cases	doc_1	doc_2	decision	id
	<table><tr><td>authors</td><td>• Lim, W. • Tayyar Madabushi, H.</td></tr><tr><td>title</td><td>UoB at SemEval-2020 Task 12: Boosting BERT with Corpus Level Information</td></tr><tr><td>publication_date</td><td>2020-01-01 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.CROSSREF</td></tr><tr><td>journal</td><td></td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td>10.18653/v1/2020.semeval-1.295</td></tr><tr><td>urls</td><td>• http://dx.doi.org/10.18653/v1/2020.semeval-1.295</td></tr><tr><td>id</td><td>id4032852201501391561</td></tr><tr><td>abstract</td><td></td></tr><tr><td>versions</td><td></td></tr></table>	<table><tr><td>authors</td><td>• Wah Meng Lim • Harish Tayyar Madabushi</td></tr><tr><td>title</td><td>UoB at SemEval-2020 Task 12: Boosting BERT with Corpus Level Information</td></tr><tr><td>publication_date</td><td>2020-08-19 16:47:15+00:00</td></tr><tr><td>source</td><td>SupportedSources.ARXIV</td></tr><tr><td>journal</td><td>None</td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td></td></tr><tr><td>urls</td><td>• http://arxiv.org/pdf/2008.08547v1 • http://arxiv.org/abs/2008.08547v1 • http://arxiv.org/pdf/2008.08547v1</td></tr><tr><td>id</td><td>id5256500666332593482</td></tr><tr><td>abstract</td><td>Pre-trained language model word representation, such as BERT, have been extremely successful in several Natural Language Processing tasks significantly improving on the state-of-the-art. This can largely be attributed to their ability to better capture semantic information contained within a sentence. Several tasks, however, can benefit from information available at a corpus level, such as Term Frequency-Inverse Document Frequency (TF-IDF). In this work we test the effectiveness of integrating this information with BERT on the task of identifying abuse on social media and show that integrating this information with BERT does indeed significantly improve performance. We participate in Sub-Task A (abuse detection) wherein we achieve a score within two points of the top performing team and in Sub-Task B (target detection) wherein we are ranked 4 of the 44 participating teams.</td></tr><tr><td>versions</td><td></td></tr></table>	DUPLICATES	281