НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра Інформаційної Безпеки

Лабораторна робота №1 дисципліни

”КРИПТОГРАФІЯ”

Підготував:

студент групи ФБ-06

Жак Костянтин

Київ 2022

Тема роботи: Експериментальна оцінка ентропії на символ джерела відкритого тексту

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 $\square_1$ та $\square_2$ за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення $\square_1$ та $\square_2$ на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення $\square_1$ та $\square_2$ на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $\square^{(10)}$, $\square^{(20)}$, $\square^{(30)}$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

За початковий текст взяв "Хроніки нарнії"
Значення частот букв з пробілами:

Буква	Її частотність
а	0.06865
б	0.01302
в	0.03577
г	0.0144
д	0.0275
е	0.0623
ж	0.00803
з	0.01467
и	0.05746
й	0.00799
к	0.02897
л	0.04541
м	0.02643
н	0.05237
о	0.09153
п	0.02322
р	0.03641
с	0.04534
т	0.04996
у	0.02378
ф	0.00066
х	0.00749
ц	0.00244
ч	0.01247
ш	0.00784
щ	0.00225
ы	0.01605
ь	0.01802
э	0.00314
ю	0.0049
я	0.0159
пробел	0.17026

Ентропія:

```
4.354901763494896
```

Надлишковість:

```
0.1290196473010209
```

Значення частот букв без пробілів:

Буква	Її частотність
а	0.08274
б	0.01569
в	0.04312
г	0.01735
д	0.03315
е	0.07508
ж	0.00968
з	0.01768
и	0.06925
й	0.00963
к	0.03491
л	0.05473
м	0.03185
н	0.06312
о	0.11031
п	0.02799
р	0.04388
с	0.05465
т	0.06021
у	0.02866
ф	0.0008
х	0.00903
ц	0.00294
ч	0.01503
ш	0.00944
щ	0.00271
ы	0.01935
ь	0.02172
э	0.00378
ю	0.0059
я	0.01917

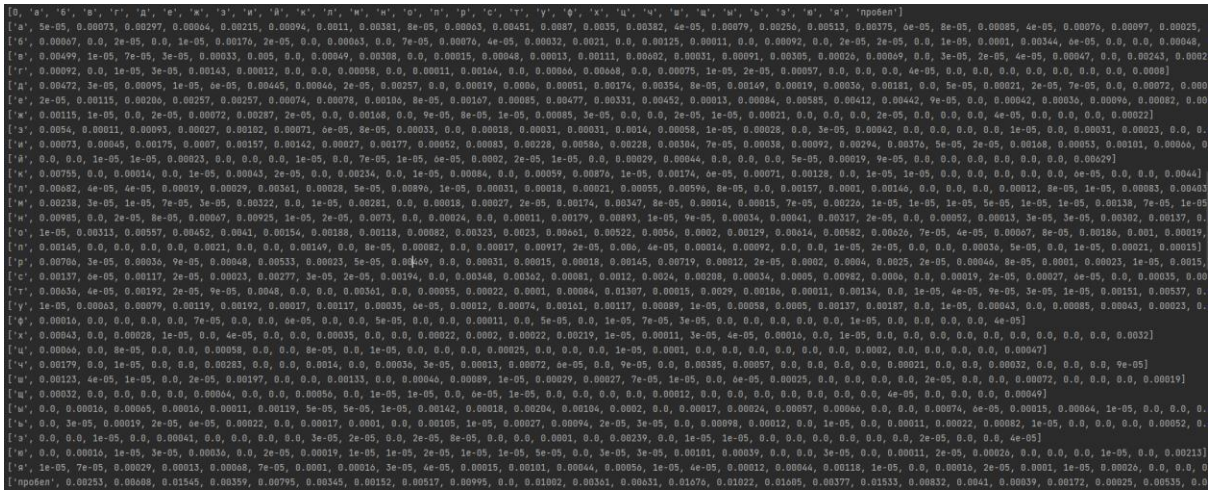
Ентропія:

```
4.45752152505319
```

Надлишковість:

```
0.10025335174796479
```

Значення частот біграм що перетинаються з пробілами:



Ентропія:

```
3.968282216187312
```

Надлишковість:

```
0.6031717783812688
```

Значення частот біграм що перетинаються без пробілів:

Ентропія:

```
3.968282216187312
```

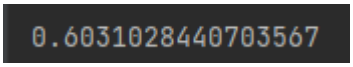
Надлишковість:

```
0.5995029296813001
```

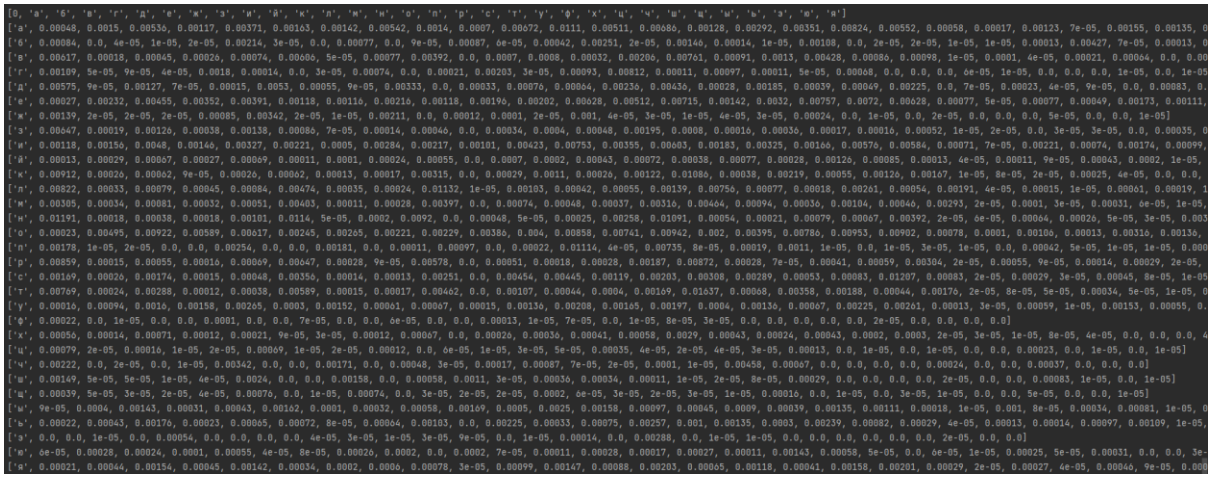
Значення частот біграм що не переитнаються з пробілами:

Ентропія:

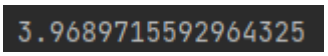
Надлишковість:



Значення частот біграм що не перетинаються без пробілів:



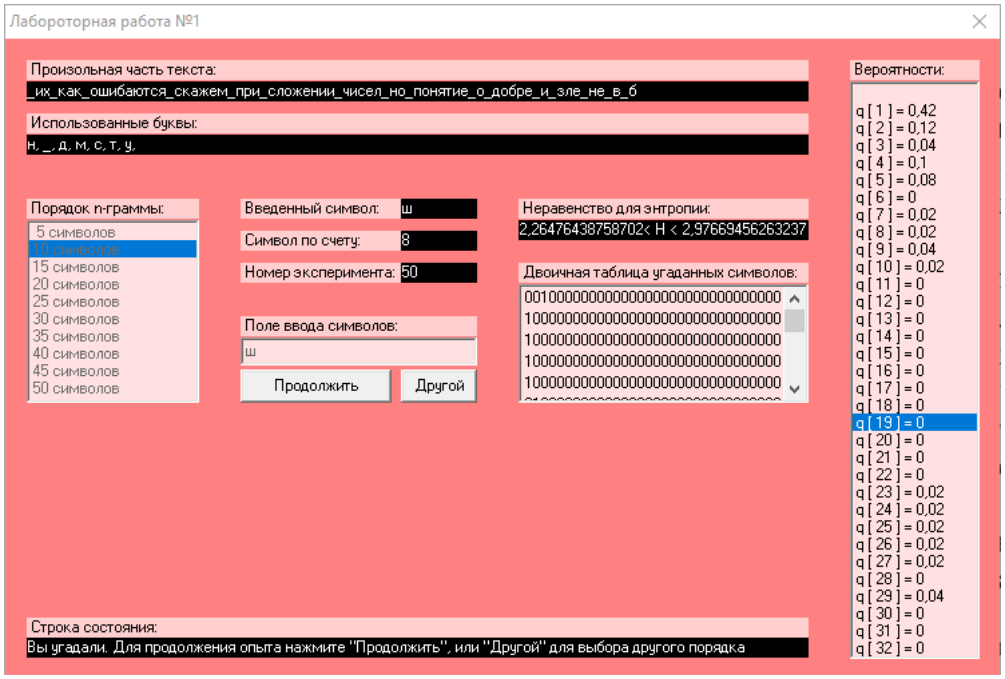
Ентропія:



Надлишковість:

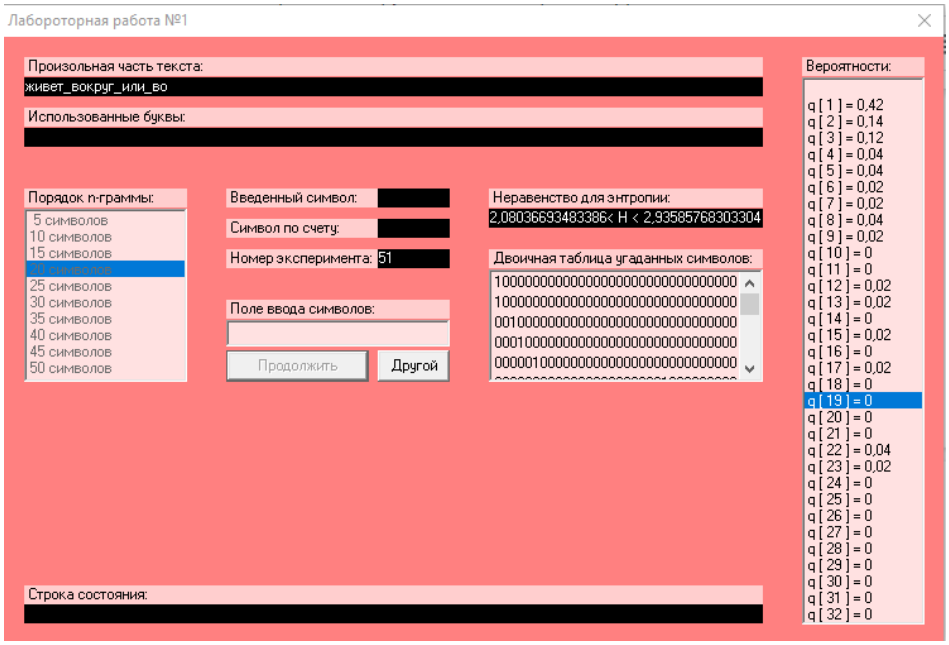
2. CoolPinkProgram

1. □ *(10)*



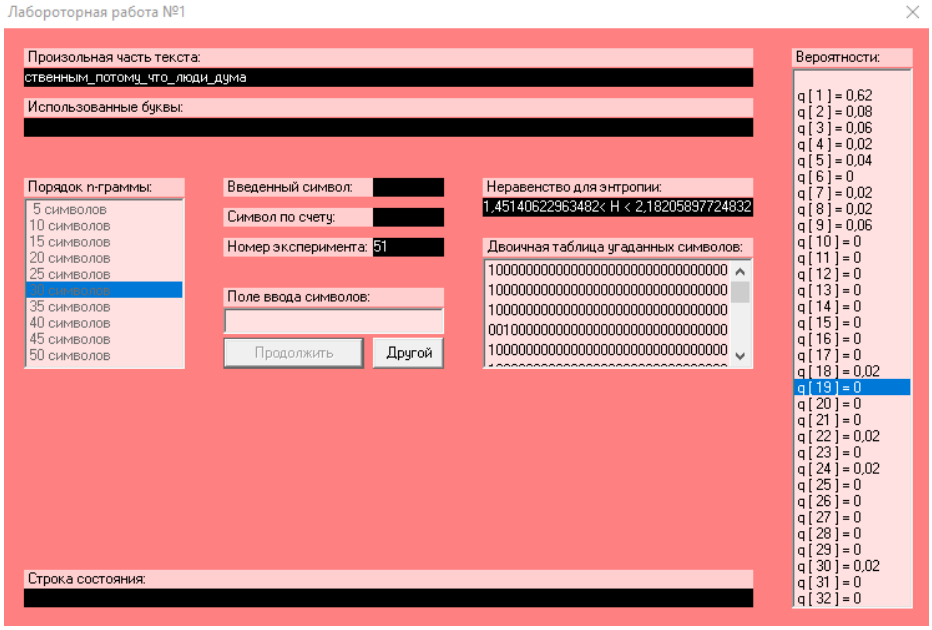
$2.265 < \square^{(10)} < 2.980$

2. $\square^{(20)}$



$2.080 < \square^{(20)} < 2.940$

3. $\square^{(30)}$



$1.451 < \square^{(30)} < 2.182$

3. Оцінимо надлишковість на основі отриманих результатів:

Для нижньої границі:

1) R1 = $1 - \frac{2.26476438758702}{5}$ = 0.547047122482596
2) R2 = 0.583926613033228
3) R3 = 0.709718754073036

Для верхньої границі:

1) R1 = 0.404661087473526
2) R2 = 0.412828463393392
3) R3 = 0.563588204550336

Висновок

Під час виконання даного лабораторного практикуму було написано програму, що обраховує частотність літер та біграм російської мови на основі тексту “Хроніки Нарнії”, на основі цих результатів було обраховано надлишковість російської мови. У шифруванні слід використовувати текст без пробілів, так як ентропія в ньому більша.

Також навчився працювати з програмою CoolPinkProgram.