	doc_1		doc_2		decision	id
cases	authors	• Beaus Iranzo, Rafael	authors	• Rafael Beaus Iranzo • Paula GÃ³mez Duran	DUPLICATES	159
	title	Article similarities using transformers	title	Article similarities using transformers		
	publication_date	2022-06-13 00:00:00	publication_date	None		
	source	SupportedSources.CORE	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal		journal			
	volume		volume			
	doi	None	doi			
	urls	• http://diposit.ub.edu/dspace/bitstream/2445/188695/2/tfg_beaus_iranzo_rafael.pdf	urls	• https://www.semanticscholar.org/paper/c7496dac7bb2da7c27ec77710ca03ed8dbe9012b		
	id	id-1062870159575154394	id	id4833935773661552929		
	abstract	Treballs Finals de Grau d'Enginyeria InformÃ tica, Facultat de MatemÃ tiques, Universitat de Barcelona, Any: 2022, Director: Paula GÃ³mez Duran i Jordi VitriÃ i Marca[en] The field of natural language processing is essential in todayâ€™s data-driven world. In 2017 the Tranformers architecture was introduced based on the concept of attention from 2014. The effects of this new structure were already changing the paradigm when the language processing model BERT marked an inflection point, in 2018. BERT makes use of the Transformersâ€™ parallelization to achieve a network that can be pretrained. In that pretraining, the model is able to learn how a language works on its own: by only feeding it with texts. An improved version came out shortly after, RoBERTa, after which most of the models were based. In this thesis, we will focus on studying BERTa (a RoBERTa-based Catalan language model) with a dataset from the Gran EnciclopÃ¨dia Catalana. That analysis will include tasks to assess how does the model perform with real-world data. The study aims to validate the quality of the resulting embeddings produced by the model in order to further use them to build an article retrieval platform. There, each article query could be related to those with similar information. The semantic textual similarity describes how alike a pair of sentences are and this will be a fundamental target for the designed experiments and development. Finally, the results will be visualized and interpreted by using a simple front- end tool also created in this work	abstract	The field of natural language processing is essential in todayâ€™s data-driven world. In 2017 the Tranformers architecture was introduced based on the concept of attention from 2014. The effects of this new structure were already changing the paradigm when the language processing model BERT marked an inflection point, in 2018. BERT makes use of the Transformersâ€™ parallelization to achieve a network that can be pretrained. In that pretraining, the model is able to learn how a language works on its own: by only feeding it with texts. An improved version came out shortly after, RoBERTa, after which most of the models were based. In this thesis, we will focus on studying BERTa (a RoBERTa-based Catalan language model) with a dataset from the Gran EnciclopÃ¨dia Catalana . That analysis will include tasks to assess how does the model perform with real-world data. The study aims to validate the quality of the resulting embeddings produced by the model in order to further use them to build an article retrieval platform. There, each article query could be related to those with similar information. The semantic textual similarity describes how alike a pair of sentences are and this will be a fundamental target for the designed experiments and development. Finally, the results will be visualized and interpreted by using a simple front-end tool also created in this work.		
	versions		versions			