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=0,159306574, о=0,093166256, е=0,071908723, а=0,069836969, т=0,053201592, н=0,053071744, и=0,052689810, с=0,045863812, л=0,041706871, р=0,038422153, в=0,037241428, м=0,028416212, к=0,028182932, д=0,027149518, у=0,023826293, п=0,021819463, я=0,017310588, ь=0,017146263, ы=0,015892106, г=0,015790914, б=0,014485713, з=0,014226016, ч=0,012566642, й=0,008880738, ж=0,008871783, ш=0,006717642, х=0,005439306, ю=0,005008568, э=0,003522027, ц=0,003069796, ф=0,002652938, щ=0,002608610

**Частота букв буз пробілу:**

о=0,110820726, е=0,085535013, а=0,083070673, т=0,063282988, н=0,063128534, и=0,062674227, с=0,054554741, л=0,049610084, р=0,045702931, в=0,044298465, м=0,033800921, к=0,033523436, д=0,032294195, у=0,028341239, п=0,025954126, я=0,020590845, ь=0,020395381, ы=0,018903569, г=0,018783201, б=0,017230672, з=0,016921764, ч=0,014947949, й=0,010563587, ж=0,010552935, ш=0,007990596, х=0,006470024, ю=0,005957663, э=0,004189431, ц=0,003651505, ф=0,003155654, щ=0,003102927

**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

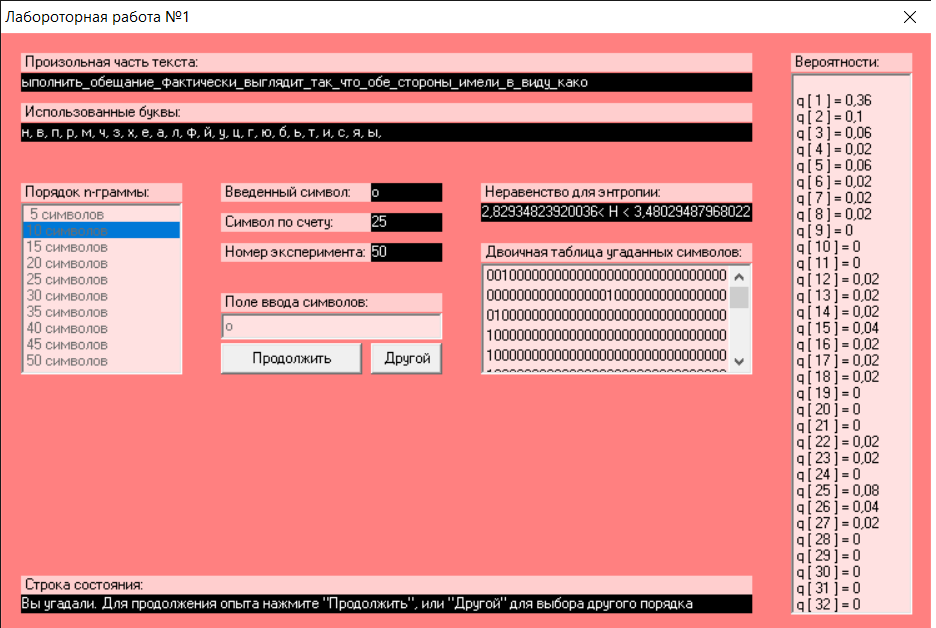
**Ентропія для біграми без пробілу з кроком 2:**

=4.645341941432777

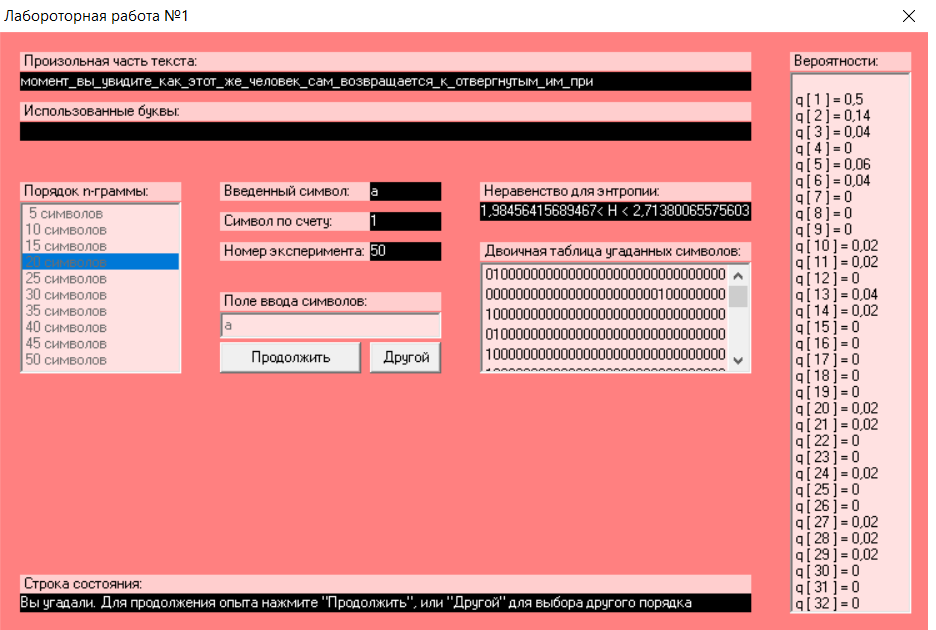
**Ентропія для біграми з пробілом з кроком 1:** =7.963104187245263

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

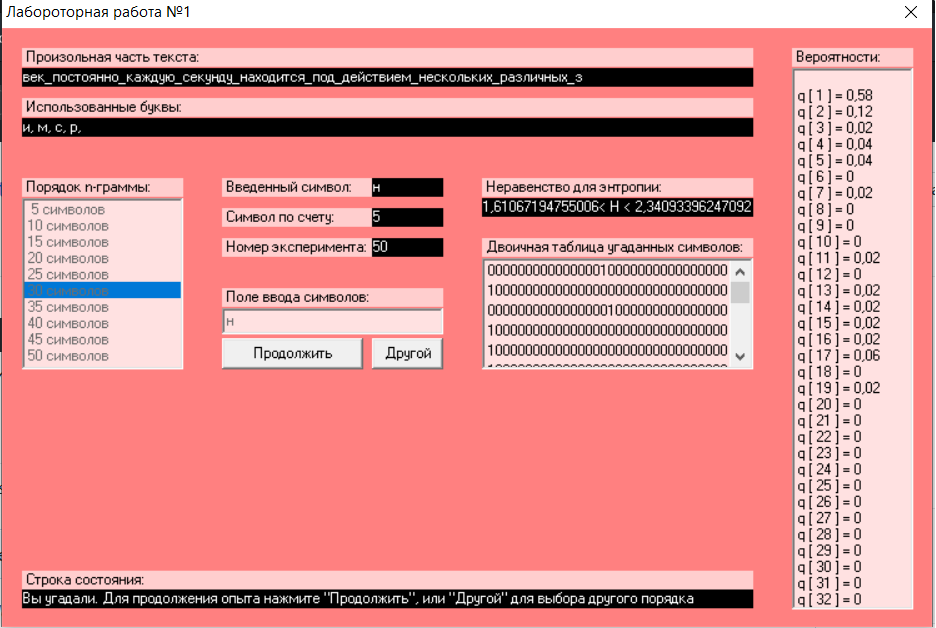
= 8.290740553097267

**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

**Надлишковість російської мови 10 символів**

0,3039412<=R<= 0,4341304‬

**Надлишковість російської мови 20 символів**

0,45724<=R<= ‬0,6030872

**Надлишковість російської мови 30 символів**

0,5318134‬<=R<= 0,6778658

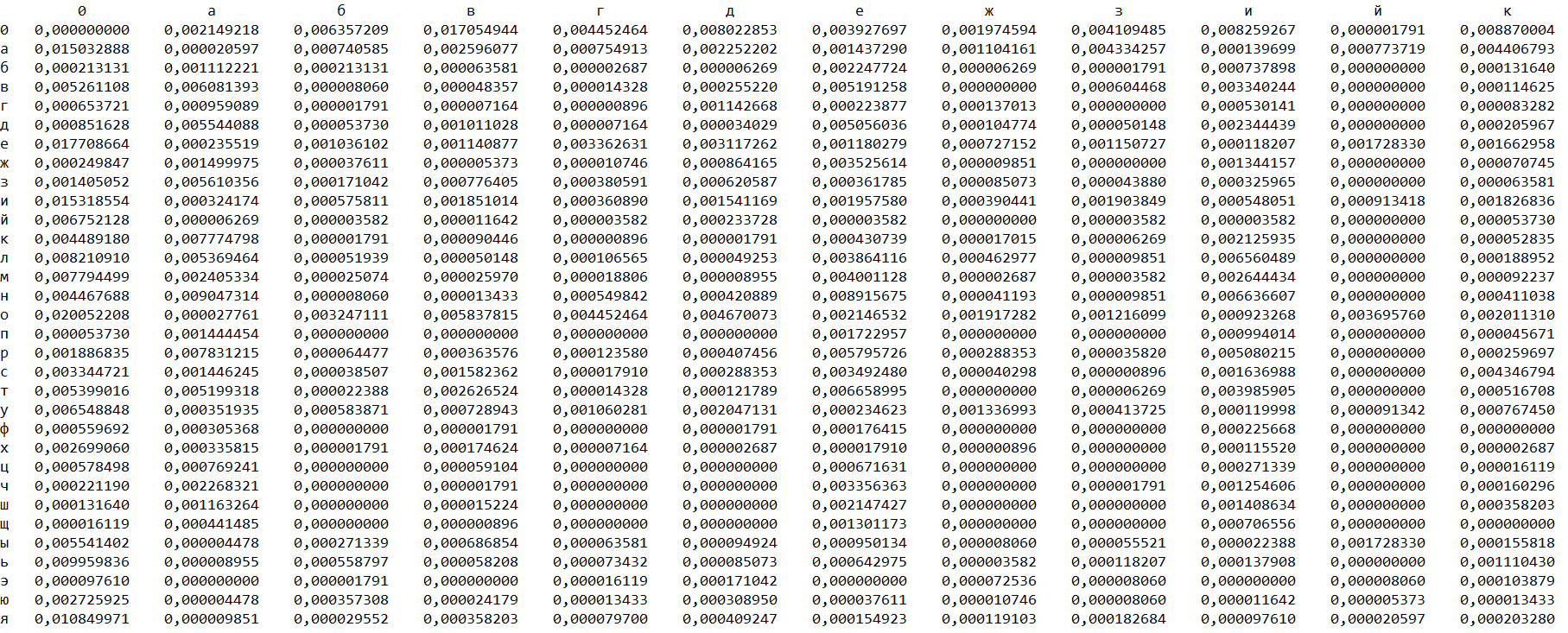
# Проблеми:

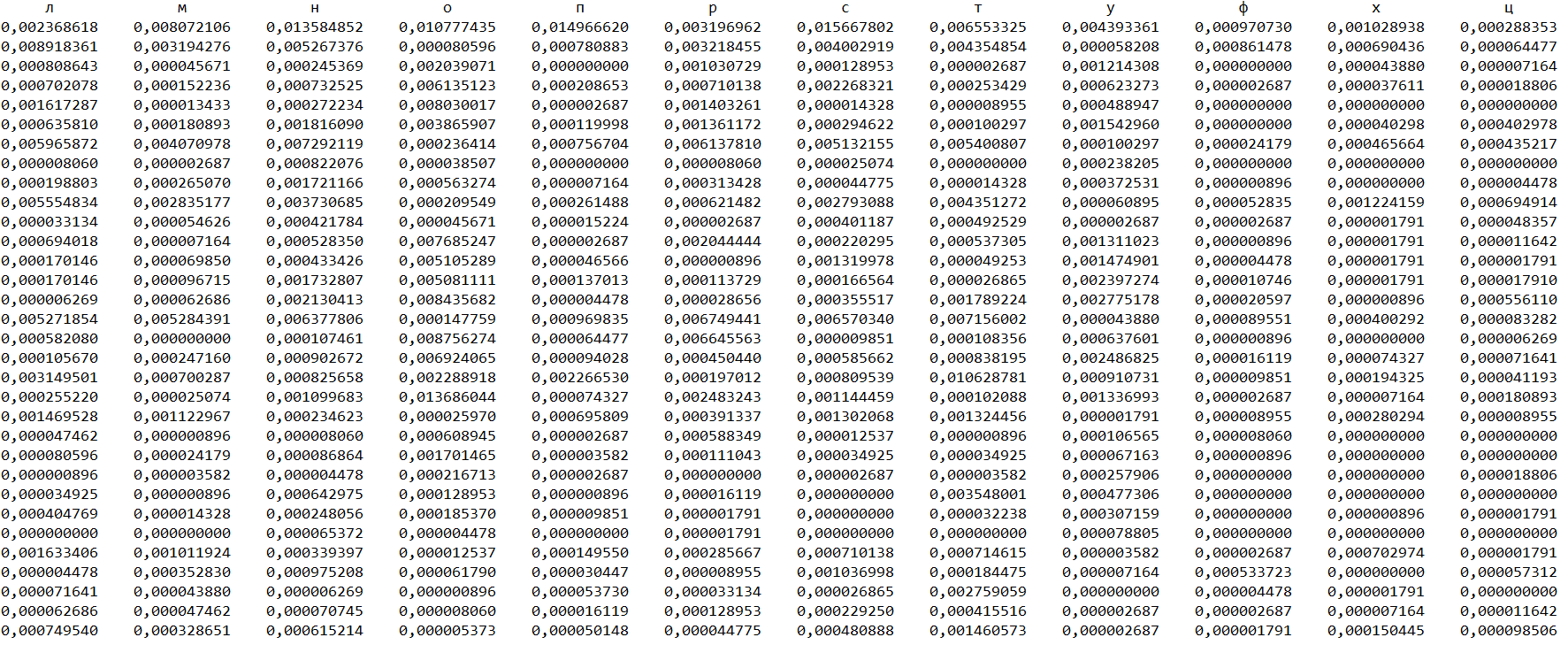
Головною проблемою було виведення кількості біграм в тексті у вигляді таблиці, оскільки для цього було використано масив величиною +1 від літер алфавіту для того, щоб створити шапку таблиці. Було проблематично співставити ключ значення біграми з відповідною ячейкою масиву.

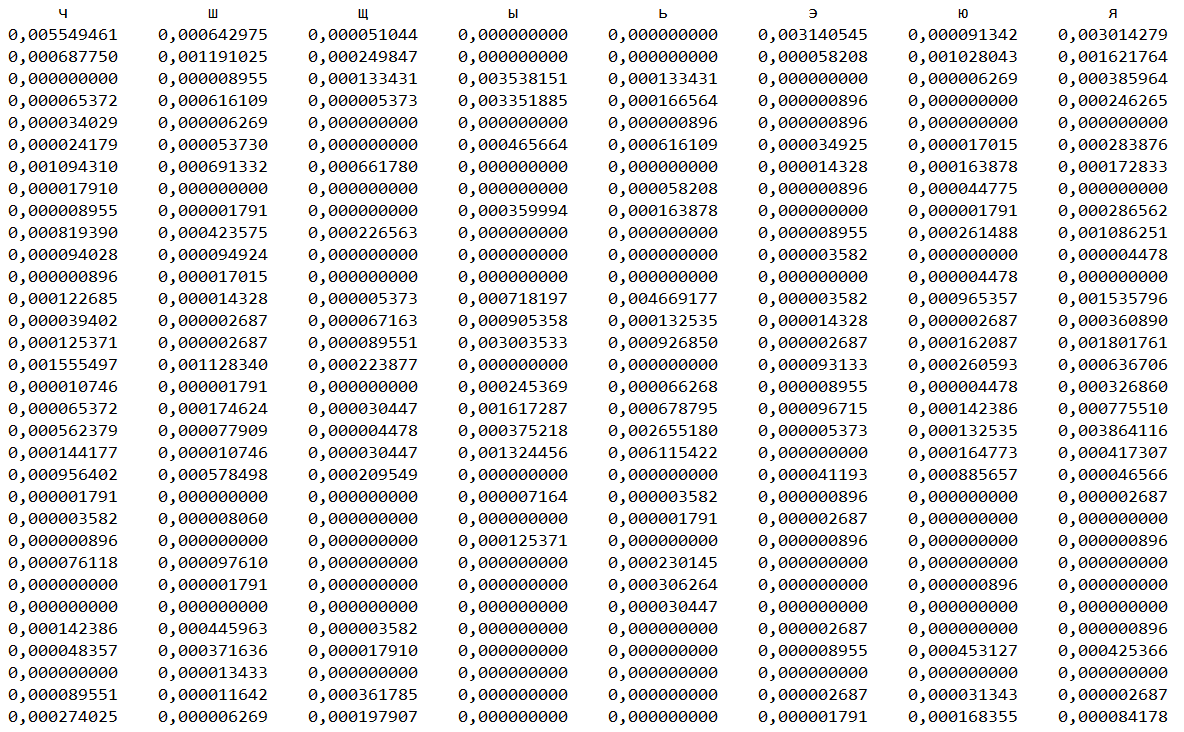
# Висновок:

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.

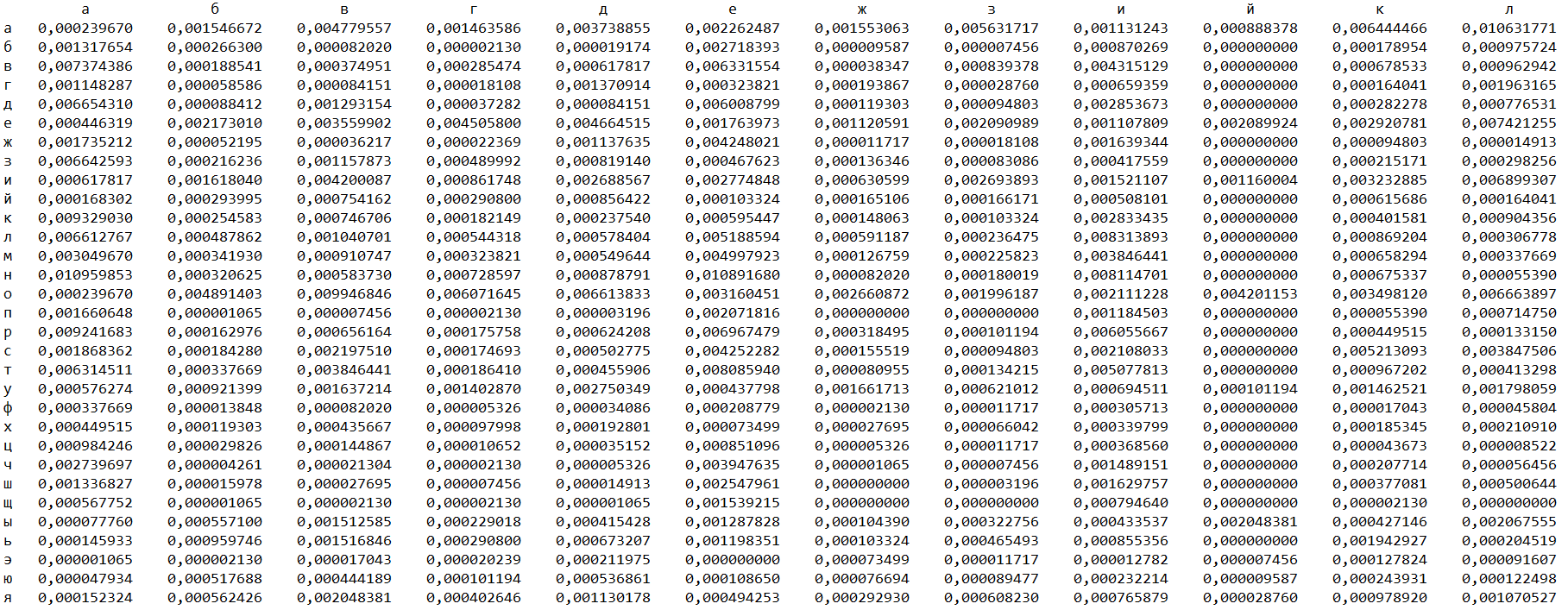
**Частота біграм з пробілом і кроком 2**

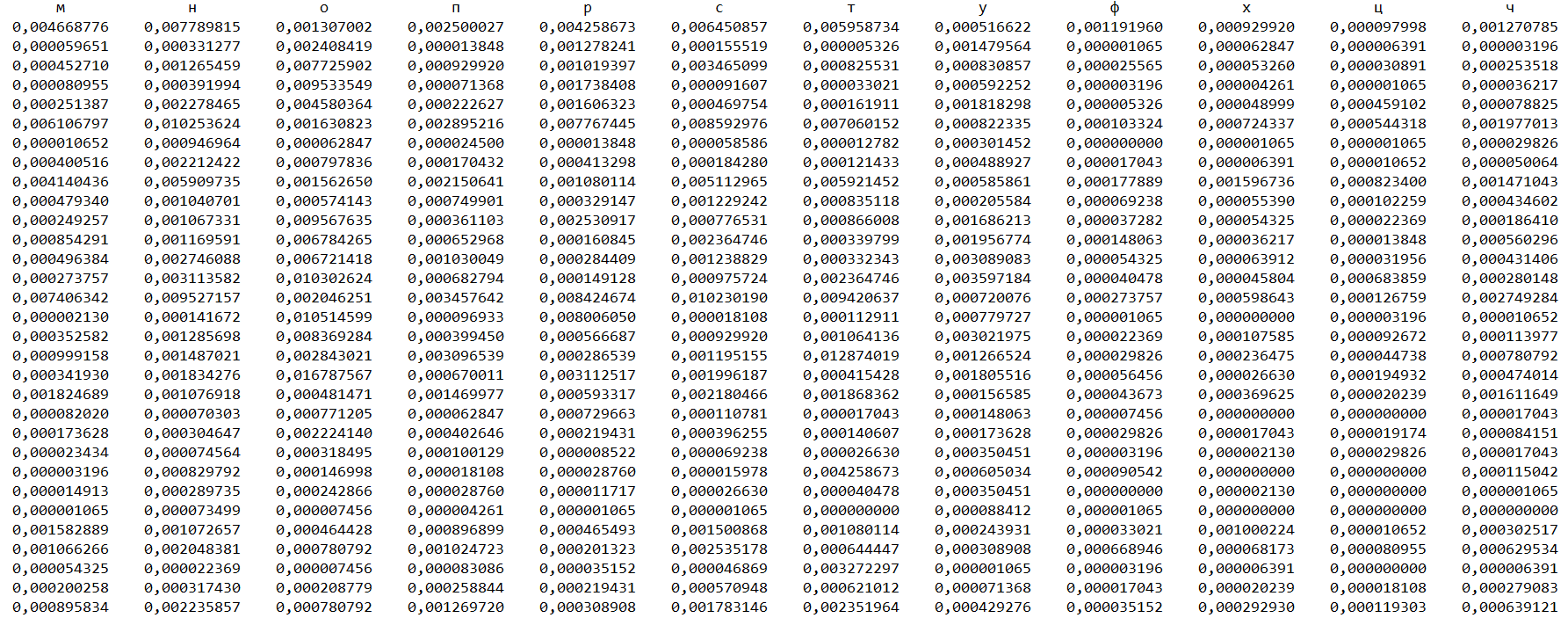
****

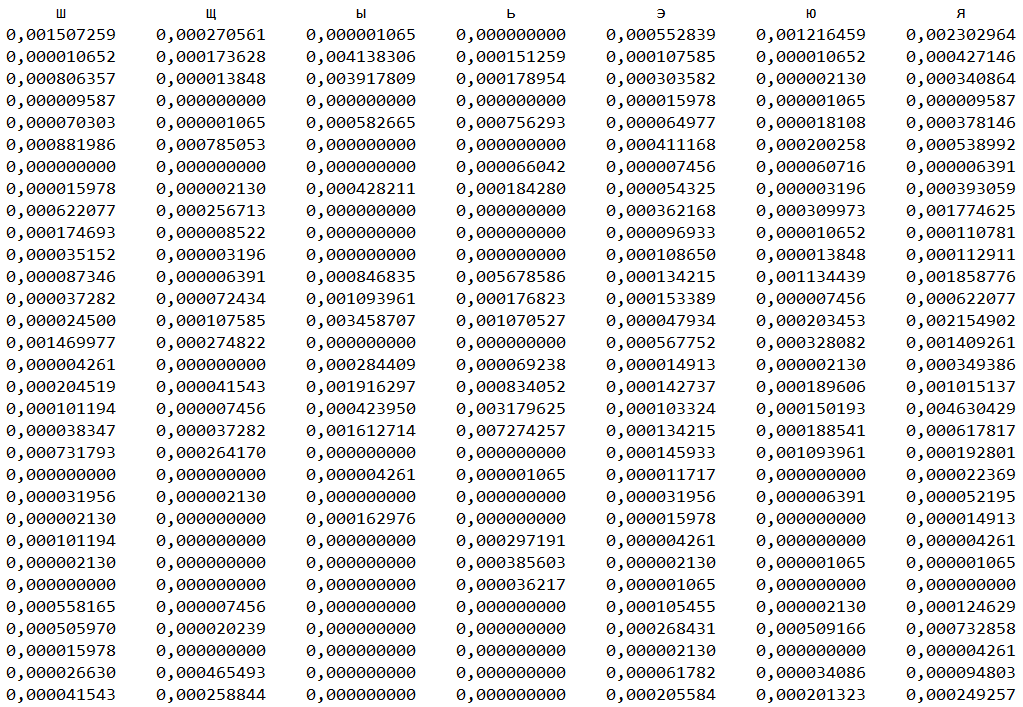
****

****

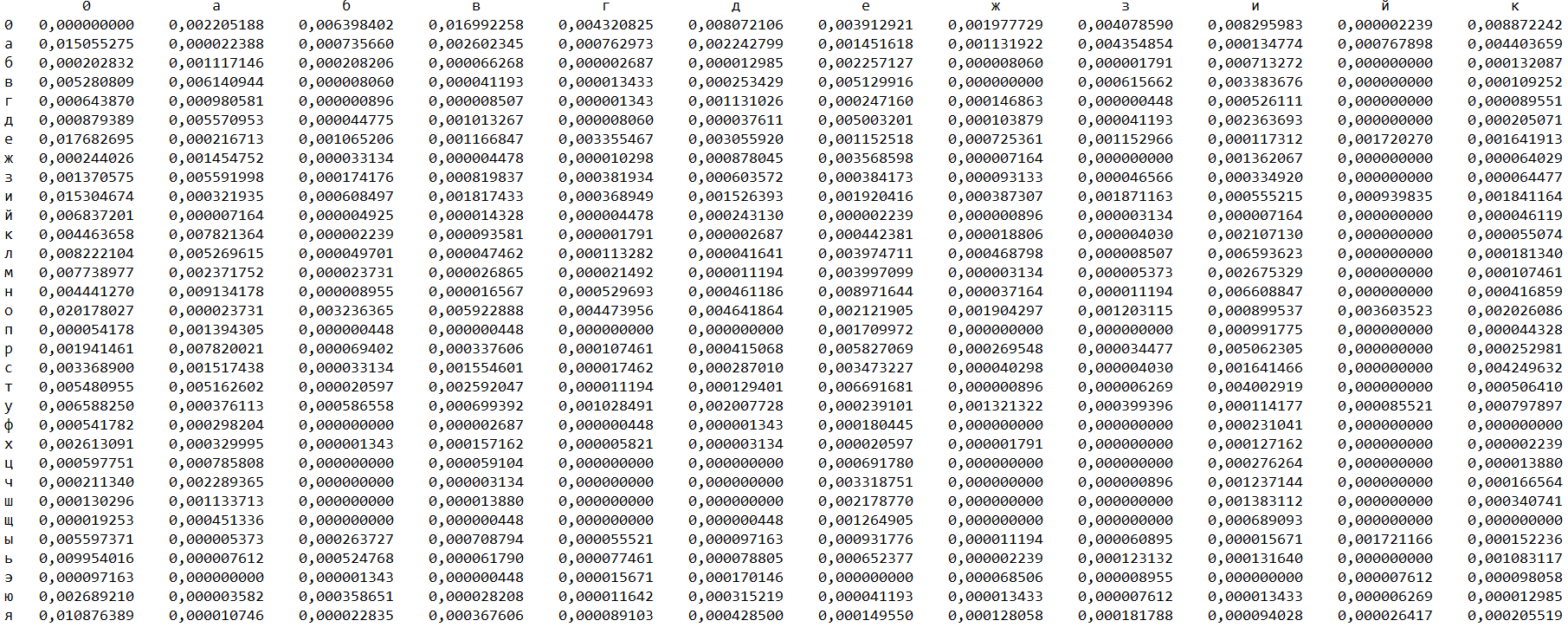
**Частота біграм без пробілу і кроком 2**

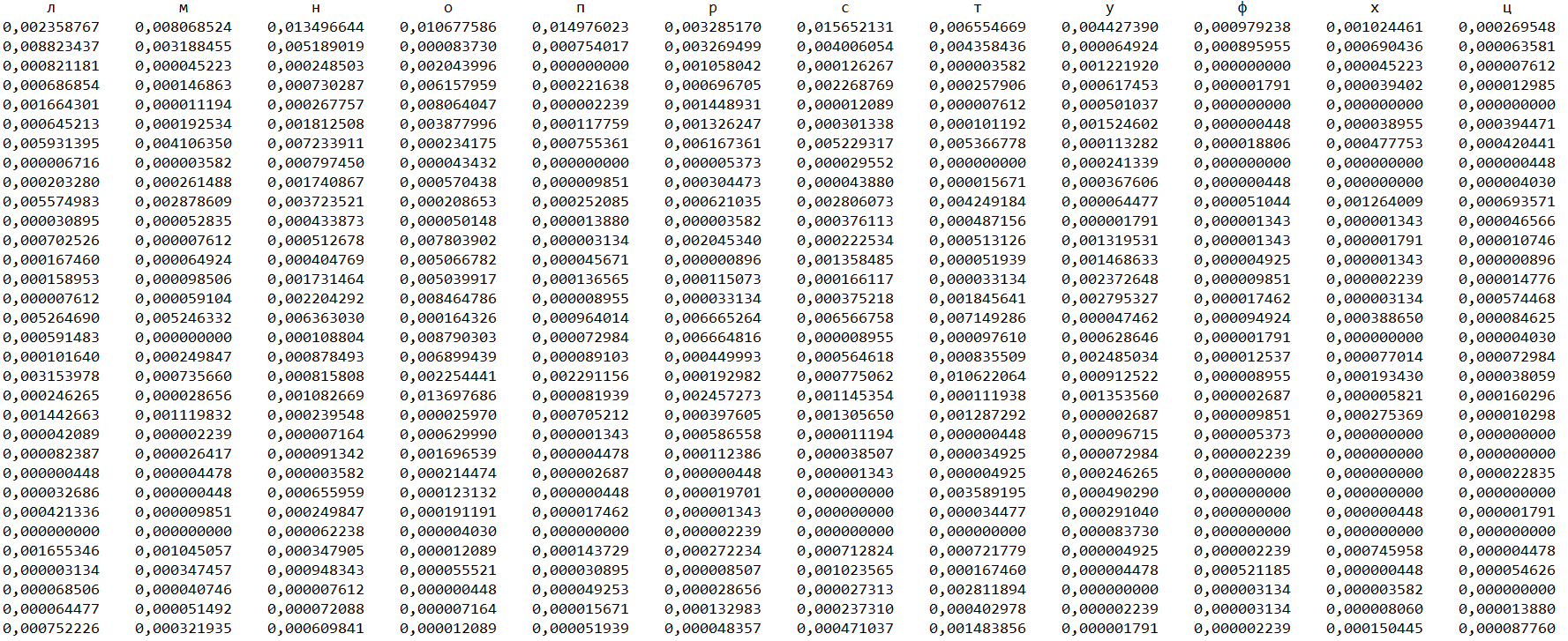
****

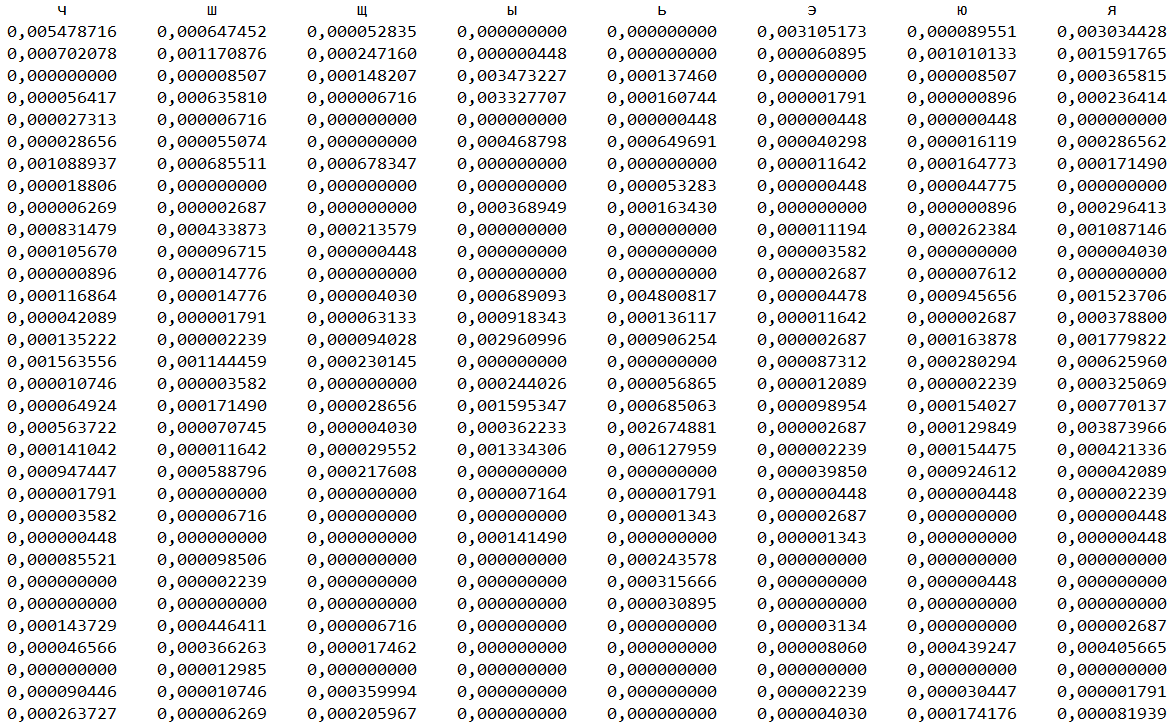
****

****

Частота біграм без пробілу з кроком 1







Частота біграм без пробілу з кроком 1

