

cases	doc_1		doc_2				decision	id
							DUPLICATES	198
	authors	<ul style="list-style-type: none">John WietingKevin GimpelNeubig, GrahamTaylor Berg-Kirkpatrick	authors	<ul style="list-style-type: none">J. WietingKevin GimpelGraham NeubigTaylor Berg-Kirkpatrick				
	title	Paraphrastic Representations at Scale	title	Paraphrastic Representations at Scale				
	publication_date	2021-04-30 00:00:00	publication_date	2021-04-30 00:00:00				
	source	SupportedSources.OPENALEX	source	SupportedSources.SEMANTIC_SCHOLAR				
	journal	arXiv (Cornell University)	journal					
	volume		volume					
	doi	10.48550/arxiv.2104.15114	doi					
	urls	<ul style="list-style-type: none">https://openalex.org/W4297795623https://doi.org/10.48550/arxiv.2104.15114http://arxiv.org/pdf/2104.15114	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/56501a3441c2074bbbbe31015d6d41c57d9d285b				
	id	id7046172893523899068	id	id-1856322892855683722				
	abstract		abstract	We present a system that allows users to train their own state-of-the-art paraphrastic sentence representations in a variety of languages. We also release trained models for English, Arabic, German, French, Spanish, Russian, Turkish, and Chinese. We train these models on large amounts of data, achieving significantly improved performance from the original papers proposing the methods on a suite of monolingual semantic similarity, cross-lingual semantic similarity, and bitext mining tasks. Moreover, the resulting models surpass all prior work on unsupervised semantic textual similarity, significantly outperforming even BERT-based models like Sentence-BERT (Reimers and Gurevych, 2019). Additionally, our models are orders of magnitude faster than prior work and can be used on CPU with little difference in inference speed (even improved speed over GPU when using more CPU cores), making these models an attractive choice for users without access to GPUs or for use on embedded devices. Finally, we add significantly increased functionality to the code bases for training paraphrastic sentence models, easing their use for both inference and for training them for any desired language with parallel data. We also include code to automatically download and preprocess training data.				
	versions		versions					