Clustering Results

Dejvis Toptani

2023-05-02

dataset	model	max_files_per_class	dimensions	NMI	Purity
bbc	tfidf	500	2	0.751	0.907
bbc	tfidf	500	4	0.787	0.923
bbc	tfidf	500	8	0.795	0.927
bbc	tfidf	500	16	0.793	0.926
bbc	Word2Vec	500	2	0.786	0.918
bbc	Word2Vec	500	4	0.767	0.903
bbc	Word2Vec	500	8	0.781	0.912
bbc	Word2Vec	500	16	0.773	0.906
bbc	BERT	500	2	0.789	0.902
bbc	BERT	500	4	0.869	0.956
bbc	BERT	500	8	0.871	0.957
bbc	BERT	500	16	0.875	0.958
20news-18828	tfidf	500	2	0.542	0.723
20news-18828	tfidf	500	4	0.601	0.797
20news-18828	tfidf	500	8	0.571	0.709
20news-18828	tfidf	500	16	0.593	0.726
20news-18828	Word2Vec	500	2	0.383	0.527
20news-18828	Word2Vec	500	4	0.449	0.630
20news-18828	Word2Vec	500	8	0.463	0.636
20news-18828	Word2Vec	500	16	0.462	0.635
20news-18828	BERT	500	2	0.163	0.351
20news-18828	BERT	500	4	0.161	0.332
20news-18828	BERT	500	8	0.155	0.331
20news-18828	BERT	500	16	0.155	0.331