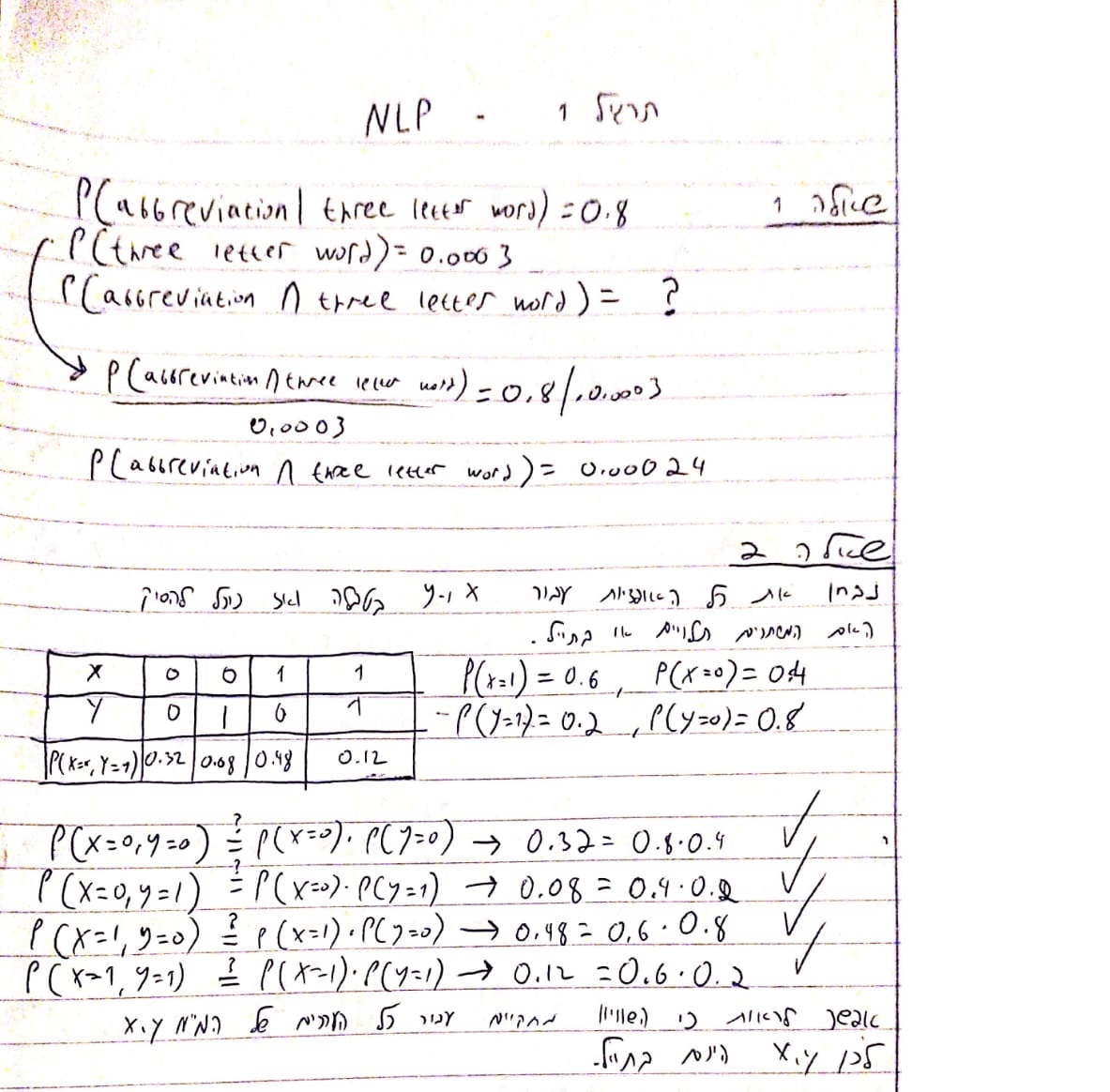
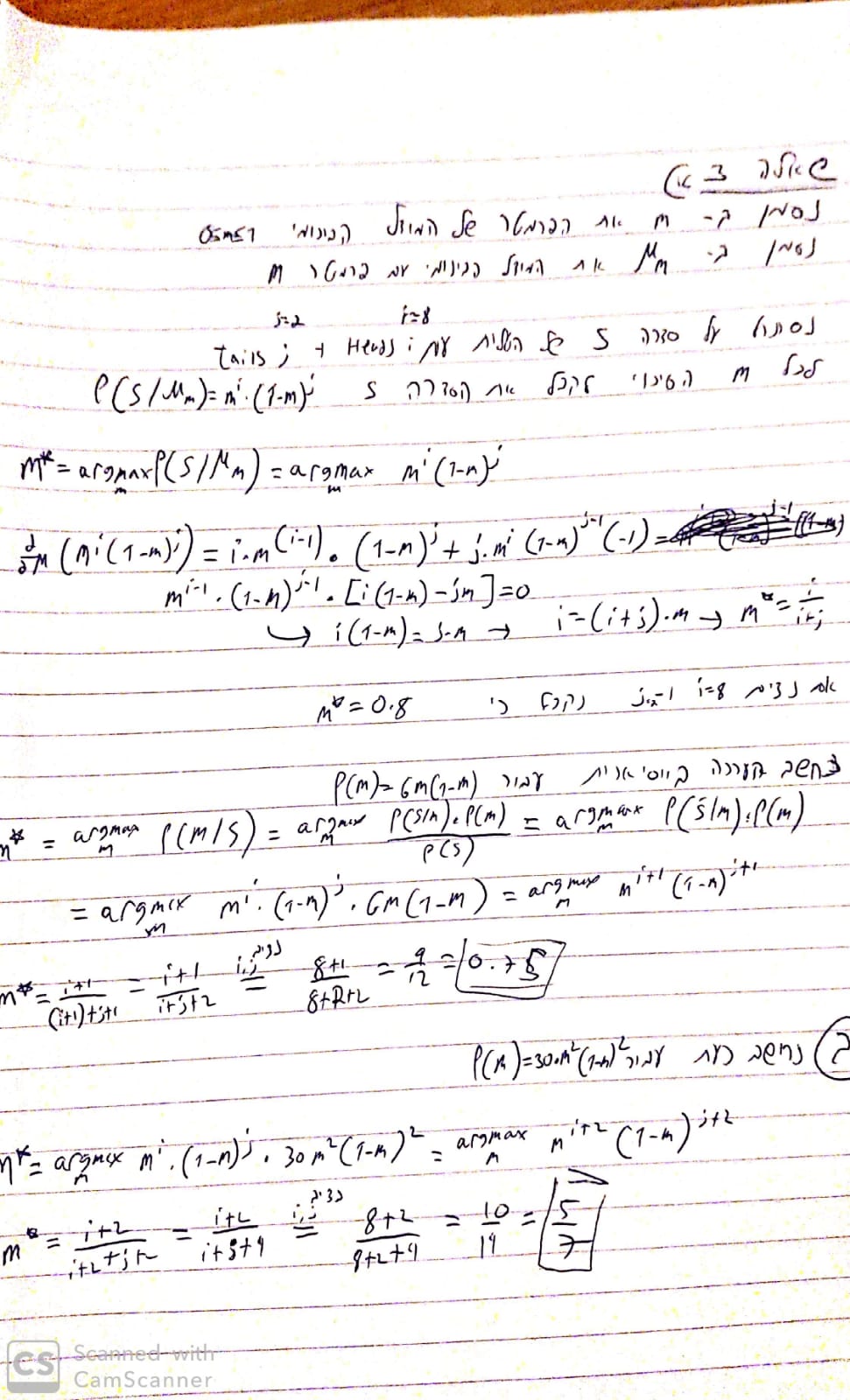
תרגיל מס' 1 NLP

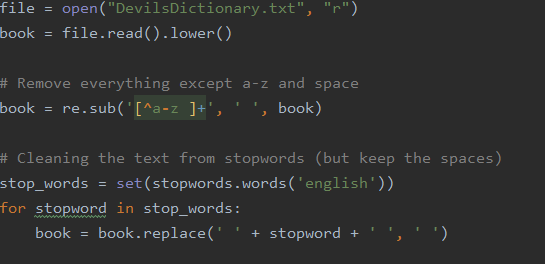
מגישים: דניאל כהן – 201191616

דביר פרדלסקי – 311503452

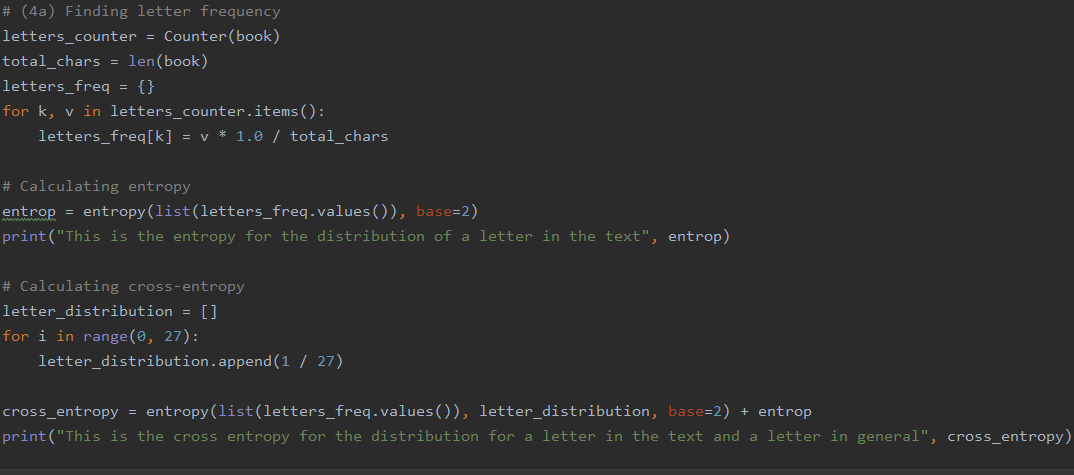




שאלה 4



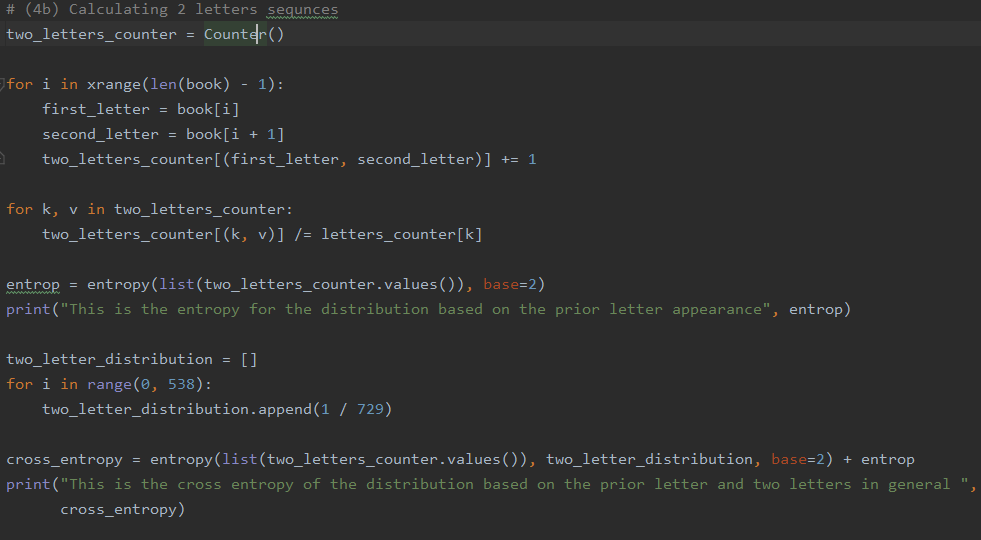
סעיף א'



This is the entropy for the distribution of a letter in the text 4.081758055081569

This is the cross entropy for the distribution for a letter in the text and a letter in general 4.754887502163469

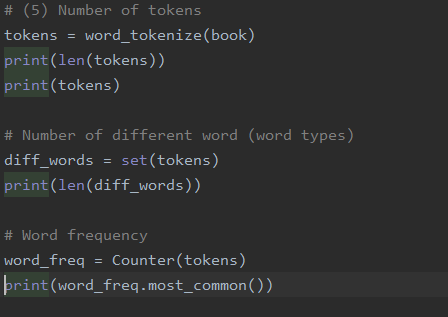
סעיף ב'



This is the entropy for the distribution based on the prior letter appearance 7.777935674541441

This is the cross entropy of the distribution based on the prior letter and two letters in general 9.071462362556627

שאלה 5



The number of tokens is 32333

The number of word types is 10757

