Legal Knowledge and Information Systems
J. Savelka et al. (Eds.)

doi:10.3233/FAIA241248

Automated Semantic Annotation Pipeline for Brazilian Judicial Decisions

Melissa ZORZANELLI COSTA [a,1], Dylan FARIA ROBSON [b], Thiago B. P. VIEIRA [b], Jean-Rémi BOURGUET [b], Giancarlo GUIZZARDI [c] and João Paulo A. ALMEIDA [a]

[a] *Federal University of Espírito Santo, Brazil*
[b] *Vila Velha University, Brazil*
[c] *University of Twente, The Netherlands*

Abstract. In recent years, Brazil's federal judicial system has embraced digitalization, making a large amount of legal process information available to citizens and legal experts. Despite the advances, a significant portion of the data produced and stored in legal systems presents itself in the form of natural language text, including numerous petitions and legal decisions. This creates barriers for automated querying and analysis of legal process data, especially considering the importance of the content of legal decisions in these tasks. In this paper, we report on an automated semantic annotation pipeline for judicial decision texts obtained from the official National Uniformization Panel (TNU) jurisprudence website. NLP models are trained in a few-shot learning context with a training set annotated by legal experts. The semantic annotation approach is evaluated using precision and recall. The results of the semantic annotation are produced into RDF-based nanopublications aligned with a reference domain ontology. The annotations are accompanied with provenance information including identification of the machine learning model used.

Keywords. Semantic Annotation, Legal Decision Classification, Machine Learning, Legal Ontology

1. Introduction

The judicial system in Brazil, particularly in its federal domain, has embraced digitalization to better handle and publicize legal process information. The body of jurisprudence formed by the decisions of several courts (e.g., the TNU[2]) is made available through public websites, allowing free access to citizens and jurists alike. However, despite the potential benefits, in many cases information is still not available in easily processable data formats. In previous works, some of us designed a domain ontology and performed the 'triplification' of semi-structured data extracted from judgements available in the official TNU jurisprudence website [1,2]. Nevertheless, an important portion of relevant information still needs to be obtained by examining the content of decisions, produced by judges and clerks and stored in natural language text. This means that key information such as types of legal procedures, subjects of legal texts, legal reasoning, etc., are not directly available for querying and analysis. In this paper, we extend our previous work

[1]Corresponding Author: Melissa Zorzanelli Costa, melissa.inf.ufes@gmail.com
[2]https://www.cjf.jus.br/jurisprudencia/tnu/

and enrich the extraction process with the automated classification of the elements of legal decision text. We adopt a learning approach and train Natural Language Processing (NLP) models on real legal decision texts labeled by legal experts.

We employ Named Entity Recognition (NER), which consists in the identification and classification of Named Entities (NEs) within unstructured text [3]. NEs are specific words or phrases that represent entities of interest in a text contributing to the understanding of the content and assisting in information retrieval. While conventional entities such as names, locations, organizations, and dates can be discerned using pre-existing datasets, the landscape of specialized domains such as legal texts requires bespoke training of NER models with domain-specific data. In this particular application, we aim to recognize the elements of discourse used to refer to types of judicial procedures (more specifically the types of appeals that can be handled by the TNU), appeal subject, *ratio decidendi* and decision outcome. We chose *spaCy* as our NLP library due to its simplicity, seamless integration into production systems, and cost-effectiveness, making it an ideal component to prototype our pipeline offering both rapid development and scalability. In addition, *spaCy* is designed to perform various NLP tasks such as NER [4]. Features such as combination of pre-trained models, customization capabilities, efficiency, rule-based approaches, and community support position *spaCy* as a strong candidate for performing NER in legal text. Its ability to handle both general and domain-specific entities, coupled with its integration into larger NLP workflows, makes it a versatile and practical tool for processing and extracting valuable information from legal documents.

Our overall goal is to automatically annotate decisions related to a specific type of appeal within the Brazilian legal system, in the Scope of Special Federal Courts termed *Request for Standardizing the Interpretation of a Federal Law* (RS). It is associated with a specialized procedure and is characterized by its verbosity and lack of transparency. The purpose of this judicial appeal is to reduce disparities in the interpretation of substantive federal law [1]. Previously, our approach involved extracting knowledge from unstructured text about judgements found on the official TNU jurisprudence website. These legal texts were made available in an unstructured textual format in Brazilian Portuguese. We first crawled the TNU website [2], upon which we proceeded with the annotation of 190 syllabi of the decisions, guided by expert clerks. After annotating the decisions, we trained a custom spaCy model on the annotated data to recognize the specific parts of the decision relevant to our task. With this trained model, we implemented an NLP pipeline to transform the extracted information into a suitable RDF-star format, aligning it with our ontology's structure. We also carried out a validation of the automatic creation of these new semantic annotations and we monitored the evolution of classic performance metrics in relation to the size of the training dataset to assess the few-shot impact of our learning process. This work addresses the critical problem of automating the extraction and semantic annotation of legal decision texts to facilitate access, searchability, and analysis of legal data. Resolving this issue is essential for improving legal research efficiency and transparency in judicial processes.

This paper is further structured as follows: Section 2 provides an overview of the NLP pipeline, covering the annotations by law clerks, model training, and its use in information extraction; Section 3 presents an evaluation of the learning approach; in Section 4, we discuss how the trained model is put to use, with automatically annotated data made available in the RDF-star semantic format; Section 5 discusses works closely related to ours; and finally, Section 6 concludes the paper, drawing perspectives for further work.

2. Pipeline overview

In this section, we present an overview of our automated semantic annotation pipeline, which extracts data from the TNU portal, includes a model training phase for domain-specific NLP tasks, applies the trained models to annotate the data, and represents these annotations through RDF-based nanopublications. Figure 1 shows an overview of the whole process, revealing the aforementioned different components of our pipeline.

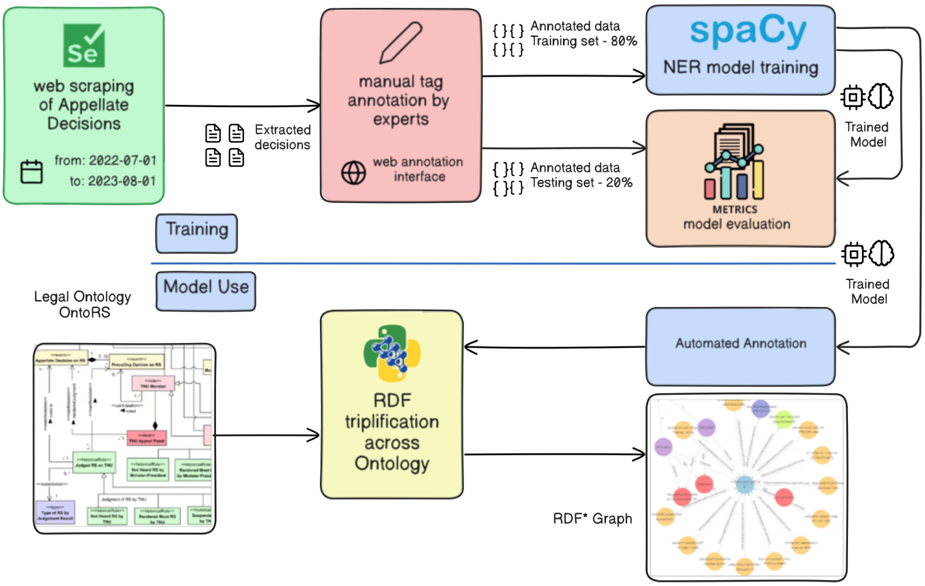

Figure 1. Pipeline overview

Our pipeline starts with data extraction from the TNU website[3]. It is well-established that jurisprudence is one of the main sources of Law and legal research. In this sense, the contents of court decisions from across the country are synthesized and disseminated through the use of syllabi, each of which consists of a summary of the content of a judicial decision. Many efforts have been made by federal agencies to suggest the standardization of judicial syllabi and and there are already some guidelines provided by the Brazilian National Justice Council (CNJ)[4]. Nevertheless, no strict rules are enforced in the formatting of syllabus text; certain ad-hoc patterns are used by judges and clerks. This explains why a simple NLP machine learning approach could be applicable for automated processing.

We curated a dataset with syllabi extracted from real judicial decisions available on the TNU website. For our purposes, we selected a specific date range (July 1st, 2022, to August 1st, 2023), retrieving all judgements (APPELLATE DECISION ON RS) within this defined period to then manually annotate 190 syllabi. The selected syllabi were manually annotated by legal specialists, including one law clerk of the Federal Judge of the Appeals Admissibility Section, who supervised the annotations made by two interns, using the vocabulary provided by a well-founded ontology for this legal domain [2].

[3]https://www.cjf.jus.br/jurisprudencia/tnu/
[4]https://shre.ink/CNJ-ementa-manual

Figure 2 shows an updated fragment of the ontology employed, dubbed OntoRS [2]. OntoRS is a reference ontology that focuses on the handling of the specific type of appeal (RS) ultimately judged at the TNU. The ontology is specified in OntoUML [5], an ontologically well-founded UML profile whose primitives reflect ontological distinctions of the Unified Foundational Ontology (UFO) [6].

The labels used in the annotation of decision syllabi correspond to classes in the reference ontology, identifying: the TYPE_OF_APPEAL, the SUBJECT (of an RS appeal), the decision's REASONING (RATIO_DECIDENDI), and the status of the appeal after a decision: NOT_HEARD, RENDERED_MOOT, SUSPENDED, NOT_ENTERTAINED, GRANTED_TO_REVOKE, NOT_GRANTED and GRANTED_AND_INDICATED. The status of an appeal after a decision follows the respective norm that deals with the subject, namely, TNU's internal Res. n. 586/2019.

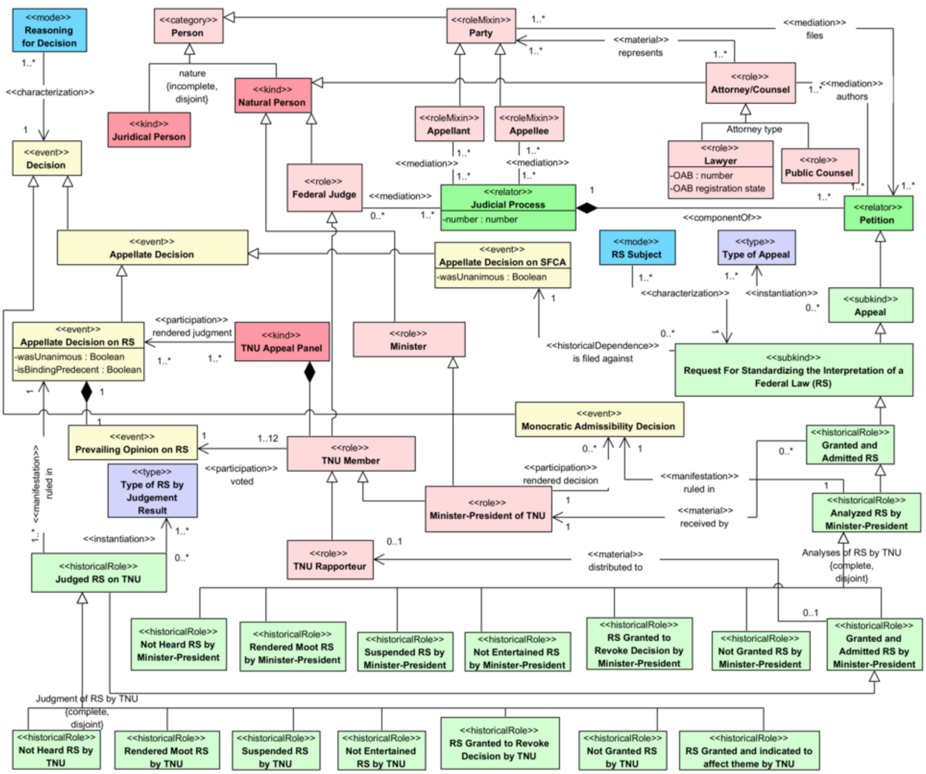

Figure 2. Fragment of Request for Standardizing Legal Ontology

Although there are several tools for manual text annotation, we have developed our own user interface to gain better control over the labelling process and eliminate potential inconsistencies in the way tags are applied. Such inconsistencies could directly impact the model's performance. We built a minimalist design, eliminating elements that do not directly support the user's task, aligning with the 8th usability heuristic presented by Nielsen and Norman [7]. A set of keyboard shortcuts streamlines the annotation process for our legal experts. The team of experts has been trained on how to annotate the syllabi correctly and consistently.

Figure 3 shows a screenshot of the tool[5], where the decisions are displayed in plain text. Experts could select fragments of the text and apply the relevant labels. The extracted fragments, along with their chosen labels, are displayed in the right pane of the interface.

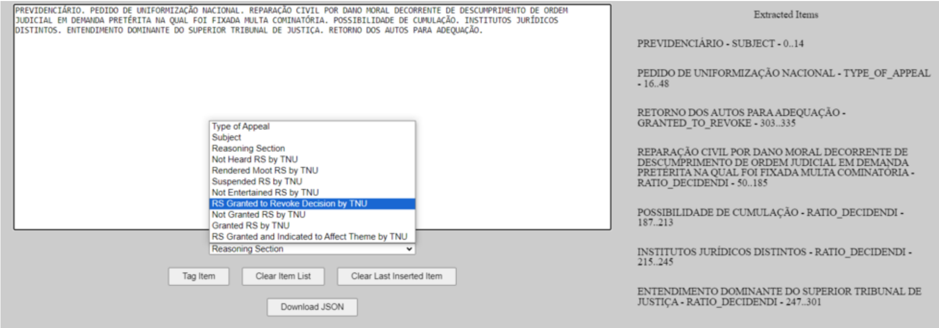

Figure 3. Annotation tool for TNU decisions

Figure 4 depicts[6] three manually annotated decisions from our dataset. Decisions may refer to the same type of appeal, although they might use varying terminology.

Similarly, each decision encompasses a subject, which may vary widely in terms of lexical descriptions. Additionally, every decision includes a reasoning (*ratio decidendi*), but the textual descriptions can exhibit considerable heterogeneity. Finally, judgement results obviously convey the essence of the decisions, which relate to the legal status of the appeal after the decision (such as NOT_HEARD, NOT_GRANTED, SUSPENDED, etc., as explained above).

Concerning automated annotations, it is well known that recognizing domain-specific entities in legal documents requires customization of standard NER libraries, as specialized vocabularies such as that used in legal documents turn inefficient the use of pre-trained models[7] (which use journalistic news or general internet texts). Because of this, we have opted to train a customized model specifically for identifying and classifying core information of the TNU decisions, including the type of appeal, its subject, the decision's reasoning and the judgement results. To start the model fitting process, a blank model was created specifying a desired language (Portuguese in our case).

We have chosen spaCy to prototype our pipeline because, unlike most popular Large Language Models (LLMs), it does not require much computational strength, with optimisations for running on widely available CPU, not requiring specialized GPUs. Furthermore, alongside other models, such as Word2Vec and GloVe, spaCy is based on word vectors, also known as word embedding, This technique consist of representing words as vectors, in which, closer vectors means words with similar meaning. This makes it possible to determine the semantic similarity of words, sentences and documents. On the other hand, spaCy uses a lexical-based strategy, meaning that the same words will have the same embedding regardless of their context, without considering it for disambiguation.

[5] https://labdeborah.github.io/jurix-2024/

[6] We used the NER Text Annotator https://tecoholic.github.io/ner-annotator/ to illustrate.

[7] https://spacy.io/models/

Figure 4. Examples of manually annotated decision syllabi

For example, the word ‘apple’ will have the same representation in both ‘I hate apple’ and ‘I ate an apple’. While LLMs generally outperform spaCy in some tasks, such as NER, we chose spaCy due to its efficiency and our specific requirements, such as fast processing on standard hardware.

In our training sessions, the model was configured with a batch size of 1,000, dropout of 0.1, learn rate of 0.001, using the Adam optimizer (short for Adaptive Moment Estimation). We employed dropout to prevent overfitting and selected Adam for its robustness and efficiency. Instead of specifying a fixed number of training iterations through the entire dataset, the model continues training until there is no improvement for 1,600 steps. This approach allows us to maximize the performance of our model for NER tasks. Additionally, the usage of a seed was adopted to ensure reproducibility. Throughout training, we have assessed the quality of the trained model by using 20% of the syllabi in the dataset as testing set, as described in the sequel.

3. Evaluation

We have evaluated the trained model throughout its development by observing the effects of the increase in the number of syllabi employed in training and calculating validation metrics. Throughout this process, we have employed 37 syllabi in a test set (20% of the final dataset). Evaluation was performed by comparing automatically annotated syllabi against the gold-standard manually annotated versions. We produced validation metrics encompassing recall, precision and the F1-score. This analysis allows to draw conclusions regarding the model's proficiency in deducing accurate annotations related to various aspects, such as the type of appeal, legal foundations, or the decision's outcome.

Recall, precision and F1-score were tracked throughout the model's training using the built-in configuration *training.score_weights*.

Per tag evaluations. We opted for a *hold-out* method for the validation, partitioning our dataset 80/20. We conducted an evaluation for a model trained in all the tags. Results for each tag are presented in Table 1.

tag	precision	recall	F1-score
TYPE_OF_APPEAL	1.00	1.00	1.00
SUBJECT	0.70	0.82	0.76
RATIO_DECIDENDI	0.68	0.63	0.65
NOT_HEARD	0.61	1.00	0.76
GRANTED_AND_INDICATED	1.00	0.20	0.33
GRANTED_TO_REVOKE	0.46	0.55	0.50
NOT_ENTERTAINED	1.00	0.67	0.80
NOT_GRANTED	1.00	0.80	0.89
SUSPENDED	0.67	1.00	0.80
RENDERED_MOOT	1.00	0.71	0.83

Table 1. Performance metrics per tag for SpaCy

An analysis of the metrics for each individual tag reveals that the most prevalent tags—SUBJECT, TYPE_OF_APPEAL, and RATIO_DECIDENDI—which are likely to appear in all decisions, generally exhibited better performance, as evidenced by their higher F1-scores. Among these, TYPE_OF_APPEAL achieved a perfect score, 100% precision, recall, and, thus, F1-score , a reflection of the usage of only one type of appeal in the dataset, which substantially boosted the performance of our model in this tag. In contrast, less common tags like GRANTED_AND_INDICATED had fewer samples and performed poorly, with an F1-score as low as 0.33.

Despite this, for decision outcomes such as RENDERED_MOOT, NOT_GRANTED, and SUSPENDED, there are more promising results, and even with the small sized dataset, we demonstrate that the approach is viable. The tag NOT_GRANTED accounted for approximately 8% of the total dataset, but RENDERED_MOOT was a surprise, as our model effectively identified the significance of the term *"prejudicado"* present in all samples (both train and test).

Although overshadowed by other tag classifications, NOT_HEARD represents approximately 45% of the dataset decisions. In nearly all cases, the sentence *"não conhecido"* was the crucial indicator for correct labeling. While this might appear to be a straightforward binary classification, unlike tag RENDERED_MOOT, these keywords are not always isolated; they are often embedded within lengthy sentences, which increases the risk of misclassification with tags like SUBJECT and RATIO_DECIDENDI.

On the other side of the spectrum, GRANTED_AND_INDICATED was never misclassified, achieving an impressive 100% precision. However, its overall performance in recognizing the tag was suboptimal, with low recall (and consequently low F1 score). Despite having no false positives, as reflected by its precision, it suffered from numerous false negatives, indicating that it failed to classify the tag either correctly or incorrectly.

Few shots evaluation. We assessed the trend in the improvement of our models' performance in relation to the change in size of the training set. These trends are plotted in Figure 5. During the training phase, performance exhibited improvement across different dataset sizes. Precision, recall and F1 improved consistently, reaching the highest point with maximum training set size. All measured metrics showed improvements following training set size, with the greatest improvement occurring from approximately 60% to 80%. This suggests that the dataset size and the overall performance of spaCY are directly linked, with great benefits coming from larger datasets in the range investigated.

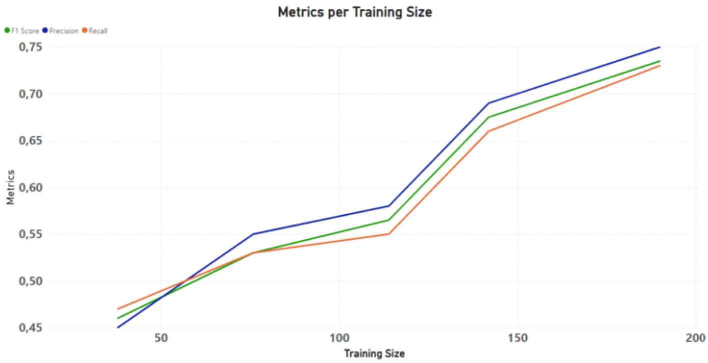

Figure 5. Model performance with respect to training set sizes

4. Representing machine learning-based annotations in RDF

One of the pillars within the realm of Big Data, known as the so-called "5Vs," concerns the concept of *veracity*, which pertains to the level of confidence one can attribute to data. Data collected directly from semi-structured formats on the web (such as the data we extracted from the TNU portal in previous work [2]), cannot be considered to have the same level of certainty as the knowledge produced through the application of statistical-based models such as the one discussed here. We therefore designed in this new iteration of our project, the publication of triples as nanopublications (which represent the smallest meaningful assertion expressed in RDF [8]). We capture both the provenance and other details such as the uncertainty associated with the models responsible for generating this data. We used RDF-star, an extension of RDF introducing the concept of quoted triples, offering a concise way to make statements about other statements [9]. Therefore, it enables the inclusion of descriptions, such as scores, weights, temporal aspects, and provenance, to be attached to edges in a graph as depicted in Figure 6.

In our scenario, the globally unique and persistent identifier that represents the provenance of the nanocertification can take one of two forms. It can either be the URL from which the information was extracted via the ETL process, as outlined in our earlier work [2], or, alternatively, it can be the identification of the specific machine learning model responsible for inducing and generating the nanocertification. A machine learning model can be identified using the hashcode of the serialized file that reifies the model. In this manner, this hashcode serves as a quasi-unique signature of the model recognizing that collisions are conceivable in certain instances with an exceedingly low probability. In

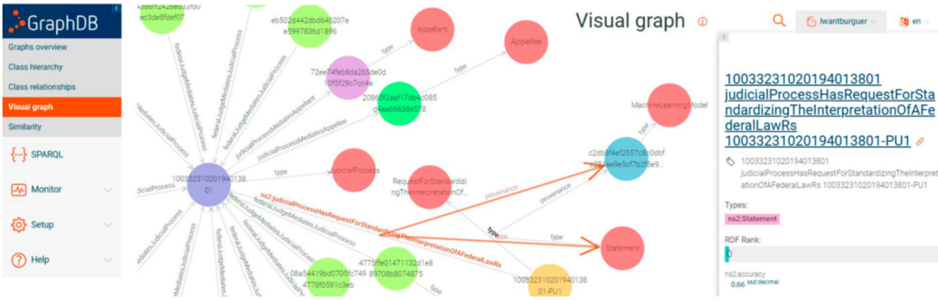

Figure 6. RDF-star representation of our machine learning-based annotations

addition to providing the origin of the nanocertification, we are also publishing[8] some other metadata associated such as the dataset identifier used for training the model, the type of machine learning model employed, some details about the model (e.g. hyperparameters values) and the performance metrics associated with the induction of the nanocertification.

5. Related work

The process of extracting relevant information for legal evidence can be challenging, especially when dealing with unstructured data (see [10]). As emphasized by the authors in [11], there exists a significant demand for NER-annotated datasets comprising legal documents. Exploring legal documents within the context of the Brazilian Judiciary faces a challenge due to the scarcity of reliably annotated corpora by experts. In [12], the authors have created an extensive Portuguese corpus dedicated to legal named entity recognition, consisting of 594 decisions annotated by 76 law students. In a more comprehensive approach, they included a platform for selecting top annotators, which is then used to generate the dataset. They utilized nine years of decisions and focused on extracting key sections from the full text of the decisions rather than syllabi. They used various technologies, including spaCy. In [13], they tested several models in a zero-shot experiment for Portuguese, and spaCy outperformed the others. In [14,15] the authors also used spaCy. In [16], the authors have assembled a dataset comprising 70 legal documents sourced from various Brazilian Courts and legislation documents. In these two efforts, the categories of tags they use are broader in nature when compared with those used in the present work (person, organization, location). Instead, we aim to assess the feasibility of identifying elements in a highly specialized niche of the Brazilian judicial system.

The labeling of legal documents and training a set of classifiers on limited data is a topic extensively addressed in the literature, as highlighted in [17]. Moreover, due to the complexities involved in creating substantial authentic datasets, the few-shot scenario has become a common context in the annotation of legal texts [18,19,20]. Further, a prevalent alternative to manually curated datasets is to depend on automated extraction methods to produce or enrich datasets, as illustrated in [21,22]. In [23], the authors introduce a system tailored for the extraction of legal knowledge, specifically targeting functional text segmentation. Leveraging contextual embedding techniques and embracing the few-shot scenario, their approach demonstrates promise in handling the scarcity of annotated data.

[8] https://github.com/LabDeborah/jurix-2024

NER plays a significant role in enhancing the accessibility of legal texts by extracting elements such as law references, jurisdictions or court decisions [11,24,25,26]. Another related task is referred to as legal text segmentation or the classification of legal sentences, as exemplified in works like [27,28]. A common subsequent task involves linking named entities, tags or references in legal texts to corresponding vocabularies, conceptual graphs or ontologies (see [29,30,31,32,33]). Recently, an increasing number of integrated annotation platforms have emerged for the classification of legal statements [30,34,35]. Such platforms are designed with the objective of generating formal representations of legal texts in legal knowledge bases [36], and could, in principle also be applied in the context of the problem we have addressed here with a bespoke tool.

In the study presented in [37,38], the authors delineate some reasons behind the relatively low scores observed in NER performance. They attribute these results to three key factors: a limited number of tokens within the entity of interest, an elevated average segment length, and a high level of semantic similarity.

The intersection between nanopublication principles and legal data is an area of increasing interest, though it remains less explored compared to other domains [39]. In [40], it was demonstrated that RDF-star is well-suited for knowledge exploration and systematic querying in knowledge graphs, effectively addressing the limitations of reification. However, the authors of [41] argue that despite the adoption of RDF-star by various libraries and graph stores, the generation of RDF-star graphs remains largely unexplored. Our present work enables richer semantic annotations that encompass both the data and its provenance and context, significantly enhancing the interpretability and usability of legal knowledge compared to previous studies that focus primarily on entity extraction without addressing the formal representation of the extracted information.

6. Final Considerations

The Brazilian Judiciary has already embraced artificial intelligence systems [42]; however, to the best of our knowledge, courts have yet to adopt ontology as a foundational artefact to promote interoperability [43] or to guide the application of learning-based approaches.

In this paper, we addressed the challenge of resemantizing data related to a specific type of appeal within the Brazilian legal system. This work demonstrates the feasibility of automating semantic annotations in a specialized legal domain using a tailored NLP pipeline, which includes the creation of a dataset, a custom NER model, an RDF-star based annotation framework with provenance and uncertainty, thereby enhancing legal data veracity and research efficiency while paving the way for AI-driven legal decision support systems.

A future challenge lies in extracting additional information and knowledge from the subjects of appeals and the reasoning of decisions, where the approach faces greater complexity due to the intricate nature of legal jargon. We also plan to conduct a thorough comparison with other LLMs on this specific task and dataset to better assess performance differences. A second evaluation by legal experts could also be valuable for assessing the system's capacity to meet legal objectives, particularly when annotations are deemed incorrect but may be considered less critical from the experts' perspective. We ultimately aim to encompass aspects such as events [44], law references [24], citations of court decisions [25] or jurisdictions [11]. Through enhanced decision annotations, our aim is to develop recommendation systems that rely on semantic evidences, as exemplified in [45].

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, CNPq (443130/2023-0, 313412/2023-5) and FAPES (2021-GL60J, 2022-NGKM5, 1022/2022). Dylan Faria Robson is supported by FAPES (615/2023 P 2023-Z3KQP). Thiago Vieira is supported by an undergraduate grant from the Institutional Scientific Initiation Program of the Vila Velha University.

References

[1] Costa MZ, Guizzardi G, Almeida JPA. On Capturing Legal Knowledge in Ontology and Process Models Combined. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2022. p. 267-72.

[2] Costa MZ, Vieira TBP, Bourguet JR, Guizzardi G, Almeida JPA. Enhancing Access to Legal Data through Ontology-based Representation: A Case Study with Brazilian Judicial Appeals. In: Proc. XVI Seminar on Ontology Research in Brazil (ONTOBRAS 2023). CEUR Workshop Proceedings; in press. .

[3] Sekine S, Ranchhod E. Named entities: recognition, classification and use. vol. 19. John Benjamins Publishing; 2009.

[4] Miranda LJ, Kádár Á, Boyd A, Landeghem SV, Søgaard A, Honnibal M. Multi hash embeddings in spaCy. CoRR. 2022;abs/2212.09255.

[5] Guizzardi G, Wagner G. Conceptual Simulation Modeling with Onto-UML. In: Proceedings of the Winter Simulation Conference. WSC '12. Winter Simulation Conference; 2012. .

[6] Guizzardi G, Wagner G. Towards ontological foundations for agent modelling concepts using the unified fundational ontology (UFO). In: Int. Bi-Conference Workshop on Agent-Oriented Information Systems. Springer; 2004. p. 110-24.

[7] Nielsen J. How to conduct a heuristic evaluation. Nielsen Norman Group. 1995;1(1):8.

[8] Mons B, Velterop J, et al. Nano-publication in the e-science era. In: Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009). vol. 523; 2009. .

[9] Hartig O, Thompson B. Foundations of an Alternative Approach to Reification in RDF. CoRR. 2014;abs/1406.3399. Available from: http://arxiv.org/abs/1406.3399.

[10] Hernandez LAM, Orozco ALS, García-Villalba LJ. Analysis of Digital Information in Storage Devices Using Supervised and Unsupervised Natural Language Processing Techniques. Future Internet. 2023;15(5):155.

[11] Leitner E, Rehm G, Schneider JM. A Dataset of German Legal Documents for Named Entity Recognition. In: Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, et al., editors. Proc. 12th Language Resources and Evaluation Conf., LREC 2020; 2020. p. 4478-85.

[12] Correia FA, de Almeida AAA, Nunes JL, Santos KG, Hartmann IA, Silva FA, et al. Fine-grained legal entity annotation: A case study on the Brazilian Supreme Court. Inf Process Manag. 2022;59(1):102794.

[13] Brugger T, Stürmer M, Niklaus J. MultiLegalSBD: A Multilingual Legal Sentence Boundary Detection Dataset. In: Grabmair M, Andrade F, Novais P, editors. Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law, ICAIL 2023, Braga, Portugal, June 19-23, 2023. ACM; 2023. p. 42-51. Available from: https://doi.org/10.1145/3594536.3595132.

[14] Mok WY, Mok JR. Legal Machine-Learning Analysis: First Steps towards A.I. Assisted Legal Research. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL 2019, Montreal, QC, Canada, June 17-21, 2019. ACM; 2019. p. 266-7. Available from: https://doi.org/10.1145/3322640.3326737.

[15] Mok JR, Mok WY, Mok RV. Sentence classification for contract law cases: a natural language processing approach. In: Maranhão J, Wyner AZ, editors. ICAIL '21: Eighteenth International Conference for Artificial Intelligence and Law, São Paulo Brazil, June 21 - 25, 2021. ACM; 2021. p. 260-1. Available from: https://doi.org/10.1145/3462757.3466074.

[16] de Araujo PHL, de Campos TE, de Oliveira RRR, Stauffer M, Couto S, Bermejo PHS. LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text. In: 13th Int Conf Computational Processing of the Portuguese Language. vol. 11122 of LNCS. Springer; 2018. p. 313-23.

[17] Contini A, Piccolo S, Zurita LL, Sadl U. Recognising Legal Characteristics of the Judgments of the European Court of Justice: Difficult but Not Impossible. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of Frontiers in A.I. and Appl. IOS Press; 2022. p. 164-9.

[18] Sarkar R, Ojha AK, Megaro J, Mariano J, Herard V, McCrae JP. Few-shot and Zero-shot Approaches to Legal Text Classification: A Case Study in the Financial Sector. In: Proc. Natural Legal Language Processing Workshop 2021. Association for Computational Linguistics; 2021. p. 102-6.

[19] Lee SM, Tan YH, Yu HT. LeArNER: Few-shot Legal Argument Named Entity Recognition. In: Proc. 19th Int Conf Artificial Intelligence and Law; 2023. p. 422-6.

[20] Chusri T, Arsaibun S, Chokesuwattanaskul P, Chuangsuwanich E, Rutherford AT. Few-Shot Law Retrieval System for Supreme Court Cases. In: 20th IEEE International Joint Conf Computer Science and Software Engineering, JCSSE 2023. IEEE; 2023. p. 84-9.

[21] Willian Sousa A, Fabro M. Iudicium textum dataset uma base de textos jurídicos para NLP. In: 2nd Dataset Showcase Workshop at SBBD (Brazilian Symposium on Databases). Fortaleza, Brazil; 2019. .

[22] Sabo IC, Billi M, Lagioia F, Sartor G, Rover AJ. Unsupervised Factor Extraction from Pretrial Detention Decisions by Italian and Brazilian Supreme Courts. In: Advances in Conceptual Modeling - ER 2022 Workshops, CMLS, EmpER, and JUSMOD. vol. 13650 of LNCS. Springer; 2022. p. 69-80.

[23] Ferrara A, Picascia S, Riva D. Few-Shot Legal Text Segmentation via Rewiring Conditional Random Fields: A Preliminary Study. In: Sales TP, Araújo J, Borbinha J, Guizzardi G, editors. Advances in Conceptual Modeling - ER 2023 Workshops, CMLS, CMOMM4FAIR, EmpER, JUSMOD, OntoCom, QUAMES, and SmartFood, Lisbon, Portugal, November 6-9, 2023, Proceedings. vol. 14319 of Lecture Notes in Computer Science. Berlin Heidelberg: Springer; 2023. p. 141-50.

[24] Dozier C, Kondadadi R, Light M, Vachher A, Veeramachaneni S, Wudali R. Named Entity Recognition and Resolution in Legal Text. In: Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language. vol. 6036 of LNCS. Springer; 2010. p. 27-43.

[25] Peikert S, Birle C, Qundus JA, Vu LDS, Paschke A. Extracting References from German Legal Texts Using Named Entity Recognition. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2022. p. 231-6.

[26] Branting K, Brown B, Giannella C, Van Guilder J, Harrold J, Howell S, et al. Decision support for detecting sensitive text in government records. Artificial Intelligence and Law. 2023.

[27] Licari D, Comandé G. ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law. In: Companion Proc 23rd Int Conf Knowledge Engineering and Knowledge Management. vol. 3256 of CEUR Workshop Proceedings. CEUR-WS.org; 2022. .

[28] Wu T, Kao B, Cheung MMK. Judgment Retrieval Made Easier Through Query Analysis. In: Sileno G, Spanakis J, van Dijck G, editors. Legal Knowledge and Information Systems - JURIX 2023: The Thirty-sixth Annual Conference, Maastricht, The Netherlands, 18-20 December 2023. vol. 379 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2023. p. 299-304. Available from: https://doi.org/10.3233/FAIA230978.

[29] Cardellino C, Teruel M, Alemany LA, Villata S. A low-cost, high-coverage legal named entity recognizer, classifier and linker. In: Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law, ICAIL 2017, London, United Kingdom, June 12-16, 2017. ACM; 2017. p. 9-18.

[30] Tamper M, Oksanen A, Tuominen J, Hietanen A, Hyvönen E. Automatic Annotation Service APPI: Named Entity Linking in Legal Domain. In: The Semantic Web: ESWC 2020 Satellite Events - ESWC 2020 Satellite Events, Revised Selected Papers. vol. 12124 of LNCS. Springer; 2020. p. 208-13.

[31] Ren Y, Han J, Lin Y, Mei X, Zhang L. An Ontology-Based and Deep Learning-Driven Method for Extracting Legal Facts from Chinese Legal Texts. Electronics. 2022;11(12).

[32] Wu T, Kao B, Chan H, Cheung MM. Judgment Tagging and Recommendation Using Pre-Trained Language Models and Legal Taxonomy. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of Frontiers in Artificial Intelligence and App. IOS Press; 2022. p. 255-60.

[33] Ferrara A, Picascia S, Riva D. Context-Aware Knowledge Extraction from Legal Documents Through Zero-Shot Classification. In: Advances in Conceptual Modeling - ER 2022 Workshops, CMLS, EmpER, and JUSMOD. vol. 13650 of LNCS. Springer; 2022. p. 81-90.

[34] Huang Y, Lin H, Liu C. Toward an Integrated Annotation and Inference Platform for Enhancing Justifications for Algorithmically Generated Legal Recommendations and Decisions. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of FAIA. IOS Press; 2022. p. 281-5.

[35] Mahmoudi SA, Zambrano G, Condevaux C, Mussard S. Scribe: A Specialized Collaborative Tool for Legal Judgment Annotation. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual

Conf. vol. 362 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2022. p. 290-3.
[36] Libal T. The LegAi Editor: A Tool for the Construction of Legal Knowledge Bases. In: Legal Knowledge and Information Systems - JURIX 2022: 35th Annual Conf. vol. 362 of FAIA. IOS Press; 2022. p. 286-9.
[37] Moreno-Acevedo SA, Escobar-Grisales D, Vásquez-Correa JC, Orozco-Arroyave JR. Comparison of Named Entity Recognition Methods on Real-World and Highly Imbalanced Business Document Datasets. In: Proc. 9th Workshop on Engineering Applications. vol. 1685 of CCIS. Springer; 2022. p. 41-53.
[38] Bönisch K, Abrami G, Wehnert S, Mehler A. BUNDESTAG-MINE: Natural Language Processing for Extracting Key Information from Government Documents. In: Sileno G, Spanakis J, van Dijck G, editors. Legal Knowledge and Information Systems - JURIX 2023: The Thirty-sixth Annual Conference, Maastricht, The Netherlands, 18-20 December 2023. vol. 379 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2023. p. 391-4. Available from: https://doi.org/10.3233/FAIA230996.
[39] Oliveira B, Sousa C. Towards a KOS to Manage and Retrieve Legal Data. In: Rocha Á, Adeli H, Dzemyda G, Moreira F, Colla V, editors. Information Systems and Technologies - WorldCIST 2023, Volume 2, Pisa, Italy, April 4-6, 2023. vol. 800 of Lecture Notes in Networks and Systems. Springer; 2023. p. 75-84.
[40] Iglesias-Molina A, Ahrabian K, Ilievski F, Pujara J, Corcho Ó. Comparison of Knowledge Graph Representations for Consumer Scenarios. In: Payne TR, Presutti V, Qi G, Poveda-Villalón M, Stoilos G, Hollink L, et al., editors. The Semantic Web - ISWC 2023 - 22nd International Semantic Web Conference, Athens, Greece, November 6-10, 2023, Proceedings, Part I. vol. 14265 of Lecture Notes in Computer Science. Springer; 2023. p. 271-89.
[41] Arenas-Guerrero J, Iglesias-Molina A, Chaves-Fraga D, Garijo D, Corcho O, Dimou A. Declarative generation of RDF-star graphs from heterogeneous data. Submitted to Semantic Web. 2024.
[42] Saldanha PM. Processo Judicial e Pós-humanidade: transformação do Judiciário e a preservação da jurisdição humana pelo 2o grau de jurisdição. Universidade Católica de Pernambuco; 2021. Ph.D. thesis.
[43] Bourguet J, Costa MZ. About the Exposition of Brazilian Jurisprudences. In: Proc. IX ONTOBRAS Brazilian Ontology Research Seminar. vol. 1862 of CEUR Workshop Proceedings; 2016. p. 138-43.
[44] Navas-Loro M, Rodríguez-Doncel V. WhenTheFact: Extracting Events from European Legal Decisions. In: Legal Knowledge and Information Systems - JURIX 2022: 35h Annual Conf. vol. 362 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2022. p. 219-24.
[45] Bourguet J, Costa MZ. Scoring Judicial Syllabi in Portuguese. In: Legal Knowledge and Information Systems - JURIX 2017: 30th Annual Conf. vol. 302 of FAIA. IOS Press; 2017. p. 119-24.