**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

**КРИПТОГРАФІЯ**

**КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1**

**Експериментальна оцінка ентропії на символ джерела відкритого тексту**

Виконали:

Орлов Дмитро ФБ-14

Макуха Андрій ФБ-14

Перевірила

Селюх П. 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:**

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 (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H(1) та H(2) на тому ж тексті, в якому вилучено всі пробіли.

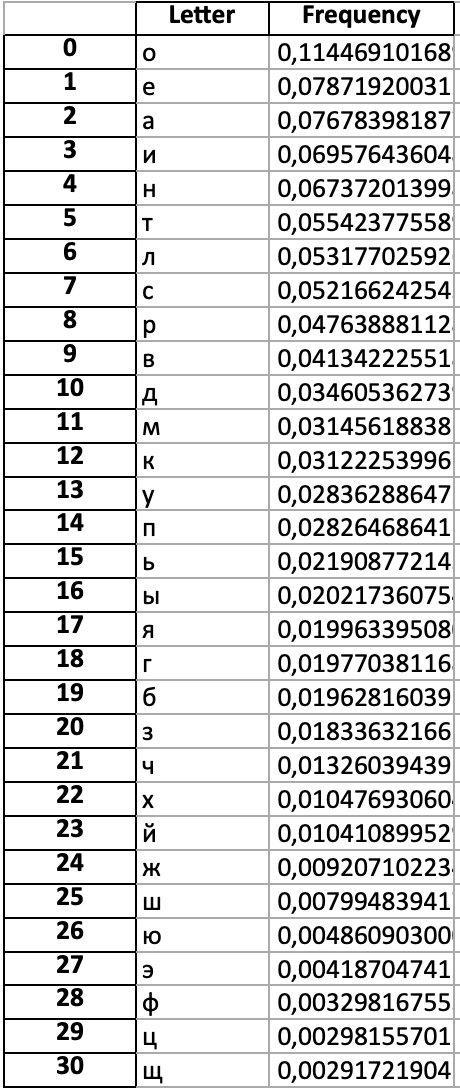
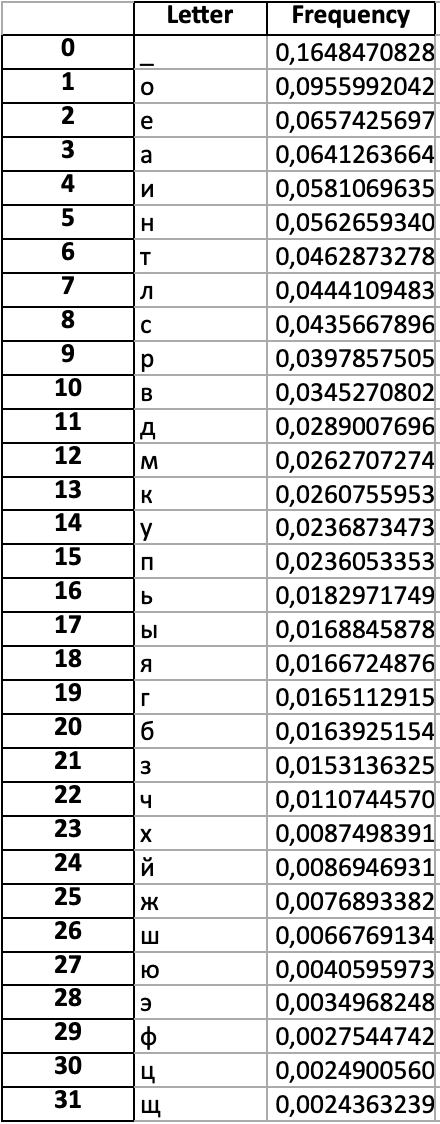
2. За допомогою програми CoolPinkProgram оцінити значення H(10), H(20), H(30)

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

**Хід роботи:**

При обрахуванні будь-якого з варіантів значень ентропії та частоти літер необхідно очистити текст від зайвих символів, а також для зручності відображення замінити пробіл на нижнє підкреслення.

Наступним етапом є обрахування ентропії для окремих літер (монограм), тому завдяки через проходження тексту рахуємо кількість літер та відповідно їх частоту для випадку з пробілом та без:

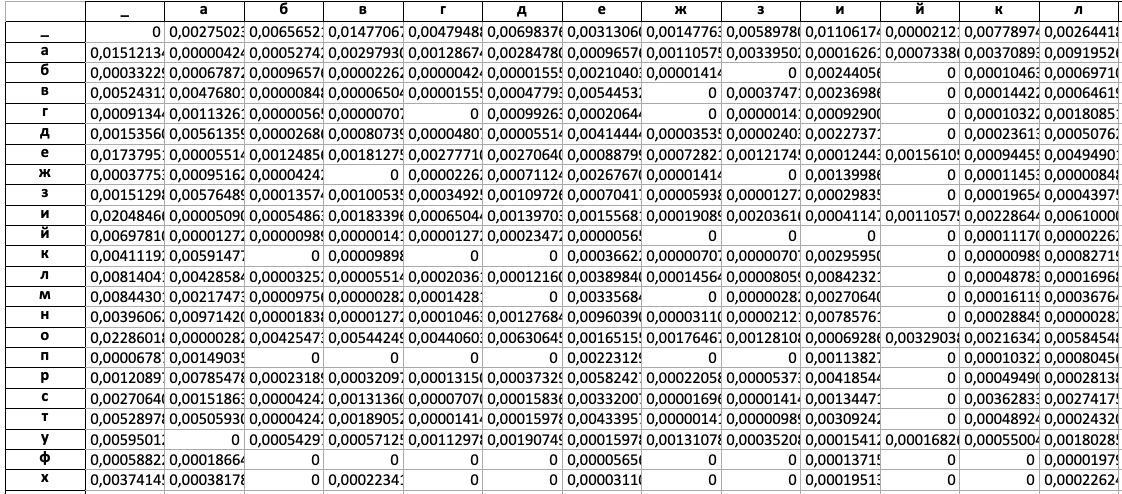


Без пробілів: 

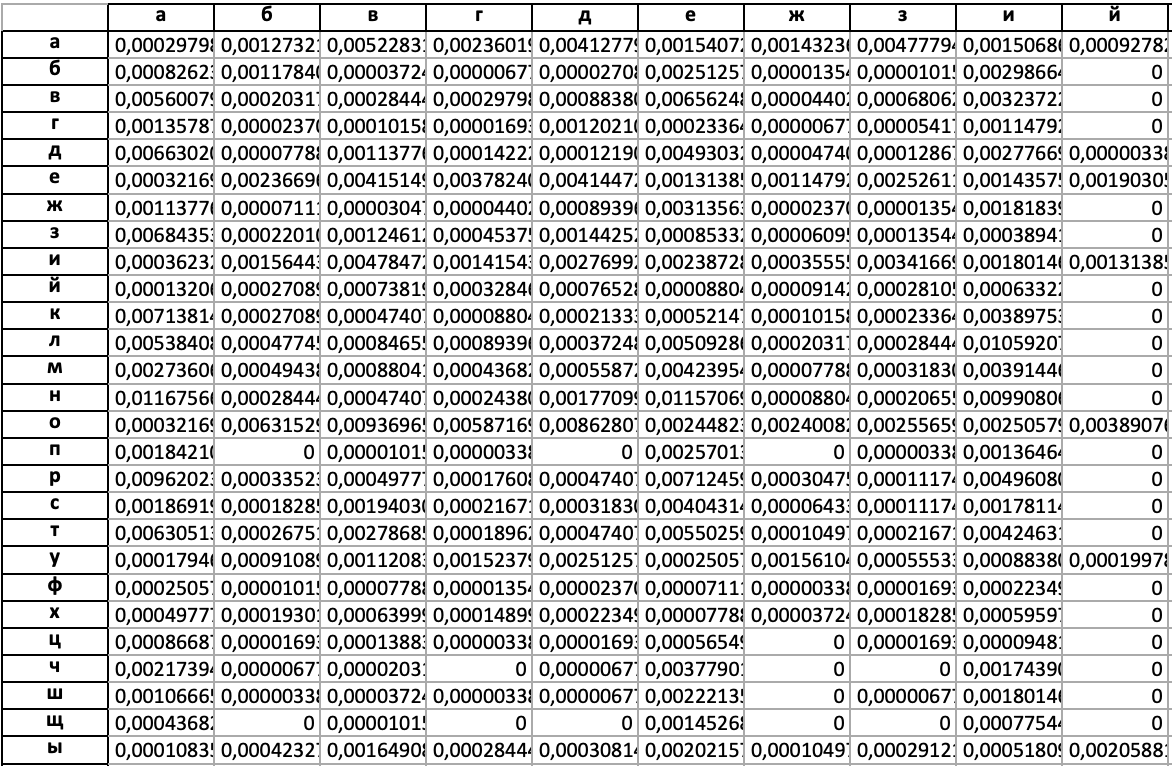
З пробілами: 

Наступним кроком рахуємо ті ж значення для 4 можливих випадків біграм (повні результати збережено в репозиторії):

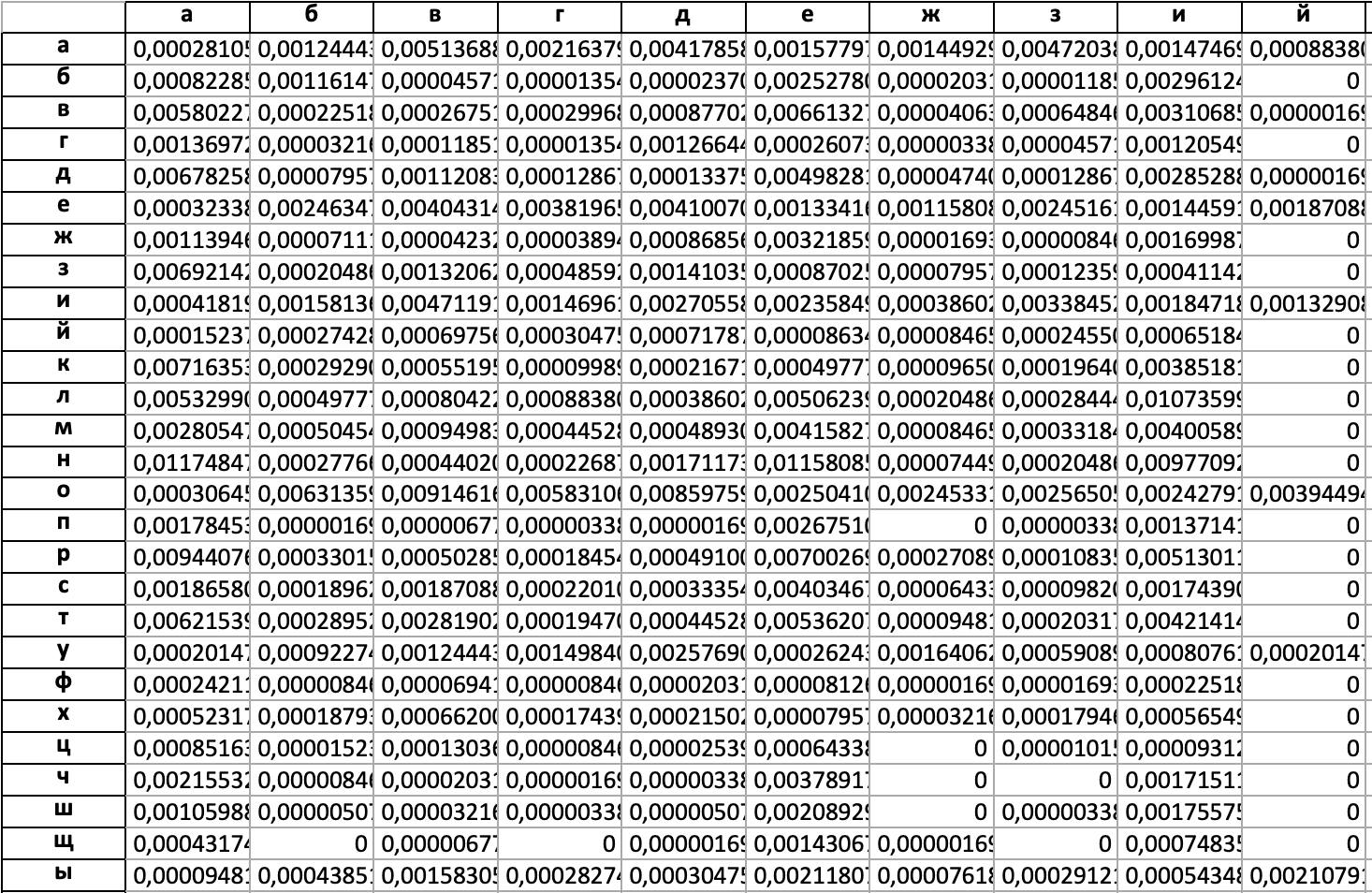
З перетином та пробілами



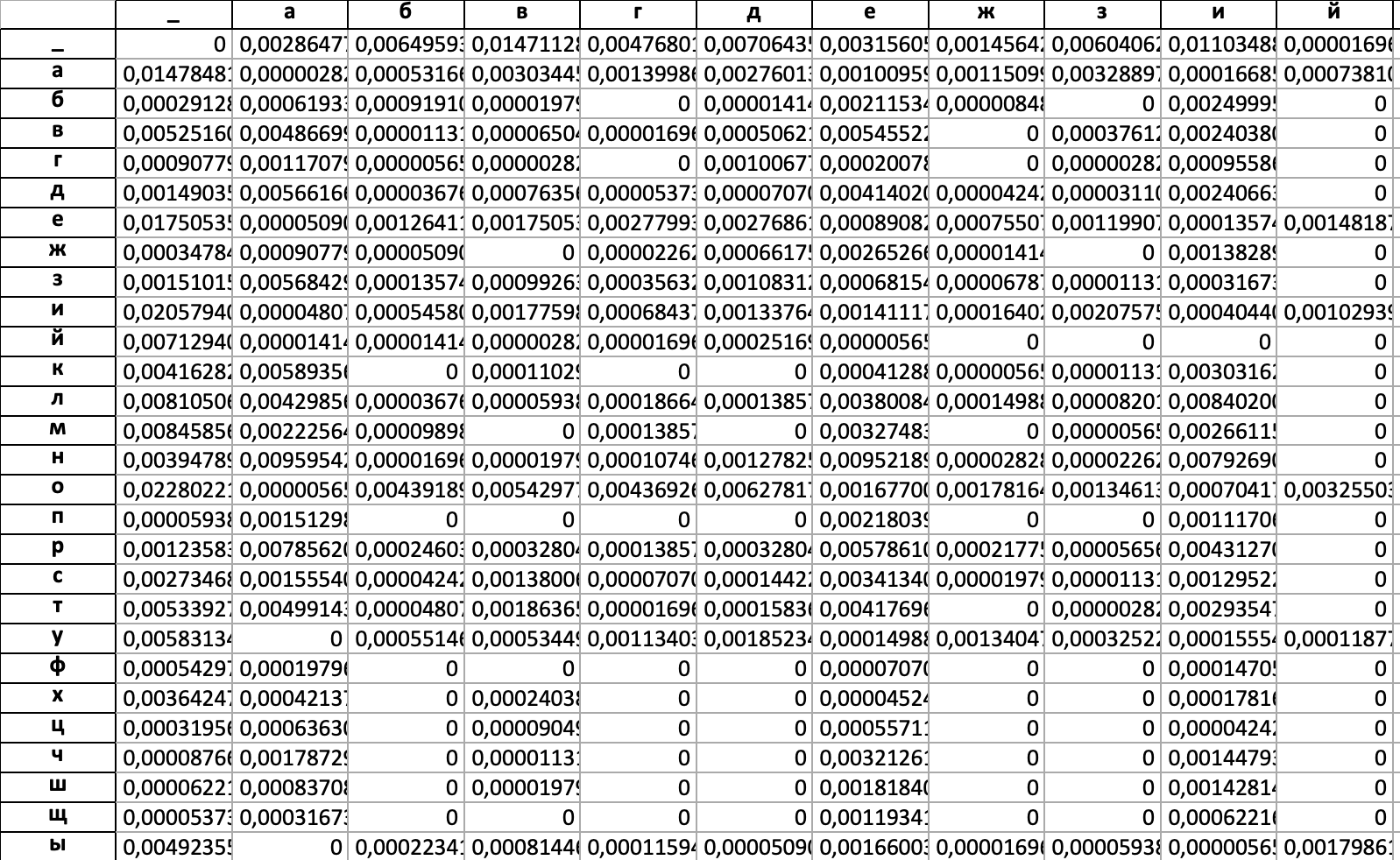
Без перетину та пробілів

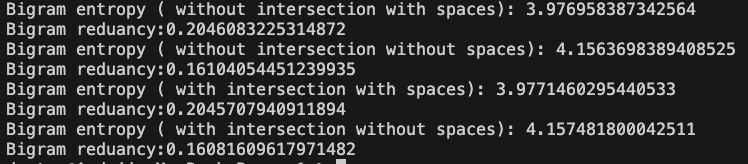


З перетином та без пробілів:

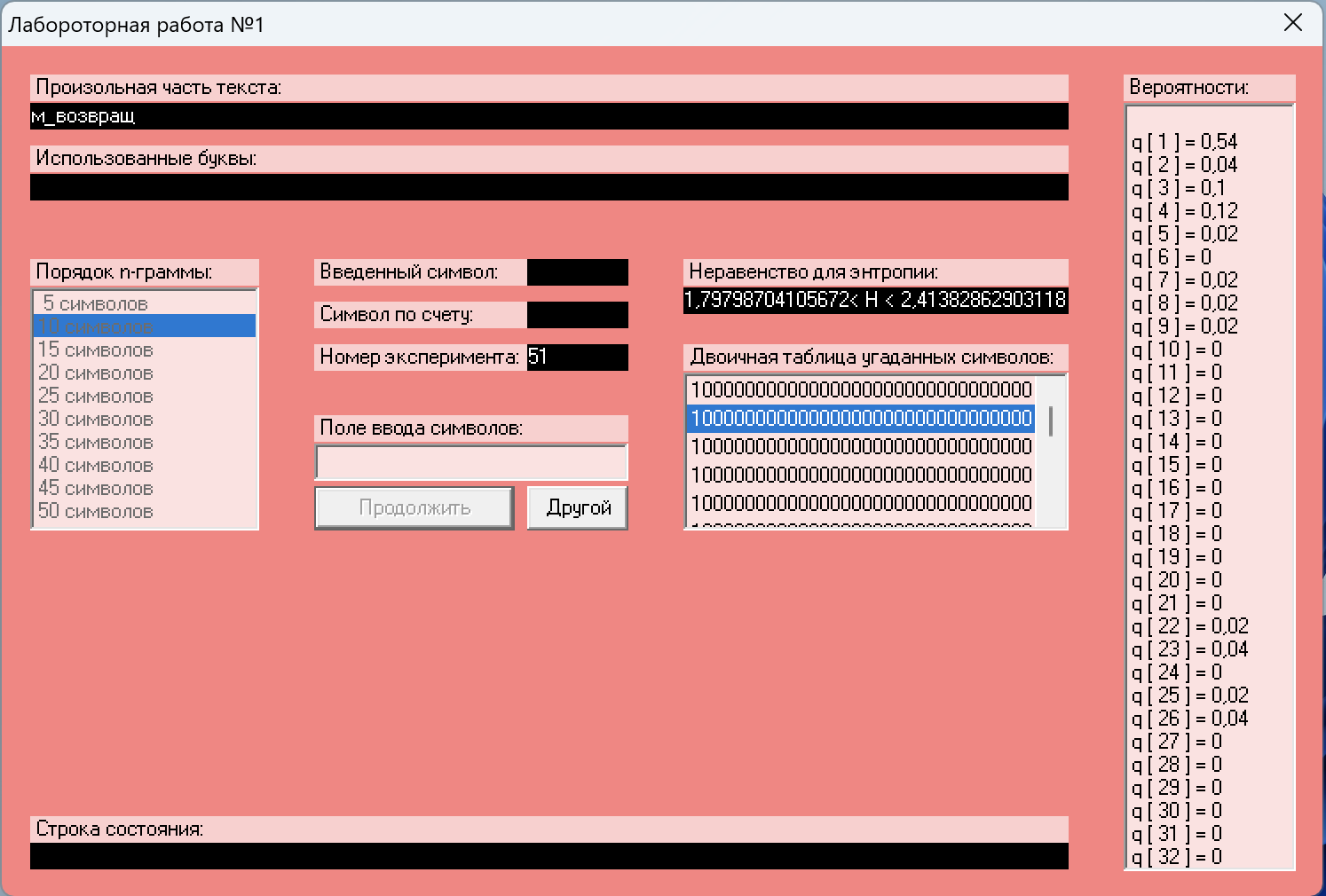


Без перетину та з пробілами:

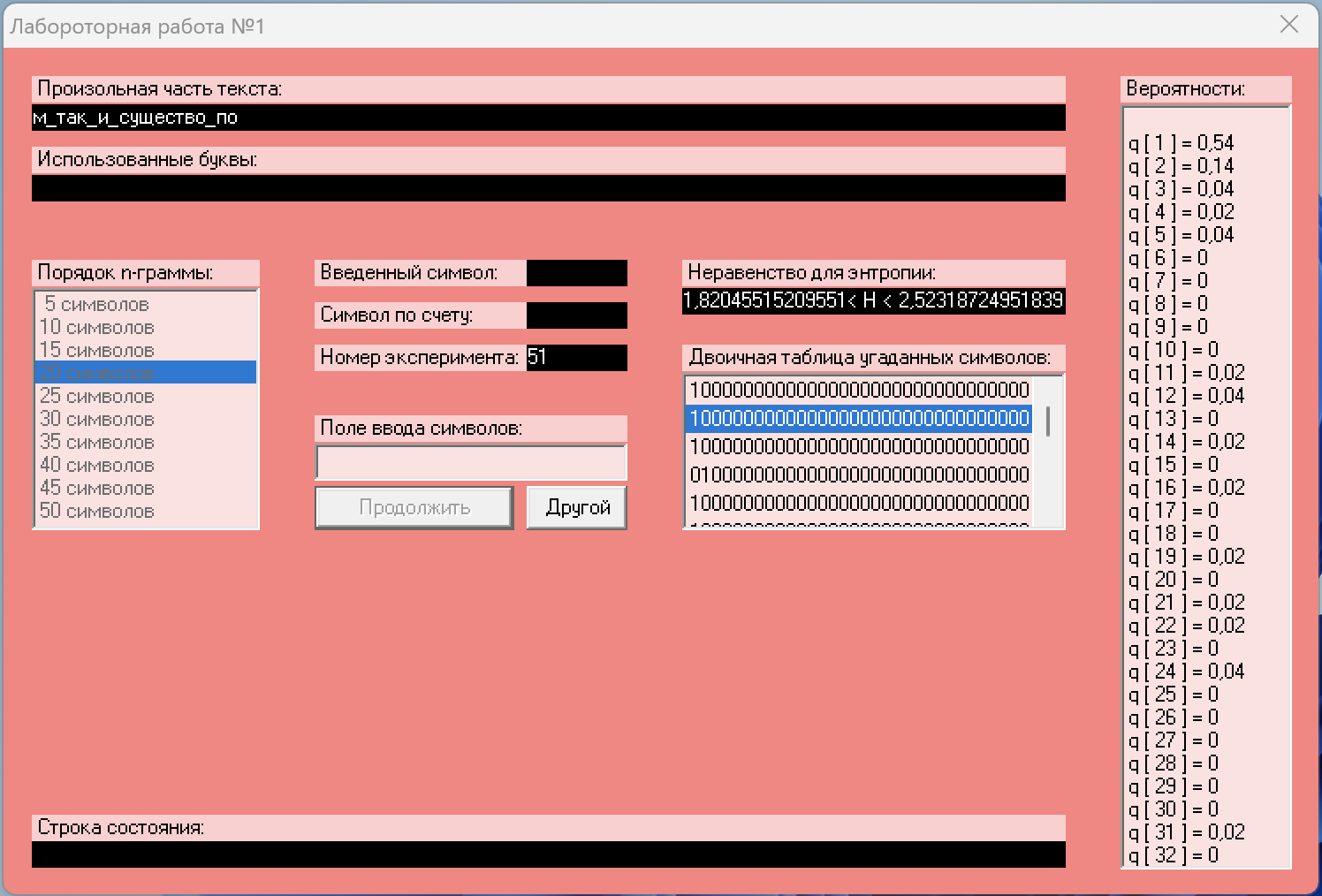


Ентропія для біграм та надлишковість 

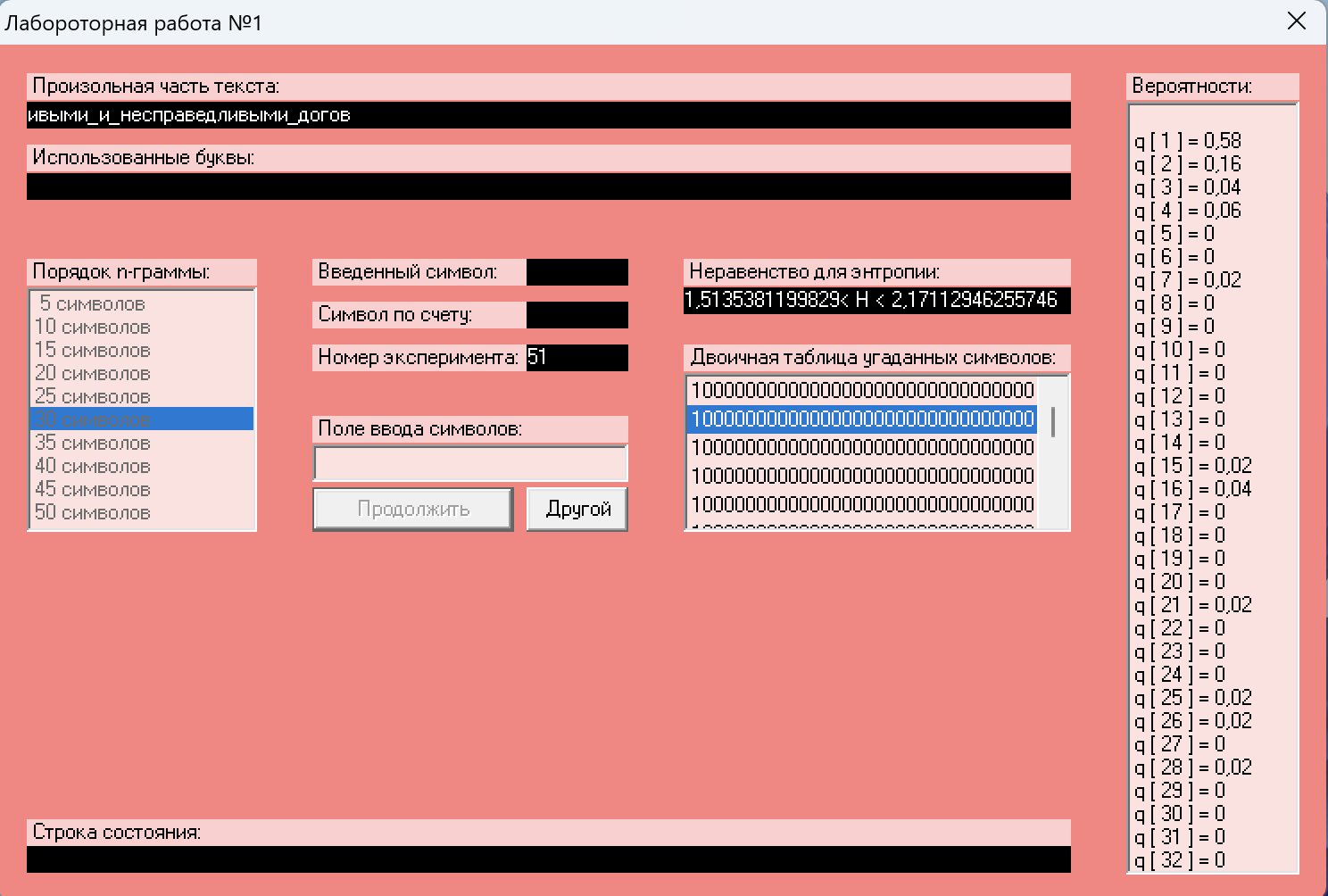
Далі обрахуємо Н(10), Н(20), Н(30) за допомогою CoolPinkProgmar:



1,797987041<<2,41382862

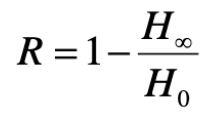


1,8204551520<<2,52318724



1,51353811<<2,17112946

Останнім кроком обрахуємо надлишковість рос мови за формулою



H(10):

0,6404025918

0,517234276

H(20):

0,6359089696

0,495362552

H(30):

0,697292378

0,565774108

**Висновки:**

Практично змогли порахувати ентропію на символ джерела та надлишковість. Також засвоїли поняття частот монограм та біграм, та знайшли їх на практиці.

За допомогою наведеної програми CoolPinkProgram змогли також обрахувати значення H10, H20, H30.