Міністерство освіти і науки України

Національний технічний університет України

“Київський політехнічний інститут ім. Ігоря Сікорського”

Фізико-технічний інститут

**КРИПТОГРАФІЯ**

**КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1**

**Експериментальна оцінка ентропії на символ джерела**

**відкритого тексту**

Виконали:

Студенти групи ФБ-05

Акбаров Елшодбек, Олифіренко Іван

Варіант 7

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

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

**Хід роботи**

Пишемо програму 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. Для коду використовуємо модулі: re – для модифікації тексту, math – отримати логарифм у формулі надлишковості, collections – використання Counter для підрахунків в тексті, pandas – запису в файл.

В результаті роботи було створено програму з функціями: очистки тексту, обчислення ентропії, частоту букв, підрахунок біграм, частота біграм, надлишковість.

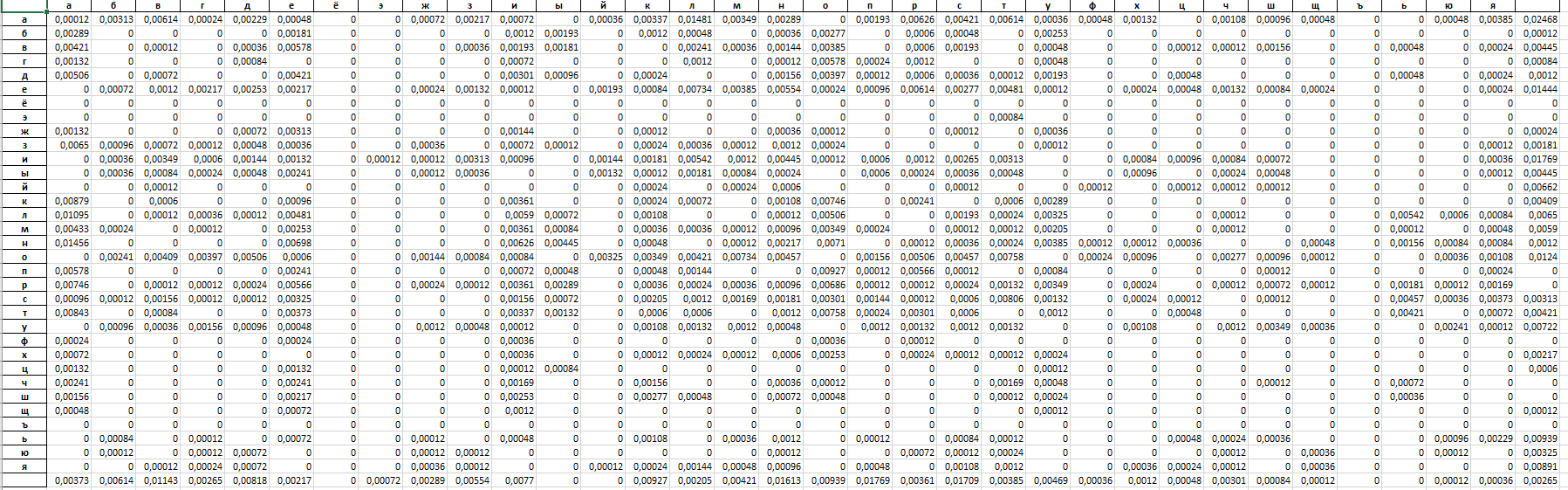
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **а** | 0,10937 |  | **а** | 0,09315 |
| **б** | 0,01922 |  | **б** | 0,01637 |
| **в** | 0,03829 |  | **в** | 0,03262 |
| **г** | 0,01498 |  | **г** | 0,01276 |
| **д** | 0,02967 |  | **д** | 0,02527 |
| **е** | 0,07376 |  | **е** | 0,06282 |
| **ё** | 0 |  | **ё** | 0 |
| **э** | 0,00099 |  | **э** | 0,00084 |
| **ж** | 0,00933 |  | **ж** | 0,00794 |
| **з** | 0,0171 |  | **з** | 0,01456 |
| **и** | 0,06458 |  | **и** | 0,055 |
| **ы** | 0,02006 |  | **ы** | 0,01709 |
| **й** | 0,00989 |  | **й** | 0,00842 |
| **к** | 0,03928 |  | **к** | 0,03346 |
| **л** | 0,05652 |  | **л** | 0,04814 |
| **м** | 0,03066 |  | **м** | 0,02612 |
| **н** | 0,06133 |  | **н** | 0,05223 |
| **о** | 0,09368 |  | **о** | 0,07979 |
| **п** | 0,0325 |  | **п** | 0,02768 |
| **р** | 0,04621 |  | **р** | 0,03935 |
| **с** | 0,04931 |  | **с** | 0,042 |
| **т** | 0,04974 |  | **т** | 0,04236 |
| **у** | 0,0366 |  | **у** | 0,03117 |
| **ф** | 0,00155 |  | **ф** | 0,00132 |
| **х** | 0,0089 |  | **х** | 0,00758 |
| **ц** | 0,00509 |  | **ц** | 0,00433 |
| **ч** | 0,01357 |  | **ч** | 0,01155 |
| **ш** | 0,01342 |  | **ш** | 0,01143 |
| **щ** | 0,00311 |  | **щ** | 0,00265 |
| **ъ** | 0 |  | **ъ** | 0 |
| **ь** | 0,02317 |  | **ь** | 0,01974 |
| **ю** | 0,00749 |  | **ю** | 0,00638 |
| **я** | 0,02063 |  | **я** | 0,01757 |
|  |  |  |  | 0,14827 |

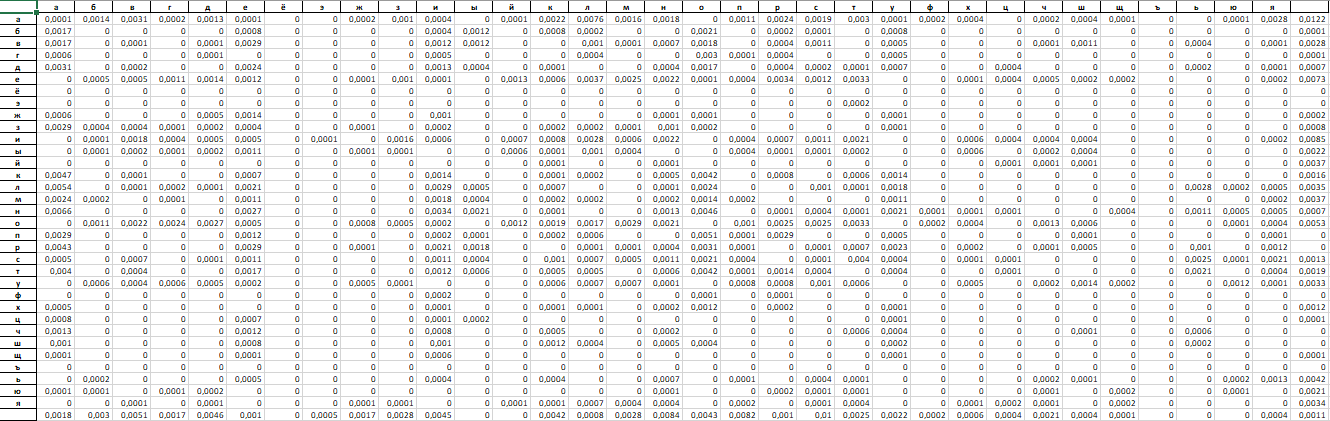
Зліва частота літер в тексті без пробілів, справа – з.

Ентропія монограм без пробілами: H1 = 4.47548549103316, R = 0.11278036863575747

Ентропія монограм з пробілів: H1 = 4.417222324314814, R = 0.13174356999741765

Частота перехресниї біграм з пробілами:

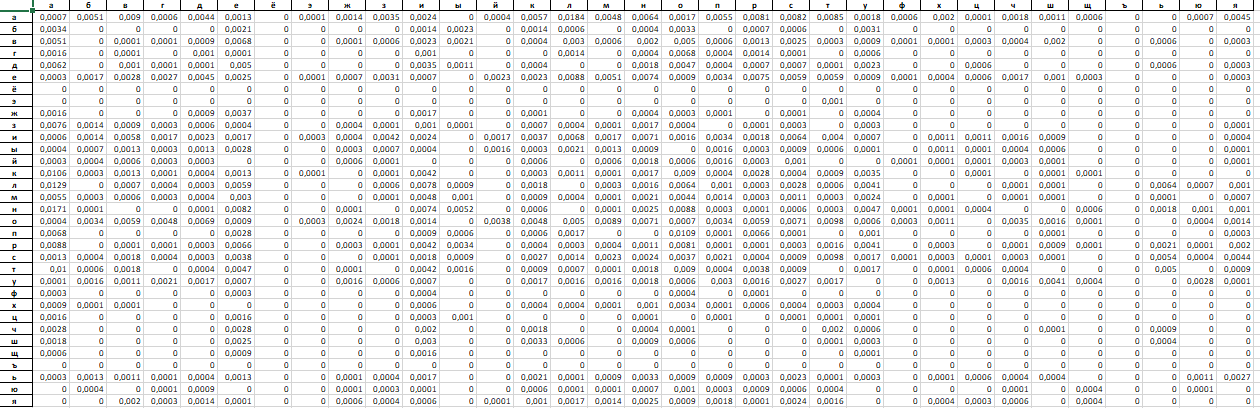
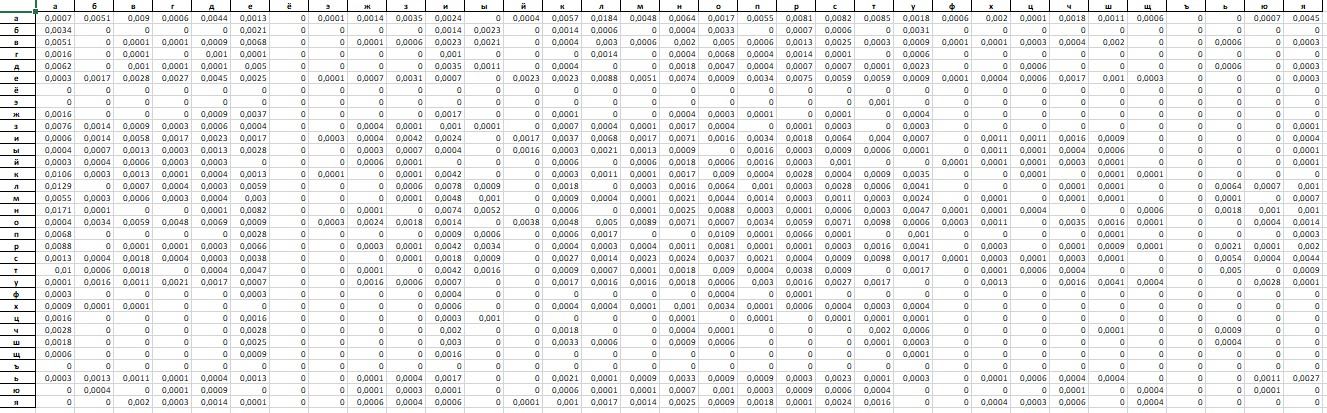




H = 2.2311330606229256

R = 0.5614448438753445

Частота перехресних біграм без пробілів:

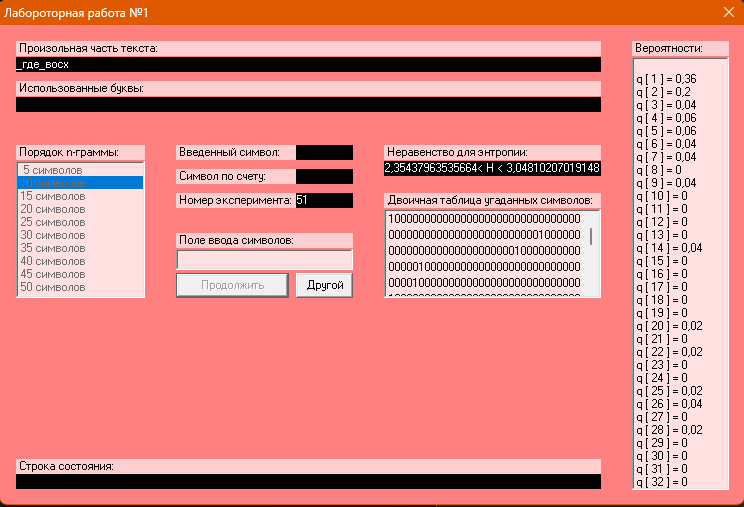
 

H = 2.2843090057347566

R = 0.5471588952638626

Робота з PinkProgramm:

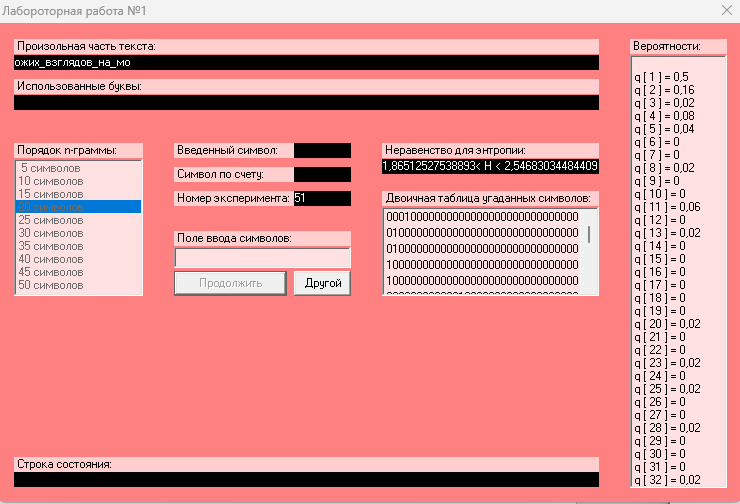
10 символів:



2,3543796353 < H10 < 3,0481020701

0,39037958598 < R < 0,52912407294

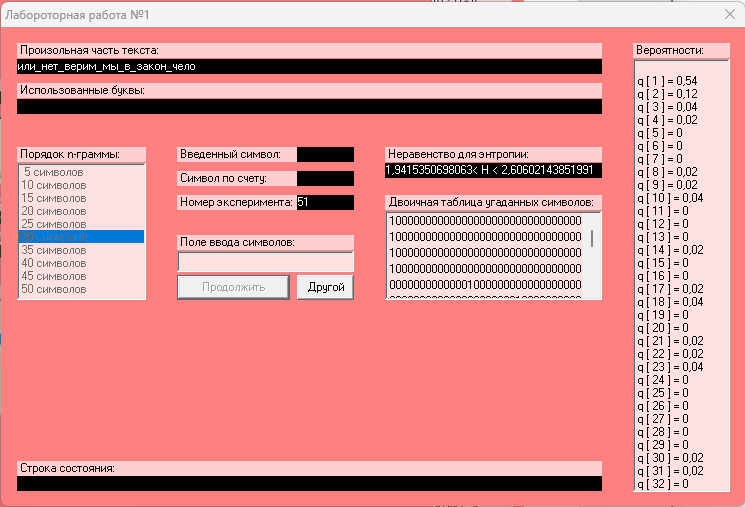
20 символів:



1,8651252753 < H20 < 2,5468303448

0,49063393104 < R < 0,62697494494

30 символів:



1,9415350698 < H30 < 2,6060214385

0,4787957123 < R < 0,61169298604

**Висновки:**

У ході виконання лабороторної роботи, ми здобули навички обраховувати ентропію, надлишковість. Ознайомились з программою, за допомогою якої можна було провести експеримети на текстах. Через яку рахувалися частоти монограм та біграм.