cases	doc_1		doc_2		decision	id
	authors	• Ruan Chaves Rodrigues • JÃ©ssica Rodrigues da Silva • Pedro Vitor Quinta de Castro • NÃ¡dia FÃ©lix Felipe da Silva • Anderson da Silva Soares	**authors**	• Evandro Fonseca • JoÃ£o Paulo Reis Alvarenga	DUPLICATES	351
	title	Multilingual Transformer Ensembles for Portuguese Natural Language Tasks	**title**	Multilingual Transformer Ensembles for Portuguese Natural Language Tasks		
	publication_date	2019-01-01 00:00:00	**publication_date**	2019-01-01 00:00:00		
	source	SupportedSources.INTERNET_ARCHIVE	**source**	SupportedSources.INTERNET_ARCHIVE		
	journal		**journal**			
	volume		**volume**			
	doi		**doi**			
	urls	• https://web.archive.org/web/20210616054435/http://ceur-ws.org/Vol-2583/3_DLB.pdf	**urls**	• https://web.archive.org/web/20210616054454/http://ceur-ws.org/Vol-2583/7_Stilingue.pdf		
	id	id-5928943429701835305	**id**	id-2458278284339229113		
	abstract	Due to the technical gap between the language models available for low-resource languages and the state-of-the-art models available in English and Chinese, a simple approach that deploys automatic translation and ensembles predictions from Portuguese and English models is competitive with monolingual Portuguese approaches that may demand task-specific preprocessing and hand-crafted features. We performed our experiments on ASSIN 2 -the second edition of the AvaliaÃ§Ã£o de Similaridade SemÃ¢ntica e InferÃªncia Textual (Evaluating Semantic Similarity and Textual Entailment). On the semantic textual similarity task, we performed multilingual ensemble techniques to achieve results with higher Pearson correlation and lower mean squared error than BERTmultilingual, and on the textual entailment task, BERT-multilingual could be surpassed by automatically translating the corpus into English and then fine-tuning a large RoBERTa model over the translated texts.	**abstract**	In this paper we present our approach to deal with semantic relatedness and textual entailment, two tasks proposed in ASSIN-2 (Second evaluation of semantic relatedness and textual entailment). We develop 18 features that explore lexical, syntactic and semantic information. To train the models we applied both supervised machine learning and an architecture based in Wide and Deep learning. Our proposal demonstrated to be competitive with the current state-of art models and with other participant models for Portuguese, mainly when the mean square error is considered.		
	versions		**versions**			