Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 2, May 2024, pp. 1081~1091
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i2.pp1081-1091

Enhancing Moroccan legal cases analysis through ontology-based information extraction

Kaoutar Belhoucine, Nadia Zame, Mohammed Mourchid
Laboratory of Research in Informatics (LaRI), Faculty of Science, Ibn Tofail University, Kénitra, Morocco

Article Info

Article history:

Received Nov 28, 2023
Revised Feb 3, 2024
Accepted Feb 4, 2024

Keywords:

Arabic text
Moroccan legal analytics
OBIE
Rule-based extraction
Statistical extraction

ABSTRACT

The efficient organization of diverse disorder cases within a unified memory necessitates an adaptable representation. This study introduces an ontology-based approach for extracting facts from Moroccan legal cases. Leveraging ontological frameworks, a comprehensive case architecture is established, enabling advanced information extraction. Utilizing rules, patterns, and knowledge modeling harmonizes cases and identifies pervasive legal concepts. Statistical techniques unveil latent entities within complex legal textual discourse. Empirical validation demonstrates proficiency, extracting up to 25 regular entities. The rule-based mechanism achieves an F1-score of 99.5%, highlighting precision, while the statistical extractor achieves an 88.3% F1-score, revealing concealed entities. This work presents an innovative ontology-based paradigm for legal information extraction, contributing to advanced knowledge management in the legal domain.

Corresponding Author:

Kaoutar Belhoucine
Laboratory of Research in Informatics (LaRI), Faculty of Sciences, Ibn Tofail University
Kenitra, Morocco
Email: kaoutar.belhoucine@gmail.com

1. INTRODUCTION

The rapid digitization of court judgments and legal cases has generated a vast and ever-growing repository of legal knowledge in Arabic. This treasure trove of information presents both formidable challenges and immense opportunities for advancing artificial intelligence (AI) systems. While the sheer volume and complexity of Arabic legal texts pose significant obstacles to automated processing, the potential benefits of harnessing this rich resource are transformative. AI systems equipped with the ability to extract legal information from Arabic texts effectively could revolutionize legal research, case-based reasoning, and other AI-powered legal applications [1].

In early 2023, the Moroccan justice portal Adala [2] launched a groundbreaking digitization initiative, meticulously transcribing Moroccan legal cases into a digital format. This significant achievement has unlocked new possibilities for research and innovation in the legal domain. Despite adhering to a standardized format, the cases in modern standard arabic (MSA) reveal distinctive linguistic features and typically span two to five pages, containing complex structural and linguistic markers that judicial personnel must navigate to understand and find relevant cases quickly. One approach to address this need is to extract legal information from Moroccan legal cases using automated methods [3], [4].

Researchers have explored a variety of techniques for legal text information extraction, including:

– Rule-based approaches: these methods rely on rules and language techniques for information extraction. The effectiveness of rule-based systems depends on the quality of handcrafted rules and patterns. For instance, Kowsrihawat *et al.* [5] used regular expressions to summarize Thai supreme court cases. Moens *et al.* [6] employed paragraph classification and sentence analysis to extract data like dates and

case names from Belgian legal texts. Zhuang *et al*. [7] utilized regular expressions and feature dictionaries to extract basic case details from Chinese judgment documents.

- Machine learning approaches: machine learning models can extract more complex information than rule-based systems but require substantial labeled data for training. For instance, Bach *et al*. [8] extracted key information from Vietnamese transport legal texts using conditional random fields (CRF). Sil and Roy [9] introduced a novel machine-learning approach for classifying legal clauses from the Indian Penal Code based on support vector machines (SVM). Jasim *et al*. [10] proposed a framework for automatically detecting and identifying argumentative components in Arabic legal texts, employing SVM and random forest algorithms.
- Deep learning models: deep learning models have revolutionized natural language processing (NLP), surpassing traditional machine learning methods in various tasks. Son *et al*. [11] demonstrated their effectiveness to identify logical patterns in Vietnamese legal datasets. They experimented with recurrent neural network models, including long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM), and their combination with CRFs. Similarly, Chalkidis and Kam-pas [12] devised a system for extracting contract elements using a deep learning model, specifically the Bi-LSTM model, which integrated word embeddings, part-of-speech tags, and token shape embeddings. Nuranti and Yulianti [13] assessed deep learning techniques on Indonesian legal texts, such as CNN, LSTM+CRF, and Bi-LSTM+CRF. Notably, the Bi-LSTM+CRF model achieved the highest accuracy in their study.
- Knowledge-driven approaches: these methods leverage the power of knowledge resources, such as ontologies, to improve precision and efficiency in extracting legal information from texts. By incorporating domain-specific knowledge, they reduce the dependence on expensive labeled data and enhance interpretability. For instance, Buey *et al*. [14] employed ontologies and keyword-based rules to identify relevant information in Spanish legal documents. Araujo *et al*. [15] constructed a domain ontology of legal events to extract information such as formal charges and convictions from Brazilian legal texts. Ren *et al.* [16] proposed an ontology-based and deep learning-driven method for extracting legal facts from Chinese legal texts. Thomas and Sangeetha [17] adopted a semi-supervised, knowledge-integrated pattern-learning approach to extract domain-specific facts from judicial text.

Legal information extraction (LIE) from Arabic legal texts presents significant challenges owing to the intricacies of the Arabic language [18], the limited availability of labeled data, and the fragmented nature of legal systems in Arabic-speaking countries. Machine learning (ML) systems are not well-suited for LIE in Arabic due to their substantial data requirements for training. Instead, we propose utilizing an ontology and rule-based system to extract legal information from Moroccan legal cases in Arabic.

Our proposed system capitalizes on the Moroccan legal case ontology (MLCO) to organize legal cases into predefined logical segments and to extract common entities such as individuals, organizations, and dates. The MLCO also serves as the foundation for identifying a vector of associated concepts for each case by applying a statistical method. The key contributions of our study include:

- MLCO: a novel framework addressing both the structural and knowledge modeling aspects of Moroccan legal cases.
- Annotation module: meticulously designed for the legal documents on the Adala portal. This module processes legal cases in PDF format, and generates structured XML files enriched with tags for common and specific entities, ensuring accurate and comprehensive annotation.
- Linguistic resources on the NooJ [19] platform include an Arabic dictionary derived from the MLCO and morphological and syntactic grammars tailored to the intricacies of the Arabic legal language.
- Empirical evaluation: a rigorous assessment of our annotation module's performance, offering valuable insights for the research community.

The remaining sections of the paper are organized as follows: Section 2 sheds light on established methodology. Section 3 assesses the performance of the proposed methodology and discusses the results. Finally, the paper's conclusion is in section 4.

2. METHOD

We employed a rigorous methodology to efficiently extract relevant information from Moroccan legal cases. This involved a thorough examination of legal case structures and linguistic markers, which was pivotal in understanding the nuances of Moroccan jurisprudence. Additionally, the analysis identified key patterns and variations within the legal documents, informing the development of specialized modules for information extraction.

2.1. Corpus description

Our main source of data is the Adala portal database, which houses Moroccan legal cases [2]. We focused specifically on cases from the cassation court, curating a subset of 13,259 cases that cover various legal domains. This selection aligns perfectly with our ontology-based approach, providing the elaborate legal reasoning required for analysis.

Apprehending the structure of these cases was paramount. In close collaboration with legal experts, we identified five key meta-elements that actually capture the logical segments within court of cassation cases:

- Header: includes case number, type, release date, folder number, and keywords.
- Principle: represents the essence of the case, encapsulated abstractly.
- Background/Context: outlines the case's factual backdrop, capturing events leading to the dispute, a precis of the inferior court's judgment, and the challenged decision.
- Legal reasoning: this element bifurcates into:
  a. Grounds for cassation: highlights the factual areas of contention raised by the appellant.
  b. Response to the grounds: showcases the judicial counterarguments and legal references underpinning the verdict.
- Conclusion: encompasses the crux and a brief declaration of the judgment's nature.

We identified these elements by analyzing linguistic markers, which can manifest as sentences or trigger words. These markers act as crucial indicators, signaling the presence of generalized entities or specific entities related to the legal subdomain under consideration. Figure 1 provides an example of how these markers help us understand legal cases.

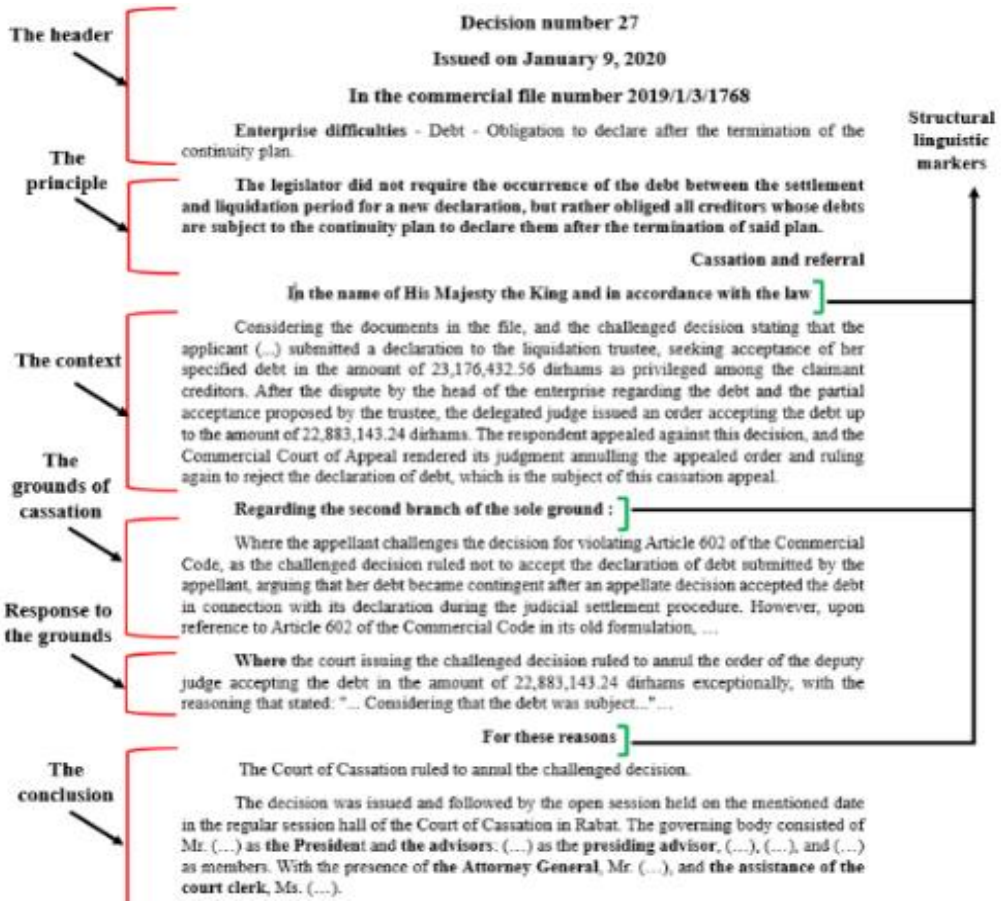


Figure 1. Legal case about insolvency proceedings

2.2. Knowledge modeling

Ontologies have emerged as powerful tools for extracting domain-specific information from textual data. Their ability to explicitly represent domain knowledge makes them adaptable across different fields, supporting system updates. The effectiveness of ontologies is widely acknowledged in the research, as seen

in studies [20], [21], which have applied ontologies to various domain-specific information extraction tasks. This trend extends to the judicial domain, as demonstrated by prior research [14]–[17]. In our specific context, we have developed a knowledge-driven system that leverages the MLCO in conjunction with a set of linguistic rules to extract domain-specific information from Moroccan legal cases. This subsection provides an overview of our approach to constructing the MLCO framework and developing the associated extraction patterns.

2.2.1. Ontology construction

The development of domain ontologies adheres to established methodologies, such as DEF5 [22], TOVE [23] and the Stanford seven-step [24] methods. These methodologies typically include core steps like determining the scope, acquiring knowledge, conceptualizing, implementing, and evaluating the ontology. In our research, we employed the Stanford seven-step method to develop the MLCO. Our approach to knowledge acquisition and conceptualization was top-down, involving collaboration with legal experts to adapt the ontology, initially proposed by [16], to the Moroccan legal context.

As a result, the MLCO (shown in Figure 2) consists of two primary components: general case ontology and statuary ontologies. The general ontology encompasses structural concepts and common entities encountered across legal cases. Conversely, statuary ontologies are customized to capture concepts and relationships derived from laws (statutes and regulations) and are unique to particular legal domains. For instance, concepts like "Court," "Parties," and "Governing body" are universally present in the cases and, therefore, included in the general ontology. In contrast, terms such as "Insolvency procedure," "Liquidation," and "Settlement" are exclusive to Insolvency law and are represented within a dedicated statuary ontology for Insolvency law. Mathematically, Let O be the MLCO, O_g, and O_s represent the general case ontology and the set of the specific statutory ontologies, respectively. Then,

$$O = Og \cup Os \quad (1)$$

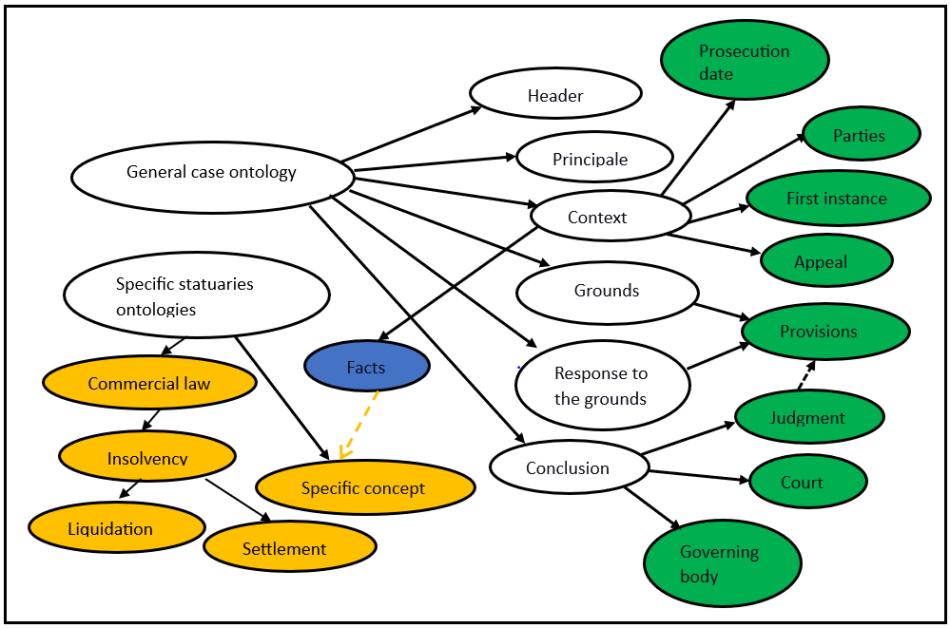


Figure 2. Schematic representation of the MLCO framework

2.2.2. Rule/Pattern identification

The standardized structure of Moroccan legal cases simplified the process of defining regular expressions to extract common information from texts. We identified a set of pattern, leveraging characteristics such as keywords, punctuation, position, and recursivity, for capturing the general entities specified during ontology development. Their performance was rigorously validated using a representative sample of 100 legal cases. Table 1 presents the patterns identified for extracting logical segments. Table 2 outlines the patterns identified for general entities.

Table 1. Linguistic characteristics of logical segments

Logical segment	Start marker	End marker	Recursiviy
Header	Case number (القرار عدد)	Punctuatin	No
The principle	Punctuation	Punctuatin	No
The context	In the name of his majesty the king (باسم جلالة الملك)	Punctuatin	No
Grounds	Grounds regarding the means/whereby it is lamented (في شأن الوسيلة / حيث ينعى)	Punctuatin	Yes
Responses	Responses however, since it was reasoned (لكن، حيث عللت / حيث قضت)	Punctuatin	Yes
The conclusion	The conclusion for these reasons/for these (لأجله، لهذه الأسباب)	Punctuatin	No

Table 2. Trigger words of the identified general entities

Segment	Entity	Trigger
Header	Case no.	Case number (القرار عدد)
	Release date	Issued on the date (الصادر بتاريخ)
	Folder no.	Number (عدد)
	Type	In the file (في الملف)
Context	Prosec. date	Filed an appeal with the court on (تقدم بتاريخ)
	1st Inst. court	An appeal to the court (مقالا إلى المحكمة / بمقال...المحكمة / بمقال للمحكمة)
	Appel court	The judgment issued by...Appeal (الصادر...عن استئناف)
	Appel no.	Appeal number/under appeal number (رقم / تحت رقم)
	Appel date	On the date (بتاريخ)
Grounds/Responses	Provisions	By decree/article...from (بمرسوم / مادة...من / فصل...من)
Conclusion	Result	With rejection/with annulment (برفض / بنقض)
	Court	Decided by (قضت)
	Advisors	Advisor (عضو)
	Judge	Chief Juge (رئيس)
	Rapporteur	Rapporteur (مقرر)
	Attorney gen.	In the presence of the public prosecutor's attorney (وبمحضر المحامي العام)
	Court clerk	With the assistance of the court clerk (وبمساعدة كاتب الضبط)

2.3. Annotation module

The annotation module, depicted in Figure 3, comprises a service that orchestrates a variety of linguistic resources to convert legal cases from PDF format into structured XML files enriched with tags for common and specific entities. This conversion process is required for further analysis and machine learning tasks, as it transforms unstructured legal text into a format that information systems can easily process. Additionally, the enriched tags provide valuable context about the legal entities, permitting more accurate and efficient information retrieval and analysis.

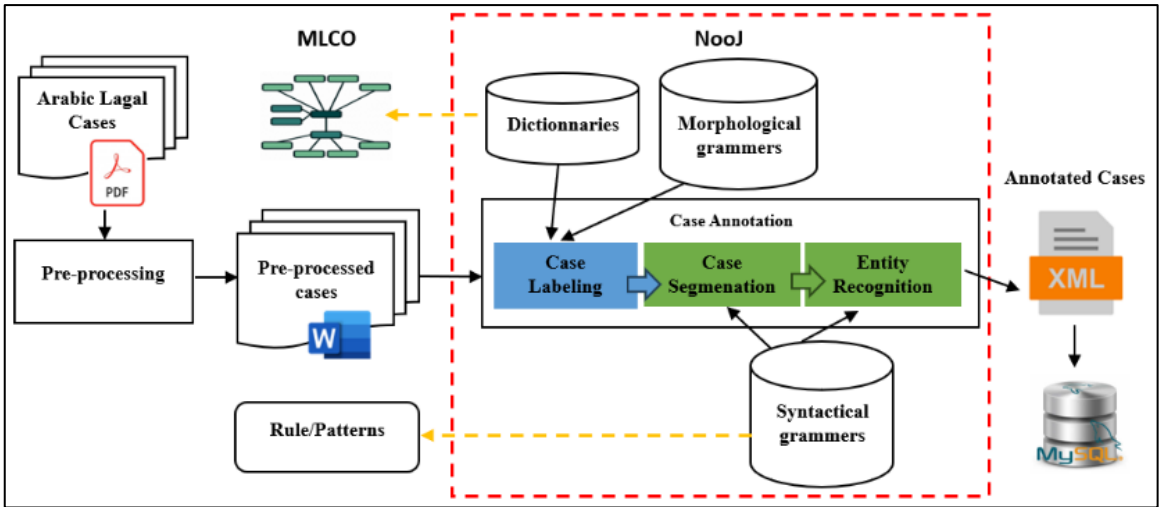


Figure 3. The annotation module

2.3.1. Pre-processing

During this stage, all PDF cases were converted to word format and normalized for easier manipulation. This process involved suppressing special characters and applying replacement rules to certain frequently misspelled Arabic letters. For instance, the replacement of the Hamzated forms of Alif (أ, إ) with

Alif (ا), the replacement of non-Alif forms of Hamza (ئ، ؤ) with the Hamza letter (ء), the replacement of Ta-Marbuta (ة) with Ha (ه), and the replacement of Alif-Maqsura (ى) with Ya (ي).

2.3.2. Dictionaries

We developed a custom Arabic-centric dictionary using the NooJ platform, employing the extensive terminology provided by the MLCO. This specialized dictionary covers specific concepts from the Arabic legal domains, trigger words associated with common entities and trigger words associated with structural concepts. NooJ's dictionary, as shown in Figure 4, integrates inflectional and derivational paradigms for each lexical entry, enabling a nuanced comprehension of each word's morphological, and syntactic facets.

تَصْفِيَة,N+FLX=FlxF+DRV=DrvF
قَاضِي,ADJ+FLX=FlxM+DRV=Qodat:Flx
مِسْطَرَة التَّسْوِيَة,N+N+PREF+N+FLX=FlxNC

Figure 4. Sample of dictionary entries

2.3.3. Morphological grammers

Based on the work of [25], morphological grammars have been modeled to recognize the constituent morphemes of agglutinative forms in Arabic. For instance, the morphological grammar in Figure 5 allows identifying the agglutinative word, including the prefixes (definite article (the, ال)), (prepositions (for, ل), (by, ب)), and (conjunctions (and, و)), and the suffix (possessive adjectives (his/her/its, ه)). This approach enables a comprehensive analysis of Arabic linguistic structures, allowing a deeper interpretation of the legal text.

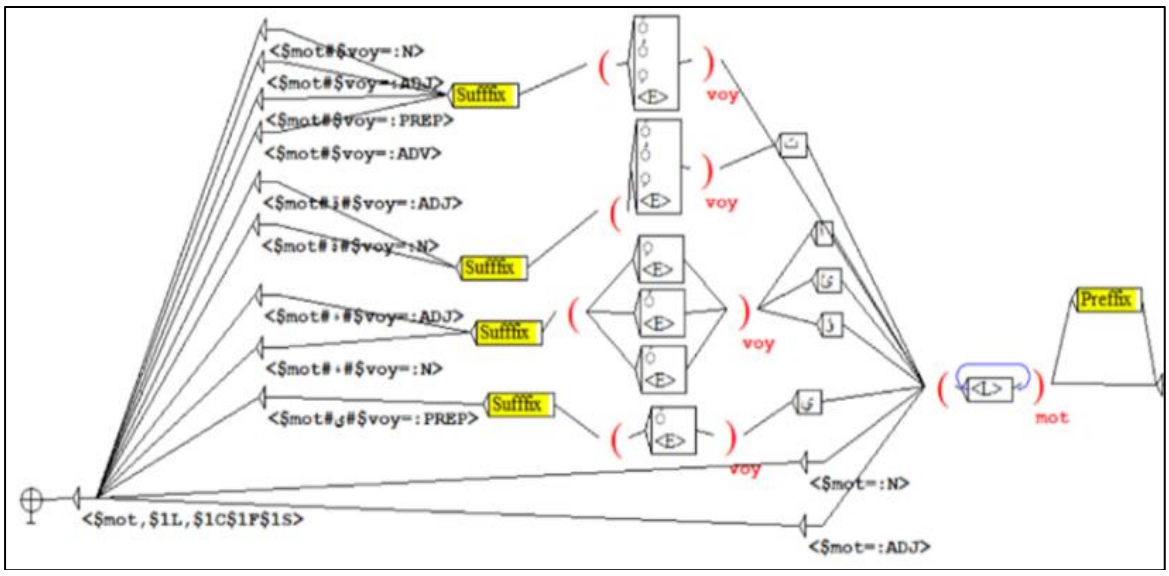


Figure 5. Arabic morphological grammar in NooJ

2.3.4. Syntactic grammers

Subsequent to the identification of regular expressions (subsection 2.3.2), we constructed corresponding syntactic grammars using the NooJ linguistic platform. Figure 6 depicts the main syntactic grammar, enabling logical segment identification and entity recognition. It comprises five nested grammars: header, context, grounds, response, and conclusion. Among these, grounds and response are recurrent grammars. Figure 7 provides an example of the conclusion grammar, which permits determining the logical segment "conclusion" and identifying all its embedded regular entities, such as the judge, general antony, and the judgment.

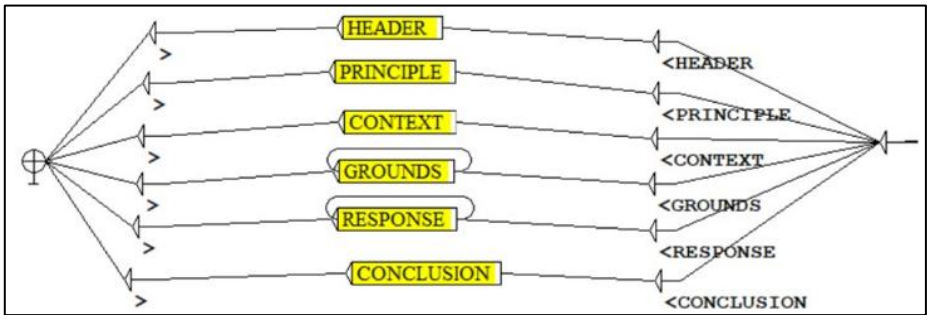


Figure 6. The main syntactic grammar

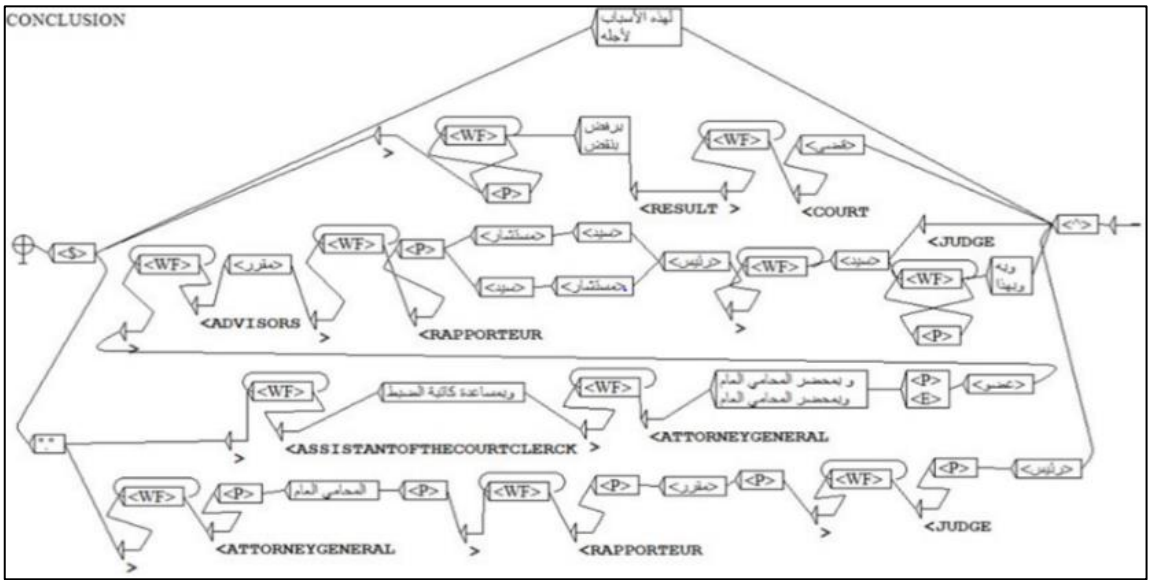


Figure 7. The conclusion syntactic grammar

2.3.5. Case annotation

After the case preprocessing stage, the annotation module embarks on the following steps:

- Case labeling: employing the NooJ dictionaries, morphological grammar, and linguistic analysis, the system meticulously labels the trigger words and specific entities within the case text.
- Case segmentation: utilizing the main syntactic grammar, each case is carefully dissected into well-defined sections, ensuring precise subsequent labeling.
- Entity recognition: tailored to each section, corresponding syntactic grammars identify and categorize distinct entities, encompassing organizations, dates, persons, locations, and specialized legal concepts.
- XML generation: structured XML files are generated for each case, encapsulating essential elements and entities to facilitate subsequent analysis and reasoning tasks.

2.4. Statistical service

To present each legal case with a list of associated specific and pertinent entities, we adopted a statistical method inspired by [26]. This method utilizes the XML files generated by the annotation module to construct a ”concept fingerprint” for each case. The concept fingerprint is derived by calculating concept relevance based on two key criteria.

- Concept frequency in the case: the relevance of a concept can be linked to its frequency of occurrence within the document. Concepts that appear more often might be considered more crucial to the document’s main subject.
- Concept depth in the statuary ontologies: the ontology structures knowledge within specific domains. A concept’s depth within such an ontology reflects its specificity. Concepts situated deeper in the ontology hierarchy tend to be more specific, suggesting a deeper or niche subject in a particular topic.

In practical applications, this method can rank concepts by considering their document frequency and their depth in the ontology. The highest ranked concepts are then selected for inclusion in the concept fingerprint. This method ensures that the fingerprint accurately captures the text's most pertinent and informative entities. Mathematically, we denote the case fingerprint as f_k, which consists of a set of specific concepts:

$$f_k = \{C_1, C_2, \dots, C_n\} \quad (2)$$

where, k is the number of the case, C is an extracted specific concept.

3. RESULTS AND DISCUSSION

The performance of our system was rigorously evaluated using two carefully curated datasets, as depicted in Figure 8. The first dataset, referred to as dataset 1, consisted of 500 cases spanning from 2016 to 2023. These meticulously selected cases ensured a diverse range of legal issues. They were targeted to assess the rule-based extractor's ability to identify logical segments and common entities.

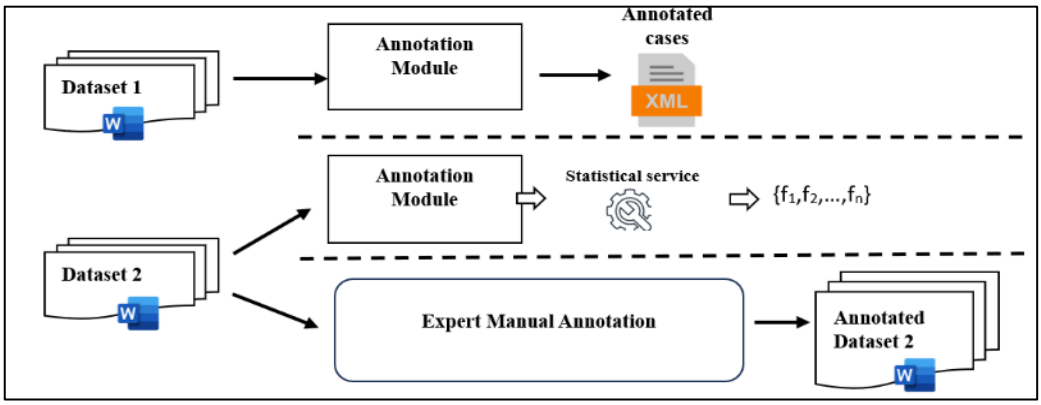


Figure 8. Diagram illustrating the testing strategy

The second dataset, known as dataset 2, was designed to evaluate the performance of the statistical module. It consisted of 140 cases related to insolvency law, which were categorized into three groups: settlement proceedings (40 cases), liquidation proceedings (30 cases), and creditor claims (70 cases). The topics were chosen due to their centrality to the insolvency law ontology developed for Morocco [27]. The cases had been previously manually annotated by a legal expert, providing a valuable baseline for comparison.

In this testing strategy, we employed a combination of three metrics: precision, recall, and F1-score, to measure the accuracy and relevance of the extracted information.

$$\text{Precision} = \frac{\text{No.of legal information correctly extracted}}{\text{Total no.of legal information extracted by the system}}$$

$$\text{Recall} = \frac{\text{No.of legal information correctly extracted}}{\text{Total.no of actual legal information}}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.1. Analysis on dataset 1

The rule-base extractor demonstrated impressive capabilities in identifying legal information, achieving an average precision of 99.96%, recall of 99.07%, and F1-score of 99.5% see in Table 3. However, certain challenges remain:

– The rule-based extractor's accuracy was particularly high for legal entities such as the court, case number, release date, folder number, and folder type.

- Shortcomings were encountered in extracting some legal entities, namely the court of first instance, the court of appeal, and assistance of the court clerk, due to their occasional absence in the data.
- The exact match strategy employed miscategorized a few logical segments during paragraph classification. This strategy also impacted the extraction of legal provisions and the prosecution date.

3.2. Analysis on dataset 2

For dataset 2, the focus was on evaluating the effectiveness of the statistical approach. The primary objective was to assess the consistency between the automatic concept assignments made by the system and the manual annotations performed by a legal expert. Concepts deeply integrated into the MLCO ontology and frequently appearing in the cases were considered highly relevant. The results are summarized in Table 4.

The precision rates align with the existing literature [16], [26], ontology-driven methods incorporating statistical techniques have demonstrated high precision in various domain. The exhaustive representation of the selected topics in insolvency law within the developed ontology has further strengthened this accuracy.

Table 3. Results of the rule-based extractor on the judicial dataset

No.	Legal entities	Precision	Rappel	F1-score
1	Header	100	100	100
2	The principle	100	100	100
3	The context	100	100	100
4	The grounds	99.72	99.24	99.47
5	Response to the grounds	99.66	99.55	99.6
6	The conclusion	100	100	100
7	Case number	100	100	100
8	Release date	100	100	100
9	Folder number	100	100	100
10	Folder type	100	100	100
11	Date of prosecution	99.72	97.44	98.56
12	The court of first instance	100	98.66	99.32
13	The court of appeal	100	98.75	99.37
14	Case number of appel	100	98.68	99.33
15	Release date of appel	100	98.68	99.33
16	Legal provision number	100	100	100
17	Legal provision name	100	97.2	98.58
18	Judgment result	100	100	100
19	Court	100	100	100
20	Chief Juge	100	100	100
21	Committee	100	95	97.43
22	Advisors	100	97.5	98.73
23	Rapporteur	100	97.5	98.73
24	Attorney general's	100	100	100
25	Assistance of the court clerk	100	98.57	99.27
	Average	99.96	99.07	99.5

Table 4. Comparison of annotation precision

Topics	Settlement	Liquidation	Creditor claims	Total (%)
Annotation precision rate (%)	86.3	87.6	91.2	88.3

4. CONCLUSION

The research successfully establishes a pioneering ontology-based paradigm, particularly for the Moroccan legal context. The proposed system combines rule-based and statistical techniques to overcome the challenges posed by the diverse and complex nature of legal discourses. Although the exact match strategy had some drawbacks, the overall results demonstrate the feasibility and utility of automated legal information extraction in Arabic, suggesting many avenues for future improvement: Enhanced logic segmentation: refine the rule set and explore adaptive segmentation to address rule-based extractor issues. Expanded ontologies: given the success of the insolvency law ontology, develop ontologies for other Moroccan legal domains. Adaptive learning models: improve statistical extractor with adaptive learning models that evolve based on new data, making the system more resilient to growing legal terminology. Cross-jurisdictional analysis: adapt and test system in Arabic legal contexts outside the Moroccan scope to gauge adaptability and performance in diverse environments.

REFERENCES

[1] S. Goźdź-Roszkowski, "Corpus linguistics in legal discourse," *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, vol. 34, no. 5, pp. 1515–1540, Nov. 2021, doi: 10.1007/s11196-021-09860-8.

[2] EUCTR2018-001013-32-ES, "ADALA," *https://trialsearch.who.int/Trial2.aspx?TrialID=EUCTR2018-001013-32-ES*, 2018.

[3] S. Sharma, S. Gamoura, D. Prasad, and A. Aneja, "Emerging legal informatics towards legal innovation: current status and future challenges and opportunities," *Legal Information Management*, vol. 21, no. 3–4, pp. 218–235, Dec. 2021, doi: 10.1017/S1472669621000384.

[4] K. Belhoucine, M. Mourchid, S. Mbarki, and A. Mouloudi, "A bottom-up approach for Moroccan legal ontology learning from Arabic texts," in *Communications in Computer and Information Science*, vol. 1389, 2021, pp. 230–242. doi: 10.1007/978-3-030-70629-6_20.

[5] K. Kowsrihawat, P. Vateekul, and P. Boonkwan, "Predicting judicial decisions of criminal cases from Thai supreme court using bi-directional gru with attention mechanism," in *Proceedings of the 5th Asian Conference on Defence Technology, ACDT 2018*, IEEE, Oct. 2018, pp. 50–55. doi: 10.1109/ACDT.2018.8592948.

[6] M. F. Moens, C. Uyttendaele, and J. Dumortier, "Information extraction from legal texts: the potential of discourse analysis," *International Journal of Human Computer Studies*, vol. 51, no. 6, pp. 1155–1171, Dec. 1999, doi: 10.1006/ijhc.1999.0296.

[7] C. Zhuang *et al.*, "Information extraction from Chinese judgment documents," in *Proceedings - 2017 14th Web Information Systems and Applications Conference, WISA 2017*, IEEE, Nov. 2017, pp. 240–244. doi: 10.1109/WISA.2017.67.

[8] N. X. Bach, L. T. N. Cham, T. H. N. Thien, and T. M. Phuong, "Question analysis for Vietnamese legal question answering," in *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, Oct. 2017, pp. 154–159. doi: 10.1109/KSE.2017.8119451.

[9] R. Sil and A. Roy, "A novel approach on argument based legal prediction model using machine learning," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, IEEE, Sep. 2020, pp. 487–490. doi: 10.1109/ICOSEC49089.2020.9215310.

[10] K. Jasim, A. T. Sadiq, and H. S. Abdullah, "A framework for detection and identification the components of arguments in arabic legal texts," in *1st International Scientific Conference of Computer and Applied Sciences, CAS 2019*, IEEE, Dec. 2019, pp. 67–72. doi: 10.1109/CAS47993.2019.9075650.

[11] N. T. Son, L. M. Nguyen, H. B. Quoc, and A. Shimazu, "Recognizing logical parts in legal texts using neural architectures," in *Proceedings - 2016 8th International Conference on Knowledge and Systems Engineering, KSE 2016*, IEEE, Oct. 2016, pp. 252–257. doi: 10.1109/KSE.2016.7758062.

[12] I. Chalkidis and D. Kampas, "Deep learning in law: early adaptation and legal word embeddings trained on large corpora," *Artificial Intelligence and Law*, vol. 27, no. 2, pp. 171–198, Jun. 2019, doi: 10.1007/s10506-018-9238-9.

[13] E. Q. Nuranti and E. Yulianti, "Legal entity recognition in indonesian court decision documents using Bi-LSTM and CRF approaches," in *2020 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2020*, IEEE, Oct. 2020, pp. 429–434. doi: 10.1109/ICACSIS51025.2020.9263157.

[14] 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 2016 - Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, SCITEPRESS - Science and and Technology Publications, 2016, pp. 438–445. doi: 10.5220/0005757204380445.

[15] D. A. de Araujo, S. J. Rigo, and J. L. V. Barbosa, "Ontology-based information extraction for juridical events with case studies in Brazilian legal realm," *Artificial Intelligence and Law*, vol. 25, no. 4, pp. 379–396, Dec. 2017, doi: 10.1007/s10506-017-9203-z.

[16] Y. Ren, J. Han, Y. Lin, X. Mei, and L. Zhang, "An ontology-based and deep learning-driven method for extracting legal facts from Chinese legal texts," *Electronics (Switzerland)*, vol. 11, no. 12, p. 1821, Jun. 2022, doi: 10.3390/electronics11121821.

[17] A. Thomas and S. Sangeetha, "Semi-supervised, knowledge-integrated pattern learning approach for fact extraction from judicial text," *Expert Systems*, vol. 38, no. 3, May 2021, doi: 10.1111/exsy.12656.

[18] I. B. Mezghanni and F. Gargouri, "Information retrieval from unstructured arabic legal data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9810 LNCS, 2016, pp. 44–54. doi: 10.1007/978-3-319-42911-3_4.

[19] K. Kocijan, "Formalizing natural languages: the NooJ approach, by Max Silberztein. London, cognitive science series, Wiley-ISTE, UK, 2016. ISBN 9781848219021," *Natural Language Engineering*, vol. 24, no. 1, pp. 161–162, Jan. 2018, doi: 10.1017/s1351324917000419.

[20] D. Celik *et al.*, "Towards an information extraction system based on ontology to match resumes and jobs," in *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, IEEE, Jul. 2013, pp. 333–338. doi: 10.1109/COMPSACW.2013.60.

[21] B. Paassen *et al.*, "Ontology-based extraction of structured information from publications on preclinical experiments for spinal cord injury treatments," in *SWAIE 2014 - 3rd Workshop on SemanticWeb and Information Extraction, Proceedings of the Workshop*, Stroudsburg, PA, USA: Association for Computational Linguistics and Dublin City University, 2014, pp. 25–32. doi: 10.3115/v1/w14-6204.

[22] Knowledge Based Systems Inc, "IDEF: integrated definition methods," *IDEF3 Process DEscription Capture Method*, pp. 3–5, 2014, [Online]. Available: http://www.idef.com/idef3.htm

[23] M. Grüninger, M. S. Fox, and M. Gruninger, "Methodology for the design and evaluation of ontologies," *International Joint Conference on Artificial Inteligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing*, pp. 1–10, 1995, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.8723

[24] N. F. Noy and D. L. McGuinness, "Ontology development 101: a guide to creating your first ontology," *Stanford Knowledge Systems Laboratory*, vol. 9, no. 12, p. 25, 2001, doi: 10.1016/j.artmed.2004.01.014.

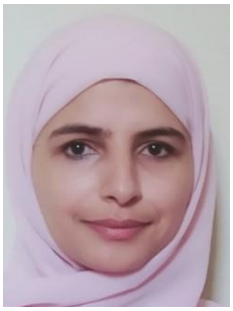
[25] S. Mesfar, "Named entity recognition for arabic using syntactic grammars," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4592 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 305–316. doi: 10.1007/978-3-540-73351-5_27.

[26] C. C. van der Eijk, E. M. van Mulligen, J. A. Kors, B. Mons, and J. van den Berg, "Constructing an associative concept space for literature-based discovery," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 5, pp. 436–444, Mar. 2004, doi: 10.1002/asi.10392.

[27] K. Belhoucine, N. Zame, M. Mourchid, and A. Mouloudi, "Enhancing legal case retrieval in Moroccan insolvency law using ontology," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 17, pp. 7119–7129, 2023.

BIOGRAPHIES OF AUTHORS

Kaoutar Belhoucine is a Ph.D. student at the Department of Computer Sciences, Ibn Tofail University in Morocco. Her research is focused on the field of computer science, specifically computational linguistics, with a specialization in Arabic natural language processing. Her research interests include software engineering, information retrieval, and ontologies. She can be contacted at email: kaoutar.belhoucine@gmail.com.



Nadia Zame is a Ph.D. student at the Department of computer sciences, Ibn Tofail University, Morocco. She is interested to create a search system based on an ontology extracted from Moroccan legal texts. She can be contacted at email: nadia.zame@uit.ac.ma.

Mohammed Mourchid Doctorate Degree in Computer Science In 1999; Associate Professor at The Computer Science Department at The Faculty of Sciences, Ibn Tofail University in Kenitra Morocco He is a supervisor of computational linguistic projects. He has been working on an Arabic dictionary since 2017 in the same department. On going research interests: natural language processing, web semantic, and information systems. He can be contacted at email: mourchidm@hotmail.com.