Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

January 2025, Special Issue on AI 4 All- 2

Inferring organizational duties from Persian administrative and employment laws using Large Language Models (LLMs) and few-shot learning

Hojjat Hajizadeh Nowkhandan✉, Code ORCID: 0009-0005-2610-2149
Department of Computer Engineering Ferdowsi University of Mashhad Email: hajizadeh@mail.um.ac.ir
Mohsen Kahani, Code ORCID: 0000-0002-2603-6066
Department of Computer Engineering Ferdowsi University of Mashhad Email: kahani@um.ac.ir

Abstract—Extracting organizational duties from legal documents is a critical yet challenging task, particularly in low-resource languages like Persian. This paper presents an innovative approach that integrates state-of-the-art Named Entity Recognition (NER) with advanced segmentation techniques and Large Language Models (LLMs) to accurately identify and extract duties assigned to organizations from Persian legal texts. Leveraging the power of the BERT-based model for NER, we enhance the recognition of relevant entities and ensure precise linkage to target organizations. Our method involves segmenting documents into sentences with an enhanced POS-based tokenizer, followed by the retrieval of contextually relevant segments based on the detected entities. We then explore the effectiveness of different LLM configurations, including a hierarchical approach that leverages both small and large models. Our experiments demonstrate that the hierarchical approach, combining 'Llama-3.1-8B' and 'gpt-4o', achieves an F1-score of 0.7901, significantly outperforming single-model approaches. This research underscores the potential of LLMs in legal text analysis, paving the way for more advanced tools in Natural Language Processing. Future work will include testing on a broader range of organizations, refining prompt engineering techniques, and enhancing model interpretability.

Index Terms—NLP, Large Language Models, Few-shot Learning, Duty Extraction, Document Segmentation, Legal Informatics

1. Introduction

The application of Natural Language Processing (NLP) in the legal domain has gained significant attention in recent years. Legal and legislative texts, characterized by their complexity and precision, present unique challenges for information extraction and inferential tasks. In the context of administrative and employment laws, accurately inferring the duties assigned to the organizations is crucial for ensuring compliance and understanding the legal obligations.

This paper addresses the task of inferring organizational duties from Persian legal documents, leveraging advanced machine learning methods and Large Language Models (LLMs). The inherent complexity of legal texts in Persian, combined with the necessity for high accuracy in legal interpretations, imposes significant challenges. Furthermore, the diversity of legal sources and the need for standardization add to the complexity of the task. Existing NLP models often fall short in handling the nuances of the Persian language, requiring the development of more robust and accurate models.

Our proposed method involves several key steps to tackle these challenges. Initially, a Named Entity Recognition (NER) model is employed to detect organizations mentioned within the legal documents. The documents are then segmented to include the context around these mentions, capturing K sentences before and after the organization's name. Each segment is processed through a two-step LLM framework: the first LLM, a smaller model, performs the initial inference of duties, and the second, larger LLM evaluates the accuracy and quality of the inferred duties. This two-step process aims to enhance the precision and reliability of the inferred duties.

This study contributes to the field by providing a comprehensive framework for inferring duties from Persian legal documents, addressing key challenges such as language complexity, accuracy requirements, and standardization of legal rules. The results demonstrate the effectiveness of the proposed method, paving the way for future research and applications in legal NLP for low- resource languages like Persian.

The structure of this paper is as follows: Section 2 reviews related works in legal NLP and the challenges in this field. Section 3 details the proposed method, including the NER model, segmentation process, and the dual LLM framework. Section 4 presents the experimental results and evaluates the performance of the method. Finally, Section 5 concludes the paper and outlines potential directions for future research.

Submit Date: 2024-12-01
Accept Date: 2025-05-05
✉ Corresponding author

2. Literature Review

Comprehension of legal documents is an important issue in the field of law. The techniques utilized have ranged from relatively straightforward to highly complex,
as well as spanning from more established to the latest innovative methods. There are three main points of view in recent researches for understanding legal documents that we discuss them, here.

2.1. Legal Information Retrieval & Knowledge Extraction

Early studies in legal information retrieval primarily focused on keyword-based models to retrieve relevant documents. For instance, Tian et al. (2017) [1] used classification-based and rank-based methods for catchphrase extraction and precedence retrieval, employing classical IR models like BM25 and the vector space model. These methods, while effective, were limited by their reliance on predefined keyword sets and ranking functions.

Ontologies have also played a significant role in guiding information extraction from legal documents. Buey et al. (2016) [2] presented the AIS project, which leveraged ontologies to capture structural and linguistic conventions of legal texts. Their approach involved text preprocessing, chunking, and section processing, enabling accurate extraction from unstructured and semi-structured documents.

Recognizing the importance of document structure and layout, Garc´ıa-Constantino et al. (2017) [3] introduced CLIEL, a context-based information extraction system for commercial law documents. By utilizing NLP techniques, JAPE rules, and GATE modules, CLIEL effectively annotated and extracted data points, demonstrating the benefits of a layout-aware approach.

Entity and relation extraction has also been a critical area of research. Andrew (2018) [4] developed a method combining statistical methods like Conditional Random Fields [5] with rule-based techniques to automatically identify and annotate entities and their relationships in legal documents. This hybrid approach facilitated better information retrieval and knowledge discovery.

With the advent of deep learning, attention-based models significantly improved the representation of legal texts. Nguyen et al. (2024) [6] introduced attentive deep neural networks for legal document retrieval, proposing two hierarchical architectures, Attentive CNN and Paraformer, that utilized attention mechanisms to highlight important parts of sentences and articles. This approach enhanced retrieval performance by focusing on relevant information, with the Paraformer model achieving the highest recall and F2 scores.

Recent research by Hwang et al. (2022) [7] highlighted data-efficient end-to-end information extraction systems. Framing information extraction (IE) as a generation task, their system achieved competent scores with minimal training examples on Korean precedents, enhancing statistical analysis capabilities for legal practitioners. This flexible approach underscored the importance of scalable IE systems for diverse legal tasks.

2.2. Legal QA

Legal question answering (LQA) [8], [9] is the process of providing answers to legal questions. Usually, a lawyer or another legal professional with expertise and knowledge in the relevant area of law does this.

Early research in legal question answering systems began with retrieval-based models that emphasized effective text representation. Kien et al. (2020) [10] introduced a model that used neural attentive text representation for answering legal questions at the article level. This approach employed convolutional neural networks and attention mechanisms to accurately align questions with relevant legal articles, significantly outperforming previous methods in recall and NDCG metrics.

In subsequent years, advancements in deep learning further improved these systems. Khazaeli et al. (2021) [11] developed a versatile legal question answering system that combined sparse vector search, embeddings, and a BERT- based [12] answer re-ranking system. Their model trained on both general and legal domain data, enabled the system to accurately answer diverse legal questions without being limited to predefined patterns, and demonstrated its practical effectiveness in a commercial setting.

The integration of knowledge graphs into legal question answering marked another significant milestone. Thomas and Sangeetha (2022) [13] proposed a judicial knowledge graph-based system designed to support natural language queries, translating them into cypher queries to extract relevant information from the knowledge graph. This innovative approach improved the completeness and utility of case law analysis for legal professionals.

In the same year, Van et al. (2022) [14] achieved top rankings in the Automated Legal Question Answering Competition (ALQAC 2022) [15] by utilizing a deep learning approach based on the XLM-RoBERTa model. They fine-tuned the model with extensive legal text data and framed the task as a binary classification problem, effectively enhancing performance in legal document retrieval and question answering with minimal labeled data.

2.3. Textual Entailment

The COLIEE competition has been a focal point for advancements in legal information extraction and textual entailment, drawing attention to the complexities of processing legal texts. Nguyen et al. (2021) [16] explored deep learning methods in legal document processing for COLIEE 2021, highlighting both the potential and difficulties of these techniques in handling the intricate structure and semantics of legal texts. Their research emphasized the need for specialized approaches to tackle the unique challenges of legal language, showcasing the effectiveness and limitations of deep learning in this domain. In subsequent competitions, various teams employed innovative strategies to enhance

performance in legal information retrieval and entailment tasks. Kim et al. (2022) [17] utilized a combination of transformer-based methods, the BM25-ranking algorithm, and semantic thesaurus techniques for COLIEE 2021. Their approach achieved high rankings, demonstrating the effectiveness of integrating these methods to improve legal information retrieval and natural language inference in both case law and statute law tasks. Bui et al. (2022) [18] continued this trend by addressing the challenges of long and ambiguous legal documents in COLIEE 2022, proposing document-level attention mechanisms and passage mining techniques. Their methods aimed to handle the complexity and ambiguity inherent in legal texts, reflecting the competitive nature of the COLIEE tasks.

The competition in 2023 saw further advancements. Nguyen et al. (2024) [19] presented the Captain team's strategies for COLIEE 2023, leveraging state-of-the-art deep learning methods and meticulous engineering practices to excel in Tasks 2, 3, and 4. Their success in securing first places in multiple tasks underscored the importance of domain-specific observations and rigorous methodologies in processing legal texts. Vuong et al. (2024) [20] explored multi-task and ensemble approaches, employing models like BERT, Longformer, and BM25. Although their results were not state-of-the-art, their findings provided valuable insights and highlighted the potential for future improvements in legal information processing.

Bui et al. (2024) [21] introduced a novel approach combining data augmentation with a large language model for legal case retrieval and entailment. Their use of techniques such as back-translation and paraphrasing to enhance training data diversity, alongside a pre- trained language model on legal corpora, demonstrated significant improvements in task performance. This approach highlighted the importance of domain-specific pre- training and data augmentation in legal textual entailment.

3. Proposed System

In this section, we present the proposed system for extracting duties assigned to specific organizations from Persian legal documents. The methodology comprises three main steps: Named Entity Recognition (NER), segmentation and retrieval, and the use of Large Language Models (LLMs). The following flowchart (Figure 1) provides an overview of these steps:

3.1. Named Entity Recognition

The first step in our methodology involves the recognition and extraction of named entities from the legal documents. For this purpose, we employ a BERT-based model, *"HooshvareLab/bert-base-parsbert- ner-uncased"* [22] which has been specifically trained on Persian documents and has demonstrated an average F1 score of 0.95. The key reason for using the NER approach is to accurately identify and classify entities within the text, ensuring that we capture the relevant organizations without including sub-organizations that do not match our target list.

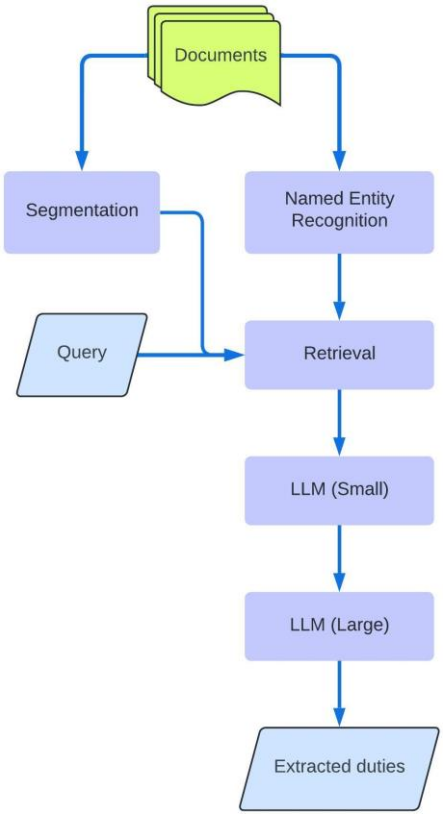

Figure 1: Flowchart of the proposed system.

3.1.1. Steps for Named Entity Recognition.

Running the NER Model:. We apply the *"HooshvareLab/bert-base-parsbert-ner-uncased"* model to the text to recognize named entities. This model, trained on Persian documents, ensures high accuracy in entity recognition. The recognized entities include names of organizations, sub-organizations, and other relevant entities within the text.

Post-processing Tasks:. After recognizing the named entities, we perform several post-processing tasks to refine the extracted entities:

- **Normalization:** We normalize the organization names using the Hazm [1] library to ensure consistency. This involves standardizing different forms of the same organization name to a single format.
- **Arabic to Persian Conversion:** Special Arabic letters are converted to their Persian equivalents. This step ensures uniformity in the text, addressing variations due to the use of Arabic script in Persian documents.
- **Merging Consecutive Named Entities:** We check for two consecutive named entities and merge them if the combined entity exists in our target list. This is crucial for correctly identifying entities that are split in the text, such as *"Ministry of Economy"* and *"and Finance."*
- **Separating Extracted Named Entities:** Where possible, we separate a single extracted named entity into its constituent parts to improve matching accuracy. For instance, if *"Ministry of Economy and Finance, Ministry of Agriculture"* is extracted as one entity but needs to be considered separately, we split it accordingly.

1. https://github.com/roshan-research/hazm

Linking Extracted Named Entities to Target Organizations:. To ensure that the recognized named entities correspond to our list of target organizations, we employ string matching algorithms:

- **String Matching with Jaccard Distance:** Initially, we use the Jaccard distance algorithm [23] for string matching. This method provides a recall of 83.5% and a precision of 98.4%. While fast, it does not achieve the desired level of recall.
- **Improved Matching with Gestalt Pattern Matching:** To enhance recall, we employ the Gestalt pattern matching algorithm. This more complex algorithm improves recall to 90.1% while maintaining a precision of 98.5%. The improved recall ensures that more relevant entities are correctly linked to the target organizations.

The NER step is fundamental to our methodology as it filters out irrelevant entities and ensures that only the named entities corresponding to our target organizations are considered in subsequent steps. This precise identification process prevents the inclusion of sub-organizations, thereby maintaining the accuracy of our duty extraction task.

3.2. Segmentation and Retrieval

In the second step of our methodology, we focus on segmenting the documents into sentences and retrieving relevant segments based on the presence of target organizations detected in the previous step. This process involves two main tasks: segmentation and retrieval.

3.2.1. Segmentation. To accurately process legal documents, it is essential to divide them into their constituent sentences. While the Hazm library provides a sentence tokenizer designed to split text into sentences based on punctuation marks like dots or question marks, this approach is often insufficient for legal texts. Legal documents frequently contain complex structures where punctuation alone does not reliably indicate sentence boundaries.

Improved Sentence Tokenization:. To enhance the accuracy of sentence segmentation, we utilize the Part-of-Speech (POS) tagger provided by the Hazm library. Instead of relying solely on punctuation marks, our approach identifies verbs as indicators of sentence boundaries. This method better handles the intricacies of legal language, where verbs often signify the conclusion of a sentence or clause.

The steps for improved sentence tokenization are as follows:

- **Apply POS Tagging:** We apply the Hazm POS tagger to the entire document to obtain the POS tags for each word.
- **Identify Sentence Boundaries:** Using the POS tags, we mark verbs as potential sentence-ending points. This method captures sentence boundaries more accurately than the simple punctuation-based tokenization.
- **Split Sentences:** Based on the identified boundaries, we split the document into sentences, ensuring that each segment represents a coherent and complete sentence.

3.2.2. Retrieval. Once the document is segmented into individual sentences, the next step is to construct contextual segments around the sentences that mention the target organizations. Each segment is defined as a window consisting of K sentences before and K sentences after the sentence containing the organization's name. This approach ensures that the relevant contextual information is preserved, which is essential for accurately extracting the duties assigned to the organization. The retrieval process is illustrated in Figure 2.

Creating Segments:. The steps for creating segments are as follows:

- **Detect Target Organization:** From the NER step, we identify sentences that contain the target organization. The target organization is specified in the user's query and is part of our predefined list of target organizations.
- **Define Window Size:** We define a window size K that determines the number of sentences before and after the target organization sentence to include in each segment.
- **Extract Segments:** For each sentence containing a target organization, we extract a segment that includes K sentences before and K sentences after the target organization sentence. This windowed approach ensures that we capture the necessary context for understanding the duties associated with the organization.

By segmenting the documents into sentences and retrieving relevant segments based on the presence of target organizations, we effectively prepare the text for the next step, which involves analyzing these segments using Large Language Models (LLMs). This segmentation and retrieval process ensures that we focus on the most relevant portions of the text, enhancing the accuracy and efficiency of our duty extraction methodology.

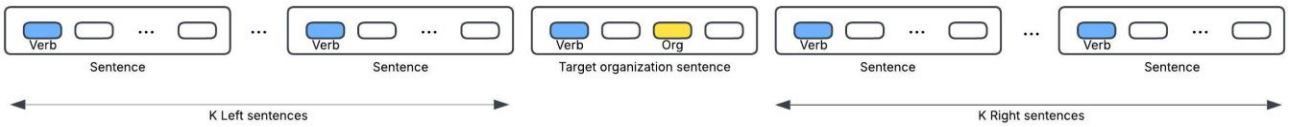

Figure 2: Illustration of the segmentation and retrieval process, where a window of K sentences before and after the organization's sentence is selected to preserve contextual information.

3.3. Use of Large Language Models (LLMs)

In the final step of our methodology, we leverage Large Language Models (LLMs) to extract the duties assigned to specific organizations from the segmented text. We explore four different approaches to accomplish this task, employing various models and configurations. In all these approaches, we use a system message to guide the behavior of the model and provide examples to perform few-shot learning.

Use of a Small Open Source Model:. We utilize the `Meta-Llama-3.1-8B-Instruct` [2] model in this approach, which has 8 billion parameters and a context length of 128k. The model is prompted with the following system message to extract the duties of the target organization:

```
You are a duty extraction system
  for Persian documents. Your
main task is to identify and
extract duties, which are specific
responsibilities that individuals
or organizations must fulfill. As
you analyze the text in Persian
paragraphs, you will look for
any mentions of these obligations
related to a specific organization.
It's important to note that a
paragraph may not contain any duties
for the input organization. When you
do find duties, you will extract
that information clearly
to help users understand what tasks
```

Use of the Small Open Source Model in Two Stages:. In this approach, we employ the `meta-llama/Llama-3.1-8B-Instruct` model in two stages. In the first stage, the model is tasked with extracting duties for all organizations in the input segment. In the second stage, it narrows down these duties to those specifically assigned to the target organization.

In this approach, the first stage system message is the same as the system message of the first approach (Use of a Small Open Source Mode).

Second stage system message:

2. https://github.com/meta-llama/llama-models

```
You are a knowledge interpreter
system in Persian documents. You
are able to detect whether a duty
is assigned to an organization or
not. The user will give you some
duties of a specific organization,
and you should infer which
duties are assigned to the given
organization.
```

Use of a Larger Commercial Model:. In this approach, we use the `gpt-4o` model, which has a context length of 128k. The model is prompted with the following system message to extract the duties of the target organization in JSON format

```
You are a duty extraction system
for Persian documents. Duty is a
task that must be performed by a
person or organization that has
a responsibility to perform it.

You extract duties in JSON format
for a specific organization in the
paragraphs if there are any.
```

Use of Small and Large Models Hierarchically:. In this hierarchical approach, we first use the small model (`Llama-3.1`) to extract the duties of the target organization. The results are then fed into the larger model to review, correct, and verify the output.

First prompt for the small model:

```
You are a duty extraction system
for Persian documents. Duty is a
task that must be performed by a
person or organization that has
a responsibility to perform it.

You extract duties for a specific
organization in the paragraphs if
there are any.
```

Second prompt for the large model:

```
You are a duty extraction system
for Persian documents. Duty is a
task that must be performed by a
person or organization that has a
responsibility to perform it. The
user has extracted some duties for
a specific organization himself,
which are not necessarily correct.
So you go through them and first
extract the other duties that he
didn't mention, then remove the
duties that are not assigned to the
input organization or are not
correct. Finally, you output the
assigned duties (in JSON format)
for a specific organization in the
paragraphs, if there are any.
```

By employing these four approaches, we aim to maximize the accuracy and comprehensiveness of the extracted duties assigned to the target organizations. Each method leverages the capabilities of LLMs differently, allowing us to assess their performance and suitability for the task at hand.

4. Experiments and Results

In this section, we present the experimental setup and results of our proposed method. We report the precision, recall, and F1-score for three of the four approaches discussed previously. The results for the first and second approaches are very close, so we only report the results of the first approach. The K parameter in all approaches is set to 4. We use 2 shots (2 examples) in our few-shot learning, one containing a segment (2*K + 1 sentences) without any assigned duties for the target organization and another containing a segment with some assigned duties for the target organization, to help the LLMs understand that segments may not necessarily have duties for the target organization. For the larger (commercial) model, we used the latest OpenAI model `gpt-4o-2024-08-06` available at the time of writing, accessed via their APIs. The temperature parameter was set to 0.01 for the small model and 0 for the large model, while the top_p parameter was set to 1 for both models.

4.1. Dataset

This study utilizes a dataset of over 17,300 Persian legal documents, with our approaches tested on a select subset of segments retrieved from these documents. All datasets are publicly available and can be downloaded by anyone who wishes to reproduce the results. The primary sources of the dataset include:

- Laws, regulations, and resolutions obtained from the website of the Research Center of the Parliament. This source constitutes the main portion of the dataset due to its extensive collection of documents related to organizations and their duties. Many of these laws are directly relevant to administrative and employment laws.
- Legal documents available on the website of the Legal Deputy of the Presidency.
- Laws and regulations published on the official newspaper website of the Islamic Republic of Iran.

To ensure the quality and compatibility of the data for further processing, we applied several preprocessing steps, including:

- Detecting and removing duplicate laws to maintain a clean and non-redundant dataset.
- Eliminating special characters that do not affect sentence meaning but might disrupt data storage and processing.
- Standardizing numerical formats by converting English digits to Persian digits.
- Correcting spacing in punctuation marks and affixes to ensure uniform formatting.
- Normalizing letter repetitions by reducing instances where a letter appears more than twice to exactly two occurrences, in accordance with Persian linguistic rules.

4.2. Evaluation Metrics

We define the evaluation metrics as follows:

- True Positive (TP): The number of extracted duties in segments that are actually assigned to the target organization.
- False Positive (FP): The number of extracted duties by our system that are not assigned to the target organization but are recognized as such by the system.
- False Negative (FN): The number of duties assigned to the target organization in the retrieved segments that our system failed to identify.

Using these values, we calculate precision, recall, and F1-score with the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

4.3. Results

The experimental results show that the hierarchical approach using both small and large LLMs achieves the best performance, as detailed in Table 1.

We compare our results with the closest research in the legal domain, particularly Task 2 of the COLIEE 2023

TABLE 1: Comparison of our approaches with the best performing teams in COLIEE 2023 Task 2.

Team/Method	F1-Score	Recall	Precision
Llama-3.1-8B + gpt-4o (best approach)	**0.8387**	**0.8966**	0.7879
gpt-4o (third approach)	0.7089	0.8750	0.5957
Llama-3.1-8B (first approach)	0.4340	0.7931	0.2987
CAPTAIN Team [19]	0.7456	0.7083	0.7870
THUIR Team [24]	0.7182	0.6583	**0.7900**
JNLP Team [21]	0.6818	0.6250	0.7500

competition [25]. Task 2 focuses on case law entailment, where the objective is to determine which paragraphs of a relevant case entail a specific fragment of a base case. This task is similar to our problem as it involves entailment detection in legal texts, albeit in different languages and contexts.

The top-performing systems in Task 2 of COLIEE 2023 achieved the following results:

- **CAPTAIN [19]:** This team used the monoT5 model fine-tuned with hard negation mining and ensemble techniques, achieving a precision of 0.7870, recall of 0.7083, and an F1-score of 0.7456.
- **THUIR Team [24]:** This team employed contrastive learning on pre-trained models, combined with traditional retrieval models like BM25 and QLD [26], achieving a precision of 0.7900, recall of 0.6583, and an F1-score of 0.7182.
- **JNLP Team [21]:** This team used an ensemble of transformer models with different loss functions to compute similarity scores, achieving a precision of 0.7500, recall of 0.6250, and an F1-score of 0.6818.

Our proposed hierarchical approach, utilizing small and large LLMs in a sequential framework, achieved a precision of 0.7879, recall of 0.8966, and an F1-score of 0.8387. These results surpass the best-performing teams in Task 2 of COLIEE

2023, particularly in terms of F1-score, demonstrating the effectiveness of our method in the complex task of extracting organizational duties from Persian legal texts. While the COLIEE datasets are extensively researched in English and Japanese, achieving such performance on less-resourced Persian datasets highlights the potential of our approach.

Our study illustrates the robustness of combining advanced segmentation and hierarchical configurations for legal text analysis, paving the way for further advancements in multilingual legal NLP.

4.4. Quality of Extracted Duties

To further analyze our method, we examined the quality of extracted duties in terms of relevance. We classified extracted duties into three categories:

- **Relevant**: Duties correctly assigned to the target organization (equivalent to precision).
- **Irrelevant**: Duties incorrectly assigned to the target organization.
- **Non-duty**: Extracted content that does not constitute a valid duty.

The results are presented in Figure 3.

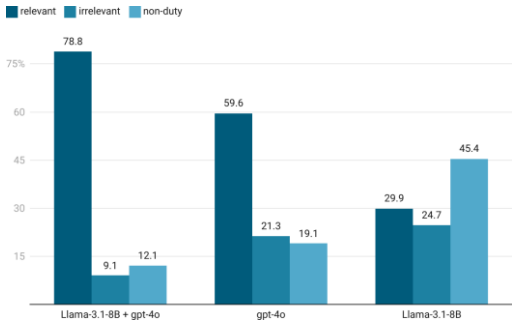

Figure 3: The quality of extracted duties by each approach.

These results indicate that our best approach (Llama-3.1- 8B + gpt-4o) achieves the highest proportion of relevant duties (78.8%), with relatively low irrelevant (9.1%) and non-duty extractions (12.1%). In contrast, using only Llama- 3.1-8B results in a much lower relevant rate (29.9%) and a significantly higher non-duty extraction rate (45.4%).

4.5. Granularity of Extracted Duties

Another aspect of evaluation is the granularity of extracted duties. We categorized duties into two types:

- **Divisible**: Duties that can be split into two or more sub-duties.
- **Non-divisible**: Duties that are atomic and cannot be further divided.

We excluded non-duty extractions from this analysis.

The results are shown in Figure 4.

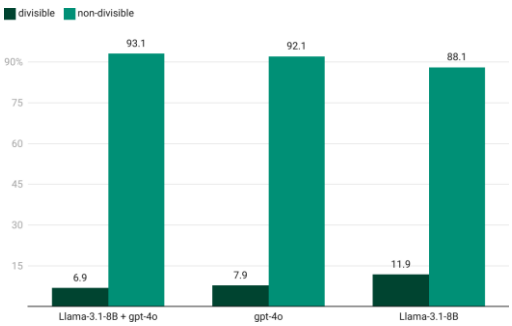

Figure 4: The granularity of extracted duties by each approach.

Our best approach produced the highest percentage of non-divisible duties (93.1%), indicating a more atomic extraction of duties. Conversely, Llama-3.1-8B alone yielded a significantly higher proportion of divisible duties (11.9%), suggesting that its extracted duties often encompass multiple responsibilities that should be split further.

4.6. Interpretation of Results

Our best approach, combining Llama-3.1-8B and gpt-4o, not only achieves superior precision, recall, and F1-score but also excels in extracting high-quality and granular duties. The significantly higher proportion of relevant extractions in our best approach confirms that hierarchical processing reduces false positives. Additionally, the reduced percentage of divisible duties suggests that our approach extracts more atomic and well-structured duties, making it more suitable for practical applications in legal and administrative NLP tasks.

Comparing our results to the COLIEE 2023 Task 2 competition, our approach demonstrates competitive performance. The best-performing system, CAPTAIN, achieved a precision of 0.7870, recall of 0.7083, and an F1- score of 0.7456. In comparison, our method outperforms CAPTAIN in recall

and F1-score, indicating its superior ability to handle complex entailment tasks in Persian legal texts. Similarly, it surpasses THUIR and JNLP in F1-score, showcasing the robustness of our methodology even in the context of less-resourced languages.

These findings emphasize the potential of hierarchical configurations with LLMs for extracting organizational duties, paving the way for more accurate and reliable legal document analysis in Persian. Our approach sets a strong benchmark for further research in multilingual legal NLP.

5. Conclusion and Future Works

In this paper, we presented a novel approach for extracting organizational duties from Persian legal documents using advanced NLP techniques and large language models (LLMs). Our methodology involved Named Entity Recognition (NER), segmentation and retrieval, and the use of LLMs to infer and extract relevant duties. The experiments demonstrated that our hierarchical approach, combining small and large LLMs, achieved the best performance in terms of precision, recall, and F1-score.

The results indicate that our approach is effective in handling the complexities of Persian legal texts and can significantly aid in automating the extraction of organizational duties. This work also highlights the challenges of working with low-resource languages like Persian, where the availability of annotated datasets and pre- trained models is limited compared to languages like English and Japanese.

5.1. Future Works

To further enhance the effectiveness and robustness of our method, we propose the following potential directions for future research:

- **Test results on a wider range of organizations:** Expanding the dataset to include a broader array of organizations will help evaluate the generalizability
  of our approach and identify any domain-specific challenges.
- **Explore various prompts through prompt engineering:** Testing different prompt formulations can help optimize the performance of LLMs, potentially leading to better extraction accuracy and understanding of the context.
- **Incorporate explanations for LLM inferences:** Asking LLMs to explain their inference processes and provide justifications for detecting each duty can enhance transparency and trust in the system, as well as provide insights for further refinement.
- **Increase the number of examples in few-shot learning:** Adding more examples (shots) in the few- shot learning approach and comparing the results can provide a deeper understanding of the model's learning capabilities and the impact of additional context on performance.

In conclusion, our research demonstrates the potential of leveraging advanced NLP and LLM techniques for extracting organizational duties from legal texts in Persian. By addressing the outlined future work directions, we aim to further refine and improve our approach, ultimately contributing to more efficient and accurate legal document analysis tools.

References

[1] L. Tian, H. Ning, L. Kong, Z. Han, R. Xiao, and H. Qi, "Hljit2017@ irled-fire2017: Information retrieval from legal documents.," in *FIRE (Working Notes)*, pp. 82–85, 2017.

[2] M. G. Buey, A. L. Garrido, C. Bobed, and S. Ilarri, "The ais project: Boosting information extraction from legal documents by using ontologies.," in *ICAART (2)*, pp. 438–445, 2016.

[3] M. Garc´ıa-Constantino, K. Atkinson, D. Bollegala, K. Chapman,
F. Coenen, C. Roberts, and K. Robson, "Cliel: context-based information extraction from commercial law documents," in *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*, pp. 79–87, 2017.

[4] J. J. Andrew, "Automatic extraction of entities and relation from legal documents," in *Proceedings of the Seventh Named Entities Workshop*,
pp. 1–8, 2018.

[5] J. Lafferty, A. McCallum, F. Pereira, *et al.*, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Icml*, vol. 1, p. 3, Williamstown, MA, 2001.

[6] H.-T. Nguyen, M.-K. Phi, X.-B. Ngo, V. Tran, L.-M. Nguyen, and M.-
P. Tu, "Attentive deep neural networks for legal document retrieval,"
Artificial Intelligence and Law, vol. 32, no. 1, pp. 57–86, 2024.

[7] W. Hwang, S. Eom, H. Lee, H. J. Park, and M. Seo, "Data-efficient end-to-end information extraction for statistical legal analysis," *arXiv preprint arXiv:2211.01692*, 2022.

[8] B. Fawei, J. Z. Pan, M. Kollingbaum, and A. Z. Wyner, "A semi- automated ontology construction for legal question answering," *New Generation Computing*, vol. 37, no. 4, pp. 453–478, 2019.

[9] A. Morimoto, D. Kubo, M. Sato, H. Shindo, and Y. Matsumoto, "Legal question answering system using neural attention.," *COLIEE@ ICAIL*, vol. 2017, pp. 79–89, 2017.

[10] P. M. Kien, H.-T. Nguyen, N. X. Bach, V. Tran, M. Le Nguyen, and T. M. Phuong, "Answering legal questions by learning neural attentive text representation," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 988–998, 2020.

[11] S. Khazaeli, J. Punuru, C. Morris, S. Sharma, B. Staub, M. Cole,
S. Chiu-Webster, and D. Sakalley, "A free format legal question answering system," in *Proceedings of the Natural Legal Language Processing Workshop 2021*, pp. 107–113, 2021.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[13] A. Thomas and S. Sangeetha, "Knowledge graph based question- answering system for effective case law analysis," in *Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021)*, pp. 291–300, Springer, 2022.

[14] H. N. Van, D. Nguyen, P. M. Nguyen, and M. Le Nguyen, "Miko team: Deep learning approach for legal question answering in alqac 2022," in *2022 14th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1–5, IEEE, 2022.

[15] C. Nguyen, M.-Q. Bui, D.-T. Do, N.-K. Le, D.-H. Nguyen, T.-

T. Nguyen, H.-T. Nguyen, V. Tran, L.-M. Nguyen, N.-C. Le, *et al.*, "Alqac 2022: A summary of the competition," in *2022 14th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1–5, IEEE, 2022.

[16] H.-T. Nguyen, P. M. Nguyen, T.-H.-Y. Vuong, Q. M. Bui, C. M. Nguyen, B. T. Dang, V. Tran, M. L. Nguyen, and K. Satoh, "Jnlp team: Deep learning approaches for legal processing tasks in coliee 2021," *arXiv preprint arXiv:2106.13405*, 2021.

[17] M.-Y. Kim, J. Rabelo, K. Okeke, and R. Goebel, "Legal information retrieval and entailment based on bm25, transformer and semantic thesaurus methods," *The Review of Socionetwork Strategies*, vol. 16, no. 1, pp. 157–174, 2022.

[18] Q. M. Bui, C. Nguyen, D.-T. Do, N.-K. Le, D.-H. Nguyen, T.-T.-T. Nguyen, M.-P. Nguyen, and M. L. Nguyen, "Jnlp team: Deep learning approaches for tackling long and ambiguous legal documents in coliee 2022," in *JSAI International Symposium on Artificial Intelligence*, pp. 68–83, Springer, 2022.

[19] C. Nguyen, P. Nguyen, T. Tran, D. Nguyen, A. Trieu, T. Pham, A. Dang, and L.-M. Nguyen, "Captain at coliee 2023: efficient methods for legal information retrieval and entailment tasks," *arXiv preprint arXiv:2401.03551*, 2024.

[20] T.-H.-Y. Vuong, H.-L. Nguyen, T.-M. Nguyen, H.-T. Nguyen, T.-B. Nguyen, and H.-T. Nguyen, "Nowj at coliee 2023: Multi-task and ensemble approaches in legal information processing," *The Review of Socionetwork Strategies*, vol. 18, no. 1, pp. 145–165, 2024.

[21] M.-Q. Bui, D.-T. Do, N.-K. Le, D.-H. Nguyen, K.-V.-H. Nguyen, T. P. N. Anh, and M. Le Nguyen, "Data augmentation and large language model for legal case retrieval and entailment," *The Review of Socionetwork Strategies*, vol. 18, no. 1, pp. 49–74, 2024.

[22] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "Parsbert: Transformer-based model for persian language understanding," *Neural Processing Letters*, vol. 53, pp. 3831– 3847, 2021.

[23] P. Jaccard, "E´ tude comparative de la distribution florale dans une portion des alpes et des jura," *Bull Soc Vaudoise Sci Nat*, vol. 37, pp. 547–579, 1901.

[24] H. Li, W. Su, C. Wang, Y. Wu, Q. Ai, and Y. Liu, "Thuir@ coliee 2023: Incorporating structural knowledge into pre-trained language models for legal case retrieval," *arXiv preprint arXiv:2305.06812*, 2023.

[25] R. Goebel, Y. Kano, M.-Y. Kim, J. Rabelo, K. Satoh, and M. Yoshioka, "Overview and discussion of the competition on legal information, extraction/entailment (coliee) 2023," *The Review of Socionetwork Strategies*, vol. 18, no. 1, pp. 27–47, 2024.

[26] C. Zhai, "Statistical language models for information retrieval," in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Tutorial Abstracts*, pp. 3–4, 2007.