

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО» ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра Інформаційної Безпеки

Лабораторна робота №1 дисципліни

"КРИПТОГРАФІЯ"

Підготували:

студенти групи ФБ-03

Борох Іван

Жигун Анастасія

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 H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення $\boldsymbol{H}^{(10)}$, $\boldsymbol{H}^{(20)}$, $\boldsymbol{H}^{(30)}$
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

Для виконання даного лабораторного практикуму було вибрано книгу Джорджа Мартіна "Полум'я та кров. Кров драконів" розміром 1,05 МБ.

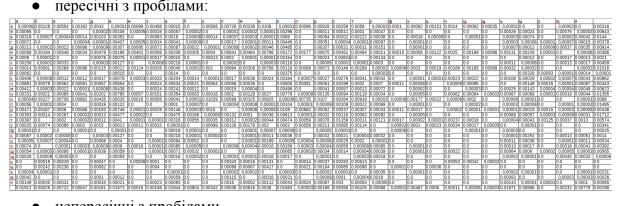
- 1. Програми для підрахунку
- 1. Частоти букв:
 - *whitespace пробіл, *intersecting/non-intersecting пересічні/непересічні (біграми).

```
Table of letter`s with whitespace:
  Probability
     0.157673
     0.093792
     0.076806
     0.068440
                                          0.013589
     0.058849
                                          0.013497
     0.053085
                                          0.013085
     0.047041
                                          0.010028
     0.045121
                                          0.008836
     0.043605
     0.038900
                                          0.008494
     0.028464
                                          0.006367
     0.027058
     0.025147
                                          0.004527
     0.020575
                                          0.004408
     0.019246
                                          0.002643
     0.015694
     0.015570
                                          0.002146
     0.015459
                                          0.000720
     0.013608
```

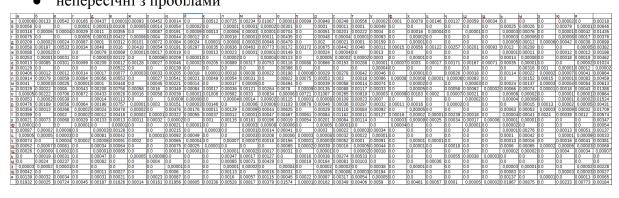
```
Table of letter`s without whitespace:
   Probability
      0.111349
0
      0.091183
      0.081251
      0.069864
      0.063022
      0.056425
p
      0.055847
      0.053568
                                                 0.016133
      0.051767
                                                 0.016023
      0.046181
                                                 0.015534
      0.033792
                                                 0.011905
Д
      0.032123
                                                0.010490
      0.029854
                                                0.010084
      0.024426
                                                0.007559
y
      0.022849
                                                0.005375
                                          Ю
      0.018631
                                                0.005233
      0.018485
                                                0.003138
      0.018353
                                                 0.002547
      0.016155
                                          φ
                                                 0.000855
```

2. Частоти біграм:

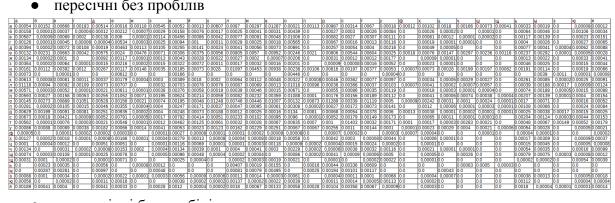
пересічні з пробілами:



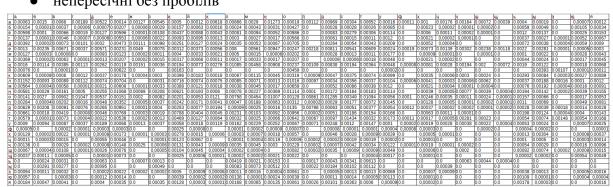
• непересічні з пробілами



пересічні без пробілів



непересічні без пробілів



3. Ентропія:

*redundancy - надлишковість

entropy for letters with whitespaces 4.363475420506432

Redundancy: 0.12730491589871362

entropy for intersecting bigrams with whitespaces 3.967018415628689

Redundancy: 0.20659631687426216

entropy for non-intersecting bigrams with whitespaces 3.9647482371162353

Redundancy: 0.20705035257675297

entropy for letters without whitespaces 4.433861511514542

Redundancy: 0.10502910386926101

entropy for intersecting bigrams without whitespaces 4.117760762303694

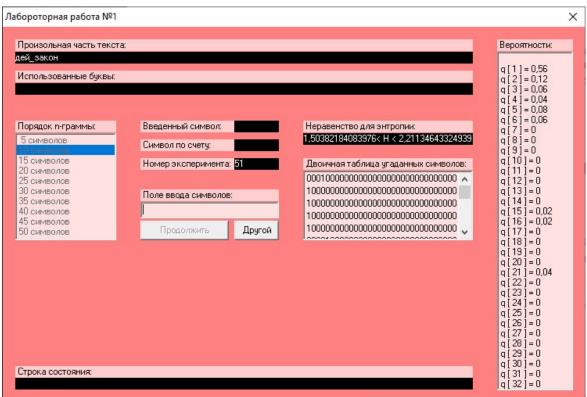
Redundancy: 0.16883375136538825

entropy for non-intersecting bigrams without whitespaces 4.116783483021081

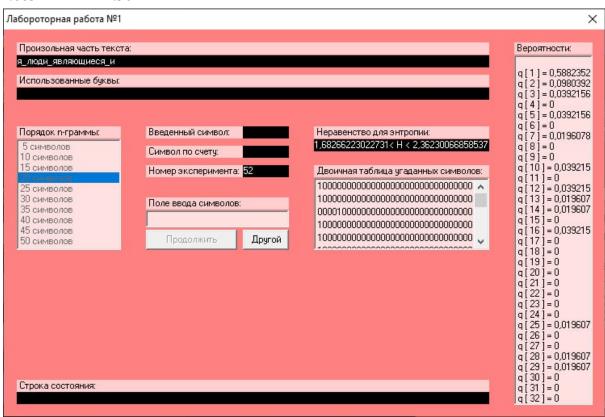
Redundancy: 0.16903101429591927

2. CoolPinkProgram

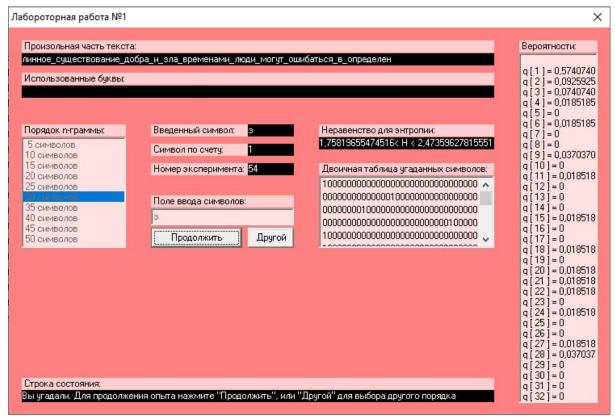
 $1.509 < H^{(10)} < 2.211$



 $1.683 < H^{(20)} < 2.362$



$1.758 < H^{(30)} < 2.473$



3. PinkCoolProgram

Для нижньої границі:

1)
$$R1 = 1 - \frac{1.50382184083976}{5} = 0,699235631832048$$

- 2) R2 = 0.663467553954538
- 3) R3 = 0.648360689050968

Для верхньої границі:

- 1) R1=0,557730713350122
- 2) R2=0.527539866282926
- 3) R3=0.505280744368898

Труднощі, що виникали

- 1) Проблеми з підрахунком ентропії n-грами відкритого тексту та звичайних n-грам, але це чисто через неуважність, пізніше перевірили та знайшли загублені моменти.
- 2) Розрахунок надлишковості джерела відкритого тексту: спочатку старались повторити формулу з методички, використовуючи H_{∞} , але пізніше уважніше перечитали та відшукали правильне подання формули.

Висновки

В ході виконання даної лабораторної роботи ми навчилися правильно форматувати текст за допомогою мови програмування Python та на основі цього тексту вирахували значення ентропій для алфавітів з різними умовами, засвоїли поняття ентропії на символ джерела та його надлишковості, здійснили порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, покращили власні навички з користування Git та GitHub та використання різних бібліотек Python.