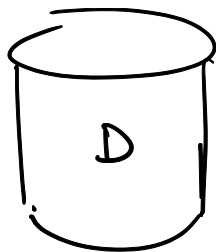
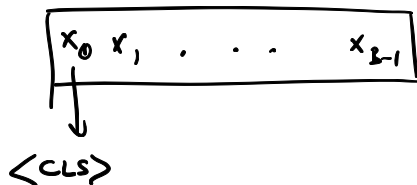


BERT Contextual Model



S_1
 S_2
 \vdots
 S_N

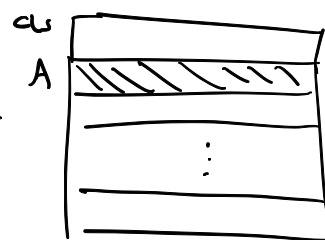
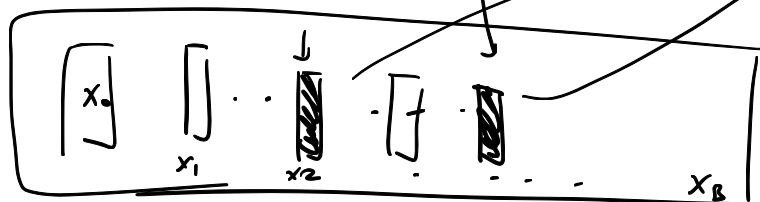
$N \approx 500k$
 $|S_i| \leq 1000$
 $k\text{-mer} = 1$



$B = \text{block size}$

$B = 1000$

M A I G L A . . .



$d_{\text{model}} = ? 128$

initial token embeddings

