мета: оцінити ентропію тексту

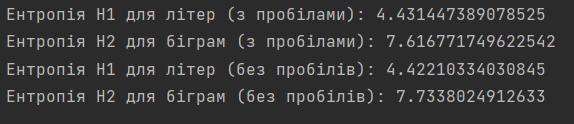
1. Знаходимо текст достатньої довжини

2. Запускаємо код для видалення знаків пунктуації та заміни великих літер на малі

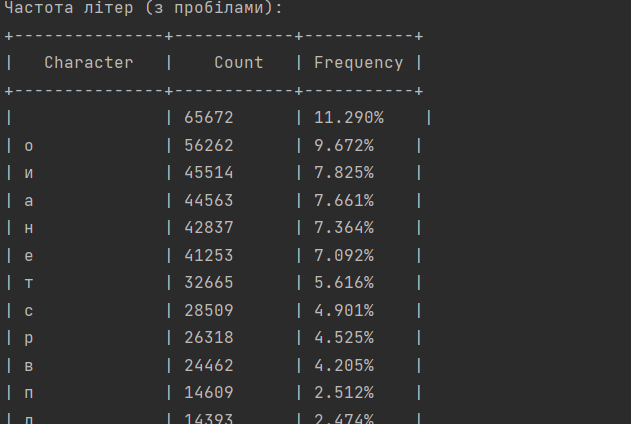
import string  
  
def process\_text\_file(input\_file\_path, output\_file\_path):  
 with open(input\_file\_path, 'r', encoding='utf-8') as input\_file, open(output\_file\_path, 'w', encoding='utf-8') as output\_file:  
 text = input\_file.read()  
 cleaned\_text = ''.join([char.lower() for char in text if char not in string.punctuation])  
 output\_file.write(cleaned\_text)  
  
input\_file\_path = r'C:\Users\Sasha\Desktop\lab1.txt'  
output\_file\_path = r'C:\Users\Sasha\Desktop\lab1\_1.txt'  
  
process\_text\_file(input\_file\_path, output\_file\_path)

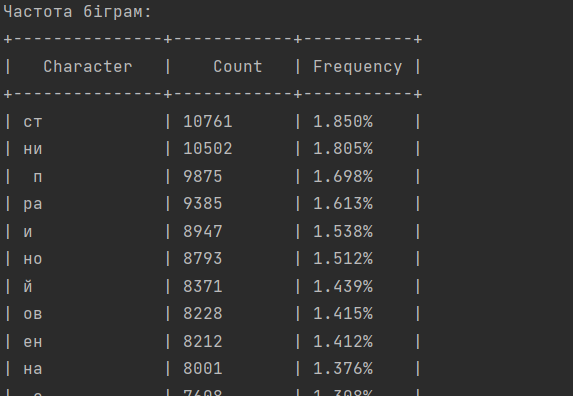
3. Пишемо код для знаходження ентропії(lab1\_1.py)

4. Отримуємо значення ентропії

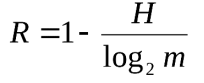


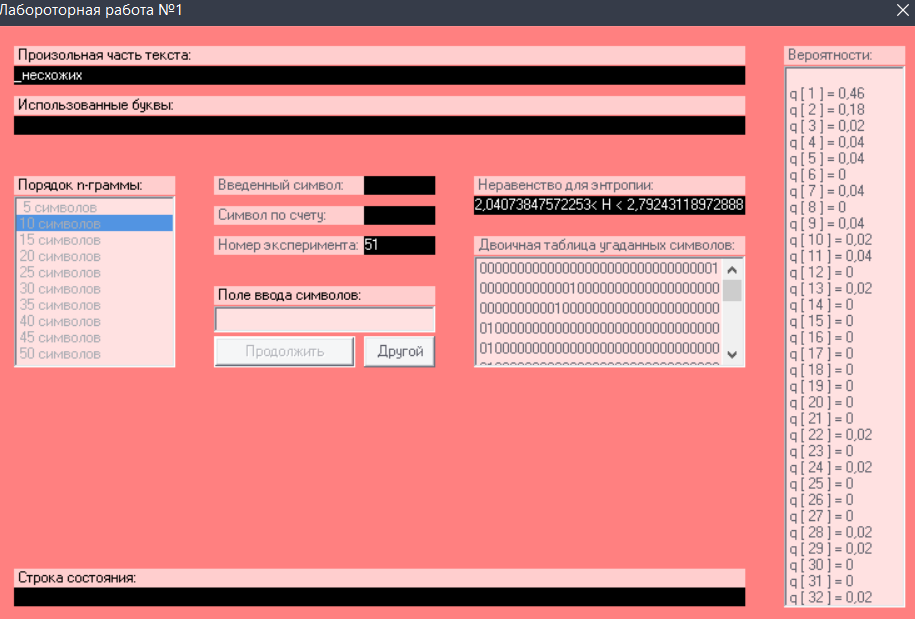
та таблиці частот літер та біграм



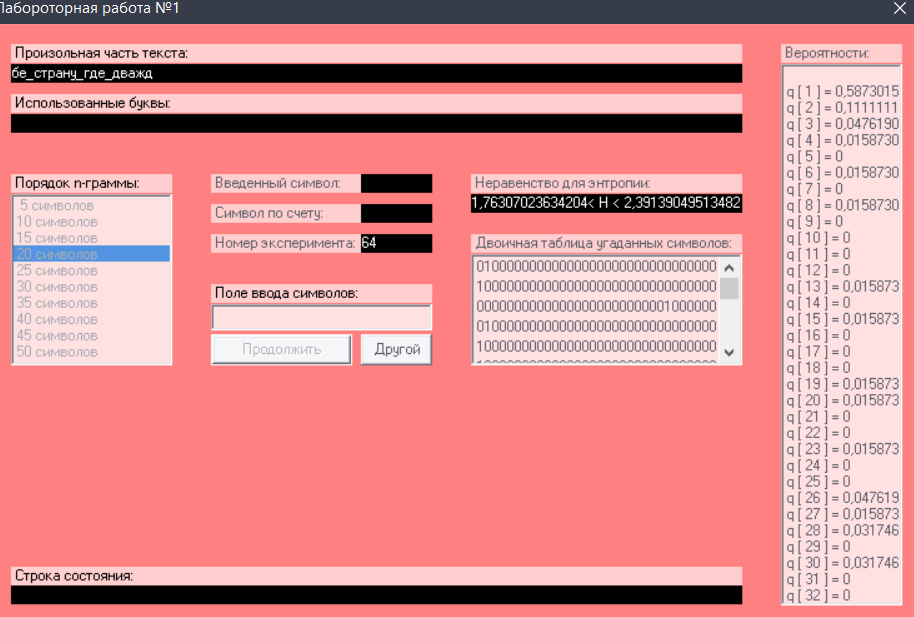


5. За допомогою програми CoolPinkProgram оцінити значення (10) H , (20) H , (30)

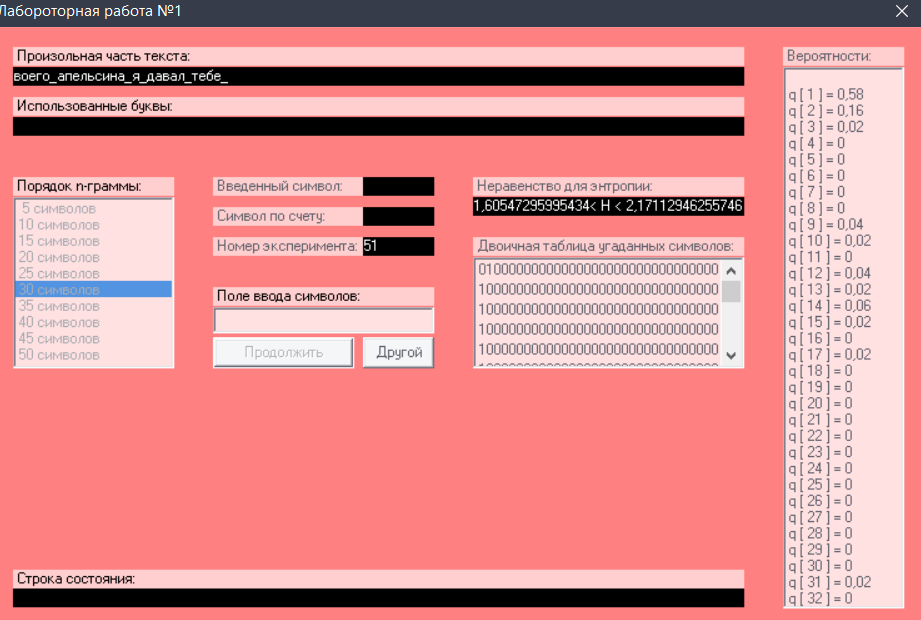




0,15<R<0,38



0,44<R<0,59



0,55<R<0,66

Висновки:

У ході роботи ми засвоїли поняття ентропії на символ джерела та його надлишковості та практично обрахували значння ентропії для різних моделей джерела тексту. У мене виникли деякі труднощі з CoolPinkProgram, спочатку я просто рандомно нажимав літери на клавіатурі, а потім зрозумів що треба вгадувати на основі частини тексту, яка вже є)