МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО"
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Виконали:

Студенти групи ФБ-03
Митрофанова М.М. та Мец Є.В.

Київ – 2022

Тема: «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Мета: 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.

ХІД РОБОТИ

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H^1та H^2за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H^1та H^2на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H^1та H^2на тому ж тексті, в якому вилучено всі пробіли.

- Частоти букв

letters_frequency_no_spaces: Блокнот

Файл Редагування Формат Вигляд Довідка

о - 0.110756
е - 0.084749
а - 0.084186
н - 0.067809
и - 0.064174
т - 0.060555
с - 0.052438
л - 0.049874
р - 0.047218
в - 0.041883
м - 0.034313
к - 0.032182
д - 0.030327
п - 0.02831
у - 0.027785
я - 0.023777
ь - 0.023373
ы - 0.017331
г - 0.016374
з - 0.016098
б - 0.014998
ч - 0.014988
ж - 0.011449
й - 0.010862
ш - 0.009188
х - 0.008726
ю - 0.005188
щ - 0.003617
э - 0.003557
ц - 0.002301
ф - 0.001613

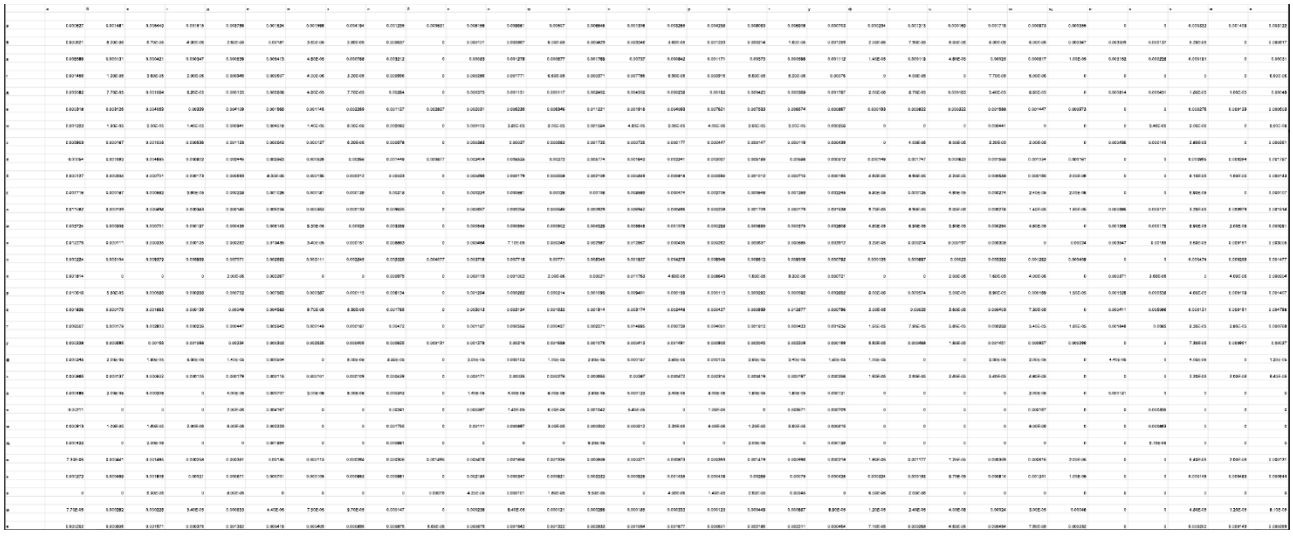
letters_frequency_with_spaces: Блокнот

Файл Редагування Формат Вигляд Довідка

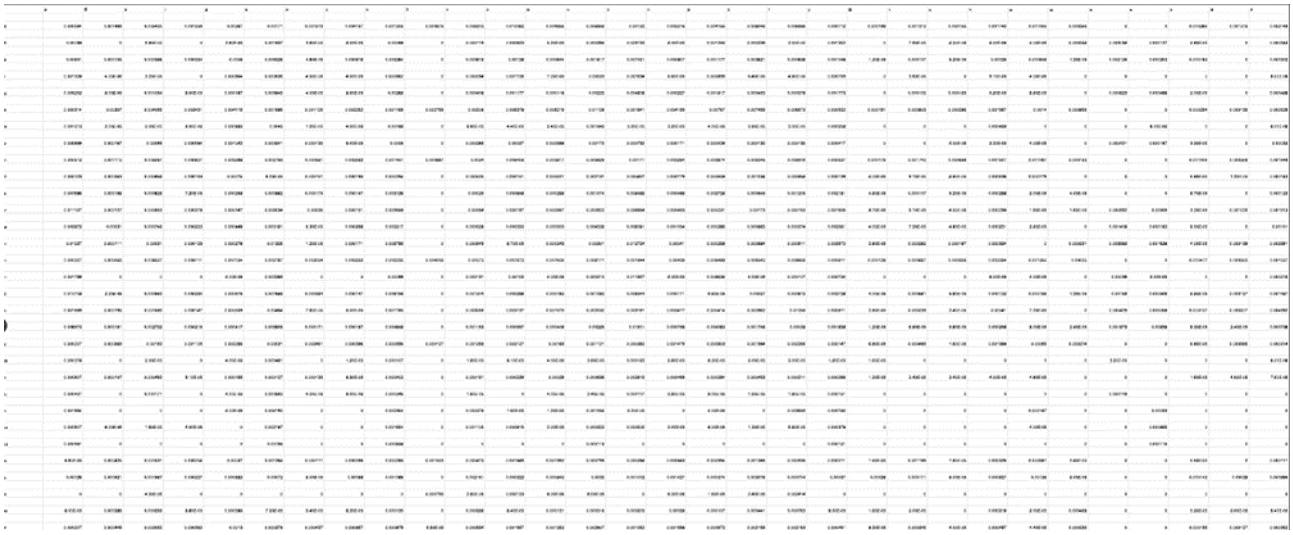
- 0.162192
о - 0.092792
е - 0.071004
а - 0.070532
н - 0.056811
и - 0.053766
т - 0.050734
с - 0.043933
л - 0.041785
р - 0.03956
в - 0.03509
м - 0.028747
к - 0.026963
д - 0.025408
п - 0.023718
у - 0.023278
я - 0.019921
ь - 0.019582
ы - 0.01452
г - 0.013718
з - 0.013487
б - 0.012566
ч - 0.012557
ж - 0.009592
й - 0.0091
ш - 0.007697
х - 0.007311
ю - 0.004347
щ - 0.00303
э - 0.00298
ц - 0.001928
ф - 0.001352

- Частоти біграм

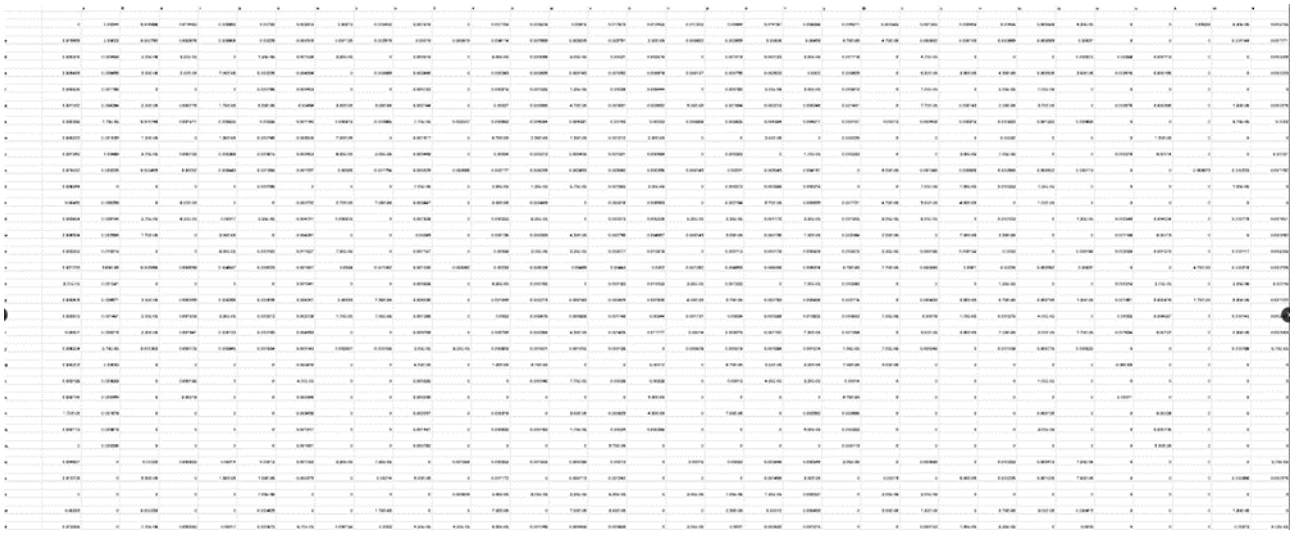
1) bigrams_no_spaces



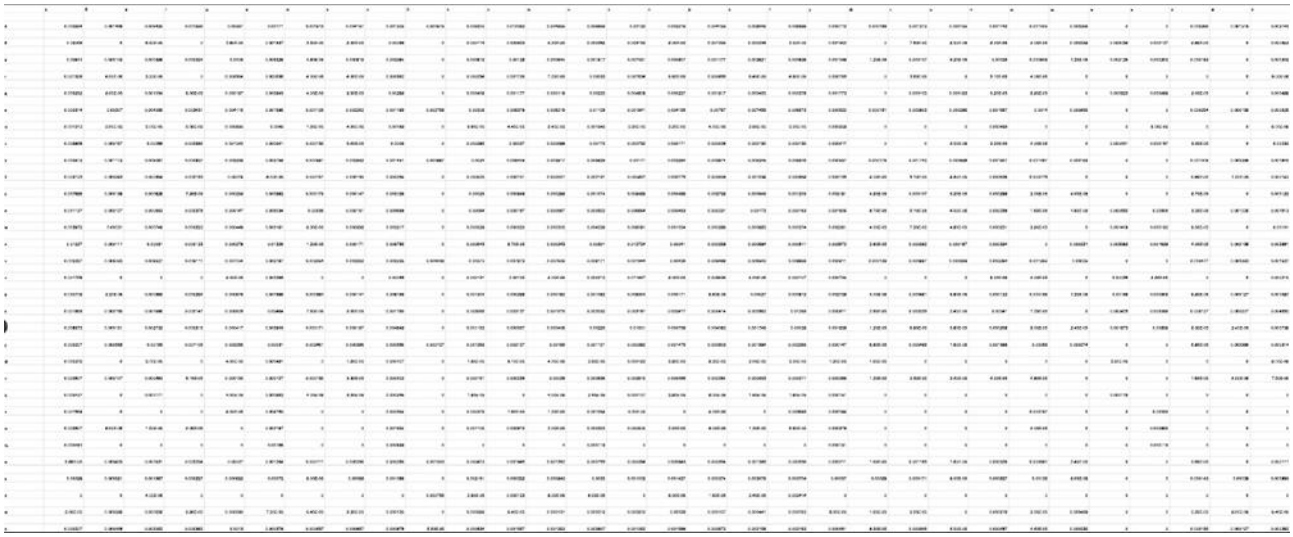
2) bigrams_no_spaces_crossing



3) bigrams_spaces_crossing



4) bigrams_spaces_no_crossing



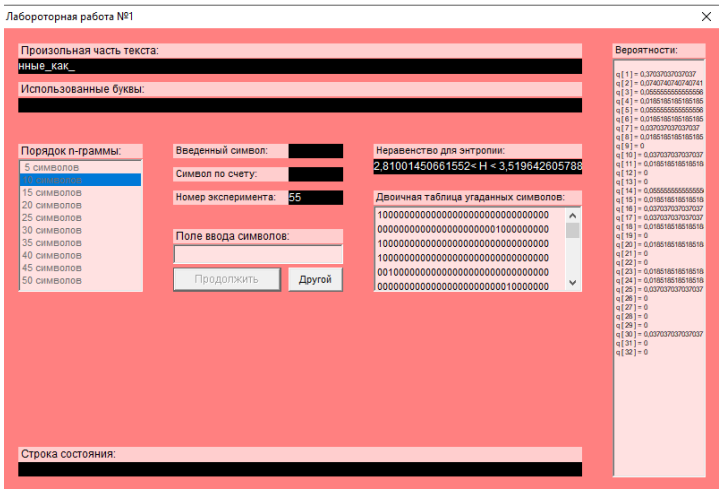
Посилання на папку з табличками:
https://drive.google.com/drive/folders/1pEcjDZcUFTM9kulQN3bzGm1Kw1_mdcdh?usp=sharing

- Ентропія

```
Letters with whitespaces entropy: 4.371059 ; redundancy: 0.125788
Letters without whitespaces entropy: 4.453914 ; redundancy: 0.100982
Bigrams with whitespaces with crossing entropy: 3.9558 ; redundancy: 0.20884
Bigrams with whitespaces without crossing entropy: 3.951953 ; redundancy: 0.209609
Bigrams without whitespaces with crossing entropy: 4.126059 ; redundancy: 0.167159
Bigrams without whitespaces without crossing entropy: 4.12453 ; redundancy: 0.167467
```

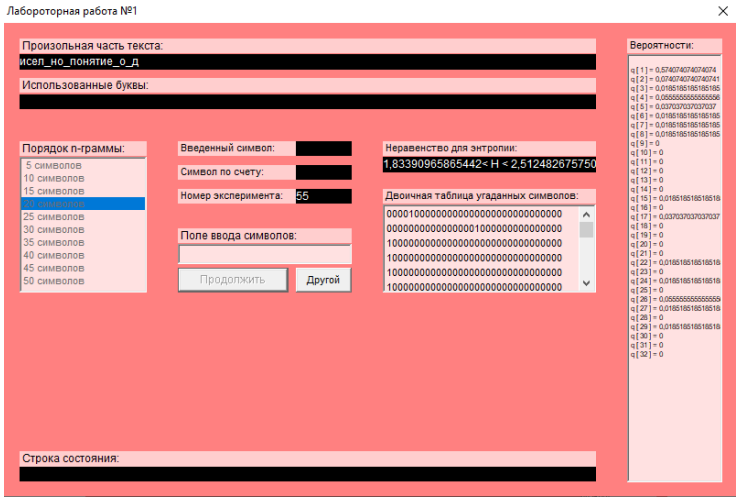
2. За допомогою програми CoolPinkProgram оцінити значення H^{10}, H^{20}, H^{30}.

H^{10}:



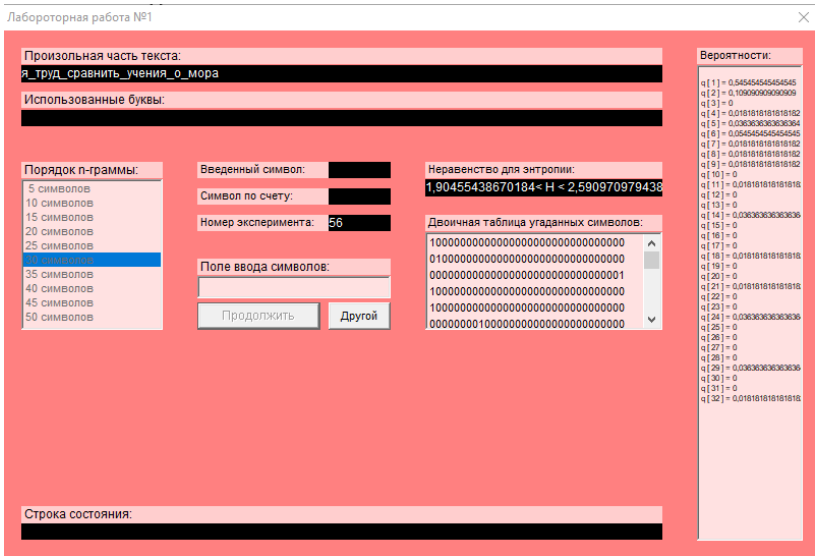
$2{,}8100 < H^{10} < 3{,}5196$

H^{20}:



$1{,}8339 < H^{20} < 2{,}51248$

H^{30}:



$1{,}90455 < H^{20} < 2{,}59097$

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

$$R = 1 - \frac{H_\infty}{H_0}$$

1) $R = 1 - \frac{2{,}81001450661552}{3{,}519642605788} \approx 0{,}201619$

2) $R = 1 - \frac{1{,}83390965865442}{2{,}512482675750} \approx 0{,}270081$

3) $R = 1 - \frac{1{,}90455438670184}{2{,}590970979438} \approx 0{,}264926$

Труднощі:

- Вивід матриці біграм

ВИСНОВКИ

Під час виконання комп'ютерного практикуму ми ознайомились з поняттям ентропії на символ джерела та його надлишковості, написали програми для підрахунку частот букв і частот біграм в тексті, попрацювали з CoolPinkProgram, де оцінювали та порівнювали H^{10}, H^{20}, H^{30}. В результаті експериментів ми отримали значення ентропії та надлишковості власного джерела, які наближаються до значень, отриманих з прикладеної програми.