Міністерство освіти і науки України Національний технічний університет України «Київський політехнічний інститут» Фізико-технічний інститут

**Компютерний практикум №1**

**З дисципліни:”Криптографія”**

**Тема:**” Експериментальна оцінка ентропії на символ джерела відкритого тексту”

|  |  |
| --- | --- |
| Перевірила:  Селюх К. І.  \_\_\_\_\_\_\_\_\_\_\_ | Виконали:  студенти ІІІ курсу  групи ФБ-95  Корольова В.Р.  групи ФБ-96  Гуменюк О.О. |

**Мета**: Застосовувати практичні навички,при оцінюванні ентропії на символ джерела,з дослідженням ентропіїї на символ джерела та надлишковості,Досягти розуміння та правильного порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії.

**Постановка задачі**: У даному практичному практикумi,досліджуємо частотності букв та біграм російської мови ,ентропії відповідно до букв(Н1) та біграм(Н2),щодо різних ситуацій по відношенню до тексту,тобто чи ми буднмо обробляти з/без пробілами/ів з заміною ймовірностей на частотами.

**Хід роботи:**

1)Написали програму для підрахунку частот букв/біграм.

2)Підрахували H1 та H2

3)Підрахували частоти біграм та H1 та H2 відповідно для тексту з/без пробілами/ів.

4)Оцінили значення Н(10), Н(20), Н(30)

4)Оцінили надлишковість російської мови в різних мовних джерелах

**Труднощі\_1:**

Вивід файлу-некоректне перенесення даних до заданого масиву.

**Рішення:**

Ми повинні були зчитувати текстовий файл та переносити дані в масив.Після цього обробляли масив зручним чином.

**Труднощі\_2:**

Проблема з переведення у нижній регістр букви верхнього регістру російської мови.Ми вирішили взяти функцію tolower(),проте вона спрацьовувала неадекватним чином на російських літерах.

**Рішення:**

Написали свою функцію, **tolower\_rus()**

Що приймала значення,оброблювала та виводила коректно у інший файл. .

**Труднощі\_3:**

Проблема з видалення лишніх символів.

**Рішення:**

Зробили масив з літер з циклом,при якому перевірялись дані нашого масиву і ,якщо елемент не є однією з літер масиву,то ми його видаляли .Тут було чимало складнощів ,тому прийняли рішення ,що треба зробити через булівську функцію,що приймає значення і порівнює символи ,якщо вона виводить значення **true**,то функція не видаляє літеру,якщо **false**,то відбувається видалення.Далі ми це ставимо у функцію вводу до файлу ,використовуємо ту ж самц функцію,яку використовували при видалені регістрів.Записуємо значення масиву даних до файлу.

**Труднощі\_4:**

Проблема з біграмами.

Складнощі з коректним записом даних ,адже було декілька варіантів аби записувати дані, як елементи масиву та робити це циклічно ,щоб воно записувало,проте виникла проблема з адекватним записом до файлу.

**Рішення:**

Функція для передачі виводу з консолі до файлу ,що мала назву **freeopen().**

**Труднощі\_5:**

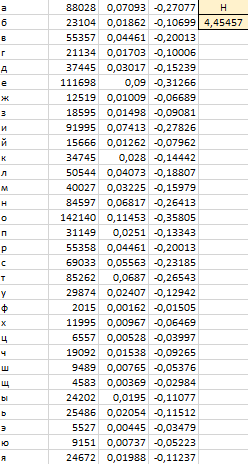
Проблема:після того,як ми закінчували вивід значень з консолі до файлу,навіть ,коли сама функція скінчилась і поток її також закривався.У нас після цього,було не можливо вивести щось у консоль,воно не записувалось у файл, ні виводилось у консоль.

**Рішення:**

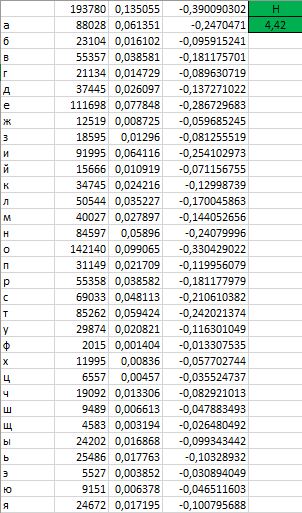
Використання буферу ,за допомогою функцій ifstream/ofstreambuff тим самим способом.

Таблиці частот букв

**Букви без пробілів**

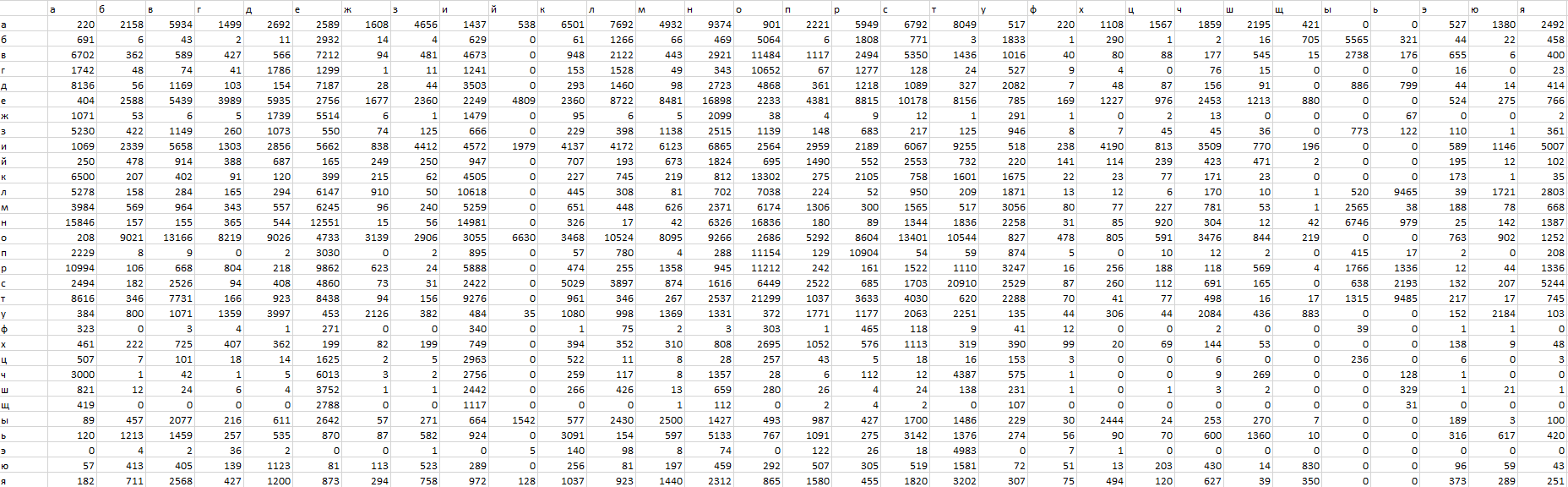


**Букви з пробілами**

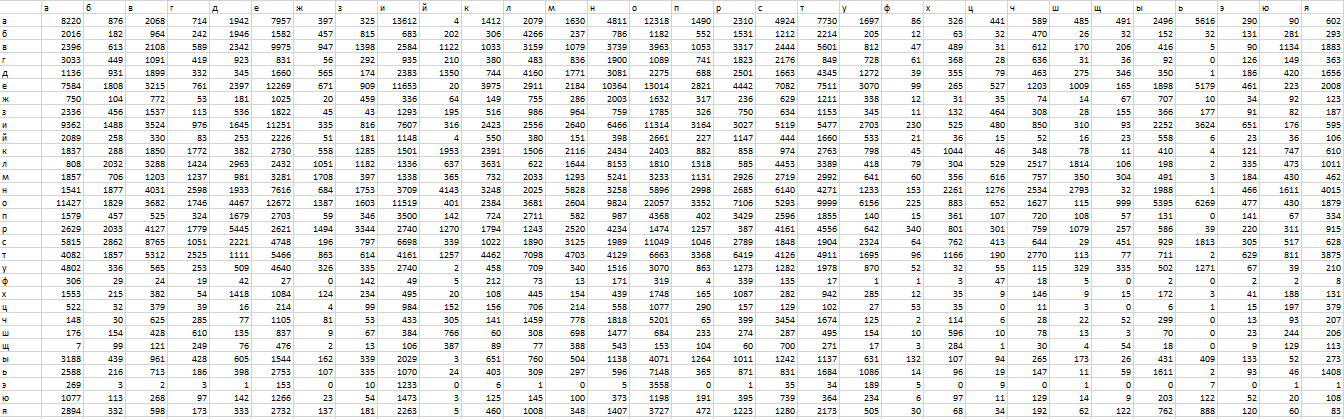


Таблиці частот біграм

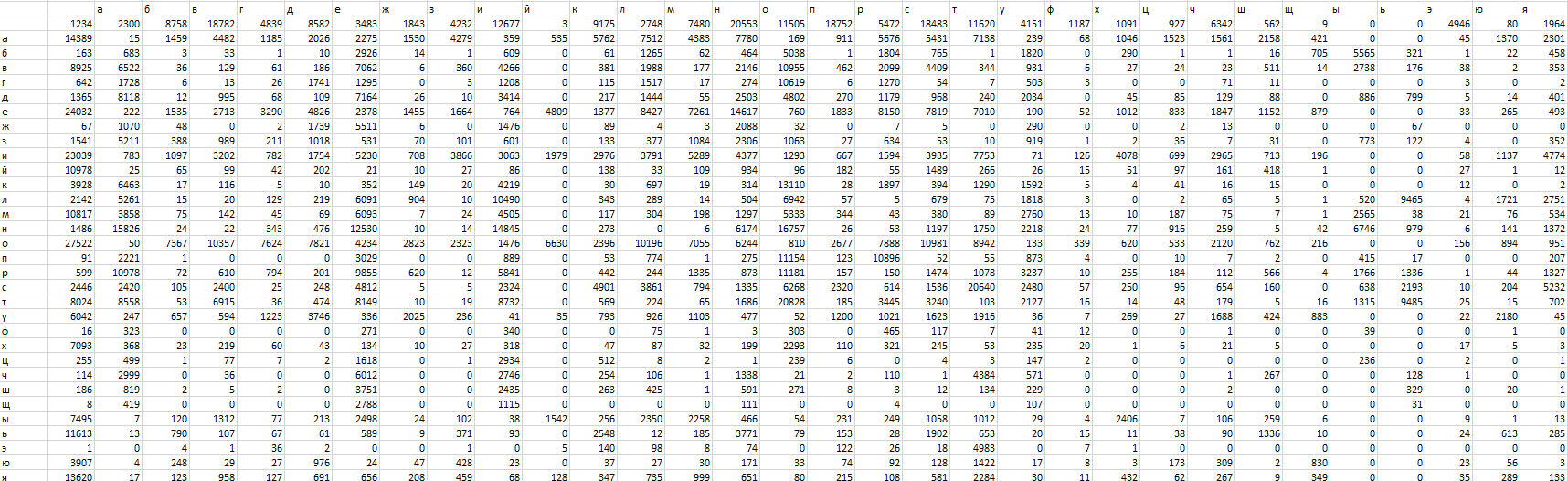
**Мал.Біграми без пробілів з кроком\_1(перетинається)**



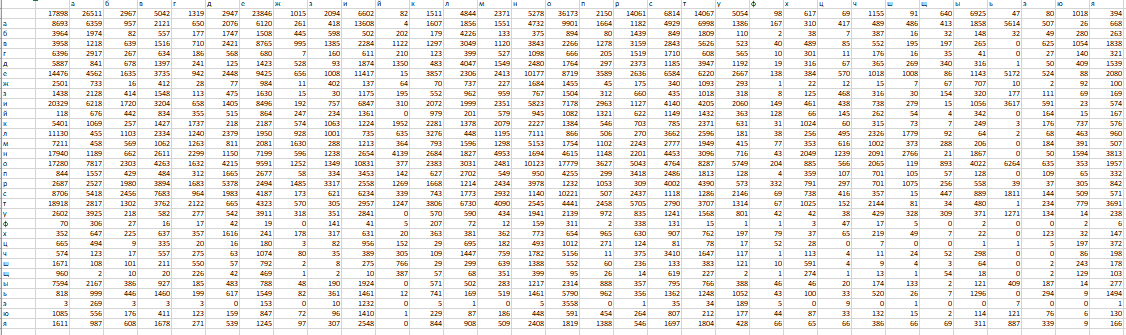
**Мал.Біграми без пробілів з кроком\_2(не перетинається)**



**Мал. Біграми з пробілами з кроком\_1(перетинається)**



**Мал. Біграми з пробілами з кроком\_2(не перетинається)**



Одержані значення H1 та H2

**Биграмка с пробелами з шагом\_2**



**Биграмка с пробелами з шагом\_1**

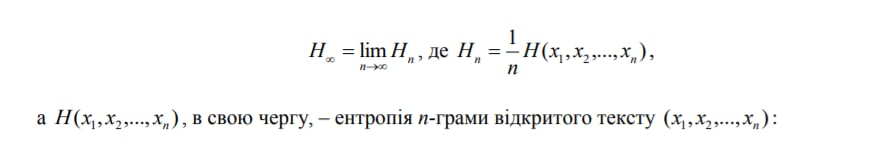


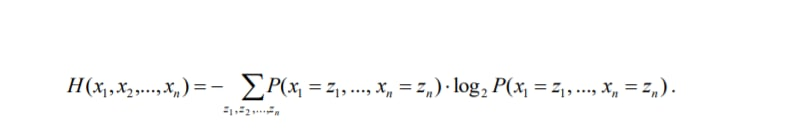
**Биграмка\_без\_пробелов\_з\_шагом\_1**



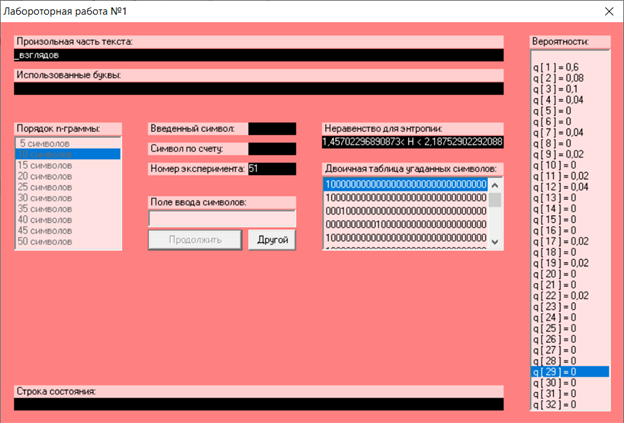
**Биграмка\_без\_пробелов\_з\_шагом\_2**





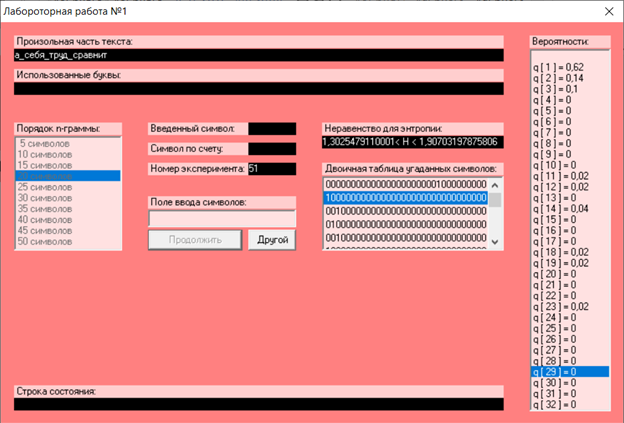


Н10



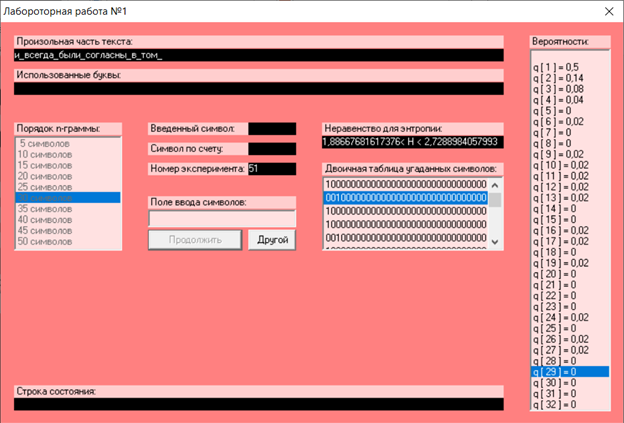
**R0,708595<R10<0,562494**

Н20



**0,739490 <R20< 0,618594**

Н30



**0,622664637<R30< 0,454220319**

Оцінку надлишковості R російської мови у різних моделях відкритого тексту;

**Висновки:**Було засвоєнно навички з підрахунком ентропії на символ джерела,що довзволило розумітиссь на невизначеності інформації у відповідному тексті ,досліджено ентропію,значення якої =0,у випадку ,коли ми знали б яка була наступна буква в біграмі ,відповідно при ентропії=1,ми не можемо знати результат.А також практично дослідили та проаналізували значення надлишковості російської мови,відповідно при різних конфігураціях.