cases	doc_1		doc_2		decision	id
			authors	• Vladimir Ivanov • Andrey Sadovykh • Alexandr Naumchev • Alessandra Bagnato • Kirill Yakovlev	DUPLICATES	114
	authors	• V. Ivanov • A. Sadovykh • A. Naumchev • A. Bagnato • K. Yakovlev	**title**	Extracting Software Requirements from Unstructured Documents		
	title	Extracting Software Requirements from Unstructured Documents	**publication_date**	2022-02-04 00:00:00		
	publication_date	2022-02-04 00:00:00	**source**	SupportedSources.INTERNET_ARCHIVE		
	source	SupportedSources.SEMANTIC_SCHOLAR	**journal**			
	journal		**volume**			
	volume		**doi**			
	doi	10.1007/978-3-031-15168-2_2	**urls**	• https://web.archive.org/web/20220208155525/https://arxiv.org/pdf/2202.02135v1.pdf		
	urls	• https://www.semanticscholar.org/paper/a6e77c9890b90035986f202c14901de68a23e986	**id**	id3556725635840589375		
	id	id6350016892214582653	**abstract**	Requirements identification in textual documents or extraction is a tedious and error prone task that many researchers suggest automating. We manually annotated the PURE dataset and thus created a new one containing both requirements and non-requirements. Using this dataset, we fine-tuned the BERT model and compare the results with several baselines such as fastText and ELMo. In order to evaluate the model on semantically more complex documents we compare the PURE dataset results with experiments on Request For Information (RFI) documents. The RFIs often include software requirements, but in a less standardized way. The fine-tuned BERT showed promising results on PURE dataset on the binary sentence classification task. Comparing with previous and recent studies dealing with constrained inputs, our approach demonstrates high performance in terms of precision and recall metrics, while being agnostic to the unstructured textual input.		
	abstract	None	**versions**			
	versions					