**cp1**

**Виконали:   
Пошивак Н.  
Боровков Н.**

**Мета роботи:** засвоєння поняття ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

**Постановка задачі**: Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 на тому ж тексті, в якому вилучено всі пробіли. За допомогою програми CoolPinkProgram оцінити значення H(10) , H(20) , H(30) . Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

**Хід роботи:**# Всі функції мають інтуїтивно зрозумілі імена і   
# короткий опис функціоналу, тому тут їх не вказуємо

1. Обрати текст достатнього об’єму
2. Відформатувати його, прибравши вске зайве
3. Написати функцію для монограм
4. Написати функцію для біграм
5. Написати допоміжні функції для ентропії та надлишковості
6. Вивести дані у зручному для читанні вигляді

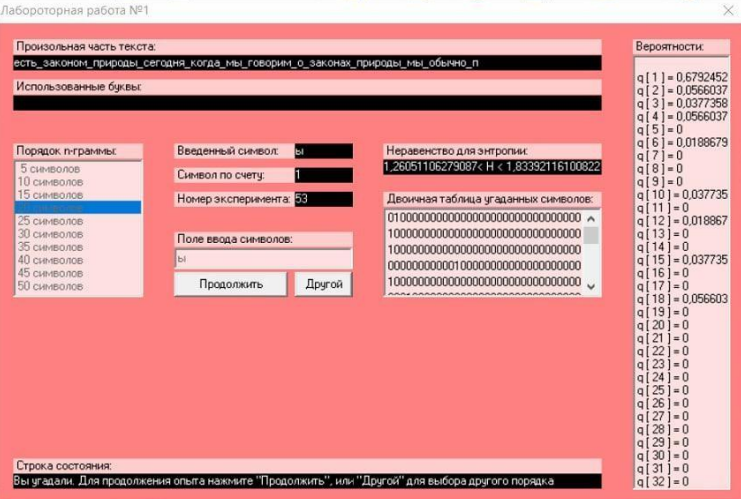
**Основні проблеми:**

1. Залити результат на GitHub
2. ~~Сісти робити лабораторну з штрафом -5 балів~~

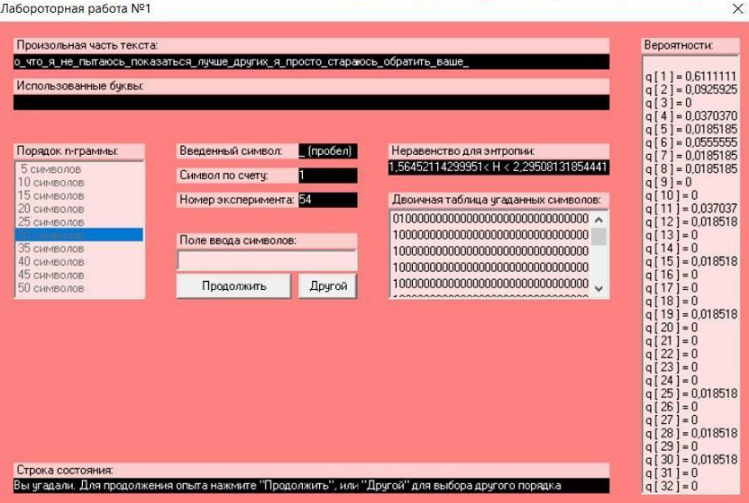
# Тут вивід результатів   
> python main.py > result\_values.txt

|  |  |  |
| --- | --- | --- |
|  | R (з пробілами) | R (без пробілів) |
| Н1 | 0.1422978 | 0.1161406 |
| Н2 | 0.5630196 | 0.5412049 |
| Н2 (крок 2) | 0.2242879 | 0.1813656 |

  
1,18 < H(10) < 1,75

0,768057 > R > 0,656017

1,26 < H(20) < 1,83  
0,752332 > R > 0,640292



1,56 < H(30) < 2,29  
0,693364 > R > 0.549874

**Висновки**:   
У ході виконання лабораторної роботи ми ознайомились з поняттям ентропії та навчились її визначати, освоїли вимірювання частоти повторювання символів в тексті. Ми дійшли такого висновку, що коли оброблений текст містить пробіл, ентропія є меншою. За допомогою програми CoolPinkProgram ми впевнилися, що чим більшою є ентропія, тим меншою є надлишковість мови.