

# BERT Pre-training Corpus: 3.3 Billion Words

2500M words

Millions of Words

800M words

BooksCorpus

Total: 3.3B

Wikipedia