

BERT Pre-training Corpus: 3.3 Billion Words

2500M words

Millions of Words

2500
2000
1500
1000
500
0

BooksCorpus

Wikipedia

800M words

Total: 3.3B

