

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

Leonardo Pacheco e Marcus Borela

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

Projeto Final da disciplina IA368DD_2023S1: Deep Learning Aplicado a Sistemas de Buscas

						Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	
TAREFA	ATRIBUÍDO PARA	PROGRESSO	INÍCIO	TÉRMINO	DIAS	q s s d s t q	q s s d s t q	q q s s d s t q	q q s s d s t q	q s s d s t q	q s s d s t q	q
Planejamento												
Plano básico	LM	100%	18/5/23	2/6/23	16							
1a Apresentação	LM	100%	22/5/23	24/5/23	3							
Definir base de dados	LM	100%	23/5/23	31/5/23	9							
Definir métricas	LM	100%	25/5/23	18/6/23	25							
Construção do Dataset de Indexação (Juris-TCU-Index)												
Carga dos dados do sistema Juris (incluindo indexações)	LM	100%	27/5/23	2/6/23	7							
Explorar dados do sistema Juris	LM	100%	1/6/23	5/6/23	5							
Explorar dados de indexações	LM	100%	5/6/23	7/6/23	3							
Tratar dados	LM	100%	27/5/23	13/6/23	18							
Gerar dataset (com grels)	M	100%	7/6/23	14/6/23	8							
Validar formato dataset	M	100%	7/6/23	9/6/23	3							
Construção do Dataset de Avaliação de Busca (Juris-TCU)												
Explorar dados de log da busca	L	100%	1/6/23	14/6/23	14							
Criar queries	L	100%	5/6/23	15/6/23	11							
Definir prompt LLM para qrel	LM	100%	15/6/23	17/6/23	3							
Montar base de avaliação (qrel)	L	100%	16/6/23	25/6/23	10							
Construção do pipeline de indexação												
Criar estrutura busca (elastic search, etc)	M	100%	4/6/23	7/6/23	4							
Criar estrutura de registro e avaliação (métrica)	M	100%	4/6/23	12/6/23	9							
Realizar experimentos: indexador como pipeline de busca	LM	100%	12/6/23	21/6/23	10							
Treinar modelo	M	80%	16/6/23	29/6/23	14							
Frente pipeline de busca												
Criar estrutura busca (elastic search, etc)	M	100%	17/6/23	26/6/23	10							
Criar estrutura de registro e avaliação (métrica)	M	100%	17/6/23	26/6/23	10							
Experimentar: indexadores como expansão (cadastro)	LM	80%	20/6/23	29/6/23	10							
Frente Avaliação e Fechamento												
Relatório	LM	40%	16/6/23	2/7/23	17							
Preparar apresentação final	LM	100%	24/6/23	28/6/23	5							
Publicar datasets com grels (indexação e busca da jurisprudência)	LM	100%	4/6/23	30/6/23	27							