**КРИПТОГРАФІЯ  
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1**

5 варіант

ФБ-01 Пітель Богдан  
ФБ-01 Ширий Віталій

**Експериментальна оцінка ентропії на символ джерела**

**відкритого тексту**

**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**

0. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.   
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**

1. Знаходження потрібного за розміром тексту (щонайменше 1МБ).

2. Текст «Унесенные ветром» (рос. м.)

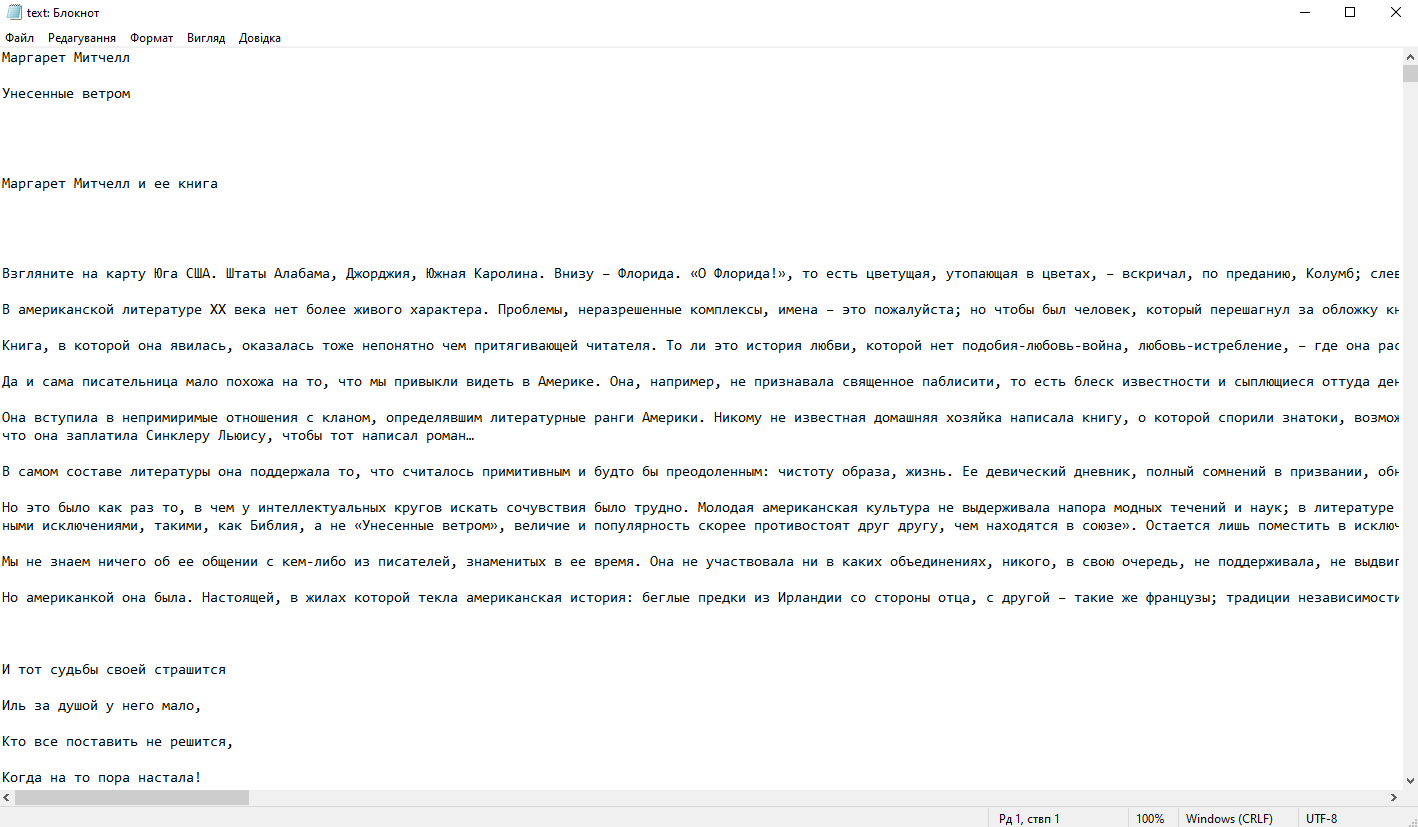
3. Почистимо наш текст від символів, які відсутні у алфавіті.

4. Підрахуємо кількість частот букв і частот біграм в тексті.

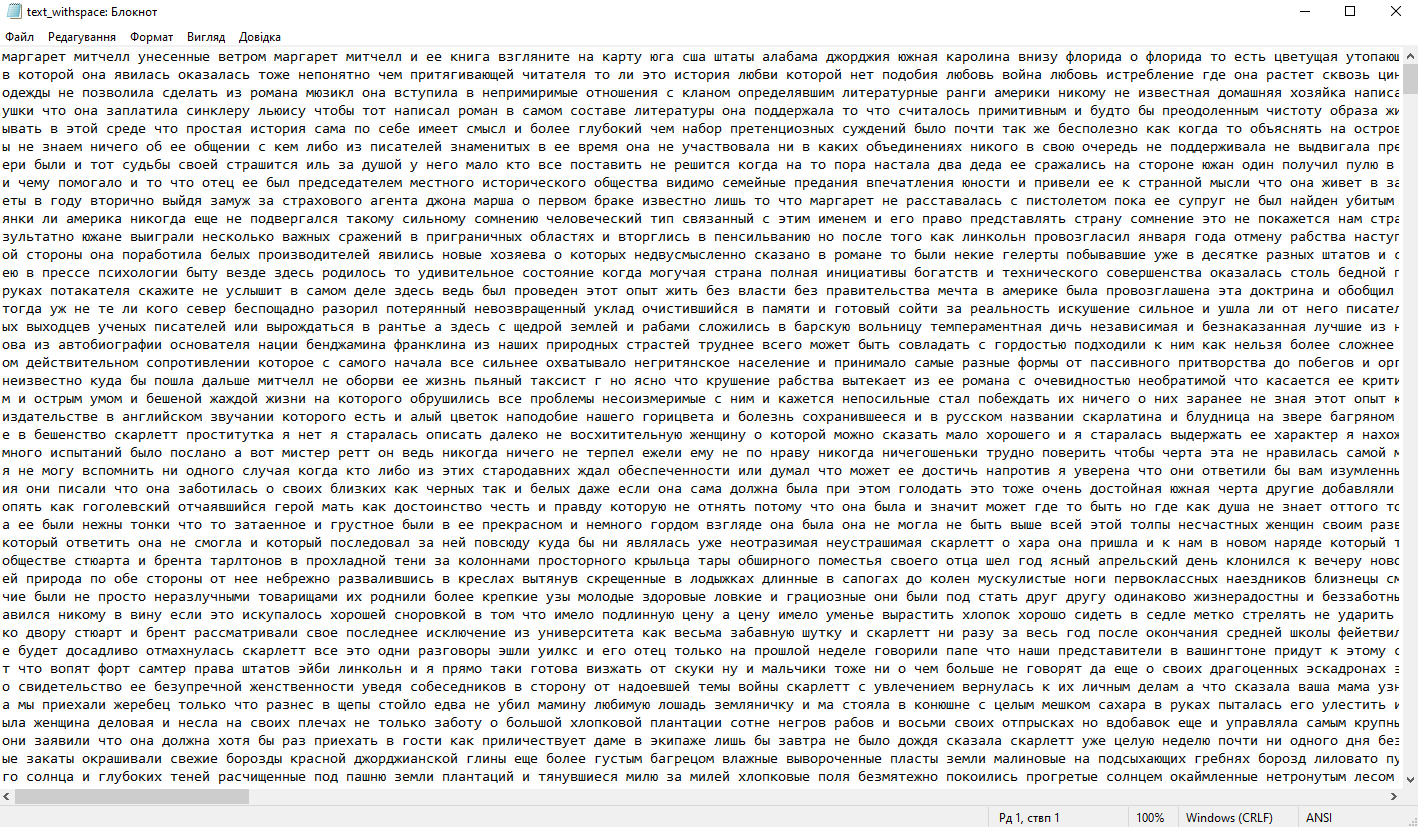
5. Підрахуємо H1 та H2, також частоти букв та біграм.

6. За допомогою програми CoolPinkProgram оцінимо значення (10)H , (20)H , (30)H.

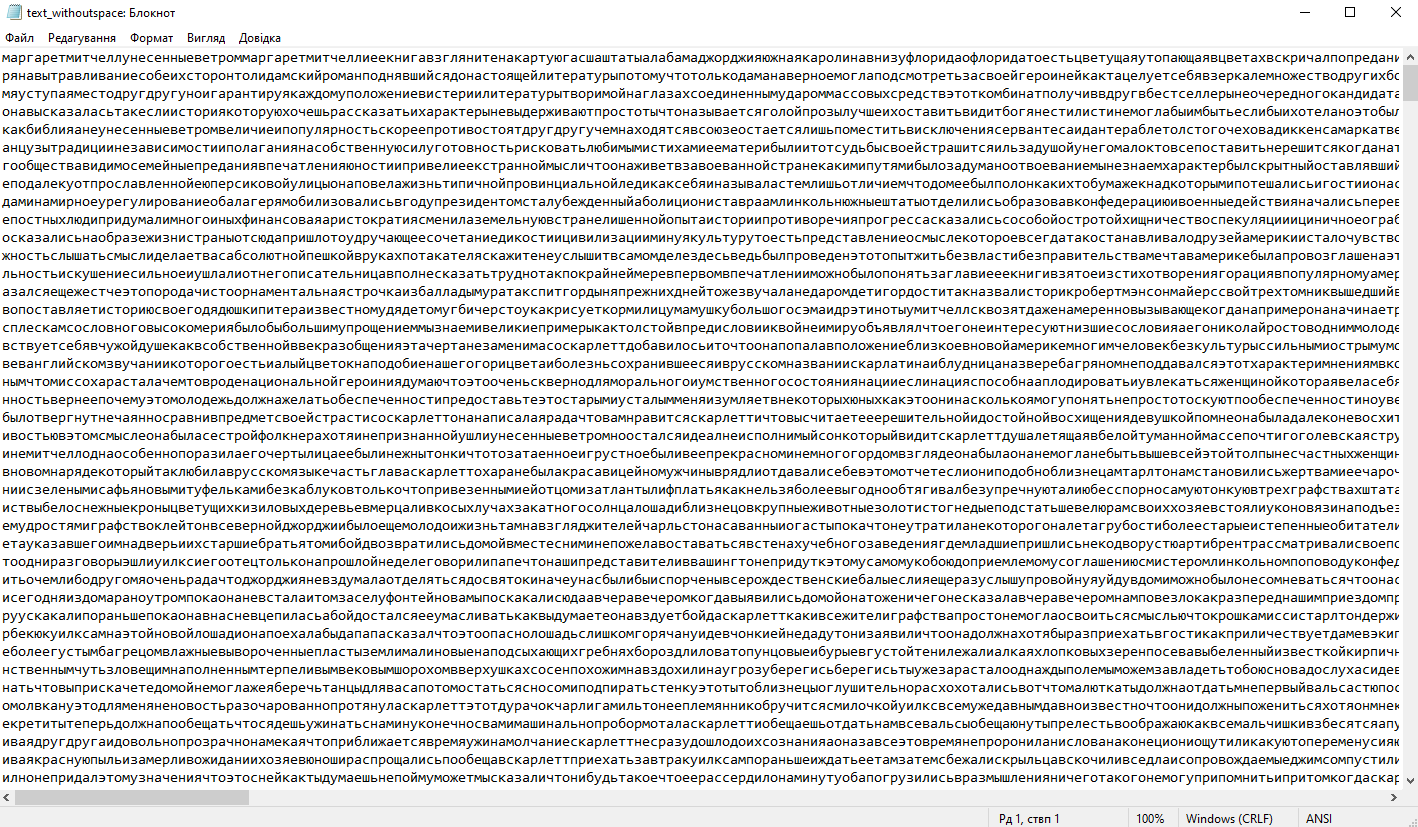
7. Оцінимо надлишковість російської мови в різних джерелах.

1-3. Текст не відредагований (1.81 МБ)  


Використаємо python код, для того, аби почистити текст від символів.  
Збережемо цей файл, як withspace.txt.



У наступному кроці приберемо з тексту пробіли, і також збережемо, але під назвою withoutspace.txt

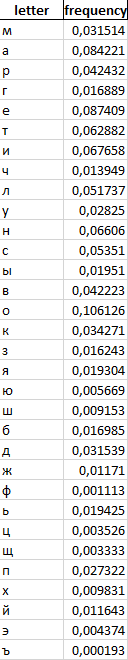


4. Підрахуємо кількість літер у тексті, частоти букв біграм.

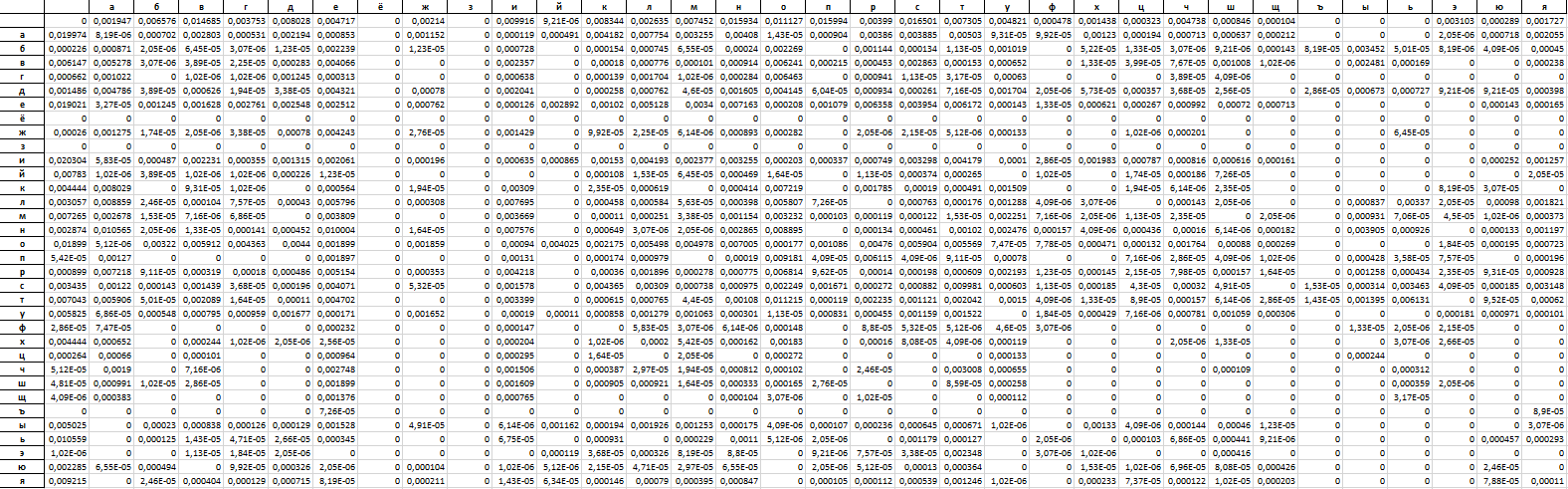
Кількість літер:



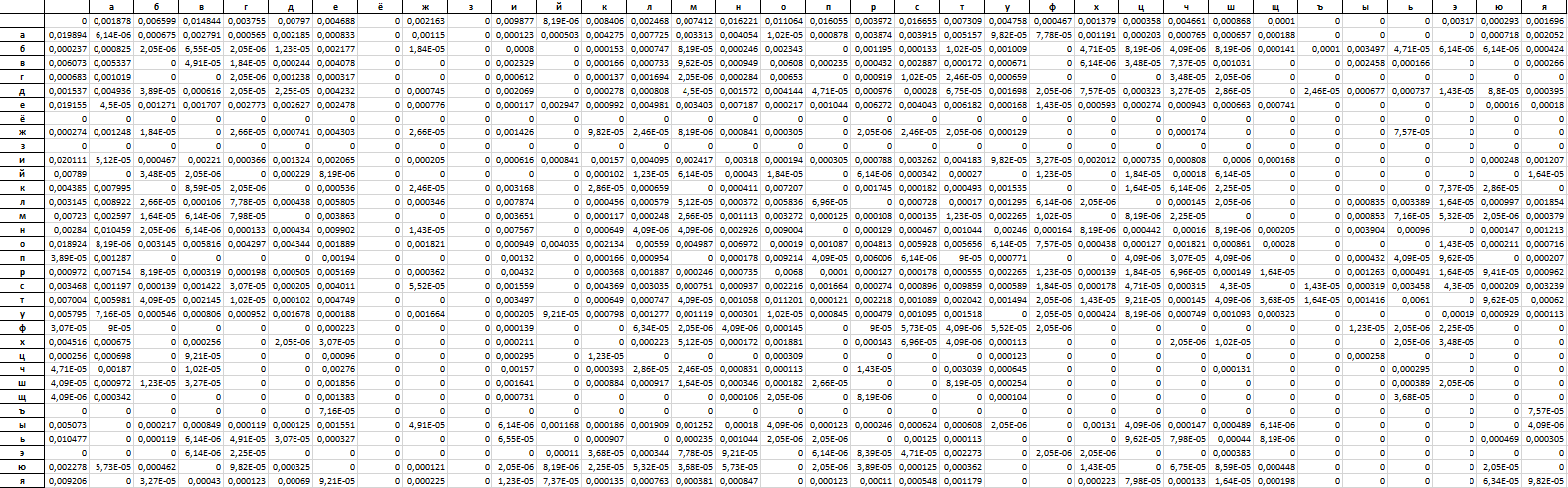
Частота, для зручності додали список у letter\_frequencies.xlsx



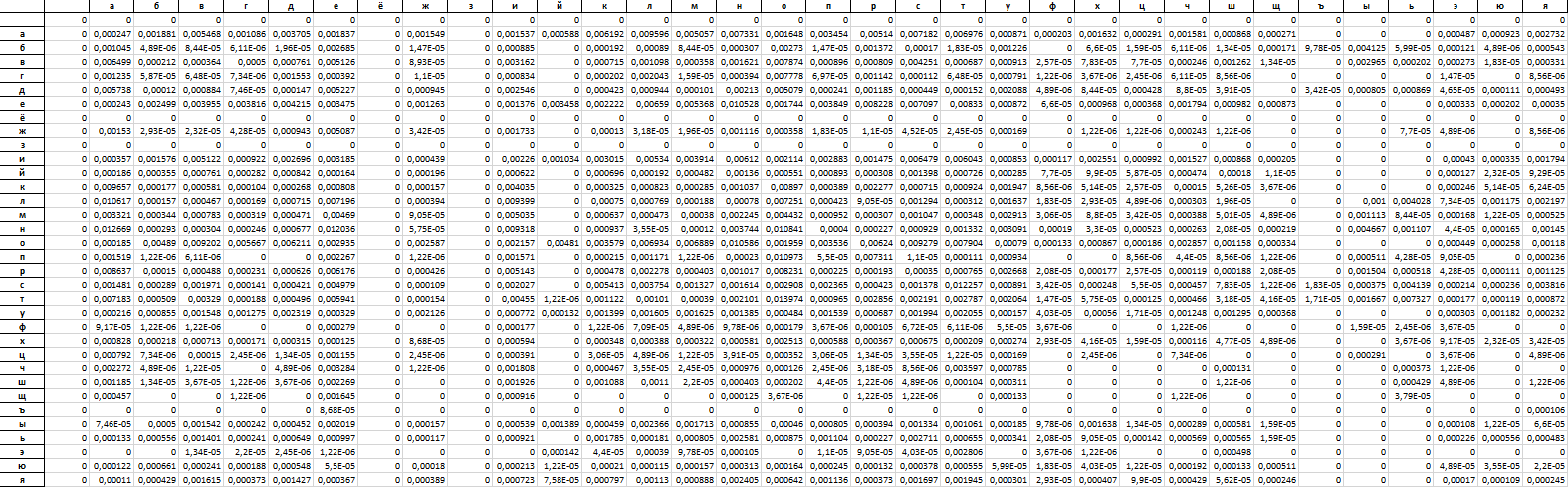
*Freqbigramwithspace:*



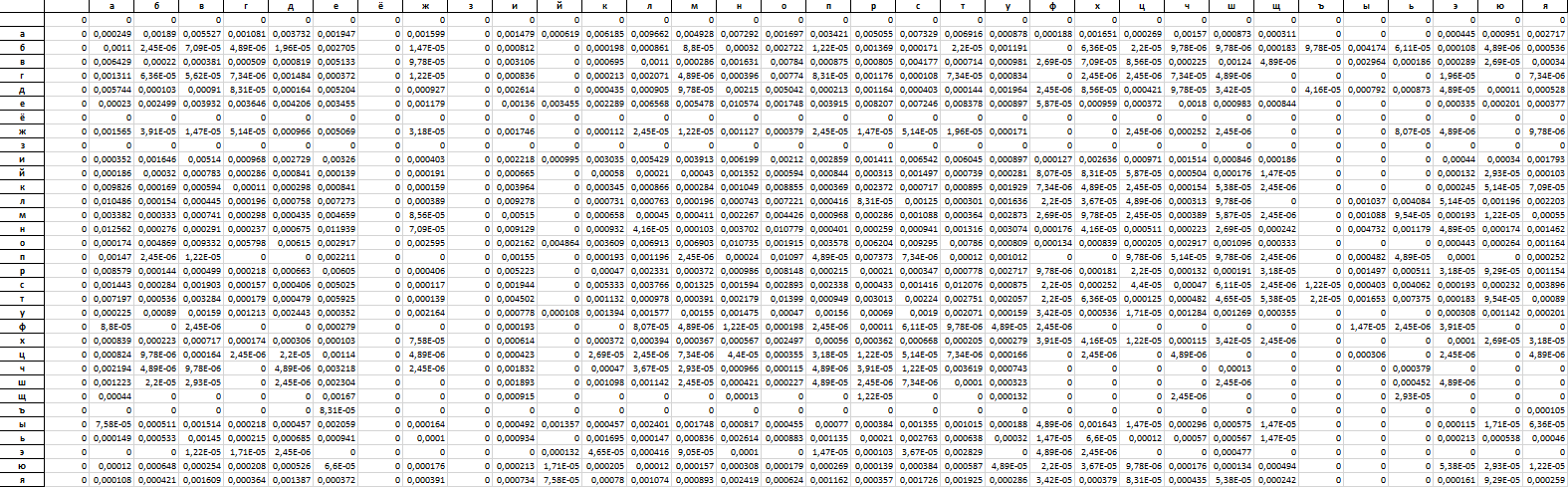
*Freqbigram\_2(withspace):*



*Freqbigramwithoutspace:*

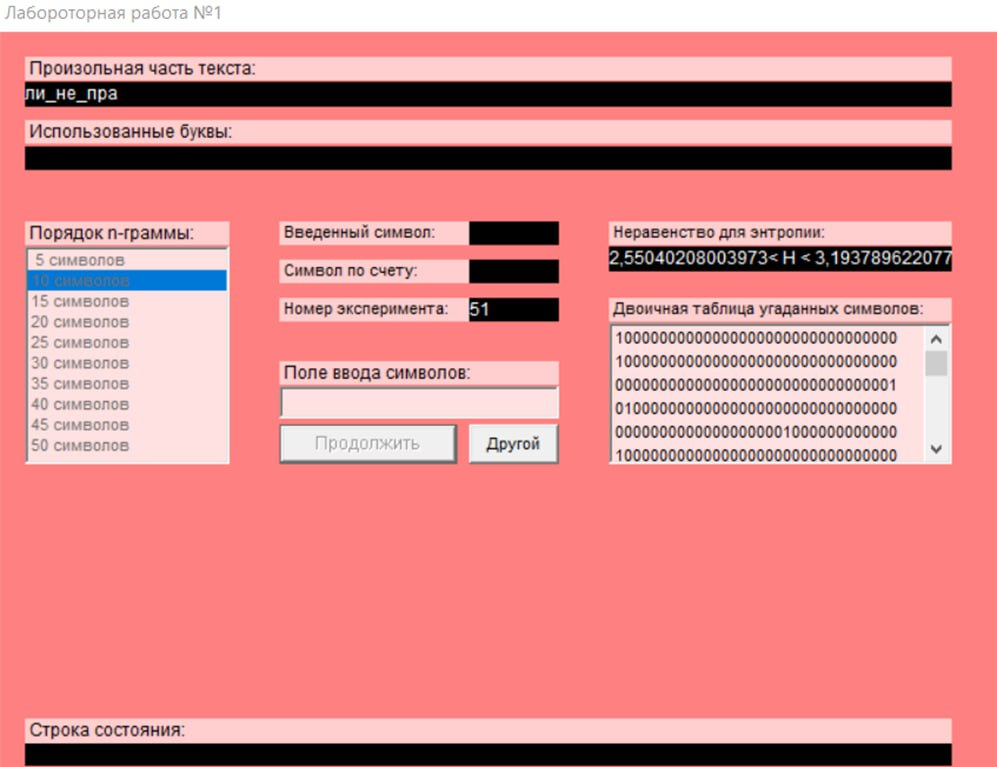


*Freqbigram\_2(withoutspace):*



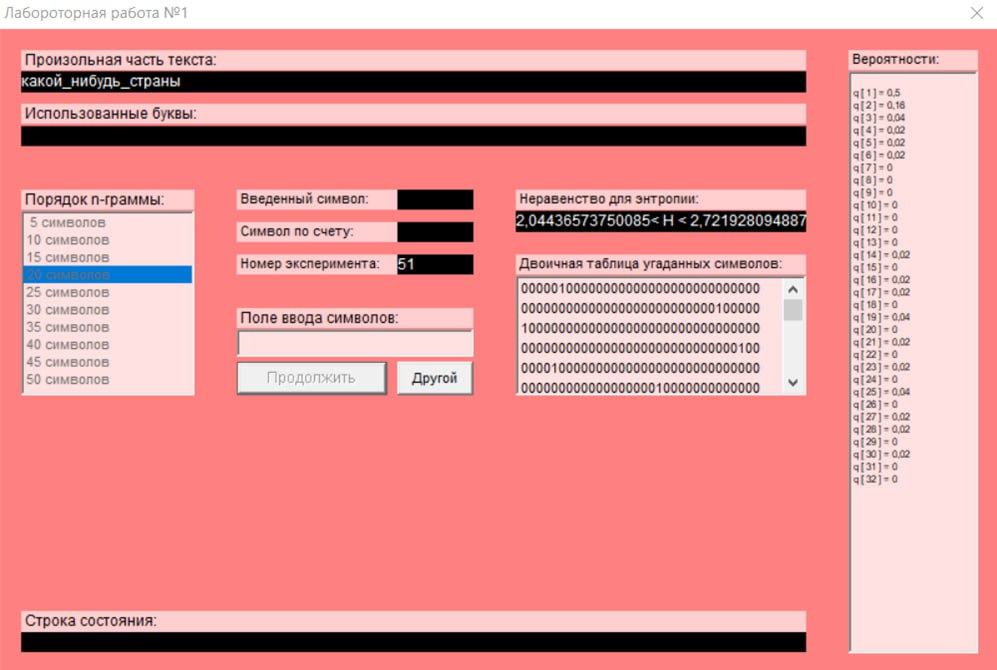
|  |  |  |
| --- | --- | --- |
|  | With space | Without space |
| H1 | 4.376164965459297 | 4.462591812647137 |
| H2 | 7.955477072143008 | 8.30571926842366 |

H(10)



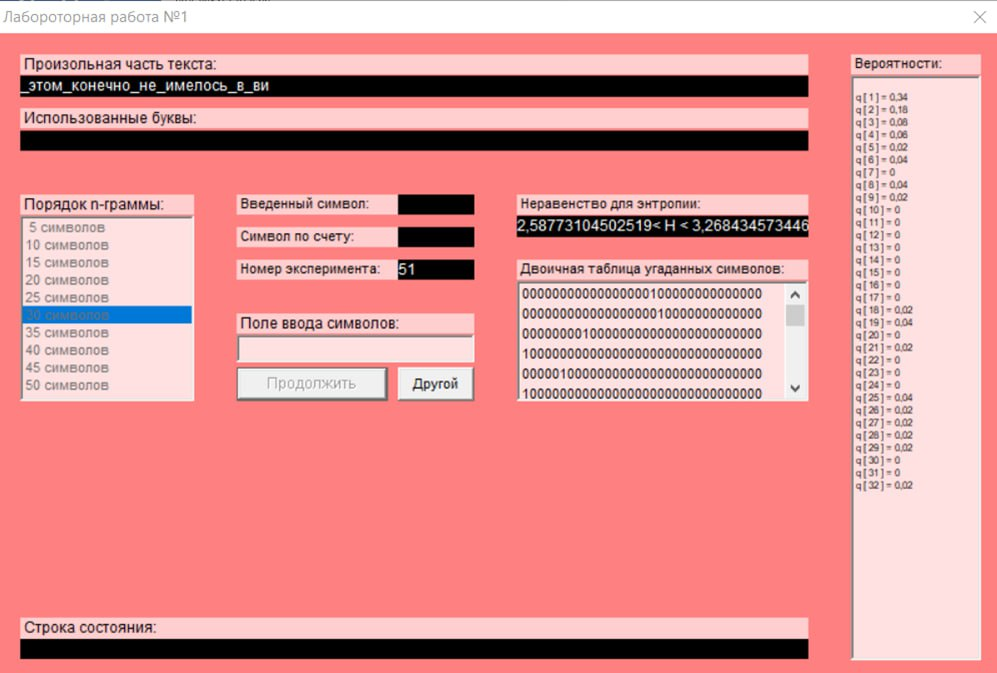
2.55 < H < 3.193

H(20)



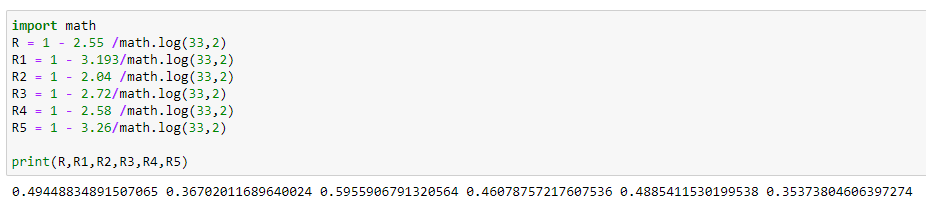
2.04 < H < 2.72

H(30)



2.58 < H < 3.26

Надлишковість:



0,494 > R(10) > 0,367

0,595 > R(20) > 0,460

0,488 > R(30) > 0,353

Висновок: виконуючи цю лабораторну роботу ми дізнались про поняття ентропії на символ джерела та його надлишковості.