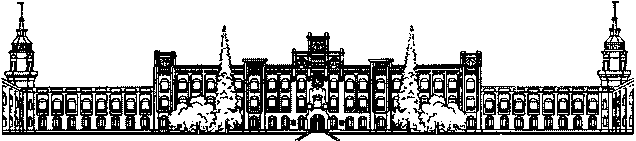
****

**Міністерство освіти і науки 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:

студенти групи ФБ-32

Кошеленко Нікіта Едуардович

Кухарук Ірина Андріївна

**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 H (10), H (20), H (30).

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

Для виконання даної лабораторної роботи було взято текст художнього твору **«Приключения Оливера Твиста»** (автор Чарльз Дікенс).

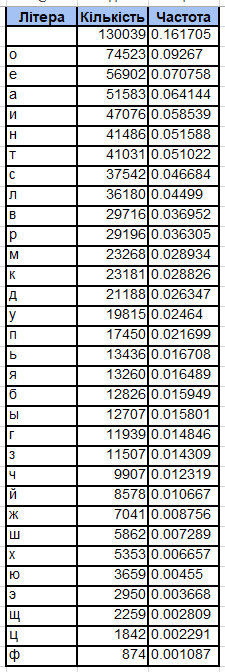
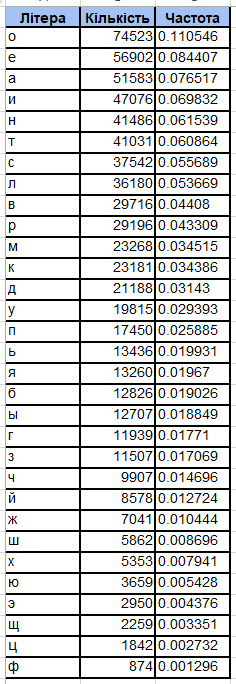
На першому етапі виконано **попередню фільтрацію тексту за допомогою програмного коду,** згідно з методичними вказівками. У результаті отримано очищений текст, який відповідає необхідним умовам:

* усі символи, окрім текстових (літер та пробілів), вилучено або замінено на пробіли;
* усі **прописні літери** замінено на відповідні **рядкові**;
* **послідовності пробілів** та розділових знаків (зокрема символів кінця рядка) замінено одним пробілом, або вилучено, якщо пробіл не входить до алфавіту;
* буква **«ё»** замінена на **«е»**, а **«ъ»** - на **«ь»** для уніфікації алфавіту.

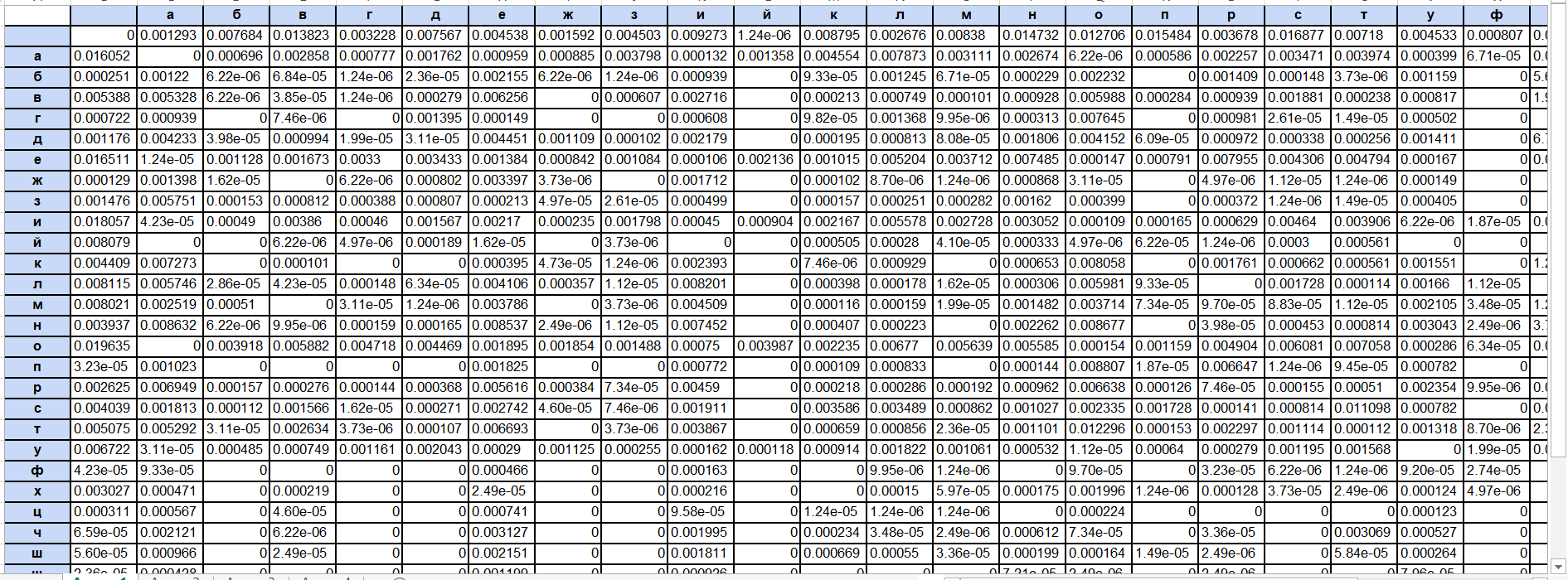
Отриманий текст збережено у двох варіантах:

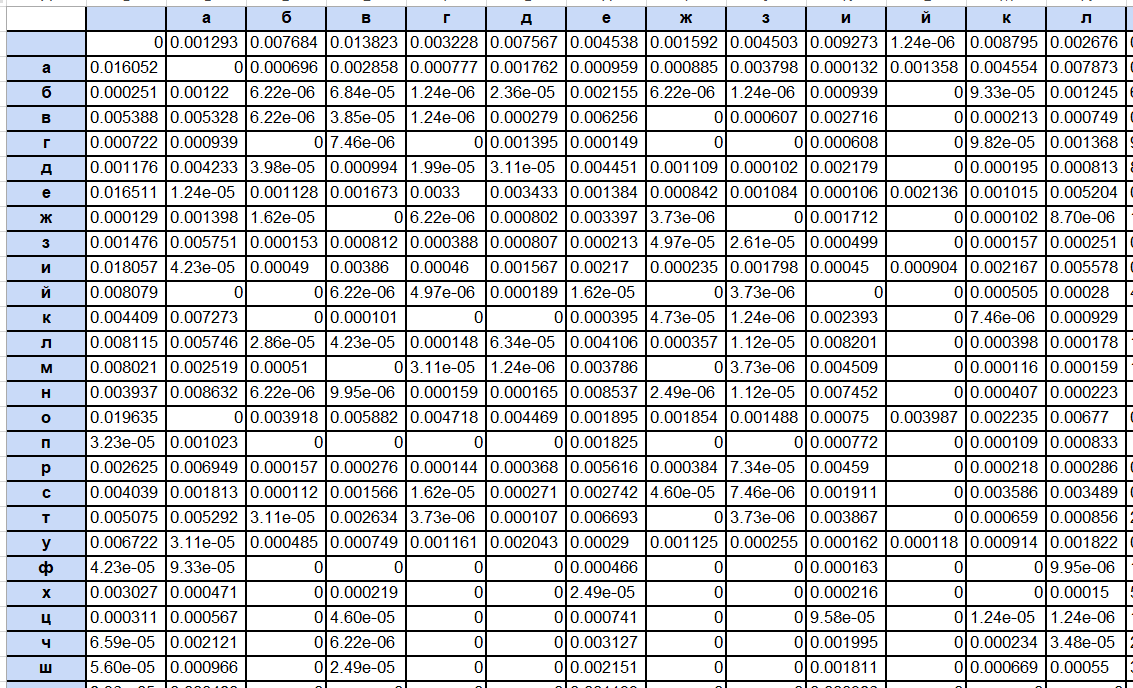
1. **із пробілами**, де пробіл входить до алфавіту (m = 32);
2. **без пробілів**, де пробіл повністю вилучено (m = 31).

Частоти монограм з пробілом та без пробілів.

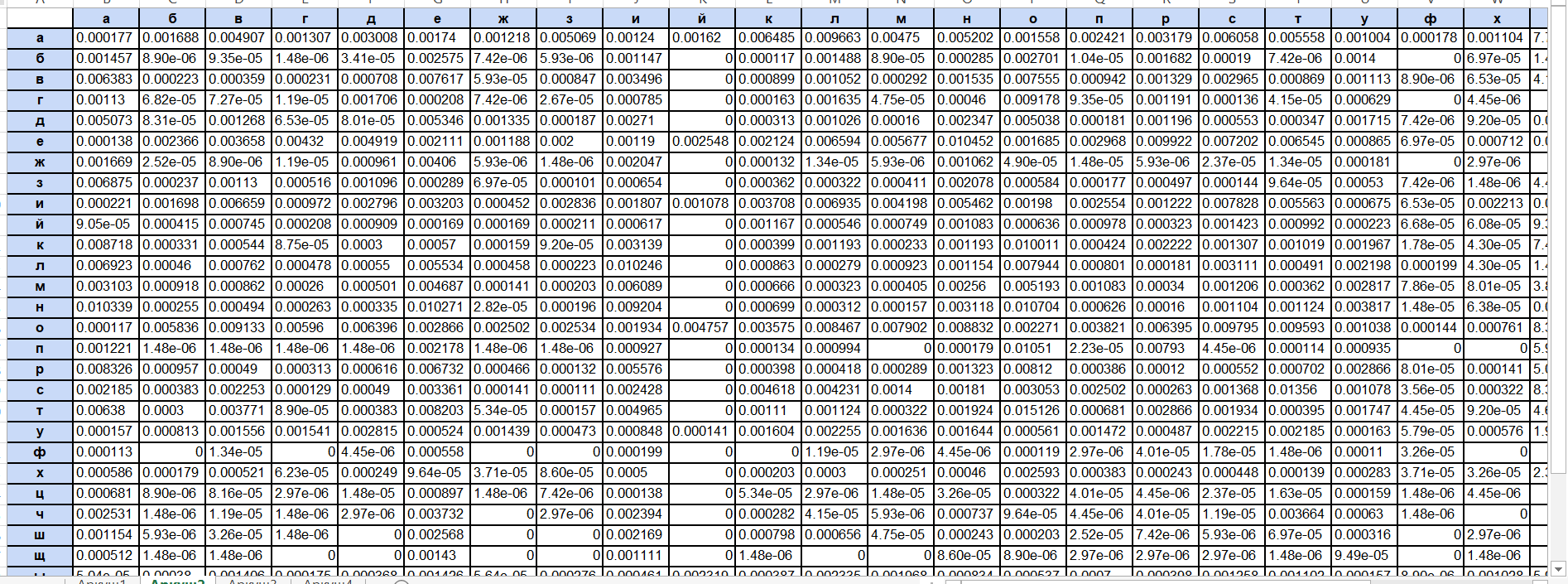
 

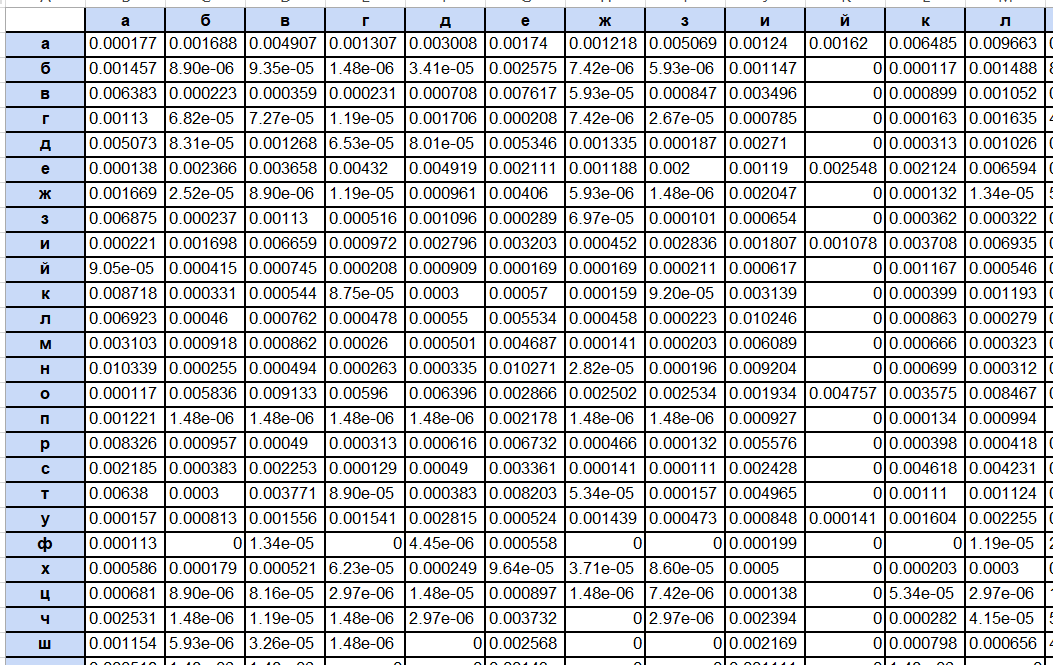
Біграми з пробілами та перекриттям (with\_sp\_bi\_overlap).



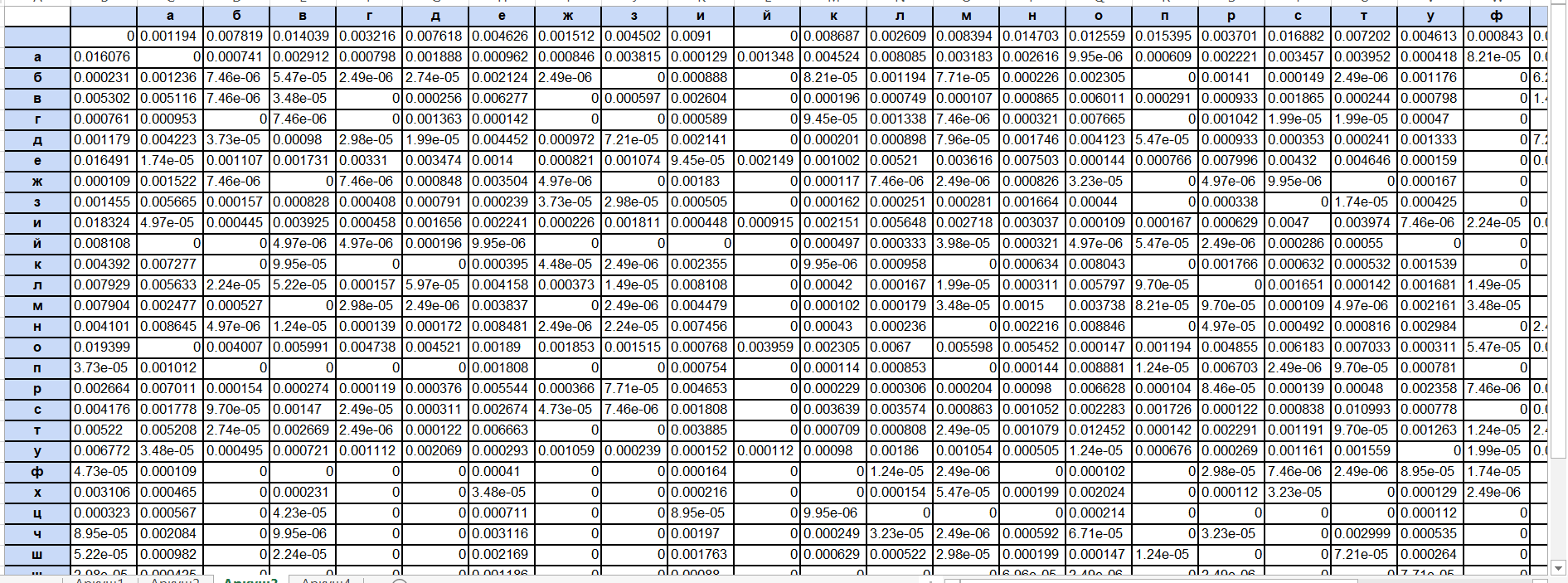


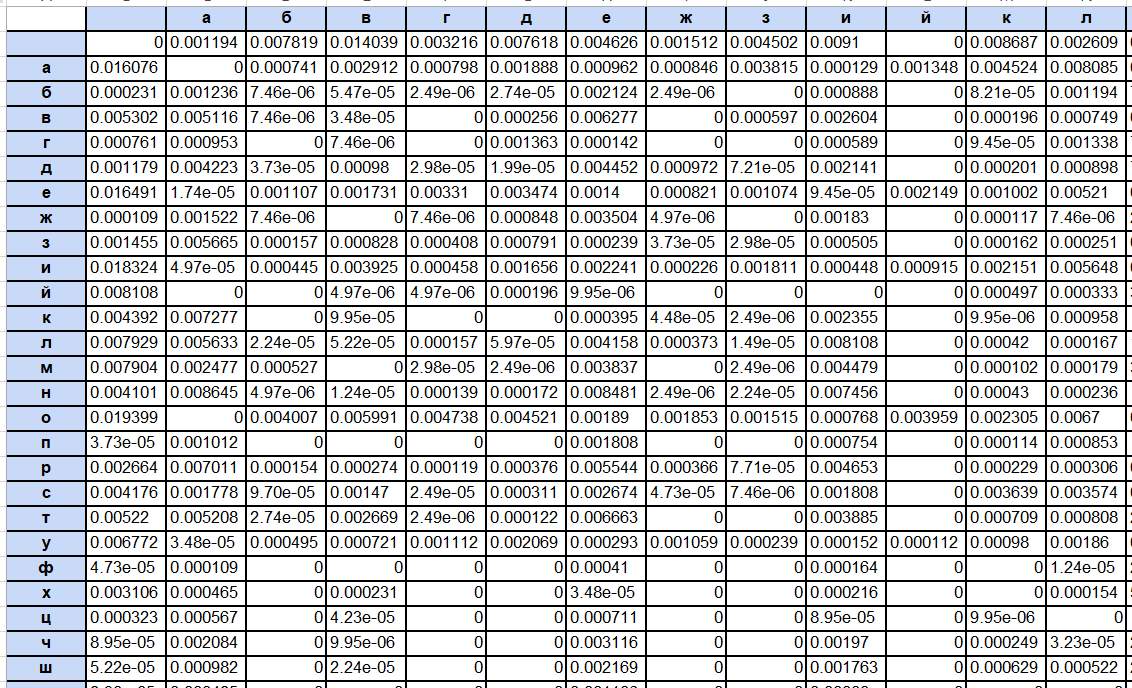
Біграми без пробілів та з перекриттям (no\_sp\_bi\_overlap).



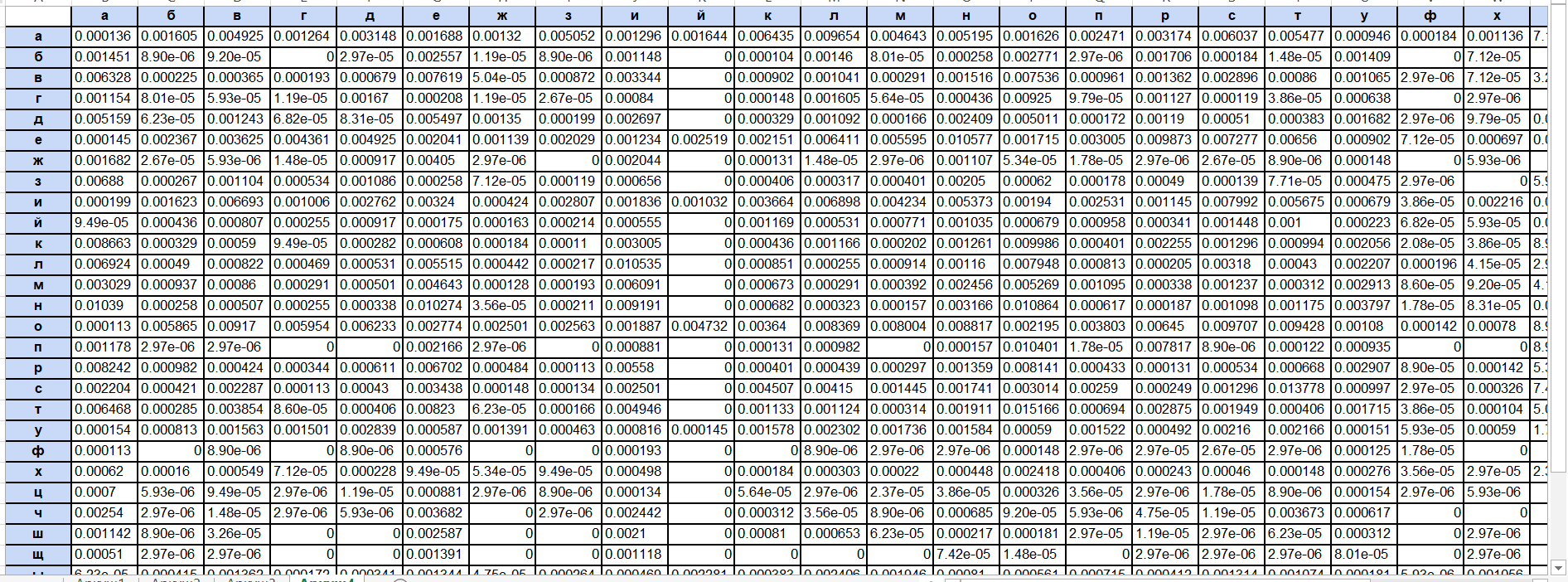


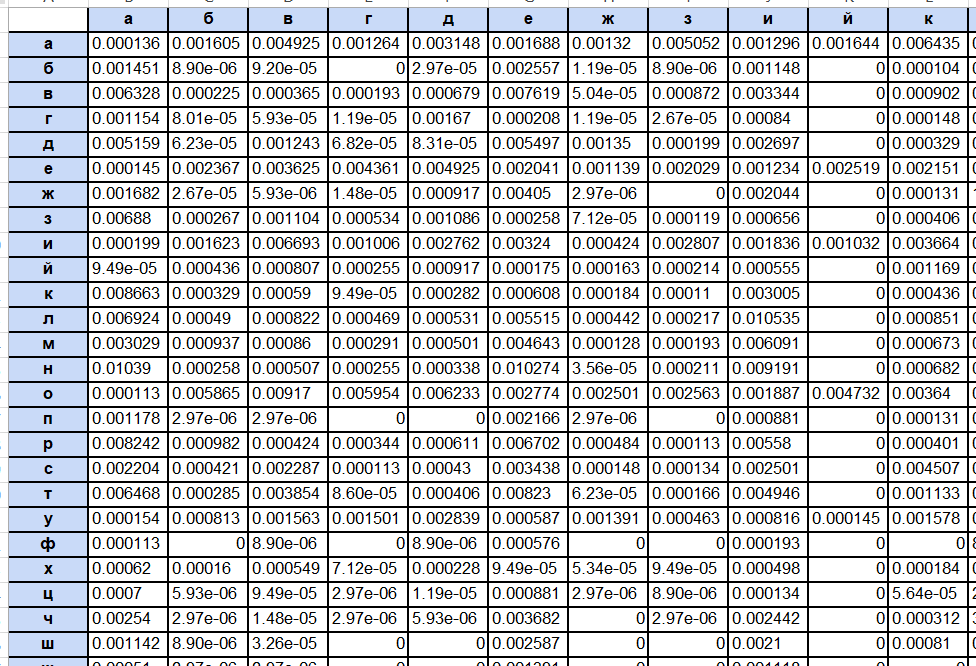
Біграми з пробілами без перекриття (with\_sp\_bi\_nooverlap).





Біграми без пробілів та без перекриття (no\_sp\_bi\_nooverlap).





Таблиці:

with\_sp\_bi\_overlap – з пробілами та перекриттям.

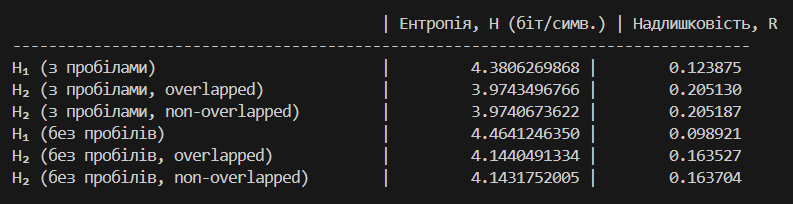
no\_sp\_bi\_overlap – без пробілів та з перекриттям.

with\_sp\_bi\_nooverlap – з пробілами без перекриття.

no\_sp\_bi\_nooverlap – без пробілів та без перекриття.

**Оцінка значень ентропії та надлишковості**

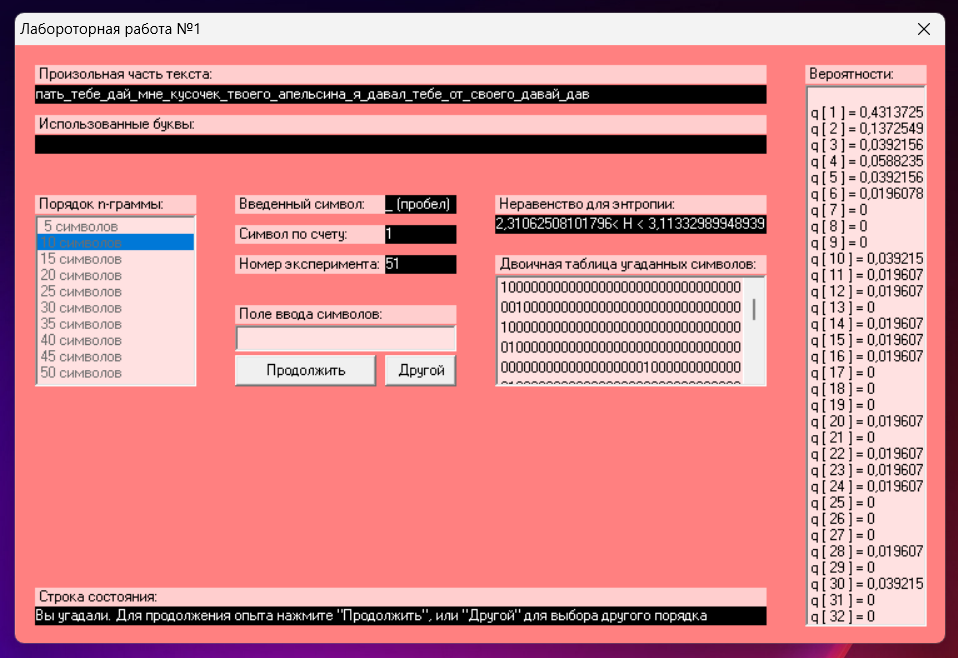
H1 – монограми, H2 – біграми.

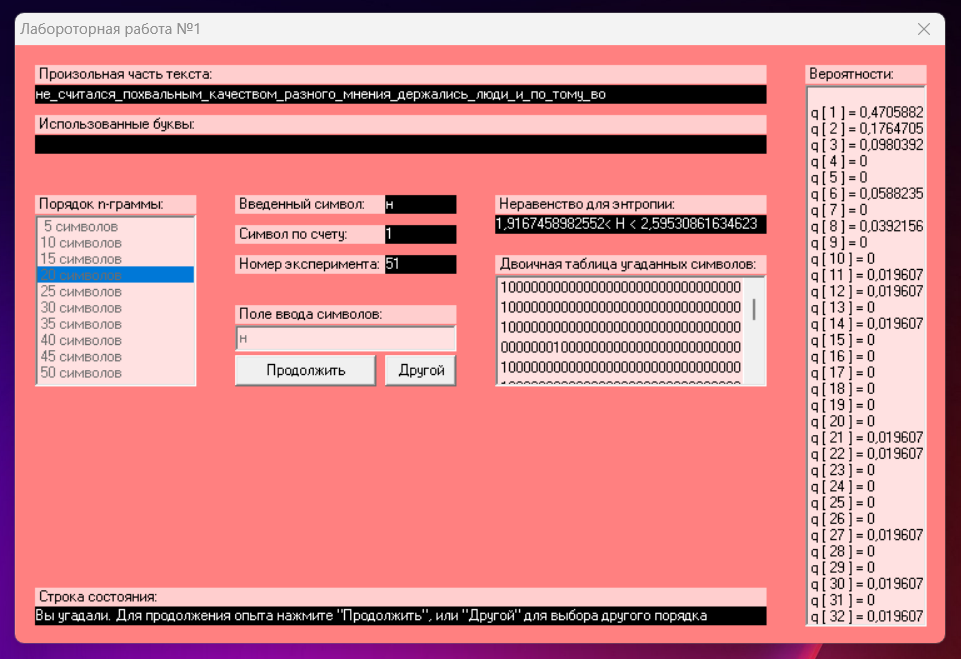


3. CoolPinkProgram

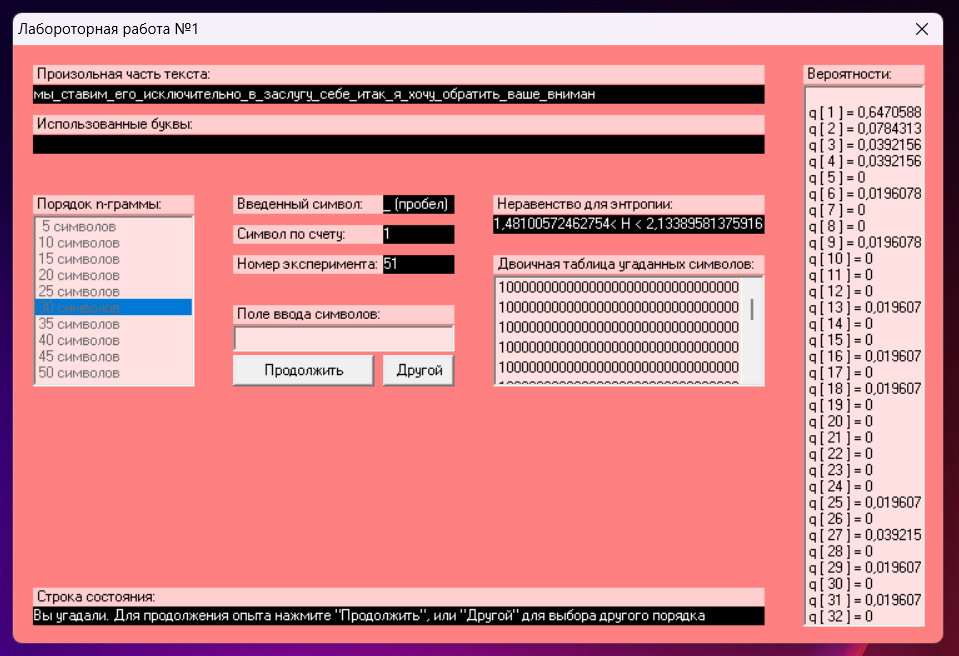
Проводимо щонайменше 50 експериментів у CoolPinkProgram.

H10:

Н20:



H30:



**Для 10 символів**

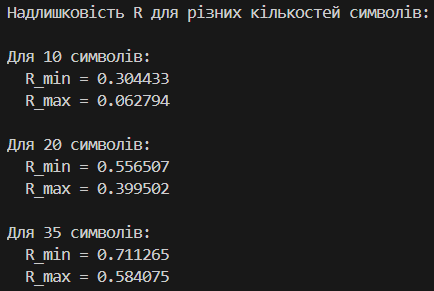
2.31062508101796<H<3.11332989948939

**Для 20 символів**

1.9167458982552<H<2.59530861634623

**Для 35 символів**

1.48100572462754<H<2.13339581375916

****

Висновки: Виконуючи дану лабораторну роботу