

MISLA²: A System to Information Retrieval in Labour Lawsuits using Legal Ontologies and Regular Expressions

Cleyton M. O. Rodrigues¹, Bruno J. T. Fernandes¹, Leandro H. S. Silva¹,
David J. Barrientos¹, Allana L. S. Rocha¹, Paulo Christiano Sobral²,
Bruno Souza², Dionizio Feitosa², Mabel Guimarães², Juliana Barreto²

¹University of Pernambuco, Recife-PE, Brazil

²NeuroLaw Desenvolvimento de Tecnologia LTDA, Recife-PE, Brazil

cleyton.rodrigues@upe.br, {bjtjf, lhss, djbr, alsr}@ecomp.poli.br,
{christiano.sobral, bruno.ribeiro, dionizio.feitosa, mabel.guimaraes,
juliana.barreto}@neuro.law

Abstract. *Electronic Legal Proceedings are a worldwide legal phenomena, allowing the use of computerized systems for the creation and monitoring of procedural acts in the most diverse legal bodies. On one hand, it allows greater transparency in the conduct of procedural acts, on the other, it has contributed to the bottleneck of open but unresolved lawsuits each year. Nowadays, Information Retrieval to automate the processing of these procedural objects is at the forefront of computer systems for Law. In this study, we present MISLA², a system to retrieve orders and preliminaries from judicial labour sentences through ontological models built from previous cases. Instead of tied and difficult-to-maintain domain specification models, we demonstrate how light ontologies, in conjunction with regular expressions for extracting significant portions of the text, can achieve the desired results. In addition, empirical experiments carried out with real labour lawsuits evidence that results are quite promising.*

1. Introduction

Information Technologies have been largely employed to streamline traditional procedures in different sectors of the economy. Particularly, in Law, there is a growing trend in the use of digital platforms to create and monitor legal and procedural acts in a standardized, fast, efficient, and economical way, in addition to being accessible from any location and temporality [Lupo and Bailey 2014]. In Brazil, Electronic Legal Proceedings (PJe) emerge as an alternative to the pile of papers of traditional procedures. Paradoxically, the facility incorporated by PJe brought, on the other hand, a dizzying increase in the number of processes generated, creating a bottleneck due to the delay in processing this growing volume [Neman de Novaes and Bissoli 2019]. Given this scenario, it is suggested that tasks to automate (albeit partially) the processing of electronic lawsuits is a path of no return.

Automation has been consolidated in the most diverse sectors of the economy [Acemoglu and Restrepo 2019]. Companies in the legal field are already beginning to experiment with the use of artificial intelligence tools, both to obtain more agility in the conduction of their work, as well as to allow a better understanding of their models business, even providing support for their decisions [Berk and Hyatt 2015]. Automating the processing of electronic lawsuits is therefore of utmost importance. A legal process is a

complex set of procedures involving a series of documents. Each document or decision unfolds in very extensive textual elements, making a single process have a very wide set of information, although they are sometimes redundant. Thus, the interpretation and extraction of information from lawsuits is non-trivial and even if more experienced lawyers, when analyzing each case individually, can indicate the most likely results and the best associated strategies, there is no clear methodology on how to achieve this analysis. In other words, there is usually a straight dependence on the subjective assessment of lawyers.

Information retrieval is one of the most requested tasks in the processing of textual content [Harman 2019]. Information Retrieval Systems are tools that establish a set of well-defined actions to retrieve necessary information, while smoothing out the extraction of non-relevant ones. To maximize performance, sophisticated semantic or contextual retrieval strategies are not enough, but must also involve correctly specification of the domain information and its relationships. In this perspective, ontological models [Napoleon 2013] can be used as semantic specifications, with support for reasoning tasks. Ontologies of different magnitudes, from the lightest to the heaviest ones, are being proposed in the literature, for the most varied realms.

Extracting information from labour lawsuits is, in this endeavor, a necessary automation task for law firms, for example. Although they tend to follow almost standardized templates, human extraction of elements such as orders and preliminaries tends to be a repetitive and error-prone task. Therefore, in this study, we describe MISLA² (acronym for Mining Intelligent System for labour Lawsuits): a system based on lightweight Ontologies with the addition of rules designed to support this process. MISLA² already has modules delivered for end-users, which has been reducing the time and cost required in decision-making by law firms. Among the components provided, this study focuses on the ontologies and rule-based module to extract and classify the merits of orders and preliminaries in labour lawsuits. For this case, the option of having the conceptual model instead of a hard-coded specification is justified by the ease with which the Ontology can naturally grow from new information, without losing semantic rigor.

This study is organized as follows. The theoretical framework of the research is presented in Section 2. Section 3 highlights the architecture of a part of the MISLA² system responsible for extracting orders and preliminaries, accompanied by the ontological specifications described in Section 4. The experiments performed in the study are described in Section 5 and, finally, in Section 6 we highlight the research considerations and the work in progress.

2. Theoretical Background

2.1. Labour Lawsuits

In the Brazilian Labour Law [Brasil 2017], when the employer, whether a person or a company, fails to respect labour rights, the employee can seek them in court through an action proposed in the labour Court called labour Claim. The labour process has five stages, initial petition, hearings, sentence, appeals and execution, and takes place in three procedures: Ordinary, Summary and Supreme. The first stage is the Initial Petition, the inaugural piece that gives impetus to jurisdictional protection. The initial may be verbal or written, requires addressing and qualification of the parties, presentation of the facts, orders, the value of the case, date, the claimant's signature, or his representative, according

to Art. 840 of the CLT (in Portuguese, Consolidação das Leis Trabalhistas).

After entering the Judiciary through the initial petition, the case is distributed and sent to the labour Court, where the inaugural hearing is scheduled, which is the moment of conciliation attempts between the parties. As soon as the process is filed, a conciliation hearing is scheduled to allow the employee and employer to discuss the possibility of an agreement. If so, the judge will ratify the agreement through a sentence, and the employer will be obliged to comply with the terms under penalty of fine and sentence execution. If there is no agreement, another hearing will be scheduled, for instruction and trial, to hear witnesses and have the judge's decision. In some cases, depending on the value of the action, the procedure is different, with only one hearing for conciliation, instruction and judgment, which can be Ordinary, Summary or Supreme.

Then comes the sentence, which is the judge's decision on the rights claimed by the worker. The action will be upheld, partially upheld or groundless. The valid decision is the one that establishes that the worker is entitled to everything that was requested in the process. The partially valid one recognizes some rights but denies others. In the case of an unfounded decision, the judge does not recognize any of the claimed rights. An appeal from the sentence will always be possible, by one or both parties, to the second instance - Regional labour Court (in Portuguese, Tribunal Regional do Trabalho).

After the sentence, the party may file an Appeal case it disagrees with the court decision. If the employee chooses to appeal the decision, he must present all the arguments for the award to be modified. If the appeal is from the employer, it will be up to the employee to contest it, seeking that the sentence is upheld. The sentence execution occurs immediately after the appeal and after the decision is final and unappealable. When the decision (sentence or judgment) can no longer be appealed against, the action becomes final, ending the process in which rights are sought to be recognized.

After the final decision, a court accountant will be appointed to settle the sentence and calculate the amounts due. Once this is done, the judge will ratify the calculations, and the employer will be summoned to make the payment. If payment is not made or a proposal for instalment payment is made, the sentence execution will proceed, with the inclusion of the employer in the National Bank of labour Debtors (in Portuguese, Banco Nacional de Devedores Trabalhistas) and the search for values and assets for the pledge. Finally, with the payment, permits will be issued favouring the parties, and the case will be filed.

Given the voluminous corpus of legal documents, the varied possibilities that a case can follow, and the legal information that may be extracted (for example, name of the judge, values, justification, provisions, among others), in this study we focus on recovery of the claimant's potential rights and in the defendants's objections, which hereafter will be called orders and preliminaries, respectively. Preliminaries are impeding facts raised by the defendant, aiming to end the process without resolution of merit. The orders address the merits of the case, where their assessment will imply in full approval, rejection or partial concession. Unlike other components in labour lawsuits, orders and preliminaries are not limited to just one occurrence per document.

2.2. Information Retrieval

With greater emphasis on Computer Science, the Information Retrieval (IR) area is, in fact, a multidisciplinary field involving other consolidated realms, such as cognitive psychology, linguistics, semiotics, library science, statistics and information sciences [Hersh 2021]. IR involves a set of actions that involves structuring and organizing, indexing and storing, as well as searching and disseminating information. Information Retrieval Systems have been massively produced in the last couple of decades to make available to the general population collections of information of the most varied natures, and of the most distinct types. This information ranges from excerpts to complete documents, including other types of media, such as metadata, audio, video, and images.

Relevance stands out as an essential property of information retrieval systems (IRS). Relevance means the ability to extract information that the user needs, without returning unrelated content. For this, semantic optimization and contextual search techniques have been used to improve the performance of these systems. The use of these techniques, however, is directly associated to the type and variability of the information stored. Unstructured texts with weak semantic tend to be more difficult to deal with, while texts with controlled vocabulary make use of simpler strategies, such as searching for keywords, less complex and simpler to maintain. IRS is based on domain conceptualization and structuring actions, which involves the specification of terms, as well as storage, indexing and query models strategies [Dong et al. 2008]. The latter can be broken down into some types, such as probabilistic models and semantic models, in which the recovery takes place through the semantic combination of the concepts involved, optionally improved with recovery heuristics.

2.3. Ontologies

Within Information Sciences, in particular, Ontology is an explicit and formal specification of a shared conceptualization [Guarino et al. 2009]. Ontological models are, therefore, built following the prerogatives of formal representations, where their components are explicitly defined by logical-mathematical resources, allowing their sharing and understanding by men and machines in an unequivocal way. In particular, Domain Ontologies are useful for describing concepts of a particular realm, usually specialized from more abstract foundation ontologies.

Casellas [Casellas 2011] lists the most common purposes of ontologies, especially those developed in the legal field: organization and information structuring, reasoning and problem solving, semantic indexing and search, semantic integration and interoperability, and domain understanding. Another distinction is in relation to light and heavy ontologies. While the former are characterized by the specification of simpler concepts and roles, with occasional simpler inference tasks, their counterpart is formed by complex ontologies enriched with more robust axioms [Corcho et al. 2006].

In order to model the knowledge of an arbitrary domain, avoiding redundancies and inconsistencies, during the conceptualization activity intermediate representations are engineered built around the main ontological components: concepts, relationships, instances, constants, attributes, axioms and rules [Corcho et al. 2006]. The literature has provided several ontological languages, such as Description Logics (DLs). DLs [Baader et al. 2010] are a set of formalisms to represent knowledge and for reasoning ser-

vices. DLs can be seen as subsets of classical First-order Logic, where predicates are limited by up to 2 variables, avoiding decidability issues. Among the distinct flavours of DL, we focus on the sublanguage \mathcal{ALCI} [Baader et al. 2010], which is the traditional Attributive Concept Language with Complements, plus inverse relations. The \mathcal{ALCI} grammar is also composed of elementary constructors, such as conjunction, disjunction, negation, existential and universal restriction quantifiers. Taking N_C as the set of atomic concepts, and N_R as the set of roles, the syntax of \mathcal{ALCI} is arranged in terms of the following constructors: $N_C \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C \mid N_R^-$.

Through these constructors, it is possible to conceptualize a domain in terms of concept expressions and role expressions. An \mathcal{ALCI} -DL knowledge base (\mathcal{KB}) is structured in terms of the following axioms: (i) Concept assertion, $a : C$; (ii) Role assertion, $(a, b) : R$; (iii) Concept inclusion, $C \sqsubseteq D$, to describe concepts; (iv) Concept equivalence, $C \equiv D$ to define concepts; and (v) Inverse roles, $(R) = R^-$.

DLs semantics are commonly specified in terms of FOL interpretations. An Interpretation \mathcal{I} is a tuple $\langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, where $\Delta^{\mathcal{I}}$ represents the non-empty set known as the domain of \mathcal{I} , while $\cdot^{\mathcal{I}}$ is a function that maps concepts to subsets of $\Delta^{\mathcal{I}}$, relations to subsets of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and instances to elements of $\Delta^{\mathcal{I}}$. For additional details, please check the reference [Baader et al. 2010]. From the first-order interpretations, some reasoning tasks are available in DL, such as satisfiability and logical implication. Given an arbitrary concept C , C is satisfiable iff it admits a model. An interpretation \mathcal{I} is a model of a concept C if $C^{\mathcal{I}} \neq \emptyset$. Likewise, an interpretation \mathcal{I} is a model of a general concept subsumption ($C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. Finally, given a knowledge base \mathcal{KB} , and two concepts C and D , D subsumes C , if for all models \mathcal{I} of \mathcal{KB} , $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, that is, the logical implication $\mathcal{KB} \models C \sqsubseteq D$ is assured.

Therefore, based on the possibilities of semantic specification, as well as on the reasoning tasks employed by ontologies, this study unfolds on the idea that these conceptual models can be used to axiomatize the information and assist in indexing and information retrieval tasks, in particular, in legal lawsuits where documents tend to follow templates standardized by law firms, with a more controlled vocabulary.

2.4. Related Work

Asim [Asim et al. 2019] carried out a survey on the use of ontologies in information retrieval, whether in textual or multimedia format. The aim of the study was to categorize the possible strategies for calculating the semantic similarity between the user's query and the indexed data. The following retrieval approaches have been identified: Vector Space based, Probability based, Context based, Semantic based, Semantic Similarity based, and Semantic Association based. What is common in these approaches is the need to augment the user's query, before seeking similarity with the stored corpus, either by calculating the cosine between the representative vectors, by probabilistic distribution, by the context of the concepts present in the query, or by the semantics of indexed documents. Asim (2019) highlights, however, that in general, these approaches tend to be much slower than classical extraction techniques, making their use in real contexts unfeasible.

Machine learning strategies are also used to retrieve legal information [Savelka et al. 2019], notably those to retrieve documents using Document Vector Embeddings [Sugathadasa et al. 2018]. With respect to the types of elements retrieved in

this present study, machine learning strategies were discarded due to the size of the database. As highlighted in Section 4, nearly three hundred of orders and preliminaries were mapped. Nevertheless, while some types appear almost frequently, other more specific types are rarer to occur in legal proceedings. This imbalance would harm the training of learning models, affecting the accuracy of the extraction.

Unlike the related works, in this study, we seek to deal with the trade-off between extraction accuracy and performance. It also takes into account the need to propose alternatives that can be easily understood and even expanded by non-technicians in the area. Therefore, taking into account that the retrieval will be triggered by somewhat standardized legal sentence excerpts, instead of user queries with a high degree of variability, and that the response time is a decisive factor, we propose to use lightweight ontological models, where elements domain keys, such as legal norms, are determining factors for recovery.

3. General Architecture of MISLA²

The architecture presented in this study is the MISLA² module addressing strategies based on regular expressions and ontological models with domain specifications. While the former are mostly used to identify the parties that make up a labour law sentence, ontologies, in turn, serve as a vocabulary of orders and preliminaries, supporting their classification. Figure 1 illustrates an activity diagram with all the actions carried out.

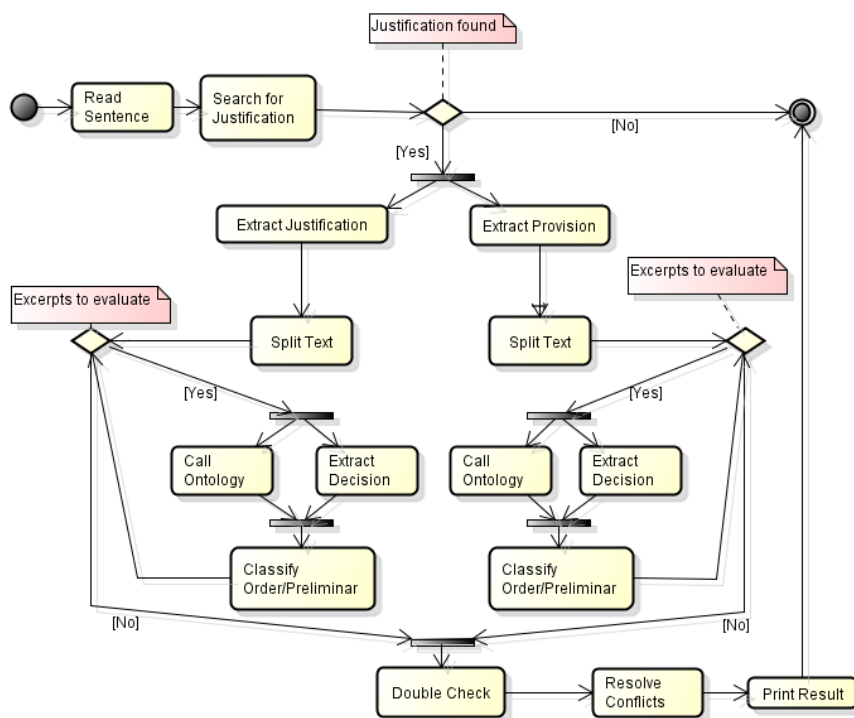


Figure 1. General Flow of Classification Activities

Briefly, a labour sentence, when analyzed, leads to two paths. The first tries to extract the basis of the sentence with the reasons on which the applicable legal consequences were based (known as the Justification), while the second seeks to extract the Provision, with the results of the judgment. Therefore, it is common for orders and preliminaries to

be described in both parts of the document, which leads us to the possibility of double-checking the classifications. Back to the diagram, after identifying each part, it is broken into excerpts (the paragraphs) and forwarded to the respective Ontologies. In the fragment where the order/preliminary is identified, new rules seek to verify its approval (or not). It is suggested that the analysis of the merits is close to its type. At the end of the parallel paths of analysis, they are unified, seeking to compare the elements and their approvals. In case of divergence, it is decided to grant the order/preliminary, considering that even those classified with partial approvals cannot be considered as rejected. In the following subsections, these tasks are described in detail.

3.1. Rules for Extraction

The justification of labour sentences typically bring information from the lawsuit orders and preliminaries separately, as exemplified in the sentence excerpt illustrated in Figure 2. These headings are then extracted, as are the paragraphs below them.

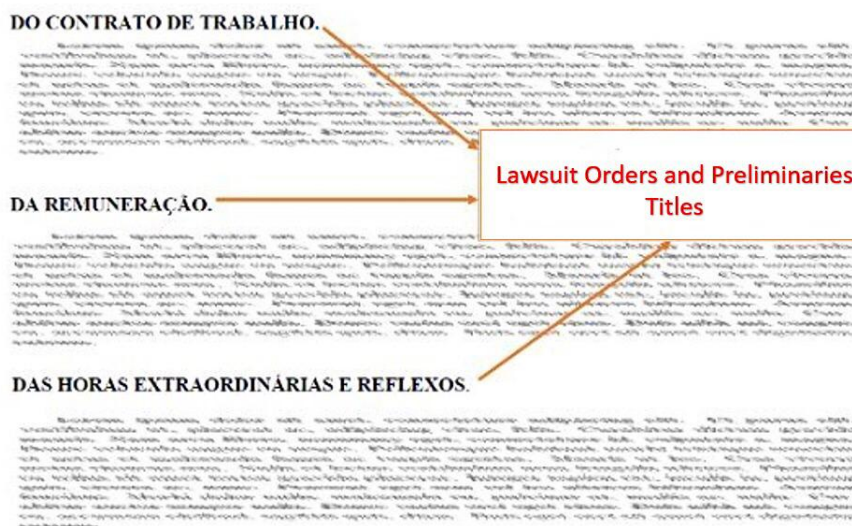


Figure 2. Example of Orders and Preliminaries in labour Sentences

Extraction is based on textual structure. That is, the lawsuit orders and preliminaries will only be extracted, correctly, if they are arranged in a single line and between two unfilled (empty) lines. This rule was implemented along with a filter for some stop-words, which eliminates some supposed titles that can be caught by the previous rule. The combination of these two methods guarantees the best extraction of titles, based on the structure of the text. The legal provision presents the results in a more linear and compact structure, imposing different rules for extraction. Therefore, other actions for text processing were used in order to build the excerpts for analysis.

3.2. Ontologies for Legal Sentences

Considering the uniqueness of labour lawsuits in Brazilian Law, two domain ontologies were developed: one to map the orders (OntoPedido) and the other to deal with the possible preliminaries (OntoPreliminar). Both ontologies follow the same structural scheme and can be easily integrated. The option for splitting is twofold: isolating the classification services and mitigating time-consuming problems, as highlighted later on.

For Ontological Engineering, we use a bottom-up construction approach. This recommends directing efforts to survey the constituent elements of the ontology from the knowledge of the domain, whether explicit as the available textual corpus, or even implicit, as the tacit knowledge of specialists in the area. Ontologies driven by the bottom-up method tend to be more specific and less reusable, compromising a broader scope and use [Corcho et al. 2006]. To tackle this problem, the ontological conceptualization was broken into three complementary parts: the extraction through real labour lawsuits, through documentary research, and with specialists from law firms. In this way, we sought to exhaust the types of preliminaries and orders described in the Ontologies of Section 4.

Ontologies were developed through Protégé editor, which follows the methodology Methontology [Corcho et al. 2005]. To handle the Ontologies, the Python OWL-Ready2 library was used. Also, to manage the simultaneous execution of Ontologies, a Thread pool was created. Figure 3 pictures the execution in the following steps: reading the ontology, followed by calling the threads, leading to the invocation of the engine for reasoning tasks.

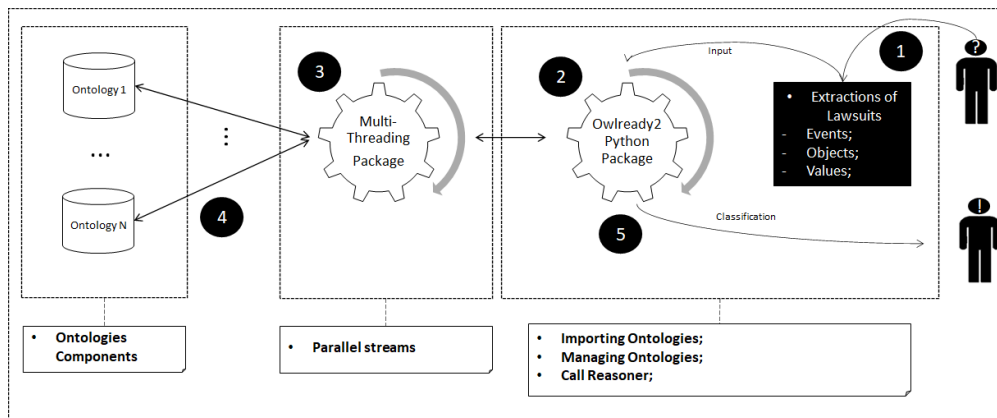


Figure 3. General Architecture for Ontology-Based Classification

4. Ontologies Specifications

In this section, we present the models engineered in this study. It is worth to mention that we chose to keep the ontologies in Portuguese to safeguard the concepts involved, which may have slightly new connotations in other languages.

4.1. Ontology for labour Lawsuit Preliminaries

Preliminaries are objections that can be raised and that are provided for in the Brazilian labour Law. These are types of claims judged even before talking about the merits of the case. Based on available labour sentences, interviews with experts in the field and bibliographical sources, we engineered the OntoPreliminar ontology with 21 preliminaries. For each one, a vocabulary was built with possible variations on how they might appear in the texts.

As an already well-established area, preliminaries in labour Law are composed of expressions that involve an object or event, and a qualification/quantification. Another situation mapped is the one where the preliminary is related to a legal norm. From these analysis, we conceptualize the ontology through DL-axioms described in \mathcal{AP} . We have,

for example, the preliminary of an initial (object) that is inept (value) and the prescription (event) that is biennial (value) or five-yearly (value). On the other hand, “Listipendência”, which is the repetition of a legal action already in progress, is brought to light in the justifications from the article that defines it, specifically art. 337 of the Civil Procedure Code. The last pair of axioms in \mathcal{AP} defines the inverse property of roles, in order to find out the objection from its constituent elements. Indexing occurs, therefore, by the constituent elements, together with citations of legal norms related to the preliminary. It is noteworthy that this method of indexing is possible due to the standards in the writings of lawsuits, where vocabulary is controlled to some extent.

$$\mathcal{AP} : \left\{ \begin{array}{ll} \text{Preliminar} & \equiv \exists \text{hasObject.Object} \sqcap \exists \text{hasValue.Value}, \\ \text{hasNorm.Article} & \sqsubseteq \text{Preliminar}, \\ \text{hasObject} & \equiv \text{isObject}^{-}, \\ \text{hasValue} & \equiv \text{isValue}^{-} \end{array} \right\}$$

Figure 4 illustrate a portion of the ontology knowledge base, considering the preliminary of five-year prescription. In particular, we present its description in the Protégé tool, and, on its right, the axiomatization in Description Logic (\mathcal{IP}). The basic idea of classification lies in exhausting the variations of the preliminaries in the sentences, enabling their identification. Unlike complex models that are difficult to maintain, the simplicity of the proposed ontology allows it to be naturally evolved, even by legal professionals.

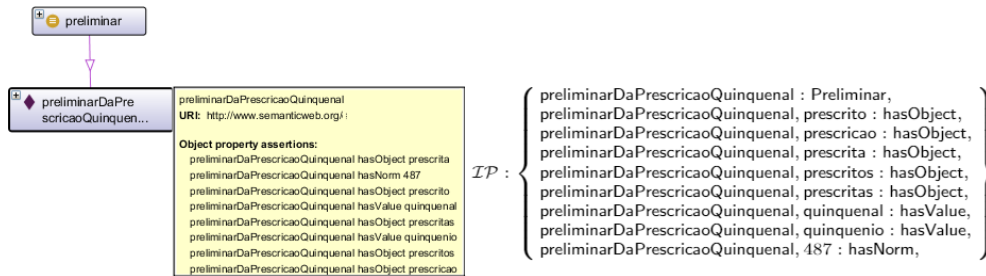


Figure 4. Metrics for the designed Ontologies

4.2. Ontology for labour Lawsuit Orders

For the registration of the types of orders that commonly appear in labour sentences, the same methodological procedure was considered. Currently, OntoPedido has approximately 250 cataloged orders. As highlighted above, the model has the same structure as OntoPreliminar, allowing its combination. Nevertheless, due to time-consuming problems to access the integrated Ontology and retrieve the classifications, we proceeded to separate the designed models. Even separated, the OntoPedidos itself also had problems regarding the time issues, which led us to multi thread processing into smaller sub ontologies. Section 5 details the experiments with and without the parallel streams. Comparing both models, OntoPedido is, in terms of axioms and instances, approximately 3 times larger than OntoPreliminar.

5. Experiments

The experiments were performed with a corpus of real labour lawsuits retrieved from the end-user database. Precision measurements were carried out w.r.t. the order and prelim-

inary extractions, as well as in the recovery of their approvals. The results achieved by MISLA² were forwarded for evaluation by law firms specialists. In order to perform an assessment, a database was created by deploying the system on a test workspace to read labour processes for 6.5 hours straight. During this period, it read a total of 2,482 different labour processes, classifying 69 of them as first-degree labour sentences. Mentioned sentences are to be recorded as documents into the final database.

A qualitative assessment was done over the sentences in which a justification is found, due to it being a requirement to obtain lawsuit preliminaries and orders. From the labour sentences database, 55 documents (79.71%) had the justification section extracted. Results were verified by a law specialized group, checking preliminaries and orders for the labour sentences (each sentence may have many different ones on itself). Table 1 presents the amount of sentences that had: full, partial or no extractions at all.

Table 1. Preliminaries and orders extraction on the first-degree labour sentences database.

	Full Extraction	Partial Extraction	No Extraction
Preliminaries	45 (81.9 %)	10 (18.1%)	0 (0 %)
Orders	50 (90.9%)	5 (9.1%)	0 (0 %)

With respect to the orders extraction, other additional results were achieved. On the 50 sentences with full orders extraction, 431 orders were identified, of which 424 (98.4%) were correctly identified and 7 (1.6%) erroneously or not identified. As for approved orders, 46 (92.0%) sentences had them correctly identified and 4 (8.0%) were partially correct. Finally, on the 5 sentences in which orders were partially extracted, 13 (65.0%) orders were correctly identified and 7 (35.0%) were not identified. In the entire process, only two orders were retrieved as false positives.

On the 45 sentences with full preliminaries extractions, in 16 sentences (35.6%), all extracted preliminaries were confirmed with the expert team. On the other hand, in 29 sentences (64.4%), false positives were also found. This means that MISLA² extracted unnecessary data in these cases. On the 10 sentences with partial preliminaries extraction, 7 (38.9%) preliminaries were correctly identified and 11 (61.1%) were not found.

Another important aspect of MISLA² is the time needed to carry out the extractions and classifications, as it must function as a crawler, recovering and handling labour processes. On the client side, the response time needs to be reasonable enough for the system to be more accurate and responsive than the human counterpart. Therefore, a new set of court cases (33 documents) was extracted to assess time consumption. This analysis took place in two scenarios: The first one is a single Ontology and the second one is a multi thread version with a set of sub ontologies. Figure 5 shows the processing time for each court cases in both scenarios. In the Single Ontology model, the total processing time was 6 hours and 39 minutes of uninterrupted processing. Nine processes took more than 20 minutes, with almost half the time being consumed by just 5 processes. In this setting, it was realized that lateness was due to lawsuits over 70 pages.

After identifying the bottlenecks in the system, the solution was, therefore, to divide the ontology of orders into smaller structures, allowing separate execution flows for each sub model. The Sub-Ontologies Multi-Thread Model required 1 hour and 31

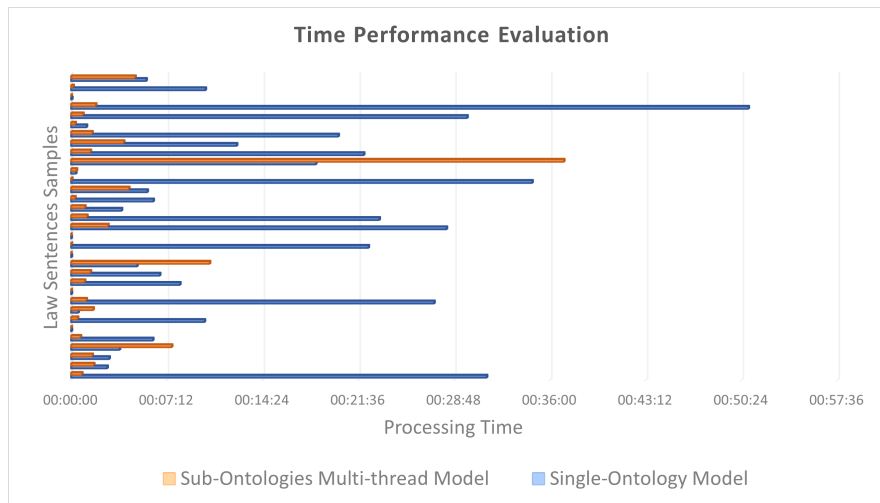


Figure 5. Time Evaluation

minutes to retrieve the same information of the Single-Ontology Model - a reduction of 77% of processing time.

6. Final Remarks

This work presents a study about information retrieval in labour lawsuits using domain legal ontologies and rules. In opposition to the hard-coded specification, the conceptual model enables a natural growth from new information while preserving semantic rigor. As presented in the experiments, the proposed system can correctly extract and identify almost all the requested information presented at labour law decisions. The results highlight that as long as MISLA² is able to identify the justification, it will extract preliminaries and orders. The system is able to identify preliminaries on 82.9% of the sentences. However, it identified preliminaries that did not exist (false positives) on more than half of them, being the only downside on the system shown after assessment. For the orders, the system is able to find and identify approved orders for 90.9% of the sentences with a 98.8% accuracy for 424 orders. Time consumption satisfactorily met end-user expectations as well.

As future work, to mitigate the recovery of non-relevant data, new follow-ups with experts are currently being set up. In particular, two preliminaries appeared more frequently as false positives, which leads us to believe that the definitions in descriptive logic need improvement. In addition, we intend to expand the analysis and the ontology for other Law areas. The simple structure of the Ontologies allows, above all, that new elements can be added, including by legal professionals, without any change in the MISLA² extraction and classification module.

References

- Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.
- Asim, M. N., Wasim, M., Ghani Khan, M. U., Mahmood, N., and Mahmood, W. (2019). The use of ontology in retrieval: A study on textual, multilingual, and multimedia retrieval. *IEEE Access*, 7:21662–21686.

- Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors (2010). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2 edition.
- Berk, R. and Hyatt, J. (2015). Machine learning forecasts of risk to inform sentencing decisions. *Federal Sentencing Reporter*, 27(4):222–228.
- Brasil (2017). Lei nº 13.467, de 13 de julho de 2017. *Diário Oficial [da] República Federativa do Brasil*.
- Casellas, N. (2011). *Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge*. Law, Governance and Technology. Springer Netherlands, Barcelona, Spain.
- Corcho, O., Fernandez-Lopez, M., and Gómez-Pérez, A. (2006). Ontological engineering: Principles, methods, tools and languages. In Calero, C., Ruiz, F., and Piattini, M., editors, *Ontologies for Software Engineering and Software Technology*, pages 1–48. Springer Berlin Heidelberg.
- Corcho, O., Fernández-López, M., Gómez-Pérez, A., and López-Cima, A. (2005). *Building Legal Ontologies with METHONTOLOGY and WebODE*, pages 142–157. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Dong, H., Hussain, F. K., and Chang, E. (2008). A survey in traditional information retrieval models. In *2008 2nd IEEE International Conference on Digital Ecosystems and Technologies*, pages 397–402.
- Guarino, N., Oberle, D., and Staab, S. (2009). What is an ontology? In Staab, S. and Studer, R., editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 1–17. Springer.
- Harman, D. (2019). Information retrieval: The early years. *Foundations and Trends in Information Retrieval*, 13(5):425–577.
- Hersh, W. (2021). *Information Retrieval*, pages 755–794. Springer International Publishing, Cham.
- Lupo, G. and Bailey, J. (2014). Designing and implementing e-justice systems: Some lessons learned from eu and canadian examples. *Laws*, 3(2):353–387.
- Napoleon, S. A. (2013). A survey of web ontology languages and semantic web services. *Annals of the Alexandru Ioan Cuza University - Economics*, 60(1):42–53.
- Neman de Novaes, R. and Bissoli, L. G. (2019). Justiça em números: estudos acerca da (in)eficiência do processo judiciário. *Revista Vianna Sapiens*, 10(1):20.
- Savelka, J., Xu, H., and Ashley, K. D. (2019). Improving sentence retrieval from case law for statutory interpretation. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL '19*, page 113–122, New York, NY, USA. Association for Computing Machinery.
- Sugathadasa, K., Ayesha, B., de Silva, N., Perera, A. S., Jayawardana, V., Lakmal, D., and Perera, M. (2018). Legal document retrieval using document vector embeddings and deep learning. In *Computing Conference 2018*.