PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

William Paulo Ducca Fernandes

Extracting and connecting plaintiff's legal claims and judicial provisions from Brazilian Court decisions

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática.

Advisor : Prof. Hélio Côrtes Vieira Lopes
Co-advisor: Profª. Simone Diniz Junqueira Barbosa

Rio de Janeiro
April 2020

Pontifícia Universidade Católica
do Rio de Janeiro

William Paulo Ducca Fernandes

Extracting and connecting plaintiff's legal claims and judicial provisions from Brazilian Court decisions

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática. Approved by the Examination Committee.

Prof. Hélio Côrtes Vieira Lopes
Advisor
Departamento de Informática – PUC-Rio

Profª. Simone Diniz Junqueira Barbosa
Co-advisor
Departamento de Informática – PUC-Rio

Profª. Clarisse Sieckenius de Souza
Departamento de Informática – PUC-Rio

Profª. Caitlin Sampaio Mulholland
Departamento de Direito – PUC-Rio

Dr. Vinícius Costa Villas Bôas Segura
IBM Research Brazil

Prof. Alex Laier Bordignon
UFF

Prof. Marcus Vinicius Soledade Poggi de Aragão
Departamento de Informática – PUC-Rio

Prof. Carlos Nelson de Paula Konder
Departamento de Direito – PUC-Rio

Prof. Marcos de Oliveira Lage Ferreira
UFF

Rio de Janeiro, April 8^{th}, 2020

William Paulo Ducca Fernandes

Bachelor's in Computer Science (2008) at Universidade Federal de Juiz de Fora (UFJF). Master's in Informatics (2012) at Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Started his doctorate at PUC-Rio in 2016, focusing his research on Machine Learning, Natural Language Processing, and Information Extraction applied to Law.

Bibliographic data

Fernandes, William Paulo Ducca

Extracting and connecting plaintiff's legal claims and judicial provisions from Brazilian Court decisions / William Paulo Ducca Fernandes; advisor: Hélio Côrtes Vieira Lopes; co-advisor: Simone Diniz Junqueira Barbosa. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2020.

v., 96 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Processamento de linguagem natural. 2. Aprendizado profundo. 3. Redes neurais recorrentes. 4. Long short-term memory. 5. Aprendizado de máquina. 6. Conditional random fields. 7. Extração de informação. 8. Direito. 9. Provisões modificatórias. I. Lopes, Hélio Côrtes Vieira. II. Barbosa, Simone Diniz Junqueira. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Acknowledgments

Firstly, I would like to thank God for guiding me through this 4-year journey. It has been a long way. Only by His grace, I'm here finishing this step, with my heart full of gratitude. Also I thank my parents, Ricardo and Delza, for being so present in my life. Thank you for all dedication to your children. This moment is a result of such dedication.

I thank my family – André, Jesana, Rebeca, Tiago, Isabella, and Manuella – for all the moments together. With you, my life is happier. I want to say thank you to Hélio, my advisor, and to Simone, my co-advisor, for counselling me along these years, and for teaching me many lessons. I would like to thank Isabella Frajhof for the strong collaboration, and enthusiasm, which have been very important to produce this work. Also I thank Carlos Konder, Guilherme Almeida, Ariane, Jonatas, and Jefry who also worked with me in parts of this project, from theoretical discussions to code implementation. I thank my laboratory friends for the discussions, the friendship, the assistance with the annotation, and for making the laboratory a place I loved to be. Also I thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the scholarship. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Abstract

Fernandes, William Paulo Ducca; Lopes, Hélio Côrtes Vieira (Advisor); Barbosa, Simone Diniz Junqueira (Co-Advisor). **Extracting and connecting plaintiff's legal claims and judicial provisions from Brazilian Court decisions**. Rio de Janeiro, 2020. 96p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

In this work, we propose a methodology to annotate Court decisions, create Deep Learning models to extract information, and visualize the aggregated information extracted from the decisions. We instantiate our methodology in two systems we have developed. The first one extracts Appellate Court modifications, a set of legal categories that are commonly modified by Appellate Courts. The second one (i) extracts plaintiff's legal claims and each specific provision on legal opinions enacted by lower and Appellate Courts, and (ii) connects each legal claim with the corresponding judicial provision. The system presents the results through visualizations. Information Extraction for legal texts has been previously addressed using different techniques and languages. Our proposals differ from previous work, since our corpora are composed of Brazilian lower and Appellate Court decisions. To automatically extract that information, we use a traditional Machine Learning approach and a Deep Learning approach, both as alternative solutions and also as a combined solution. In order to train and evaluate the systems, we have built Kauane Junior corpus for the first system, and three corpora for the second system – Kauane Insurance Report, Kauane Insurance Lower, and Kauane Insurance Upper. We used public data disclosed by the State Court of Rio de Janeiro to build the corpora. For Kauane Junior, the best model, which is a Bidirectional Long Short-Term Memory network combined with Conditional Random Fields (BILSTM-CRF), obtained an $F_{\beta=1}$ score of 94.79%. For Kauane Insurance Report, the best model, which is a Bidirectional Long Short-Term Memory network with character embeddings concatenated to word embeddings combined with Conditional Random Fields (BILSTM-CE-CRF), obtained an $F_{\beta=1}$ score of 67.15%. For Kauane Insurance Lower, the best model, which is a BILSTM-CE-CRF, obtained an $F_{\beta=1}$ score of 89.12%. For Kauane Insurance Upper, the best model, which is a BILSTM-CRF, obtained an $F_{\beta=1}$ score of 83.66%.

Keywords

Natural language processing; Deep learning; Recurrent neural networks; Long short-term memory; Machine learning; Conditional random fields; Information extraction; Law; Modificatory provisions.

Resumo

Fernandes, William Paulo Ducca; Lopes, Hélio Côrtes Vieira; Barbosa, Simone Diniz Junqueira. **Extração e conexão entre pedidos e decisões judiciais de um tribunal brasileiro**. Rio de Janeiro, 2020. 96p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Neste trabalho, propomos uma metodologia para anotar decisões judiciais, criar modelos de Deep Learning para extração de informação, e visualizar de forma agregada a informação extraída das decisões. Instanciamos a metodologia em dois sistemas. O primeiro extrai modificações de um tribunal de segunda instância, que consiste em um conjunto de categorias legais que são comumente modificadas pelos tribunais de segunda instância. O segundo (i) extrai as causas que motivaram uma pessoa a propor uma ação judicial (causa de pedir), os pedidos do autor e os provimentos judiciais dessas ações proferidas pela primeira e segunda instância de um tribunal, e (ii) conecta os pedidos com os provimentos judiciais correspondentes. O sistema apresenta seus resultados através de visualizações. Extração de Informação para textos legais tem sido abordada usando diferentes técnicas e idiomas. Nossas propostas diferem dos trabalhos anteriores, pois nossos corpora são compostos por decisões de primeira e segunda instância de um tribunal brasileiro. Para extrair as informações, usamos uma abordagem tradicional de Aprendizado de Máquina e outra usando Deep Learning, tanto individualmente quanto como uma solução combinada. Para treinar e avaliar os sistemas, construímos quatro corpora: Kauane Junior para o primeiro sistema, e Kauane Insurance Report, Kauane Insurance Lower e Kauane Insurance Upper para o segundo. Usamos dados públicos disponibilizados pelo Tribunal de Justiça do Estado do Rio de Janeiro para construir os corpora. Para o Kauane Junior, o melhor modelo ($F_{\beta=1}$ de 94.79%) foi uma rede neural bidirecional Long Short-Term Memory combinada com Conditional Random Fields (BILSTM-CRF); para o Kauane Insurance Report, o melhor ($F_{\beta=1}$ de 67,15%) foi uma rede neural bidirecional Long Short-Term Memory com embeddings de caracteres concatenados a embeddings de palavras combinada com Conditional Random Fields (BILSTM-CE-CRF). Para o Kauane Insurance Lower, o melhor ($F_{\beta=1}$ de 89,12%) foi uma BILSTM-CE-CRF; para o Kauane Insurance Upper, uma BILSTM-CRF ($F_{\beta=1}$ de 83,66%).

Palavras-chave

Processamento de linguagem natural; Aprendizado profundo; Redes neurais recorrentes; Long short-term memory; Aprendizado de máquina; Conditional random fields; Extração de informação; Direito; Provisões modificatórias.

Table of Contents

List of Figures

List of Tables

I'm so glad to know that He [Jesus] can be found amongst the brokenhearted. In our sorrows, He's not one that would leave us. He stands by us when all has failed, and the last hopes of earthly reaching has come to its end, He is still God and He loves us.

William Marrion Branham, *Living, Dying, Buried, Rising, Coming.*

1
Introduction

Artificial Intelligence and Law, a research field since the 1980s, is facing a revolution. Teams of researchers of question answering, information extraction, and argument mining are responsible for the beginning of such revolution. New legal systems are being developed, using technologies developed by those teams (Ashley, 2017).

Now, legal texts – such as cases, Court decisions, statutes, regulations – can be processed automatically, and useful information can be extracted from them. Judges and lawyers can benefit from this by using aggregated information extracted from past cases to determine what decision or approach would be the most recommended for a given situation.

Most legal systems rely on an artificial simplification of the legal process. Usually legal cases deal with multiple issues: a criminal procedure may include several different charges, whilst a civil procedure may include several different requests. When those multiple charges and requests are boiled down to a single case-level variable, valuable information is lost. For instance, cases brought before the European Court of Human Rights may involve alleged violations of several different provisions of the European Union. Much useful legal data gets lost when we reduce these multiple different legal issues to a single measure by treating cases where any provisions were deemed to be violated as cases of human rights violations (+1), and cases where no provisions were found to be infringed upon as the complementary category (-1) (Medvedeva et al., 2019, Aletras et al., 2016). Which provisions were violated? How many of the allegations were accepted by the Court? What legal rights were altered? Those are questions that the simplifying approach traditionally used in the literature cannot answer. However, they are some of the central questions that occupy the minds of lawyers in their practice.[1] Reflecting the importance of more granular information regarding judicial decision-making, recent work at the intersection of artificial intelligence and the practice of law has focused on formally conceptualizing (Walton, 2019) and identifying legal arguments within broader legal documents. However, the technical challenge involved in correlating specific legal claims and their judicial solution is

[1]For an account of those questions, see (Schauer, 2009).

acknowledged to be hard (Correia et al., 2019, Fernandes et al., 2019).

In light of such challenges and opportunities, this thesis aims at answering the research questions presented below.

Q1: How to know what the Court holds to be the law?

Q2: How to see trends of judgments in Court decisions?

Q3: How to predict judicial outcomes for specific situations?

In order to answer those questions, we propose a methodology to annotate Court decisions, create Deep Learning models to extract information, and visualize the aggregated information extracted from the decisions. Our methodology is instantiated in two systems we have developed. The first one extracts Appellate Court modifications, which consists of a set of legal categories that are commonly modified by Appellate Courts, in a narrow legal domain of civil suits regarding *negativação indevida*, which is a business practice that unlawfully downgrades one's credit score.

This system can help lawyers to determine what the Court holds to be the law, by extracting the standards proposed by the Appellate Court modifications, answering Q1. Moreover, it can assist judges in finding whether their decisions conform to the Appellate Court's overall stance or not, leading to judicial consistency and certainty. Furthermore, it can assist lawyers to verify trends of judgments by analyzing aggregated Appellate Court modifications, helping them to decide whether to appeal from a lower Court decision or not, answering Q2.

The second system (i) extracts plaintiff's legal claims and each specific provision on legal opinions enacted by lower and Appellate Courts, and (ii) connects each legal claim with the corresponding judicial provision. The proposed system then presents the results through visualizations that let one verify the consistency of Court decisions for a given group of claims, as well as the variation between different groups of legal claims. We applied the system to the narrow legal domain of civil suits involving the refusal of insurance companies on providing assistance to their customers.

Like the first system, the second one allows answering Q1 and Q2. In addition, the second system can help lawyers in predicting judicial outcomes for very specific situations. For instance, given that the plaintiff suffered cancer and was denied coverage by his healthcare provider, what is the usual legal relief? The proposed system leverages information regarding the historic trends of the Court to render an answer to that specific question, answering Q3.

Our proposals differ from previous work (Nguyen et al., 2018, de Araujo et al., 2017, Dragoni et al., 2016, Angelidis et al., 2018, Trompper and Winkels, 2016, Garcia et al., 2017, Boella et al., 2014, Nanda et al., 2017b, Gianfelice et al., 2013), since our corpora are composed of Brazilian lower and Appellate Court decisions, in which, for the first system, we look for a set of modifications commonly provided by the Court, and for the second system, we look for a set of plaintiff's legal claims and judicial provisions commonly judged by the Court. To automatically extract that information, we use a traditional Machine Learning approach and a Deep Learning approach, both as alternative solutions and also as a combined solution. We use five algorithms in one or both systems: (i) Long Short-Term Memory network (Hochreiter and Schmidhuber, 1997, Graves and Schmidhuber, 2005); (ii) Bidirectional Gated Recurrent Units network (Graves et al., 2013); (iii) Conditional Random Fields (Lafferty et al., 2001); (iv) a combination of Bidirectional Long Short-Term Memory network and Conditional Random Fields (Huang et al., 2015); and (v) a combination of Bidirectional Gated Recurrent Units network and Conditional Random Fields (Huang et al., 2015).

Since we use supervised Machine Learning algorithms, we need annotated datasets to train and evaluate the systems. For the first system, we have built the KAUANE JUNIOR corpus, using public jurisprudence data disclosed by the Appellate State Court of Rio de Janeiro. For the second system, we have built three corpora – KAUANE INSURANCE REPORT, KAUANE INSURANCE LOWER, and KAUANE INSURANCE UPPER – also using public jurisprudence data disclosed by the Appellate State Court of Rio de Janeiro. We generated gold standard annotations for the corpora, produced by human annotators and considered as the groundtruth for training and test. In addition, we have included part-of-speech (POS) annotation in the corpora.[2]

The remainder of this thesis is organized as follows. In chapter 2, we present related work in the literature, laying out our contribution. The models which are the basis for our systems are explained in chapter 3. In chapter 4, we describe the extraction tasks for each corpus of our systems. The experiments are reported in chapter 5. Finally, we present our conclusion and future work in chapter 6.

[2]We used an in-house BILSTM-based tagger trained for Portuguese.

2
Related work

In this chapter, we present an overview of the existing work concerning information extraction applied to legal texts. In each case, we show similarities and differences from our approaches. Moreover, we review the neural network approach which is the basis for our systems and present other references.

2.1
Information Extraction Applied to Legal Texts

Table 2.1 presents a summary of the related work on information extraction in the context of legal texts and our work. The works are ordered by the date of publication.

Table 2.1: Summary of the related work on information extraction in the context of legal texts and our work.

Work	*Purpose*	*Source*	*Language*	*Techniques*
(Gianfelice et al., 2013)	Extract information from legal documents	Laws, decree-laws, other sorts of decrees, and regulation statuses	Italian	Regular expressions
(Boella et al., 2014)	Extract information from textual data	Legal dataset, definitional sentences from Wikipedia	English	Traditional ML
(Dragoni et al., 2016)	Structure legal information and organize it in logical statements	Consumer protections code	English	Manual rules
(Trompper and Winkels, 2016)	Assign a section structure to Court judgments	Case law	Dutch	Traditional ML
(de Araujo et al., 2017)	Extract information from Court decisions	Appellate Court decisions	Portuguese	Manual rules
(Garcia et al., 2017)	Extract information from legal documents	Commercial law documents	English	Gazetteer for NER, grammar rules
(Nanda et al., 2017b)	Extract entities from legal documents	European directives and the UK national law	English	Traditional ML
(Angelidis et al., 2018)	Extract named entities from legal documents	Legislation documents	Greek	Deep Learning
(Nguyen et al., 2018)	Separate the main parts of legal documents	Code of laws	English, Japanese	Deep Learning
Ours	Extract and connect information from Court decisions	Lower and Appellate Court decisions	Portuguese	Deep Learning

Gianfelice et al. propose a system to automatically annotate modificatory provisions in Italian normative texts (Gianfelice et al., 2013). A modificatory provision is a change made to one or more clauses within a text, to the whole text, or to the relations among the constituent provisions of a legal system.

The sources of information are legal, such as laws, decree-laws, other sorts of decrees, and regulation statuses. The system is mainly concerned with assisting legal annotators in locating and qualifying modificatory provisions.

In Italy, they use a standard format for Legal Text, the NormeInRete (NIR) format. The NIR format encodes the structural elements used to mark up the main partitions of legal texts, as well as their atomic parts (such as articles, paragraphs, and numbered items) and any unstructured text fragment. That format is implemented in XML.

Their system couples two techniques: deep parsing and regular expressions. The first step of the system consists of extracting specific XML nodes where modifications may be found. The extracted text is slightly rewritten in order to simplify the input. For instance, thousands of separators are removed; typos due to wrong characters encoding are converted to proper characters; and NIR constants are capitalized in order to ease their recognition by the parser.

Then, the text is split into sentences and each sentence is processed. All the implemented modifications are tested using a set of regular expressions. Sentences where a modification may be found are parsed. The parsing information is used to identify the subject and object that indicate the norms involved in the modification. Finally, the extracted information is assigned to its respective semantic frame (Fillmore, 1977).

They evaluated the system in a legal corpus annotated by legal experts of the University of Bologna. The whole dataset contains more than 12,000 files. The system presented a precision[1] of 47%, a recall[2] of 61%, and an $F_{\beta=1}$[3] of 53.09%.

The purpose of their work was to extract pieces of information that correspond to modifications in legal documents. Their approach uses documents in the context of law creation, and their documents are in Italian. They apply an algorithm with several sets of regular expressions in order to identify modifications in the text. The objective of our first system is to extract pieces of information that correspond to modifications in legal documents. The purpose in our second system is to extract legal claims and each specific provision on legal opinions enacted by lower and Appellate Courts. For both systems, we use documents of Court decisions; all documents are in Portuguese. We use Machine Learning algorithms to extract the proposed entities.

De Araujo et al. propose a system for ontology-based information extraction from Appellate Court decisions (de Araujo et al., 2017). The system

[1]Precision is the proportion of positive identifications that was actually correct.
[2]Recall is the proportion of actual positives that was correctly identified.
[3]$F_{\beta=1}$ is the harmonic mean of the precision and recall.

is based on a domain ontology of legal events and a set of linguistic rules, which are integrated through an inference mechanism.

Their approach comprises two phases. The first phase, called linguistic phase, is the corpus study performed to describe the domain ontology and the linguistic rules. In that phase, a linguistic expert and a law expert study the target corpus and generate two ontologies: the extraction ontology and the domain ontology. Both ontologies are formalized in OWL.[4] The second phase, called computational phase, comprises the description of the linguistic rules in ontologies, the construction and update of the domain ontologies with law concepts, and the use of an inference mechanism to annotate the corpus. In this latter phase, the documents are first parsed with a deep linguistic parser, which provides syntactic, morphologic, and shallow semantic annotation. Next, that information is translated into an OWL representation with the POWLA data model (Chiarcos, 2012), and the ontologies are loaded and merged. The logical inference system is then started: it finds, through logical inference, the textual terms that reference the domain ontology concepts. The inferred references to concepts are stored in a knowledge base.

They conducted a case study to automatically identify the legal events *formal charges*, *acquittal*, *conviction*, and *questioning*. To elaborate the linguistic rules, they built a corpus with ten decisions from the jurisprudence database of the Appellate State Court of Rio Grande do Sul. In order to evaluate their method, they built a corpus with 200 decisions from the same database. Their system obtained an overall precision of 98.5%, an overall recall of 91.5%, and an overall $F_{\beta=1}$ of 94.87%.

Their approach uses Court decisions from the Appellate Court of Rio Grande do Sul, establishing some legal events of interest. Their work is based on a time-consuming corpus analysis, in which they need experts to create both the domain ontology and the linguistic rules. Our first system uses Court decisions from the Appellate Court of Rio de Janeiro, focusing on one specific legal event – *decision* – which is not covered by their ontology. Our second system uses Court decisions from the lower and Appellate Court of Rio de Janeiro, focusing on two specific legal events – *legal claims* and *decision* – which are also not covered by their ontology. For both systems, we use Deep Learning to automatically create rules to identify participants of a legal event. We need only to provide annotated corpora to our Deep Learning methods.

Dragoni et al. present a framework for the automated extraction of rules from legal texts (Dragoni et al., 2016). The framework uses syntax-based patterns extracted by the Stanford Parser (Klein and Manning, 2003) and

[4]OWL is a language for authoring ontologies.

logic-based patterns extracted by the Boxer framework (Curran et al., 2007). The idea is to extract rules that express deontic meaning (*i.e.*, prohibition, permission, obligation) from legal texts. They combined different natural language processing (NLP) approaches to automatically extract a set of rules from legal texts. They use two ontologies, one to identify deontic linguistic elements and another to describe how the natural language text is structured.

Their framework uses two parallel branches, implementing two different analysis techniques. The first branch tags the sentence using the Stanford Parser, building the related tree for extracting the terms contained in each sentence. Then, the deontic ontology is applied to annotate each term with the appropriate label (obligation, permission, prohibition). Finally, the system looks for patterns within the terms set of each sentence to compose the rules. The second branch applies a Combinatory Categorical Grammar (CCG) parser to the sentence to extract logical relationships between terms. The output of the CCG parser is then used to confirm the rules extracted by the first analysis technique, and to discover new relationships between terms (*i.e.*, relationships which had not been detected yet).

They evaluated the system in the section Compliant Management of the Australian Telecommunications Consumer Protections Code. They compared the handcrafted rules created by an analyst to the automatically generated rules. The second branch presented the best quality, achieving a precision of 80.56%. The recall measure was not objectively informed.

Their work aims at structuring legal information and organizing it in logical statements. The authors point as the main advantage of their work not having to manually annotate a dataset in order to train a Machine Learning model. However, they had to manually create rules to extract terms, and to manually define a set of patterns for creating the rules. The created rules combine the extracted terms. The objective of our systems is to extract pieces of legal information. We have spent time in creating annotations in order to train Machine Learning models. Our sets of documents for annotation are the result of searches in the jurisprudence database of the State Court of Rio de Janeiro. So they are composed of a great variety of documents of different writing styles. If one would create rules manually for those documents, it would require a big number of rules to treat all different cases, and it would be error prone.

Boella et al. propose a technique for the automatic extraction of semantic information from textual data (Boella et al., 2014). The purpose of their work is to identify semantic units that can be used to extract structured knowledge and to efficiently compute semantic searches in texts belonging to different domains. They first apply a dependency parser in the text in order to obtain lexical and syntactic information. They then transform that information into more generalized features with the purpose of capturing the linguistic variability of the syntactic dependents of the target label. That transformed dataset is used to train an SVM classifier to semantically annotate the text.

The technique was evaluated in two datasets in English, using a 10-fold cross-validation. The first was a legal dataset with 560 legal texts annotated with various semantic information, from which the work focused on identifying three types: active role, passive role, and object involved. The second was an annotated dataset of definitional sentences containing more than 4,000 sentences extracted from Wikipedia. On the first dataset, the system obtained a global precision of 85.13%, a global recall of 74.17%, and a global $F_{\beta=1}$ of 79.27%. On the second dataset, the system presented a precision of 83.05%, a recall of 68.64%, and an $F_{\beta=1}$ of 75.16%.

Their work focuses on identifying general semantic units, which are present in texts of any domain. They use the tree generated by a dependency parser as an auxiliary feature to an SVM classifier. By contrast, our systems focus on extracting specific information from Court decisions. We use a Deep Learning approach that does not require auxiliary features to generate the annotation. Another advantage of our systems is that, with the information generated by our models, it is possible to extract specific knowledge of Court decisions, such as the mean value paid for moral damage.

Garcia et al. propose a system and methodology called CLIEL (Commercial Law Information Extraction based on Layout) for annotating legal documents using XML tags (Garcia et al., 2017). Their approach uses NLP through some GATE (General Architecture for Text Engineering)[5] modules and JAPE (Java Annotation Patterns Engine)[6] rules. Their idea is to provide a mechanism for annotating legal documents using XML tags in order to facilitate the information extraction of dates, legal framework, named entities, and so on. They extract two kinds of information: structural, where they delimit sections, subsections, appendices, etc. within a document; and semantic, where they extract the following fields: date of document, name of party, name of counterparty, governing law, and jurisdiction.

Their system first applies a tokenizer and a sentence splitter to the input using GATE in order to split the text into units (tokens and sentences) and then it uses gazeteers to identify names of entities. The system applies JAPE grammar rules to the text in order to identify and generate XML annotations

[5]https://gate.ac.uk/sale/tao/splitch8.html
[6]https://gate.ac.uk/

for document sections. The result is parsed into a tree structure, where each node represents a text unit, such as title, heading, or paragraph. In the sequence, the system processes the tree structure to generate one XML file per document. JAPE grammar rules are then applied to the text in order to generate XML annotations for specific information. Finally, the annotated information is extracted and stored properly.

They built a corpus of a set of 97 commercial law documents in English. They used 20 documents as a training set to generate JAPE rules. The other 77 were used for testing. They compared their approach with two other systems, outperforming them when comparing the $F_{\beta=1}$ scores. Their system achieved a general precision of 80.37%, a general recall of 61.14%, and a general $F_{\beta=1}$ of 69.45%.

Their work has the purpose of extracting information from legal documents, which are commercial law documents. They use a gazetteer to identify named entities and apply rules to generate annotations. Our systems also have the objective of extracting information from legal documents, which are Court decisions. However, our approach does not use any fixed source of information, such as a gazetteer, and the rules are automatically learned by Machine Learning models.

Nguyen et al. propose some systems for recognizing requisite and effectuation (RE) parts in legal sentences using Deep Learning models (Nguyen et al., 2018). A legal sentence can be separated into two main parts: a requisite part and an effectuation part. Those parts are used to create legal structures of law provisions in legal sentences. Depending on the annotation scheme, there may be an overlap between the requisite and effectuation parts.

In their first system, they propose a modification of BILSTM-CRF in order to integrate external features to facilitate the recognition of non-overlapping RE parts. Their second system uses a sequence of n separate BILSTM-CRF models, each model recognizing labels in its layer. These labels are used as features for predicting labels at higher layers. According to the authors, that approach is inconvenient, since many single models have to be trained.

Their two latest systems – a multilayer BILSTM-CRF and a multilayer BILSTM-MLP-CRF – recognize labels of all layers at the same time. The multilayer BILSTM-CRF model is constructed from many BILSTM-CRF components. The input of a component at a certain layer is a sequence of vectors. Each vector is the concatenation of word embedding, feature embedding, and tag score vector from previous layers. The multilayer BILSTM-MLP-CRF has only one BILSTM component, which is used to encode the input sentence into a sequence of hidden states. This information is used as the input for the following MLP-CRF (Multilayer Perceptron followed by CRF) components, which also use tag score vectors from previous layers.

They work with two datasets, Japanese National Pension Law RRE (JPL-RRE) and Japanese Civil Code RRE (JCC-RRE). The first one is a non-overlapping dataset in Japanese. Thus they recognize RE parts using a single BILSTM-CRF. The second one is a version of the Japanese Civil Code translated into English. Different from JPL-RRE, RE parts may be overlapped. They generate syntactic information using a state-of-the-art parser.

In the JPL-RRE dataset, they compared BILSTM-CRF with four systems, outperforming all of them. Their system achieved a precision of 92.77%, a recall of 93.77%, and an $F_{\beta=1}$ of 93.27%. In the JCC-RRE dataset, they compare the proposed models that deal with overlapping RE parts with one system, outperforming it. Using only word features, the two best models, multilayer BILSTM-CRF and sequential BILSTM-CRF, achieved the same $F_{\beta=1}$ score of 77.40%. Using word features combined with syntactic features, the two best models, multilayer BILSTM-MLP-CRF and sequential BILSTM-CRF, achieved an $F_{\beta=1}$ score of 78.24% and 78.23%, respectively.

Their work aims at separating the main parts of legal documents, which are code of laws. They use Deep Learning methods to eliminate the need for manually engineering features. The purpose of our systems is to extract specific entities from legal documents, which are Court decisions. We also use Deep Learning techniques.

Angelidis et al. propose a named entity recognition (NER) system and a named entity linker (NEL) system for Greek legislation (Angelidis et al., 2018). Both systems are used in Nomothesia[7], a Web platform that makes Greek legislation available on the Web as linked open data.

The NER system extracts six entity types: person, organization, geopolitical entity, geographical landmark, legislation reference, and public document reference. The NEL system represents entity references identified by the NER system using the RDF specification. As the interlink of entities is not the purpose of our work, we focus only on the first system.

To build the NER system, they experimented with three LSTM-based methods. All the methods receive word embeddings and token shape embeddings as input. The first method, called BILSTM-LR, uses a Bidirectional LSTM connected with a Logistic Regression (LR) layer to estimate the probability that a token belongs to a category. The second method, called BILSTM-BILSTM-LR, has two Bidirectional LSTM layers in sequence, followed by a

[7]http://legislation.di.uoa.gr

LR layer. The third method, called BILSTM-CRF, uses a Bidirectional LSTM followed by a Conditional Random Fields (CRF) layer.

They built a corpus to evaluate their approaches. The corpus is composed of 254 daily issues of the Greek Government Gazette from 2000 to 2017. Every issue contains multiple legal acts. They annotated all the documents for the six entity types studied. They split the corpus in training set (162 issues), validation set (45 issues), and test set (47 issues).

They compared the quality of the proposed methods to each other. BILSTM-BILSTM-LR presented the best quality in the test set, achieving a general precision of 91%, a general recall of 85%, and a general $F_{\beta=1}$ of 88%.

Their techniques have the goal of extracting named entities from legal documents, which are a set of legislation documents. They use Deep Learning methods to eliminate the need of manually engineering features. Our systems have the purpose of extracting specific entities related to the judicial decision-making from legal documents, which are Court decisions. We also use Deep Learning techniques.

Trompper and Winkels propose a system to assign a section structure to Dutch Court judgments (Trompper and Winkels, 2016). The proposed system comprises two steps: (i) they label text elements with their roles in the document (text, title, numbering, or newline) using linear-chain Conditional Random Fields (CRF) and, (ii) based on the identified list of labels, they generate a parse tree using Probabilistic Context-Free Grammars, which represents the section hierarchy of a document. As step (ii) lies outside the scope of our work, we focus only on step (i).

They defined around 250 features to provide the CRF model. Those features consist mostly of regular expressions for known section title patterns. They built a corpus from an online open dataset of Dutch case law published by the Council for the Judiciary in the Netherlands[8]. They do not report the amount of data used for training and testing in step (i).

They compared their approach using CRF to a manually written tagger that uses many of the same features used by CRF. The proposed technique outperforms the manually written tagger when comparing recall and $F_{\beta=1}$ scores. The system achieved a general precision of 91%, a general recall of 91%, and a general $F_{\beta=1}$ of 91%.

Their purpose is to structure Court decisions. They use a traditional Machine Learning method, providing manually created features to it. Our systems aim at extracting pieces of legal information from Court decisions. We

[8]https://www.rechtspraak.nl

use Deep Learning models that automatically learn the features, eliminating the feature engineering step.

Nanda et al. propose a system for concept recognition in European directives and national law (statutory instruments of the United Kingdom) (Nanda et al., 2017b). They used the Inter-active Terminology for Europe[9] (IATE) vocabulary for developing their concept recognition system. IATE is the EU's inter-institutional terminology database, which consists of 1.3 million entries in English. Every concept is mapped onto a subject domain, which comprises the concepts to be learned.

They created a corpus of 2,884 directives and 2,884 statutory instruments (SIs) for their experiments. They used a semi-supervised approach to annotate the corpus. They manually annotated a few documents with IATE subject domains. Then, they developed a dictionary lookup program to tag terms in the text with IATE subject domains. They also employed a state-of-the-art NER system to generate annotation for time, date, and monetary units. They divided the dataset in 80% for training (2307 directives and 2307 SIs) and 20% for testing (577 directives and 577 SIs).

They used Conditional Random Fields (CRF) to build the concept recognition system. They provided CRF with the following features: word suffix, word identity (whether a word represents a subject domain/named entity or not), word shape (capitalized, lowercase, or numeric), and POS tags.

They compared their approach to a baseline model – which assign tags only to the most frequent class – and the Stanford NER. For the directive corpus, the proposed CRF model outperformed the baseline model and achieved the same quality compared to Stanford NER, presenting a general precision of 80%, a general recall of 71%, and a general $F_{\beta=1}$ of 75%. For the SIs corpus, the proposed model outperformed both baseline and Stanford NER, achieving a general precision of 73%, a general recall of 61%, and a general $F_{\beta=1}$ of 66%.

Their goal is to extract entities (concepts) from legal documents, which are legislation documents. They use a traditional Machine Learning method, providing manually created features to it. Our systems also aim at extracting entities from legal documents, which are Court decisions. We use Deep Learning models that automatically learn the features, eliminating the feature engineering step.

We now review the neural network approach which is the basis for our systems and present other references.

[9]http://iate.europa.eu

2.2 Neural Network Methods

Huang et al. propose a variety of LSTM based models for sequence tagging (Huang et al., 2015). The first one is composed of a LSTM layer followed by a CRF layer, denoted LSTM-CRF. The next one is composed of a Bidirectional LSTM (BILSTM) layer followed by a CRF layer, denoted BILSTM-CRF. They show that the BILSTM-CRF model can use past and future input features thanks to a BILSTM layer. In addition, that model uses sentence-level information thanks to a CRF component. That work was the first to apply a BILSTM-CRF model to NLP benchmark sequence tagging datasets.

They evaluated the system on three data sets for sequence labeling tasks – Penn Treebank for POS tagging, CoNLL 2000 for chunking, and CoNLL 2003 for NER. On the test sets, the system presented an accuracy of 97.55% for POS tagging, 94.46% for chunking, and 90.10% for NER. It represented the state-of-the-art for POS tagging and close to state-of-the-art for chunking, which at that time was 95.23%, and for NER, which at that time was 90.90%. Such models are explained in more detail in Chapter 3.

Beyond the representation of words using word embeddings (Mikolov et al., 2013, Pennington et al., 2014, Bojanowski et al., 2016), we also looked for inspiration in the work of (Ling et al., 2015), who propose a model for constructing vector representations of words by its composing characters using Bidirectional LSTMs. The proposed model produces a character embedding representation, aiming at capturing intra-word morphological and shape information.

They evaluated the system on POS tagging tasks for five languages. The system presented an accuracy of 97.36% for English, 97.47% for Portuguese, 98.92% for Catalan, 98.08% for German, and 91.59% for Turkish. The proposed system reached state-of-the-art performance for all datasets, except Portuguese, which was 97.54% at the time. We explain that model in more detail in Chapter 3.

There are other neural network techniques to approach the sequence labeling task, such as (Ma and Hovy, 2016), (Strubell et al., 2017), (Peters et al., 2017), (Peters et al., 2018a), (Peters et al., 2018b), just to cite a few.

In this chapter, we presented an overview of research on information extraction in the context of legal texts. We compared each related work to our approaches, presenting similarities and differences. To the best of our knowledge, those are the most similar works to ours in the literature, focusing on information extraction for legal text. We also reviewed the neural network approach which is the basis for our systems and presented other references. Next, we present the models applied to our systems, and also show the word representations adopted by them.

3
Models

In this chapter, we present five models applied to one or both of our systems: (i) a Bidirectional Long Short-Term Memory network; (ii) a Bidirectional Gated Recurrent Units network; (iii) a Conditional Random Fields model; (iv) a combination of Bidirectional Long Short-Term Memory network and Conditional Random Fields; and (v) a combination of Bidirectional Gated Recurrent Units network and Conditional Random Fields. We also show the word representations adopted by our systems.

3.1
Recurrent Neural Networks

A Recurrent Neural Network (RNN) is an extension of a feedforward neural network that can handle inputs of different sizes. It treats those different sizes through a recurrent state whose activation at each time depends on the activation at the previous time (Chung et al., 2014). However, RNNs are not good at capturing long-term dependencies, because of gradient vanishing and exploding problems (Bengio et al., 1994).

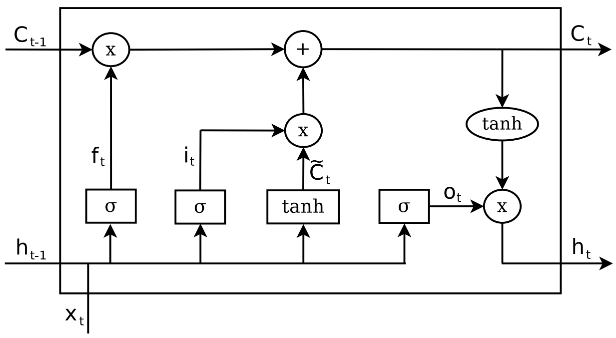


Figure 3.1: Illustration of the LSTM architecture. Each line carries one vector, from the output of one node to the input of one of more nodes. Circles mean point-wise operations and boxes are neural network layers. A line merge represents a concatenation of vectors, while a line fork symbolizes a copy of vectors. Image inspired by http://colah.github.io/posts/2015-08-Understanding-LSTMs/

One approach to tackle such problems is to use different activation functions, composed of gating units. A very successful attempt at this is the Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), which made it possible to capture distant dependencies between data. LSTMs have been successfully applied to NLP tasks (Dyer et al., 2015, Hermann et al., 2015, Wen et al., 2015, Yang et al., 2016, Ma and Hovy, 2016, Dozat and Manning, 2017).

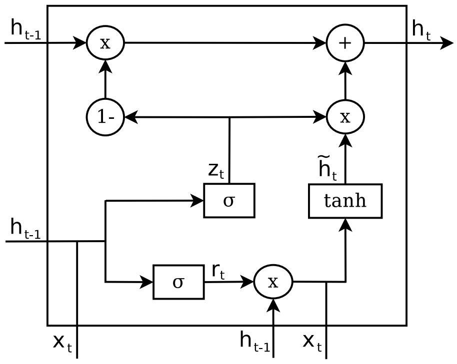


Figure 3.2: Illustration of the GRU architecture. Each line carries one vector, from the output of one node to the input of one of more nodes. Circles mean point-wise operations and boxes are neural network layers. A line merge represents a concatenation of vectors, while a line fork symbolizes a copy of vectors. Image inspired by http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Figure 3.1 shows the architecture of an LSTM. Each line carries one vector, from the output of one node to the input of one or more nodes. Circles mean point-wise operations and boxes are neural network layers. Line merge represents concatenation of vectors, whilst line fork symbolizes copy of vectors. An LSTM model has a cell state represented as C. It receives the cell state from the previous time, $t-1$, and performs two operations: deciding what information to keep from the previous time, and adding some information from the new candidate cell $\tilde{C}$. The LSTM model receives the output from the previous time, h_{t-1}, and the input from the current time, x_t. It concatenates those two pieces of information and provides them as an input vector to the forget gate, f_t, input gate, i_t, output gate, o_t, and also for the new candidate cell, $\tilde{C}_t$. The three gates are responsible for adding or removing information. A gate is composed by a sigmoid function, which sets the vector values between zero and one, and a point-wise multiplication. f_t is responsible for eliminating unnecessary information from C_{t-1}. i_t is responsible for selecting the new information that will be added to C_{t-1} to constitute C_t. o_t selects which information to output from the point-wise operation *tanh* over C_t, which sets the vector values between minus one and one, as the hidden output, h_t. The equations below describe each operation in the cell.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_{\tilde{C}} * [h_t - 1, x_t] + b_{\tilde{C}})$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$
$$h_t = tanh(C_t) * o_t$$

Similar to LSTMs, the Gated Recurrent Unit (GRU) (Cho et al., 2014) has a simpler structure and produces models with comparable quality (Chung et al., 2014). Figure 3.2 presents the architecture of a GRU. A GRU combines the forget gate and the input gate into a single gate, the update gate, z_t. It does not have the concept of cell state, only the hidden output, h. There is a new gate in GRU, the reset gate, r_t, which allows to completely forget the previous state. The equations below describe each operation in the unit.

$$r_t = \sigma(W_r * [h_{t-1}, x_t] + b_r)$$
$$z_t = \sigma(W_z * [h_{t-1}, x_t] + b_z)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$
$$\tilde{h}_t = tanh(W_{\tilde{h}} * [r_t * h_{t-1}, x_t] + b_{\tilde{h}})$$

As we see in Figures 3.1 and 3.2, the resulting model generated using GRUs is simpler compared to LSTMs. The main advantage of GRUs over LSTMs is that they require less computational time to generate similar quality models (Chung et al., 2014).

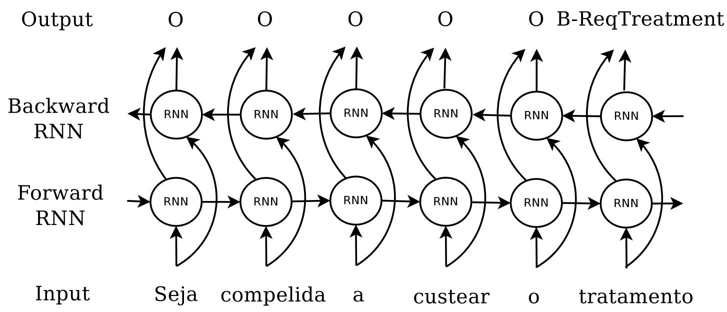


Figure 3.3: Example of a Bidirectional RNN.

In standard RNNs, predictions are made based on past sequence states. However, the relationship among words in a sentence is bidirectional. Bidirectional RNNs (Graves et al., 2013) have been proposed to handle that. They are a special kind of RNNs that use past and future states to predict the label of the current state. They combine two RNNs, one of which runs forward and another one that runs backward, so it can use past and future input features to predict the current tag. Figure 3.3 presents an example of a Bidirectional RNN.

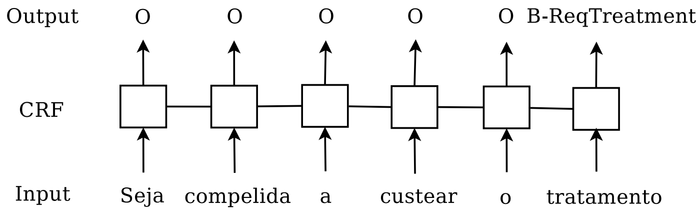


Figure 3.4: Example of CRF. Image inspired by (Huang et al., 2015).

3.2 Conditional Random Fields

Conditional Random Fields (CRF) is a framework proposed for building probabilistic models to segment and label sequence data (Lafferty et al., 2001). CRF has been successfully applied to different NLP tasks (Sha and Pereira, 2003, McCallum and Li, 2003, Cohn and Blunsom, 2005, Gimpel et al., 2011) and has proved to be more effective than Hidden Markov models (HMMs) and Maximum Entropy Markov models (MEMMs) (Lafferty et al., 2001). Figure 3.4 presents an example of a CRF. Note that CRF works at a sentence level, thus inputs and outputs are directly connected.

We provided CRF with a feature set to solve our tasks. The features are based on the word itself and also on the POS tag of the word. The proposed feature set is: the word itself in lowercase, the last four characters of the word in lowercase, the characters that precede the last four ones of the word in lowercase, the last three characters of the word in lowercase, the characters that precede the last three ones of the word in lowercase, the first three characters of the word in lowercase, a boolean that indicates whether the word is in uppercase, a boolean indicating whether the first character of the word is capitalized, a boolean that indicates whether the word is composed of only numbers, and the POS tag of the word. Those features are generated for each token and for a window of five tokens to the left and five tokens to the right of each token.

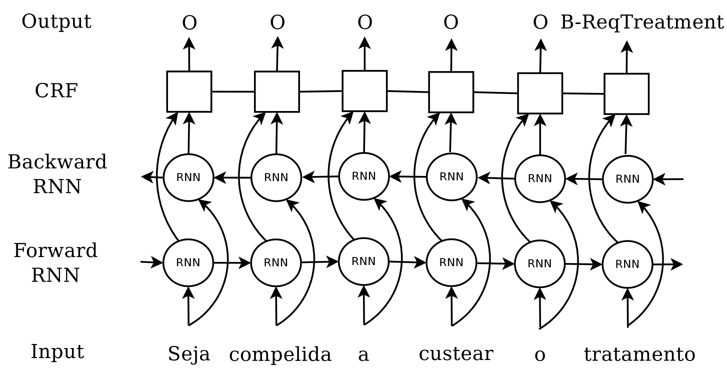


Figure 3.5: Example of a Bidirectional RNN combined with a CRF. Image inspired by (Huang et al., 2015).

3.3 Bidirectional Recurrent Neural Network with Conditional Random Fields

A Bidirectional Recurrent Neural Network (RNN) with Conditional Random Fields (CRF) is a combination of a bidirectional RNN layer with a CRF layer. With Bidirectional RNNs, it is possible to use past and future input features to predict the current tag. However, the tagging decision is performed locally, that is, the prediction of the current tag does not depend on past or future tags. Adding a CRF layer lets the model use past and future tags to predict the current tag. Therefore, the resulting network can efficiently use past and future input features via a bidirectional RNN layer and use sentence level tag information via a CRF layer (Huang et al., 2015). Figure 3.5 presents an example of a Bidirectional RNN combined with a CRF.

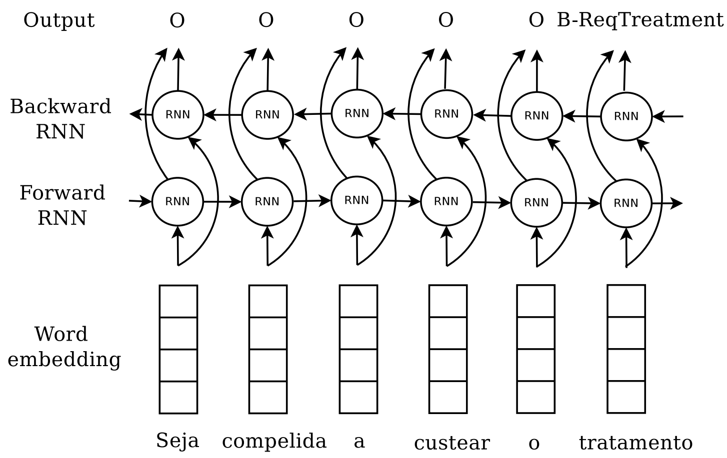


Figure 3.6: Example of a 4-dimensional word embedding used as input to a Bidirectional RNN.

3.4
Distributed Representation of Words

Distributed representations of words in a vector space, called word embeddings, help learning algorithms to achieve better performance in NLP tasks by grouping similar words. For our first system, we use Word2vec (Mikolov et al., 2013). For our second system, we use Word2vec (Mikolov et al., 2013), FastText (Bojanowski et al., 2016), and GloVe (Pennington et al., 2014). All of them are well-known methods to learn vector representations of words. Figure 3.6 presents an example of a 4-dimensional word embedding used as input to a Bidirectional RNN.

3.5
Char-level Representation of Words

For our second system, beyond the representation of words using the well-known methods mentioned above, which have a lookup table for the word vocabulary, we also use a model that generates a character embedding representation for the words. Such model, proposed by (Ling et al., 2015), benefits from being sensitive to lexical aspects within words.

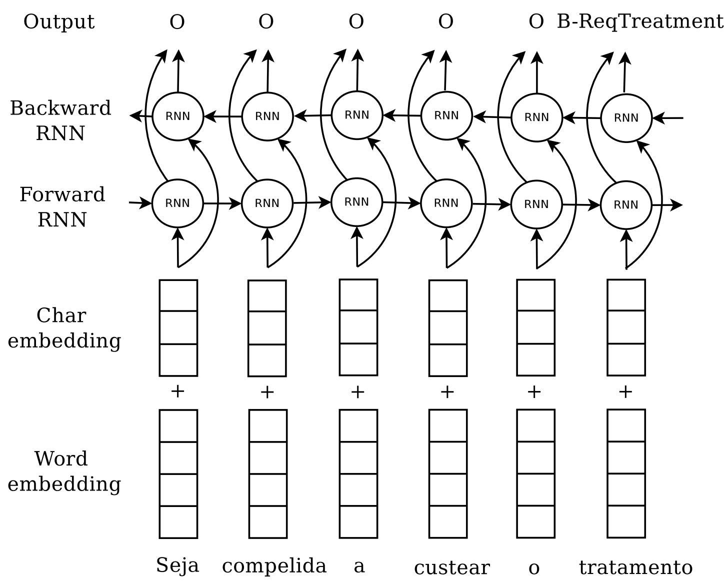


Figure 3.7: Example of a 4-dimensional word embedding concatenated with a 3-dimensional char embedding. The resulting vector is used as input to a Bidirectional RNN.

The input of the model is a single word w, and what is obtained is a d-dimensional vector used to represent w. As input, they define an alphabet of characters, containing an entry for each uppercase and lowercase letter, numbers and punctuation. The input word w is decomposed into a sequence of characters c_1, ..., c_m, where m is the length of w. Each c_i is represented as a one-hot vector, with one on the index c_i that corresponds to the entry at the alphabet of characters. Then the input vectors of the word w are passed to a Bidirectional LSTM that generates the representation of that word.

Figure 3.7 shows an example of a 4-dimensional word embedding concatenated with a 3-dimensional char embedding. The concatenated vector is used as input to a Bidirectional RNN.

In this chapter, we presented the models created to perform the identification of the entities that are going to be defined in chapter 4. We also introduced the word representations adopted by our systems. Next, we describe the extraction tasks for each corpus we have built. We also show the annotation guidelines that oriented our annotation process for each entity of our tagsets.

4
Tasks

In this chapter, we describe the extraction tasks for identifying (i) Appellate Court modifications; and (ii) plaintiff's legal claims and judicial provisions. We also show the annotation guidelines that oriented our annotation process for each entity of our tagsets. This chapter comprises the first step of our methodology, which is the annotation of Court decisions.

Brazilian Court decisions are well structured. They can be made by lower Court judges or Appellate Court (also known as upper Court) judges. A decision is composed of three parts as illustrated in Figure 4.1. The first part is called the *report*, which is a brief description made by the lower or Appellate Court judge on their decision, which includes: the plaintiff's claims, the defendant defense, and a summary of what happened during the proceedings until that point.

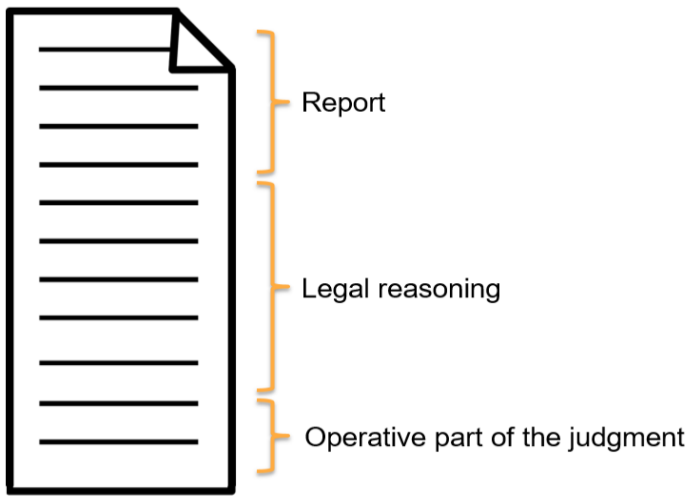


Figure 4.1: Illustration of the structure of Brazilian Court decisions.

The second part is named the *legal reasoning*. In that part, the lower or Appellate Court judge sets out in greater detail the facts involved on the law suit, indicating which events have motivated their legal understanding of the controversy. In other words, the legal reasoning is when the judge presents a logic argument analyzing the factual-legal reasons that justify their decision-making.

The last part is called the *operative part of the judgement*, where the lower or Appellate Court judge presents their judicial solution to the law suit. The operative part of the judgement, thus, consists of the final judicial solution to the case, and is the part of the judicial decision which precludes ("res judicata").

We identify those three parts through the application of a set of regular expressions in the lower and Appellate Court decisions.

We now detail the extraction tasks for each system.

4.1 First system

For our first system, we have built one corpus, KAUANE JUNIOR[1], composed of the operative part of the judgement of Appellate Court decisions in a narrow legal domain of civil suits regarding *negativação indevida*, which is a business practice that unlawfully downgrades one's credit score. We applied a set of regular expressions to select only the decisions granting or partially upholding the appeal to compose the corpus.

In KAUANE JUNIOR, we identified what the upper Court modified from the lower Court judge's decision.

Table 4.1 presents the tagset entities that compose the corpus. Such entities correspond to the most popular legal categories that were modified by the Appellate Court in claims of a narrow legal domain of civil suits regarding *negativação indevida*.

Table 4.1: Tagset entities that compose the Kauane Junior corpus.

Kauane Junior Entities
The value of the moral damage
The increase of the moral damage value
The decrease of the moral damage value
The initial date of interest of arrears
The initial date of monetary correction
The value of legal fee due by the defeated party

We now detail each entity for Kauane Junior corpus. The first entity is *the value of the moral damage*. Moral damage is a pecuniary compensation for the violation of the personality rights of an individual, such as one's right to honor, intimacy, privacy, image, and human dignity, which cause the victim

[1] *Kauane* is an indigenous word that refers to a type of hawk. The word choice indicates that the annotated data gives us a broad view of the Court decisions, such as when a hawk is flying and it has a broad view of what is going under his watch. *Junior* indicates that this corpus, as well as this system, is an initial step to have a more granular information extracted from the Court decisions.

sadness, humiliation, or shame (Gonçalves, 2018). For this entity, we decided to annotate both the numeric value and the value in full.

The second entity is *the increase of moral damage value*. The Appellate Court, when reviewing the lower Court judge's decision, can increase or decrease the moral damage value if it understands that the amount of such compensation is not coherent with the overall case law stance for similar cases. In the case of increase, we annotated just the term that indicates that the moral damage value established by the lower Court judge was increased by the Appellate Court.

The third entity is *the decrease of moral damage value*. We annotated only the term that indicates that the moral damage value established by the lower Court judge was decreased by the Appellate Court.

The fourth entity is *the initial term of interest of arrears*. The interest of arrears is an obligation foreseen in law (article 395 of (Brasil, 2008)), which is due whenever an illicit act is practiced (article 398 of (Brasil, 2008)) or when a pecuniary value is fixed by the Court (article 407 of (Brasil, 2008)); it is applied from the date of the illicit act or from the date of the Court's decision. For this entity, we decided to annotate the initial date of interest of arrears established by the Appellate Court.

The fifth entity is *the initial term of monetary correction*. The monetary correction, as the interest of arrears, is also a legal determination ((Brasil, 1981) and article 395 of (Brasil, 2008)), which is due to any judicial debt enacted by a Court. We chose to annotate the initial date of monetary correction set by the Appellate Court.

Finally, the sixth entity is *the value of legal fee due by the defeated party*. This is also an obligation foreseen by the law (article 395 of (Brasil, 2008) and article 85 of (Brasil, 2015)), in which the defeated party must pay a legal fee to the winning party lawyer. For this entity, we decided to annotate the fixed value of the legal fee due by the defeated party.

We now show an example of an Appellate Court decision in which all entities were annotated. The annotated entities are in boldface and the subscript after them identifies the entity. *DecMorDm* is the decrease of the moral damage value; *ValMorDm* is the value of the moral damage; *InitIntArr* is the initial date of interest of arrears; *InitMonCo* is the initial date of monetary correction; and *ValLegFee* is the value of legal fee due by the defeated party.

> Do exposto, voto para DAR PARCIAL PROVIMENTO ao recurso, para **reduzir**$_{DecMorDm}$ o valor arbitrado a título de indenização por danos morais ao patamar de **R$ 5.000,00 (cinco mil reais)**$_{ValMorDm}$, acrescido de **ju-**

> **ros de mora de 1% ao mês a contar do evento danoso**$_{InitIntArr}$ e **correção monetária a partir da presente**$_{InitMonCo}$, honorários advocatícios, os quais fixo em **R$ 1.000,00 (um mil reais)**$_{ValLegFee}$, mantendo no mais, a sentença tal como lançada.
>
> (In light of the above, I vote for the partial upholding of the appeal, to **reduce**$_{DecMorDm}$ the arbitrated value of the moral damages to **R$ 5,000.00 (five thousand reais)**$_{ValMorDm}$, increased with **interest of arrears of 1% per month to be applied since the harmful event**$_{InitIntArr}$ and **monetary correction beginning from the present date**$_{InitMonCo}$, legal fee due by the defeated party, fixed at **R$ 1,000 (one thousand reais)**$_{ValLegFee}$, maintaining, on the most, the lower court judge decision.)

4.2 Second system

For our second system, we have built three corpora based on parts of lower and Appellate Court decisions in the narrow legal domain of civil suits involving the refusal of insurance companies on providing assistance to their customers. The first corpus, Kauane Insurance Report, is composed of the report of lower Court decisions. The second corpus, Kauane Insurance Lower, is composed of the operative part of the judgement of lower Court decisions. The last corpus, Kauane Insurance Upper, is composed of the operative part of the judgement of Appellate Court decisions.

In order to compose the latter corpus, we applied a set of regular expressions to select only the decisions granting or partially upholding the appeal.

In Kauane Insurance Report, we identified the plaintiff's claims which motivated the filing of the law suit. In Kauane Insurance Lower, we identified which of the plaintiff's claims were granted, dismissed or upholded by the lower Court judge. In Kauane Insurance Upper, we identified what the Appellate Court has modified or maintained from the lower Court judge's decision.

Table 4.2 presents the entities of the tagsets that compose the three corpora. The entities in the same line have direct correspondence among the corpora.

We now detail the tagset entities for each corpus.

Table 4.2: Entities of the tagsets that compose Kauane Insurance Report corpus, Kauane Insurance Lower corpus, and Kauane Insurance Upper corpus. The entities in the same line have direct correspondence among the corpora.

Kauane Insurance Entities		
Report	*Lower*	*Upper*
The health cause that motivated the judicial procedure		
The claimed value of the moral damage	The value of the moral damage	The value of the moral damage
The claimed value of the material damage	The value of the material damage	The value of the material damage
The claimed value of legal fee due by the defeated party	The value of legal fee due by the defeated party	The value of legal fee due by the defeated party
Claims material damage	Material damage	Material damage
Claims treatment	Treatment	
Claims procedure	Procedure	
Claims medicine/exam	Medicine/exam	
		Moral damage has been excluded

4.2.1 Kauane Insurance Report

We have built a tagset with eight entities, which constitute the most common legal claims presented by plaintiffs on law suits involving insurance companies. Such law suits are commonly motivated by the denial of the company in providing a specific good or service which is allegedly due according to a legal obligation previously assumed by both parties on the health care plan agreement. Insurance companies can be held liable for acting in such a manner, due to allegedly breach of contract and for causing harm to their clients, causing plaintiffs to seek for compensation.

The first entity is *the health cause that motivated the judicial procedure*, referring to the medical condition or disease that affected the plaintiff, for which the insurance company denied medical coverage, motivating the legal claim. For this entity, we decided to annotate the name of the medical condition or disease, and the health problems developed by such disease.

The second entity is *the claimed value of the moral damage*. Moral damage is a violation caused to one of the personality rights of an individual (such as honour, image, privacy or intimacy) or human dignity. When such harm causes damage to the victim, causing her sadness, humiliation, or shame (Gonçalves, 2018), one can be financially compensated for such inconvenience (article 186 of (Brasil, 2008)). For this entity, we decided to annotate both the numeric value and the value in full of the moral damage that has been expressly claimed by the plaintiff.

The third entity is *the claimed value of the material damage*, and corresponds to the financial expenditure that the plaintiff had to unduly disburse to have access to a specific good or service. Such value should have been afforded by the defendant due to a legal obligation previously assumed by both parties on the health care plan agreement. Hence, the plaintiff can claim for the reimbursement of such value. For this entity, we decided to annotate both the numeric value and the value in full of the material damage that has been expressly claimed by the plaintiff.

The fourth entity is *the claimed value of legal fee due by the defeated party*. The Brazilian Civil Code of 2002 establishes that the defeated party of any law suit shall pay a legal fee to the counsel of the winning party (article 395 of (Brasil, 2008) and article 85 of (Brasil, 2015)), as a compensation for their work during the proceeding. For this entity, we decided to annotate the fixed value of the legal fee due by the defeated party that has been expressly claimed by the plaintiff.

The fifth entity is *claims material damage*, which has already been described above. For this entity, we decided to annotate only the word "reimbursement" or the most similar word related to such term.

The sixth entity is *claims treatment*, relating to a specific service requested by the plaintiff, allegedly foreseen on the plaintiff's health care plan agreement, and denied by the defendant. Thus, *claims treatment* is a legal claim presented by the plaintiff demanding the insurance company to cover all costs related to a hospitalization, which includes regular hospitalization, hospitalization on Specialized Unities, or home care. For this entity, we decided to annotate the word "treatment", or the most similar word related to such term, referring to what was expressly claimed by the plaintiff.

The seventh entity is *claims procedure*, relating to a specific service requested by the plaintiff, allegedly foreseen on the plaintiff's health care plan agreement, and denied by the defendant. The entity *claims procedure* is a legal claim presented by the plaintiff to cover any medical procedure, involving any kind of hospitalization related to surgery needs and its inherent costs, such as: medical exams, remedies, medical fees, anesthesia. For this entity, we decided to annotate the word "procedure", or the most similar word related to such term, referring to what was expressly claimed by the plaintiff.

The eighth entity is *claims medicine/exam*, relating to the plaintiff's legal claim to the insurance company to cover any medicines or exams which were previously denied, an obligation that was due since it was allegedly foreseen on the plaintiff's health care plan agreement. For this entity, we decided to annotate the word "medicine" or "exam" expressly claimed by the plaintiff.

We now show an example of a lower Court report in which all entities were annotated. The annotated entities are in boldface and the subscript after them identifies the entity. *HltCausMot* is the health cause that motivated the judicial procedure; *ClmTreatment* is claims treatment; *ClmValMatDm* is the claimed value of the material damage; *ClmMatDm* is claims material damage; and *ClmValMorDm* is the claimed value of the moral damage.

> Trata-se de uma ação que, pelo procedimento ordinário, NURIA MANSUR move em face de BRADESCO SAÚDE, ambos já devidamente qualificados, objetivando, em síntese, autorização para realização de tratamento e indenização por danos morais. Narra a inicial que a parte autora, associada ao plano de saúde da ré através da empresa Mansur Advogados Associados, foi diagnosticada com um **tumor no úmero**$_{HltCausMot}$ esquerdo, submetendo-se a uma cirurgia em 17 de julho de 2015. Conta que a cirurgia foi autorizada pela ré que também reembolsou a autora de parte das despesas incorridas, contudo, a ré não aprovou solicitação para que a mesma se submetesse a dez sessões de radioterapia prescritas pela médica Dr. Celia Viégas, tendo negado, ainda, autorização para realização do exame PET-CT SCAN, que acabou sendo custeado pela autora no valor de R$ 3.600,00 (três mil e seiscentos reais). Pugna pela concessão de tutela antecipada para que a ré seja compelida a custear o **tratamento**$_{ClmTreatment}$ de radioterapia, bem como seja a ré condenada ao pagamento da quantia de **R$ 3.600,00**$_{ClmValMatDm}$ a título de **reembolso**$_{ClmMatDm}$ e ao pagamento de uma indenização a título de dano moral em quantia não inferior a **R$ 5.000,00**$_{ClmValMorDm}$.
>
> (This law suit runs through an ordinary procedure filed by NURIA MANSUR before BRADESCO SAÚDE, both duly qualified, claiming, in summary, for the authorization of the treatment and moral damages. The plaintiff, associated to the health care plan of the defendant by means of Mansur Advogados Associados, was diagnosed with a **tumor on her left humerus**$_{HltCausMot}$, going through surgery on July 17th, 2015. The surgery was authorized by the defendant, which partially reimbursed the plaintiff, however, the defendant did not approved the ten sessions of radiotherapy prescribed by the plaintiff's doctor, Dr. Celia Viégas, and also denied, the authorization for the PET - CT SCAN exam, which

was, anyhow, done by the plaintiff for the value of R$3.600,00 (three thousand and six hundred reais). The plaintiff claims for the prior of relief to demand the defendant to pay for the radiotherapy **treatment**$_{ClmTreatment}$, as well as for the payment of **R$3.600,00**$_{ClmValMatDm}$ for the **reimbursement**$_{ClmMatDm}$ and moral damage not lower than **R$5.000,00**$_{ClmValMorDm}$.)

4.2.2 Kauane Insurance Lower

Those entities constitute legal categories that corresponds to the lower Court decision when analyzing the plaintiff's legal claim, and the defendant's defense. The analyzed decision is given on the context of law suits motivated by the denial of the insurance company in providing a specific good or service which is allegedly due according to a legal obligation previously assumed by both parties on the health care plan agreement. Insurance companies can be held liable for acting in such a manner, due to allegedly breach of contract and for causing harm to their clients, causing plaintiffs to seek for compensation.

The first entity is *the value of the moral damage*, which corresponds to the lower Court decision of *the claimed value of the moral damage* entity described above. For this entity, we decided to annotate both the numeric value and the value in full of the moral damage that has been granted by the lower Court.

The second entity is *the value of the material damage*, which corresponds to the lower Court decision of *the claimed value of the material damage* entity described above. For this entity, we decided to annotate both the numeric value and the value in full of the material damage that has been granted by the lower Court.

The third entity is the *the value of legal fee due by the defeated party*, which corresponds to the lower Court decision of *the claimed value of legal fee due by the defeated party* entity described above. For this entity, we decided to annotate the fixed value of the legal fee due by the defeated party that has been granted by the lower Court.

The fourth entity is *material damage*, which corresponds to the lower Court decision of *claims material damage* entity described above. For this entity, we decided to annotate only the word "reimbursement" or the most similar word related to such term granted by the lower Court.

The fifth entity is *treatment*, which corresponds to the lower Court decision of *claims treatment* entity described above. For this entity, we decided to annotate the word "treatment", or the most similar word related to such term, referring to what was granted by the lower court.

The sixth entity is *procedure*, which corresponds to the lower Court decision of *claims procedure* entity described above. For this entity, we decided to annotate the word “procedure”, or the most similar word related to such term, referring to what was granted by the lower Court.

The seventh entity is *medicine/exam*, which corresponds to the lower Court decision of *claims medicine/exam* entity described above. For this entity, we decided to annotate the word “medicine” or “exam” expressly granted by the lower Court.

We now show an example of a lower Court operative part of the judgment in which all entities were annotated. The annotated entities are in boldface and the subscript after them identifies the entity. *Treatment* is treatment; *ValMatDm* is the value of the material damage; *MatDm* is material damage; *ValMorDm* is the value of the moral damage; and *ValLegFee* is the value of legal fee due by the defeated party.

> Ante o exposto, JULGO PROCEDENTE o pedido inicial e, por consequência, RATIFICO a tutela antecipada antes concedida e, em caráter definitivo, CONDENO a ré a custear o **tratamento**$_{Treatment}$ de radioterapia ao qual a autora foi submetida. CONDENO, ainda, a ré, ao pagamento da quantia de **R$3.600,00**$_{ValMatDm}$ a título de **reembolso**$_{MatDm}$ das despesas com o exame Pet Scan, quantia essa que deverá ser atualizada desde a data do desembolso até o efetivo pagamento e acrescida de juros de 1% ao mês, contados da citação, e ao pagamento da quantia de **R$8.000,00 (oito mil reais)**$_{ValMorDm}$ a título de dano moral, quantia essa que deverá ser atualizada desde a data desta sentença até o efetivo pagamento e de juros de 1% ao mês, contados da citação. Por fim, CONDENO a ré ao pagamento das custas processuais e honorários advocatícios, estes fixados em **15% sobre o valor da condenação**$_{ValLegFee}$. P.R.I.
>
> (In light of the above, I GRANT the plaintiff’s claim and, therefore, confirm the prior relief, to sentence the defendant to pay for the plaintiff’s radiotherapy **treatment**$_{Treatment}$. Also, the defendant shall pay the plaintiff **R$3,600.00**$_{ValMatDm}$ to **reimburse**$_{MatDm}$ expenses related to the PetScan exam, monetary adjustment of such value from the disbursement date until the payment date, with interest of arrears of 1% per month, to be applied since the subpoena date, and moral damage on the value of **R$8,000.00 (eight thousand reais)**$_{ValMorDm}$, which should be monetary adjusted from this decision, with interest of arrears of 1% per month,

to be applied since the subpoena date. At last, the defendant shall be responsible for the defeated party's expenses and legal fee arbitrated on **15% of the matter of controversy**$_{ValLegFee}$. P.R.I.)

4.2.3 Kauane Insurance Upper

We have built a tagset with five entities. Those entities constitutes the most common legal categories that the Appellate Court modifies or maintains in regards to the lower court decision. Such decisions are given in the context of law suits involving the denial of insurance companies in providing a specific good or service which is allegedly due according to a legal obligation previously assumed by both parties on the health care plan agreement.

The first entity is *the value of the moral damage*, which corresponds to the Appellate Court decision of *the claimed value of the moral damage* entity described above. For this entity, we decided to annotate both the numeric value and the value in full of the moral damage that has been granted by the Appellate Court.

The second entity is *the value of the material damage*, which corresponds to the Appellate Court decision of *the claimed value of the material damage* entity described above. For this entity, we decided to annotate both the numeric value and the value in full of the material damage that has been granted by the Appellate Court.

The third entity is the *the value of legal fee due by the defeated party*, which corresponds to the Appellate Court decision of *the claimed value of legal fee due by the defeated party* entity described above. For this entity, we decided to annotate the fixed value of the legal fee due by the defeated party that has been granted by the Appellate Court.

The fourth entity is *material damage*, which corresponds to the Appellate Court decision of *claims material damage* entity described above. For this entity, we decided to annotate only the word "reimbursement" or the most similar word related to such term granted by the Appellate Court.

The fifth entity is *moral damage has been excluded*. The Appellate Court can review the lower Court judge's decision and reverse the moral damage value if it understands that the fixed value does not fit with similar cases of the overall case law stance. For this entity, we decided to annotate only the word "exclude" or the most similar word related to such term used by the Appellate Court.

We now show an example of an Appellate Court operative part of the judgment in which all entities were annotated. The annotated entities are in boldface and the subscript after them identifies the entity. *MatDm* is material damage; *ValMatDm* is the value of the material damage; and *ValMorDm* is the value of the moral damage.

> Por tais razões e fundamentos, DÁ-SE PARCIAL PROVIMENTO À APELAÇÃO DA RÉ para que a ré seja reduzido o quantum da indenização pelo dano **material**$_{MatDm}$ para o valor de **R$9.164,40 (nove mil e cento e sessenta e quatro reais e quarenta centavos)**$_{ValMatDm}$ quantia esta que deverá ser acrescida de juros de mora de 1% (um por cento) ao mês, a contar da citação, e correção monetária a contar da data do desembolso. DÁ-SE PROVIMENTO AO RECURSO DA AUTORA, para condenar a parte ré a pagar indenização no valor de **R$ 10.000,00 (dez mil reais)**$_{ValMorDm}$, a título de danos morais, quantia esta que deverá ser acrescida de juros de mora de 1% (um por cento) ao mês, a contar da citação, e correção monetária a partir deste julgado. No mais, mantém-se a sentença nos termos em que foi lançada. Rio de Janeiro, na data da assinatura digital. Desembargador
>
> (In light of the above, I PARTIALLY UPHOLD THE DEFENDANT APPEAL to reduce the **material**$_{MatDm}$ damage to the value of **R$9,164.40 (nine thousand and one hundred and sixty four reais and forty cents)**$_{ValMatDm}$, which should be increased with interest of arrears of 1% (one percent) per month, since the subpoena date, and monetary adjustment from the disbursement date. I GRANT THE PLAINTIFF'S APPEAL, to award moral damages on the value of **R$ 10,000.00 (ten thousand reais)**$_{ValMorDm}$, increased with interest of arrears of 1% (one percent) per month, from the subpoena date, and monetary adjustment from this decision. On the most, I maintain the lower court judge decision. Rio de Janeiro, signature of digital date. Appelate Court Judge.)

In this chapter, we defined the extraction tasks and entities extracted for each one of them. Moreover, we presented the guidelines that drove the annotation process. Next, for each system of ours, we show the composition of our datasets, the experimental setup, and the quality of the proposed models. In addition, for our second system, we perform a visual analysis of our datasets.

5
Experiments

In this chapter, for each of our systems, we present the composition of our datasets, the experimental setup, and the observed quality of our models. Those models comprise the second step of our methodology, which is the creation of Deep Learning models to extract information. For our second system, we also perform a visual analysis in our datasets using the charts that constitute the third step of our methodology, which is the visualization of the aggregated information extracted from the decisions.

5.1
First system

For our first system, we have built one dataset, KAUANE JUNIOR, and we have built models upon five algorithms to solve the proposed extraction task: Long Short-Term Memory network (Hochreiter and Schmidhuber, 1997, Graves and Schmidhuber, 2005); Bidirectional Gated Recurrent Units network (Graves et al., 2013); Conditional Random Fields (Lafferty et al., 2001); a combination of Bidirectional Long Short-Term Memory network and Conditional Random Fields (Huang et al., 2015); and a combination of Bidirectional Gated Recurrent Units network and Conditional Random Fields (Huang et al., 2015). Below we present the composition of KAUANE JUNIOR, the experimental setup, and the observed quality of our models.

5.1.1
Dataset

To build our dataset, we used public data disclosed by the Appellate State Court of Rio de Janeiro's jurisprudence database[1]. The theme searched was *negativação indevida*, a business practice that could be considered as an unlawful downgrading of one's credit score. Our dataset only included decisions of the Appellate State Court from 2016, excluding interlocutory appeals, decisions that were not taken by the collegiate body, or that analyzed motions for clarifying the judgment.

[1]We have built a crawler to collect data from the State Court of Rio de Janeiro website at http://www.tjrj.jus.br

With the use of the ERAS annotation system (Grosman et al., 2020), we conducted the annotation process. We had two groups of people for annotating documents. The first one was composed of 6 undergraduate Law students, a Master Law student, and a PhD Law student, a total of 8 annotators, all from the Law area. The second group was composed of 6 undergraduate Law students, 5 undergraduate Computer Engineering students, 3 undergraduate Design students, an undergraduate Economics student, an undergraduate Information Systems student, an undergraduate Production Engineering student, a Law professional, and a PhD Law student, a total of 19 annotators.

We explained the task separately to each group, taught them how to use the annotation system, provided them with annotation examples and presented a document with the annotation guidelines. We also provided each group with a small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. We used the kappa metric (Carletta, 1996) to evaluate the quality of the annotations. The first group presented a mean agreement of 0.91 with standard deviation of 0.058, while the second one presented a mean agreement of 0.84 with standard deviation of 0.088. The difference of agreement may be due to the fact the first group is specialized in Law, while the second one is mostly composed of people of different areas.

After certifying that the annotation agreement was acceptable, we concluded the annotators were ready to the annotation task itself. Then, we provided each user with a set of documents to annotate. The first group annotated 800 documents and the second group annotated 2,222 documents.

By the end of the annotation process, we joined the annotated sets of each user and composed our dataset, which comprises 3,022 documents and 221,820 tokens, an average of 73 tokens per document. This final dataset contains gold standard annotations considered as the groundtruth for training and test.

5.1.1.1 Adopted Annotation

The corpus information is codified on a per token basis. In Table 5.1, we exemplify the annotation of a sentence with its three basic features. The first one is the word. Next, we have the part-of-speech (POS) annotation provided by an in-house tagger.[2] We use the IOB format (Ramshaw and Marcus, 1995) to annotate the entities established in our tagset. The IOB format is a common tagging format for named entity recognition, where the *B-* prefix indicates that the tag is the beginning of a chunk, the *I-* prefix indicates that the tag is inside

[2]We used a BILSTM based tagger trained for Portuguese.

a chunk, and the O tag indicates that the token belongs to no chunk. In the example of Table 5.1, we annotated two entities: the decrease of the moral damage value (DecMorDm), and the value of the moral damage (ValMorDm).

Table 5.1: Annotated corpus example. *DecMorDm* is the decrease of the moral damage value, and *ValMorDm* is the value of the moral damage, both annotated using the IOB format. Each column corresponds to a token.

Word	reduzir	o	valor	de	indenização	a	r$	5.000,00
POS	V	ART	N	PREP	N	PREP	CUR	NUM
Entity	B-DecMorDm	O	O	O	O	O	B-ValMorDm	I-ValMorDm

5.1.2 Word Embedding

In Table 5.2, we show different configurations tested for word2vec. First we trained each word2vec with the CBOW algorithm (Mikolov et al., 2013) using a dataset of 92,122 appeal decisions. The CBOW algorithm produces a distributed representation of words, where the model is trained to predict the current word from a window of surrounding words. Then, with each trained word2vec, we trained a Gated Recurrent Units network with 32 hidden layers for 20 epochs using approximately 80% of the training set of our task. Next, we selected the word2vec with the best performance on the remaining 20% of the training set. We also present the performance of each model in Table 5.2. We notice that the combination of a large embedding size with a large window produces noisy embeddings. Moreover we verify that generally a large window size produces lower quality models. Finally, we see that a small embedding size produces more robust models. The lines in boldface show the models which presented the best quality. Based on the Occam's razor principle, we selected the simplest word2vec configuration, that is, the one that has vector dimension of 100, window of 5, and minimum word count of 2.

Table 5.2: word2vec configurations tested for our task. *VD* is Vector Dimension and *MWC* is Minimum Word Count.

VD	*Window*	*MWC*	*Precision (%)*	*Recall (%)*	$F_{\beta=1}$*(%)*
100	**5**	**2**	**55.42**	**69.55**	**61.69**
100	5	4	52.66	68.47	59.53
100	10	2	49.92	66.52	57.04
100	10	4	50.24	66.95	57.41
300	**5**	**2**	**55.42**	**69.55**	**61.69**
300	5	4	50.33	66.31	57.22
300	10	2	22.54	42.12	29.37
300	10	4	24.35	44.28	31.42

5.1.3 Experimental Setup

In order to assess the proposed models, we divided the annotated corpus into training set and test set. Table 5.3 shows the annotated corpus sizes. Table 5.4 presents the number of annotated entities by set, and Table 5.5 shows number of documents in which each entity appeared, per set.

Table 5.3: Annotated corpus set sizes.

Part	*#Documents*	*#Tokens*
Training	2,568	187,857
Test	454	33,963

In our work, we create five models to tackle the Appellate Court Modifications Extraction task: (i) a Bidirectional Long Short-Term Memory network (BILSTM); (ii) a Bidirectional Gated Recurrent Units network (BIGRU); (iii) a Conditional Random Fields (CRF); (iv) a combination of BILSTM and CRF (BILSTM-CRF); and (v) a combination of BIGRU and CRF (BIGRU-CRF).

Table 5.4: Annotated entities by set.

Entity	*Training*	*Test*
Increase of Moral Damage	267	37
Initial Term of Monetary Correction	380	63
Initial Term of Interest of Arrears	503	83
Decrease of Moral Damage	269	46
Value of Legal Fee	347	60
Value of Moral Damage	834	147

Table 5.5: Number of documents in which each entity appeared by set.

Entity	*Training*	*Test*
Increase of Moral Damage	255	37
Initial Term of Monetary Correction	359	60
Initial Term of Interest of Arrears	476	80
Decrease of Moral Damage	267	46
Value of Legal Fee	333	57
Value of Moral Damage	818	141

We used Keras 2.2.4, Keras-contrib 2.0.8, and Tensorflow-gpu 1.15.0 to run the deep neural network experiments.

In order to calibrate our models, we used a 5-fold cross-validation over the training set. For the first two models, we tested several neural network architectures preserving the first and last layer of our neural network – a pre-trained word2vec layer with 100 units and a Softmax layer with 13 units (six tags with the *B*- prefix, six tags with the *I*- prefix, and the *O* tag), respectively.

Table 5.6 presents the grid search step we have performed. We created a dummy element called *None* to represent the absence of an element in that level. We used RMSProp[3] algorithm as optimizer, with learning rate of 0.001, ρ of 0.9, and batch size of 32. We established 50 epochs for each cross-validation run. We executed grid search twice, where BIRNN is replaced by BILSTM at the first time and BIRNN is replaced by BIGRU at the second time.

Table 5.6: Combination of neural network levels in order to find the best architecture.

Level	*Layers*
1	None
	BIRNN with 64 hidden units
	BIRNN with 128 hidden units
	BIRNN with 256 hidden units
	BIRNN with 512 hidden units
2	BIRNN with 64 hidden units
	BIRNN with 128 hidden units
	BIRNN with 256 hidden units
	BIRNN with 512 hidden units

For the CRF model, we performed a grid search to find the best C1 and C2 hyperparameters. The candidate values for each of them were 0.001, 0.01 and 0.1.

Now we show the configuration that presented the best quality for each model. For models (i), (ii), (iv) and (v), we chose the number of epochs based on the mean of the lowest losses in the validation set of the cross-validation run of the best architecture.

The Bidirectional LSTM (BILSTM) architecture that presented the best quality starts with a pre-trained word2vec layer with 100 units, followed by a Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Next, there is another Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 13 units with Softmax activation. The number of epochs for training is 2.

The Bidirectional GRU (BIGRU) architecture that presented the best quality starts with a pre-trained word2vec layer with 100 units, followed by a

[3]We performed a previous step testing stochastic gradient descent and RMSProp optimizers. In that step, RMSProp performed better. Then we chose it to run all the experiments.

Table 5.7: Performance of the Kauane Junior Extraction task.

Model	*Precision (%)*	*Recall (%)*	$F_{\beta=1}$*(%)*
BILSTM-CRF	95.71	93.89	94.79
BIGRU-CRF	92.69	90.74	91.70
BILSTM	86.34	91.79	88.98
BIGRU	84.74	91.16	87.83
CRF	89.34	82.95	86.03

Bidirectional GRU with 128 units – a forward GRU layer with 64 units, and a backward GRU layer with 64 units. Next, there is another Bidirectional GRU with 64 units – a forward GRU layer with 32 units, and a backward GRU layer with 32 units. Finally, a dense layer with 13 units with Softmax activation. The number of epochs for training is 2.

The best combination of hyperparameters for the CRF model was C1 = 0.1 and C2 = 0.1.

In order to choose the fourth and fifth models, by computing time restrictions, we could not perform a grid search through all architectures of Table 5.6. Thus we selected the top 6 architectures with higher quality found in the grid search performed to choose the first and second models. For each candidate architecture, we replaced the last layer by a dense layer of 13 units with ReLU activation and connected that layer to a CRF layer. We used RMSProp algorithm as optimizer, with learning rate of 0.001, ρ of 0.9, and batch size of 32.

The Bidirectional LSTM combined with a CRF layer (BILSTM-CRF) that presented the best quality starts with a pre-trained word2vec layer with 100 units, followed by a Bidirectional LSTM with 256 units – a forward LSTM layer with 128 units, and a backward LSTM layer with 128 units. Next, there is another Bidirectional LSTM with 256 units – a forward LSTM layer with 128 units, and a backward LSTM layer with 128 units. Finally, a dense layer with 13 units with ReLU activation followed by a CRF layer. The number of epochs for training is 3.

The Bidirectional GRU combined with a CRF layer (BIGRU-CRF) that presented the best quality starts with a pre-trained word2vec layer with 100 units, followed by a Bidirectional GRU with 256 units – a forward GRU layer with 128 units, and a backward GRU layer with 128 units. Next, there is another Bidirectional GRU with 64 units – a forward GRU layer with 32 units, and a backward GRU layer with 32 units. Finally, a dense layer with 13 units with ReLU activation followed by a CRF layer. The number of epochs for training is 2.

Table 5.8: Detailed performance of the Kauane Junior Extraction task for BILSTM-CRF, BIGRU-CRF, BILSTM, BIGRU, and CRF. *IncMorDm* is the increase of the moral damage value; *InitMonCo* is the initial date of monetary correction; *InitIntArr* is the initial date of interest of arrears; *DecMorDm* is the decrease of the moral damage value; *ValLegFee* is the value of legal fee due by defeated party; and *ValMorDm* is the value of the moral damage.

Entity	BILSTM-CRF			BIGRU-CRF			BILSTM			BIGRU			CRF		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
IncMorDm	97.37	92.50	94.87	90.48	95.00	92.68	94.74	90.00	92.31	85.71	90.00	87.80	94.74	90.00	92.31
InitMonCo	97.06	91.67	94.29	86.57	80.56	83.45	89.19	91.67	90.41	82.67	86.11	84.35	90.32	77.78	83.58
InitIntArr	95.51	91.40	93.41	92.13	88.17	90.11	72.97	87.10	79.41	76.19	86.02	80.81	85.71	77.42	81.36
DecMorDm	96.36	96.36	96.36	98.18	98.18	98.18	96.30	94.55	95.41	94.64	96.36	95.50	93.88	83.64	88.46
ValLegFee	90.48	91.94	91.20	88.14	83.87	85.95	79.71	88.71	83.97	87.69	91.94	89.76	73.33	70.97	72.13
ValMorDm	96.73	96.73	96.73	96.08	96.08	96.08	91.82	95.42	93.59	86.31	94.77	90.34	94.59	91.50	93.02
General	95.71	93.89	94.79	92.69	90.74	91.70	86.34	91.79	88.98	84.74	91.16	87.83	89.34	82.95	86.03

5.1.4 Experiment Results

Table 5.7 presents the quality of our models assessed in the test set. BILSTM-CRF obtained an $F_{\beta=1}$ score of 94.79%, which is an increase in quality of 3.09% compared to the second best model, BIGRU-CRF. Compared to BILSTM, BIGRU and CRF, BILSTM-CRF increases the quality by 5.81%, 6.96%, and 8.76%, respectively. We employed the *conlleval* tool to evaluate the performance of the proposed models. The performance of our models cannot be directly compared to previous work, since their purposes, corpora, and tagsets are different. Table 5.8 presents the detailed quality of our models for each entity type assessed in the test set.

5.2 Second system

For our second system, we have built three datasets – Kauane Insurance Report, Kauane Insurance Lower, and Kauane Insurance Upper. We have built models upon three algorithms to solve the proposed extraction tasks: Long Short-Term Memory network (Hochreiter and Schmidhuber, 1997, Graves and Schmidhuber, 2005); Conditional Random Fields (Lafferty et al., 2001); and a combination of Bidirectional Long Short-Term Memory network and Conditional Random Fields (Huang et al., 2015). Below we present the composition of our datasets, the experimental setup, and the observed quality of our models. We also perform a visual analysis in our datasets.

5.2.1 Datasets

To build our dataset, we used public data disclosed by the Appellate State Court of Rio de Janeiro's jurisprudence database[4]. Our datasets included State Court decisions from the lower and Appellate Court from 2016 and 2017, excluding interlocutory appeals – decisions that were not taken by the collegiate body – or that analyzed motions for clarifying the judgment.

With the use of the ERAS annotation system (Grosman et al., 2020), we conducted the annotation process for each dataset. In the next sections, we present the annotation setup for each of them.

[4]We have built a crawler to collect data from the State Court of Rio de Janeiro website at http://www.tjrj.jus.br

5.2.1.1
Kauane Insurance Report

We had 14 annotators to build the dataset – an undergraduate Computer Science (CS) student, four Master CS students, a PhD Law student, seven PhD CS students, and a PhD CS researcher.

We explained the task to the annotators, taught how to use the annotation system, provided them with annotation examples and presented a document with the annotation guidelines. Moreover, we provided them with a small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. We used the kappa metric (Carletta, 1996) to evaluate the quality of the annotations. Until the annotation agreement reached an acceptable value, we repeated the steps above, expanding the document with the annotation guidelines. By the end of the iterations, the annotators presented a mean agreement of 0.88 with standard deviation of 0.046. At that point, we provided each user with a set of documents to annotate.

By the end of the annotation process, we joined the annotated sets of each user and composed KAUANE INSURANCE REPORT corpus, which comprises 3,180 documents and 646,840 tokens, an average of 203.41 tokens per document. This final corpus contains gold standard annotations considered as the groundtruth for training and test.

5.2.1.2
Kauane Insurance Lower

We had nine annotators to build the dataset – eight Law undergraduate students and a PhD Law student.

We explained the task to the annotators, taught them how to use the annotation system, provided them with annotation examples, and presented a document with the annotation guidelines. Moreover, we provided a small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. We used the kappa metric (Carletta, 1996) to evaluate the quality of the annotations. By the end of that step, the annotators presented a mean agreement of 0.87 with standard deviation of 0.079. Then, we provided each user with a set of documents to annotate. When each user finished the annotation, we provided them with another small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. The annotators then presented of mean agreement of 0.92 with standard deviation of 0.045. Finally we provided each user with another set of documents to annotate.

By the end of the annotation process, we joined the annotated sets of each user and composed KAUANE INSURANCE LOWER corpus, which comprises 3,317 documents and 590,569 tokens, an average of 178.04 tokens per document. This final corpus contains gold standard annotations considered as the groundtruth for training and test.

5.2.1.3 Kauane Insurance Upper

We had five annotators to build the dataset – all Law undergraduate students.

We explained the task to the annotators, taught how to use the annotation system, provided them with annotation examples, and presented a document with the annotation guidelines. Moreover, we provided a small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. We used the kappa metric (Carletta, 1996) to evaluate the quality of the annotations. By the end of that step, the annotators presented a mean agreement of 0.86 with standard deviation of 0.087. Then, we provided each user with a set of documents to annotate. When each user finished the annotation, we provided them with another small annotation set in order to verify the annotation agreement, comparing each annotator agreement to the annotations produced by a specialist. The annotators then presented of mean agreement of 0.91 with standard deviation of 0.044. Finally we provided each user with another set of documents to annotate.

By the end of the annotation process, we joined the annotated sets of each user and composed KAUANE INSURANCE UPPER corpus, which comprises 1,250 documents and 133,203 tokens, an average of 106.56 tokens per document. This final corpus contains gold standard annotations considered as the groundtruth for training and test.

5.2.1.4 Adopted Annotation

The corpora information is codified on a per token basis. In Table 5.9, we present the annotation of a sentence with its three basic features. The first one is the word. The next one is the part-of-speech (POS) provided by an in-house tagger.[5] We use the IOB format (Ramshaw and Marcus, 1995) to annotate

[5] We used a BILSTM based tagger trained for Portuguese.

the entities of our tagsets. In Table 5.9, we have two annotated entities: *claims procedure* (ClmProcedure), and *the health cause that motivated the judicial procedure* (HltCausMot).

Table 5.9: Annotated corpus example. *ClmProcedure* is claims procedure, and *HltCausMot* is the health cause that motivated the judicial procedure, both annotated using the IOB format. Each column corresponds to a token.

Word	pediu	cirurgia	de	câncer	de	próstata
POS	V	N	PREP	N	PREP	N
Entity	O	B-ClmProcedure	O	B-HltCausMot	I-HltCausMot	I-HltCausMot

5.2.2 Word Embedding

We tested three types of word embedding to be used as the basis for our tasks: GloVe (Pennington et al., 2014), FastText (Bojanowski et al., 2016), and Word2vec (Mikolov et al., 2013). We evaluated several pretrained models (Hartmann et al., 2017), trained on a big Portuguese corpus of several sources and textual genres, and local models, trained using a dataset of 96,691 Court decisions. For Word2vec, we also loaded the pretrained models, and trained the models using the dataset of Court decisions. All models have vector dimension of 100, and all local models were trained using window size of 5, minimum word count of 2, and for 100 epochs. In Table 5.10, we present all different models evaluated.

Table 5.10: Word embeddings tested to be used as the basis for our tasks. *Prec.* means precision, and *Rec.* means recall.

Type	*Algorithm*	*Training*	*Prec. (%)*	*Rec. (%)*	$F_{\beta=1}$*(%)*
Word2vec	**Skipgram**	**pretrained**	**70.73**	**72.73**	**71.71**
Word2vec	Skipgram	pretrained and local	69.03	74.02	71.43
GloVe	–	local	69.42	73.43	71.36
FastText	Skipgram	pretrained	69.82	72.80	71.23
FastText	Skipgram	local	70.45	72.14	71.23
Word2vec	Skipgram	local	70.19	72.36	71.23
Word2vec	CBOW	pretrained	69.04	69.83	69.39
GloVe	–	pretrained	66.62	69.82	68.15
Word2vec	CBOW	local	54.24	60.13	57.01
FastText	CBOW	pretrained	53.38	58.55	55.76
Word2vec	CBOW	pretrained and local	43.49	49.97	46.27
FastText	CBOW	local	21.86	20.72	21.16

With each trained word embedding, we performed a 5-fold cross-validation of a BILSTM network with 32 hidden layers, using the training set of Kauane Insurance Report. We established 50 epochs for each cross-validation run. We used RMSProp algorithm as optimizer, with learning rate of 0.01, ρ of 0.9, and batch size of 32. We present the mean cross-validation quality for each trained word embedding in Table 5.10. We notice that models trained using the Skipgram algorithm performed better than the models trained using CBOW algorithm. The Skipgram algorithm produces a distributed representation of words, where the model is trained to predict a surrounding window of words for the current word. Moreover, we verify that all different types of word embeddings evaluated showed competitive quality. Finally, using a pretrained model, a pretrained model further trained on a judicial dataset, or a model trained specifically on a judicial dataset does not impact the quality of the next tasks significantly, considering Word2vec and FastText. GloVe performs better when trained specifically on a judicial dataset. The line in boldface shows which model presented the best results.

5.2.3 Char Embedding

We generated character embedding representation for the words using the model proposed by (Ling et al., 2015). The Bidirectional LSTM of the model has 32 units, the input dimension and the output dimension are the same for each dataset. The model is trained jointly with our Deep Learning methods. For computing time restrictions, we had to establish a maximum word size that corresponds to the input dimension, truncating the words that had more characters than the maximum allowed. The maximum word size was established as the percentile 90% of the length of the words of each dataset. For Kauane Insurance Report, the maximum word size is 10. For Kauane Insurance Lower, the maximum word size is 9. For Kauane Insurance Upper, the maximum word size is 9.

Table 5.11: Annotated corpora set sizes. *KI* means Kauane Insurance.

Corpus	*Part*	*#Documents*	*#Tokens*
KI Report	Training	2,544	520,202
	Test	636	126,638
KI Lower	Training	2,653	471,346
	Test	664	119,223
KI Upper	Training	1,000	106,479
	Test	250	26,724

5.2.4 Experimental Setup

In order to assess the proposed models, we divided the annotated corpora into training set and test set. Table 5.11 presents the annotated corpora sizes. Table 5.12 shows the number of annotated entities by set, and Table 5.13 shows number of documents in which each entity appeared by set.

In our work, we create five models to tackle each extraction task: (i) a Conditional Random Fields (CRF); (ii) a Bidirectional Long Short-Term Memory network (BILSTM); (iii) a combination of BILSTM and CRF (BILSTM-CRF); (iv) a Bidirectional Long Short-Term Memory network with char embeddings concatenated to word embeddings (BILSTM-CE); and (v) a combination of BILSTM-CE and CRF (BILSTM-CE-CRF).

Table 5.12: Annotated entities by set for each corpus. *KI* means Kauane Insurance.

Corpus	*Entity*	*Training*	*Test*
KI Report	The health cause that motivated the judicial procedure	1,736	443
	Claims material damage	344	102
	Claims medicine/exam	142	31
	Claims procedure	699	156
	Claims treatment	486	123
	The claimed value of legal fee due by defeated party	20	6
	The claimed value of the material damage	167	45
	The claimed value of the moral damage	419	80
KI Lower	Material damage	285	64
	Medicine/exam	85	14
	Procedure	308	77
	Treatment	262	62
	The value of legal fee due by defeated party	2,598	642
	The value of the material damage	281	77
	The value of the moral damage	2,094	547
KI Upper	Material damage	40	19
	Moral damage has been excluded	137	28
	The value of legal fee due by defeated party	423	91
	The value of the material damage	108	18
	The value of the moral damage	516	129

We used Keras 2.2.4, Keras-contrib 2.0.8, and Tensorflow-gpu 1.15.0 to run the deep neural network experiments.

In order to calibrate our models, we used a 5-fold cross-validation over the training set. For models (ii) to (v), we performed a grid search in order to find the best neural network architecture, preserving the first layer – a pre-trained

Table 5.13: Number of documents in which each entity appeared by set for each corpus. *KI* means Kauane Insurance.

Corpus	*Entity*	*Training*	*Test*
	The health cause that motivated the judicial procedure	1,416	333
	Claims material damage	336	106
	Claims medicine/exam	138	30
	Claims procedure	699	155
KI Report	Claims treatment	484	122
	The claimed value of legal fee due by defeated party	20	6
	The claimed value of the material damage	158	39
	The claimed value of the moral damage	411	80
	Material damage	276	76
	Medicine/exam	83	1
	Procedure	300	70
KI Lower	Treatment	256	61
	The value of legal fee due by defeated party	2,451	624
	The value of the material damage	262	78
	The value of the moral damage	2,019	535
	Material damage	37	15
	Moral damage has been excluded	137	27
KI Upper	The value of legal fee due by defeated party	387	87
	The value of the material damage	100	7
	The value of the moral damage	507	139

word2vec layer with 100 units – and the last layer of our neural network. For models (ii) and (iv), the last layer is a Softmax layer. For models (iii) and (v), the last layer is a Dense layer with ReLU activation, followed by a CRF layer. Table 5.14 presents the grid search step performed for models (ii) and (iii). We created a dummy element called *None* to represent the absence of an element in that level. Table 5.15 presents the grid search step performed for models (iv) and (v). We work only with one layer for those models, since it is costly to compute the char embedding representation. For all models, we used RMSProp algorithm as optimizer, with learning rate of 0.01, ρ of 0.9, batch size of 32, and dropout rate of 0.2. We established 60 epochs for each cross-validation run.

For the first model, CRF, we performed a grid search to find the best C1 and C2 hyperparameters. The candidate values for each of them were 0.001, 0.01 and 0.1.

In the next sections, we show the configurations that presented the best qualities for each of the extraction tasks. For models (ii) to (v), we chose the number of epochs based on the mean of the lowest losses in the validation set of the cross-validation run of the best architecture.

5.2.4.1
Kauane Insurance Report

In this section, we show the best architectures and hyperparameters for Kauane Insurance Report.

The BILSTM-CE-CRF architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 10 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 17 units with ReLU activation followed by a CRF layer. The number of epochs for training is 23.

Table 5.14: Combination of neural network levels in order to find the best architecture for BILSTM and BILSTM-CRF.

Level	*Layers*
1	None
	BILSTM with 64 hidden units
	BILSTM with 128 hidden units
	BILSTM with 256 hidden units
	BILSTM with 512 hidden units
2	BILSTM with 64 hidden units
	BILSTM with 128 hidden units
	BILSTM with 256 hidden units
	BILSTM with 512 hidden units

The BILSTM-CRF architecture that displayed the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 64 units – a forward LSTM layer with 32 units, and a backward LSTM layer with 32 units. Next, there is another Bidirectional LSTM with 64 units – a forward LSTM layer with 32 units, and a backward LSTM layer with 32 units. Finally, a dense layer with 17 units with ReLU activation followed by a CRF layer. The number of epochs for training is 40.

Table 5.15: Different neural network layers validated in order to find the best architecture for BILSTM-CE and BILSTM-CE-CRF.

Layers
BILSTM with 64 hidden units
BILSTM with 128 hidden units
BILSTM with 256 hidden units
BILSTM with 512 hidden units

The BILSTM architecture that presented the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 256 units – a forward LSTM layer with 128 units, and a backward LSTM layer with 128 units. Next, there is another Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 17 units with Softmax activation. The number of epochs for training is 3.

The BILSTM-CE architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 10 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 512 units – a forward LSTM layer with 256 units, and a backward LSTM layer with 256 units. Finally, a dense layer with 17 units with Softmax activation. The number of epochs for training is 3.

The best combination of hyperparameters for the CRF model was C1 = 0.1 and C2 = 0.1.

Table 5.16: Performance of the extraction tasks assessed in the test set. The models are ordered by $F_{\beta=1}$ for each task. *KI* means Kauane Insurance.

Task	*Model*	*Precision (%)*	*Recall (%)*	$F_{\beta=1}$*(%)*
KI Report	BILSTM-CE-CRF	64.69	69.80	67.15
	BILSTM-CRF	67.76	63.90	65.77
	CRF	63.72	48.30	54.95
	BILSTM	30.76	21.90	25.58
	BILSTM-CE	25.55	9.30	13.64
KI Lower	BILSTM-CE-CRF	90.30	87.97	89.12
	CRF	92.27	85.26	88.63
	BILSTM-CRF	89.14	79.78	84.20
	BILSTM-CE	80.60	82.35	81.46
	BILSTM	81.63	76.93	79.21
KI Upper	BILSTM-CRF	83.22	84.10	83.66
	CRF	83.83	78.80	81.24
	BILSTM-CE	75.68	79.15	77.37
	BILSTM-CE-CRF	75.08	78.80	76.90
	BILSTM	39.94	51.24	44.89

5.2.4.2 Kauane Insurance Lower

In this section, we show the best architectures and hyperparameters for Kauane Insurance Lower.

The BILSTM-CE-CRF architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 9 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 512 units – a forward LSTM layer with 256 units, and a backward LSTM layer with 256 units. Finally, a dense layer with 15 units with ReLU activation followed by a CRF layer. The number of epochs for training is 53.

Table 5.17: Detailed performance of the Kauane Insurance Report Extraction task for CRF, BILSTM-CE-CRF, BILSTM-CE, BILSTM-CRF, and BILSTM. *HltCausMot* is the health cause that motivated the judicial procedure; *ClmMatDm* is claims material damage; *ClmMedExam* is claims medicine/exam; *ClmProcedure* is claims procedure; *ClmTreatment* is claims treatment; *ClmValLegFee* is the claimed value of legal fee due by defeated party; *ClmValMatDm* is the claimed value of the material damage; and *ClmValMorDm* is the claimed value of the moral damage.

Entity	CRF			BILSTM-CE-CRF			BILSTM-CE			BILSTM-CRF			BILSTM		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
HltCausMot	61.31	45.47	52.22	63.53	58.06	60.67	23.78	9.71	13.79	71.05	53.64	61.13	28.28	36.20	31.75
ClmMatDm	68.33	37.96	48.81	76.34	65.74	70.65	0.00	0.00	0.00	64.81	64.81	64.81	0.00	0.00	0.00
ClmMedExam	85.00	56.67	68.00	60.98	83.33	70.42	0.00	0.00	0.00	58.54	80.00	67.61	0.00	0.00	0.00
ClmProcedure	62.32	55.13	58.50	67.35	84.62	75.00	0.00	0.00	0.00	71.26	79.49	75.15	0.00	0.00	0.00
ClmTreatment	54.44	39.52	45.79	53.37	83.06	64.98	0.00	0.00	0.00	65.41	70.16	67.70	0.00	0.00	0.00
ClmValLegFee	100.00	16.67	28.57	100.00	33.33	50.00	0.00	0.00	0.00	50.00	33.33	40.00	0.00	0.00	0.00
ClmValMatDm	51.85	32.56	40.00	55.77	67.44	61.05	0.00	0.00	0.00	47.92	53.49	50.55	0.00	0.00	0.00
ClmValMorDm	80.23	86.25	83.13	82.95	91.25	86.90	27.68	61.25	38.13	70.97	82.50	76.30	47.41	68.75	56.12
General	63.72	48.30	54.95	64.69	69.80	67.15	25.55	9.30	13.64	67.76	63.90	65.77	30.76	21.90	25.58

The BILSTM-CRF architecture that displayed the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Next, there is another Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 15 units with ReLU activation followed by a CRF layer. The number of epochs for training is 41.

The BILSTM architecture that presented the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 64 units – a forward LSTM layer with 32 units, and a backward LSTM layer with 32 units. Next, there is another Bidirectional LSTM with 64 units – a forward LSTM layer with 32 units, and a backward LSTM layer with 32 units. Finally, a dense layer with 15 units with Softmax activation. The number of epochs for training is 5.

The BILSTM-CE architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 9 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 64 units – a forward LSTM layer with 32 units, and a backward LSTM layer with 32 units. Finally, a dense layer with 15 units with Softmax activation. The number of epochs for training is 4.

The best combination of hyperparameters for the CRF model was C1 = 0.1 and C2 = 0.1.

5.2.4.3
Kauane Insurance Upper

In this section, we show the best architectures and hyperparameters for Kauane Insurance Upper.

The BILSTM-CE-CRF architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 9 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 256 units – a forward LSTM layer with 128 units, and a backward LSTM layer with 128 units. Next, a dense layer with 11 units with ReLU activation followed by a CRF layer. The number of epochs for training is 55.

Table 5.18: Detailed performance of the Kauane Insurance Lower Extraction task for CRF, BILSTM-CE-CRF, BILSTM-CE, BILSTM-CRF, and BILSTM. *MatDm* is material damage; *MedExam* is medicine/exam; *Procedure* is procedure; *Treatment* is treatment; *ValLegFee* is the value of legal fee due by the defeated party; *ValMatDm* is the value of the material damage; and *ValMorDm* is the value of the moral damage.

Entity	CRF			BILSTM-CE-CRF			BILSTM-CE			BILSTM-CRF			BILSTM		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
MatDm	79.55	45.45	57.85	69.23	58.44	63.38	87.50	54.55	67.20	69.77	38.96	50.00	93.33	18.18	30.43
MedExam	68.75	61.11	64.71	62.50	83.33	71.43	0.00	0.00	0.00	30.00	33.33	31.58	0.00	0.00	0.00
Procedure	77.36	58.57	66.67	78.85	58.57	67.21	70.18	57.14	62.99	48.65	25.71	33.64	100.00	14.29	25.00
Treatment	65.22	49.18	56.07	67.86	62.30	64.96	57.89	18.03	27.50	85.71	19.67	32.00	60.00	19.67	29.63
ValLegFee	95.67	94.64	95.15	94.62	94.33	94.48	86.22	93.87	89.88	90.22	86.22	88.18	82.23	87.14	84.61
ValMatDm	80.00	51.76	62.86	75.64	69.41	72.39	31.82	41.18	35.90	82.86	68.24	74.84	46.53	55.29	50.54
ValMorDm	94.98	93.08	94.02	94.34	94.17	94.26	84.03	91.99	87.83	95.24	94.72	94.98	88.58	93.26	90.86
General	92.27	85.26	88.63	90.30	87.97	89.12	80.60	82.35	81.46	89.14	79.78	84.20	81.63	76.93	79.21

The BILSTM-CRF architecture that displayed the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Next, a dense layer with 11 units with ReLU activation followed by a CRF layer. The number of epochs for training is 41.

The BILSTM architecture that presented the best results starts with a pre-trained word2vec layer with 100 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 256 units – a forward LSTM layer with 128 units, and a backward LSTM layer with 128 units. Next, there is another Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 11 units with Softmax activation. The number of epochs for training is 7.

The BILSTM-CE architecture that presented the best results starts with a pre-trained word2vec layer with 100 units concatenated with a char embedding layer with 9 units, succeeded by a Dropout layer, and followed by a Bidirectional LSTM with 128 units – a forward LSTM layer with 64 units, and a backward LSTM layer with 64 units. Finally, a dense layer with 11 units with Softmax activation. The number of epochs for training is 5.

The best combination of hyperparameters for the CRF model was C1 = 0.1 and C2 = 0.1.

5.2.5 Experiment Results

Table 5.16 presents the performance of our models assessed in the test sets of our three corpora. For Kauane Insurance Report, BILSTM-CE-CRF obtained an $F_{\beta=1}$ score of 67.15%, which is an increase in quality of 1.38% compared to the second best model, BILSTM-CRF. Compared to CRF and BILSTM and BILSTM-CE, BILSTM-CE-CRF increases the quality by 12.2%, 41.57%, and 53.51%, respectively. Table 5.17 presents the detailed quality of those models for each entity type assessed in the test set.

For Kauane Insurance Lower, BILSTM-CE-CRF obtained an $F_{\beta=1}$ score of 89.12%, which is an increase in quality of 0.49% compared to the second best model, CRF. Compared to BILSTM-CRF and BILSTM-CE and BILSTM, BILSTM-CE-CRF increases the quality by 4.92%, 7.66%, and 9.91%, respectively. Table 5.18 shows the detailed quality of those models for each entity type assessed in the test set.

Table 5.19: Detailed performance of the Kauane Insurance Upper Extraction task for CRF, BILSTM-CE-CRF, BILSTM-CE, BILSTM-CRF, and BILSTM. *MatDm* is material damage; *MorDmExc* is moral damage has been excluded; *ValLegFee* is the value of legal fee due by defeated party; *ValMatDm* is the value of the material damage; and *ValMorDm* is the value of the moral damage.

Entity	CRF			BILSTM-CE-CRF			BILSTM-CE			BILSTM-CRF			BILSTM		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
MatDm	100.00	13.33	23.53	70.00	46.67	56.00	0.00	0.00	0.00	54.55	40.00	46.15	0.00	0.00	0.00
MorDmExc	71.88	85.19	77.97	68.42	96.30	80.00	85.71	66.67	75.00	80.77	77.78	79.25	57.14	59.26	58.18
ValLegFee	88.24	82.42	85.23	88.51	84.62	86.52	85.19	75.82	80.23	83.51	89.01	86.17	4.40	8.79	5.86
ValMatDm	25.00	57.14	34.78	10.00	28.57	14.81	0.00	0.00	0.00	22.22	28.57	25.00	0.00	0.00	0.00
ValMorDm	90.84	83.22	86.86	78.17	77.62	77.89	70.62	95.80	81.31	89.51	89.51	89.51	79.08	84.62	81.76
General	83.83	78.80	81.24	75.08	78.80	76.90	75.68	79.15	77.37	83.22	84.10	83.66	39.94	51.24	44.89

For KAUANE INSURANCE UPPER, BILSTM-CRF obtained an $F_{\beta=1}$ score of 83.66%, which is an increase in quality of 2.42% compared to the second best model, CRF. Compared to BILSTM-CE and BILSTM-CE-CRF and BILSTM, BILSTM-CRF increases the quality by 6.29%, 6.76%, and 38.77%, respectively. Table 5.19 presents the detailed quality of those models for each entity type assessed in the test set.

We employed the *conlleval* tool to evaluate the performance of the proposed models. The performance of our models cannot be directly compared to previous work, since their corpora and tagsets are different.

5.2.6 Visual Analysis

The third step of our methodology is the visualization of the aggregated information extracted from the decisions. Therefore we generated a set of charts presenting them in a dashboard.

In this section, we perform a visual analysis using the gold standard annotations for our three corpora. Our datasets include decisions of the State Court of Rio de Janeiro from the lower and Appellate Court from 2016 and 2017, excluding interlocutory appeals – decisions that were not taken to the collegiate body – or that analyzed motions for clarifying or amending the judgment.

We decided to use the gold standard annotations instead of the annotations extracted by our models to eliminate any noise in our analysis. However, when performing another visual analysis with new data, it will be done using the annotations extracted by our models.

5.2.6.1 Data Preprocessing

We performed a preprocessing step in order to prepare data for visualization. For KAUANE INSURANCE REPORT, we applied regular expressions to the entity *the claimed value of the moral damage* in order to extract its value. When the value is specified in term of minimum wage, we performed the multiplication using the minimum wage value as one thousand reais. We also applied regular expressions to extract the value from the entity *the claimed value of the material damage*. Moreover, we converted entities *claims material damage*, *claims medicine/exam*, *claims procedure*, and *claims treatment* to boolean values. Finally, we manually created rules to convert the entity *the* *health cause that motivated the judicial procedure* to the Brazilian version[6] of the International Classification of Diseases[7] (ICD), 10th revision, provided by the World Health Organization. The created rules try to generate the most faithful conversion, but there are known mismatches. After converting to the corresponding ICD, we used the grouping of ICDs suggested by the ICD, 10th revision, which divides the diseases into 22 chapters as presented in Table 5.20. We added an extra column, *Short Description*, which will be used in the visualizations.

Table 5.20: Grouping of the International Classification of Diseases, 10th revision, into chapters. We added an extra column, *Short Description*.

Chapter	*Description*	*Short Description*
I	Certain infectious and parasitic diseases	Infectious
II	Neoplasms	Neoplasm
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	Blood
IV	Endocrine, nutritional and metabolic diseases	Endocrine
V	Mental and behavioural disorders	Mental
VI	Diseases of the nervous system	Nervous
VII	Diseases of the eye and adnexa	Eye
VIII	Diseases of the ear and mastoid process	Ear
IX	Diseases of the circulatory system	Circulatory
X	Diseases of the respiratory system	Respiratory
XI	Diseases of the digestive system	Digestive
XII	Diseases of the skin and subcutaneous tissue	Skin
XIII	Diseases of the musculoskeletal system and connective tissue	Musculoskeletal
XIV	Diseases of the genitourinary system	Genitourinary
XV	Pregnancy, childbirth and the puerperium	Childbirth and post-childbirth
XVI	Certain conditions originating in the perinatal period	Perinatal
XVII	Congenital malformations, deformations and chromosomal abnormalities	Congenital
XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	Exams
XIX	Injury, poisoning and certain other consequences of external causes	Poisoning
XX	External causes of morbidity and mortality	Death
XXI	Factors influencing health status and contact with health services	Health services
XXII	Codes for special purposes	Special

For Kauane Insurance Lower, we applied regular expressions to the entity *the value of the moral damage* in order to extract its value. When the value is specified in term of minimum wage, we performed the multiplication

[6]The Brazilian version of the International Classification of Diseases, 10th revision, can be found at the following URL: http://www.datasus.gov.br/cid10/V2008/principal.htm

[7]The International Classification of Diseases, 10th revision, can be found at the following URL: https://www.who.int/classifications/icd/en/

using the minimum wage value as one thousand reais. We also applied regular expressions to extract the value from the entity *the value of the material damage*. Furthermore, we converted entities *material damage*, *medicine/exam*, *procedure*, and *treatment* to boolean values.

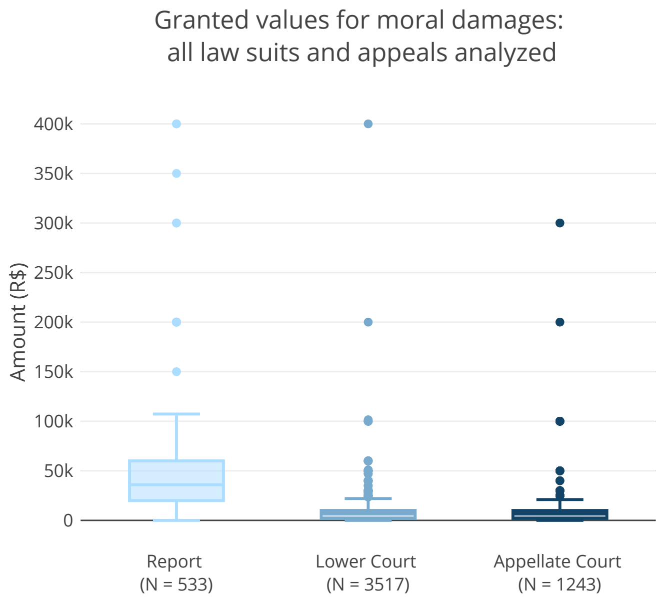
For Kauane Insurance Upper, we applied regular expressions to the entity *the value of the moral damage* in order to extract its value. When the value is specified in term of minimum wage, we performed the multiplication using the minimum wage value as one thousand reais. We also applied regular expressions to extract the value from the entity *the value of the material damage*. Moreover, we converted entities *material damage*, and *moral damage has been excluded* to boolean values.

A lower Court report can have several entities *the health cause that motivated the judicial procedure*. We performed the conversion of those entities to their respective chapters, as described above. In order to present a more detailed analysis that represents the number of health care causes that motivated the law suit, thus, representing each chapter, we replicated each report to the number of different chapters found in the initial report (e.g. a report with the chapters Eye and Ear is replicated to two reports, one with the chapter Eye, and another with the chapter Ear). Then, we duplicated the lower and Appellate Court decisions related to those new reports. The dataset for analysis is composed of 3,517 lower Court reports, 3,517 lower Court decisions, and 1,243 Appellate Court decisions.

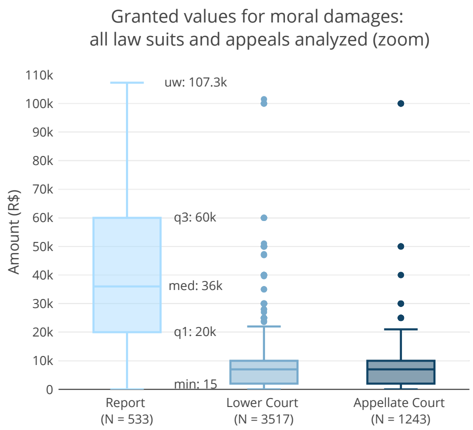
5.2.6.2 Analysis

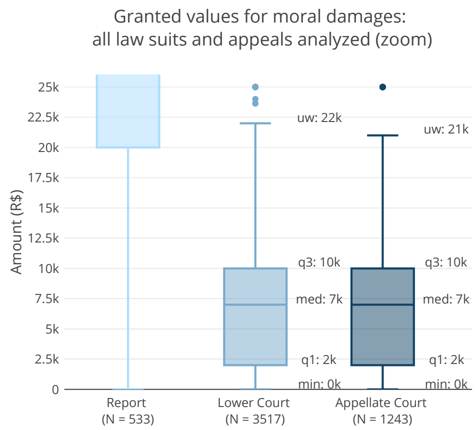
In our analysis, firstly we notice that the entity *the health cause that motivated the judicial procedure*, reported on the Appellate Court decisions denying the appeals, does not appear in 944 lower Court reports, and it was not converted by the manual rules 56 times. The three most common chapters, composed of the identified entities converted to their corresponding ICDs, are the following diseases: Infectious with 244 occurrences, Neoplasm with 203 occurrences, and Circulatory with 134 occurrences.

The entity *the health cause that motivated the judicial procedure*, reported on the Appellate Court's decisions that grants or partially upholds the appeals, does not appear in 492 lower Court reports, and it was not converted by the manual rules 36 times. The three most common chapters are Infectious with 150 occurrences, Neoplasm with 129 occurrences, and Circulatory with 64 occurrences. They are the same chapters compared to the Appellate Court decisions that denied the appeals.



(a)

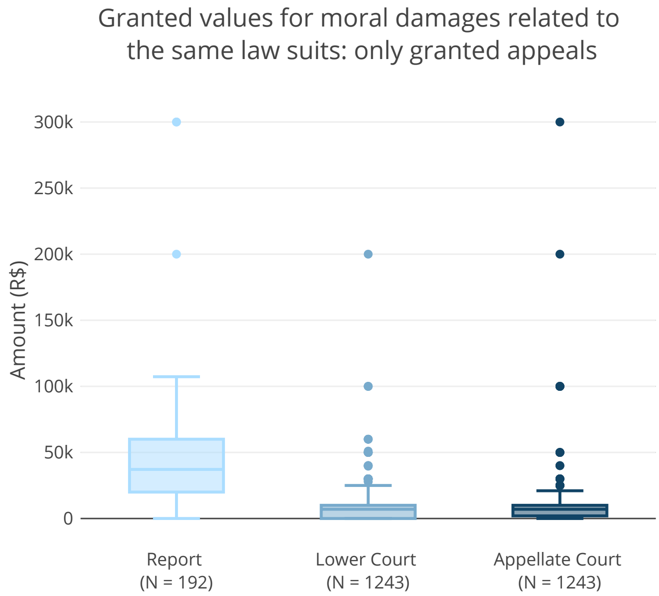




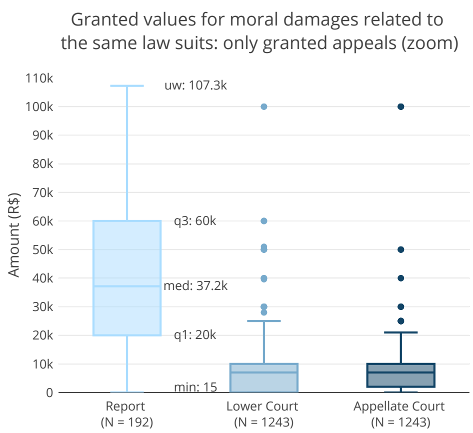
(b) (c)

Figure 5.1: Box plots of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions granting, partially upholding or denying the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

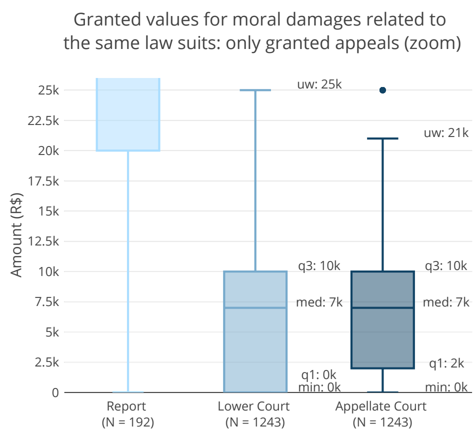
We separated the visualizations into three groups: (i) occurrences that are reported on the Appellate Court's decisions granting, partially upholding or denying the appeal; (ii) occurrences that are reported on the Appellate Court's decisions granting or partially upholding the appeal; and (iii) occurrences that are reported on the Appellate Court's decisions denying the appeal.



(a)



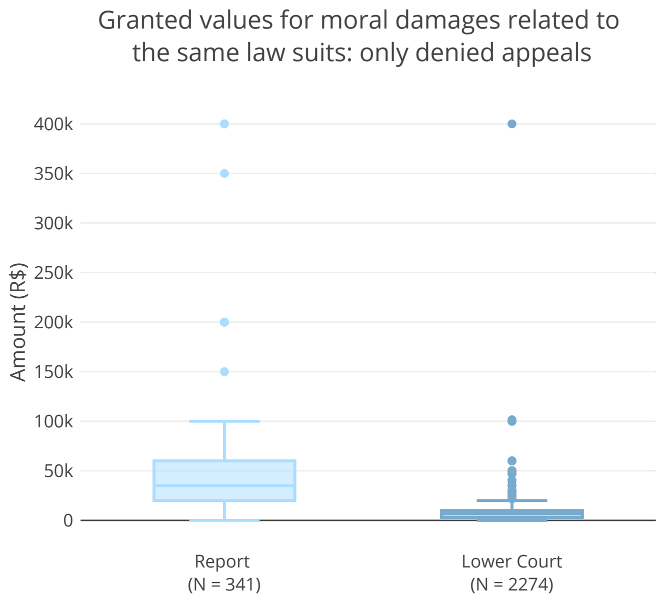
(b)



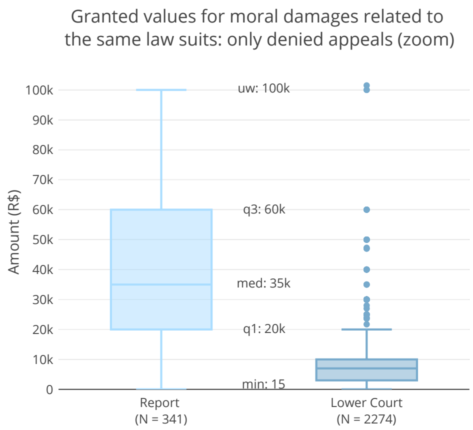
(c)

Figure 5.2: Box plots of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions granting or partially upholding the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

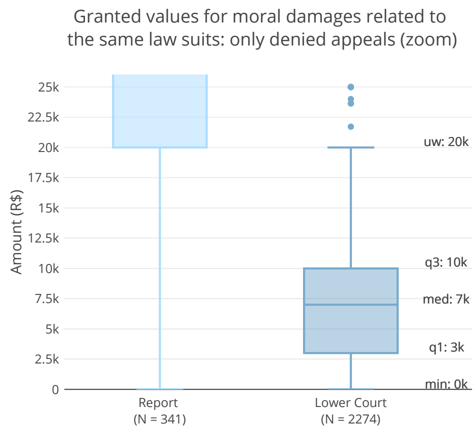
We generated scatter plots relating the entities *the claimed value of the*



(a)



(b)



(c)

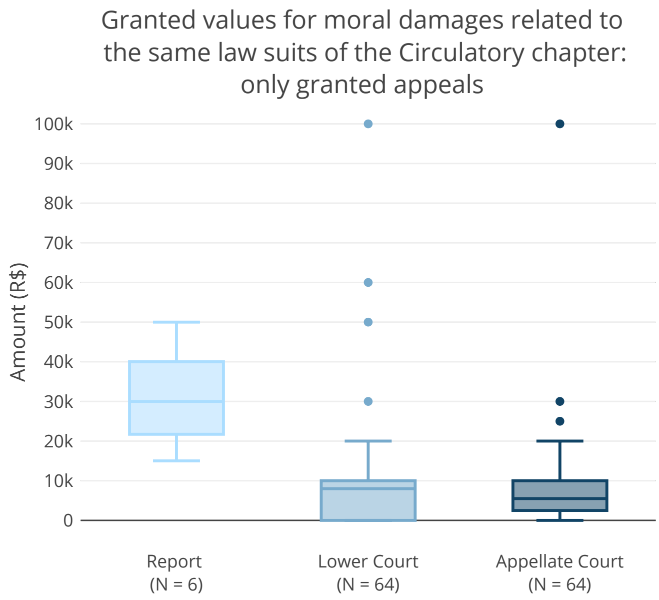
Figure 5.3: Box plots of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions denying the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

moral damage on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. However, as most of the claimed values are empty – 2,984 in total – we were not able to perform an interesting analysis. Then, we generated box plots to analyze each variable separately.

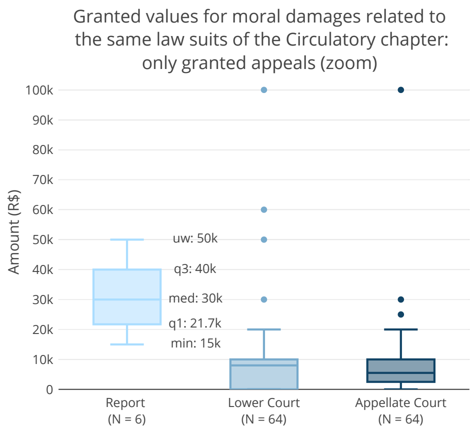
Figure 5.1 presents three box plots from group (i) of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. Those entities are placed in the same axis. In the lower Court reports, the first quartile, which is 20 thousand reais, is almost the same comparing to the upper whiskers of lower and Appellate Court decisions – 22 thousand reais and 21 thousand reais, respectively. That could indicate that the moral damage values claimed by the plaintiffs are higher compared to the values that are usually granted by the Courts. Furthermore, comparing lower and Appellate Court decisions, they only differ in the upper whisker, although the values are similar – 22 thousand reais and 21 thousand reais, respectively. That indicates Appellate Courts usually do not provide values out of the range usually provided by lower Courts.

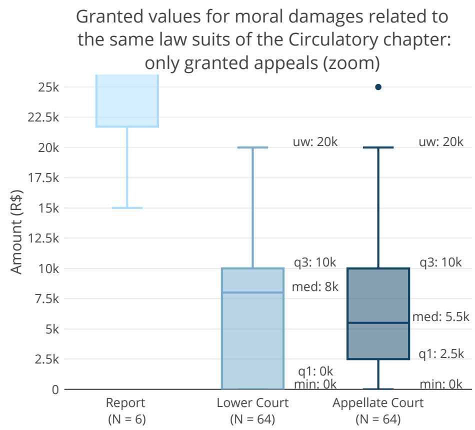
Figure 5.2 presents three box plots from group (ii) of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. Those entities are placed in the same axis. In the lower Court reports, the first quartile, which is 20 thousand reais, is almost the same comparing to the upper whisker of Appellate Court decisions – 21 thousand reais. Moreover, the first quartile and the third quartile – 20 thousand reais and 60 thousand reais, respectively – are the same comparing to lower Court reports from Figure 5.1, and the median is almost the same – 37.2 thousand reais and 36 thousand reais, respectively. Furthermore, the median is the same comparing lower and Appellate Court decisions, which is 7 thousand reais. The upper whisker of lower Court decisions is 25 thousand reais, and the first quartile is zero, which is also the lower whisker. In the Appellate Court decisions, the upper whisker is 21 thousand reais, the lower whisker is zero, and the first quartile is 2 thousand reais. That indicates Appellate Court decisions make the values come closer to the median.

Figure 5.3 presents two box plots from group (iii) of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower Court decisions. Those entities are placed in the same axis. In the lower Court reports, the first quartile, which is 20 thousand reais, is the same comparing to the upper whisker of lower Court decisions. Moreover, the first quartile and the third quartile – 20 thousand reais and 60 thousand reais, respectively – are the same comparing to lower Court reports from Figure 5.1, and the median is almost the same – 35 thousand reais and 36 thousand reais, respectively. In addition, we verify lower Court decisions have the same median compared to Appellate Court decisions from Figure 5.2, which is 7 thousand reais. Furthermore, the first and the third quartile are



(a)





(b) (c)

Figure 5.4: Box plots of the Circulatory chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions granting or partially upholding the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

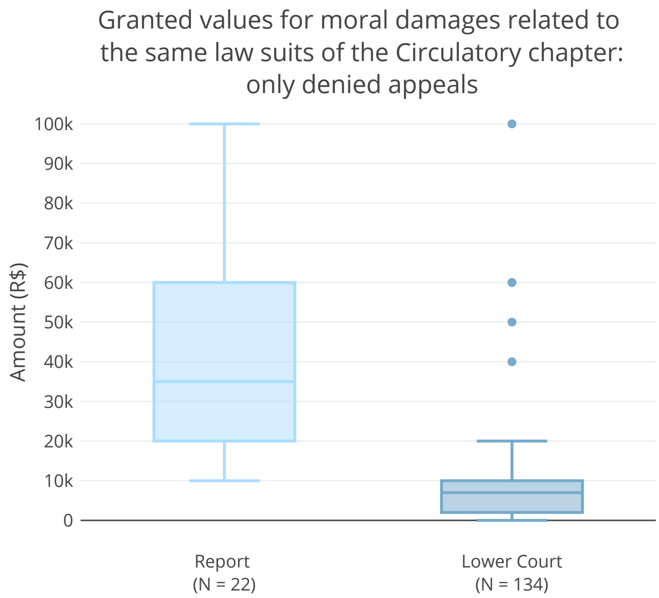
similar – for Appellate Court decisions from Figure 5.2, they are 2 thousand reais and 10 thousand reais, respectively; and for lower Court decisions from Figure 5.3, they are 3 thousand reais and 10 thousand reais, respectively.

Figure 5.4 presents three box plots from group (ii) of the Circulatory chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. Those entities are placed in the same axis. For lower Court reports, the median is 30 thousand reais, the first quartile is 21.7 thousand reais, and the third quartile is 40 thousand reais. For lower Court decisions, the median is 8 thousand reais, the first quartile is zero, and the third quartile is 10 thousand reais. For Appellate Court decisions, the median is 5,500 reais, the first quartile is 2,500 reais, and the third quartile is 10 thousand reais. We notice that the median decreased from the lower to the Appellate Court decisions, but the first quartile increased.

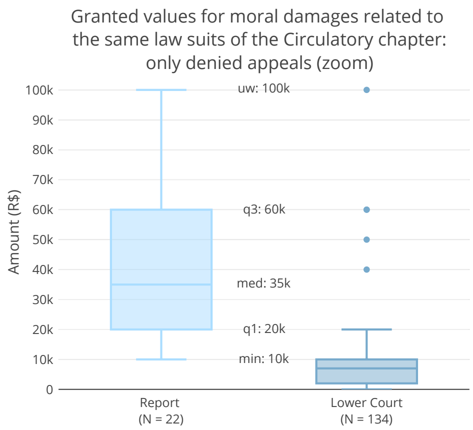
Figure 5.5 presents two box plots from group (iii) of the Circulatory chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower Court decisions. Those entities are placed in the same axis. For lower Court reports, the median is 35 thousand reais, the first quartile is 20 thousand reais, and the third quartile is 60 thousand reais. For lower Court decisions, the median is 7 thousand reais, the first quartile is 2 thousand reais, and the third quartile is 10 thousand reais. We notice those lower Court decisions have the first quartile similar compared to Appellate Court decisions from Figure 5.4 – 2 thousand reais and 2,500 reais, respectively – and the third quartile is the same – 10 thousand reais.

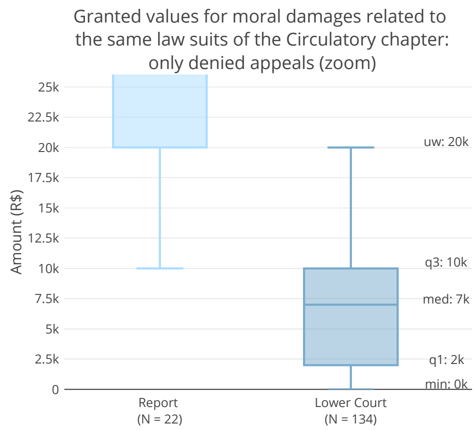
Figure 5.6 presents three box plots from group (ii) of the Infectious chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. Those entities are placed in the same axis. For lower Court reports, the median is 40 thousand reais, the first quartile is 30 thousand reais, and the third quartile is 67.8 thousand reais. For lower and Appellate Court decisions, the median is 6 thousand reais, the first quartile is zero, and the third quartile is 10 thousand reais.

Figure 5.7 presents two box plots from group (iii) of the Infectious chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower Court decisions. Those entities are placed in the same axis. For lower Court reports, the median is 26.7 thousand reais, the first quartile is 15 thousand reais, and the third quartile is 50 thousand reais. For lower Court decisions, the median is 7 thousand reais, the first quartile is 3,500 reais, and the third quartile is 10 thousand reais. We notice that those lower Court decisions have the median, and the first quartile higher than the Appellate Court decisions, according to Figure 5.6 – for Appellate Court decisions from Figure 5.6, the values are 6 thousand reais



(a)



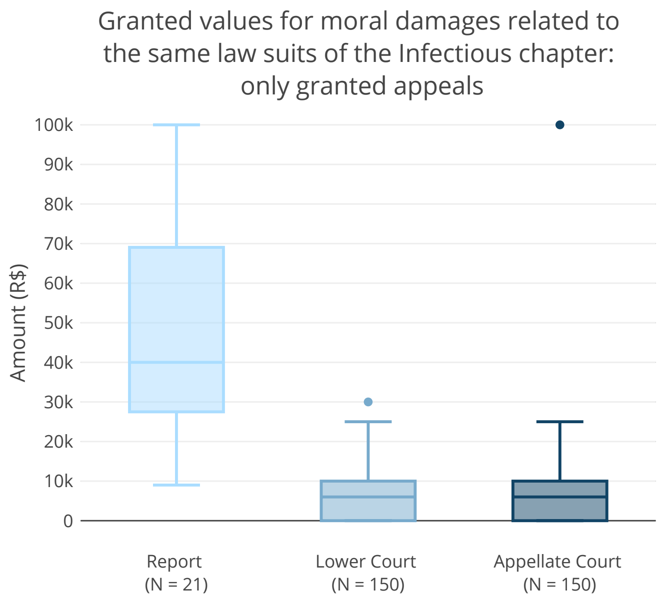


(b) (c)

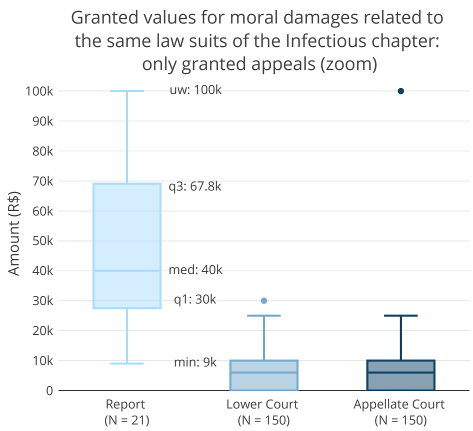
Figure 5.5: Box plots of the Circulatory chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions denying the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

and zero, respectively; for lower Court decisions from Figure 5.7, the values are 7 thousand reais and 3,500 reais, respectively. However, the third quartile is the same, corresponding to 10 thousand reais. That could indicate that lower

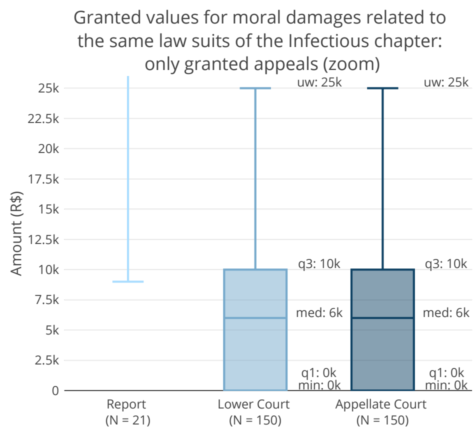
Court decisions that grant higher values for moral damage, respecting a limit of 10 thousand reais, are less reversed by the Appellate Court.



(a)



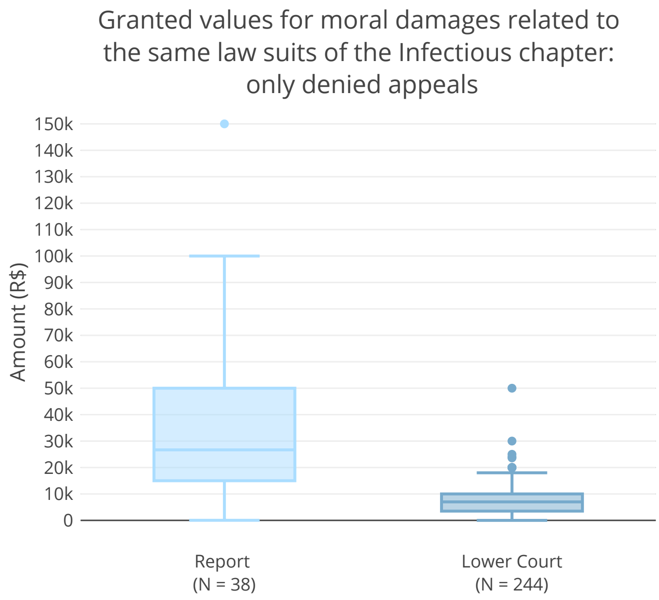
(b)



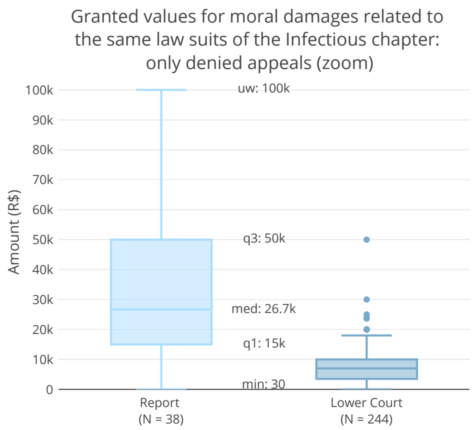
(c)

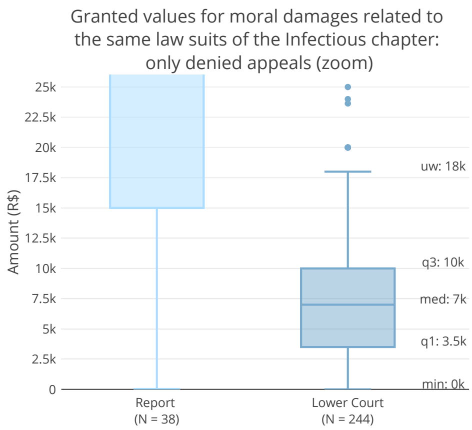
Figure 5.6: Box plots of the Infectious chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions granting or partially upholding the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

We generated box plots for the entities *the value of the material damage*



(a)





(b) (c)

Figure 5.7: Box plots of the Infectious chapter of the entities *the claimed value of the moral damage* on the lower Court reports, and *the value of the moral damage* on the lower and Appellate Court decisions. **N** corresponds to the number of elements in a box plot. Those occurrences are part of the Appellate Court's decisions denying the appeal. Part (a) contains all the data points, whilst part (b) and part (c) zoom it in order to present part of the data distribution in more detail.

and *the claimed value of the material damage* for groups (i), (ii) and (iii). However, as those values are casuistic, since it depends on each plaintiff's health care plan agreement with the insurance company, and the coverage foreseen on such agreement, we were not able to perform an interesting analysis. Then, we decided to analyze the decisions where the pecuniary value of the material damage was expressly described, verifying if each provided value is lower, equal, or higher than the value that was firstly claimed by the plaintiff.

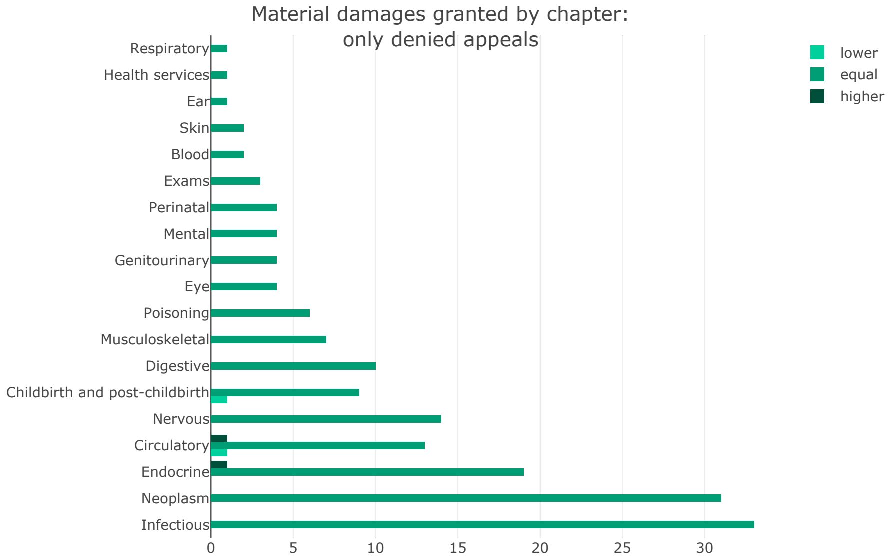


Figure 5.8: Material damages granted by chapter. Decisions that provided a *lower* value than the claimed value by the plaintiff appears in light green. Decisions that provided the *equal* value as claimed by the plaintiff are in medium green. Decisions that provided a *higher* value than the claimed value by the plaintiff are in dark green. Those occurrences are part of the Appellate Court's decisions that deny the appeal.

Figure 5.8 presents the provided material damage decisions by chapter for group (iii). Decisions that provided a lower value than the claimed value by the plaintiff are in light green. Decisions that provided the equal value than the one claimed by the plaintiff are in medium green. Decisions that provided a higher value than the claimed value by the plaintiff are in dark green. It was suppressed the bar *Not Informed*, with 125 occurrences, in which 5 occurrences appeared as the lower value claimed by the plaintiff, 116 received the same value as the one claimed by the plaintiff, and 4 received a higher value than the one claimed by the plaintiff. The Infectious chapter, with 33 occurrences, has the biggest number of material damage provisions, followed by Neoplasm with 31, and Endocrine with 20. In 16 chapters, the decisions provided exactly the same value as claimed by the plaintiff. In the Childbirth and post-childbirth chapter, there is one provision that is lower than the claimed material damage value by the plaintiff, and the remaining ones have the same value as initially claimed by the plaintiff. In the Endocrine chapter, there is one provision that is higher than the claimed material damage value by the plaintiff, and the remaining ones has the same value as initially claimed by the plaintiff. In the Circulatory chapter, there is one provision that is lower than the claimed material damage value by the plaintiff, a provision that has a higher value than the one claimed by the plaintiff, and the remaining ones has the same value as initially claimed by the plaintiff. We noticed that judges are more inclined in granting the material damage value claimed by the plaintiff, when compared to decisions that either defer a higher or lower value than the one claimed by the plaintiff.

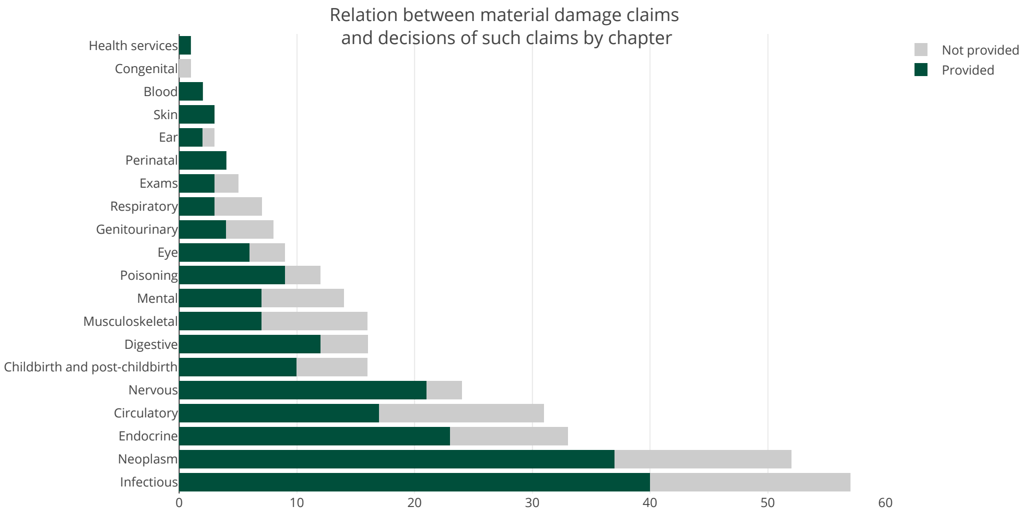


Figure 5.9: Relation between material damage claims, and decisions of such claims by chapter. Decisions that provided material damage are in dark green, whilst decisions that did not provide material damage claim are in grey. The sum of the decisions that provided material damage and the decisions that did not provide material damage corresponds to the number of material damage claims.

For group (ii), we were not able to generate a visualization for each chapter regarding decisions that granted material damage claims, since there are few data points, 49 in total, in which 12 received a higher value than the one claimed by the plaintiff, 35 received the same value as the one initially claimed by the plaintiff, and 2 received a lower material damage value than the one claimed by the plaintiff.

Figure 5.9 presents the relation by chapter between the entities *claims material damage* on the lower Court reports, and *material damage* on the lower and Appellate Court decisions, only considering precluded decisions of either Lower or Appellate Court which are related to material damage claims. Those entities correspond to a material damage claim, and a provision for such material damage, respectively. Decisions that provided material damage are in dark green, whilst decisions that did not provide material damage are in grey.

The sum of the decisions that provided material damage, and the decisions that did not provide material damage corresponds to the number of material damage claims. The figure omits the bar for *Not Informed*, with 225 occurrences of *claims material damage* and 155 occurrences of *material damage*, which correspond to lower Court reports that do not have the entity *the health cause that motivated the judicial procedure* informed. The Infectious chapter, with 57 occurrences, has the biggest number of material damage claims, followed by Neoplasm with 52, and Endocrine with 33. In general, most decisions provide material damage when it is claimed. Moreover, for chapters Blood, Skin and Perinatal, all material damage claims were provided. The Congenital chapter, in contrast, did not have any provision for material damage.

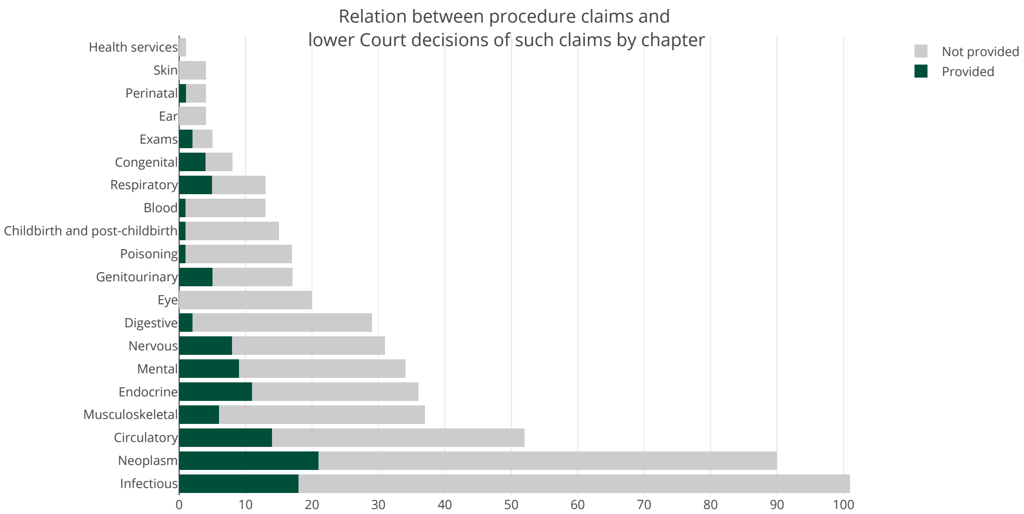


Figure 5.10: Relation between procedure claims and lower Court decisions of such claims by chapter. Decisions that provided a procedure are in dark green, whilst decisions that did not provide a procedure are in grey. The sum of the decisions that provided a procedure and the decisions that did not provide a procedure corresponds to the number of procedure claims.

Figure 5.10 presents the relation by chapter between the entities *claims procedure* on the lower Court reports, and *procedure* on the lower Court decisions. Those entities correspond to a procedure claim, and a provision for such procedure, respectively. Decisions that provided a procedure are in dark green, whilst decisions that did not provide a procedure are in grey. The sum of the decisions that provided a procedure, and the decisions that did not provide a procedure corresponds to the number of procedure claims. The figure omits the bar for *Not Informed*, with 380 occurrences of *claims procedure* and 68 occurrences of *procedure*, that corresponds to lower Court reports that do not have the entity *the health cause that motivated the judicial procedure* informed. The Infectious chapter, with 101 occurrences, has the biggest number of procedure claims, followed by Neoplasm with 90, and Circulatory with 52. In general, the number of procedure claims is considerable bigger compared to the number of provisions for procedures. In addition, four chapters did not have any provision for procedure.

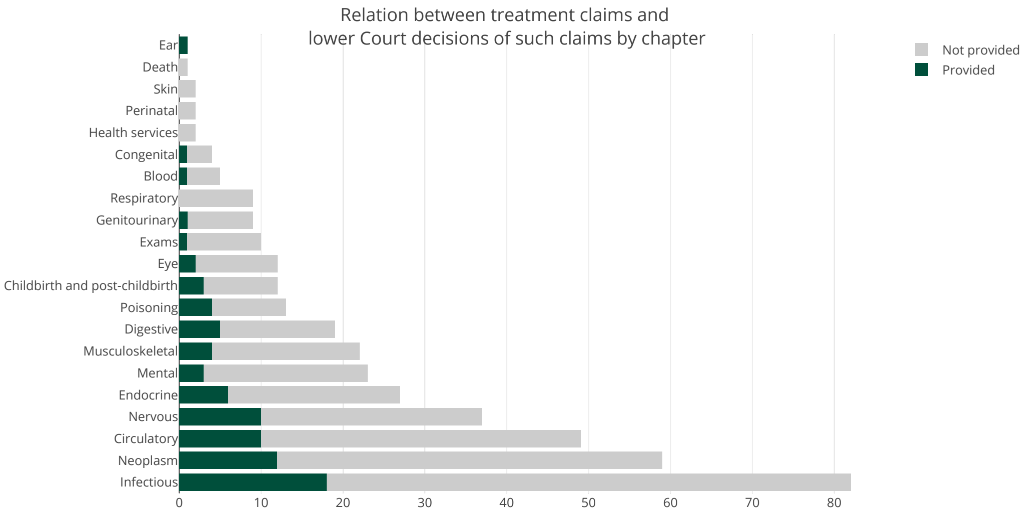


Figure 5.11: Relation between treatment claims and lower Court decisions of such claims by chapter. Decisions that provided a treatment are in dark green, whilst decisions that did not provide a treatment are in grey. The sum of the decisions that provided a treatment and the decisions that did not provide a treatment corresponds to the number of treatment claims.

Figure 5.11 presents the relation by chapter between the entities *claims treatment* on the lower Court reports, and *treatment* on the lower Court decisions. Those entities correspond to a treatment claim and a provision for such treatment, respectively. Decisions that provided a treatment are in dark green, whilst decisions that did not provide a treatment are in grey. The sum of the decisions that provided a treatment and the decisions that did not provide a treatment corresponds to the number of treatment claims. The figure omits the bar for *Not Informed*, with 263 occurrences of *claims treatment* and 47 occurrences of *treatment*, that corresponds to lower Court reports that do not have the entity *the health cause that motivated the judicial procedure* informed. The Infectious chapter, with 82 occurrences, has the biggest number of treatment claims, followed by Neoplasm with 59, and Circulatory with 49. In general, the number of treatment claims is considerable bigger compared to the number of provisions for treatments. Moreover, five chapters did not have any provision for treatment.

Figure 5.12 presents the relation by chapter between the entities *claims* *medicine/exam* on the lower Court reports, and *medicine/exam* on the lower Court decisions. Those entities correspond to a medicine/exam claim, and a provision for such medicine/exam, respectively. Decisions that provided a medicine/exam are in dark green, whilst decisions that did not provide a medicine/exam are in grey. The sum of the decisions that provided a medicine/exam, and the decisions that did not provide a medicine/exam corresponds to the number of medicine/exam claims. It was suppressed the bar *Not Informed*, with 87 occurrences of *claims medicine/exam* and 12 occurrences of *medicine/exam*, that corresponds to lower Court reports that do not have the entity *the health cause that motivated the judicial procedure* informed. The Infectious chapter, with 20 occurrences, has the biggest number of medicine/exam claims, followed by Neoplasm with 14, and Endocrine with 8. In general, the number of medicine/exam claims is considerable larger compared to the number of provisions for medicine/exam. Furthermore, five chapters did not have any provision for medicine/exam.

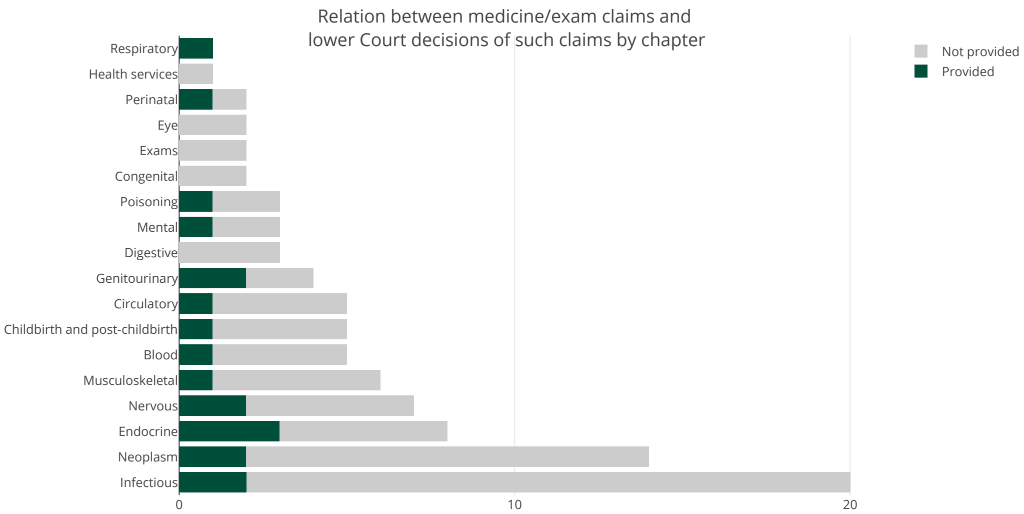


Figure 5.12: Relation between medicine/exam claims and lower Court decisions of such claims by chapter. Decisions that provided a medicine/exam are in dark green, whilst decisions that did not provide a medicine/exam are in grey. The sum of the decisions that provided a medicine/exam, and the decisions that did not provide a medicine/exam corresponds to the number of medicine/exam claims.

In this chapter, for each of our systems, we presented the composition of our datasets, the experimental setup, and the quality of the proposed models. We also performed a visual analysis in our datasets for our second system. Next, we present our conclusions and future work.

6
Conclusion

New legal systems are being developed, using technologies developed by teams of researchers of question answering, information extraction, and argument mining.

Now legal texts can be processed automatically, and useful information can be extracted from them. Judges and lawyers can benefit from it by using aggregated information extracted from past cases to decide what decision, or approach would be the most recommended for a given situation.

Most legal systems rely on an artificial simplification of the legal process. Usually legal cases deal with multiple issues: a criminal procedure may include several different charges, whilst a civil procedure may include several different requests. When those multiple charges and requests are boiled down to a single case-level variable, valuable information is lost.

Reflecting the importance of more granular information regarding judicial decision-making, recent work at the intersection of artificial intelligence and the practice of law has focused on formally conceptualizing, and identifying legal arguments within broader legal documents. However, the technical challenge involved in correlating specific legal claims, and their judicial solution is acknowledged to be hard.

In light of such challenges and opportunities, this thesis aimed at answering the research questions presented below.

Q1: How to know what the Court holds to be the law?

Q2: How to see trends of judgments in Court decisions?

Q3: How to predict judicial outcomes for specific situations?

In order to answer those questions, we have proposed a methodology to annotate Court decisions, create Deep Learning models to extract information, and visualize the aggregated information extracted from the decisions. We instantiated our methodology in two systems we have developed. The first one extracts Appellate Court modifications, which consists of a set of legal categories that are commonly modified by Appellate Courts.

This system can help lawyers to determine what the Court holds to be the law, by extracting the standards proposed by the Appellate Court modifications, answering Q1. Moreover, it can assist judges in finding whether their decisions conform to the Appellate Court's overall stance or not, leading to judicial consistency and certainty. Furthermore, it can assist lawyers to verify trends of judgments by analysing aggregated Appellate Court modifications, helping them to decide whether appeal a lower Court decision or not, answering Q2.

The second system extracts plaintiff's legal claims and each specific provision on legal opinions enacted by lower and Appellate Courts, and connects each legal claim with the corresponding judicial provision. The proposed system presents the results through visualizations that let one verify the consistency of Court decisions for a given group of claims, as well as the variation between different groups of legal claims.

This system can help lawyers to determine what the Court holds to be the law, by extracting the standards proposed by the Appellate Court decisions, answering Q1. Moreover, it can assist judges in finding whether their decisions conform to the Appellate Court's overall stance or not, leading to judicial consistency and certainty. Furthermore, it can assist lawyers and judges to verify trends of judgments by analysing aggregated information extracted from lower and Appellate Court decisions, answering Q2. In addition, it can help lawyers in predicting judicial outcomes for very specific situations. For instance, given that the plaintiff suffered cancer and was denied coverage by his healthcare provider, what is the usual legal relief? The proposed system leverages information regarding the historic trends of the Court to render an answer to that specific question, answering Q3.

This research has a main limitation: we have not performed a study with lawyers and judges, who are the parties that will effectively use the systems. Though we realize the research questions are correctly answered as presented above, only those users could respond whether they are properly answered.

Information Extraction for legal texts has been previously addressed using different techniques and for several languages. Our proposals differ from previous work since our corpora are composed of Brazilian lower and Appellate Court decisions, in which for the first system, we look for a set of modifications commonly provided by the Court, and for the second system, we look for a set of plaintiff's legal claims and judicial provisions commonly judged by the Court. To automatically extract that information, we use a traditional Machine Learning approach and a Deep Learning approach, both as alternative solutions and also as a combined solution.

For our first system, we have built the KAUANE JUNIOR corpus, using public jurisprudence data disclosed by the Appellate State Court of Rio de Janeiro. For the second system, we have built three corpora – KAUANE INsurance Report, Kauane Insurance Lower, and Kauane Insurance Upper – using also public jurisprudence data disclosed by the Appellate State Court of Rio de Janeiro. We generated gold standard annotations for the corpora, produced by human annotators and considered as the groundtruth for training and test. In addition, we have included part-of-speech (POS) annotation in the corpora.

For our first system, we created five models to tackle the extraction task of Kauane Junior: a Bidirectional Long Short-Term Memory network (BILSTM); a Bidirectional Gated Recurrent Units network (BIGRU); a Conditional Random Fields (CRF); a combination of BILSTM and CRF (BILSTM-CRF); and a combination of BIGRU and CRF (BIGRU-CRF).

BILSTM-CRF presented the best quality compared to BIGRU-CRF, CRF, BIGRU and BILSTM. The performance of our models cannot be directly compared to previous work, since their corpora and tagsets are different.

For our second system, we created five models built upon three algorithms to solve the extraction tasks of each corpus: a Conditional Random Fields (CRF); a Bidirectional Long Short-Term Memory network (BILSTM); a combination of BILSTM and CRF (BILSTM-CRF); a Bidirectional Long Short-Term Memory network with char embeddings concatenated to word embeddings (BILSTM-CE); and a combination of BILSTM-CE and CRF (BILSTM-CE-CRF).

For Kauane Insurance Report, BILSTM-CE-CRF presented the best quality compared to CRF, BILSTM, BILSTM-CE and BILSTM-CRF. For Kauane Insurance Lower, BILSTM-CE-CRF showed the best quality compared to CRF, BILSTM, BILSTM-CE and BILSTM-CRF. For Kauane Insurance Upper, BILSTM-CRF presented the best quality compared to CRF, BILSTM, BILSTM-CE and BILSTM-CE-CRF. The performance of our models cannot be directly compared to previous work, since their corpora and tagsets are different.

However, the quality of our best models for both systems are as expected, comparing with related works that use similar BILSTM-CRF methods, which are: (Chalkidis et al., 2017a), who tackle the task of extracting elements of contracts; (Nguyen et al., 2018), who deal with recognizing requisite and effectuation parts in legal sentences; and (Angelidis et al., 2018), who propose a named entity recognition system and a named entity linker system for Greek legislation.

As future work in the short term, we intend to increase the size of Kauane Junior, Kauane Insurance Report, Kauane Insurance Lower, and Kauane Insurance Upper. That would probably increase the quality of our models. Furthermore, we plan to obtain decisions of the State Court of Rio de Janeiro from the lower and Appellate Court from 2018 and 2019, apply our second system, and release the charts produced from the extracted information in a web page.

As future work in the medium term, we plan to conduct a study with lawyers and judges to certify whether the research questions are properly answered. In addition, we could develop a visual exploration system that could let users explore subsets of the extracted information, producing charts in real-time. Moreover, we could approach another legal domain of civil suits in addition to the ones covered by our systems at present. Furthermore, we could add other entities that could be useful to increase the understanding of Court decisions. In addition, we could evaluate the proposed systems for decisions from other Brazilian Courts. Finally, we could devise and test other Recurrent Neural Network architectures to increase the quality of our models.

As future work in the long term, we could increase the coverage of legal domains in order to apply our methods to a broad range of domains. Moreover, we could create models to identify the rationales of the Court decisions, which are subsets of the text that justifies the extracted entities, in order to produce explainable and consequently more trustworthy models.

Bibliography

[Aletras et al., 2016] ALETRAS, N.; TSARAPATSANIS, D.; PREOTIUC-PIETRO, D. ; LAMPOS, V.. **Predicting judicial decisions of the european court of human rights: a natural language processing perspective**. PeerJ Computer Science, 2016.

[Angelidis et al., 2018] ANGELIDIS, I.; CHALKIDIS, I. ; KOUBARAKIS, M.. **Named entity recognition, linking and generation for greek legislation**. In: JURIX, p. 1–10, 2018.

[Ashley, 2017] ASHLEY, K. D.. **Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age**. 2017.

[Bengio et al., 1994] BENGIO, Y.; SIMARD, P. ; FRASCONI, P.. **Learning long-term dependencies with gradient descent is difficult**. IEEE transactions on neural networks, 5(2):157–166, 1994.

[Boella et al., 2014] BOELLA, G.; DI CARO, L.; RUGGERI, A. ; ROBALDO, L.. **Learning from syntax generalizations for automatic semantic annotation**. Journal of Intelligent Information Systems, 43(2):231–246, 2014.

[Bojanowski et al., 2016] BOJANOWSKI, P.; GRAVE, E.; JOULIN, A. ; MIKOLOV, T.. **Enriching word vectors with subword information**. arXiv preprint arXiv:1607.04606, 2016.

[Brasil, 1981] BRASIL. **Lei 6.899, de 8 de abril de 1981**. Senado Federal, 1981. [Online; accessed 01/05/2020].

[Brasil, 2008] BRASIL. **Código civil brasileiro e legislação correlata**. Senado Federal, Subsecretaria de Edições Técnicas, 2 edition, 2008. [Online; accessed 01/05/2020].

[Brasil, 2015] BRASIL. **Código de processo civil**. Senado Federal, Secretaria de Editoração e Publicações, 2015. [Online; accessed 01/05/2020].

[Carletta, 1996] CARLETTA, J.. **Assessing agreement on classification tasks: the kappa statistic**. Computational linguistics, 22(2):249–254, 1996.

[Chalkidis et al., 2017a] CHALKIDIS, I.; ANDROUTSOPOULOS, I. ; MICHOS, A.. **Extracting contract elements**. In: PROCEEDINGS OF THE 16TH EDITION OF THE INTERNATIONAL CONFERENCE ON ARTICIAL INTELLIGENCE AND LAW, p. 19–28. ACM, 2017.

[Chiarcos, 2012] CHIARCOS, C.. **Powla: Modeling linguistic corpora in owl/dl**. In: EXTENDED SEMANTIC WEB CONFERENCE, p. 225–239. Springer, 2012.

[Cho et al., 2014] CHO, K.; VAN MERRIËNBOER, B.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H. ; BENGIO, Y.. **Learning phrase representations using rnn encoder-decoder for statistical machine translation**. arXiv preprint arXiv:1406.1078, 2014.

[Chung et al., 2014] CHUNG, J.; GULCEHRE, C.; CHO, K. ; BENGIO, Y.. **Empirical evaluation of gated recurrent neural networks on sequence modeling**. arXiv preprint arXiv:1412.3555, 2014.

[Cohn and Blunsom, 2005] COHN, T.; BLUNSOM, P.. **Semantic role labelling with tree conditional random fields**. In: PROCEEDINGS OF THE NINTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING, p. 169–172. Association for Computational Linguistics, 2005.

[Correia et al., 2019] CORREIA, F. A.; NUNES, J. L.; DE ALMEIDA, G. F. C. F.; ALMEIDA, A. A. A. ; LOPES, H.. **An exploratory analysis of precedent relevance in the brazilian supreme court rulings**. In: PROCEEDINGS OF THE ACM SYMPOSIUM ON DOCUMENT ENGINEERING 2019, DocEng '19, New York, NY, USA, 2019. Association for Computing Machinery.

[Curran et al., 2007] CURRAN, J. R.; CLARK, S. ; BOS, J.. **Linguistically motivated large-scale nlp with c&c and boxer**. In: PROCEEDINGS OF THE 45TH ANNUAL MEETING OF THE ACL ON INTERACTIVE POSTER AND DEMONSTRATION SESSIONS, p. 33–36. Association for Computational Linguistics, 2007.

[Dozat and Manning, 2017] DOZAT, T.; MANNING, C. D.. **Deep biaffine attention for neural dependency parsing**. In: PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 2017.

[Dragoni et al., 2016] DRAGONI, M.; VILLATA, S.; RIZZI, W. ; GOVERNATORI, G.. **Combining nlp approaches for rule extraction from legal documents**. In: 1ST WORKSHOP ON MINING AND REASONING WITH LEGAL TEXTS (MIREL 2016), 2016.

[Dyer et al., 2015] DYER, C.; BALLESTEROS, M.; LING, W.; MATTHEWS, A. ; SMITH, N. A.. **Transition-based dependency parsing with stack long short-term memory**. In: PROCEEDINGS OF THE 53RD ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS AND THE 7TH INTERNATIONAL JOINT CONFERENCE ON NATURAL LANGUAGE PROCESSING (VOLUME 1: LONG PAPERS), volumen 1, p. 334–343, 2015.

[Fernandes et al., 2019] FERNANDES, W. P. D.; SILVA, L. J. S.; FRAJHOF, I. Z.; DE ALMEIDA, G. D. F. C. F.; KONDER, C. N.; NASSER, R. B.; DE CARVALHO, G. R.; BARBOSA, S. D. J. ; LOPES, H. C. V.. **Appellate court modifications extraction for portuguese**. Artificial Intelligence and Law, Jul 2019.

[Fillmore, 1977] FILLMORE, C. J.. **Scenes-and-frames semantics**. Linguistic structures processing, 59:55–88, 1977.

[Garcia et al., 2017] GARCÍA-CONSTANTINO, M.; ATKINSON, K.; BOLLEGALA, D.; CHAPMAN, K.; COENEN, F.; ROBERTS, C. ; ROBSON, K.. **Cliel: context-based information extraction from commercial law documents**. In: PROCEEDINGS OF THE 16TH EDITION OF THE INTERNATIONAL CONFERENCE ON ARTICIAL INTELLIGENCE AND LAW, p. 79–87. ACM, 2017.

[Gianfelice et al., 2013] GIANFELICE, D.; LESMO, L.; PALMIRANI, M.; PERLO, D. ; RADICIONI, D. P.. **Modificatory provisions detection: a hybrid nlp approach**. In: PROCEEDINGS OF THE FOURTEENTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, p. 43–52. ACM, 2013.

[Gimpel et al., 2011] GIMPEL, K.; SCHNEIDER, N.; O'CONNOR, B.; DAS, D.; MILLS, D.; EISENSTEIN, J.; HEILMAN, M.; YOGATAMA, D.; FLANIGAN, J. ; SMITH, N. A.. **Part-of-speech tagging for twitter: Annotation, features, and experiments**. In: PROCEEDINGS OF THE 49TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES: SHORT PAPERS (VOLUME 2), p. 42–47. Association for Computational Linguistics, 2011.

[Gonçalves, 2018] GONÇALVES, C. R.. **Direito civil brasileiro v. 6–Responsabilidade civil**. Editora Saraiva, 2018.

[Graves and Schmidhuber, 2005] GRAVES, A.; SCHMIDHUBER, J.. **Framewise phoneme classification with bidirectional lstm and other neural network architectures**. Neural Networks, 18(5-6):602–610, 2005.

[Graves et al., 2013] GRAVES, A.; MOHAMED, A.-R. ; HINTON, G.. **Speech recognition with deep recurrent neural networks**. In: ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2013 IEEE INTERNATIONAL CONFERENCE ON, p. 6645–6649. IEEE, 2013.

[Grosman et al., 2020] GROSMAN, J. S.; FURTADO, P. H. T.; RODRIGUES, A. M. B.; SCHARDONG, G. G.; BARBOSA, S. D. J. ; LOPES, H. C. V.. **Eras: Improving the quality control in the annotation process for Natural Language Processing tasks**. Information Systems, 93:101553, Nov. 2020.

[Hartmann et al., 2017] HARTMANN, N.; FONSECA, E. R.; SHULBY, C.; TREVISO, M. V.; RODRIGUES, J. S. ; ALUÍSIO, S. M.. **Portuguese word embeddings: Evaluating on word analogies and natural language tasks**. Proceedings of the Brazilian Symposium in Information and Human Language Technology (STIL), p. 122–131, 2017.

[Hermann et al., 2015] HERMANN, K. M.; KOCISKY, T.; GREFENSTETTE, E.; ESPEHOLT, L.; KAY, W.; SULEYMAN, M. ; BLUNSOM, P.. **Teaching machines to read and comprehend**. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 1693–1701, 2015.

[Hochreiter and Schmidhuber, 1997] HOCHREITER, S.; SCHMIDHUBER, J.. **Long short-term memory**. Neural computation, 9(8):1735–1780, 1997.

[Huang et al., 2015] HUANG, Z.; XU, W. ; YU, K.. **Bidirectional lstm-crf models for sequence tagging**. arXiv preprint arXiv:1508.01991, 2015.

[Klein and Manning, 2003] KLEIN, D.; MANNING, C. D.. **Accurate unlexicalized parsing**. In: PROCEEDINGS OF THE 41ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2003.

[Lafferty et al., 2001] LAFFERTY, J.; MCCALLUM, A. ; PEREIRA, F. C.. **Conditional random fields: Probabilistic models for segmenting and labeling sequence data**. 2001.

[Ling et al., 2015] LING, W.; DYER, C.; BLACK, A. W.; TRANCOSO, I.; FERMANDEZ, R.; AMIR, S.; MARUJO, L. ; LUIS, T.. **Finding function in form: Compositional character models for open vocabulary word representation**. In: PROCEEDINGS OF THE 2015 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, p. 1520–1530, 2015.

[Ma and Hovy, 2016] MA, X.; HOVY, E.. **End-to-end sequence labeling via bi-directional lstm-cnns-crf**. In: PROCEEDINGS OF THE 54TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (VOLUME 1: LONG PAPERS), volumen 1, p. 1064–1074, 2016.

[McCallum and Li, 2003] MCCALLUM, A.; LI, W.. **Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons**. In: PROCEEDINGS OF THE SEVENTH CONFERENCE ON NATURAL LANGUAGE LEARNING AT HLT-NAACL 2003 (VOLUME 4), p. 188–191. Association for Computational Linguistics, 2003.

[Medvedeva et al., 2019] MEDVEDEVA, M.; VOLS, M. ; WIELING, M.. **Using machine learning to predict decisions of the european court of human rights**. Artificial Intelligence and Law, Jun 2019.

[Mikolov et al., 2013] MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G. S. ; DEAN, J.. **Distributed representations of words and phrases and their compositionality**. In: Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z. ; Weinberger, K. Q., editors, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 26, p. 3111–3119. 2013.

[Nanda et al., 2017b] NANDA, R.; SIRAGUSA, G.; DI CARO, L.; THEOBALD, M.; BOELLA, G.; ROBALDO, L. ; COSTAMAGNA, F.. **Concept recognition in european and national law**. In: JURIX, p. 193–198, 2017.

[Nguyen et al., 2018] NGUYEN, T.-S.; NGUYEN, L.-M.; TOJO, S.; SATOH, K. ; SHIMAZU, A.. **Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts**. Artificial Intelligence and Law, p. 1–31, 2018.

[Pennington et al., 2014] PENNINGTON, J.; SOCHER, R. ; MANNING, C. D.. **Glove: Global vectors for word representation**. In: EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), p. 1532–1543, 2014.

[Peters et al., 2017] PETERS, M. E.; AMMAR, W.; BHAGAVATULA, C. ; POWER, R.. **Semi-supervised sequence tagging with bidirectional language models**. In: PROCEEDINGS OF THE 55TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (VOLUME 1: LONG PAPERS), volumen 1, p. 1756–1765, 2017.

[Peters et al., 2018a] PETERS, M.; NEUMANN, M.; YIH, W.-T. ; ZETTLEMOYER, L.. **Dissecting contextual word embeddings: Architecture and representation**. empirical methods in natural language processing, p. 1499–1509, 2018.

[Peters et al., 2018b] PETERS, M. E.; NEUMANN, M.; IYYER, M.; GARDNER, M.; CLARK, C.; LEE, K. ; ZETTLEMOYER, L. S.. **Deep contextualized word representations**. In: NAACL HLT 2018: 16TH ANNUAL CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES, volumen 1, p. 2227–2237, 2018.

[Ramshaw and Marcus, 1995] RAMSHAW, L.; MARCUS, M.. **Text chunking using transformation-based learning**. In: PROCEEDINGS OF THE 3RD ACL WORKSHOP ON VERY LARGE CORPORA, p. 82–94, 1995.

[Schauer, 2009] SCHAUER, F.. **Thinking like a lawyer: a new introduction to legal reasoning**. Harvard University Press, Cambridge, MA, 2009.

[Sha and Pereira, 2003] SHA, F.; PEREIRA, F.. **Shallow parsing with conditional random fields**. In: PROCEEDINGS OF THE 2003 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS ON HUMAN LANGUAGE TECHNOLOGY (VOLUME 1), p. 134–141. Association for Computational Linguistics, 2003.

[Strubell et al., 2017] STRUBELL, E.; VERGA, P.; BELANGER, D. ; MCCALLUM, A.. **Fast and accurate entity recognition with iterated dilated convolutions**. In: PROCEEDINGS OF THE 2017 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, p. 2670–2680, 2017.

[Trompper and Winkels, 2016] TROMPPER, M.; WINKELS, R.. **Automatic assignment of section structure to texts of dutch court judgments**. In: JURIX, p. 167–172, 2016.

[Walton, 2019] WALTON, D.. **When expert opinion evidence goes wrong**. Artificial Intelligence and Law, 27(4):369–401, Dec 2019.

[Wen et al., 2015] WEN, T.-H.; GASIC, M.; MRKŠIĆ, N.; SU, P.-H.; VANDYKE, D. ; YOUNG, S.. **Semantically conditioned lstm-based natural language generation for spoken dialogue systems**. In: PROCEEDINGS OF THE 2015 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, p. 1711–1721. Association for Computational Linguistics, 2015.

[Yang et al., 2016] YANG, Z.; YANG, D.; DYER, C.; HE, X.; SMOLA, A. ; HOVY, E.. **Hierarchical attention networks for document classification**. In: PROCEEDINGS OF THE 2016 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES, p. 1480–1489, 2016.

[de Araujo et al., 2017] DE ARAUJO, D. A.; RIGO, S. J. ; BARBOSA, J. L. V.. **Ontology-based information extraction for juridical events with case studies in brazilian legal realm**. Artificial Intelligence and Law, 25(4):379–396, 2017.