MUSER: A Multi-View Similar Case Retrieval Dataset

Qingquan Li*
Yiran Hu*
{liqq20,huyr21}@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Feng Yao
yaof20@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Chaojun Xiao
xiaocj20@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Zhiyuan Liu†
liuzy@tsinghua.edu.cn
Tsinghua University
Beijing, China

Maosong Sun
sms@tsinghua.edu.cn
Tsinghua University
Beijing, China

Weixing Shen†
wxshen@tsinghua.edu.cn
Tsinghua University
Beijing, China

ABSTRACT

Similar case retrieval (SCR) is a representative legal AI application that plays a pivotal role in promoting judicial fairness. However, existing SCR datasets only focus on the fact description section when judging the similarity between legal cases, ignoring other valuable sections (e.g., the court's opinion) that can provide insightful reasoning process behind. Furthermore, the case similarities are typically measured solely by the textual semantics of the fact descriptions, which may fail to capture the full complexity of legal cases from the perspective of legal knowledge. In this work, we present MUSER, a similar case retrieval dataset based on multi-view similarity measurement and comprehensive legal element knowledge. Specifically, we select three perspectives (legal fact, dispute focus, and law statutory) and build a comprehensive and structured label system of legal elements for each of them, to enable accurate and knowledgeable evaluation of case similarities. The constructed dataset originates from Chinese civil cases and contains 100 query cases and 4,024 candidate cases. We implement several text classification algorithms for legal element prediction and various retrieval methods for retrieving similar cases on MUSER. The experimental results indicate that incorporating legal element labels can benefit the performance of SCR models, but further efforts are still required to address the remaining challenges posed by MUSER. The source code and dataset are released at https://github.com/THUlawtech/MUSER.

CCS CONCEPTS

• **Applied computing** → **Law**.

KEYWORDS

datasets, domain-specific, similar case retrieval

[*] Both authors contributed equally to this research.
[†] Corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA
© 2018 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Qingquan Li, Yiran Hu, Feng Yao, Chaojun Xiao, Zhiyuan Liu, Maosong Sun, and Weixing Shen. 2018. MUSER: A Multi-View Similar Case Retrieval Dataset. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

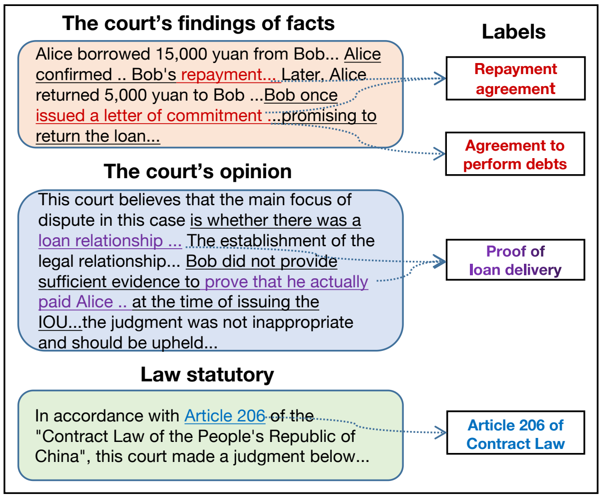
The court's findings of facts
Labels
Alice borrowed 15,000 yuan from Bob... Alice confirmed .. Bob's repayment... Later, Alice returned 5,000 yuan to Bob ...Bob once issued a letter of commitment ...promising to return the loan...
Repayment agreement
Agreement to perform debts
The court's opinion
This court believes that the main focus of dispute in this case is whether there was a loan relationship ... The establishment of the legal relationship... Bob did not provide sufficient evidence to prove that he actually paid Alice .. at the time of issuing the IOU...the judgment was not inappropriate and should be upheld...
Proof of loan delivery
Law statutory
In accordance with Article 206 of the "Contract Law of the People's Republic of China", this court made a judgment below...
Article 206 of Contract Law

Figure 1: The valuable sections of a judgment document.

1 INTRODUCTION

Similar case retrieval (SCR) is a vital legal AI [4, 19, 20] application that plays a strong role in achieving the consistent judgment for maintaining judicial fairness. Given a case, the goal of the SCR task is to retrieve similar cases from the candidate pool according to the judgment criteria, and judges refer to similar cases to assist in the current case judgment. Previous SCR works [3, 5, 6, 7, 11, 13, 16, 18] have achieved some success, [3, 18] measured the similarity of cases using textual information, [6] develop similar case measurement strategies considering both subjective and objective evaluation, [16] attend legal elements in cases through contrastive learning for SCR. However, two main challenges of existing SCR datasets have constrained the progress of the case retrieval models:

(1) Single-View. Existing SCR datasets [6, 13] only analyzed cases from the perspective of legal fact. However, as shown in Figure 1, in addition to the court's findings of fact, there are also sections such as "court's opinion" and "law statutory", which legal professionals are very concerned about. [14] As integral sections of cases, they are very valuable in SCR and should not be ignored. For example, As shown in Figure 1, the legal fact of the present case pertains to “repayment agreement”, which is similar to one of the candidate cases. However, the dispute focus of the candidate case concerns “limits on interest rates”, governed by “Article 205 of the Contract Law”, which is different from the case in Figure 1. Although the legal fact suggests a high similarity, the two cases diverge in terms of their focus on different aspects of the repayment agreement, which leads to distinct dispute focuses and law statutory. Consequently, it is not appropriate to classify these two cases as strongly similar. The limitations of previous works have led to an incomplete measurement of case similarity. To make the judgment of case similarity more comprehensive, it is particularly important to construct multi-view evaluation criteria for case similarity.

(2) Lack of legal knowledge. Existing SCR works [12, 18] only judge similar cases through semantic similarity. This makes it difficult to consider deep-level legal knowledge in SCR. In judicial practice, cases are very complex. For example, the legal effects of loans between couples, loans between employer and employee, and loans between relatives are completely different, and cannot be inferred solely from the text. Research in the law field [1] also attempts to design legal-element-based SCR methods for specific case types. Therefore, it is necessary to consider structural legal knowledge when retrieving similar cases.

To alleviate the above issues, we present MUSER, a similar case retrieval dataset based on multi-view similarity measurement with sentence-level legal element annotations. We highlight MUSER with the following advantages:

(1) Multi-View. To address the issue that current SCR works mainly focus on fact description, we propose dividing judgment information into three dimensions: legal fact, dispute focus, and law statutory. Each dimension provides a summary and description of the case from a unique perspective. Integrating multi-view information into SCR can comprehensively summarize all the information of the case and meet a wide range of SCR needs, including those of ordinary people, lawyers, and judges. Our hypothesis is confirmed by the Supreme People's Court's issuance of guidance documents that divide the definition of similar cases into three dimensions. [1]

(2) Fine-grained legal element label schema. As Figure 1 shows, we propose a sentence-level legal element label system for the legal fact and dispute focus to represent the deep legal knowledge implied in cases. Our legal expert team presents a three-level label schema, which contains 22 1st-level labels, 190 2nd-level labels, and 505 3rd-level labels. We annotated labels at the sentence level for both query and candidate cases. Following manual labeling, we trained a prediction model to automatically extract labels.

In addition, to our knowledge, previous SCR datasets mainly focus on criminal cases, whereas civil cases are more numerous and complex, necessitating an SCR dataset in this field. In this paper, we propose MUSER, a similar case retrieval dataset based on multi-view similarity measurement with sentence-level legal element annotations. Our dataset contains 100 query cases, each query has 100 candidates. All the cases in MUSER are adopted from Chinese Civil Law, published by the Supreme People's Court of China.

We implement several legal element prediction models and SCR models to explore the challenges of MUSER. Meanwhile, we use our legal element labels to design an SCR model. Experimental results show that the performance of our model is better than those baseline models, which provides our introduction of multiple views and legal knowledge in SCR is effective. The source code and dataset are available to the public at https://github.com/THUlawtech/MUSER.

2 DATASET CONSTRUCTION

In this section, we introduce the process of the dataset construction. Our goal is to construct a multi-view similar case retrieval dataset with sentence-level legal element annotation. Therefore, we need to define a label schema, select the query cases and develop relevant judgment criteria. In this paper, we focus on private lending cases, as they are the most complex and voluminous among all civil cases.

2.1 Label Schema Construction

To comprehensively construct the legal element schema that describes private lending cases, we propose two types of labels: legal fact and dispute focus. For each type of label, our legal expert team designs a large-scale, hierarchical three-level label schema, covering the legal elements that may appear in the vast majority of private lending cases. These legal elements mainly involve private lending contracts, private lending relationships, private lending amounts, private lending litigation, etc. Among them, the legal fact labels have 11 1st-level categories, 69 2nd-level categories, and 211 3rd-level categories; the dispute focus labels have 11 1st-level categories, 121 2nd-level categories, and 294 3rd-level categories. Details can be found in our GitHub link.

All labels are annotated at the sentence level. The legal fact labels are annotated in the “court's findings of fact” section of the judgment document, while the dispute focus labels are annotated in the “court's opinion” section.

2.2 Query and Candidate Selection

Following [6], we collect 100 query cases from 7000 cases published by the Supreme People's Court of China. Considering the diversity of label coverage, we sort the cases based on the number and types of labels and conduct uniform sampling. Since each case may cover multiple labels, we prioritize sampling cases with more label types in order to obtain a query set that covers more categories of labels.

For the candidate selection, we combine three strategies:

- `strategy 1`: Take the top 30 cases based on the cosine similarity of the text.
- `strategy 2`: Rank the document based on TF-IDF [10] and BM25 [9], and select the top 70 cases that appear in both rankings.
- `stratrgy 3`: If the candidate has less than 100 cases at this point, the leftover cases will be selected as the average of the two rankings.

2.3 Relevant Judgement Criteria and Annotation

As discussed above, we introduce 'multi-view' similarity measurement in our dataset. The relevance between query and candidate

[1] https://www.chinacourt.org/law/detail/2020/07/id/150187.shtml

Table 1: Case relevance criteria on different dimensions.

Dimension	Description	Relevance	Score
Legal fact	Key facts relevance, general facts relevance.	strong relevance	3
	Key facts relevance, general facts irrelevance.	relevance	2
	Key facts irrelevance, general facts relevance.	weak relevance	1
	Key facts irrelevance, general facts relevance.	irrelevance	0
Dispute focus	Identical key dispute focuses exist.	strong relevance	3
	Identical dispute focuses exist, but not key.	weak relevance	1
	No identical dispute focuses.	irrelevance	0
Law statutory	Identical articles are cited and have an impact on the judgement.	relevance	2
	Identical articles are cited but have no impact on the judgement.	irrelevance	0
	No identical articles are cited.	irrelevance	0

cases will be judged by the following three dimensions: legal fact, dispute focus and law Statutory.

For legal facts, we define two types: key facts and general facts. Key facts refer to facts that involve the plaintiff's requested rights, determine whether the case is a private lending case, and are typical facts in private lending cases. General facts are usually incidental facts other than key facts, such as the calculation of interest. For dispute focus, we also define two types: key dispute focus and general dispute focus. The key disputed focus includes the disputed focus generated around the plaintiff's core claims and the key facts of the case. The difference between key focus and general focus is whether they influence the plaintiff's basic claims and whether they involve the establishment of private lending. For the law statutory, the relevance judgment criteria are whether the same articles are cited and whether they have an impact on the judgment. Due to the large number of articles, our legal expert team selects 35 common key articles in private lending cases as the relevance judgment criteria. Table 1 shows in detail the case relevance judgment criteria on different dimensions. The candidates for each query will be sorted in descending order based on the sum of the three dimension relevance scores. Candidates with a relevance score of no less than 7 are considered relevant to the query.

Legal elements label and case relevance annotation experts are students in civil law from globally renowned law schools. They received a three-day training and trail annotation before the formal annotation. During annotation, we grouped the annotation experts, and the results of each group were regularly checked by the group leader. For case relevance annotation, we only reveal the legal element labels of the query case and hide them of the candidate cases, in order to help annotation experts better understand the query case and avoid the influence of legal element labels on the case relevance judgment.

3 DATA ANALYSIS

In this section, we aim to provide a deep understanding of MUSER through data analysis.

Table 2: Dataset statistics of MUSER.

Statistic	Number
Total documents	4,024
Total querys	100
Candidate cases per query	100
Avg. relevant cases per query	10.38

3.1 Data Size

Statistics. Table 2 shows the statistics of our dataset. We consider candidate cases with relevance scores to the query cases that are not less than 6 as relevant cases for the query.

Table 3: Statistics of text size of MUSER.

Document Section	Avg. Len.	Avg. #Token	Avg. #Sentence
The court's findings of fact	1,369.71	790.92	28.07
The court's opinion	1,629.34	900.94	27.89
Total	2,999.05	1,691.86	55.96

Text Size. Table 3 shows the text size of MUSER, including the average text length, average token number, and average sentence number. For token statistics, we utilize jieba [2] for tokenization. It is obvious that civil cases in MUSER are particularly long, which poses challenges for SCR, including extracting important information from long texts and encoding long texts with deep neural models.

Table 4: Statistics of legal element annotation of MUSER.

Label Type	Document Section	#Sentence	#Label	#Negative.
Legal Fact	The court's findings of fact	112,951	40,636	78,916
Dispute focus	The court's opinion	112,229	24,501	91,930

Legal Elements. Table 4 shows detailed statistics of legal element annotations of MUSER. Statistical results indicate the imbalance of our dataset, particularly in dispute focus, which poses challenges for legal element prediction.

3.2 Data Distribution

MUSER has data imbalance issues in legal element label distribution. For the 1st-level labels, 3/11 of legal fact labels and 2/11 of dispute focus labels account for more than 50% of the total instances. However, most legal elements still have sufficient instances, with more than 1000 instances for 8/11 of legal fact labels and 6/11 of disputed focus labels. Additionally, in actual judicial practice, the occurrence of different legal facts and disputed focuses follows a long-tail distribution, which proves that MUSER can serve as a real-world SCR dataset. Therefore, inspired by [15], we do not perform data augmentation or balancing during dataset construction.

4 EXPERIMENT

4.1 Legal Element Prediction

We evaluate several baseline legal element prediction models. We fine-tune the neural network models, including BERT [2] and Lawformer [17], as a multi-label classification task. We first encode the

[2] https://github.com/fxsjy/jieba

Table 5: Evaluation of the legal element prediction task.

Label Type	Model	Precision	Recall	F1
Legal Fact	**BERT**	**64.07**	56.43	60.01
	Lawformer	63.56	**59.65**	**63.10**
Dispute focus	**BERT**	**54.70**	41.50	46.86
	Lawformer	53.25	**44.62**	**48.83**

sentence with those deep neural network models, and then use a fully-connected layer for classification. We randomly sample 80% of the overall sentences as the train set, and the rest of them as the test set. We do not consider the hierarchy of our label schema and directly used the 3rd-level legal elements as the classification labels. We evaluate models on metrics including micro Precision, Recall, and F1. The legal element prediction results are shown in Table 5. The experimental results show that due to the complexity of the label schema and data imbalance, legal element classification is a challenging task. In addition, how to incorporate the hierarchical information between legal elements is a key focus of future work.

4.2 Similar Case Retrieval

Baseline Models. We evaluate several competitive retrieval models, which are widely used in the SCR task, on MUSER. Three types of retrieval models are involved, including (1) Traditional bag-of-words IR models. Following [6], we select BM25 [9], TF-IDF [10], and LMIR [8] as bag-of-words similar case retrievers. (2) Deep neural models. We consider Lawformer [17], a law-specific pre-trained language model capable of processing long texts, as the deep neural retrieval model. Specifically, we fine-tune Lawformer with a text pair relevance classification task. The task input contains the court's findings of fact and opinion of the query and candidate cases. A `[CLS]` token is added to the beginning of the input, and two `[SEP]` tokens are used, one to separate the query-candidate pair, and the other added to the end of the input. Global attention is added to `[CLS]` and query case. The hidden state vector of `[CLS]` outputted by Lawformer is fed into a fully-connected layer for relevance classification. Finally, the candidates are sorted in descending order based on the probability of being predicted as relevant to the query. (3) SCR models based on legal elements. Tradition IR models only consider textual information, lacking attention to professional legal concepts. Aiming at this issue, we present an SCR model based on our annotated legal elements including legal fact, dispute focus, and law statutory. Specifically, labels of a certain level in a specific dimension of a case are represented as a vector $\mathcal{L} = (l_1, \ldots, l_i, \ldots, l_n)$, where l_i is 1 when label i is annotated, otherwise 0. For label vectors of the same level and dimension in the query and candidate cases, we calculate their cosine similarity as the relevance score. For the three dimensions, we compute the weighted sum of the relevance scores by the number of labels for each level. Finally, the weighted sum of the relevance scores for three dimensions between query and candidate cases represents their final similarity score. Assigning different weights to different dimensions can adjust the degree of attention to them during similar case retrieval.

Settings. We test bag-of-words models and the legal-element-based model on the overall queries. To compare the deep neural model with other retrieval models, we randomly sample 80% of the queries as the train set for Lawformer, and test all baselines on the rest 20% queries as the test set. We evaluate baseline models on precision metrics including P@5, P@10, and Mean Average Precision (MAP), and ranking metrics including NDCG@10, NDCG@20, and NDCG@30.

Table 6: Evaluation of retrieval models on different query sets of MUSER. "N@(10/20/30)" is NDCG@(10/20/30), "LFM" is Lawformer, "LE" is our proposed legal-element-based model.

	Model	P@5	P@10	MAP	N@10	N@20	N@30
Overall Query Set	**BM25**	63.60	48.60	79.24	23.68	21.98	20.53
	TF-IDF	72.20	59.80	81.52	23.96	22.35	21.47
	LMIR	68.00	53.70	**84.40**	26.33	23.54	21.89
	LE	**77.20**	**65.50**	83.23	**28.96**	**26.02**	**24.51**
Test Set	**BM25**	78.00	53.50	91.76	21.80	19.54	17.48
	TF-IDF	80.00	63.50	85.23	20.61	18.30	17.85
	LMIR	**83.00**	63.50	**92.55**	28.57	24.43	22.04
	LFM	28.00	17.50	65.00	3.83	4.01	3.93
	LE	81.00	**71.50**	87.01	**31.82**	**27.01**	**25.29**

For baseline implementation, we use gensim [3] to implement BM25, TF-IDF, and LMIR, and set all parameters to default values. We fine-tuned Lawformer on 8 Nvidia RTX 2080Ti GPUs. Due to memory limitations, we set the maximum length of both query and candidate text to 600. For the legal-element-based model, we set the weights of legal fact, dispute focus, and law statutory to 0.5, 0.4, and 0.1, to attend more to the first two dimensions.

Results. The similar case retrieval results are shown in Table 6. We can observe that (1) Our proposed legal-element-based SCR model can outperform other baselines significantly, especially achieving comprehensive superiority in ranking metrics. We attribute this to our fine-grained annotated labels that can make the retrieval model focus on the key legal elements in the case, while other retrieval models only pay attention to the textual information. (2) The performance of the deep neural model is comparatively poor. We suspect that the possible reason is the truncation of case text causes the model to miss important information. (3) Overall, the civil case retrieval task is challenging, especially in terms of ranking metrics. Compared to criminal cases, civil cases have more complex legal relationships and diverse legal elements, making similar civil case retrieval more difficult. Future research is needed to explore SCR models specifically designed for civil cases.

5 CONCLUSION

In this paper, we propose MUSER, a similar case retrieval dataset for Chinese civil law systems. Compared with other SCR datasets that only consider legal facts, we propose a multi-view definition of similar cases that incorporates dispute focuses and law statutory. To incorporate legal knowledge into SCR, we design a large-scale label schema to represent the legal elements in the cases and conducted sentence-level annotation. The experimental results show the challenge in legal element prediction and civil SCR tasks, providing direction for future work.

In the future, we will continue to expand our data scale and optimize our legal element label schema. The latest resources will be updated to https://github.com/THUlawtech/MUSER.

[3] https://radimrehurek.com/gensim/

REFERENCES

[1] Layman E Alen. 1962. Beyond document retrieval toward information retrieval. *MINNESOTA LAW REVIEW*, 47, 713. https://core.ac.uk/download/pdf/72834475.pdf.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, 4171–4186. DOI: 10.18653/v1/n19-1423.

[3] Yoshinobu Kano, Mi-Young Kim, Masaharu Yoshioka, Yao Lu, Juliano Rabelo, Naoki Kiyota, Randy Goebel, and Ken Satoh. 2019. COLIEE-2018: Evaluation of the competition on legal information extraction and entailment. In *JSAI-isAI Workshops*. Springer, 177–192. https://link.springer.com/chapter/10.1007/978-3-030-31605-1_14.

[4] Qingquan Li, Qifan Zhang, Junjie Yao, and Yingjie Zhang. 2020. Event extraction for criminal legal text. In *Proceedings of ICKG*, 573–580. DOI: 10.1109/ICBK50248.2020.00086.

[5] Bulou Liu, Yiran Hu, Yueyue Wu, Yiqun Liu, Fan Zhang, Chenliang Li, Min Zhang, Shaoping Ma, and Weixing Shen. 2023. Investigating conversational agent action in legal case retrieval. In *European Conference on Information Retrieval*. Springer, 622–635. https://link.springer.com/chapter/10.1007/978-3-031-28244-7_39.

[6] 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 SIGIR*, 2342–2348. DOI: 10.1145/3404835.3463250.

[7] Shubham Kumar Nigam, Navansh Goel, and Arnab Bhattacharya. 2022. Nigam@COLIEE-22: Legal case retrieval and entailment using cascading of lexical and semantic-based models. In *JSAI International Symposium on Artificial Intelligence*. Springer, 96–108. https://link.springer.com/chapter/10.1007/978-3-031-29168-5_7.

[8] Jay M Ponte and W Bruce Croft. 1998. A language modeling approach to information retrieval. In *Proceedings of SIGIR*, 275–281. DOI: 10.1145/290941.291008.

[9] Stephen E. Robertson, Steve Walker, Micheline Hancock-Beaulieu, Mike Gatford, and A. Payne. 1995. Okapi at TREC-4. In *Proceedings of TREC*. http://trec.nist.gov/pubs/trec4/papers/city.ps.gz.

[10] Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24, 5, 513–523. DOI: 10.1016/0306-4573(88)90021-0.

[11] 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. *arXiv preprint arXiv:2012.13102*. https://arxiv.org/abs/2012.13102.

[12] Yunqiu Shao, Yueyue Wu, Yiqun Liu, Jiaxin Mao, and Shaoping Ma. 2023. Understanding relevance judgments in legal case retrieval. *ACM Transactions on Information Systems*, 41, 3, 1–32. DOI: 10.1145/3569929.

[13] Vu Tran, Minh Le Nguyen, and Ken Satoh. 2019. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In *Proceedings of ICAIL*, 275–282. DOI: 10.1145/3322640.3326740.

[14] David M Trubek. 1980. The construction and deconstruction of a disputes-focused approach: an afterword. *Law and society review*, 727–747. DOI: 10.2307/3053510.

[15] Xiaozhi Wang et al. 2020. MAVEN: A massive general domain event detection dataset. In *Proceedings of EMNLP*, 1652–1671. DOI: 10.18653/v1/2020.emnlp-main.129.

[16] Zhaowei Wang. 2022. Legal element-oriented modeling with multi-view contrastive learning for legal case retrieval. In *Proceedings of IJCNN*, 01–10. DOI: 10.1109/IJCNN55064.2022.9892487.

[17] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2, 79–84. DOI: 10.1016/j.aiopen.2021.06.003.

[18] Chaojun Xiao et al. 2019. CAIL2019-SCM: A dataset of similar case matching in legal domain. *arXiv preprint arXiv:1911.08962*. https://arxiv.org/abs/1911.08962.

[19] Feng Yao et al. 2022. LEVEN: A large-scale chinese legal event detection dataset. In *Findings of ACL*, 183–201. DOI: 10.18653/v1/2022.findings-acl.17.

[20] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How does nlp benefit legal system: a summary of legal artificial intelligence. In *Proceedings of ACL*, 5218–5230. DOI: 10.18653/v1/2020.acl-main.466.