**Міністерство освіти і науки України**

**Національний технічний університет України**

**"Київський політехнічний інститут імені Ігоря Сікорського"**

**Фізико-технічний інститут**

Комп’ютерний практикум №1

З дисципліни «Криптографія»

Виконали:

Студенти групи ФБ-33

Рудий А.О., Шкуропінський М.М.

Київ – 2025

**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:**

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.

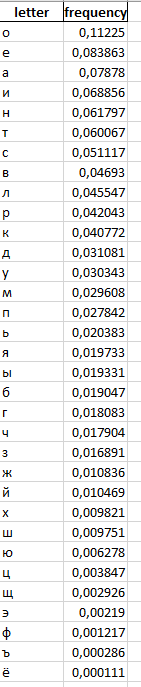
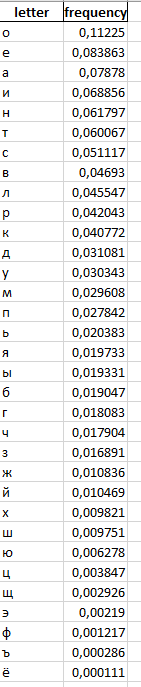
2. За допомогою програми CoolPinkProgram оцінити значення H(10) , H(20) , H(30) .

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

**Хід роботи:**

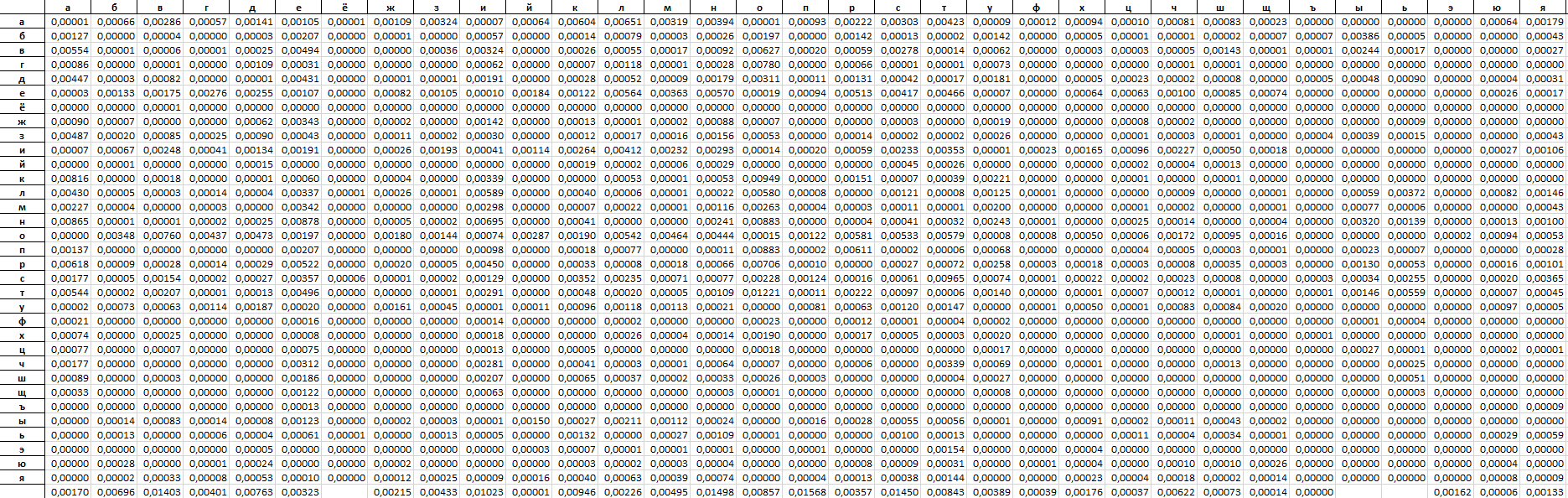
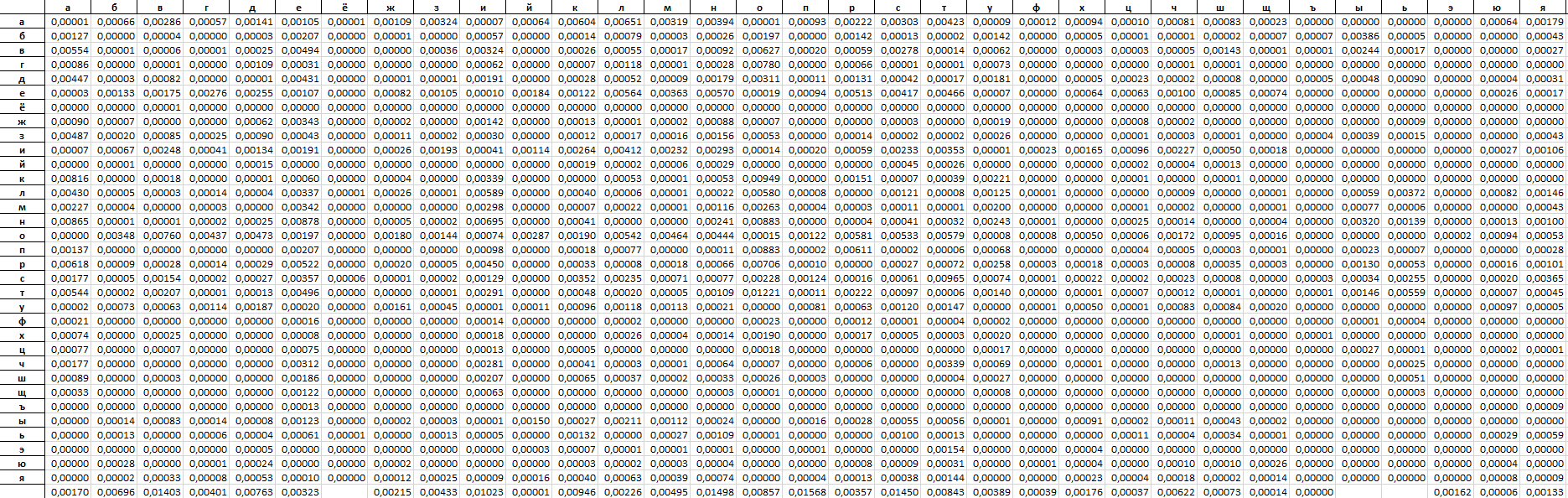
1. Для виконання лабораторної роботи використали два тексти: «Тарас Бульба» та «Мертвые души» Миколи Гоголя.

**Текст з пробілами:**

Частота букв:

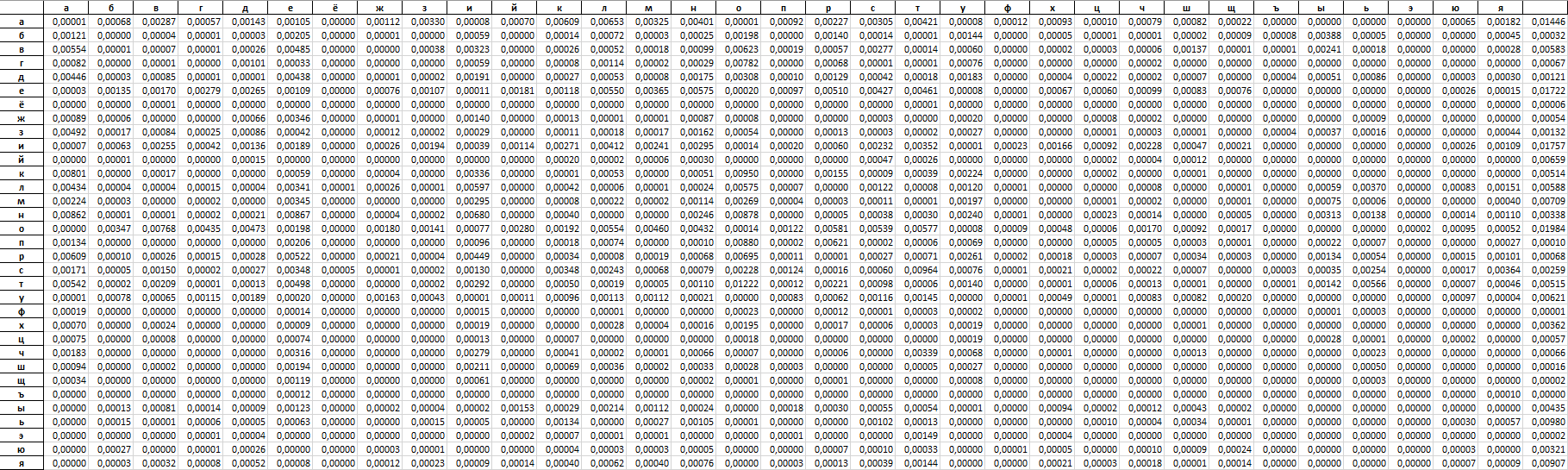
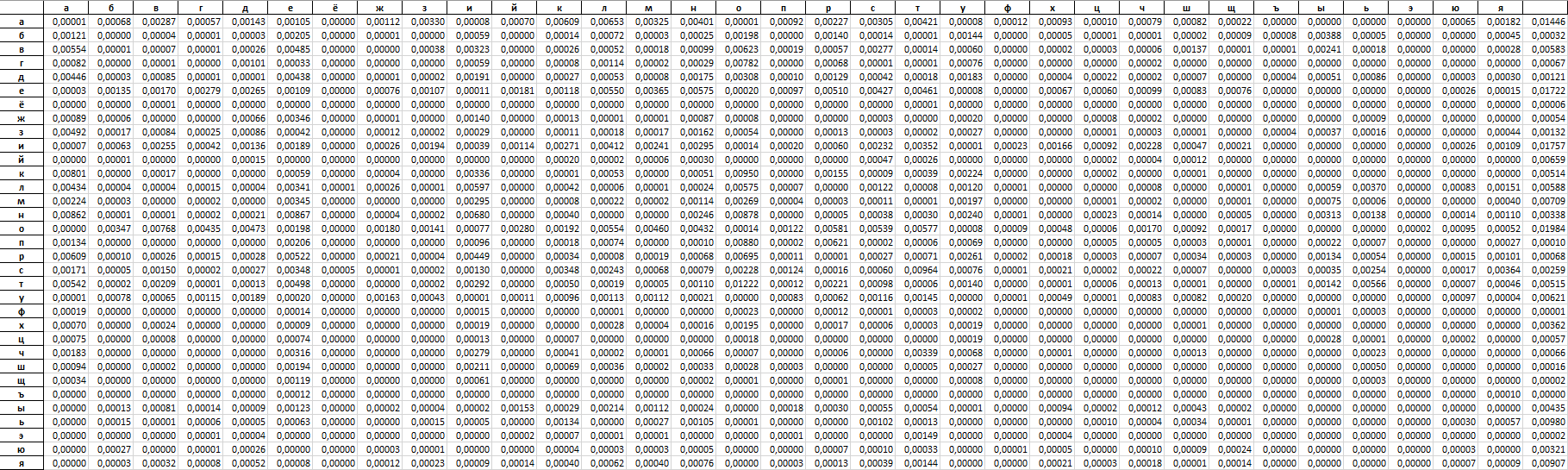
Надлишквість: 0.11306918112118969

Ентропія: 4.474028607030048

Частота біграм без перетенів H2:

Надлишковість: 0.22656859801448703

Ентропія: 3.901492815902886

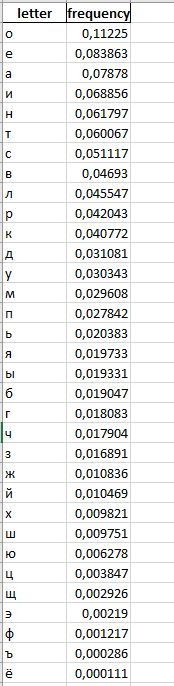
Частота біграм без перетенів H2:

Надлишковість: 0.226655806350349

Ентропія: 3.9010529026863043

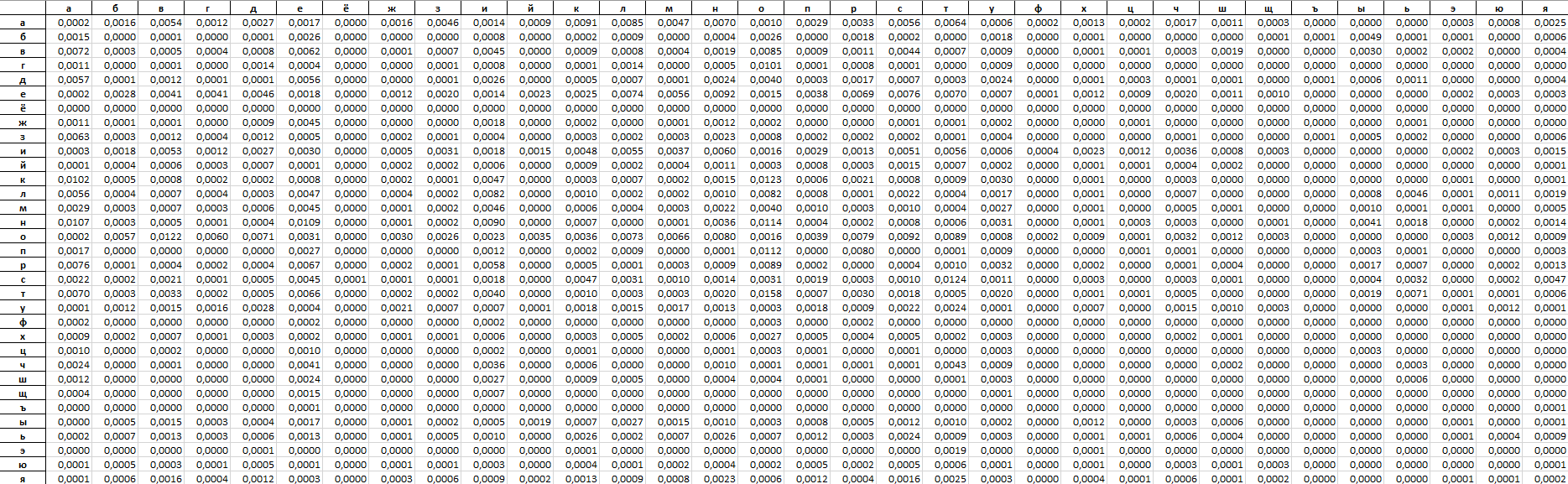
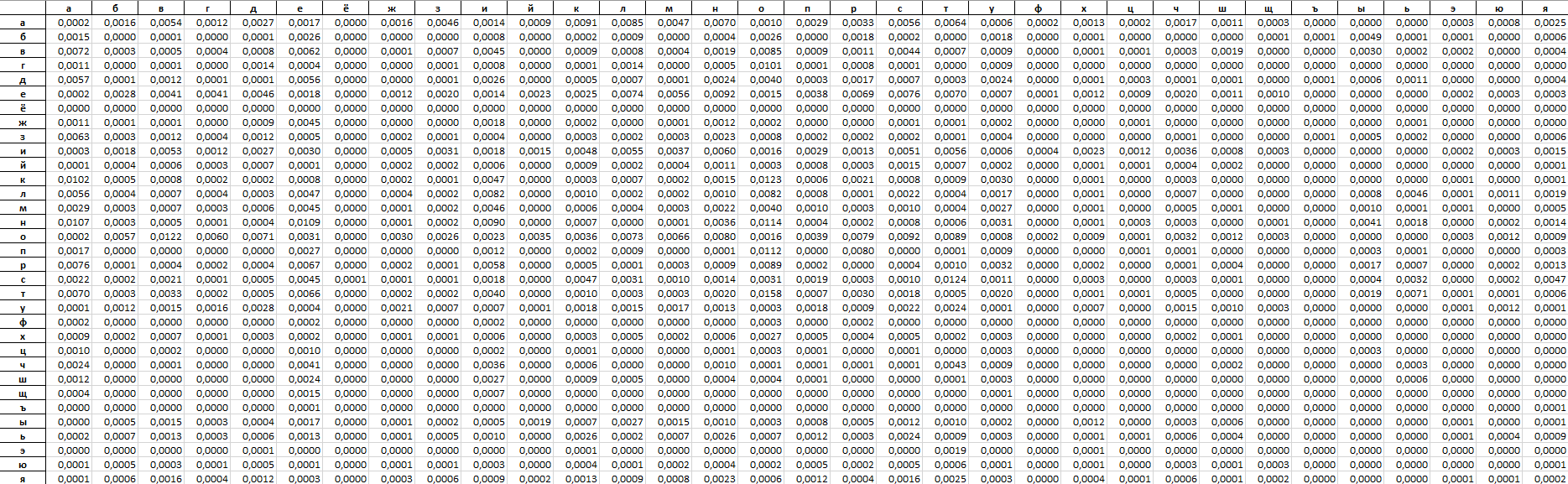
**Текст без пробілів:**

Частота букв:



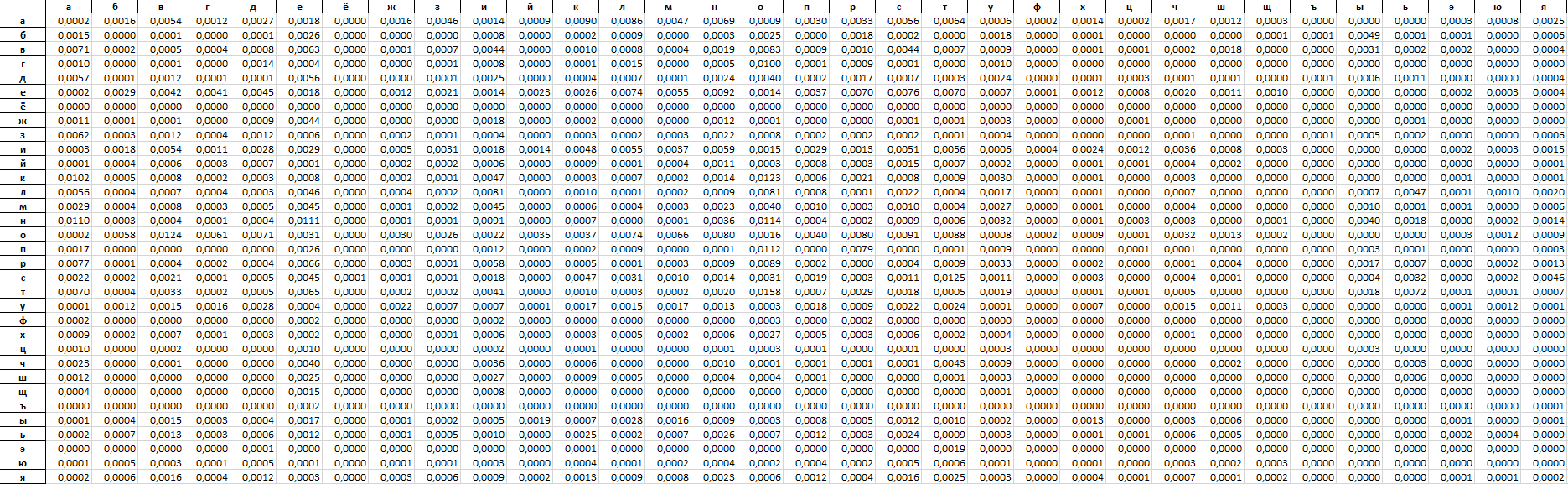
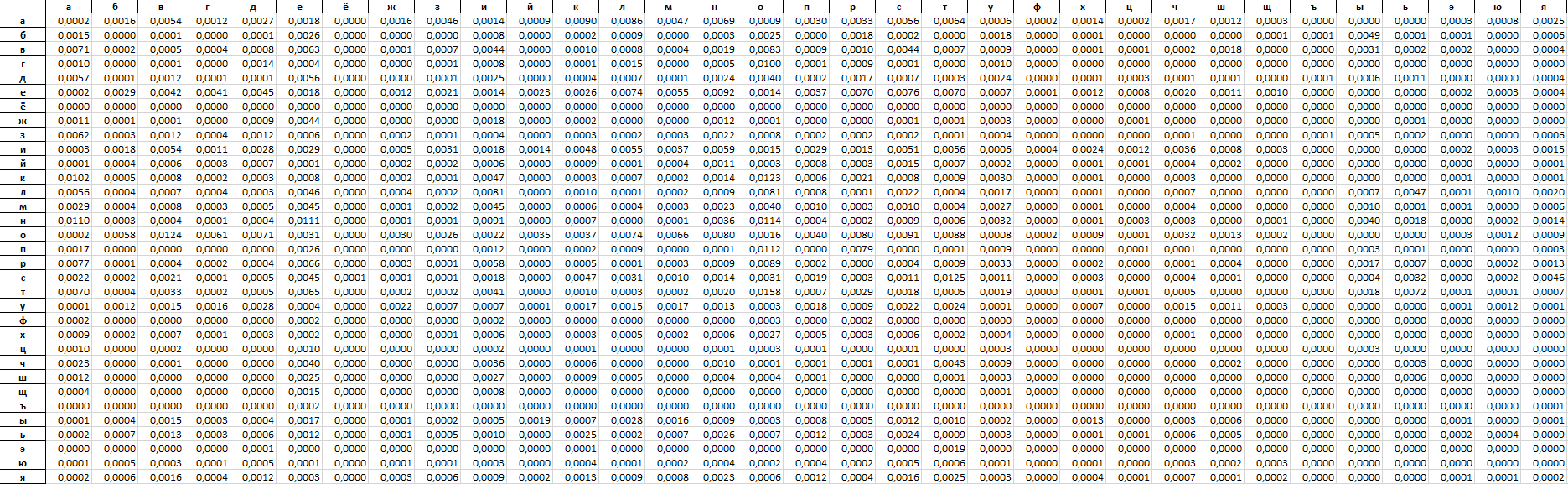
Надлишковість: 0.11306918112118969

Ентропія: 4.474028607030048

Частота біграм без перетенів H2:

Надлишковість: 0.17821406080436575

Ентропія: 4.145412159049921

Частота біграм без перетенів H2:

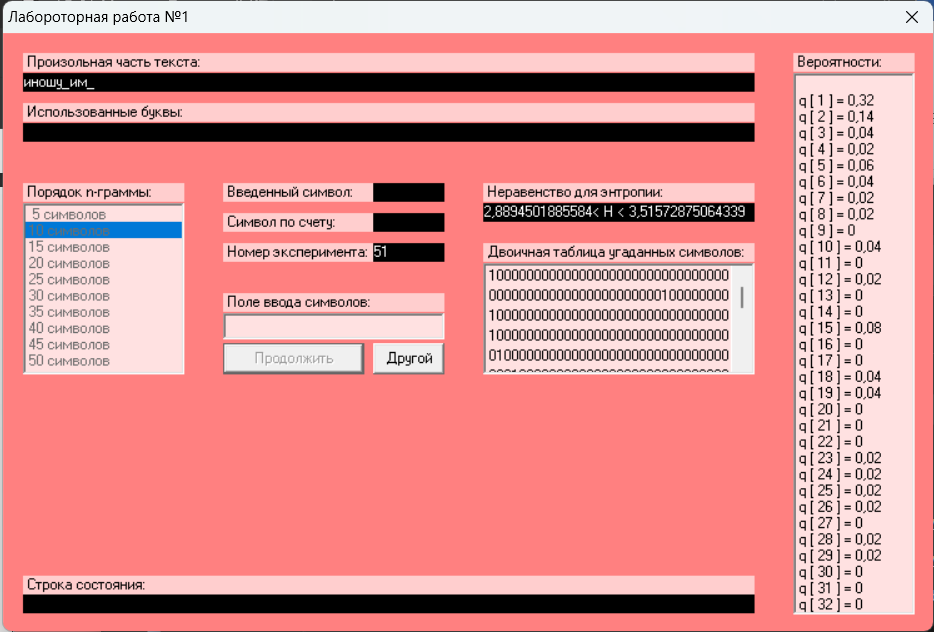
Надлишковість: 0.17817224954972866

Ентропія: 4.145623071496935

2.

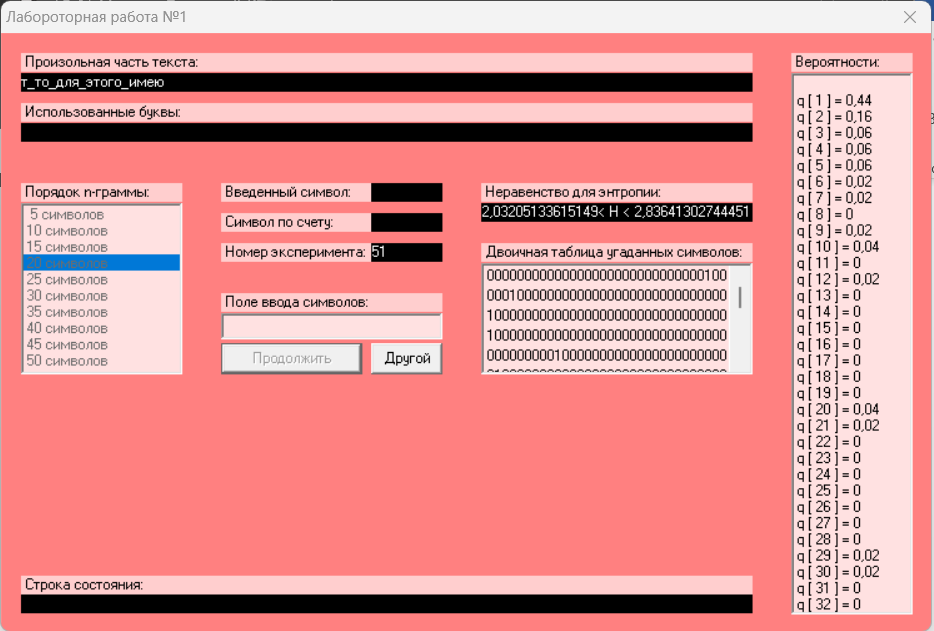
2,889<H(10)<3,515

30,3%<R(10)<42,7%



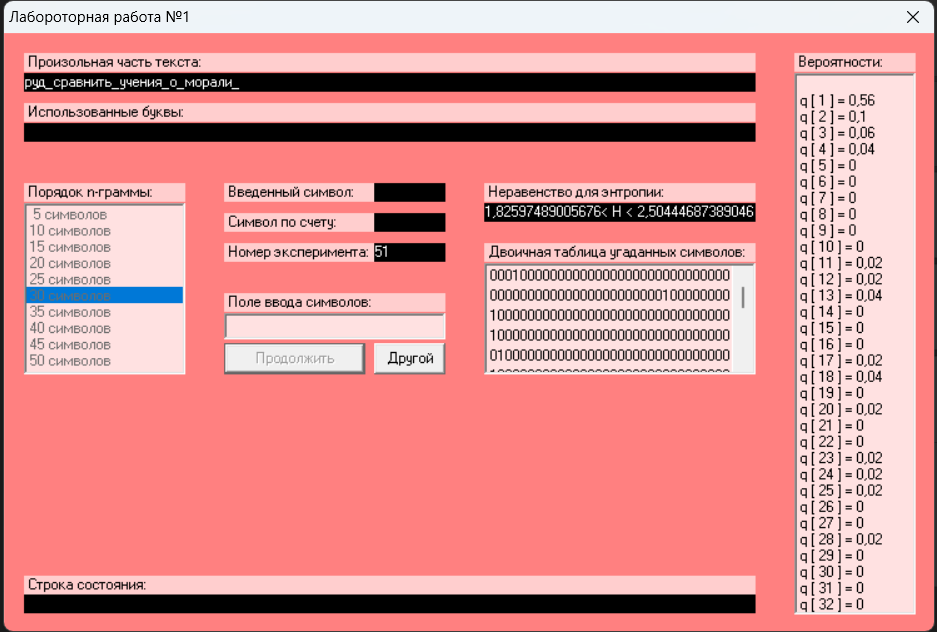
2,032<H(20)<2,836

43,7%<R(20)<59,7%



1,825<H(30)<2,504

50,3%<R(30)<63,8%



3. Надлишковість:

|  |  |  |
| --- | --- | --- |
|  | Текст з пробілами | Текст без пробілів |
| Н1 | 0.113069 | 0.113069 |
| H2 (перетин) | 0.226568 | 0.178214 |
| H2 (без перетину) | 0.226655 | 0.178172 |

**Висновок:**

У ході виконання лабораторної роботи, ми ознайомилися з поняттям ентропії, навчилися визначати її та також за допомогою її значень - надлишковість мови. Ми дізналися, що для тексту без пробілів значення виявилися меншими, також з 2 п. що зі зростанням довжини контексту ентропія зменшується.

Загалом, робота продемонструвала як на практиці вимірюється ентропія та надлишковість, і показала, що російська мова має помітну надлишковість.