НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені Ігоря Сікорського»

Фізико-Технічний Інститут

Звіт

із лабораторної роботи №1

із дисципліни «Криптографія»

на тему

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:  
студент групи ФБ-13

Берчук В.В.

Київ – 2023

**Мета роботи**

Засвоєння понять 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.

**Порядок виконання роботи**

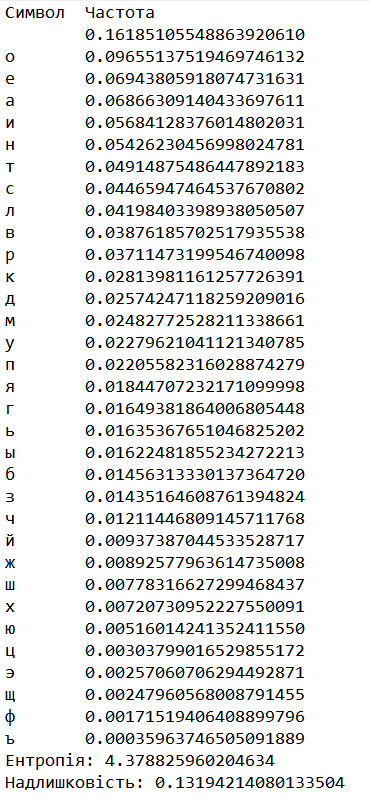
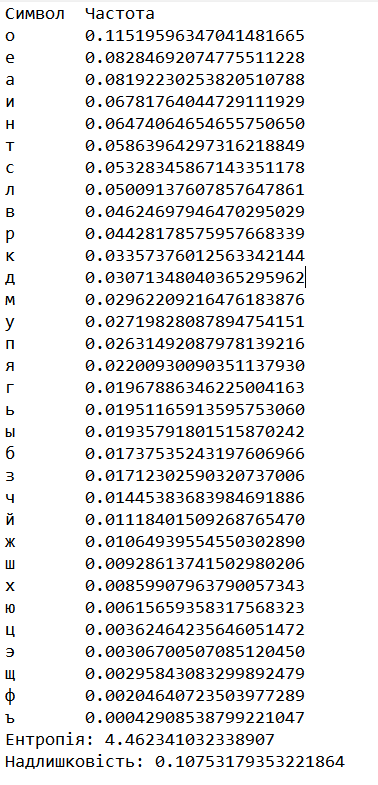
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними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

**Хід роботи**

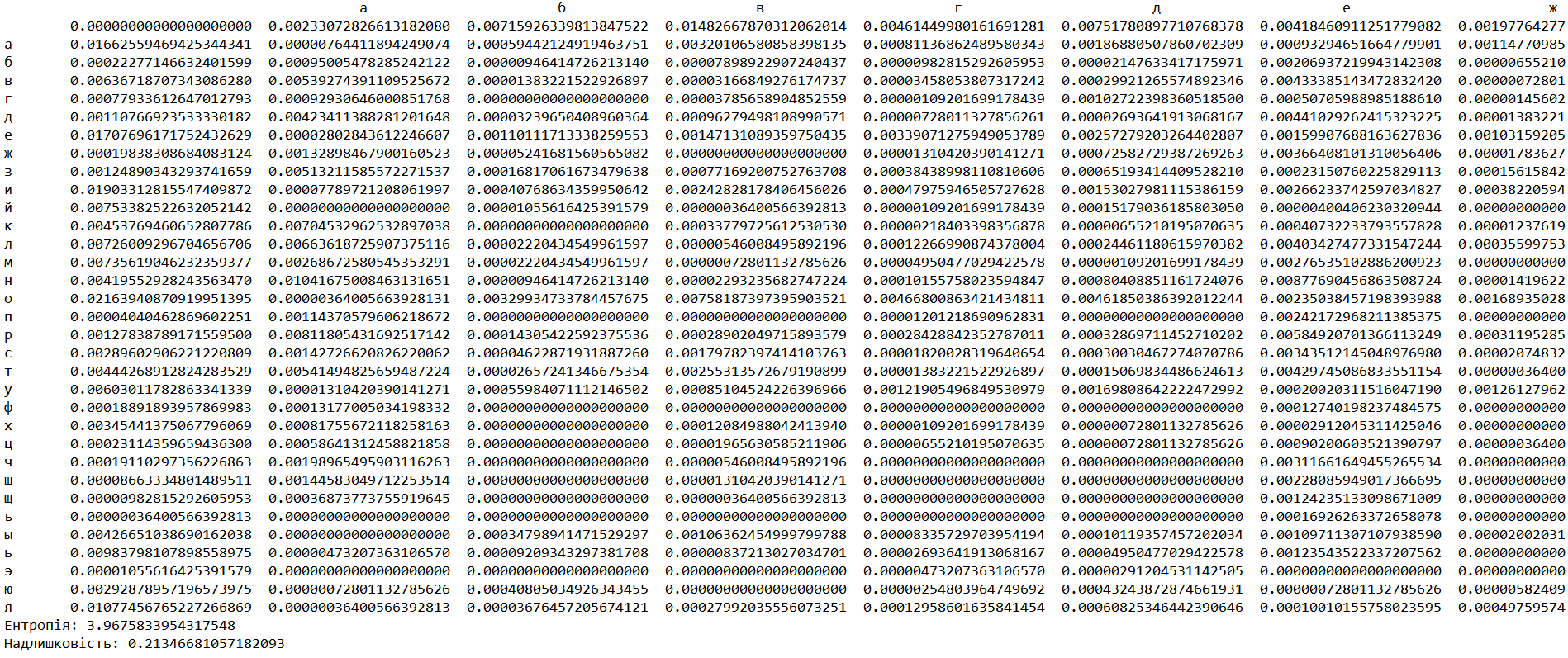
Підраховуємо частоту, ентропію та надлишковість у відфільтрованому тексті:

* text.txt – файл з текстом який містить тільки прописні букви та пробіли.
* text\_no\_spaces.txt – такий самий текст, але без пробілів.

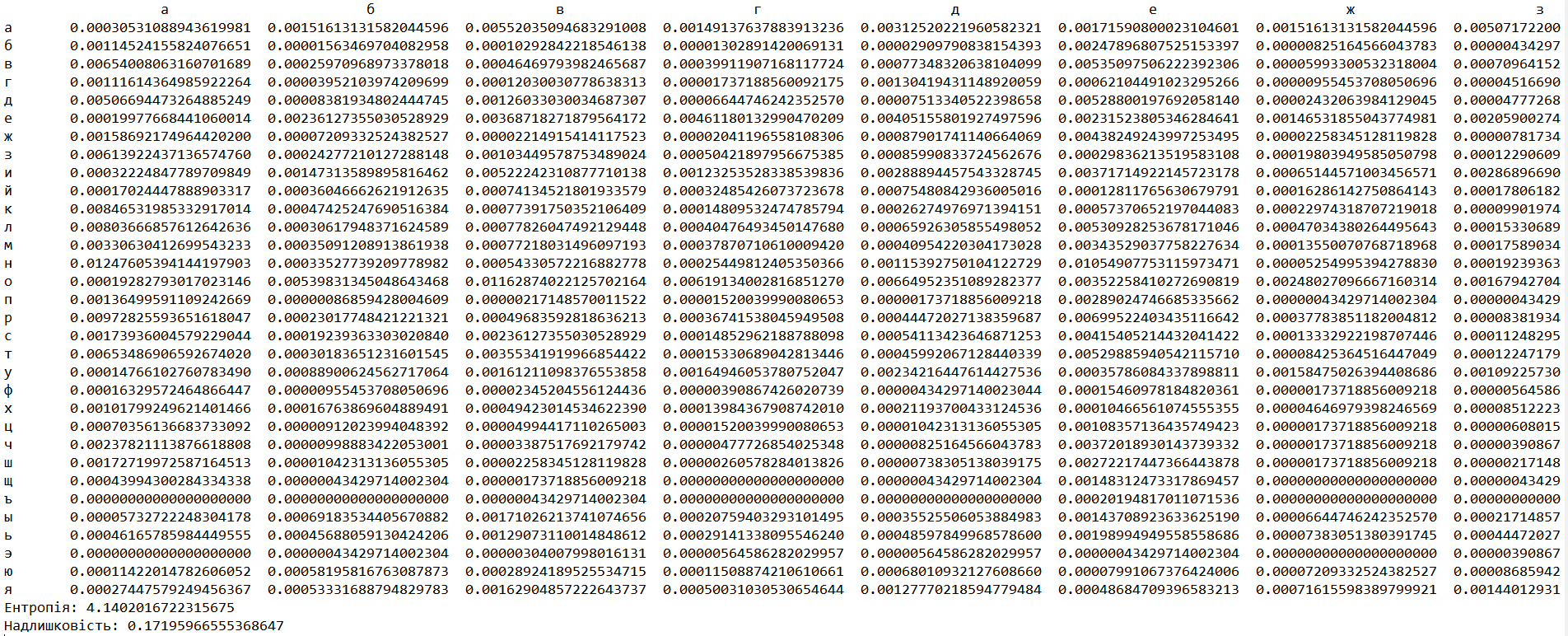
1. Розрахунки для букв у:   
     
    тексті з пробілами тексті без пробілів  
    (result1.txt) (result2.txt)

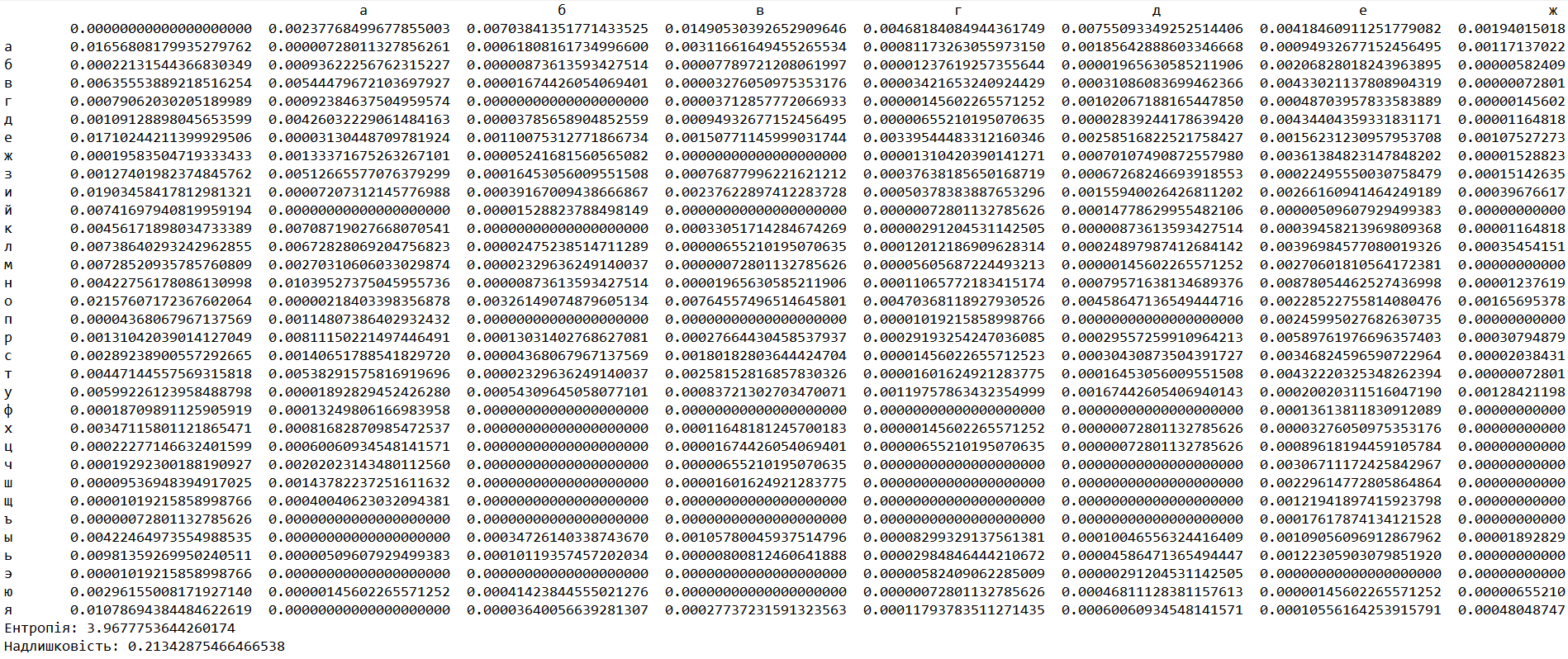
 

Розрахунки для біграм з перетинами у:

тексті з пробілами (result3.txt)  


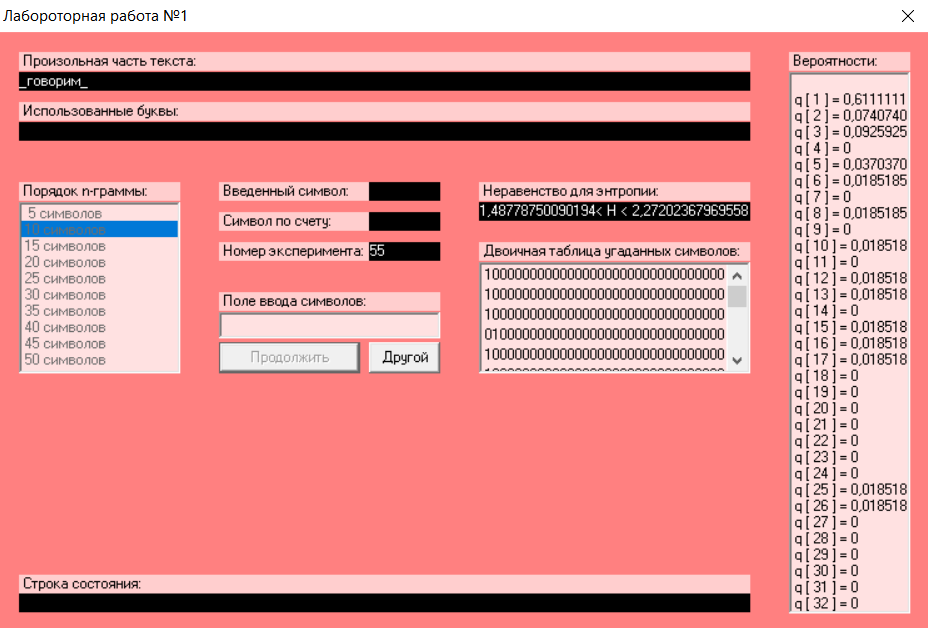
тексті без пробілів (result5.txt)

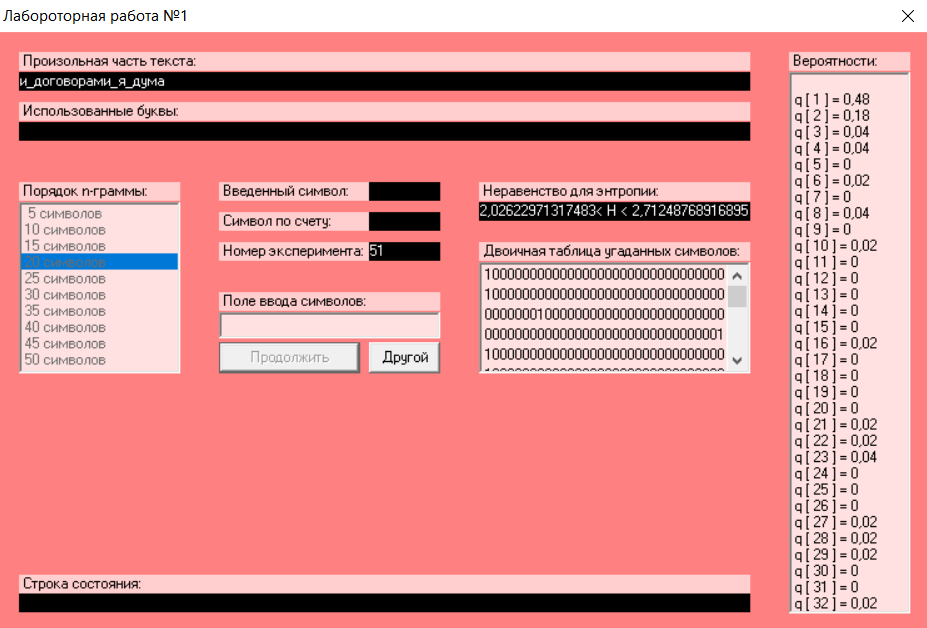


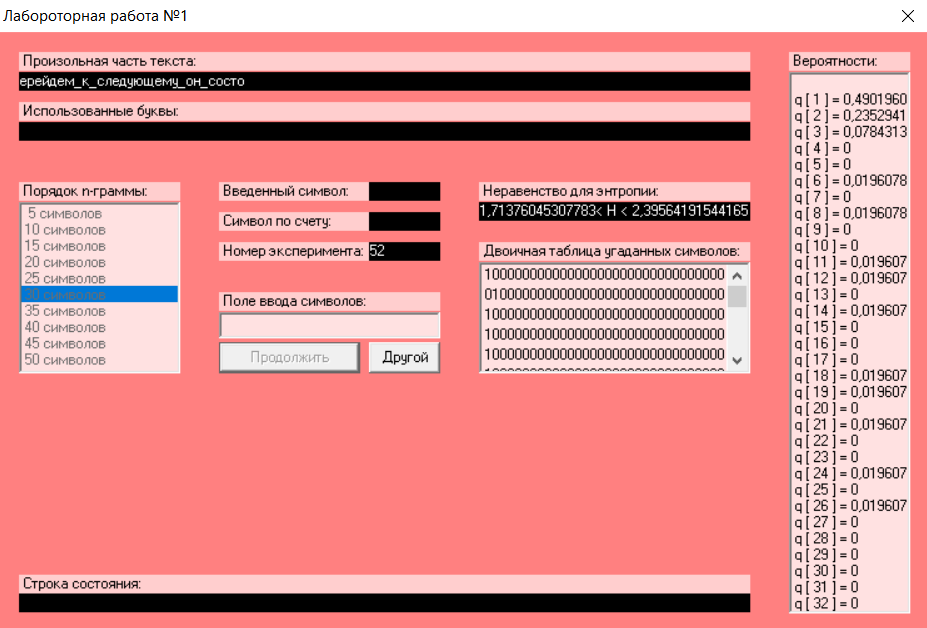
Розрахунки для біграм без перетинів у:  
тексті з пробілами (result6.txt)  


тексті без пробілів (result4.txt)



1. Оцінюємо значення *H*(10), *H*(20), *H*(30) за допомогою CoolPinkProgram   
   *H*(10): 1,48778750090194 < *H* < 2,27202367969558

  
*H*(20): 2,02622971317483 < *H* < 2,71248768916895



*H*(30): 1,71376045307783 < *H* < 2,39564191544165

Розрахуємо значення надлишковості *R*(10), *R*(20), *R*(30).

Для *R*(10):

< *R*(10) <

Для *R*(20):

< *R*(20) <

Для *R*(30):

< *R*(30) <

**Висновок**

У лабораторній роботі я вивчав ентропію та надлишковість символів у відкритому тексті, аналізував різні моделі для цих оцінок та проводив оцінку надлишковості російської мови в цих моделях. В результаті отримав практичні навички оцінки текстів.