**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

**ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ**

**Кафедра інформаційної безпеки**

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

з дисципліни

Криптографія

З теми: «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав студент  
групи ФБ-91  
Братунець Дмитро

**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.

Код програми знаходиться у файлі ***main.py***

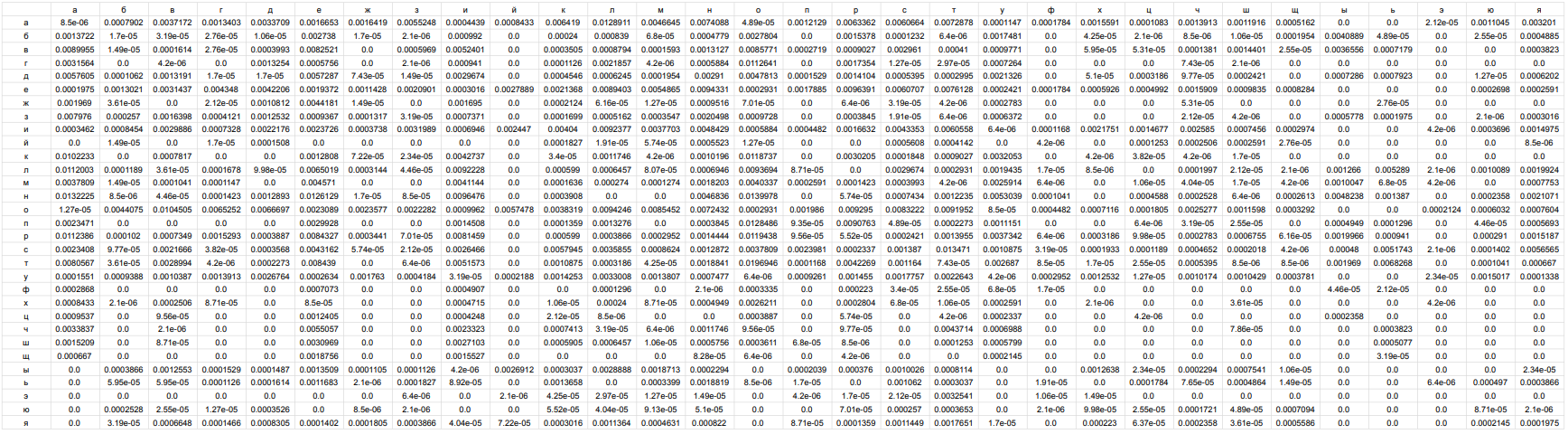
|  |  |
| --- | --- |
| H1 з пробілами | 4,27860 |
| H2 з пробілами | 3,97619 |
| H1 без пробілів | 4,88888 |
| H2 без пробілів | 3,49650 |

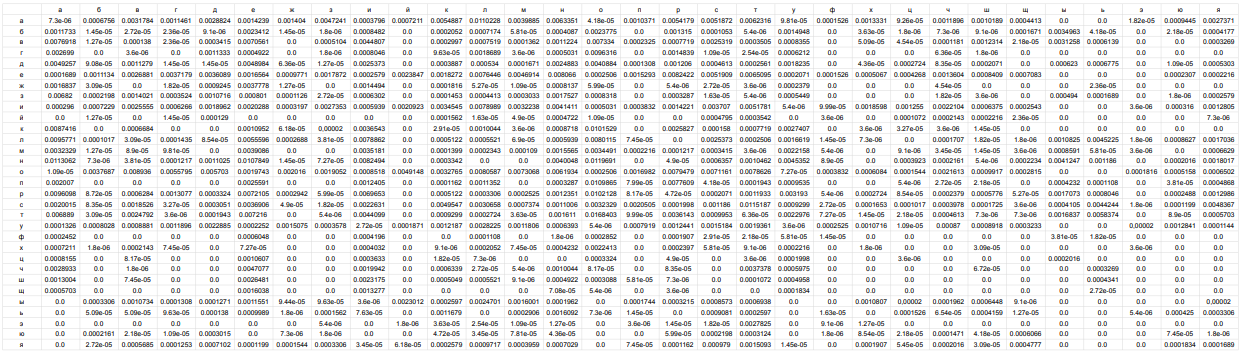
Частота символів з пробілом

|  |  |
| --- | --- |
| **Символ** | **Частота** |
| Пробіл | 0.18014488865038905 |
| О | 0.10445082782240882 |
| А | 0.08154169099891097 |
| Е | 0.07674205716630629 |
| И | 0.06460802727315772 |
| Н | 0.06052968007702142 |
| Т | 0.05692009027920264 |
| Л | 0.04960305560378162 |
| С | 0.04784797739863319 |
| Р | 0.044805000078915386 |
| В | 0.04445619406871952 |
| К | 0.03453021670165249 |
| У | 0.02837165990624852 |
| М | 0.028332202212787448 |
| П | 0.026839123092220522 |
| Д | 0.026534509698701053 |
| Г | 0.018082671759339634 |
| Я | 0.01800059975694061 |
| З | 0.016780567875124292 |
| Ь | 0.016771098028693636 |
| Ы | 0.016327593554191198 |
| Ч | 0.015009706592591423 |
| Б | 0.014619864581196042 |
| Й | 0.011065515554222763 |
| Ж | 0.008505500402468473 |
| Ш | 0.00824192301014852 |
| Х | 0.007768430688615666 |
| Ю | 0.0049842958380024935 |
| Щ | 0.0033396991745450527 |
| Ц | 0.0031313625530705975 |
| Э | 0.0029009296232579427 |
| Ф | 0.0020375952903297083 |

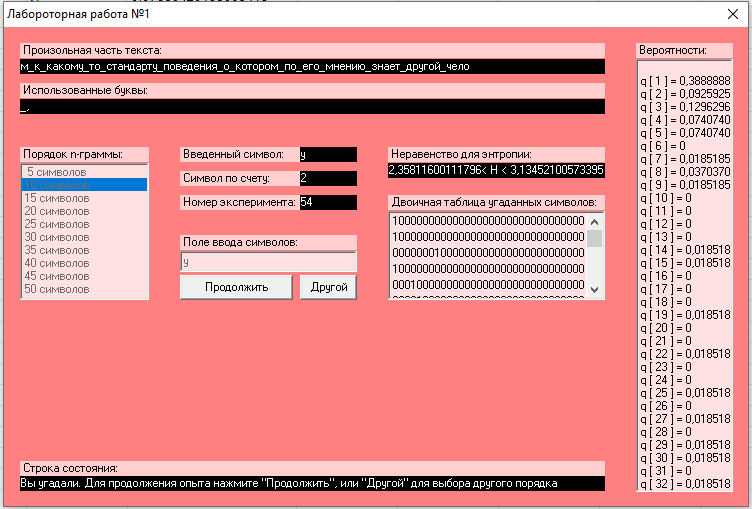
Частота символів без пробіла

|  |  |
| --- | --- |
| **Символ** | **Частота** |
| О | 0.1274015693461571 |
| А | 0.09945866028044939 |
| Е | 0.09360441388232214 |
| И | 0.07880420135065415 |
| Н | 0.07382972825208103 |
| Т | 0.06942701154293371 |
| Л | 0.060502221571964304 |
| С | 0.05836150404657216 |
| Р | 0.05464990027952535 |
| В | 0.054224451922410544 |
| К | 0.042117462248677455 |
| У | 0.034605699852921924 |
| М | 0.03455757221071437 |
| П | 0.032736422229580404 |
| Д | 0.03236487683173806 |
| Г | 0.022055935870879312 |
| Я | 0.021955830375087592 |
| З | 0.020467723678029924 |
| Ь | 0.02045617304390011 |
| Ы | 0.01991521834548717 |
| Ч | 0.018307755095754757 |
| Б | 0.01783225399074409 |
| Й | 0.01349691598068734 |
| Ж | 0.01037439455426103 |
| Ш | 0.010052901904314546 |
| Х | 0.00947537019782386 |
| Ю | 0.006079483763658625 |
| Щ | 0.004073523636447641 |
| Ц | 0.003819409685591739 |
| Э | 0.0035383442550996048 |
| Ф | 0.0024853114435982536 |

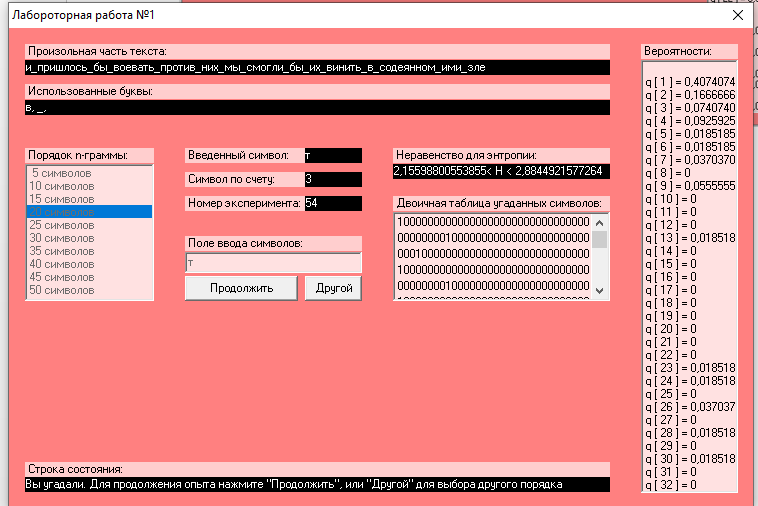
Частота біграм з пробілами (Файл - ЧастотаБіграмЗПробілами.pdf)

Частота біграм без пробілів (Файл - ЧастотаБіграмБезПробілів.pdf)

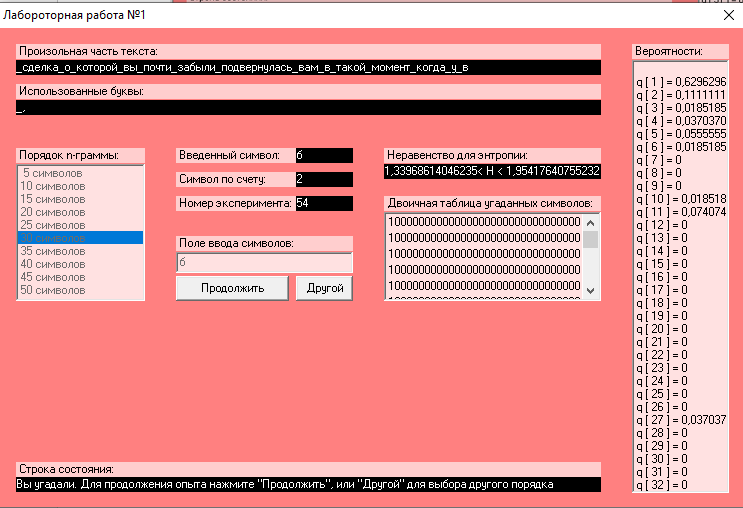
H10



H20



H30



Оцінка надлишковості мови

|  |  |
| --- | --- |
| **H** | **R** |
| H1 з пробілами | 0,1380192581 |
| H2 з пробілами | 0,1282640645 |
| H1 без пробілів | 0,1577059355 |
| H2 без пробілів | 0,1127902581 |
| H10 | 0,07606825806<R<0,1011135806 |
| H20 | 0,069548<R<0,09304812903 |
| H30 | 0,04321567742<R<0,06303793548 |

Висновок:  
Під час виконання цієї лаюораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію та надлишковість мови.