LegEx: Dataset for Legal Case Retrieval Based on Negation and Exclusion Conditions

Anonymous ACL submission

Abstract

In the legal domain, queries involving negation or exclusion conditions such as "not having done " or "excluding " frequently arise in legal case retrieval. However, existing studies rarely address such expressions systematically. To bridge this gap, this study constructs a dataset explicitly tailored for legal case retrieval based on negation and exclusion conditions, consisting of queries, corresponding relevant cases, and challenging negative examples. This work also experimentally evaluates the limitations of existing information retrieval models and the performance improvements achieved through fine-tuning in case retrieval given such conditions. Experimental results demonstrate that pretrained information retrieval models initially fail to properly handle negation and exclusion expressions, whereas their ability to respond to these conditions significantly improves after fine-tuning. By introducing a specialized dataset for negation and exclusion queries in the previously unexplored legal domain, this study highlights the limitations of current retrieval models and validates that a dataset-driven approach can effectively overcome these challenges.

1 Introduction

Case-law retrieval—the task of locating precedents relevant to a query—is more challenging in the legal domain than elsewhere because of specialised terminology and phraseology. For instance, case law frequently contains negative and exclusionary phrases such as "*did not*," "*is not A but is B*," as well as words with negative prefixes like "*disallowance*" and "*non-existence*." This characteristic necessitates a consideration of the overall context rather than specific expressions or individual words during case-law retrieval, as a single-word difference can alter the relevant precedent that must be found.

While prior research exists on information retrieval methods for negation and exclusion conditions in general domains (e.g. Weller et al. (2024); Zhang et al. (2025); Barkhof et al. (2025)), these studies show that existing models perform poorly on such queries because current datasets lack query–document pairs that explicitly encode negation or exclusion. In response, Weller et al. (2024) constructed a dataset containing these conditions.

However, their dataset primarily deals with relatively easy documents, where the correct answer can be inferred from the mere presence of a negative expression between otherwise similar passages. In contrast, legal texts exhibit much richer linguistic complexity. The simple presence of a negation or exclusion term does not automatically make a document irrelevant; understanding the entire context is essential.

Therefore, we propose a legal-case retrieval dataset focused on negation and exclusion conditions that better captures the unique characteristics of legal discourse.

First, we utilise a large language model (LLM) to generate queries with negation and exclusion conditions from publicly available civil-case precedents of the Supreme Court of Korea. The prompt templates and outputs were reviewed by domain experts (lawyers) to ensure high factual reliability.

Next, for each generated query, we apply the BM25 algorithm (Robertson and Zaragoza, 2009) to retrieve the top-10 precedents. We randomly select three of them as hard-negative candidates, then use an LLM again to verify whether each candidate is a true or false negative.

Through this process, we build triplets {query, positive precedent, hard negative}. We evaluate the baseline BM25 model and four multilingual embedding models on this dataset using mean reciprocal rank (MRR; Craswell, 2009) and Recall@k. Subsequently, we fine-tune the embedding models with a contrastive-learning objective (Chen et al., 2020) and re-evaluate them. The results confirm substantial performance gains on queries containing negation and exclusion expressions after fine-tuning.

2 Related Work

2.1 IR with Negation and Exclusion Conditions in General Domains

General natural-language queries often include negation or exclusion (e.g. "*not X*", "*excluding Y*") to filter out unwanted information. Such expressions remain challenging for many information-retrieval (IR) systems: Weller et al. (2024) show that state-of-the-art models can perform no better than random retrieval when faced with negated queries. They attribute this drop largely to the scarcity of training data that explicitly encodes negation. Later work confirms that providing a dedicated evaluation set and additional training pairs can partially mitigate this weakness (Zhang et al., 2025; Barkhof et al., 2025).

Weller et al. (2024) evaluate ranking behaviour by constructing nearly identical document pairs that differ only in the presence or absence of a negative expression. When various IR models rank these pairs with the same negated query, most models fail to prefer the truly relevant document, again illustrating the difficulty of handling negation.

2.2 Legal-Domain IR and Its Limitations

In the legal domain, research has largely focused on improving retrieval via semantic similarity. For Korean case law, Park and Kim (2022) employ Sentence-BERT to surpass traditional baselines such as TF–IDF and Doc2Vec. Deep-learning approaches that incorporate logical or semantic structures (e.g. rules) have also been explored (Sun et al., 2024), yet most concentrate on relevance-score computation and do not explicitly model negation or exclusion.

Recent efforts leverage larger datasets and generative models. Kim et al. (2025) introduce LEGAR BENCH, a benchmark of 1.2 million precedents across 411 crime types, and propose LEGALSEARCHLM, which retrieves cases by generating key legal elements from a query, achieving superior performance to prior methods.

To date, however, no specialised study has addressed legal-case retrieval queries that contain negation and exclusion conditions. Even in the broader IR literature, work on such queries remains sparse. Our study fills this gap by constructing and releasing a dataset specifically designed for negation- and exclusion-based retrieval in the legal domain.

3 Methodology

3.1 Dataset for Negation and Exclusion-Based Case Law Retrieval

In this study, we aim to construct a case law retrieval dataset for queries containing negation and exclusion conditions within the legal domain. This dataset is a triplet dataset composed of a query with negation/exclusion conditions, a corresponding positive document, and a negative document. An example of such a query is provided below.

> Query Example 1: "Show me a precedent where an act was recognized as a commercial act even though it was not a basic commercial act."

Here, the negation/exclusion condition is 'not a basic commercial act,' and the precedent sought by the query is one that reflects this condition while being 'recognized as a commercial act.'

The positive document is the correct precedent that reflects the query's negation/exclusion condition. An example is shown in Figure A1. In this example, we have excerpted only the "Gist of Judgment" and "Summary of Judgment" sections from the full Supreme Court precedent.

As the underlined portion of Figure A1 indicates, this precedent is a case where, despite the act not being a basic commercial act, the act of a merchant entering into a contract for business purposes was deemed a commercial act, thereby recognizing the commercial claim and the commercial statute of limitations. This corresponds to the query's condition 'not a basic commercial act' and also fulfills the requirement of being a 'precedent where an act was recognized as a commercial act.'

By contrast, a negative document is a precedent that fails to satisfy the negation or exclusion constraint. An example is shown in Figure A2.

As the underlined portion of Figure A2 indicates, this precedent was not a case where something other than a basic commercial act was recognized as a commercial act; rather, it was judged to be a typical basic commercial act. This is a precedent where a basic commercial act was recognized as a commercial act and thus does not meet the conditions of the query.

Let's examine another example of a query with a negation/exclusion condition, a positive document, and a negative document.

Query Example 2: “Please show me a case where the act of a debtor selling real estate to a specific creditor was recognized as a fraudulent act, even though the real estate was not sold at a price significantly lower than the market value.”

In this query, the negation/exclusion condition is ‘not sold at a price significantly lower than the market value,’ and the sought precedent is a ‘case where the act of a debtor selling real estate to a specific creditor was recognized as a fraudulent act.’

This precedent in Figure A3 is a case where fraudulent intent was recognized even though the real estate was sold at a price equivalent to its market value. This corresponds to the query’s condition ‘not sold at a price significantly lower than the market value’ and also meets the requirement of being a ‘case where the act of a debtor selling real estate to a specific creditor was recognized as a fraudulent act.’

Figure A4 is a case where fraudulent intent was ‘recognized’ because the real estate was sold at a low price. Although it is ultimately a case where a fraudulent act was recognized, it does not meet the query’s condition that the real estate was not sold at a price significantly lower than the market value.

3.2 Dataset Generation Strategy

To generate the negation and exclusion-based case law retrieval dataset, we collected public case law data, utilized an LLM and existing IR models, and received reviews from legal domain experts (lawyers).

First, for the case law data, we collected a total of 16,792 Supreme Court civil case precedents (case classification ‘’) from the “Legal/Regulatory Text Analysis Data (Advanced) – Case Law Data by Situation” publicly available on AI-Hub[1].

From this collected data, we used an LLM to generate queries with negation and exclusion conditions, creating query–positive document pairs. Subsequently, using the generated queries and the BM25 search model (Robertson and Zaragoza, 2009), we selected hard negative documents. When utilizing the LLM, we applied techniques such as few-shot prompting (Brown et al., 2020 (Brown et al., 2020)), and the entire data generation process was accompanied by expert review.

[1] https://www.aihub.or.kr

3.3 Source data Collection

For the source data collection, we extracted the “Case Number” and “Precedent Content” from a total of 16,792 Supreme Court civil case precedents (case classification ‘’) from AI-Hub’s “Legal/Regulatory Text Analysis Data (Advanced) – Case Law Data by Situation”. The format of the data actually used is shown in Figure A5.

3.4 Generation of Query–Positive Document Pairs with Negation/Exclusion Conditions

Based on the collected case law data, we used Google’s LLM model, Gemini 2.5 Pro, to generate queries with negation and exclusion conditions. The prompt used for query generation consisted of an explanation and conditions for such queries, as well as example case law data for few-shot learning (Brown et al., 2020 (Brown et al., 2020)), and was reviewed by legal domain experts.

First, let’s look at the query explanation part of the prompt described in Figure 1. This part includes the definition of a negation/exclusion query, example expressions and structures, and sample questions.

Next is the conditions part for the negation/exclusion queries. This is composed of the requirements that the queries must meet and corresponding example case law data.

In Figure B1, (Condition 2) is to prevent the query from being applicable to only one specific precedent. (Condition 3) is to ensure the negation/exclusion condition is relevant to what the query asks for and simultaneously narrows the search scope meaningfully.

Using the prompt in Figure B1 with the Google Gemini 2.5 Pro model to generate negation/exclusion queries resulted in a total of 16,791 query–positive document pairs, excluding one null value.

3.5 Selection of Hard Negative Documents for Queries with Negation/Exclusion Conditions

For the 16,791 generated queries, we selected hard negative precedents from the remaining 16,790 precedents (excluding the single source precedent for each query) using expert review, BM25, and an LLM.

First, for each query, we performed a BM25 search on the 16,790 precedents and randomly sampled three documents from the top 10 results to

A 'negation and exclusion condition query' refers to a question that, when searching for case law, excludes a specific element or asks only for situations where that element was not established.

1. Example Expressions and Structure

- "Show me B that is not A."
- "Excluding situations like A, show me cases where B is recognized."
- "Excluding A, show me B."

→ B must be a concept that encompasses both the case of A and the case of not-A. In other words, in a 'negation and exclusion condition query,' the negation/exclusion condition (~A) must substantively narrow the scope of the query (B). To achieve this, B should be established as a comprehensive concept.

2. Example Questions

- "Excluding cases where the fraudulent act was a form of repayment, show me cases where a claim for the revocation of the fraudulent act was granted."
- "Excluding cases where a preliminary legal issue was in question, show me precedents where *res judicata* was recognized."
- "Show me a precedent where the crime of bribery was applied to a person who is not a public official."
- "Are there cases where liability for damages was imposed even when the patent right was confirmed to be non-existent?"
- "Show me a precedent where a claim was recognized as a commercial claim even though it did not arise from a basic commercial act."

Figure 1: Query generation prompt explaining negation and exclusion condition queries

serve as hard negative candidates.

For these three selected hard negative candidates, we used the Google Gemini 2.5 Pro model, along with expert review, to make a final determination of whether they were correct or incorrect answers for the query, i.e., to identify false negatives. The prompt used for this task is shown in Figure B2as follows.

As a result of this final verification shown in Table 1, a total of 16,769 dataset entries were created (excluding 22 LLM generation errors), categorized into four types based on the number of false negatives among the three hard negative candidates.

False Negatives	Number of Datasets
3	1,841
2	3,455
1	5,582
0	5,891

Table 1: Distribution of datasets by the number of false negatives.

3.6 Preparation of Train, Dev, and Test Datasets

To conduct testing and training with the generated data, we partitioned the data into train, dev, and test sets, distributing them proportionally according to the number of false negative precedents.

Ultimately, we successfully created the "Negation and Exclusion-Based Case Law Retrieval Dataset" as initially intended, consisting of queries with negation/exclusion conditions, corresponding positive documents, and hard negative documents.

Split	Size
train	11,928
dev	1,500
test	2,000

Table 2: Train, dev, and test dataset split.

As shown in Table 2, the dataset is divided into 11,928 training, 1,500 development, and 2,000 test instances.

4 Experiments and Analysis

4.1 Experimental Design

We conducted two experiments with the created dataset. First, we evaluated the performance of existing IR models in handling queries with negation and exclusion conditions using the test dataset. Subsequently, we fine-tuned the existing IR models on the train and dev datasets using a contrastive learning approach and then verified whether their performance on such queries improved, using the test dataset.

The evaluation covers BM25 and four sentence embedding models—gte-multilingual-base, bge-m3, KURE-v1, and Qwen3-0.6B.

The evaluation metrics used were MRR (Mean Reciprocal Rank) and Recall@k. MRR is the average of the reciprocal ranks of the first correct document for each query, indicating how quickly the IR model surfaces the correct answer. A value closer to 1 indicates a higher proportion of correct answers appearing first. Recall@k is the average proportion of all correct documents found within the top k documents for each query, showing how well the IR model finds the correct answers without missing them in the top k results. A value closer to 1 indicates a higher proportion of correct answers included within the top k documents.

4.2 Baseline Performance Verification on Queries with Negation/Exclusion Conditions

We measured the MRR and Recall@k values for BM25 and the embedding models (gte-multilingual-base, bge-m3, KURE-v1, Qwen3-0.6B) on the test dataset. The measurement results are shown in Table 3.

First, BM25 shows the highest MRR and Recall@k values across all k values. Starting from 0.181 at k=1, the values gradually increase to an MRR of 0.283 and Recall@k of 0.792. Legal terms appearing in case law data often include uncommon yet crucial words like ’auxiliary commercial act’ or ’fraudulent act’. BM25 assigns high scores to such words, allowing it to rank the correct documents higher than the embedding models.

In contrast, the results for the embedding-based models are generally lower than those for BM25. The gte-multilingual-base model has an MRR of 0.117 at k=1, lower than BM25’s 0.181. As k increases, the MRR gradually rises to 0.189 and Recall@k to 0.614, but these values still fall short of BM25’s (MRR 0.283, Recall@k 0.792). This is the result of using a multilingual embedding model without fine-tuning, suggesting that the gte-multilingual-base model struggled with the precise semantic understanding of difficult technical terms and queries containing negation/exclusion conditions in the Korean legal domain.

The bge-m3 model shows results similar to the gte-multilingual-base model. Starting with an MRR of 0.107 and Recall of 0.107 at k=1, it increases to an MRR of 0.178 and Recall of 0.612, but its scores are distinctly lower than BM25 across all intervals. bge-m3 is a multilingual embedding model released by the Beijing Academy of Artificial Intelligence (BAAI) and, like gte-multilingual-base, is not specifically pre-trained on Korean.

The KURE-v1 model shows the highest results among the embedding models in almost all aspects. At k=1, its MRR of 0.151 and Recall@k of 0.151 are the second highest after BM25. At k=100, its MRR of 0.244 and Recall@k of 0.739 are close to BM25’s values (MRR 0.283, Recall@k 0.792). The KURE-v1 model is a version of the bge-m3 model fine-tuned on a Korean corpus and is reputed to show superior performance in Korean search compared to multilingual-based models (Hwang et al., 2025). However, since the KURE-v1 model was also used without additional training specialized for queries with negation/exclusion conditions, its values were not higher than the BM25 model.

The Qwen3-0.6B model shows respectable results comparable to the KURE-v1 model. At k=1, its MRR of 0.145 and Recall@k of 0.145 are similar to KURE-v1, and at k=100, it records an MRR of 0.236 and Recall@k of 0.744, showing a slight difference from KURE-v1. Qwen3-0.6B is a pre-trained sentence embedding model with about 600 million parameters that has shown excellent performance in multilingual sentence embedding benchmarks, and this strength seems to be reflected in the results of this experiment. However, this model was also used without separate legal domain-specific training or fine-tuning for negation/exclusion queries, thus limiting its ability to perfectly handle them.

4.3 Performance Improvement Verification after Fine-tuning

The baseline performance verification showed that pre-trained embedding models perform worse than BM25 and are generally inadequate at handling queries with negation and exclusion conditions. Therefore, we conducted contrastive learning-based fine-tuning on the embedding models using the train and dev datasets and then verified how much their performance improved compared to before fine-tuning, using the same test dataset. The experimental results are shown in Table 4.

After fine-tuning, the performance of the gte-multilingual-base model shows a meaningful improvement across the board. The MRR and Recall@k at k=1 increased from 0.117 to 0.197. At k=100, the MRR increased from 0.189 to 0.303, and Recall@k rose significantly from 0.614 to 0.818. The bge-m3 model also showed signifi-

IR Model	Metric	k				
		1	3	5	10	100
BM25	MRR	0.181	0.238	0.258	0.272	0.283
	Recall@k	0.181	0.314	0.401	0.507	0.792
gte-multilingual-base	MRR	0.117	0.156	0.168	0.178	0.189
	Recall@k	0.117	0.208	0.263	0.335	0.614
bge-m3	MRR	0.107	0.145	0.157	0.167	0.178
	Recall@k	0.107	0.197	0.249	0.324	0.612
KURE-v1	MRR	0.151	0.204	0.220	0.233	0.244
	Recall@k	0.151	0.275	0.342	0.441	0.739
Qwen3-0.6B	MRR	0.145	0.196	0.212	0.222	0.236
	Recall@k	0.145	0.262	0.330	0.413	0.744

Table 3: Performance of IR Models on Negation and Exclusion Queries.

IR Model	Metric	k				
		1	3	5	10	100
BM25	MRR	0.181	0.238	0.258	0.272	0.283
	Recall@k	0.181	0.314	0.401	0.507	0.792
gte-multilingual-base	MRR	0.197	0.259	0.277	0.291	0.303
	Recall@k	0.197	0.337	0.413	0.514	0.818
bge-m3	MRR	0.243	0.312	0.329	0.343	0.356
	Recall@k	0.243	0.401	0.476	0.584	0.874
KURE-v1	MRR	0.273	0.351	0.368	0.383	0.395
	Recall@k	0.273	0.454	0.530	0.638	0.899
Qwen3-0.6B	MRR	0.298	0.378	0.396	0.412	0.423
	Recall@k	0.298	0.481	0.560	0.677	0.929

Table 4: Performance of IR Models After Fine-Tuning.

cant improvement, with MRR and Recall@k at k=1 increasing from 0.107 to 0.243, and at k=100, MRR increased from 0.178 to 0.356, and Recall@k from 0.612 to 0.874. This indicates that even multilingual-based models can learn the semantic relationship between negation/exclusion queries and the correct documents to some extent through fine-tuning, and it shows that the triplet dataset created in this study works effectively.

The KURE-v1 model continued to show a high level of performance after fine-tuning. The MRR increased from 0.151 to 0.273 at k=1 and from 0.244 to 0.395 at k=100, while Recall@k improved significantly from 0.739 to 0.899. Since the KURE-v1 model was already fine-tuned on a Korean corpus, it appears to have secured even more robust performance by additionally learning the characteristics of negation/exclusion queries through further fine-tuning.

The Qwen3-0.6B model showed the best performance among all models after fine-tuning. All metrics that were lower than KURE-v1 before fine-tuning became higher than KURE-v1 after fine-tuning. At k=1, the MRR and Recall@k more than doubled from 0.145 to 0.298. At k=100, the MRR increased from 0.236 to 0.423, and Recall@k from 0.744 to 0.929.

In conclusion, the embedding models fine-tuned with the negation and exclusion-based case law retrieval dataset showed consistent improvement in both MRR and Recall@k metrics. This demonstrates that a training dataset specialized for queries with negation and exclusion conditions is effective in enhancing the semantic discrimination ability of embedding models.

5 Conclusion and Future Work

In this study, we constructed the 'Negation and Exclusion-Based Case Law Retrieval Dataset,' a dataset specialized for queries containing negation and exclusion conditions in the legal domain. The generated dataset is based on Supreme Court civil case precedents and is in a triplet format, consisting of queries with negation/exclusion conditions like "B, not A," a corresponding positive document (correct precedent), and a hard negative document (a precedent that is semantically similar to the query but does not satisfy the negation/exclusion condition). Through this dataset, we were able to confirm that existing IR models do not properly handle negative and exclusionary expressions. In terms of MRR and Recall@k, BM25 showed relatively good performance across all k value ranges, while embedding models showed significantly lower performance. This result aligns with the research by Weller et al. (2024), which stated that most IR models show random-level ranking performance on queries containing negative expressions (Weller et al., 2024).

After fine-tuning the embedding models with the dataset using a contrastive learning approach, the performance of all models consistently improved. Models pre-trained on a Korean corpus like KURE-v1 and Qwen3-0.6B, which showed excellent performance on multilingual sentence embedding benchmarks, achieved high results. Additionally, other embedding models such as gte-multilingual-base and bge-m3 also showed improved performance, enabling them to better reflect the negation and exclusion conditions.

Limitations

Our study is limited to the civil law of the Republic of Korea. Furthermore, we did not investigate whether training on data with negative or exclusion conditions impairs existing retrieval performance. Future research could expand the scope of legal fields covered and explore the feasibility of creating a general-purpose case law retrieval model by utilizing both our dataset and existing ones.

References

F. Barkhof, C. van den Elsen, T. Nijdam, S. Lupart, and M. Aliannejadi. 2025. Reproducing NevIR: Negation in neural information retrieval. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Padua, Italy. Association for Computing Machinery.

T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS) 33*, pages 1877–1901, Virtual Event. Curran Associates, Inc.

T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 1597–1607, Vienna, Austria. PMLR.

N. Craswell. 2009. Mean reciprocal rank. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 1703–1703. Springer US, Boston, MA.

Y. Hwang, S. Jung, H. Lee, and S. Yu. 2025. TWICE: What advantages can low-resource domain-specific embedding model bring? – a case study on korea financial texts. *Computing Research Repository*, arXiv:2502.07131.

C. Kim, J. Lee, and W. Hwang. 2025. Legalsearchlm: Rethinking legal case retrieval as legal elements generation. *Computing Research Repository*, arXiv:2505.23832.

G. Park and J. Kim. 2022. Legal search method using S-BERT. *Journal of the Korea Society of Computer and Information*, 27(11):57–66.

S. Robertson and H. Zaragoza. 2009. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

Z. Sun, K. Zhang, W. Yu, H. Wang, and J. Xu. 2024. Logic rules as explanations for legal case retrieval. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC--COLING 2024)*, pages 10747–10759, Torino, Italy. ELRA and ICCL.

O. Weller, D. Lawrie, and B. Van Durme. 2024. NevIR: Negation in neural information retrieval. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)*, pages 2274–2287, St. Julian's, Malta. Association for Computational Linguistics.

W. Zhang, M. Zhang, S. Wu, J. Pei, Z. Ren, M. de Rijke, Z. Chen, and P. Ren. 2025. ExcluIR: Exclusionary neural information retrieval. In *Proceedings* *of the 39th AAAI Conference on Artificial Intelligence*, pages 13295–13303, Philadelphia, PA, USA. AAAI Press.

A Case Law Examples

[Supreme Court, Decision 98Da23195, rendered May 12 2000]

【Issues】
[1] Whether a claim arising from a unilateral commercial act or an accessory commercial act is also included among the "commercial claims" subject to the five-year statute of limitations under Article 64 of the Commercial Act (answered in the affirmative).
[2] Case holding that a transferee's claim for registration of ownership transfer—arising from a landfill-site transfer agreement between a profit-making corporation engaged in landfill business and a transferee who is not a merchant—constitutes a commercial claim.
[3] Case finding that, where a landfill operator agreed to transfer part of the reclaimed land immediately after registering completion of the landfill works but did not stipulate who held the right of selection, the statute of limitations on the transferee's claim for ownership-transfer registration begins to run once (i) the operator's preservation registration of ownership over the reclaimed land is completed, (ii) an urban-planning decision and cadastral notice fix the location and area of the land, and (iii) a reasonable period necessary for the operator to exercise the selection right has elapsed.

【Holding / Reasoning】
[1] "Commercial claims" under Article 64 of the Commercial Act cover not only claims arising from acts that are commercial for both parties, but also claims arising from acts that are commercial for only one party. The term "commercial act" here includes both the basic commercial acts listed in Article 46 and the accessory commercial acts performed by a merchant in the course of business. Accordingly, such claims are subject to the five-year statute of limitations set out in Article 64.
[2] The transferee's claim for registration of ownership transfer based on a landfill-site transfer agreement, concluded between a profit-oriented corporation engaged in landfill business and a transferee who is not a merchant, constitutes a commercial claim.
[3] Where a landfill operator, after registering completion of the landfill works, promised to transfer part of the reclaimed land immediately but did not agree with the transferee on who held the right of selection, the statute of limitations on the transferee's ownership-transfer registration claim starts to run after a reasonable period has passed from the time when (i) the operator completes the preservation registration of ownership for the reclaimed land and (ii) an urban-planning decision and cadastral notice finalize the location and area of that land.

(rest omitted)

Figure A1: Example of a positive document for Negation/Exclusion Condition Query Example 1

[Supreme Court, Decision 66Da1741, rendered Nov 29 1966]

【Issues】
Circumstances in which joint-and-several liability is recognized under the Commercial Act.

【Holding / Reasoning】
When the defendants jointly operated a sock-manufacturing business and continuously purchased yarn on credit from the plaintiff, any unpaid balance constitutes a debt arising from the defendants' basic commercial acts. Accordingly, under the Commercial Act the defendants are jointly and severally liable to the plaintiff for payment of that debt.

(rest omitted)

Figure A2: Example of a negative document for Negation/Exclusion Condition Query Example 1

[Supreme Court, Decision 94Da14582, rendered 30 June 1995]

【Issues】
Whether the lower court properly found a fraudulent conveyance where an over-indebted debtor, acting in collusion with a single creditor, sold substantial assets without receiving any actual payment of the purchase price.

【Holding / Reasoning】
The debtor was already insolvent when, in collusion with one of his creditors, he sold key assets—factory buildings and adjoining land—to that creditor. In reality, no part of the purchase price was paid: a portion was offset against the creditor's existing claim, the creditor assumed the bank loan secured by the land, and the balance was deemed to cover a lease deposit because the debtor immediately rented the factory back for continued use. Even if the debtor intended to keep operating the business for economic recovery, and even if the stated price was roughly equivalent to market value, the transaction was a legal act performed with intent to prejudice the other creditors. The Supreme Court affirmed the lower court's determination that this constituted a fraudulent conveyance.

(rest omitted)

Figure A3: Example of a positive document for Negation/Exclusion Condition Query Example 2

[Seoul Central District Court, Decision 2014Gahap578263, rendered 10 Sept 2015]
Plaintiff
Korea Deposit Insurance Corporation, as bankruptcy trustee of Shilla Savings Bank Co., Ltd. (debtor in bankruptcy)
Defendants
1. A 2. B
Order
1. The sale contract dated **25 Mar 2010** between Defendant A and non-party C, concerning each parcel of real property listed in the attached schedule, is **revoked**.
2. Non-party C shall, with respect to each parcel of real property listed in the attached schedule, complete the following cancellation registrations:
a. **Defendant A** shall cancel the ownership-transfer registration filed with the Gwangju District Court on **26 Mar 2010**, Receipt No. 50357.
b. **Defendant B** shall cancel the ownership-transfer registration filed with the Gwangju District Court on **15 Mar 2011**, Receipt No. 44790.
3. Litigation costs shall be borne by the Defendants.
Relief Sought
Identical to the Order.
Reasons
(omitted)
2) Determination
In light of the evidence discussed above and the entire purport of the pleadings, the following circumstances are established:
① <u>The contract price was approximately **KRW 170 million** below the then-market value of the properties.</u> Even taking into account C's cash-flow difficulties and need for a quick sale, the price was substantially low.
② Defendant B and C's father have continuously resided on the properties.
③ Defendants have failed to produce the written sale contract for the first transaction; in particular, they have not clearly stated the sale price between Defendant A and Defendant B.
④ Defendant A asserts that she purchased the properties from C for **KRW 420 million**, yet there is no evidence that Defendant B paid any amount in excess of **KRW 400 million** when acquiring them from Defendant A. Thus Defendant A resold the properties for less than she paid, contradicting Defendants' explanation that the original price was below market solely because of C's financial distress.
Taken together, the evidence submitted is insufficient to prove that Defendants were **good-faith beneficiaries or subsequent transferees** unaware that the transactions were fraudulent conveyances prejudicing other creditors, and no other evidence supports such a finding.
D. Sub-conclusion
Accordingly, <u>the sale contract between Defendant A and C constitutes a fraudulent conveyance</u> and is revoked. As restitution, C must cancel the first ownership-transfer registration, and Defendant B must cancel the second ownership-transfer registration.
4. Conclusion
For these reasons, the Plaintiff's entire claim is well-founded and is granted as set forth in the Order.
(rest omitted)

Figure A4: Example of a negative document for Negation/Exclusion Condition Query Example 2

{

“Case No.”: “2023Da217534”,
“Case Summary”: “【Plaintiffs / Appellees】Plaintiff 1 et al. (Counsel: Seohwi Law Firm; Lead Attorney Kim Ik-Hyun and five others)
【Defendant / Appellant】Defendant (Counsel: Garam Law Firm and one other)
【Lower-Court Judgment】Seoul High Court, Judgment 2021Na2035828, rendered February 1 2023
【Order】All appeals are dismissed. Appeal costs are to be borne by the Defendant.
【Reasoning】The grounds of appeal (including any supplemental brief filed after the statutory deadline, insofar as it merely supplements the original grounds) are reviewed as follows.

(omitted)

4. Conclusion
Therefore, all appeals are dismissed and the costs of appeal are assessed against the losing party, in accordance with the unanimous opinion of the participating Justices.

Justices: Kim Seon-Su (presiding), Park Jeong-Hwa, Noh Tae-Ak, Oh Kyeong-Mi (opinion author)”

}

Figure A5: Data format Example

B Prompt Details

(Condition 1) The <Target Case-Law Data> must be able to serve as an answer to the given "negation/exclusion condition query." See the example below.

<Example Target Case-Law Data 1>
[Supreme Court, Decision 2003Da26020, rendered Sept 13 2004]
(omitted)

<Example of a Negation/Exclusion Condition Query for Example Target Case-Law Data 1>
"Except for cases in which an agreement prohibiting a cooperative member's withdrawal is held invalid, show me cases in which such a prohibition was at issue for an insolvent member."
The above <Example Target Case-Law Data 1> can serve as an answer to the <Example of a Negation/Exclusion Condition Query for Example Target Case-Law Data 1>.

(Condition 2) The answer to a "negation/exclusion condition query" must be able to include several concrete examples; in other words, the documents relevant to the query must not be limited to the <Target Case-Law Data> alone.

(Condition 3) The negation or exclusion condition in the query must substantively narrow the scope of the query. See the example below.

[Example 1]
For the <Example Target Case-Law Data 1>, the following <Incorrect Negation/Exclusion Condition Query Example 1> fails to satisfy (Condition 3).
<Incorrect Negation/Exclusion Condition Query Example 1>
"Excluding cases in which a member remains in good standing without bankruptcy, show me precedents where a bankrupt member was forced to stay because of a contractual withdrawal-ban clause."
Specific reason:
In that query—"(① Excluding cases where the member remains in good standing without bankruptcy); show me (② precedents where a bankrupt member was forced to stay because of a contractual withdrawal-ban clause)"—the negation/exclusion part (① a member remains in good standing without bankruptcy) is already inherently outside the category defined by (② a bankrupt member). Because the situations "bankrupt member" and "member not bankrupt" are mutually exclusive, the negation/exclusion condition does not actually narrow the scope.
For <Example Target Case-Law Data 1>, the <Correct Negation/Exclusion Condition Query Example 1> is as follows.
<Correct Negation/Exclusion Condition Query Example 1>
"Except for cases in which an agreement prohibiting a member's withdrawal is declared invalid, show me cases where such a withdrawal-ban agreement was applied to a bankrupt member."

[Example 2]
<Example Target Case-Law Data 2>
[Supreme Court, Decision 2018Da215756, rendered Sept 13 2018]
(omitted)
For <Example Target Case-Law Data 2>, the following <Incorrect Negation/Exclusion Condition Query Example 2> fails to satisfy (Condition 3).

<Incorrect Negation/Exclusion Condition Query Example 2>
"Excluding cases where the lease contract was concluded before the fraudulent conveyance, show me precedents in which the tenant's deposit under a new lease concluded after the fraudulent conveyance was not deducted from the amount to be restituted."
Specific reason:
In that query—"(① Excluding cases where the lease was concluded before the fraudulent conveyance); show me (② precedents concerning a new lease concluded after the fraudulent conveyance where the tenant's deposit was not deducted)"—the negation/exclusion part (① lease concluded before the fraudulent conveyance) is already inherently outside the category defined by (② new lease after the fraudulent conveyance). Because "before" and "after" the fraudulent conveyance cannot coexist, the negation/exclusion condition does not actually narrow the scope.
For <Example Target Case-Law Data 2>, the <Correct Negation/Exclusion Condition Query Example 2> is as follows.

<Correct Negation/Exclusion Condition Query Example 2>
"Excluding cases in which deposits with statutory priority (preferential repayment rights) were deducted, show me precedents where the court ordered monetary compensation equal to the property's value in an action to rescind a fraudulent conveyance."
(rest omitted)

Figure B1: Query generation prompt explaining negation and exclusion condition queries requirements

Tell me whether case0, case1, and case2 are correct documents (answers) or incorrect documents for the "Negation/Exclusion Condition Query" below, and give the reason for each decision.

Below is an example of the expected answer format:

```
[
  {"case_num": 0, "decision": true/false, "reason": "Why it is correct or incorrect"},
  ...
]
```

Negation/Exclusion Condition Query: {}

case0: {}

case1: {}

case2: {}

Figure B2: Final prompt for determining hard negative documents