Міністерство освіти і науки України
Національний технічний университет України
“Київський політехнічний інститут ім. Ігоря Сікорського”
Фізико-технічний інститут

СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконала:
студентка групи ФІ-93
Меднікова Олександра

Перевірив:
Чорний О. М.

Київ – 2022

Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближенного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

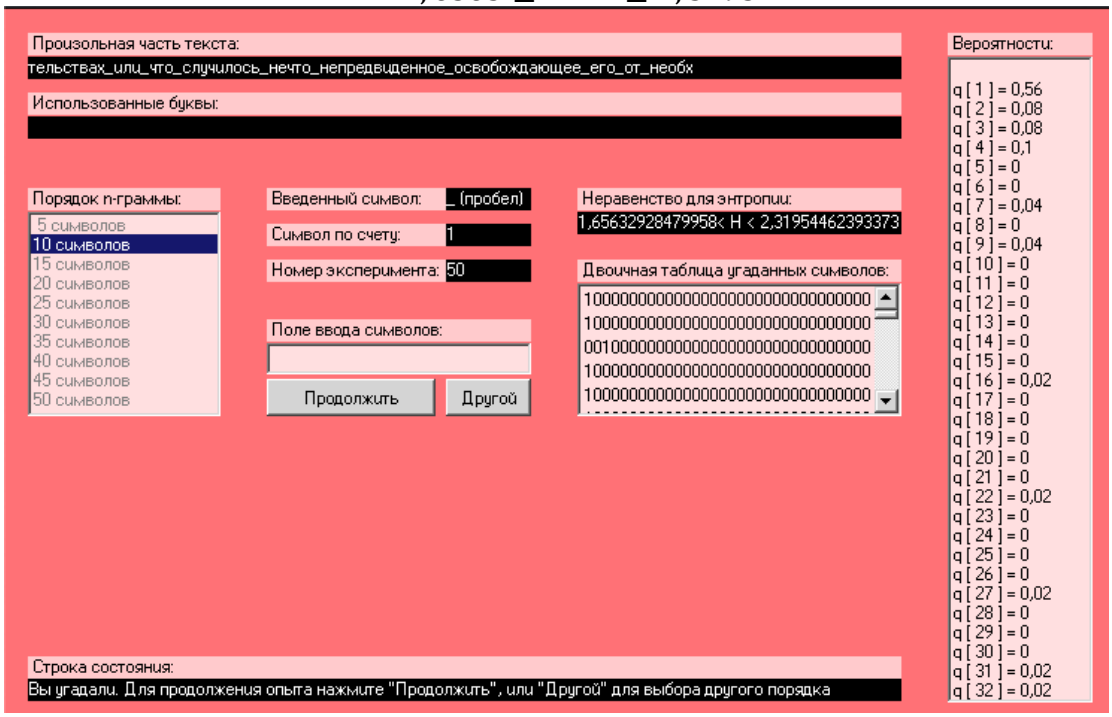
Постановка задачі

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини, де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено все пробіли.
2. За допомогою CoolPinkProgram оцінити значення $H^{(10)}, H^{(20)}, H^{(30)}$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

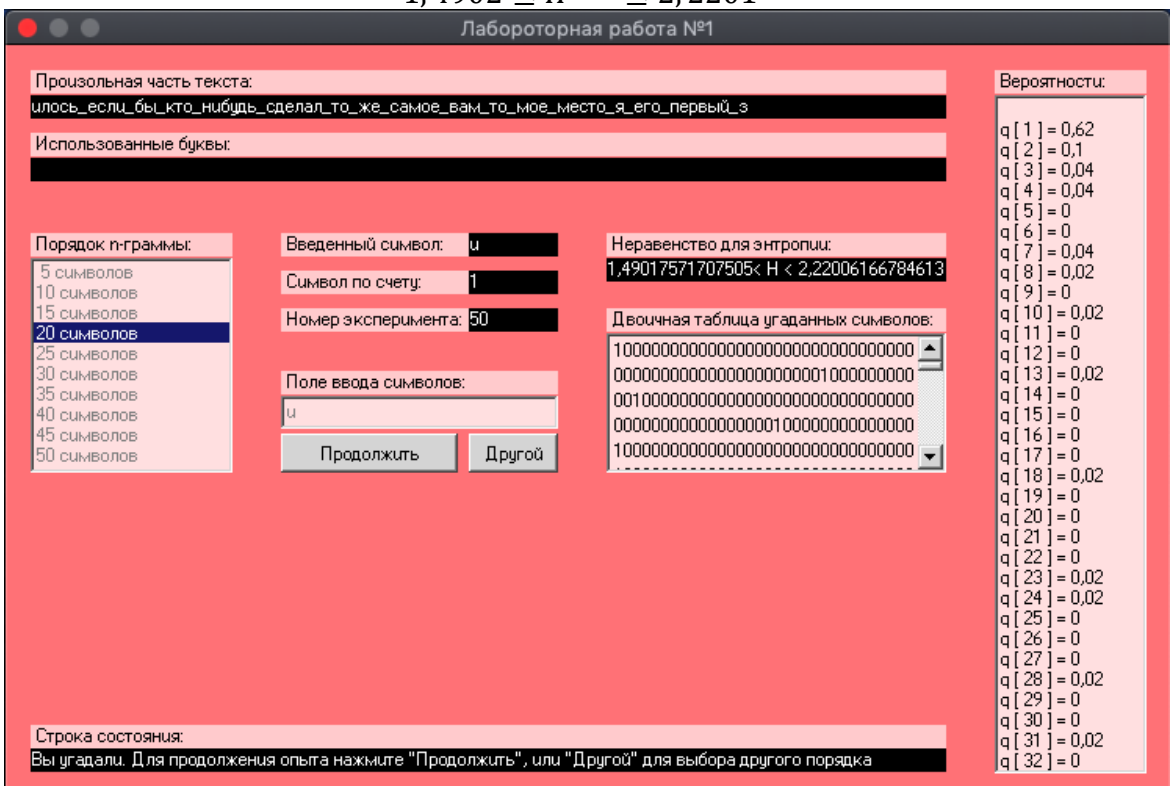
Хід роботи

1. Програмна реалізація на github: https://github.com/OleksandraMednikova/fi-labs-2022/tree/master/cp_1/mednikova_fi-93_cp1
2. Частоти букв та біграм(з перетином та без) з пробілом та без у алфавіти наведені у файлі result.txt
   Значення H_1 та H_2:
   H_1 у тексті з пробілом: 4.383831209580581
   H_2 біграм без перетину у тексті з пробілом: 3.985189847546685
   H_2 біграм з перетином у тексті з пробілом: 3.9856911068555023
   H_1 у тексті без пробілу: 4.46869156754106
   H_2 біграм без перетину у тексті без пробілу: 4.1520236191862425
   H_2 біграм з перетином у тексті без пробілу: 4.152704651455609
3. За допомогою CoolPinkProgram були отримані такі значення:

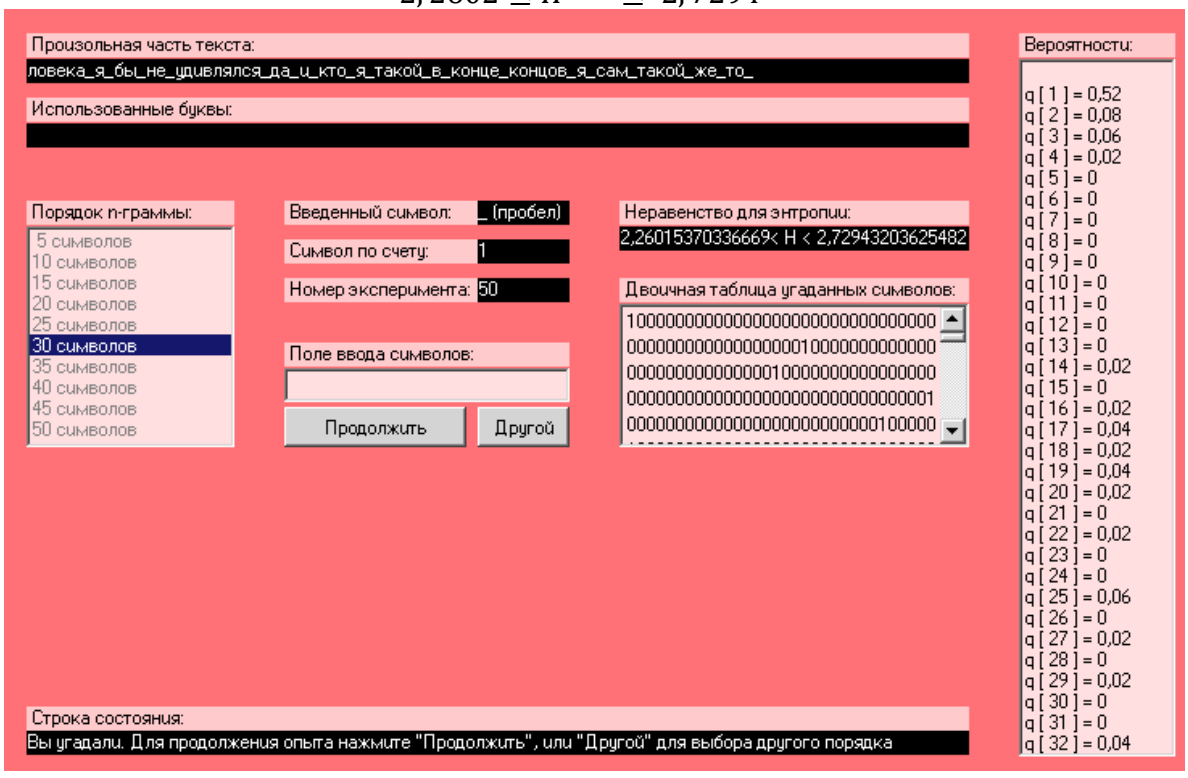
$$1,6563 \leq H^{(10)} \leq 2,3195$$



$$1,4902 \leq H^{(20)} \leq 2,2201$$



$$2,2602 \leq H^{(30)} \leq 2,7294$$



4. Оцінка надлишковості

$$H_0 = \log_2 32 = 5 \text{ (алфавіт з пробілом)}$$

$H_{inf} = H^{(10)}$, тоді $1 - \frac{1,6563}{5} \leq R \leq 1 - \frac{2,3195}{5}$, тобто $0,536 \leq R \leq 0,669$

$H_{inf} = H^{(20)}$, тоді $1 - \frac{1,4902}{5} \leq R \leq 1 - \frac{2,2201}{5}$, тобто $0,556 \leq R \leq 0,702$

$H_{inf} = H^{(30)}$, тоді $1 - \frac{2,2602}{5} \leq R \leq 1 - \frac{2,7294}{5}$, тобто $0,454 \leq R \leq 0,548$

$H_{inf} = H_1$, тоді $R = 1 - \frac{4,3838}{5} = 0, 12324$

$H_{inf} = H_{21}$, тоді $R = 1 - \frac{3,9852}{5} = 0, 2029$

$H_{inf} = H_{22}$, тоді $R = 1 - \frac{3,9856}{5} = 0, 2028$

$H_0 = \log_2 31 = 4,9542$ (алфавіт без пробілу)

$H_{inf} = H_1$, тоді $R = 1 - \frac{4,4686}{4,9542} = 0, 100$

$H_{inf} = H_{21}$, тоді $R = 1 - \frac{4,1520}{4,9542} = 0, 162$

$H_{inf} = H_{22}$, тоді $R = 1 - \frac{4,1527}{4,9542} = 0, 162$