

Projeto Final

IA368DD_2023S1: Deep Learning aplicado a Sistemas de Buscas

Student: Marcus Vinícius Borela de Castro

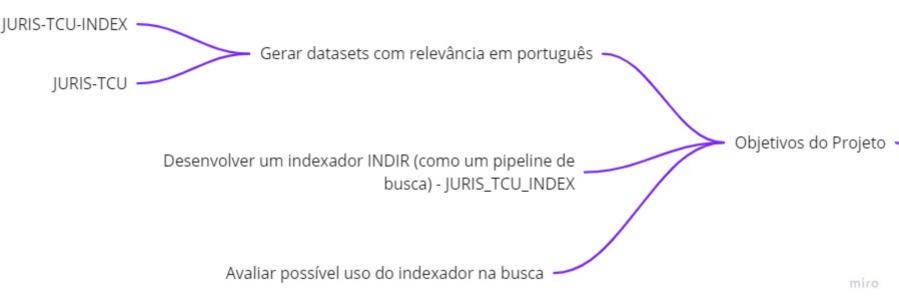
IndIR

Indexing Improving Information Retrieval in the Juris Dataset
(Jurisprudence Statements and Thesaurus of the
Federal Court of Accounts of Brazil - TCU)

Leonardo Pacheco & Marcus Borela

https://github.com/marcusborela/ind-ir







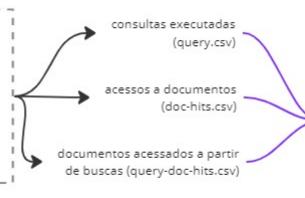




período: junho/2022 a maio/2023 queries específicas na base de Jurisprudência Selecionada

Retiradas consultas por todos documentos (*)
Retiradas consultas com operadores de proximidade

Acessos sumarizados e anonimizados



miro

Histórico de uso do sistema de

pesquisa à Jurisprudência

(LOG-JURIS-TCU)

Sistema Vocabulário de Controle Externo (VCE)

Ação de controle

DEF: Ação de controle é uma generalização dos atos do TCU, conduzida por uma de suas unidades, tendo um ministro relator, buscando investigar aspectos de um objeto de

controle, podendo levar a uma decisão (acórdão) e a uma sanção.

TE: Ação de controle externo

UP: Ações de controle

Ação de controle externo

DEF: Toda ação empreendida para a consecução da missão institucional do TCU, no

âmbito de suas funções finalísticas. (Fonte: BRASIL. Tribunal de Contas da União.

Glossário de Controle Externo. Revisão set.2017)

TG: Ação de controle

TR: Controle externo

Relatório de monitoramento

Volume de Recursos Fiscalizados

Órgão de controle externo

Ação de descumprimento de preceito fundamental

USE: Arguição de descumprimento de preceito fundamental

Ação de esbulho

USE: Ação possessória

área - 10 termos fixos (1 obrigatória por enunciado)

temas - qq termo do VCE (1 obrigatória por enunciado)

subtemas- qq termo do VCE (1 obrigatória por enunciado)

outras indexações (de 0 a 9 por enunciado)

indexações de enunciados por termos do VCE (qrel.csv) Dados dos 2 sistemas

query: JURIS

doc: VCE

qrel: JURIS x VCE

Produção do dataset de indexação (JURIS-TCU-INDEX)

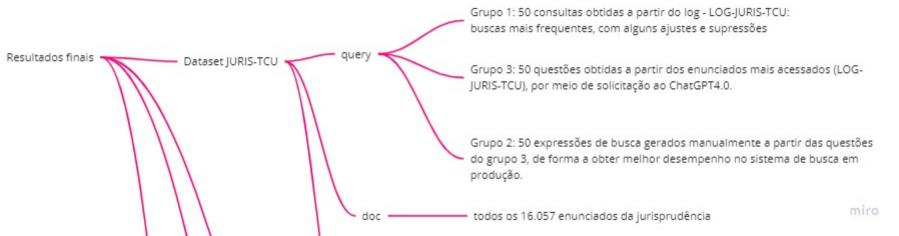
doc: JURIS

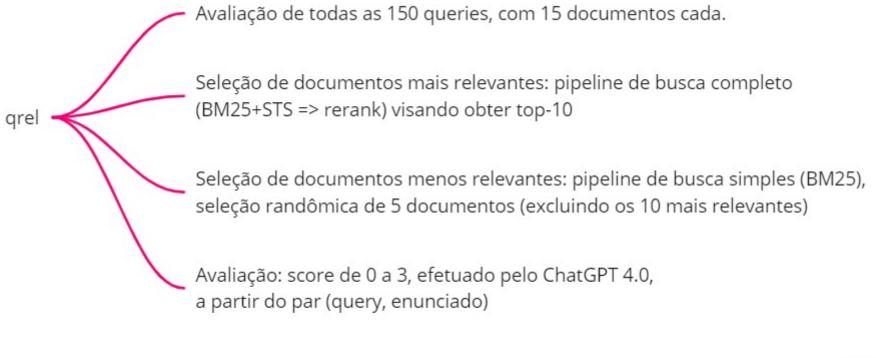
queries e grel construídos

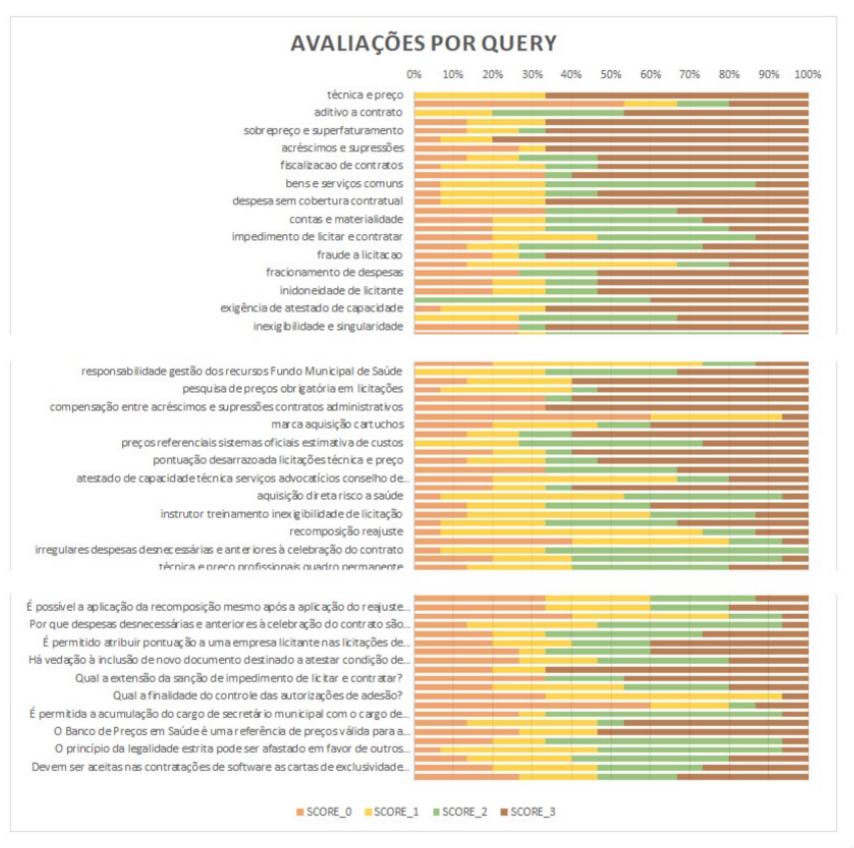
Produção do dataset de avaliação (IURIS-TCU)

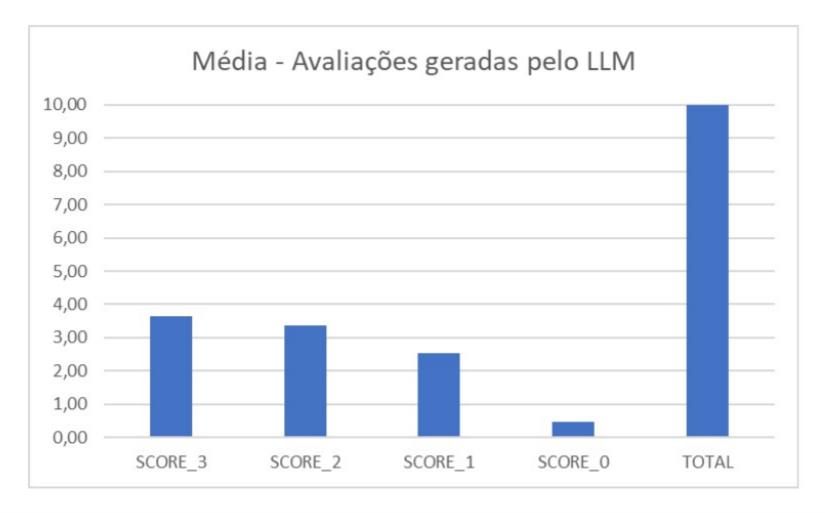
Construção do INDIR como pipeline de busca

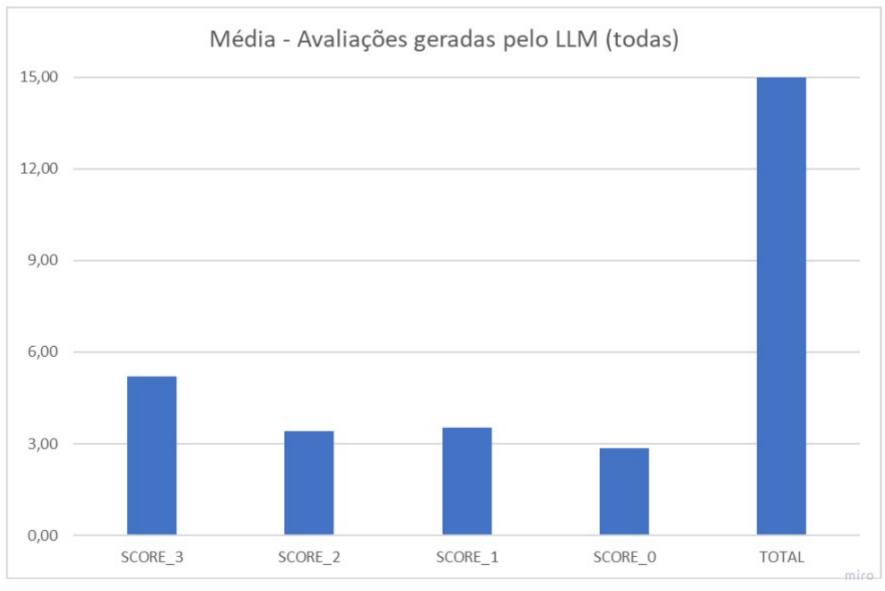
Experimentação de uso do INDIR na busca no JURIS-TCU



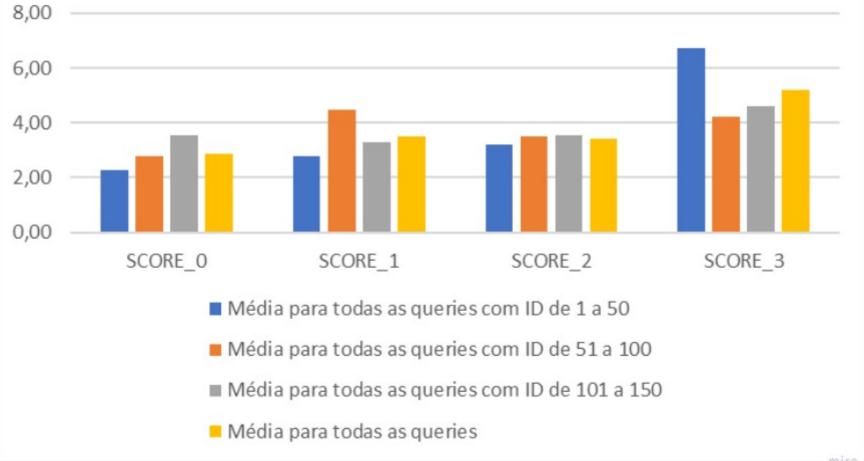






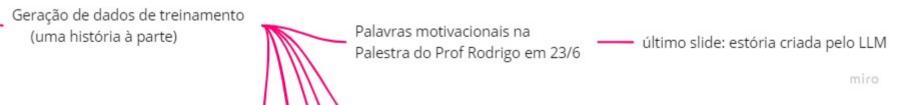


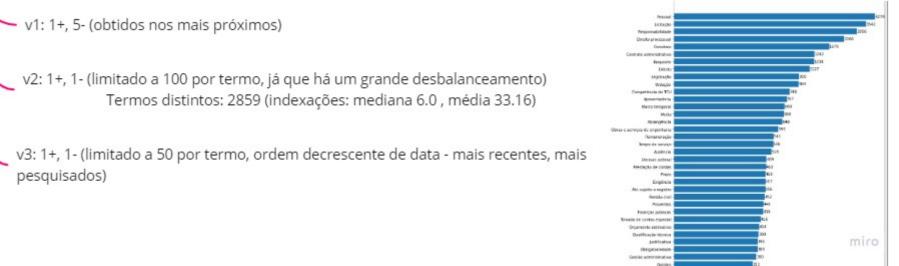
Média de enunciados por score Comparação entre os grupos de queries







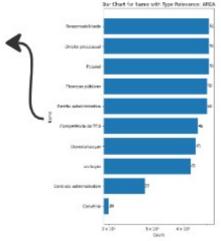




Único com bons resultados (v1 a v3): indexação com critério "AREA"

São só 10 termos (problema de classificação, na verdade)

🔪 10 termos bem diferentes; pessoal, contrato, etc 🤸



v3: 1+, 1- (sample bm25@1000; balanceamento limitando quantidade por tipo de indexação)

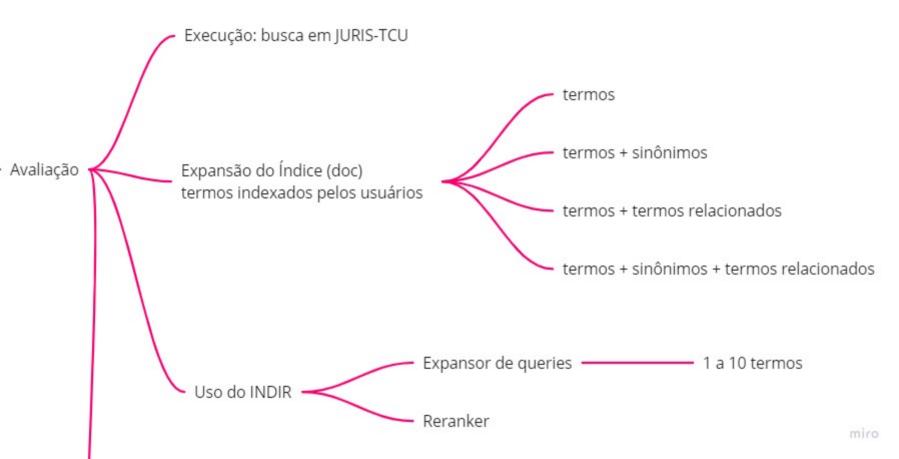
Modelos treinados

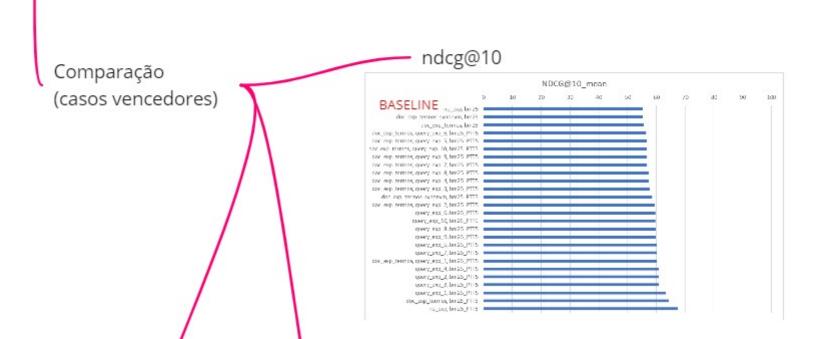
PTT5_INDIR_83 unicamp-dl/ptt5-base-pt-msmarco-100k-v2

MINILM_INDIR_400 unicamp-dl/mMiniLM-L6-v2-pt-v2

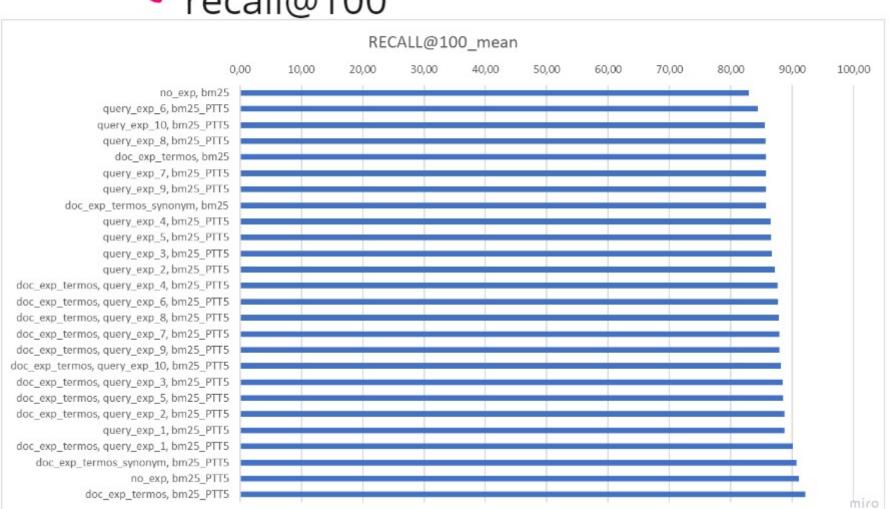
```
minilm:
ptt5:
mt5 3b:
area: 62.09; subtheme: 20,90; theme: 53,78; extra: 26,71; total gte5: 37.57; total: 37,51;
minilm indir (treino de 4 épocas)
area: 74.18; subtheme:29.20; theme: 48.58; extra: 33.85; total gte5: 43.49; total: 43.71;
ptt5 indir 25 (25% de 1 época):
area: 85,66; subthem 43,483; theme: 60,77; extra: 42.03; total gte5 54,409; total: 54,56;
```

Novo INDIR em treinamento (138%)

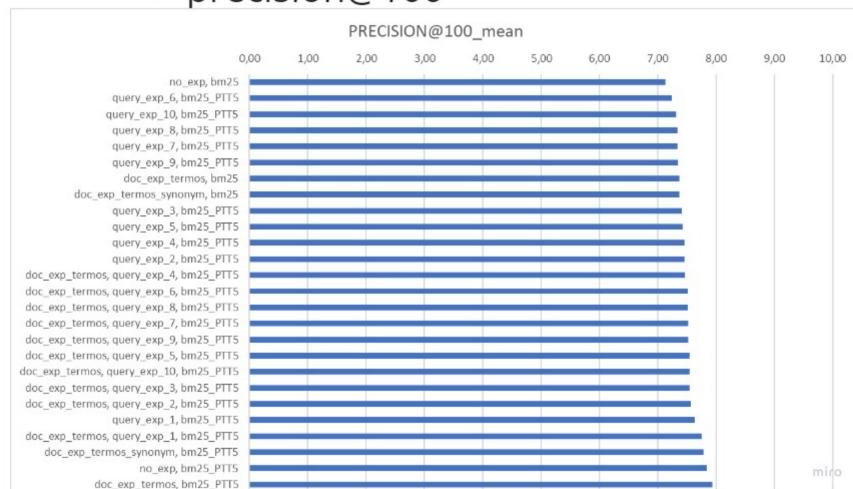


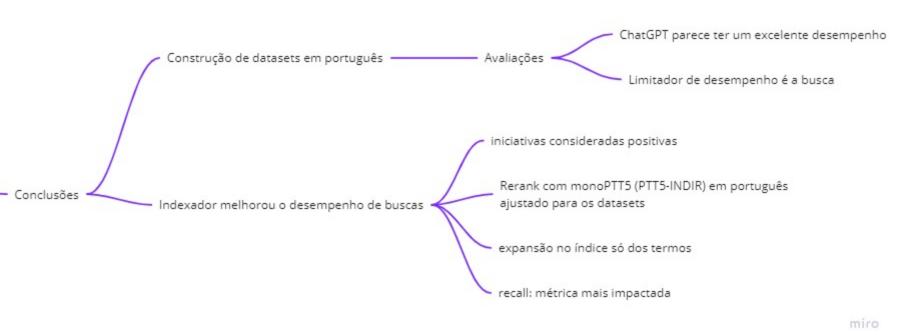


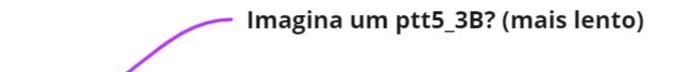
recall@100



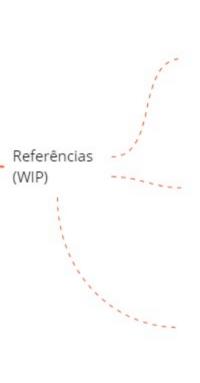
precision@100







Trabalhos futuros LLM como estágio 3?



R.Nogueira, Z.Jiang, R.Pradeep, and J.Lin. Document ranking with a pretrained sequence- to sequence model. In Findingsof the Association for Computational Linguistics: EMNLP 2020, pages 708–718. Disponível em: https://arxiv.org/pdf/2010.06467.pdf. Acesso em 25 maio 2023, 2020.

BRASIL. Tribunal de Contas da União. Vocabulário de controle externo do Tribunal de Contas da União – 3.ed. rev. e ampl. – Brasília : TCU, Instituto Serzedello Corrêa, Centro de Documentação. Disponível em:

https://portal.tcu.gov.br/data/files/F8/04/8E/5E/A0B3071068A7C107F18818A8/VCE_TCU.pdf. Acesso em 25 maio 2023. 2019.

____. Portaria-TCU - 85, de 06 de junho de 2022. Aprova o Manual de Sistematização e Divulgação da Jurisprudência do Tribunal de Contas da União. Disponível em: https://pesquisa.apps.tcu.gov.br/#/documento/ato-

normativo/Ac%25C3%25B3rd%25C3%25A3o%2520n%25C2%25BA%25202800%252F2022/% 2520/score%2520desc/1/%2520. Acesso em 25 maio 2023. 2022. INDIR: nasceu como uma hipótese... e se tornou uma realidade!



