Міністерство освіти і науки України

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

**ЛАБОРАТОРНА РОБОТА №1**

З дисципліни«Криптографія»

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:

студенти 3 курсу ФТІ

групи ФБ-73  
Дем’яненко Д.

Проноза А.

Перевірив:

Чорний О.

# 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 біграx`м в тексті, а також підрахунку та за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення та на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення та на тому ж тексті, в якому вилучено всі пробіли.

2. За допомогою програми CoolPinkProgram оцінити значення , , .

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

# Хід роботи:

1)Прочитали методичні вказівки до виконання лабораторної роботи.

2) За допомогою програми CoolPinkProgram оцінили значення , , і запушили відповідні скріншоти на github (пункт 2).

3)Створили додаток у IntelliJ IDEA для виконання завдання 1: написати програму для підрахунку частот букв і частот біграм в тексті, а також підрахунку та за безпосереднім означенням.

4)Запушили код на github

5) Використовуючи отримані значення ентропії, оцінили надлишковість російської мови в різних моделях джерела.

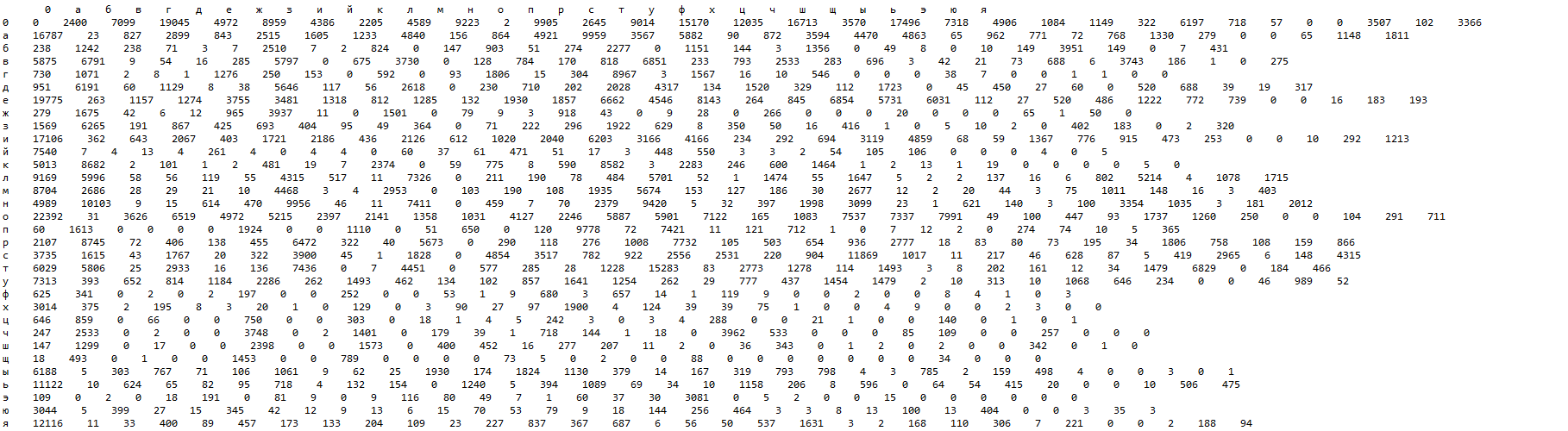
**Частота букв з пробілом:**

0=355791, а=155972, б=32352, в=83174, г=35267, д=60635, е=160599, ж=19814, з=31772, и=117676, й=19834, к=62943, л=93147, м=63464, н=118529, о=208075, п=48731, р=85811, с=102431, т=118819, у=53213, ф=5925, х=12148, ц=6856, ч=28066, ш=15003, щ=5826, ы=35493, ь=38294, э=7866, ю=11186, я=38661

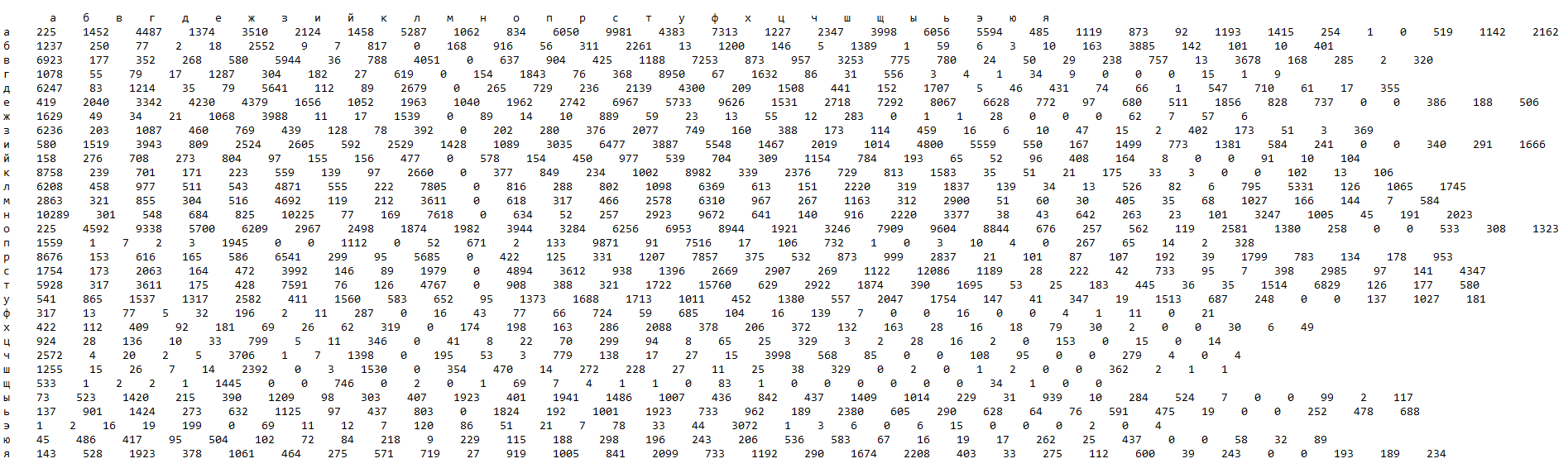
**Частота букв буз пробілу:**

а=155972, б=32352, в=83174, г=35267, д=60635, е=160599, ж=19814, з=31772, и=117676, й=19834, к=62943, л=93147, м=63464, н=118529, о=208075, п=48731, р=85811, с=102431, т=118819, у=53213, ф=5925, х=12148, ц=6856, ч=28066, ш=15003, щ=5826, ы=35493, ь=38294, э=7866, ю=11186, я=38661

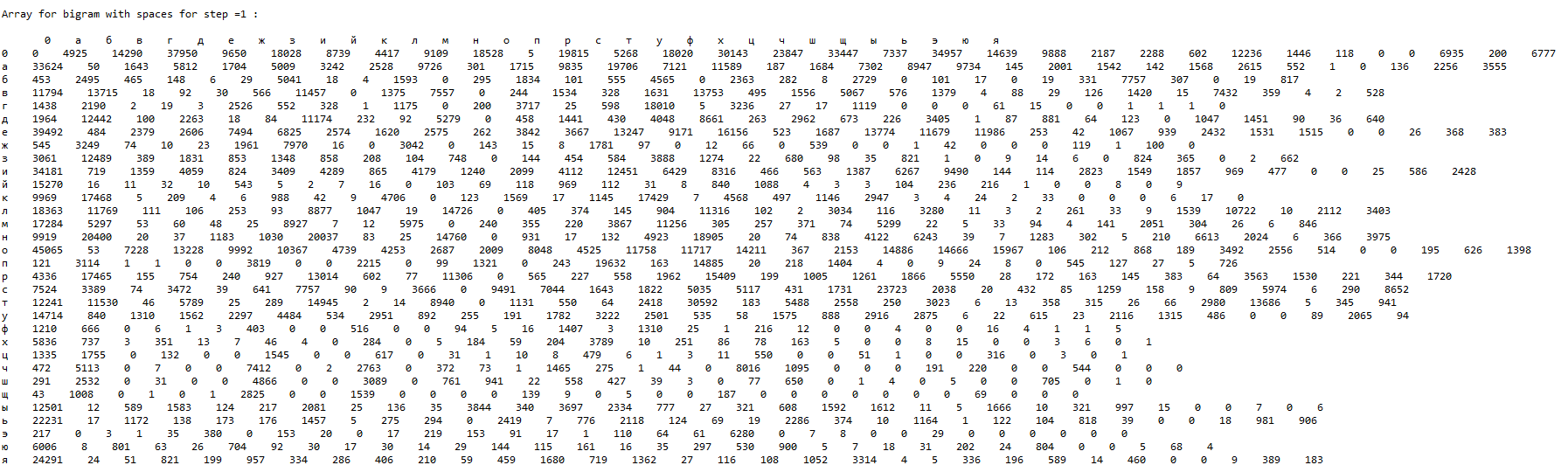
**Частота біграм з пробілом з кроком 2:**



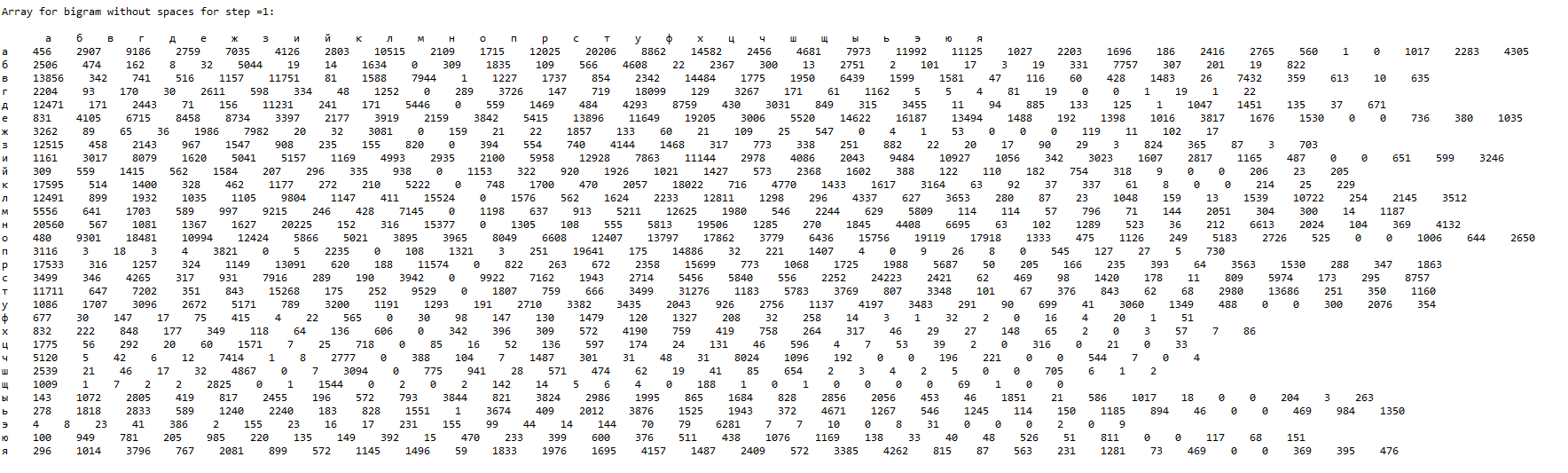
**Частота біграм без пробілу з кроком 2:**

****

**Частота біграм з пробілом з кроком 1:**

****

**Частота біграм без пробілу з кроком 1:**

****

**Ентропія для 1 символу з пробілом:**

=4.384479655591922

**Ентропія для 1 символу без пробілу:**

=4.462782106749086

**Ентропія для біграми з пробілом з кроком 2:**

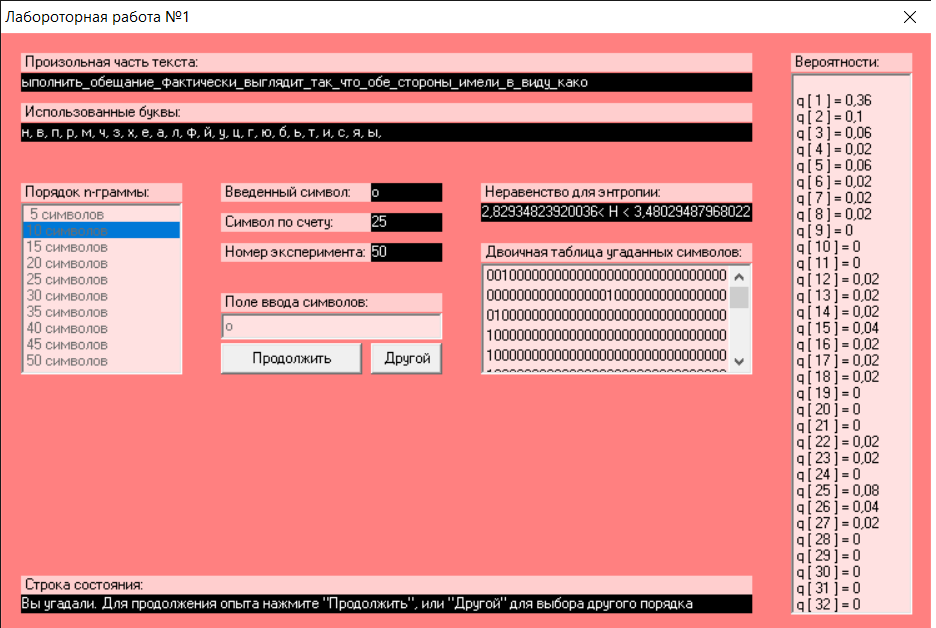
=4.4815897093739565

**Ентропія для біграми без пробілу з кроком 2:**

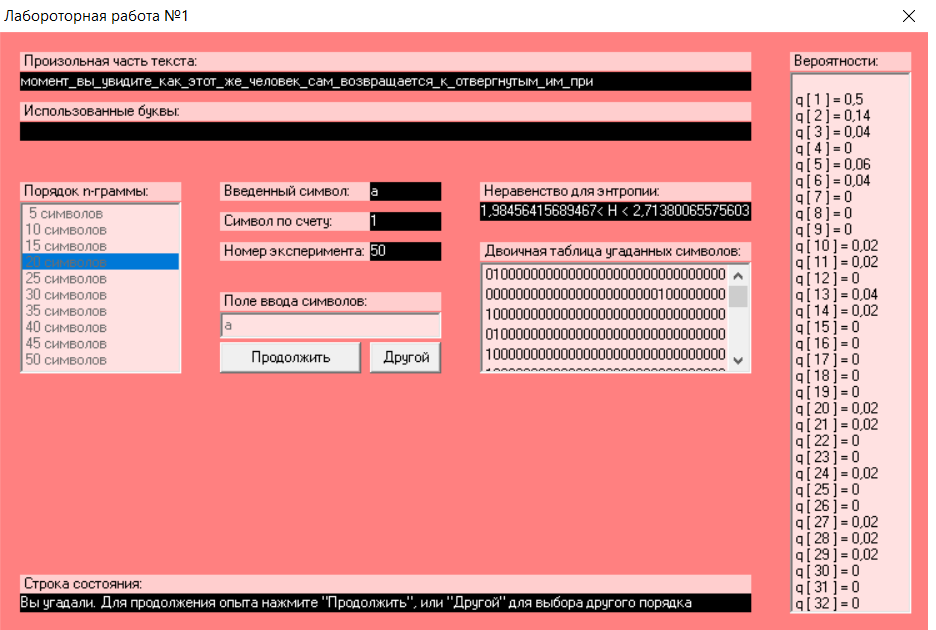
=4.645341941432777

**Ентропія для біграми з пробілом з кроком 1:** =7.963104187245263

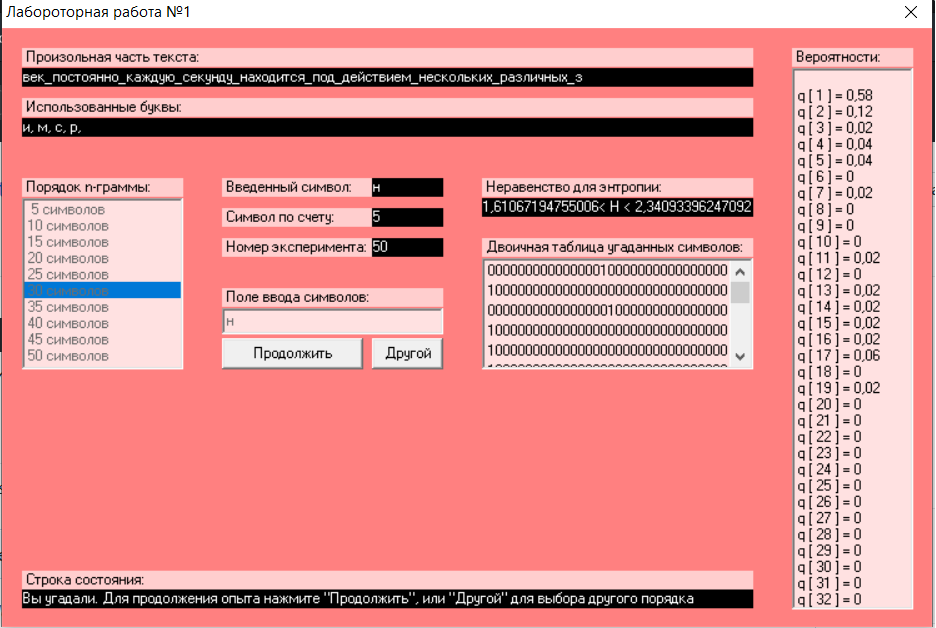
**Ентропія для біграми без пробілу з кроком 1:** = 8.290740553097267

**Умовна ентропія джерела для 10 символів:**

**Умовна ентропія джерела для 20 символів:**



**Умовна ентропія джерела для 30 символів:**



# Надлишковість російської мови:

**Ідеальна ентропія з пробілом**

=21.090792787932422

**Ідеальна ентропія без пробілу**

=20.84044448553777

**Надлишковість російської мови для символів з пробілами**

R=0.792114042384381

**Надлишковість російської мови для символів без пробілів**

R=0.7858595525711539

**Надлишковість російської мови для біграм з пробілами з кроком 2**

R=0.7875096609958541

**Надлишковість російської мови для біграм без пробілів з кроком 2**

R=0.7770996705633407

**Надлишковість російської мови для біграм з пробілами з кроком 1**

R= 0.6224369435841437

**Надлишковість російської мови для біграм з пробілами з кроком 1**

R= 0.6021802433796156

# Проблеми:

Головною проблемою було виведення кількості біграм в тексті у вигляді таблиці, оскільки для цього було використано масив величиною +1 від літер алфавіту для того, щоб створити шапку таблиці. Було проблематично співставити ключ значення біграми з відповідною ячейкою масиву.

# Висновок:

Засвоїли поняття ентропії на символ джерела та його надлишковості, вивчили та порівняли різні моделі джерела відкритого тексту для наближеного визначення ентропії, набули практичних навичок щодо оцінки ентропії на символ джерела.