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

						Seman Seman					a 6		
					21	22	23	24	25	26	27 2	.8 2	
TAREFA	ATRIBUÍDO PARA	PROGRESSO	INÍCIO	TÉRMINO	q	q	s	s	d	s	t	a a	
Planejamento										\Box	\perp		
Plano básico	LM	100%	18/5/23	2/6/23							\perp		
1a Apresentação	LM	100%	22/5/23	24/5/23									
Definir base de dados	LM	100%	23/5/23	31/5/23									
Definir métricas	LM	100%	25/5/23	18/6/23									
Construção do Dataset de Indexação (Juris-TCU-Index)													
Carga dos dados do sistema Juris (incluindo indexações)	LM	100%	27/5/23	2/6/23									
Explorar dados do sistema Juris	LM	100%	1/6/23	5/6/23									
Explorar dados de indexações	LM	100%	5/6/23	7/6/23									
Tratar dados	LM	100%	27/5/23	13/6/23									
Gerar dataset (com qrels)	M	100%	7/6/23	14/6/23									
Validar formato dataset	M	100%	7/6/23	9/6/23									
Construção do Dataset de Avaliação de Busca (Juris-TCU)													
Explorar dados de log da busca	L	100%	1/6/23	14/6/23									
Criar queries	L	100%	5/6/23	15/6/23									
Definir prompt LLM para qrel	LM	90%	15/6/23	17/6/23									
Montar base de avaliação (qrel)	L	40%	16/6/23	25/6/23									
Construção do pipeline de indexação													
Criar estrutura busca (elastic search, etc)	M	100%	4/6/23	7/6/23									
Criar estrutura de registro e avaliação (métrica)	M	100%	4/6/23	12/6/23									
Realizar experimentos: indexador como pipeline de busca	LM	100%	12/6/23	21/6/23									

					21 <mark>22</mark> 23 24 25 26 27 28 <mark>29</mark>
TAREFA	ATRIBUÍDO PARA	PROGRESSO	INÍCIO	TÉRMINO	q
Treinar modelo	M	5%	16/6/23	23/6/23	
(ou: experimentar LLM como estágio 3 dos pipes)	М	0%	16/6/23	26/6/23	
Frente pipeline de busca					
Criar estrutura busca (elastic search, etc)	M	90%	17/6/23	26/6/23	
Criar estrutura de registro e avaliação (métrica)	М	90%	17/6/23	26/6/23	
Experimentar: indexadores como expansão (cadastro)	LM	0%	20/6/23	25/6/23	
Frente Avaliação e Fechamento					
Relatório	LM	0%	16/6/23	30/6/23	
Preparar apresentação final	LM	0%	24/6/23	30/6/23	
Publicar datasets com qrels (indexação e busca da jurisprudência)	LM	60%	4/6/23	30/6/23	

Se m

Semana 6