

**Fine-tune Representations**



**Weighting scheme**



**Tokenizer**



**Clustering**



**Dimensionality Reduction**



**Embeddings**

**Optional  
Fine-tuning**

**c-TF-IDF**

**CountVectorize**

**HDBSCAN**

**UMAP**

**SBERT**