

CDJUR-BR - Uma Coleção Dourada do Judiciário Brasileiro com Entidades Nomeadas Refinadas

Maurício Brito¹, Vládia Pinheiro¹, Vasco Furtado^{1,4},
João Araújo Monteiro Neto¹, Francisco das Chagas Jucá Bomfim¹,
André Câmara Ferreira da Costa^{1,3}, Raquel Silveira²

¹Programa de Pós-Graduação em Informática Aplicada – Universidade de Fortaleza
Fortaleza – CE – Brazil

²Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Fortaleza – CE – Brazil

³Centro Universitário Christus
Fortaleza – CE – Brazil

⁴Empresa de Tecnologia da Informação do Ceará (ETICE)
Fortaleza – CE – Brazil

mauriciobrito@edu.unifor.br, {vladiacelia,vasco,joaoneto,fjuca}@unifor.br
andrecamarafc@gmail.com, raquel.silveira@ifce.edu.br

Abstract. This article presents the development of the Golden Collection of the Brazilian Judiciary (CDJUR-BR), a corpus formed by 21 fine-grained annotations of legal entities in various legal documents. CDJUR-BR aims to provide a comprehensive and robust corpus for REN, consisting of 44,526 annotations. Furthermore, a model for NER based on BERT was developed which achieved an average F1-macro of 0.58. These results indicated the importance and usefulness of the CDJUR-BR.¹

Resumo. Este artigo apresenta o desenvolvimento da Coleção Dourada do Judiciário Brasileiro (CDJUR-BR), um corpus formado por 21 entidades específicas anotadas em documentos jurídicos. A CDJUR-BR visa fornecer um corpus abrangente e robusto para REN, composto por 44.526 anotações. Além disso, foi desenvolvido um modelo para REN baseado no BERT que alcançou a F1-macro média de 0,58. Estes resultados indicaram a importância e a utilidade da CDJUR-BR.¹

1. Introdução

Uma tarefa básica para a maioria das aplicações em Inteligência Artificial Jurídica (Legal AI, em inglês) é a de Reconhecimento de Entidades Nomeadas (REN), que busca identificar e qualificar trechos de textos que se referem a entidades como pessoas, locais, organizações, datas, dentre outras, agregando ao texto informação semântica [Yadav and Bethard 2019, Schmitt et al. 2019, Li et al. 2020]. Os modelos para REN

¹**Licença:** Este trabalho foi licenciado com a licença Creative Commons CC0 1.0 Universal (CC0 1.0) - Domínio público dedicado.

são treinados a partir de coleções de textos rotulados por humanos especialistas, comumente denominadas de Coleção Dourada [Schmitt et al. 2019, Jiang et al. 2016, Atdağ and Labatut 2013].

No domínio jurídico, os documentos legais fazem referências a certas entidades que não são trivialmente reconhecidas pelos RENs disponíveis atualmente, mas que são necessárias a aplicações em Legal AI. Por exemplo, uma determinada legislação citada em um documento jurídico, trata-se de uma *norma principal* que define o assunto da petição? Ou se trata de uma *norma acessória* servindo somente para apoiar os argumentos do peticionante? Outro exemplo seria o de identificar pessoas específicas em uma peça processual: é a vítima? o réu? o juiz? As respostas a essas perguntas são de natureza interpretativa e requer do anotador conhecimento técnico para fornecê-las. Além disso, existem poucos exemplos de coleções douradas adequadas para o domínio da justiça brasileira [de Araujo et al. 2018, Leitner et al. 2020], e somente entidades básicas rotuladas. Há carência de categorizações de legislação, provas, penalidades, pessoas específicas em um processo jurídico (juiz, advogado, vítima, réu, testemunha), tipos de locais (local do crime, endereço do réu), etc. Neste sentido, persiste ainda a necessidade de uma coleção robusta, anotada com entidades refinadas do domínio jurídico, e que abranja diversos documentos de um processo legal, como petições, inquéritos, denúncias, decisões e sentenças.

Diante desta problemática, as seguintes questões de pesquisa nortearam o desenvolvimento deste trabalho: *QP1* - Como elaborar uma metodologia de anotações manuais de entidades nomeadas que contemple as especificidades e complexidades do domínio jurídico? e *QP2* - A coleção dourada gerada é adequada para o treinamento e validação de modelos de *Legal AI*?

Neste artigo, descrevemos o desenvolvimento da Coleção Dourada do Judiciário Brasileiro (CDJUR-BR) contemplando um conjunto de entidades anotadas manualmente por especialistas em documentos jurídicos. A criação da CDJUR-BR seguiu uma metodologia própria que visou atribuir o caráter de abrangência e robustez à coleção contendo 21 entidades refinadas, as quais possam servir ao processo de treinamento e validação de modelos de *Legal AI* para língua portuguesa. Especificamente para a rotulação das normas legais e seus artigos e dispositivos foram aplicadas etapas adicionais de refinamento e validação, pois estas entidades são consideradas essenciais a uma série de aplicações. A avaliação experimental visou demonstrar que a CDJUR-BR suporta o desenvolvimento de REN eficazes em documentos legais. Além deste importante recurso linguístico para área de IA Jurídica no Brasil, foi desenvolvido um REN baseado no modelo BERT e treinado com a CDJUR-BR, cujo desempenho alcançou a F1-macro média de 0,58, indicando as vantagens e prevalência do uso da CDJUR-BR.

2. Trabalhos Relacionados

A prática de criar coleções douradas para o contexto jurídico tem na Europa seus maiores exemplos [Leitner et al. 2020, Angelidis et al. 2018]. Para o domínio jurídico em português, os trabalhos desenvolvidos são fragmentados, com poucas classes específicas e de tamanho limitado, o que é um obstáculo para o desenvolvimento de classificadores REN baseados em dados. No trabalho pioneiro realizado por [de Araujo et al. 2018], os autores disponibilizaram a LeNER-Br, construída a partir de anotações manuais de 66 do-

cumentos jurídicos de diversos tribunais brasileiros, e 04 documentos legislativos, como a Lei Maria da Penha, totalizando 70 documentos anotados. As entidades categorizadas foram as genéricas: "ORGANIZACAO", "PESSOA", "TEMPO", "LOCAL", e duas específicas para o domínio jurídico - "LEGISLACAO" e "JURISPRUDENCIA". Ao todo, foram feitas 12.248 anotações de EN. O trabalho não cita as atividades realizadas durante o processo de anotação e nem se foram realizadas avaliações de concordância entre anotadores, porém relata que obteve F1-Score geral de 92,53%. Para as entidades específicas do domínio jurídico, obteve F1-Score de 97,00% e 88,82% para entidades de "LEGISLACAO" e "JURISPRUDENCIA", respectivamente.

A base UlyssesNER-Br [Albuquerque et al. 2022] foi criada a partir de 154 projetos de lei e 800 consultas legislativas da Câmara dos Deputados do Brasil, contendo dezoito tipos de entidades estruturadas em sete classes ou categorias semânticas. Baseadas no HAREM [Santos and Cardoso 2006] foram definidas 5 classes típicas: pessoa, localização, organização, evento e data; e duas classes semânticas específicas para o domínio legislativo: fundamento do direito (leis, resoluções, decretos, projetos de lei, econsultas legislativas) e produto do direito (sistemas, programas e outros produtos criados a partir da legislação). Os autores relatam que o processo de anotação ocorreu em três etapas - treinamento prático dos anotadores, anotação e avaliação, onde as anotações foram avaliadas com média kappa = 90%. Não há detalhes da quantidade de anotações realizadas. Os modelos Hidden Markov Model (HMM) e Conditional Random Fields (CRF) foram usados para avaliar o corpus, com desempenho na tarefa de REN de F1-Score = 80,8% (por categorias) e = 81,04% (por tipos).

3. Metodologia de Construção da CDJUR-BR

A metodologia de construção da CDJUR-BR consiste das etapas descritas a seguir (Figura 1). Um comitê com três professores da área do Direito e dois da Computação foi formado com o objetivo de definir, juntamente com os especialistas do domínio, os principais parâmetros da CDJUR-BR bem como zelar pela adequada aplicação da metodologia.

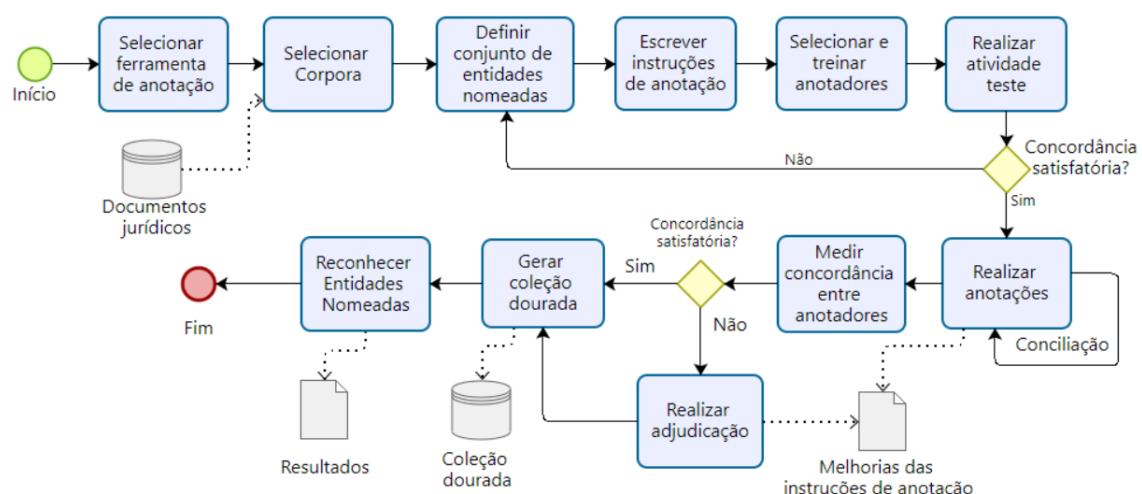


Figura 1. Fluxo geral da Metodologia de Construção da CDJUR-BR.

Seleção da Ferramenta de Anotação - Inicialmente, foi realizada uma pesquisa de mercado objetivando identificar ferramentas de anotação existentes e selecionar

aquelas que melhor atendesse aos seguintes critérios: Tipo de Dado; Perfis de Usuário; Fases do Processo de Anotação; Anotação Automática; Formato do arquivo de saída; Colaboração; Custo; Suporte Técnico. Ao final da avaliação, foram selecionadas as ferramentas Annotation Lab², Inception [Klie et al. 2018] e Tagtog [Cejuela et al. 2014]. Neste projeto, foi adotada a ferramenta Tagtog por facilidade de iniciação das atividades e disponibilidade de suporte técnico.

Seleção dos Documentos do *Corpus* - Para seleção dos documentos que compõem a CDJUR-BR, foram usados critérios de representatividade e qualidade. Quanto à representatividade, foram selecionados documentos das classes-CNJ [Silva et al. 2013] mais representativas dos processos de primeiro grau, encerrados em 2019 no TJCE (85% dos processos de primeiro grau são das classes selecionadas). Quanto à qualidade, foram filtrados os documentos com 80% ou mais de palavras válidas da língua portuguesa e com mais de 50 tokens. Em relação aos tipos de documentos, a equipe de especialistas selecionou: Petição Inicial, Petição, Denúncia, Inquérito, Decisão, Sentença, Despacho e Alegações Finais. A seleção final foi randômica dentre um conjunto de 80 mil documentos dos arquivos do TJCE que atenderam aos critérios acima explanados. A Tabela 1 apresenta a quantidade de documentos que compõem o *corpus* a ser anotado para a CDJUR-BR, por tipo de documento e classe, totalizando 1.216 documentos.

Tabela 1. Composição do *Corpus* a ser anotado.

Tipo de Documento	Proc. Comum Cível	Juizado Especial Cível	Execução Fiscal	Execução Extra-judicial	Inquérito Policial	Ação Penal	Outras
Petição Inicial	35	30	35	33	0	16	29
Petição	20	18	20	19	0	33	57
Denúncia	0	0	0	0	12	85	35
Inquérito	0	0	0	0	53	54	31
Decisão	21	21	20	0	22	71	32
Sentença	20	20	18	20	22	30	48
Despacho	11	12	13	30	2	6	22
Alegações Finais	0	0	0	0	2	115	23
Total por Classe	107	101	106	102	113	410	277
Total TJCE							1.216

Definição das Entidades Nomeadas a serem anotadas - As categorias de entidades mais comuns são pessoas, organizações, normas e localizações [Mikheev et al. 1999]. Porém, especialistas do TJCE e do Ministério Público do Ceará (MPCE) definiram um conjunto de 21 entidades mais refinadas em cada categoria. A motivação é a melhoria de tarefas e aplicações de Legal AI para o judiciário brasileiro, tais como: sumarização de textos, similaridade de processos, classificação de assunto, sugestão de penas e consulta de jurisprudência. A relação completa das entidades por categoria é apresentada na Tabela 3. Por exemplo, a categoria *Pessoa* foi especificada em 9 entidades que normalmente estão presentes em um processo judicial, quais sejam: autor, advogado, réu, vítima, testemunha, juiz, promotor, autoridade policial e outras. Os *Endereços* foram especificados em 6 entidades para identificar os diversos endereços presentes em um processo judicial. A categoria *Norma* foi especificada em três entidades. A primeira descreve as referências legais diretamente relacionadas ao assunto principal do processo. A segunda refere-se às normas que contextualizam o documento jurídico em questão (e.g. petições, decisões,

²<https://www.johnsnowlabs.com/annotation-lab/>

etc.). A terceira especifica as decisões de jurisprudências mencionadas nos processos. De forma similar, foram definidas entidades para i) *Prova*, especificando as provas usadas pela acusação e defesa durante o processo; ii) *Pena*, para identificar as sanções aplicadas; iii) *Sentença*, para identificar as sentenças proferidas pelos juízes.

InSTRUÇÕES DE ANOTaÇÕES, Seleção e Treinamento dos Anotadores - Uma vez definidas as entidades a serem anotadas, foi iniciada a elaboração das instruções de anotação, as quais foram aprimoradas continuamente, por meio de reuniões entre as equipes de anotadores e o comitê técnico e de gestão dos trabalhos. Os especialistas no domínio jurídico, de ambos os sexos, com experiência de 5 a 10 anos na área jurídica foram divididos em três equipes: Equipe 1 - 14 juízes estaduais; Equipe 2 - 19 promotores de justiça e técnicos do judiciário; Equipe 3 - 03 professores de direito, com nível de mestrado ou doutorado. Esta última trabalhou nas revisões da fase de adjudicação. Todos os anotadores foram treinados nas instruções de anotação, na metodologia e em exercícios práticos do software Tagtog.

Atividade Teste de Anotação - Nesta etapa, poucos documentos foram anotados e os anotadores praticaram as instruções de anotação e o uso do software *Tagtog*, além de interagirem com os demais discutindo as dúvidas em situações específicas. Cada anotador, rotulou pelo menos duas ocorrências de cada entidade. Ao final, houve revisão das instruções de anotação e da configuração final das entidades a serem rotuladas.

Processo de Anotação do Corpus - Cada documento foi anotado por dois anotadores diferentes. Os anotadores receberam o manual de instruções e, cada um deles teve liberdade para realizar seu trabalho na ordem que preferisse. A primeira etapa do processo de anotação ocorreu em 2,5 meses, quando todos os documentos haviam sido anotados por pelo menos dois anotadores. Depois dessa etapa, a CDJUR-BR entrou no ciclo de avaliação e refinamento.

Avaliação da Concordância entre Anotadores - A Tabela 2 mostra os resultados obtidos do coeficiente Kappa [McHugh 2012] para as categoria de entidades. A categoria Pessoa obteve Kappa 0,79, indicando a maior concordância entre os anotadores. O resultado é considerado "substancial" na interpretação do índice Kappa segundo Landis e Koch [Landis and Koch 1977]. Nesta mesma interpretação, a categoria Prova é considerada "moderada" e foi a que apresentou maior dissenso entre os anotadores.

Tabela 2. Resultados da Concordância entre Anotadores em cada Categoria

Categoria	Anotações	Kappa
Pessoa	15.149	0,79
Prova	1.696	0,46
Pena	205	0,64
Endereço	2.041	0,73
Sentença	106	0,75
Norma	6.216	0,76
Total	25.413	0,69

Processo de Adjudicação - Ao final da fase de anotações, 732 documentos obtiveram coeficiente Kappa médio de 0,69. Para 166 documentos com Kappa inferior a 0,50 foi adotada a estratégia de adjudicação em que um terceiro especialista revisa os casos de desacordo [Hovy and Lavid 2010]. Nestes casos, o revisor recebeu as anotações dos dois anotadores iniciais (união das anotações), e poderia acatar uma das anotações ou mesmo

realizar nova anotação diferente das iniciais. Além disso, no decorrer das anotações, alguns anotadores deixaram os trabalhos, e com isso, 176 documentos remanescentes foram anotados pela equipe de adjudicação.

Refinamento das Normas Legais - Dada a importância da categoria Normas para o domínio jurídico, a equipe de revisores recebeu um relatório com as normas identificadas pelo REN e analisaram se eram, de fato, normas e qual a entidade específica (principal, acessória ou uma jurisprudência). Esta etapa de melhoria e refinamento da CDJUR-BR propiciou a revisão das diretrizes de caráter jurídico destas entidades e dos limites das anotações de normas. Ao final dessa etapa, foram adicionadas 4.338 novas entidades de normas jurídicas. Por fim, a CDJUR-BR contém 1.074 documentos com 44.526 ocorrências de entidades rotuladas, conforme detalhado na Tabela 3.

Tabela 3. Estatística do corpus anotado.

Categoría	Anotações	%	Entidade Nomeada	Anotações	%
Pessoa	24.844	55,80	PES-ADVOG PES-AUTOR PES-AUTORID-POLICIAL PES-JUIZ PES-OUTROS PES-PROMOTOR-MP PES-REU PES-TESTEMUNHA PES-VITIMA	735 1.259 2.012 576 6.003 363 8.773 2.967 2.156	1,65 2,83 4,52 1,29 13,48 0,82 19,70 6,66 4,84
Prova	3.318	7,45	PROVA	3318	7,45
Pena	407	0,91	PENA	407	0,91
Endereço	2.065	4,64	END-AUTOR END-DELITO END-OUTROS END-REU END-TESTEMUNHA END-VITIMA	132 466 355 693 295 124	0,30 1,05 0,80 1,56 0,66 0,28
Sentença	172	0,39	SENTENCA	172	0,39
Norma	13.720	30,81	NOR-ACESSÓRIA NOR-JURISPRUDÊNCIA NOR-PRINCIPAL	5.767 1.823 6.130	12,95 4,09 13,77
Total	44.526	100	Total	44.526	100

4. Avaliação Experimental

A avaliação da CDJUR-BR foi guiada pelas Questões de Pesquisa definidas na introdução deste artigo. A *QP1* (Como elaborar uma metodologia de anotações manuais de entidades nomeadas que contemple as especificidades e complexidades do domínio jurídico?) foi respondida nas etapas de desenvolvimento das anotações, descritas na metodologia aqui apresentada, e validada pelas avaliações de concordância entre anotadores (Kappa geral de 0,69). A etapa de refinamento, com uso do REN treinado na CDJUR-BR, numa abordagem *humam-in-the-loop*, possibilitou a adição de 19.113 anotações à CDJUR-BR.

Para responder a *QP2* (A coleção dourada gerada é adequada para o treinamento e validação de modelos de *Legal AI*?), avaliamos a CDJUR-BR na tarefa de REN, através de alguns cenários de experimentos descritos a seguir. Em todos os cenários, os conjuntos de treino, validação e teste ficaram com 68%, 15% e 17% das amostras, respectivamente, observando-se a proporção por categoria no conjunto original.

- **C1. Reconhecimento das entidades específicas da CDJUR-BR**- os dados da CDJUR-BR foram usados para treinar REN específicas. Nossa propósito é avaliar a viabilidade da CDJUR-BR para o treinamento de modelos REN no domínio jurídico brasileiro, em língua portuguesa.

- **C2. Reconhecimento das categorias da CDJUR-BR**- o resultado do REN, treinado com a CDJUR-BR, agrupando as entidades específicas nas seguintes categorias: Pessoa, Legislação, Jurisprudência, e Local. O objetivo é uma análise comparativa com a base LENER-BR.
- **C3. Reconhecimento das entidades da LENER-BR**- neste cenário, os modelos REN foram treinados no conjunto da LENER-BR, em 6 diferentes entidades: Pessoa, Jurisprudência, Tempo, Local, Legislação e Organização.
- **C4. Reconhecimento das entidades da LENER-BR com modelo treinado na CDJUR-BR**- os modelos foram treinados com a CDJUR-BR e testados com os dados do LENER-BR, visando avaliar a capacidade de generalização do modelo REN treinado com a CDJUR-BR quando usado com outros documentos (no caso, os documentos que compuseram a LENER-BR).
- **C5. Reconhecimento das categorias de entidades da CDJUR-BR com modelo treinado na LENER-BR**- os modelos foram treinados com a LENER-BR e testados nas seguintes categorias da CDJUR-BR: Pessoa, Legislação, Jurisprudência e Local. Esse cenário visa avaliar quão capaz serão os modelos treinados com a LENER-BR em reconhecer entidades de outro *corpus* (no caso, os documentos que compuseram a CDJUR-BR).

A Tabela 4 apresenta os resultados obtidos no conjunto de teste para o C1, usando REN desenvolvidos com BI-LSTM + CRF [Graves and Schmidhuber 2005, Hochreiter and Schmidhuber 1997, Lafferty et al. 2001], SPACY [Honnibal and Montani 2017], e BERT [Devlin et al. 2018]. O modelo com o BERT, comparativamente, obteve o melhor desempenho na grande maioria das entidades, alcançando um F1-macro médio = 0,58. Das 21 entidades da CDJUR-BR, 09 (nove) alcançaram F1-Score $\geq 0,70$. Entretanto, 12 entidades tiveram F1-Score inferior a 0,70. Para alguns casos, como END-AUTOR, END-VÍTIMA e SENTENÇA, a causa foi o pequeno número de exemplos anotados para as entidades e os Falsos Positivos (FP) do tipo “O”, os quais representaram mais de 60% dos erros de predição. Para as entidades PES-ADVOG, PES-AUTOR, PES-OUTROS, PES-TESTEMNHA e PES-VÍTIMA, além dos FP com tokens do tipo “O”, também houve uma quantidade equivalente dos erros entre entidades da mesma categoria. Já para a entidade PROVA, a precisão foi excelente (0,87), com poucos FP. No entanto, a cobertura ficou baixa, em 0,33, com FN nas entidades Normas, Prova, Pessoa e, principalmente, nos tokens tipo “O”.

A Tabela 5 apresenta os resultados obtidos para os cenários de C2 a C5. Os experimentos destes cenários foram realizados com o modelo baseado no BERT, visto que este foi a melhor abordagem no cenário C1. Em uma comparação direta de C2 com C3, observamos que os resultados do REN treinado e testado com a CDJUR-BR são inferiores aos resultados do REN treinado e testado com a LENER-BR. A diversidade e quantidade de documentos jurídicos na CDJUR-BR, em relação a quantidade da base LENER-BR contribuiu para este desempenho inferior. Todavia, quando comparamos C4 (modelo treinado com a CDJUR-BR e testado com o LENER-BR) e C5 (modelo treinado com LENER-BR e testado com a CDJUR-BR), verifica-se que a CDJUR-BR obtém um desempenho superior (F1-Macro = 0,68 contra F1-Macro=0,56). Esse resultado indica que a CDJUR-BR tem maior capacidade de adaptabilidade para reconhecer entidades de outro *corpus* de documentos legais.

Tabela 4. Resultados da Tarefa de REN no Cenário C1, utilizando os modelos BI-LSTM+CRF, SPACY e BERT

Entidade Nomeada	BI-LSTM+CRF	SPACY	BERT	Suporte
END-AUTOR	0.56	0.31	0.33	18
END-DELITO	0.72	0.45	0.73	61
END-OUTROS	0.00	0.02	0.16	81
END-REU	0.55	0.59	0.71	152
END-TESTEMUNHA	0.27	0.26	0.67	68
END-VITIMA	0.06	0.00	0.22	27
NOR-ACESSÓRIA	0.79	0.79	0.82	990
NOR-JURISPRUDÊNCIA	0.90	0.87	0.89	333
NOR-PRINCIPAL	0.67	0.71	0.77	791
PENA	0.56	0.39	0.50	82
PES-ADVOG	0.54	0.22	0.63	122
PES-AUTOR	0.59	0.59	0.56	169
PES-AUTORID-POLICIAL	0.87	0.66	0.90	300
PES-JUIZ	0.79	0.50	0.78	83
PES-OUTROS	0.54	0.44	0.58	1.210
PES-PROMOTOR-MP	0.81	0.27	0.88	57
PES-REU	0.64	0.57	0.71	1.503
PES-TESTEMUNHA	0.57	0.45	0.64	519
PES-VÍTIMA	0.33	0.23	0.46	405
PROVA	0.47	0.29	0.34	461
SENTENÇA	0.00	0.29	0.00	11
F1-micro avg	0.64	0.55	0.67	7.443
F1-macro avg	0.53	0.42	0.58	7.443
F1-weighted avg	0.62	0.54	0.67	7.443

Tabela 5. Resultados de F1-score para o REN na CDJUR-BR e LENER-BR (C2, C3, C4 e C5) utilizando o modelo BERT

Entidade	Cenário de Experimento			
	C2	C3	C4	C5
JURISPRUDÊNCIA	0.89	0.96	0.79	0.48
LEGISLAÇÃO	0.92	0.97	0.92	0.86
LOCAL	0.77	0.77	0.32	0.15
PESSOA	0.83	0.97	0.69	0.76
F1-micro avg	0.85	0.96	0.81	0.60
F1-macro avg	0.85	0.92	0.68	0.56
F1-weighted avg	0.85	0.96	0.79	0.74

5. Conclusão

Neste artigo, apresenta-se a metodologia de anotação manual de documentos jurídicos, que serviu de base para a construção da CDJUR-BR, uma coleção dourada composta de mais de 1000 documentos de processos judiciais brasileiros, contendo 44.526 anotações de 21 entidades específicas das categorias Pessoa, Prova, Pena, Endereço, Sentença e Norma. A avaliação da concordância entre anotadores alcançou medida Kappa de 0,69 para 73% dos documentos, e os demais documentos passaram por revisões por especialistas e etapas de refinamento. Nos experimentos realizados com a CDJUR-BR, os resultados apontaram superioridade do modelo BERT com F1-macro média = 0,58, e testes comparativos entre CDJUR-BR e LENER-BR indicaram que o modelo REN treinado com a CDJUR-BR é superior em reconhecer entidades de outros corpus. Como trabalhos futuros, planeja-se melhorar a desambiguidade entre entidades e realizar mais anotações objetivando reduzir o desbalanceamento entre entidades. Além disso, desenvolver novos modelos REN a fim de selecionar aqueles de melhor desempenho para entidades específicas, realizando comparações com outros *corpus* no domínio jurídico.

Referências

- Albuquerque, H. O., Costa, R., Silvestre, G., Souza, E., da Silva, N. F., Vitorio, D., Moriyama, G., Martins, L., Soezima, L., Nunes, A., et al. (2022). Ulyssesner-br: a corpus of brazilian legislative documents for named entity recognition. In *International Conference on Computational Processing of the Portuguese Language*, pages 3–14. Springer.
- Angelidis, I., Chalkidis, I., and Koubarakis, M. (2018). Named entity recognition, linking and generation for greek legislation. In *JURIX*, pages 1–10.
- Atdag, S. and Labatut, V. (2013). A comparison of named entity recognition tools applied to biographical texts. In *2nd International conference on systems and computer science*, pages 228–233. IEEE.
- Cejuela, J. M., McQuilton, P., Ponting, L., Marygold, S. J., Stefancsik, R., Millburn, G. H., Rost, B., Consortium, F., et al. (2014). tagtog: interactive and text-mining-assisted annotation of gene mentions in plos full-text articles. *Database*, 2014.
- de Araujo, P. H. L., de Campos, T. E., de Oliveira, R. R., Stauffer, M., Couto, S., and Bermejo, P. (2018). Lener-br: a dataset for named entity recognition in brazilian legal text. In *International Conference on Computational Processing of the Portuguese Language*, pages 313–323. Springer.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Honnibal, M. and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.
- Hovy, E. and Lavid, J. (2010). Towards a ‘science’ of corpus annotation: a new methodological challenge for corpus linguistics. *International journal of translation*, 22(1):13–36.
- Jiang, R., Banchs, R. E., and Li, H. (2016). Evaluating and combining name entity recognition systems. In *Proceedings of the Sixth Named Entity Workshop*, pages 21–27.
- Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E., and Gurevych, I. (2018). The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9.
- Lafferty, J., McCallum, A., and Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Landis, J. R. and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, pages 363–374.

- Leitner, E., Rehm, G., and Moreno-Schneider, J. (2020). A dataset of german legal documents for named entity recognition. *arXiv preprint arXiv:2003.13016*.
- Li, J., Sun, A., Han, J., and Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.
- Mikheev, A., Moens, M., and Grover, C. (1999). Named entity recognition without gazetteers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 1–8.
- Santos, D. and Cardoso, N. (2006). A golden resource for named entity recognition in portuguese. In *International Workshop on Computational Processing of the Portuguese Language*, pages 69–79. Springer.
- Schmitt, X., Kubler, S., Robert, J., Papadakis, M., and LeTraon, Y. (2019). A replicable comparison study of ner software: Stanfordnlp, nltk, opennlp, spacy, gate. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 338–343. IEEE.
- Silva, R. L. d., Hoch, P. A., and Righi, L. M. (2013). Transparência pública e a atuação normativa do cnj. *Revista direito GV*, 9:489–514.
- Yadav, V. and Bethard, S. (2019). A survey on recent advances in named entity recognition from deep learning models. *arXiv preprint arXiv:1910.11470*.