$_{Stricter}$ in Sections 2.1 and 2.2, respectively.

2.1 LEGAR BENCH$_{Standard}$

LEGAR BENCH$_{Standard}$ is designed to provide a comprehensive assessment of most crime categories, enabling accurate identification and improvement of retrieval failure points.

2.1.1 Definition of Standard Relevance

We define standard relevance based on the charge title and statutory provision. This includes various scenarios that share identical charge titles. As shown in Figure 1 on sexual crime, for the query case **(Query)** on distributing false sexual images/videos for profit, the standard target case **(Standard)** satisfies the three statutory elements: 1. creation of false sexual images/videos, 2. intent for profit, and 3. distribution. Cases like **(A)**, which are not for profit, or **(B)**, which concern illegal sexual video filming rather than false sexual image creation, cannot be considered target cases. This is because they are distinct crimes governed by different laws.

2.1.2 Data Construction

The construction of LEGAR BENCH$_{Standard}$ begins by establishing a framework in Steps 1–3 for grouping similar cases, created through intensive collaboration between the first author and five lawyers. To ensure comprehensive coverage of the diverse and complex legal literature, we employ a top-down approach, systematically categorizing crimes based on Korean legal frameworks. An example of Steps 1–3 is illustrated in Figure 2. In Step 4, 1 million criminal cases (85.79%) out of 1.2 million are mapped to each group by all authors.

(Query) From September 2020 to March 2021, at the defendant's residence in B Building, C Unit, Seobuk District, Cheonan, the defendant created a Telegram chat room named "D." Over 95 transactions, the defendant received a total of 2,400,000 KRW in cultural gift certificates from buyers. The defendant then invited them to other chat rooms, "E," "F," and "G", which contained 101 edited photos of female celebrities, with their faces inserted into explicit content without consent. The defendant also sent 250 similar photos directly to buyers. Through these actions, the defendant sold and distributed edited, sexually suggestive images for profit, using the internet without the individuals' consent.

(Standard) On June 15, 2021, at 01:57 AM, at the defendant's residence in B Apartment, C Unit, in Gyeongbuk Gunwi, the defendant, after posting a false video sale advertisement on Telegram, was contacted by Officer E from the Jeju Police Department. The defendant accepted an offer to sell the fake videos, received 20,000 KRW via a bank transfer to an account under the defendant's name, and sent 9 edited photos of women's bodies to Officer E via Telegram. The photos were digitally altered to insert images of women's breasts or genitals. In doing so, the defendant provided edited or fabricated material for profit, using the internet to create sexually suggestive content without the consent.

(Stricter) From June 2023 to January 2024, the defendant operated a private Telegram channel from their residence in Gangseo-gu, Seoul, advertising the sale of "high-quality manipulated photos and videos." Through this platform, the defendant received payments from multiple buyers into their H Bank account and distributed approximately 320 manipulated images and videos, in which the faces of ordinary women and celebrities were superimposed onto the nude bodies of unidentified individuals. Engaging in this activity on approximately 110 occasions, the defendant accrued a total of 3,300,000 KRW. These materials were repeatedly distributed online for profit, without the victims' consent.

(A) On July 23, 2022, the defendant used a program to create 8 manipulated images by combining a photo of the victim with nude images of an unknown woman from a pornographic site. Later, on October 6, 2022, the defendant created a G account under the victim's name and posted the 8 manipulated images along with 24 other personal photos, including family photos, making them public. By doing so, the defendant illegally manipulated and distributed the victim's images against their will.

(B) On May 24, 2024, the defendant filmed the victim, E (female, 45), performing a nude massage using a hidden camera. Between February and May 2024, the defendant filmed 50 similar instances involving women, including sexual acts, without their consent. Later, on February 27, 2024, the defendant uploaded one video to a site and sold it for 300 'wood' (approximately 44,715 KRW). Between February and September 2024, the defendant sold similar videos, earning a total of 73,742,488 KRW.

Figure 1: Examples of Relevance Cases. **(Query)** is a query case on distributing false images/videos for profit. The Red Highlight indicates profit, the Yellow Highlight represents the creation of false images/videos, and the Blue Highlight denotes distribution—the three key legal elements of the crime. Both **(Standard)** and **(Stricter)** satisfy the three elements, and **(Stricter)** additionally meets the requirements concerning the scale of distributed images/videos and the total financial gains obtained. **(A)** and **(B)** are not target cases, as **(A)** distributed a false image without intending to obtain financial gains, and **(B)** committed the offense for financial gain through the unlawful filming of real footage, not the creation of false images.

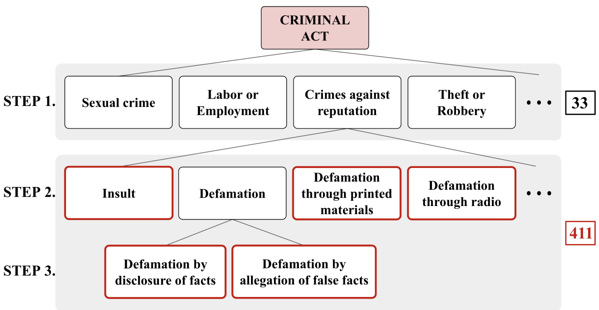

Figure 2: Examples of the construction process for each step in LEGAR BENCH$_{Standard}$.

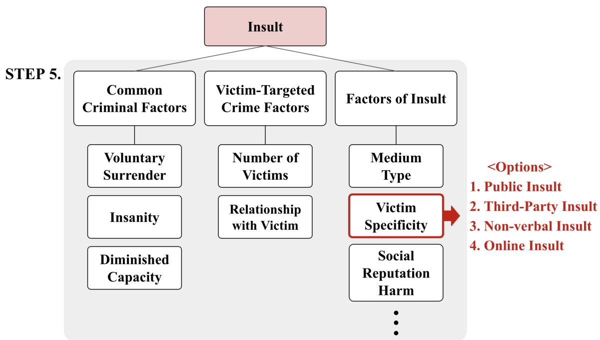

Figure 3: Examples of the construction process of LEGAR BENCH$_{Stricter}$.

Step 1: Construction of Crime Typology. We establish a crime typology to categorize various types of crimes in criminal cases. As shown in STEP 1 of Figure 2, we define major categories, including Sexual Crimes, Labor or Employment Offenses, Crimes Against Reputation, and Theft or Robbery. Appendix 4 lists a total of 33 crime categories.

Step 2: Assignment of Charge Titles. We construct specific crime charge titles that can occur within each crime category. A charge title is the official name used in legal documents, such as indictments or complaints, to describe a specific offense. As shown in Figure 2, crimes against reputation can be expanded into sub-categories, such as defamation, defamation through printed materials, defamation through radio, insults, etc., based on crime charges. While this charge title is based on a particular statute, not every statute directly translates into a single charge title. Charge titles can be further refined based on multiple statutes corresponding to them. This feature is used for further refinement in the next step.

Step 3: Refinement from Statutory Provisions. To better reflect the specific laws applied and ensure greater alignment with legal facts, we refine charge titles by categorizing them at the statute level. Figure 2 shows how defamation can be specified according to distinct laws, such as defamation by disclosure of facts and defamation by allegation of false facts. Finally, the standard similar groups are formed by combining the results of Step 2 and Step 3, as shown in the red-bordered box of Figure 2. *As a result, LEGAR BENCH$_{Standard}$ contains 411 similar groups across 33 categories.*

Step 4: Case Mapping. We automatically process 1.2 million criminal cases, mapping them to their respective groups based on the judgment title (which closely aligns with the charge title) and statutory provisions annotated for each group. This process successfully maps 1 million cases (85.79%) to the defined groups, *enabling evaluation on the majority of criminal cases through our LEGAR BENCH$_{Standard}$*.

2.2 LEGAR BENCH$_{Stricter}$

2.2.1 Definition of Stricter Relevance

For stricter case similarity, we expand the scope from facts to include claims, reasoning, sentencing factors, and conclusions sections from the case, aiming to provide a more comprehensive view of the legal context. Stricter relevance further requires *factual details* such as severity of the crime, relationship between the defendant and the victim, situational information, and arguments made by the defendants. For instance, while making only a few fake images and selling them for 20 dollars is ruled under the same crime with making hundreds of fake videos with thousands of dollars of profit, the stricter factual relevance between the two cases is low (See the pink and mint highlights in Figure 1). Also, if two assault defendants make the same claim of self-defense but only one of them was judged guilty, these two cases should also be distinguished. We hire legal experts to annotate these important factors that determine the strict relevance between cases.

2.3 Dataset Construction

Step 5: Define Detailed Factors. For the 160 similar groups across 10 crime categories in LEGAR BENCH$_{Standard}$, we construct a stricter dataset. First, we define factors to be further considered for each standard similar group, as shown in Figure 3, where Insult has specific factors such as Common Criminal Factors, Victim-Targeted Crime Factors, and Factors of Insult. Next, for each factor, we categorize sub-factors and create options for each sub-factor. Finally, based on the defined factors, sub-factors, and options for each standard group, we annotate the cases belonging to each standard group using GPT-4o.

Step 6: Case Grouping. As a result of Step 5, we obtain (sub-factor, option) pairs for each case across all sub-factors required for each standard group. The following grouping algorithm is then applied using these pairs. We created one stricter group for each standard group, resulting in a total of 160 stricter query sets (See Table 1 for details).

Algorithm 1 Stricter Relevance Group

```
Input: case_data, subfactor-option pair_list
Output: grouped_cases
for each case in case_data do
    key = generate_key(subfactor-option pair_list)
    group[key].append(case)
end for
if any group has 2 or more cases then
    return the group
end if
for r = number of subfactors to 1 do
    for each case in case_data do
        key = generate_key(subfactor-option pair_list[:r])
        group[key].append(case)
    end for
    if any group has 2 or more cases then
        return the group
    end if
end for
return None
```

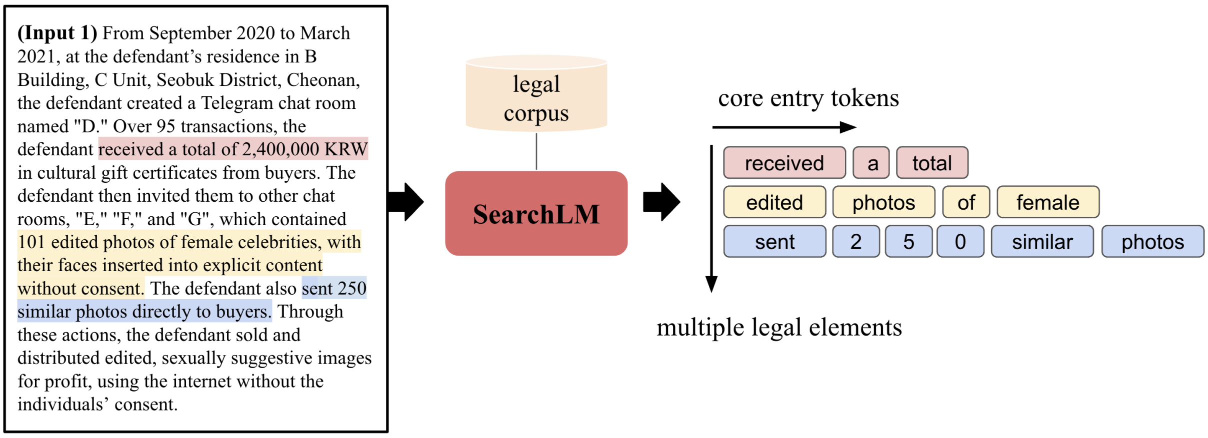

Figure 4: Approach of SearchLM. When a query case is given as input, SearchLM generates essential legal elements that start with key tokens to identify the document's identifiers.

3 LegalSearchLM

To better capture core legal elements during retrieval, we first introduce a generative retrieval approach to the LCR task, leveraging the power of language models with next-token prediction. In Section 3.1, we explain the content generation-based generative retrieval, and in Sections 3.2 and 3.3, we present our advanced approach and the specifics of our training process.

3.1 Background

LegalSearchLM is based on generative retrieval that directly generates content in documents as document identifiers (Bevilacqua et al., 2022; Li et al., 2023e,d; Kim et al., 2024). To generate content related to a query from a database, the decoding process during inference is constrained by a prefix tree-like data structure. In our work, we utilize the FM-Index (Bevilacqua et al., 2022), which is beneficial for efficiently compressing the entire content of a vast amount of documents.

Specifically, given a generated token sequence $x_1, x_2, \ldots, x_n$, the entire vocabulary V, and $\mathcal{C}(x_{<t})$ that represents the candidate token set constrained by pre-constructed FM-Index, the next token is selected only from the allowed candidate set $\mathcal{C}(x_{<t})$ at each time step $t \geqslant 2$.

$$x_1 = \arg\max_{x \in V} P(x \mid [\text{BOS}]) \quad (1)$$

$$x_t = \arg\max_{x \in \mathcal{C}(x_{<t})} P(x \mid x_{<t}), \quad \text{for } t \geqslant 2 \quad (2)$$

The final sequences serve as document identifiers, which are then aggregated to rank the documents.

3.2 Our Approach

Choice of Optimal Reference. It is difficult to construct case pairs across diverse criminal domains, as it requires exceptional precision and legal expertise. To alleviate the obstacle of data scarcity, we adopt a self-supervised fine-tuning approach. We use a query case as input, and the targets are the legal elements required for the crime described in the query case. The method for selecting the legal element targets (identifiers) is described in the next paragraph. With this approach, we can build reliable identifiers, as the LCR task is a document-to-document matching task, where all essential information is contained within the query case. By doing this, we also learn how to extract core legal factors from the query case as a reference, without relying on trained legal knowledge, which helps improve generalizability (see Section 5).

Core Entry Point-Aware Identifiers. Selecting optimal identifiers is especially crucial when handling complex documents such as precedents. To achieve this, first, we carefully select *the initial tokens of identifiers to guide meaningful subsequent token generation*. Since we use a generative model for retrieval, the informative entry token is crucial in a constrained decoding environment where the selection of previous tokens restricts the range of possible next tokens. Second, *to meet the multiple legal elements required for each query case*, we decompose complex contexts into distinct facts (Min

	LEGAR BENCH$_{Standard}$ (P@5)				LEGAR BENCH$_{Strciter}$ (P@5)			
Criminal Category	LegalSeachLM	BM25	Contriever	SAILER	LegalSeachLM	BM25	Contriever	SAILER
Fraud	0.74	0.53	0.57	0.64	0.30	0.30	0.03	0.24
Injury or Violence	0.62	0.50	0.42	0.66	0.32	0.32	0.01	0.23
Sexual crime	0.62	0.49	0.40	0.65	0.40	0.39	0.02	0.32
Finance or Insurance	0.72	0.56	0.60	0.64	0.20	0.20	0	0.2
Defamation or Insult	0.83	0.58	0.48	0.78	0.22	0.22	0	0.22
Drug	0.80	0.52	0.76	0.84	0.23	0.23	0	0.1
Murder	0.50	0.50	0.30	0.50	0.35	0.35	0	0.35
Traffic offenses	0.47	0.33	0.33	0.94	0.26	0.26	0.02	0.12
	⋮	⋮	⋮	⋮				
[Total]	**0.68**	0.51	0.48	0.62	**0.29**	0.28	0.01	0.22

Table 2: Results on LEGAR BENCH$_{Standard}$ and LEGAR BENCH$_{Strciter}$

et al., 2023; Cai et al., 2024; Chen et al., 2024; Deng et al., 2024a), as even a slight difference in a single fact (e.g., victim's age) can significantly alter the relevance between cases. Detailed selection processes are described in Section 3.3.

3.3 Training

To construct the training dataset on the above approach, we employ GPT-4o and the overall process is as below:

- First, we decompose $\mathcal{F}$, the fact description of a query case, into individual information units $f_1, f_2, \ldots, f_m \in \mathcal{F}$.
- We then construct $\mathcal{F}_{\text{relevant}}$ by eliminating irrelevant facts f_i such as event dates or region names.
- Finally, each fact in $\mathcal{F}_{\text{relevant}}$ is rephrased by positioning key tokens at the beginning, resulting in the set $\mathcal{F}_{\text{relevant, core-entry}}$.

During training, we use $\mathcal{F}$ as the input and multiple facts from $\mathcal{F}_{\text{relevant, core-entry}}$ as targets, pairing them one by one. Additionally, for generalizability, we include one fact from $\mathcal{F}_{\text{relevant, core-entry}}$ as the input and another fact from the same set as the target. Detailed information is described in Appendix AA.

4 Experiments

4.1 Baselines

We evaluate a range of baselines, including traditional lexical matching and embedding-based retrieval models on both general and legal domains.

Lexical Matching. We use BM25, a strong baseline in the legal domain (Rosa et al., 2021), widely adopted by Legal AI corporations for their RAG systems (Magesh et al., 2024).

General-Domain Dual Encoder. We select Contriever (Izacard et al., 2022), an unsupervised dual encoder model, as it aligns with our no-supervision approach and is widely used in general-purpose retrieval.

Legal-Domain Dual Encoder. We use SAILER (Li et al., 2023a), which achieves strong performance in the LCR task of the COLIEE 2023 competition, pretrains legal documents by assigning training loss at the section level (e.g., fact, interpretation (reasoning), and decision) before fine-tuning. We also consider KELLER (Deng et al., 2024b), which leverages sub-facts in fact descriptions for more comprehensive retrieval. However, KELLER focuses on the reranking process and performs passage-level retrieval and majority voting (*maxsum*), which requires multiple inferences for a case, making a fair comparison difficult. For KELLER, we refer to Appendix A.1.

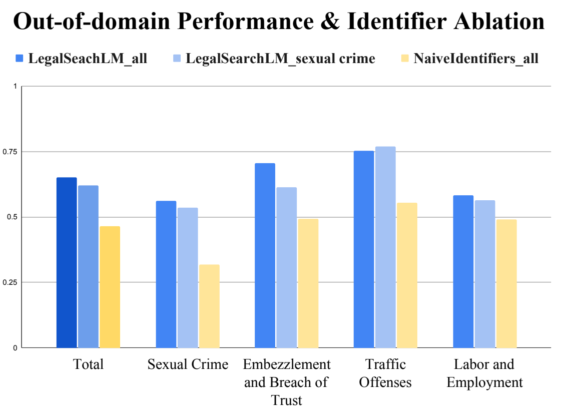

Figure 5: Performance on out-of-distribution of SearchLM.

4.2 Experimental Results

Performance on LEGAR BENCH$_{Standard}$. An evaluation on the standard version, consisting of 411 various query types across 33 crime categories, demonstrates that SearchLM outperforms BM25 by 17%, Contriever by 20%, and SAILER by 6%. Specifically, it outperformed BM25 in 28 crime categories, Contriever across all categories, and SAILER in 21 categories. In Table 2, we provide the results for 8 out of 33 criminal categories, and the full results are listed in Appendix A.1 where we also provide a comparison with the reranked model, KELLER.

Performance on LEGAR BENCH$_{Stricter}$. An evaluation on the stricter version, which includes 160 diverse query types across 8 crime categories, further demonstrates SearchLM's effectiveness in handling complex legal knowledge, achieving the highest performance in Table 2. BM25 excels at capturing fine-grained details through exact lexical matching, leading to stronger performance in LEGAR BENCH$_{Stricter}$ compared to embedding-based similarity search. SearchLM effectively captures both fine-grained details and legal semantic understanding, combining the strengths of both approaches for more robust retrieval.

5 Analysis and Discussion

- **RQ1. Can LegalSearchLM generalizes across various crime which never encounter during training?**
- **RQ2. Does element-to-element learning beyond query case-to-element improve LegalSearchLM performance?**

RQ1. LegalSearchLM exhibits stronger generalizability than naive generative retrieval. Generalizability is crucial in LCR, as legal professionals handle diverse cases. To evaluate this, we train SearchLM on a sexual crime domain and test it on other domains (embezzlement and breach of trust, traffic offenses, and labor and employment). We compare the results with a generative retrieval model using naive identifiers, trained on all crime domains in the training dataset. Figure 5 shows that LegalSearchLM$_{sexual\ crime}$ outperforms NaiveIdentifiers$_{all}$ by 15.66%. This demonstrates that effectively capturing key legal factors is more beneficial than training on various datasets with careless identifiers. Notably, it performs on par with LegalSearchLM$_{all}$, indicating that, even with limited data in certain domains, it can achieve performance similar to that of models trained on broader domains.

RQ2. Element-to-element learning slightly enhances SearchLM performance compared to training only on query case-to-element. LCR typically uses the query case as input, but training between specific factual elements within the case improves precision. Our experimental results in the sexual crime domain show that training with element-to-element pairs (37.172%) shows a slight improvement over training without them (37.103%). We randomly extract legal elements from $\mathcal{F}_{\text{relevant, core-entry}}$ described in Section 3.3 and balance the dataset scale between test sets with and without these elements. These findings align well with the research on the granularity of retrieval text in (Chen et al., 2022).

6 Related Works

6.1 Legal Case Retrieval Datasets

Legal Case Retrieval (LCR) is the task of retrieving target cases relevant to a query case (Feng et al., 2024; Ma et al., 2021).

LCR can be further divided by the definition of relevance, either *citation-based* (Shao et al., 2020; Li et al., 2023c) or *similarity-based* (Ma et al., 2021). In our work, we focus on the latter, as (1) finding precedents (similar past cases) is an important task for legal practitioners (Bhattacharya et al., 2022; Mandal et al., 2017) and (2) citation-based labels are inherently scarce (*e.g.* only one of the similar related cases might be cited), which might lead to a large number of false negatives.

In similar case search, annotating large-scale case data is extremely challenging due to the complexity of legal knowledge. To ease the annotation complexity, existing works often restrict the number of crime types, or only use *fact* section and discard other important legal issues such as defendant's claims and court's judgments about the claim (Ma et al., 2021). Furthermore, *pooling*, where classic IR methods first filter the corpus into a small (Ĩ00 documents) retrieval pool (Arora et al., 2018; Ma et al., 2021), which might cause unwanted bias (*e.g.* when pooled with BM25, models might prefer targets with higher lexical similarity).

While LeCARD-v2 (Li et al., 2023b) improves the data quantity and quality by applying diverse pooling strategies and taking account of penalty/procedural controversy beyond facts, it still fails to take account of critical legal issues and overcome the small retrieval pool. In contrast, LEGAR BENCH$_{Standard}$ has effectively scaled the number of distinct crimes and the number of documents in the retrieval pool by using statutory provisions. Furthermore, LEGAR BENCH$_{Stricter}$ can evaluate the relevance based on expert-annotated legal factors on a large scale, which was not possible before.

6.2 Legal Case Retrievers

Legal case retrievers have rapidly adapted to the recent improvements in language model-based retrieval techniques. While earlier approaches have directly applied general retriever architectures like cross-encoder rerankers (Nogueira and Cho, 2020) using models fine-tuned on legal data (Xiao et al., 2021), recent works focus on incorporating the structure and legal knowledge to improve the performance. SAILER (Li et al., 2023a) incorporates the document structure of legal cases during the pretraining strategy, improving the embedding quality. KELLER and Elem4LCR first segment the case into atomic legal elements, and apply element-wise embedding similarity (KELLER) or cross-encoder scoring (Elem4LCR) between cases to obtain a fine-grained similarity score. LegalSearchLM further improves the strategy by selecting the key phrase as the initial token, which we show its importance for generative retrieval in Section 5.

7 Conclusion

We propose a benchmark with strict relevance criteria for Korean legal case retrieval and introduce generative retrieval to overcome the limitations of existing search methods. We construct LEGAR BENCH*Standard*, which consists of 411 similar case groups across 33 criminal cases, and LEGAR BENCH*Stricter*, which comprises 160 diverse query types across 8 criminal categories. We also present a new retrieval approach in LCR using generative retrieval, which can capture the core legal elements required from given query cases. This achieves state-of-the-art performance on both LEGAR BENCH*Standard* and LEGAR BENCH*Stricter*.

8 Limitations

In this dataset, we construct the largest benchmark in the legal case retrieval task, LEGAR BENCH. However, this dataset is restricted to the cases and statutes from the Korean legal system, which might limit its applicability beyond other jurisdictions and to non-Korean speakers. Furthermore, although we hired legal experts to establish the relevance criteria in LEGAR BENCH$_{Stricter}$, they were not involved in the manual verification of case-to-case relevance. As a result, there may be undetected noise in the dataset.

References

Piyush Arora, Murhaf Hossari, Alfredo Maldonado, Clare Conran, and G. Jones. 2018. Challenges in the development of effective systems for professional legal search. In *ProfS/KG4IR/Data:Search@SIGIR*.

Michele Bevilacqua, Giuseppe Ottaviano, Patrick Lewis, Wen tau Yih, Sebastian Riedel, and Fabio Petroni. 2022. Autoregressive search engines: Generating substrings as document identifiers. *Preprint*, arXiv:2204.10628.

Paheli Bhattacharya, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. 2022. Legal case document similarity: You need both network and text. *Information Processing Management*, 59(6):103069.

Fengyu Cai, Xinran Zhao, Tong Chen, Sihao Chen, Hongming Zhang, Iryna Gurevych, and Heinz Koeppl. 2024. MixGR: Enhancing retriever generalization for scientific domain through complementary granularity. *Preprint*, arXiv:2407.10691.

Jiangui Chen, Ruqing Zhang, Jiafeng Guo, Yiqun Liu, Yixing Fan, and Xueqi Cheng. 2022. Corpusbrain: Pre-train a generative retrieval model for knowledge-intensive language tasks. In *Proceedings of the 31st*

ACM International Conference on Information amp; Knowledge Management, CIKM '22, page 191–200. ACM.

Tong Chen, Hongwei Wang, Sihao Chen, Wenhao Yu, Kaixin Ma, Xinran Zhao, Hongming Zhang, and Dong Yu. 2024. Dense x retrieval: What retrieval granularity should we use? *Preprint*, arXiv:2312.06648.

Chenlong Deng, Zhicheng Dou, Yujia Zhou, Peitian Zhang, and Kelong Mao. 2024a. An element is worth a thousand words: Enhancing legal case retrieval by incorporating legal elements. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2354–2365, Bangkok, Thailand. Association for Computational Linguistics.

Chenlong Deng, Kelong Mao, and Zhicheng Dou. 2024b. Learning interpretable legal case retrieval via knowledge-guided case reformulation. *Preprint*, arXiv:2406.19760.

Yi Feng, Chuanyi Li, and Vincent Ng. 2024. Legal case retrieval: A survey of the state of the art. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6472–6485, Bangkok, Thailand. Association for Computational Linguistics.

Wonseok Hwang, Saehee Eom, Hanuhl Lee, Hai Jin Park, and Minjoon Seo. 2022. Data-efficient end-to-end information extraction for statistical legal analysis. In *Proceedings of the Natural Legal Language Processing Workshop 2022*, pages 143–152, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. 2022. Unsupervised dense information retrieval with contrastive learning. *Preprint*, arXiv:2112.09118.

Chaeeun Kim, Soyoung Yoon, Hyunji Lee, Joel Jang, Sohee Yang, and Minjoon Seo. 2024. Exploring the practicality of generative retrieval on dynamic corpora. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13616–13633, Miami, Florida, USA. Association for Computational Linguistics.

Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Yueyue Wu, Yiqun Liu, Chong Chen, and Qi Tian. 2023a. Sailer: Structure-aware pre-trained language model for legal case retrieval. *Preprint*, arXiv:2304.11370.

Haitao Li, Yunqiu Shao, Yueyue Wu, Qingyao Ai, Yixiao Ma, and Yiqun Liu. 2023b. Lecardv2: A large-scale chinese legal case retrieval dataset. *Preprint*, arXiv:2310.17609.

Haitao Li, Weihang Su, Changyue Wang, Yueyue Wu, Qingyao Ai, and Yiqun Liu. 2023c. Thuir@coliee 2023: Incorporating structural knowledge into pre-trained language models for legal case retrieval. *Preprint*, arXiv:2305.06812.

Yongqi Li, Nan Yang, Liang Wang, Furu Wei, and Wenjie Li. 2023d. Learning to rank in generative retrieval. *Preprint*, arXiv:2306.15222.

Yongqi Li, Nan Yang, Liang Wang, Furu Wei, and Wenjie Li. 2023e. Multiview identifiers enhanced generative retrieval. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6636–6648, Toronto, Canada. Association for Computational Linguistics.

Yixiao Ma, Yunqiu Shao, Yueyue Wu, Yiqun Liu, Ruizhe Zhang, Min Zhang, and Shaoping Ma. 2021. Lecard: A legal case retrieval dataset for chinese law system. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2342–2348, New York, NY, USA. Association for Computing Machinery.

Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho. 2024. Hallucination-free? assessing the reliability of leading ai legal research tools. *Preprint*, arXiv:2405.20362.

Arpan Mandal, Raktim Chaki, Sarbajit Saha, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. 2017. Measuring similarity among legal court case documents. In *Proceedings of the 10th Annual ACM India Compute Conference*, Compute '17, page 1–9, New York, NY, USA. Association for Computing Machinery.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. *Preprint*, arXiv:2305.14251.

Rodrigo Nogueira and Kyunghyun Cho. 2020. Passage re-ranking with bert. *Preprint*, arXiv:1901.04085.

Guilherme Moraes Rosa, Ruan Chaves Rodrigues, Roberto Lotufo, and Rodrigo Nogueira. 2021. Yes, bm25 is a strong baseline for legal case retrieval. *Preprint*, arXiv:2105.05686.

Yunqiu Shao, Bulou Liu, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. 2020. Thuir@coliee-2020: Leveraging semantic understanding and exact matching for legal case retrieval and entailment. *Preprint*, arXiv:2012.13102.

Weihang Su, Qingyao Ai, Yueyue Wu, Yixiao Ma, Haitao Li, Yiqun Liu, Zhijing Wu, and Min Zhang. 2024. Caseformer: Pre-training for legal case retrieval based on inter-case distinctions. *Preprint*, arXiv:2311.00333.

Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A pre-trained language model for chinese legal long documents. *Preprint*, arXiv:2105.03887.

Kun Zhang, Chong Chen, Yuanzhuo Wang, Qi Tian, and Long Bai. 2023. Cfgl-lcr: A counterfactual graph learning framework for legal case retrieval. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 3332–3341, New York, NY, USA. Association for Computing Machinery.

A Benchmark Details

A.1 Full Results on LEGAR BENCH$_{Standard}$ and LEGAR BENCH$_{Stricter}$

A.2 Data statistics

Statistics of standard set on 33 crime categories. Table 4 presents a criminal typology that includes 33 major categories of criminal offenses. Each category is classified in detail based on charge titles and statutes, forming the standard evaluation set. The number of standard groups for each category is listed under *# of Standard Group*, while the number of unique case documents mapped to each group is listed under *# of Cases*. The total number of standard groups is 411, encompassing 1,052,506 unique cases, which constitute 85.79% of the entire corpus (1,226,814 cases). This figure underscores the broad coverage of our benchmark across a wide range of criminal offense types.

A.3 Examples of the Performance of LEGAR BENCH$_{Standard}$ on certain crime categories

A.4 List of Stricter Relevance Group

LEGAR BENCH$_{Stricter}$ further divides LEGAR BENCH$_{Standard}$ categories based on different factual details of a criminal case that do not affect the type of charge, but might affect the final judgment (guilty or innocent) or the sentence *e.g.* information about defendant/victims, methods, consequences, and claims made in court.

Five Korean lawyers specialized in the Criminal Act were hired (250$/hr) to list such factors given a specific charge in LEGAR BENCH$_{Standard}$, and provide a comprehensive list of possible options for each factor. The options are primarily based on the official sentencing guidelines from the Sentencing Commission of the Supreme Court of Korea, and annual crime statistics reports published by government/academic authorities including the Supreme Prosecutor's Office and the Korean Institute of Criminology. However, these lists are often insufficient to express existing cases, especially the defendant's claims (*e.g.*, a defendant convicted of assault might claim that the act was due to self-defense, pleading for innocence). Identifying such factors heavily relies on deep understanding and expertise in practicing law. Hence, the lawyers were instructed to add factors and options that are frequent and important in practice but not mentioned in the official documents. Full instructions for the

	Precision@5			
Criminal Category	**SeachLM**	BM25	Contriever	KELLER
[Total]	**[0.68]**	[0.51]	[0.48]	[0.70]
Traffic offenses	**0.75**	0.60	0.72	0.97
Fraud	**0.67**	0.53	0.57	0.92
Injury or Violence	**0.60**	0.50	0.42	0.79
Sexual crime	**0.56**	0.49	0.40	0.82
Theft or Robbery	**0.61**	0.48	0.41	0.86
Obstruction of Business	**0.77**	0.58	0.40	0.94
Embezzlement or Breach of trust	**0.71**	0.64	0.49	0.95
Destruction	**0.76**	0.64	0.48	0.88
Finance or Insurance	**0.84**	0.56	0.60	0.96
Threat	0.58	**0.60**	0.47	0.91
Defamation or Insult	**0.75**	0.58	0.48	0.80
Drug	**0.84**	0.52	0.76	0.92
Criminal trespass	**0.67**	0.63	0.51	0.92
Gambling	**0.89**	0.54	0.63	1.0
Negligent homicide and injury	**0.33**	0.27	0.30	0.93
Obstruction of right	**0.64**	0.52	0.52	1.0
Child abuse or School violence	**0.60**	0.48	0.38	0.64
Medical or Food drug	0.29	**0.35**	0.22	0.1
Murder	0.30	**0.50**	0.30	0.90
Corporation	**0.60**	0.33	0.33	0.33
Bribery	0.47	**0.60**	0.40	0.93
Car	**0.70**	0.60	0.60	0.90
Labor or Employment	**0.58**	0.51	0.40	0.60
Industrial or Serious accidents	**0.55**	0.45	0.20	0.25
Military duty or law	**0.60**	0.50	0.50	0.50
Consumer or Fair trade	**1.00**	0.60	0.80	0.20
Arrest or Detention	**0.80**	**0.80**	0.40	1.0
Intellectual property	**0.87**	0.33	0.67	0.67
IT or Privacy	**1.00**	0.60	0.80	1.0
Misdemeanor	**0.40**	0.20	0.20	0.20
Sexual norms	0.20	0.20	0.20	0
Tax, Administ, Const law	**0.83**	0.61	0.71	0.84
Other criminal offenses	**0.70**	0.58	0.42	1.0

Table 3

annotators can be found in <anonymized>.

Previous work in identifying such factors in the Korean Criminal Act Hwang et al. (2022) includes only 11 unique factors across 4 crime categories focusing only on facts, while this work adds 102 unique factors (including 39 defendant claims) across 8 categories.

B Implementation Details

All models are trained using 8 * A100 80GB GPUs.

SearchLM. To develop our SearchLM based on an autoregressive language model, we take the MT5-base pretrained model and train it on 170K cases for a single epoch.

Contriever. We select Contriever as a representative model for retrieval in the general domain. We perform unsupervised training on the BERT-base-multilingual-cased pretrained model with 170K cases for 10 epochs. Following the results in their work, we use the MoCo method during training rather than in-batch.

SAILER. We implement SAILER as a representative model for retrieval in the legal domain. Following their paper, we pretrain the BERT-base-

Crime categories	# of Standard group	# of Cases
Traffic offenses	13	319,527
Fraud	21	181,703
Injury or Violence	31	146,764
Sexual crime	132	104,919
Theft or Robbery	38	74,772
Obstruction of Business	13	74,722
Embezzlement or Breach of trust	15	39,835
Destruction	5	39,595
Finance or Insurance	5	32,944
Threat	11	27,496
Defamation or Insult	8	27,278
Drug	5	26,066
Criminal trespass	15	24,856
Gambling	7	11,091
Negligent homicide and injury	6	7,384
Obstruction of right	5	6,749
Child abuse or School violence	10	5,756
Medical or Food drug	11	98
Murder	2	4,306
Corporation	3	1,195
Bribery	3	1,638
Car	2	20,882
Labor or Employment	11	12,647
Industrial or Serious accidents	4	198
Military duty or law	2	9,300
Consumer or Fair trade	1	128
Arrest or Detention	1	6
Intellectual property	3	3,927
IT or Privacy	2	2,311
Misdemeanor	1	6,476
Sexual norms	1	4,140
Tax, Administ, Const law	14	40,890
Other criminal offenses	10	23,211
Total	**411**	**1,052,506**

Table 4: Statistics of Crime typology and Standard version of LEGAR BENCH. The total number of cases is reported as a unique count, excluding duplicates from cases classified under multiple categories $1,347,962 \rightarrow 1,052,506$.

multilingual-cased model on facts, interpretations, and decisions of 1.2M cases for a single epoch, using the same configuration as in SAILER. The pretrained model is then fine-tuned for a single epoch with positive and negative samples, adjusting the learning rate from the default 5e-6 to 5e-5. We retrieve 100 related cases using BM25 over the 170K cases, selecting those with the same case name as positive samples and others as negative. To ensure comparability with other baselines, we use 5 positive and 5 negative cases per query.

Crime categories	# of Stricter group	# of Cases
Fraud	8	325
Injury or Violence	19	308
Sexual crime	111	1,061
Finance or Insurance	1	28
Defamation or Insult	6	253
Drug	4	37
Murder	2	8
Traffic offenses	9	330
Total	**160**	**2,350**

Table 5: Statistics of Stricter version of LEGAR BENCH.

<Traffic offenses>
Violation of the Road Traffic Act (Driving Under the Influence)
Violation of the Road Traffic Act (Refusal to Submit to a Breathalyzer Test)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Hit-and-Run Resulting in Injury)
Violation of the Road Traffic Act (Failure to Take Measures After an Accident)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Dangerous Driving Resulting in Injury)
Violation of the Road Traffic Act (Unlicensed Driving)
Violation of the Act on Special Cases Concerning the Settlement of Traffic Accidents (Injury by Negligence)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Assault on a Driver, etc.)
Violation of the Act on Special Cases Concerning the Settlement of Traffic Accidents (Death by Negligence)
Violation of the Road Traffic Act (Reckless Joint Dangerous Driving)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Dangerous Driving Resulting in Death)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Injury in a Child Protection Zone)
Violation of the Act on the Aggravated Punishment of Specific Crimes (Death in a Child Protection Zone)

Table 6: List of query case types for Traffic offenses.

<Fraud>
Fraud
Violation of the Act on the Punishment of Violent Acts, etc. (Extortion)
Attempted Fraud
Computer-Based Fraud
Violation of the Act on the Aggravated Punishment of Specific Economic Crimes (Fraud)
Violation of the Specialized Credit Finance Act
Extortion
Attempted Extortion
Habitual Fraud
Habitual Extortion
Quasi-Fraud
Fraudulent Use of Public Facilities
Violation of the Act on the Aggravated Punishment of Specific Economic Crimes (Extortion)
Attempted Computer-Based Fraud
Aggravated Extortion
Attempted Aggravated Extortion
Habitual Quasi-Fraud
Attempted Quasi-Fraud
Attempted Habitual Extortion
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Extortion)
Attempted Habitual Fraud

Table 7: List of query case types for Fraud.

<Injury and Violence>
Aggravated Assault
Assault
Injury (Bodily Harm)
Violation of the Act on the Punishment of Violent Acts, etc. (Assault)
Violation of the Act on the Punishment of Violent Acts, etc. (Injury)
Aggravated Assault with a Deadly Weapon or Other Means
Attempted Aggravated Assault
Violation of the Act on the Punishment of Violent Acts, etc. (Organization and Activities of a Criminal Group, etc.)
Aggravated Assault Resulting in Injury
Assault Resulting in Death
Assault Resulting in Injury
Injury to a Lineal Ascendant
Habitual Assault
Habitual Infliction of Injury
Assault Against a Lineal Ascendant
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Assault)
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Injury)
Aggravated Assault Against a Lineal Ascendant
Serious Injury (Grievous Bodily Harm)
Habitual Aggravated Assault
Violation of the Act on the Punishment of Violent Acts, etc. (Injury to a Lineal Ascendant)
Assault Against a Lineal Ascendant Resulting in Injury
Habitual Assault Against a Lineal Ascendant
Habitual Aggravated Assault
Assault Against a Lineal Ascendant Resulting in Death
Attempted Injury (Attempted Bodily Harm)
Aggravated Injury to a Lineal Ascendant
Violation of the Act on the Punishment of Violent Acts, etc. (Assault Against a Lineal Ascendant)
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Injury to a Lineal Ascendant)
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Assault Against a Lineal Ascendant)

Table 8: List of query case types for Injury and Violence.

<Theft and Robbery>
Aggravated Larceny
Larceny (Theft)
Nighttime Burglary and Larceny in a Structure
Attempted Larceny
Aggravated Robbery
Attempted Nighttime Residential Burglary and Larceny
Attempted Aggravated Larceny
Nighttime Residential Burglary and Larceny
Habitual Larceny
Robbery
Robbery Resulting in Injury
Attempted Nighttime Burglary and Larceny in a Structure
Robbery and Rape
Quasi-Robbery (Larceny Escalating into Robbery)
Quasi-Aggravated Robbery
Habitual Aggravated Larceny
Attempted Aggravated Robbery
Habitual Nighttime Residential Burglary and Larceny
Nighttime Burglary and Larceny in an Occupied Room
Attempted Nighttime Burglary and Larceny in an Occupied Room
Murder During Robbery
Nighttime Ship Burglary and Larceny
Preparation for Robbery
Conspiracy to Commit Robbery
Habitual Nighttime Burglary and Larceny in a Structure
Attempted Robbery
Robbery Resulting in Bodily Injury
Attempted Murder During Robbery
Attempted Habitual Larceny
Attempted Quasi-Robbery
Robbery Resulting in Death
Attempted Habitual Aggravated Larceny
Attempted Nighttime Ship Burglary and Larceny
Attempted Quasi-Aggravated Robbery
Attempted Robbery and Rape
Attempted Habitual Nighttime Burglary and Larceny in a Structure
Habitual Nighttime Burglary and Larceny in an Occupied Room
Attempted Habitual Nighttime Residential Burglary and Larceny

Table 9: List of query case types for Theft and Robbery.

<Embezzlement and Breach of trust>
Embezzlement
Embezzlement of Lost or Misplaced Property
Breach of Trust
Breach of Trust in the Course of Duty
Violation of the Act on the Aggravated Punishment of Specific Economic Crimes (Breach of Trust)
Giving a Bribe in Relation to a Breach of Trust
Embezzlement in the Course of Duty
Violation of the Act on the Aggravated Punishment of Specific Economic Crimes (Embezzlement)
Receiving a Bribe in Relation to a Breach of Trust
Attempted Breach of Trust in the Course of Duty
Attempted Receipt of a Bribe in Relation to a Breach of Trust
Attempted Embezzlement
Attempted Breach of Trust
Attempted Giving of a Bribe in Relation to a Breach of Trust
Attempted Embezzlement in the Course of Duty

Table 10: List of query case types for Embezzlement and Breach of trust.

<Destruction>
Destruction of Property
Aggravated Destruction of Property
Violation of the Act on the Punishment of Violent Acts, etc. (Destruction of Property)
Attempted Destruction of Property
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Destruction of Property)

Table 11: List of query case types for Destruction.

<Finance and Insurance>
Violation of the Electronic Financial Transactions Act
Violation of the Act on Real Name Financial Transactions and Confidentiality
Violation of the Special Act on Prevention of Insurance Fraud
Violation of the Act on the Regulation and Punishment of Crime Proceeds Concealment
Violation of the Act on the Regulation of Similar Deposit-Like Transactions
Violation of the Act on the Registration of Loan Businesses and Protection of Financial Consumers
Violation of the Act on the Reporting and Use of Specific Financial Transaction Information

Table 12: List of query case types for Finance and Insurance.

<Threat>
Threatening
Aggravated Threatening
Violation of the Act on the Punishment of Violent Acts, etc. (Threatening)
Aggravated Threatening Against a Lineal Ascendant
Attempted Aggravated Threatening
Threatening Against a Lineal Ascendant
Habitual Threatening
Habitual Threatening Against a Lineal Ascendant
Attempted Threatening
Violation of the Act on the Punishment of Violent Acts, etc. (Repeat Offense of Threatening)
Habitual Aggravated Threatening

Table 13: List of query case types for Threat.

<Defamation and Insult>
Violation of the Act on Promotion of Information and Communications Network Utilization and Information Protection (Defamation through False Allegation)
Defamation (Insult)
Defamation by Factual Statement
Defamation by False Statement
Defamation through Publication
Violation of the Act on Promotion of Information and Communications Network Utilization and Information Protection (Dissemination of Obscene Materials)
Violation of the Act on Promotion of Information and Communications Network Utilization and Information Protection (Interference with Information and Communications Networks, etc.)
Violation of the Act on Promotion of Information and Communications Network Utilization and Information Protection (Defamation by Factual Statement)

Table 14: List of query case types for Defamation and Insult.

<Drug>
Violation of the Narcotics Control Act (Psychotropic Substances)
Violation of the Narcotics Control Act (Cannabis)
Violation of the Narcotics Control Act (Narcotic Drugs)
Violation of the Narcotics Control Act (Temporary Narcotic—Psychotropic Substances)
Violation of the Narcotics Control Act (Temporary Narcotic—Cannabis)

Table 15: List of query case types for Drug.

<Gambling>
Operation of a Gambling Facility
Violation of the National Sports Promotion Act (Operation of a Gambling Venue, etc.)
Habitual Gambling
Violation of the Game Industry Promotion Act
Establishment of a Gambling House
Violation of the National Sports Promotion Act (Gambling, etc.)
Gambling

Table 16: List of query case types for Gambling.

Crime categories	# of Standard group	Factors(# Options)	# of Stricter Group
Traffic offenses		Traffic accident type(6), Traffic accident time(2), Automobile type(3), Road type(4), Gross negligence type(18), Automobile accident insurance(3), Malpractice?(3), Hit-and-run type(3), Hit-and-run loss type(2), Aided victim?(3), Not aware of accident?(3), Blood alcohol level(3), Driving distance(4), Necessity?(3), Not driving?(3), Absorption phase?(3), Excessive extrapolation?(3), Driving without license type(5), Not aware of license suspension(3), Not aware of invalidation(3), Injury severity(8), Injury?(3), Number of victims(3), Defendant-victim relation(10), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3), Reason not reaching consummation(4), Reached consummation?(3)	
Fraud		Fraud type(14), No intent for pecuniary advantage?(3), No intent to defraud?(3), Profit(12), Defendant feeble-minded?(3), Defendant insanity?(3)	
Injury or Violence		Two-way assault(2), Motivation(7), Intent to injure?(3), Self-defense?(3), Assault method(9), Injury severity(8), Injury?(3), Special crime type(2), Number of accomplices(5), Dangerous weapon?(3), Time between injury and death(4), Injury direct cause of death?(3), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3)	
Sexual crime		Sexual assault location(6), Victim age(4), Victim disability(2), Defendant under influence(3), Victim under influence(3), Consent?(3), Intercourse type(4), Incident act type(4), Incident act by blitz(2), Victim sexual shame(3), Inability to resist cause(5), Aware of inability to resist?(3), Aware of victim's age under 13?(3), Aware of victim's age under 16?(3), Fraudulence/influence type(7), Victim under influence?(3), Covert photography filming/distribution type(7), Number of covert photography(4), Profit(4), Obscene communication medium(4), Obscene communication content(6), Object of sexual satisfaction(2), Reached the victim?(3), Assault/threat type(6), Assault method(9), Injury severity(8), Injury?(3), Special crime type(2), Number of accomplices(5), Dangerous weapon?(3), Time between injury and death(4), Injury direct cause of death?(3), No intent to defraud?(3), Number of victims(3), Defendant-victim relation(10), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3), Reason not reaching consummation(4), Reached consummation?(3)	
Finance or Insurance	1[†]	Insurance fraud type(5), No intent for pecuniary advantage?(3), No intent to defraud?(3), Profit(12), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3), Reason not reaching consummation(4), Reached consummation?(3)	
Defamation or Insult		Defamation content(5), Defamation medium(8), Insult content(4), Victim type(3), Alleged facts?, Publicly alleged?(3), Can specify victim?(3), Defaming the social status?(3), Justified(3), Number of victims(3), Defendant-victim relation(10), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3)	
Drug		Drug type(14), Drug crime type(7), Defendant role(6), Narcotic handling license(6), Drug quantity(6), Profit(12), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3)	
Murder		Motivation(7), Intent to kill?(3), Self-defense?(3), Assault method(9), Injury?(3), Number of victims(3), Defendant-victim relation(10), Surrender(2), Defendant feeble-minded?(3), Defendant insanity?(3), Reason not reaching consummation(4), Reached consummation?(3)	

Table 17: Factors for defining Stricter relevance. Each factor is presented with the number of options in parentheses. Question mark(?) indicates that the factor represents a *claim* defendant makes in a court, which always has three options (not mention, claimed but not taken, claimed and taken). As some factors only apply to certain standard groups (*e.g.* Traffic accident type(6) only applies to traffic crimes involving accidents and not crimes like Driving Under the Influence (without any traffic accident)) and not all combinations are possible (*e.g.* Killing Ascendant (killing one's own or any lineal ascendant of one's spouse) cases can only take two options (*parent, other family members*) out of 10 options (*partners, friend, ...*) provided for the Defendant-victim relation factor), the total number of stricter groups is a magnitude smaller compared to all option numbers multiplied.