Automatic identification of similar judicial precedents

Igor Stemler[1], Marcelo Ladeira[1], Thiago de Paulo Faleiros[1]

Universidade de Brasília, Brazil
igorstemler@gmail.com
mladeira@unb.br
thiagodepaulo@unb.br

Abstract. Brazilian Code of Civil Procedure has been reformulated in 2015 and created new institutes of judicial precedents to allow the Courts of Appeal to decide about similar cases based on one main case, which is considered the paradigm for similar cases that remain suspended. This mechanism aims to avoid legal uncertainty in the lower courts, but, uncertainty can be taken to the Courts of Appeal, since different courts can judge similar legal matter in the opposite way. The identification of similar judicial cases is hard because Courts of Appeal work independently and the number of cases is high. We propose the use of computational intelligence techniques to automatically identify similar judicial precedents. Our hypothesis is that algorithms based on semantic approaches, such as Latent Semantic Indexing and Latent Dirichlet Allocation, perform better than those that use only syntactic approach, as (Okapi) BM25 ranking function. The best-performing model is extended with named entities to verify if its performance increases. The performance of the models is evaluated using similarity metrics and with the assistance of a specialist. We test this approach with the database of judicial precedent of the National Council of Justice. Our approach correctly grouped more than 90% of judicial precedents and found similar precedents with divergent decisions or precedents that should be suspended due to the existence of appeals into superior courts of same subject matter. Models based on syntactic approach presented the best results, as it required lower computational cost and fewer parameter tuning compared to the others.

Categories and Subject Descriptors: I.5.3 [**Computing Methodologies**]: Pattern Recognition; I.2.6 [**Artificial Intelligence**]: Learning

Keywords: Judicial Precedents; Class Action; BM25; LSI; LDA

1. INTRODUCTION

The Brazilian Judiciary is composed of 91 courts and 3 councils. The Supreme Federal Court (STF) is responsible to maintain the Brazilian Federal Constitution. The Brazilian National Council of Justice (CNJ) is in charge of controlling the administrative and financial performance of the Judiciary. This Court has been transformed the Brazilian judiciary with the Justice 4.0 Program, that uses new technologies, including artificial intelligence, to make the Brazilian justice system more efficient and effective[UNDP and CNJ 2022]. According to the CNJ, the number of lawsuits in the Judiciary has increased gradually in recent years. It reached 77.3 million of lawsuits at the end of 2021 encompassed 15 thousand judicial units with around 18 thousand judges [CNJ 2022]. On average, each Brazilian judge pronounced 7 sentences per day in 2021. Compared to European countries, the productivity of Brazilian judges is the third largest of 31 countries, behind only Denmark and Austria [CNJ 2011].

A considerable number of lawsuits deal with the same legal matter and judges can present different solutions to similar cases, disregarding the principle of legal certainty. So, the system of mandatory precedents was created to reduce the number of divergent judgments, to reduce the caseload of the Judiciary and to reduce trial time of lawsuits. This mechanism requires that judges observe the judicial precedent before giving their judgments. Judicial precedents can be considered as legal decisions that serve as a base for the judgment of similar lawsuits subsequent to them.

The "General Repercussion" (RG) institute was created in 2004 and made it possible to suspend similar lawsuits until the judgment of a case took as a paradigm case by STF. In 2008, a procedure called "Repetitive Appeal" (RR) was created, by suspending cases with the same legal theses until the decision of the Superior Court of Justice (STJ). This institute was extended to the Higher Labor Court (TST) in 2014. The Code of Civil Procedure (CPC) was reformulated in 2015, adding two new hypotheses of suspension of lawsuits, the "Incidents of Resolution of Repetitive Claims" (IRDR) and the "Incident of Assumption of Competence" (IAC), which can be judged by 86 different Brazilian courts (Figure 1).

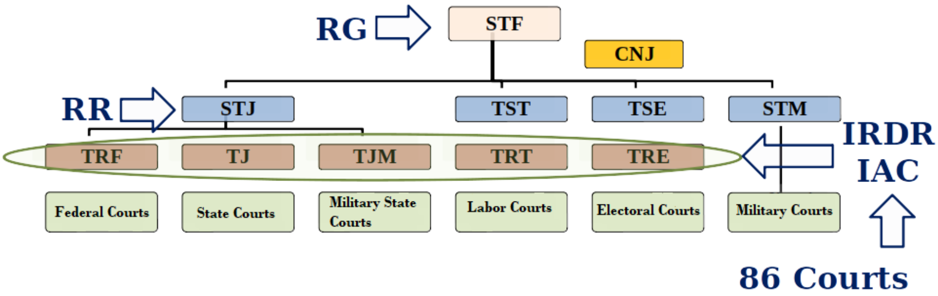

Fig. 1. Organization chart of the Judiciary

The advent of IRDR and IAC allow to the courts apply the same made decision in the main lawsuit to all other similar cases in the lower courts, thus reducing legal uncertainty. However, this insecurity can move to the appeal court, because different courts can judge similar cases in a contrary way. In order to contribute to solve this problem, we propose a software that can analyze the CNJ's precedent base and presents for each one its set of similar judicial precedents.

The main purpose of this research is to automatically identify similar judicial precedents with the use of text mining techniques. We also intend to evaluate the use of techniques based on semantic analysis to estimate similarity between precedents and to compare the results obtained by the proposed techniques. A tool that allows the visualization of similar judicial precedents represents a great contribution to Brazilian Judiciary, since the courts can identify issues already judged in other courts and can use them as reference in their decisions, or they can send the lawsuit to the higher courts, this way standardizing the understandings.

This article is structure as follow. Section 2 presents the main related work on judicial precedents and techniques for identifying similar texts. Section 3 describes the methodology used. Section 4 presents the results obtained with the application of text mining techniques. Finally, the last section brings the conclusions obtained and suggests the accomplishment of future works.

2. RELATED WORK

The articles about IRDR and IAC aim mainly to discuss the main procedural issues arising from these new institutes, such as the standardization of the judgements, legal certainty and procedural speed-up. Thus, these new institutes can help to reduce the time of processing of the lawsuits and to minimize the discrepant number of divergent decisions on the same issue of law in the Brazilian Courts[Antonio Pereira Gaio Júnior 2013][Durço 2016] [de Jesus Silva 2014].

According to Artur Souza[Artur César de Souza 2015], other countries also have mechanisms to deal with repetitive claims, and it is possible to identify similar institutes in US law (class action), Austrian law (testprozess), Danish law (class action), German law (Musterprozessfuhrung), Portuguese, Canadian and Israeli law. However, we have not found, in the specialized literature, articles that apply text mining techniques in the identification of similar judicial precedents or that relate these new institutes to legal uncertainty in the appeal courts.

The identification of similar precedents is considered a form of information retrieval, since it is desired to automatically extract information associated with these data that are unstructured in nature [Manning et al. 2008]. We use the measure of similarity of precedents to build clusters. Osei-Bryson[Osei-Bryson 2010] illustrates how a generic process model of data mining can be adapted to perform clustering analysis. According to Andreas Hotho[Hotho et al. 2005], there are in general two ways of evaluating the results of clustering. The first one is through the use of statistical measures. The second one is with the classification given by a specialist (Judicial Analyst) that can be considered a gold standard. With a gold standard is possible to calculate the precision, recall, and F-measure.

Sinoara[Sinoara et al. 2017][Sinoara et al. 2019] carried out extensive mapping of the literature on text mining studies related to semantics and identified that the most used method is Latent Semantic Indexing (LSI[Deerwester et al. 1990]). Another technique that is commonly used for topic modeling is Latent Dirichlet Allocation (LDA[Blei et al. 2003]). Chandrasekaran[Chandrasekaran and Mago 2021] traces the evolution of Semantic Similarity. Along the same lines, [Allahyari et al. 2017] conducted a survey of text mining techniques based on classification, clustering, and extraction techniques. According to this study, the most common way to represent the document is with a Bag of Word (BOW), considering the number of occurrences of each term, regardless of the order of occurrence.

This literature review was based on the Theory of the Consolidated Meta-Analytic Approach (TEMAC)[Melo and Br 2017] method. A search containing the term “document similarity” was performed on the Scopus database and returned 645 published studies in the field of computer science. Figure 2 shows the main groups identified by correlating the keywords put by the authors in their articles. There are similarities between this work with the group of studies that use Natural Language Processing in conjunction with the LSI and LDA techniques. The cosine similarity is still widely used to identify similar documents, so, these techniques were used in this article.

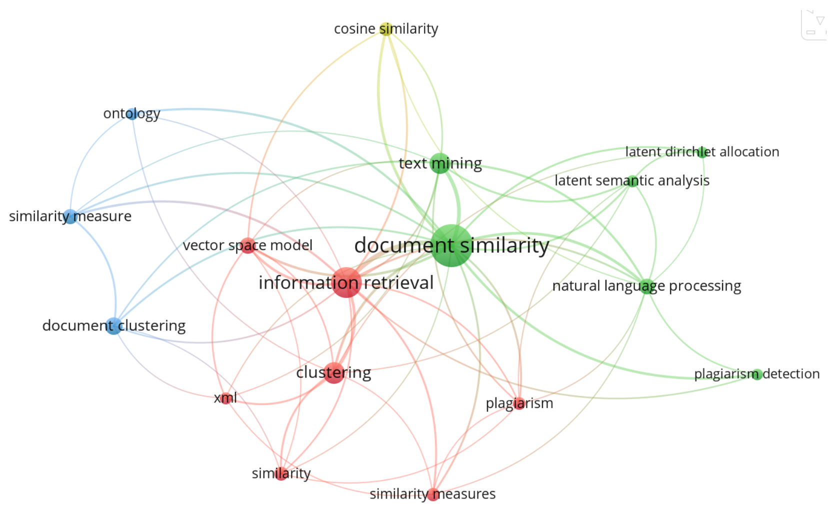

Fig. 2. Co-occurrence of keywords in articles

3. METHODOLOGY

The database used in this study is public and is available on the CNJ portal[1], forming the database of the National Data of Repetitive Claims and Precedents Required (BNPR).

The BNPR database has information about lawsuits, with 2472 precedents and more than 2.2 million lawsuits awaiting the judgment of the paradigm processes. The database used has information regarding:

—number of precedent, court, and type of institute: they form the unique theme number (NUT);

—subject of the process: established by the CNJ Unified Procedural Tables (TPU);

—situation of the precedent: field used in the graph to verify that the judicial precedent has been admitted, discontinued (suspended) or has been judged.

—precedent description: this is the main information for the identification of similar precedents. The terms of this field are used as attributes of the model.

—thesis: description of the judgment of the precedent. It is not used in the model to identify similar precedents, but it is used to verify similar precedents with divergent theses.

The data were preprocessed in order to construct a more consistent database, by deleting records of canceled precedents, rejected or without precedent description, thus, remaining 2260 precedents. The corpus of the description of the precedents was generated and then the tokenization, and removal of scores, symbols, stopwords and words with less than 3 letters were carried on. Finally, it was generated the matrix of terms by precedents.

There are several ways of weighing the terms and documents. It is usually based on the *tf-idf* function. One of the main weighting schemes is named Okapi BM25, which takes into account the size of documents. The terms were weighted according to the ordering function BM25 with standard parameters of k = 1.2 and b = 0.75.

The BM25, LSI, and LSI-weighted with BM25 and LDA techniques were then applied and the cosine similarity was used to identify similar precedents using the cosine calculation of the angle between the vector representation of two documents. Similar documents have values close to 1. Since for each document a vector containing the similarity value of it to the others, a threshold of the cosine value is set to determine which precedents should be effectively considered as similar. Precedents that present values higher than the threshold are considered similar.

The groups formed by similar precedents were presented to the specialist who checked if the precedents have been correctly grouped. The specialist also labels each group with an annotation regarding its correcting. Based on the labeled data, we can calculate precision, recall, and the F-measure. The models were then evaluated to verify which presented the best results. The evaluation was carried out experimentally, using similarity metrics and the assistance of the specialist.

The F-Measure metric was used as support to define the threshold of the groups to be labeled by the specialist, to compare the models, and to define the number of dimensions of the models that use dimensional reduction.

Finally, similar judicial precedents are visualized by graphs. The R software was used to perform the text mining analyses [Meyer et al. 2008]. The Quanteda (Quantitative Analysis of Textual Data)[Benoit et al. 2018] tool was added to R to perform the analysis faster and more efficiently.

[1] Available in: `https://www.cnj.jus.br/pesquisa-judiciarias/demandas-repetitivas`

4. RESULTS

After the pre-processing phase, with corpus construction, tokenization, stopwords removal, normalization, and attribute selection, a matrix precedents-attributes of size 2,260 by 3,370 was generated. The elements of this matrix were weighted by the rank function BM25 and the cosine similarity function was used in the weighted matrix to generate a similarity matrix between precedents.

Analyzing the frequency of similarity values obtained in each pair of precedents, the value 0.64 was chosen as an initial threshold. With this threshold, 27 judicial precedents were clustered and sent to the specialist to label each group as correct or not correct. In addition, the precedents identified by regular expressions similar to those previously grouped, but not included in the initial grouping, were also labeled. After this labeling, all precedents identified in the initial clustering were grouped correctly, resulting in the following metrics: precision of 100%, recall of 73%, and F-measure of 83%. As only 27 judicial precedents were grouped it was necessary to choose a new threshold better than 0.64.

Figure 3 presents the metrics F-measure, precision, recall, and percentage of precedents (of the BNPR base that remained unlabeled) for each evaluated threshold. After initial labeling, almost all the initially labeled documents were correctly grouped at threshold 0.4, generating 95% of recall, however, 73% of grouped precedents were not labeled at that point. This new threshold clustered 132 precedents and the results obtained were again referred to the specialist. With the new labeling, the total correct groupings went from 11 to 32 out of a total of 50 identified groups. As other similar groups were verified, we chose to label the 287 precedents grouped considering the threshold point of 0.31. These labeled precedents represented 12.7% of the total documents in the analyzed database, containing 51.2% of total IAC, 38.5% of total IRDR, 8.5% of total RG, and 5.1% of total RR.

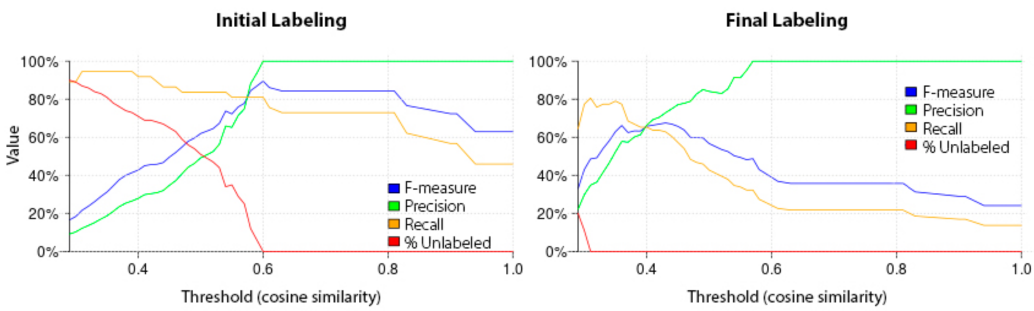

Fig. 3. Evaluation metrics by similarity threshold after initial and final labeling Lexical approach with BOW weighted by the function BM25

Although more than double the number of precedents presented at the threshold point of 0.39, only 7 new groups from a total of 79 identified groups were identified. That is, this model does not help in labeling new precedents using 287 pooled documents, because many precedents were grouped together although they were different.

4.1 Semantic Approaches

The LSI technique was then applied to both precedents matrices. In order to identify the optimal dimension (k) value of each model, the cosine similarity function was used in the low-rank matrices generated for k ranging from 50 dimensions to 1,000. The best result was F-measure of 54,2% with 500 dimensions for the LSI model using only the frequency of terms.

The Latent Dirichlet Allocation (LDA) model is a mixed Bayesian model for discrete variables, so it was not possible to weight by BM25 function. As with the LSI model, it was necessary to choose the number of dimensions/topics for the model. The cosine similarity function was used in the matrices of precedents by topics, with k varying every 50 topics up to 500. With 500 topics there wasn't improvement and the processing time increased considerably. The results were generated for the model that uses the Gibbs sampling algorithm. In addition, the standard values of *50/factors* for α and of 0.1 for β were used as a priori distribution parameters. The 150-topic model presented the highest F-measure.

The F-measure metric was used as a way of comparing the generated models against the number of grouped precedents. Thus, it was not necessary to choose a similarity threshold arbitrarily, since, given that the specialist had identified 136 similar judicial precedents, it was sufficient to select the largest F-measure model that groups quantitative precedents close to that number.

The LSI model weighted by BM25 presented the largest F-measure when grouping 100 legal precedents, remaining with an indicator above the other models until the grouping of the previous 140 (Figure 4). It is emphasized, however, that the value of the F-measure changes according to the number of dimensions chosen. The model BM25 is already much simpler and presents F-measures very close to those obtained by LSI, even with higher metrics when more than 140 precedents were grouped.

The LDA algorithm presented the worst results. The model may not have yielded good results due to the fact that most terms in judicial precedents were not attributable to a single topic. In addition, recent studies show that the LDA model is not consistent when applied to short texts[Hajjem and Latiri 2017].

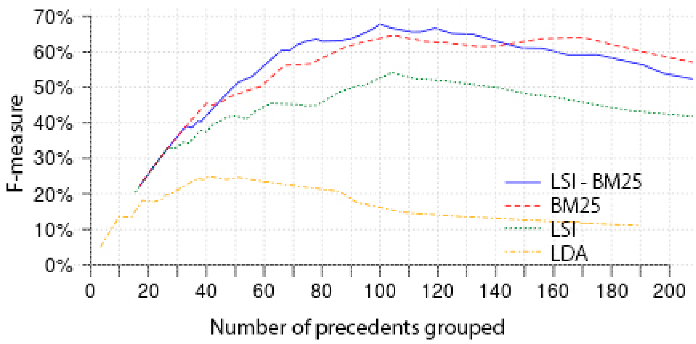

Fig. 4. F-measure using BM25, LSI, LSI with BM25 and LDA

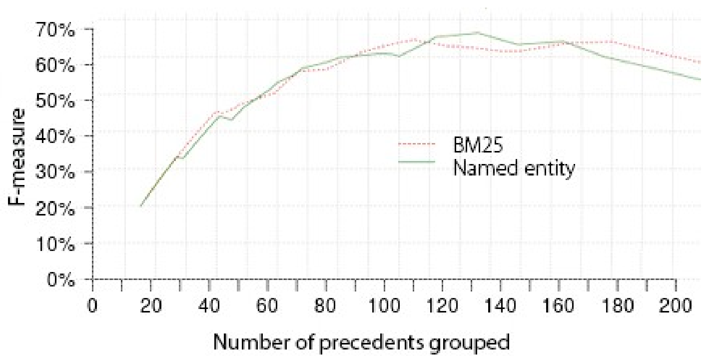

Fig. 5. F-measure using BM25, and BM25 with named entity

4.2 Named Entities

In order to verify if there were gains when using named entities, the BM25 model was induced with legal articles and norms extracted from the descriptions of judicial precedents, STF thesaurus, and named entities identified as proper names. Thus, 514 norms, 1,566 proper names, and 2,854 words composed in the STF thesaurus were identified in the descriptions of judicial precedents. However, only the STF thesaurus and the extracted norms presented better results than the BM25 model. Inducing the BM25 model with these entities, there was a small gain since the BM25 model showed a maximum F-measure of 64.7% when grouping previous ones, while the model with named entities presented F-measure of 66%, however, grouping over 65 precedents (Figure 5).

A graph was used to represent the clusters, as verified in the final model of Figure 6, which uses BM25 inducted with named entities. Each color in the graph represents a type of judicial precedent: red RG, orange RR, blue IRDR and green IAC. Each symbol represents a situation of precedent: ellipse indicates precedent already judged, star indicates preceded discontinued, and circle indicates precedent admitted.

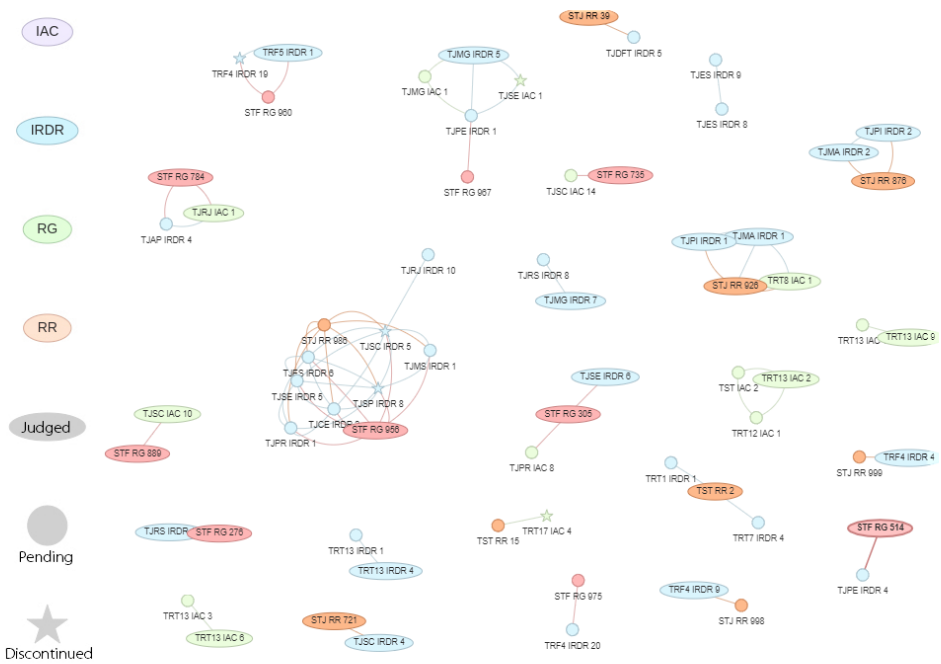

Fig. 6. Graph of similar judicial precedents

The CPC provides instruments for not admitting or suspending precedents that deal with a matter with appeal in the higher courts, however, few precedents are suspended despite the similarity with themes of RR and RG. These institutes are present in three out of four identified clusters. Of the 399 IRDRs and IACs in the database, 56 (14%) contain precedents similar to those received by STF and STJ. We observe in Figure 6 that a precedent of "the legality of the calculation base of the tax of ICMS on the tariff for the use of the transmission and distribution of electricity" is present in eight Courts of Justice and still with resources for the STF and for the STJ. The STF considered that this issue does not have general repercussions, however, the STJ recognized the repetitive appeal and informed the courts that all precedents related to this subject should be suspended, however, only two courts present this suspended topic.

5. CONCLUSION

The performance of the models proposed give us experimental evidence that our approach to automatic identify similar judicial precedents is feasible. There is no significant experimental evidence that algorithms (LSI and LDA) based on semantic approaches perform better than those based on syntactic approach (BM25) in the domain of legal matter.

The LSI model weighted by BM25 presented good results in identifier similar judicial precedents, since it obtained F-measure nearly 70% and, with the help of a specialist, it was possible to group together 136 similar judicial precedents. This model presented better results than BM25, LSI and LDA. However, the LSI model depends on the choice of the number of dimensions, and it is necessary to verify the number of dimensions with the entry of new precedents. As the model BM25 demands less processing and presented results close to the LSI, this model was chosen to verify if there were gains when inducing the model with named entities. The model with named entities presented better results than the model BM25, but with F-measure very close to the BM25 model.

In future work, we want to improve the automatic identification of named entities and keywords based on descriptions of judicial precedents. In addition, the tool will be adapted to the courts of the first instance so that the similarity of the process can be verified in relation to the existing precedents according to the content of the initial petition at the entrance of the process.

REFERENCES

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., and Kochut, K. A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*, 2017.

Antonio Pereira Gaio Júnior. Incidente de Resolução de Demandas Repetitivas no projeto do novo CPC: Breves apontamentos. *Legis Augustus* 4 (2): 1–11, 2013.

Artur César de Souza. *Resolução de demandas repetitivas: Comunicação de demanda individual incidente de resolução de demandas repetitivas recursos repetitivos*. Novo Processo Civil Brasileiro. Almedina Brasil, Brasil, 2015.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., and Matsuo, A. Quanteda: An r package for the quantitative analysis of textual data software. *J. Open Source Software* 3 (30), 2018.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of machine Learning research* vol. 3, pp. 993–1022, 2003.

Chandrasekaran, D. and Mago, V. Evolution of semanticsimilarity – a survey. *ACM Computing Surveys* vol. 54, pp. 1–37, 2021.

CNJ. Estudo comparado sobre recursos, litigiosidade e produtividade: a prestação jurisdicional no contexto internacional. Tech. rep., Conselho Nacional de Justiça, 2011.

CNJ. Justiça em Números 2022. Tech. rep., Conselho Nacional de Justiça, 2022.

de Jesus Silva, J. *O Incidente de Resolução de Demandas Repetitivas no projeto do novo Código de Processo Civil: Segurança Jurídica e legitimidade democrática das decisões judiciais no Estado Constitucional de direito*. Ph.D. thesis, Universidade de Brasília, 2014.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. Indexing by latent semantic analysis. *Journal of the American society for information science* vol. 41, pp. 391, 1990.

Durço, K. A. O Incidente de Resolução de Demandas Repetitivas: Uma das propostas centrais do projeto de novo Código de Processo Civil. *Revista Eletronica de Direito Processual* 8 (8), 2016.

Hajjem, M. and Latiri, C. Combining IR and LDA Topic Modeling for Filtering Microblogs. *Procedia Computer Science* vol. 112, pp. 761–770, 2017.

Hotho, A., Nürnberger, A., and Paass, G. A brief survey of text mining. *Journal for Language Technology and Computational Linguistics* vol. 20, pp. 19–62, 2005.

Manning, C. D., Raghavan, P., and Shutze, H. *Introduction to information retrieval.* Cambridge University Press, 2008.

Melo, A. and Br, M. A. *Revisão da Literatura: Apresentação de uma Abordagem Integradora.* AEDEM International Conference, 2017.

Meyer, D., Hornik, K., and Feinerer, I. Text mining infrastructure in r. *Journal of statistical software* 25 (5): 1–54, 2008.

Osei-Bryson, K.-M. Towards supporting expert evaluation of clustering results using a data mining process model. *Information Sciences* 180 (3): 414–431, 2010.

SINOARA, R. A., ANTUNES, J., AND REZENDE, S. O. Text mining and semantics: a systematic mapping study Journal of the Brazilian Computer Society Text mining and semantics: a systematic mapping study. *Journal of the Brazilian Computer Society* vol. 23, 2017.

SINOARA, R. A., ANTUNES, J., AND REZENDE, S. O. Knowledge-enhanced document embeddings for text classification. *Knowledge-Based Systems* vol. 163, pp. 955–971, 2019.

UNDP AND CNJ. Justice 4.0 program - understand how the brazilian justice system is using digital transformation to promote innovation, become more efficient and increase access to justice for all. Tech. rep., United Nations development Programme – Brazil and Conselho Nacional de Justiça, 2022.