Міністерство освіти і науки України

Національний технічний університет України

"Київський політехнічний інститут імені Ігоря Сікорського"

Фізико-технічний інститут

КРИПТОГРАФІЯ

Комп'ютерний практикум

Робота № 1

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав:

студент гр. ФБ-02

Шубін Д.Ю

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:

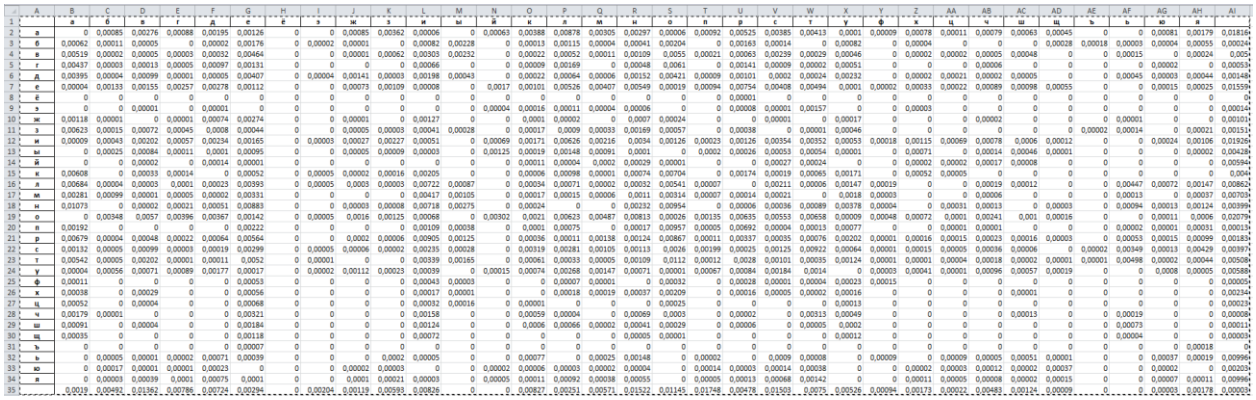
1) Частоти букв у тексті з пробілами

а	0,06953
б	0,01352
в	0,03372
г	0,0184
д	0,02625
е	0,06564
ё	0,00001
э	0,00232
ж	0,00813
з	0,01596
и	0,05901
ы	0,01364
й	0,00757
к	0,02668
л	0,04597
м	0,02668
н	0,05415
о	0,09251
п	0,02464
р	0,0472
с	0,04492
т	0,04783
у	0,02546
ф	0,00228
х	0,00696
ц	0,00236
ч	0,01241
ш	0,00718
щ	0,00248
ъ	0,00024
ь	0,01626
ю	0,00389
я	0,01658
	0,15963

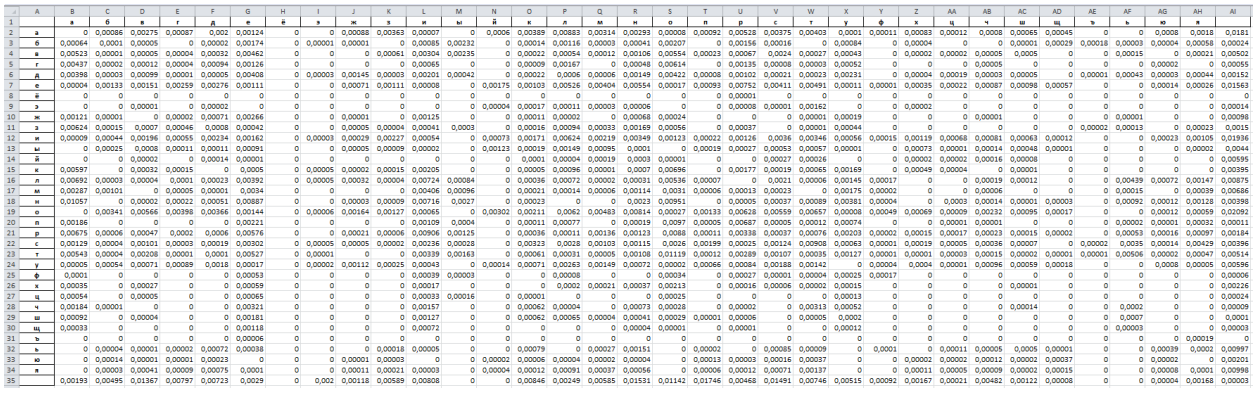
2) Частоти букв у тексті без пробілів

а	0,08274
б	0,01608
в	0,04013
г	0,02189
д	0,03123
е	0,07811
ё	0,00002
э	0,00276
ж	0,00967
з	0,019
и	0,07022
ы	0,01623
й	0,009
к	0,03175
л	0,0547
м	0,03174
н	0,06443
о	0,11009
п	0,02932
р	0,05617
с	0,05345
т	0,05691
у	0,0303
ф	0,00271
х	0,00828
ц	0,00281
ч	0,01477
ш	0,00855
щ	0,00295
ъ	0,00029
ь	0,01934
ю	0,00463
я	0,01973

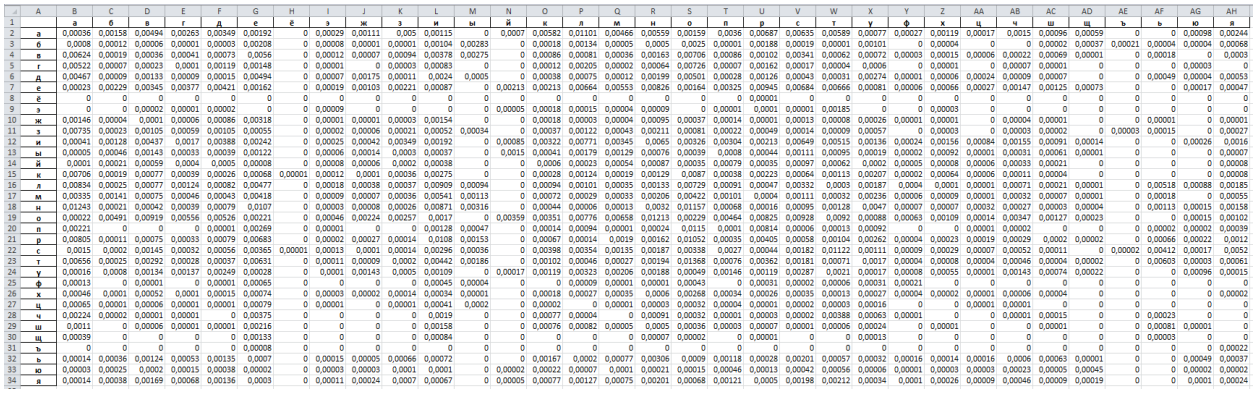
3) Частота біграм у тексті з пробілами



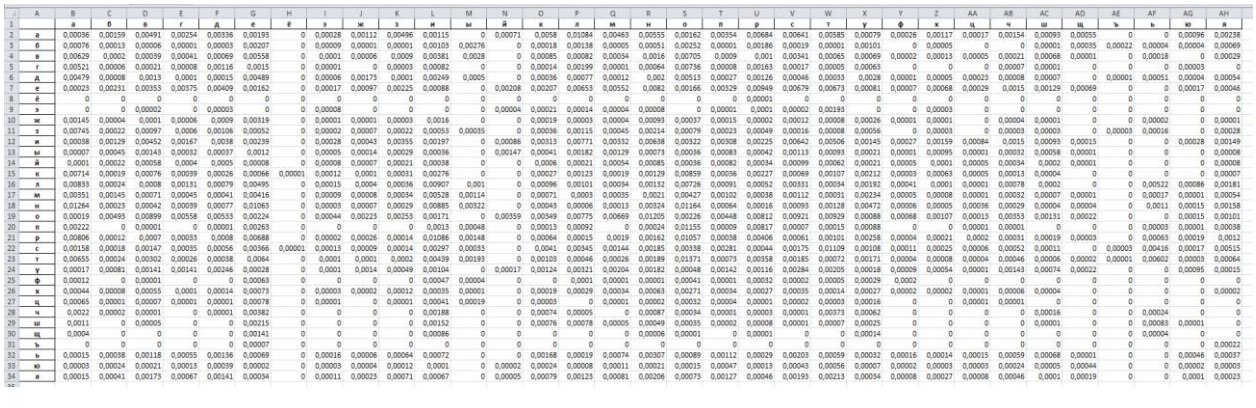
4) Частота біграм в тексті з пробілами, з 2-им кроком



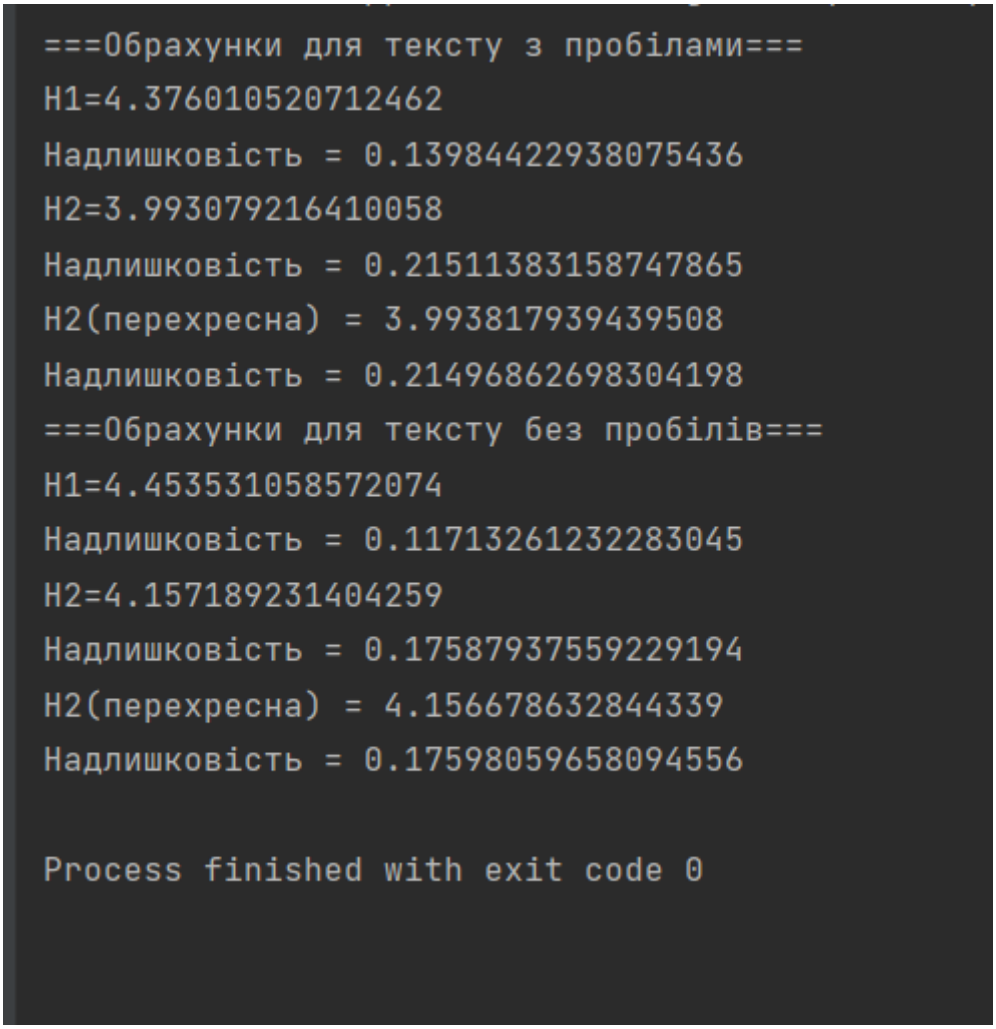
5) Частота біграм в тексті без пробілів



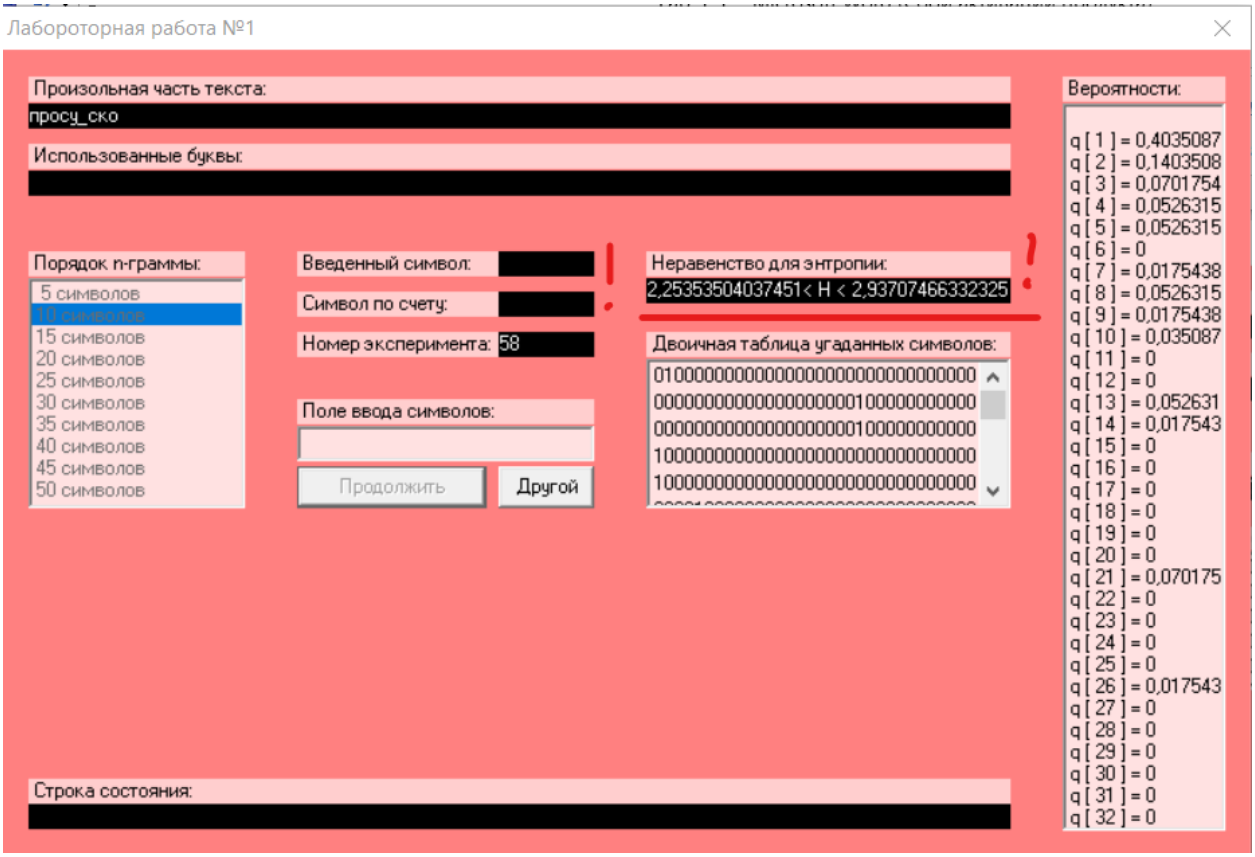
6) Частота біграм в тексті без пробілів, з 2-им кроком



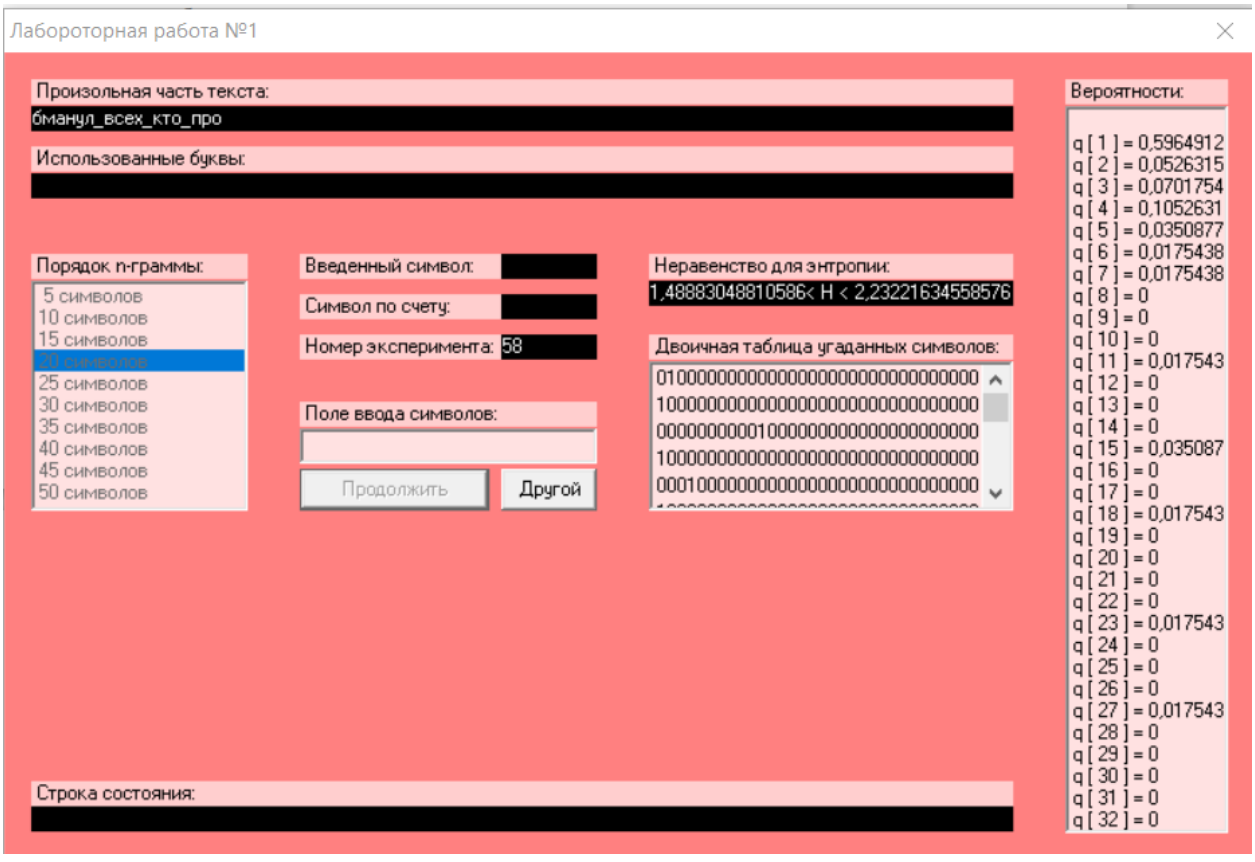
7) За допомогою скрипта Lab_1.py дізнаємось ентропію та надлишковість:



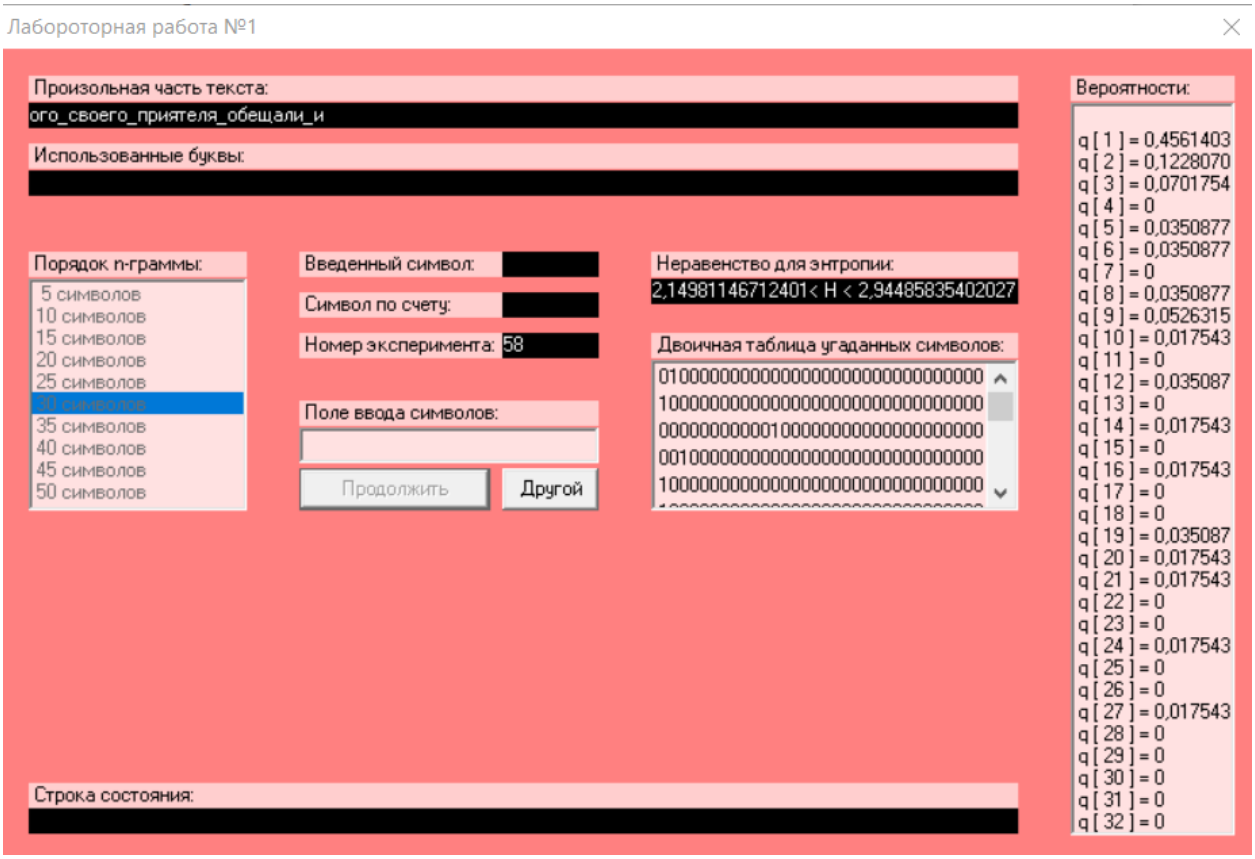
8) Заходимо у програму *CoolPinkProgram*
9) Завдяки програмі оцінили значення H10 (10 символів)



10) Завдяки програмі оцінили значення H20 (20 символів)



11) Завдяки програмі оцінили значення H30 (30 символів)



12) За знайденими значеннями знайдемо надлишковість, підстави значення у формулу, з скріпта Lab_1.py:

```
print('Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H10 (10 символів)')
h10 = ((2.25353504037451+2.93707466332325)/2)
rh10 = rozr_nadl(h10, len(alf_a))
print(rh10)

print('Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H20 (20 символів)')
h20 = ((1.48883048810586+2.23221634558576)/2)
rh20 = rozr_nadl(h20, len(alf_a))
print(rh20)

print('Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H30 (30 символів)')
h30 = ((2.14981146712401+2.94485835402027)/2)
rh30 = rozr_nadl(h30, len(alf_a))
print(rh30)
```

```
Lab_1_2
Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H10 (10 символів)
0.4855071212835862
Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H20 (20 символів)
0.6311700924188628
Розрахуємо надлишковіть для виводу программи CoolPinkProgram у випадку значення H30 (30 символів)
0.495016675878587
```

Висновок: у ході лабораторної роботи, я на практиці розібрався з поняттям ентропії, надлишковості, реалізувавши їх на мові програмування Python. Ознайомився з вивченням

та порівнянням різних моделей джерела відкритого тексту для наближеного визначення ентропії, набув практичних навичок щодо оцінки ентропії на символу джерела.