КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

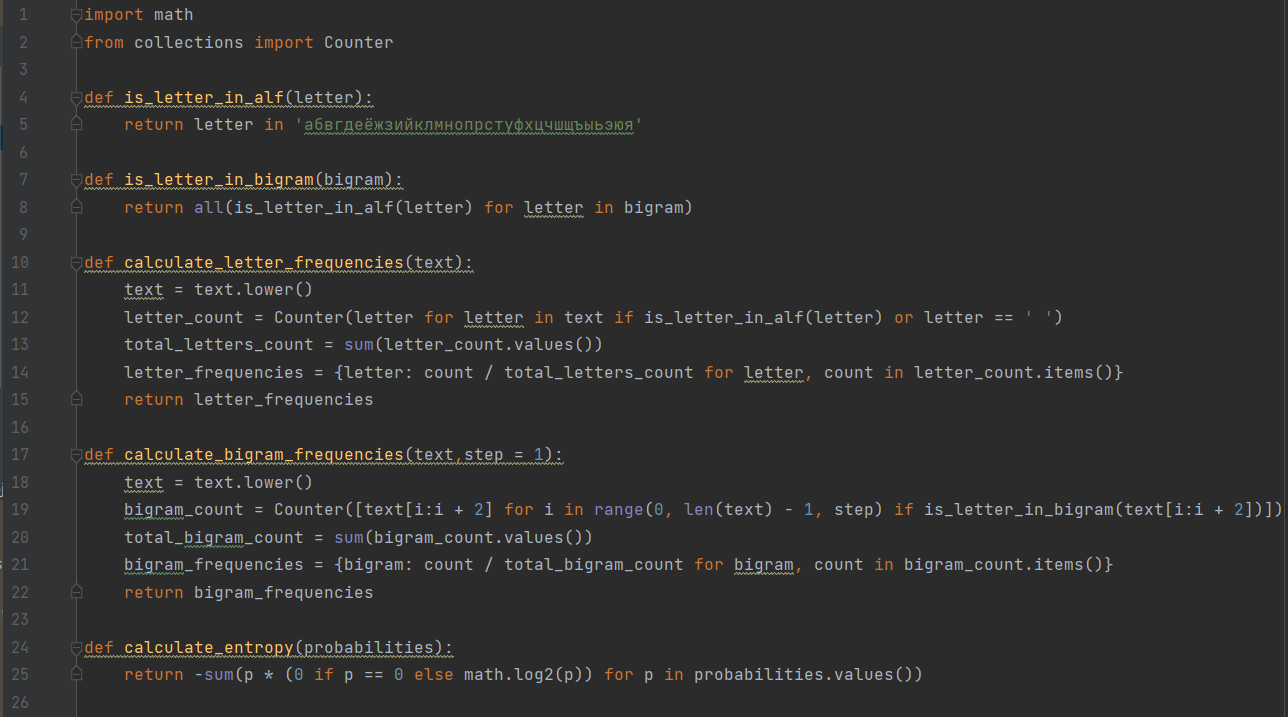
***ФБ-13 Владислав Садохін та Данило Розумовський***

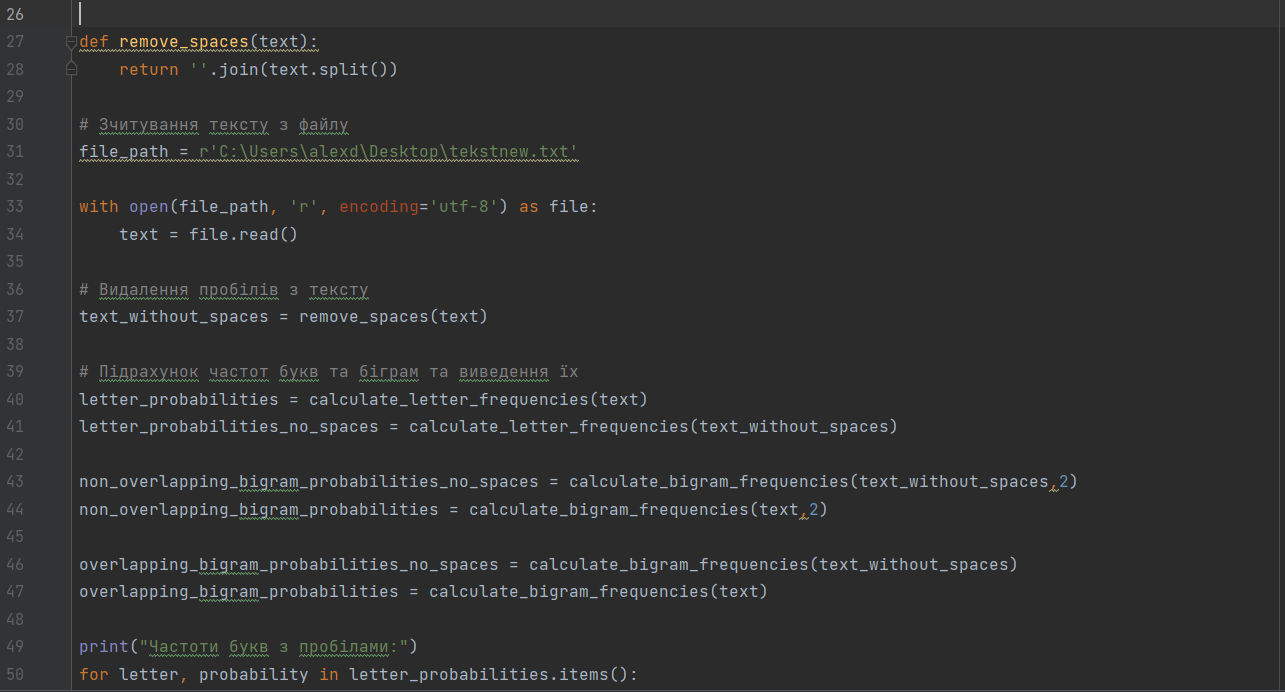
Мета роботи

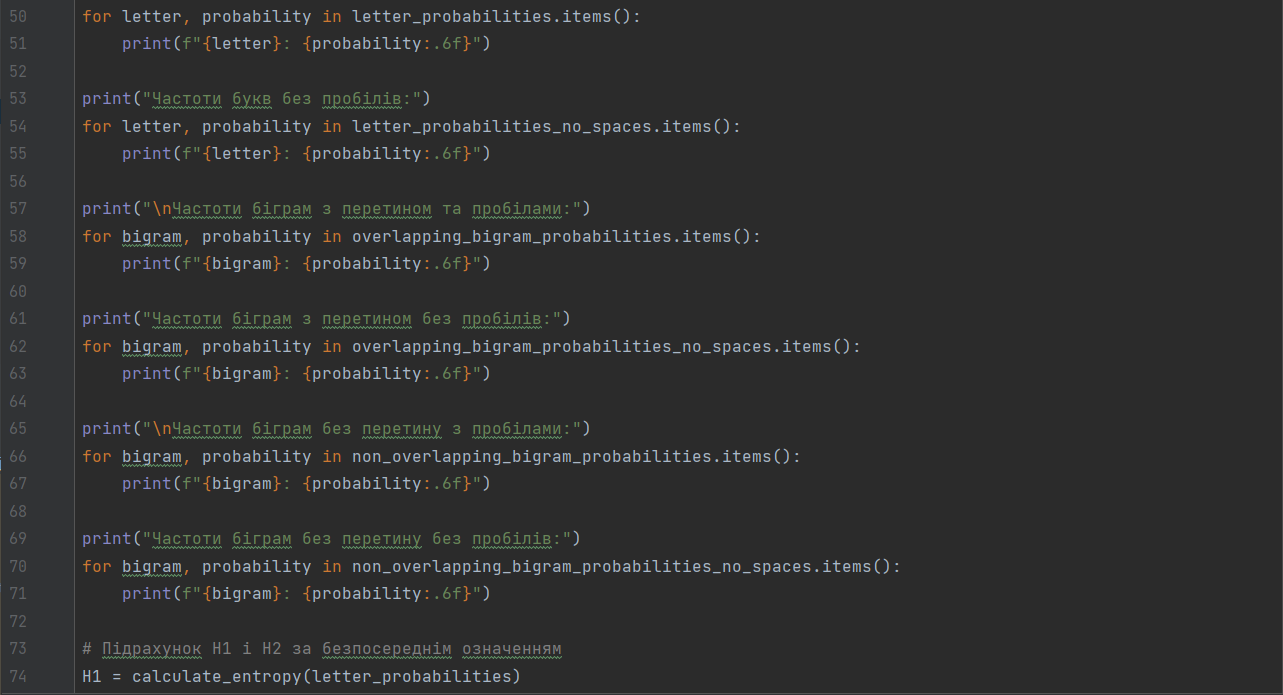
Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.  
  
Порядок виконання роботи

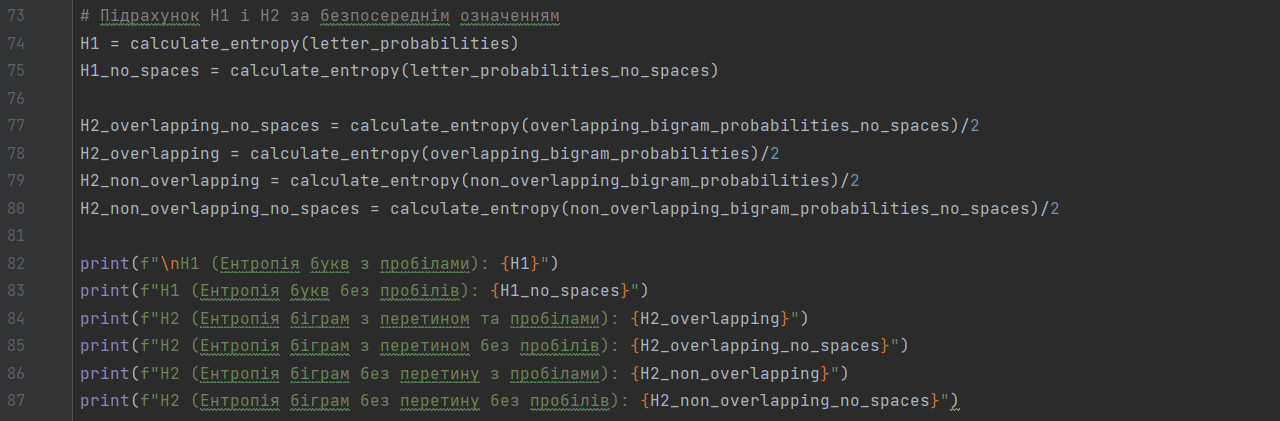
0. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.

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.









Текст був взятий з книги ,Робинзон Крузо, Даниэль Дефо перекладеної на рос.мову.

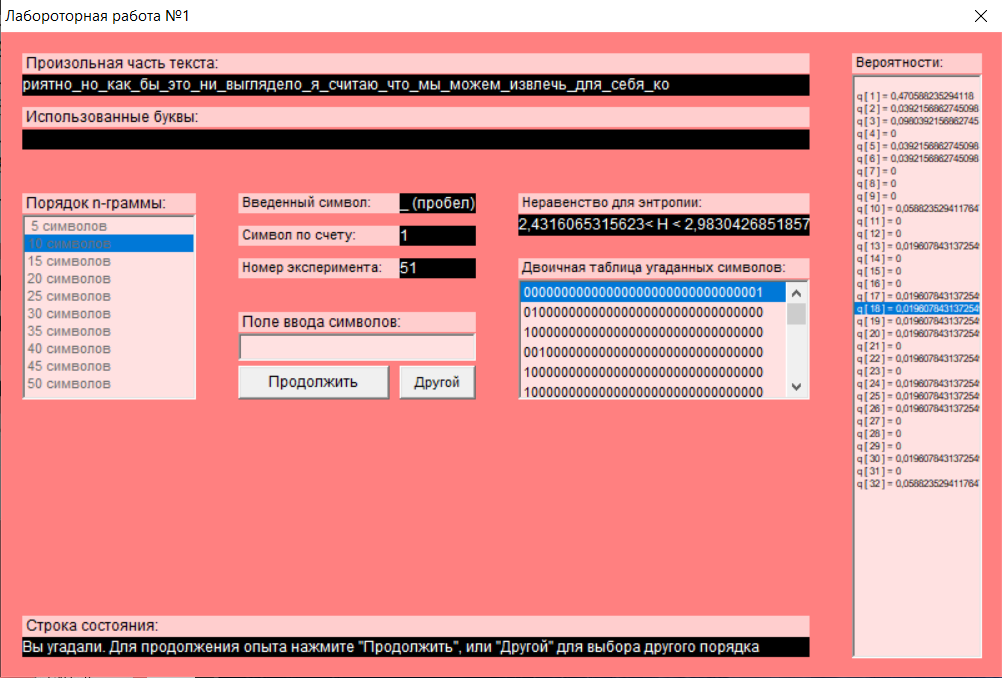
|  |
| --- |
| Частоти букв **з** пробілами: |
|  |
| е: 0.071370 |  |
| с: 0.043650 |  |
| л: 0.041818 |  |
| и: 0.056086 |  |
| : 0.169165 |  |
| у: 0.022645 |  |
| щ: 0.002502 |  |
| т: 0.049919 |  |
| в: 0.034276 |  |
| н: 0.051917 |  |
| а: 0.059080 |  |
| о: 0.097866 |  |
| р: 0.034909 |  |
| я: 0.023572 |  |
| п: 0.023700 |  |
| к: 0.027598 |  |
| ю: 0.004436 |  |
| ч: 0.012491 |  |
| й: 0.008666 |  |
| г: 0.015183 |  |
| ц: 0.002264 |  |
| з: 0.013273 |  |
| ж: 0.008418 |  |
| ь: 0.017921 |  |
| б: 0.017453 |  |
| м: 0.029682 |  |
| д: 0.027138 |  |
| ы: 0.015604 |  |
| э: 0.002321 |  |
| х: 0.007937 |  |
| ш: 0.006555 |  |
| ф: 0.000345 |  |
| ъ: 0.000238 |  |

|  |
| --- |
| Частоти букв **без** пробілів: |
|  |
| е: 0.085902 |  |
| с: 0.052538 |  |
| л: 0.050332 |  |
| и: 0.067506 |  |
| у: 0.027256 |  |
| щ: 0.003012 |  |
| т: 0.060083 |  |
| в: 0.041255 |  |
| н: 0.062488 |  |
| а: 0.071109 |  |
| о: 0.117792 |  |
| р: 0.042017 |  |
| я: 0.028371 |  |
| п: 0.028525 |  |
| к: 0.033217 |  |
| ю: 0.005339 |  |
| ч: 0.015035 |  |
| й: 0.010431 |  |
| г: 0.018274 |  |
| ц: 0.002725 |  |
| з: 0.015975 |  |
| ж: 0.010132 |  |
| ь: 0.021570 |  |
| б: 0.021007 |  |
| м: 0.035725 |  |
| д: 0.032664 |  |
| ы: 0.018782 |  |
| э: 0.002794 |  |
| х: 0.009553 |  |
| ш: 0.007889 |  |
| ф: 0.000415 |  |
| ъ: 0.000286 |  |

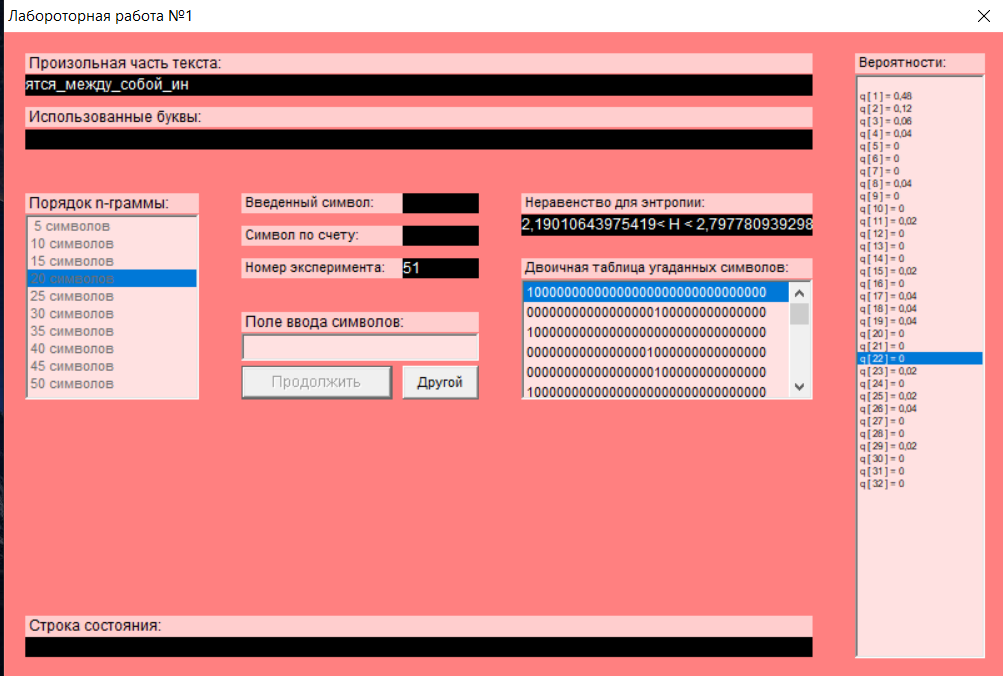
|  |  |
| --- | --- |
| Частоти біграм **з** перетином **та** пробілами: | Частоти біграм **з** перетином **без** пробілів: |
|  |
|  |
| ес: 0.007359 | ес: 0.007794 |  |
| сл: 0.004622 | сл: 0.003726 |  |
| ли: 0.010647 | ли: 0.009107 |  |
| су: 0.001380 | ис: 0.004986 |  |
| ущ: 0.000371 | су: 0.001147 |  |
| ще: 0.001973 | ущ: 0.000295 |  |
| ст: 0.017542 | ще: 0.001571 |  |
| тв: 0.003167 | ст: 0.014250 |  |
| ву: 0.000894 | тв: 0.002897 |  |
| уе: 0.000215 | ву: 0.000799 |  |
| ет: 0.004498 | уе: 0.000239 |  |
| на: 0.013901 | ет: 0.004717 |  |
| св: 0.002450 | тн: 0.002047 |  |
| ве: 0.007581 | на: 0.011094 |  |
| те: 0.006076 | ас: 0.005801 |  |

|  |  |
| --- | --- |
| Частоти біграм **без** перетину **з** пробілами: | Частоти біграм **без** перетину **без** пробілів: |
|  |
|  |
| ес: 0.007456 | ес: 0.007957 |  |
| ли: 0.010523 | ли: 0.009091 |  |
| ущ: 0.000399 | су: 0.001170 |  |
| тв: 0.003218 | ще: 0.001598 |  |
| уе: 0.000204 | ст: 0.014150 |  |
| на: 0.013896 | ву: 0.000809 |  |
| ве: 0.007518 | ет: 0.004797 |  |
| те: 0.006067 | на: 0.011004 |  |
| ст: 0.017442 | св: 0.002050 |  |
| ор: 0.008814 | еи: 0.001534 |  |
| ия: 0.002144 | ор: 0.007402 |  |
| ри: 0.006325 | ия: 0.002860 |  |
| кл: 0.000595 | пр: 0.007451 |  |
| юч: 0.000178 | ик: 0.003450 |  |
| ен: 0.012516 | лю: 0.000898 |  |

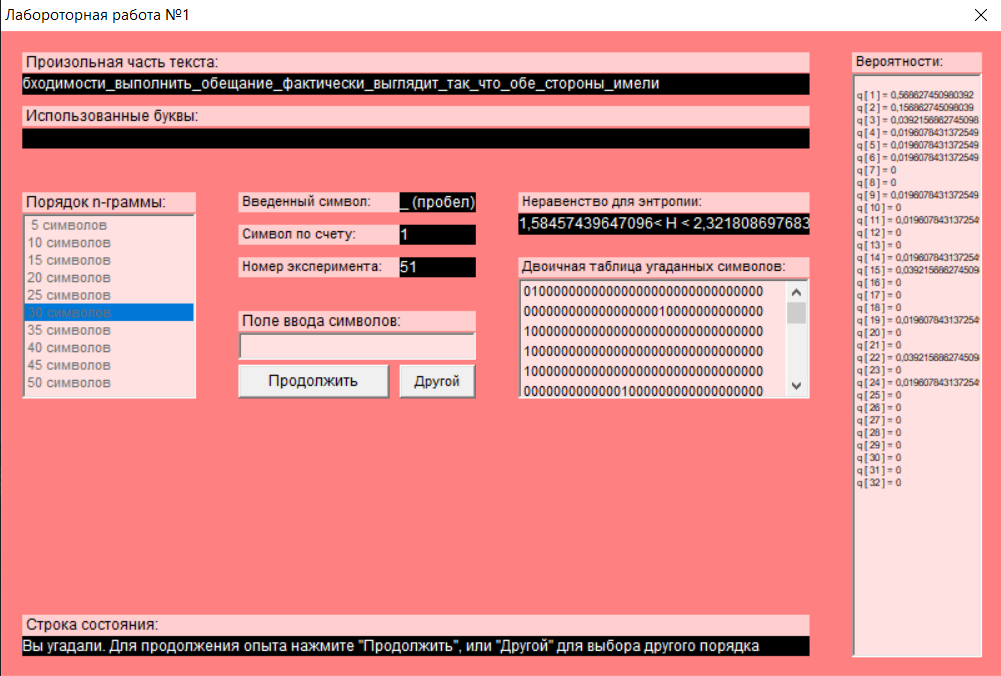
2. За допомогою програми CoolPinkProgram оцінити значення H(10), H(20) , H(30) .



Результати для H(10)



Результати для H(20)



Результати для H(30)

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

|  |
| --- |
| H1 (Ентропія букв з пробілами): 4.360373178797449 |
| H1 (Ентропія букв без пробілів): 4.458867100342456 |
| H2 (Ентропія біграм з перетином та пробілами): 3.9195543933732493 |
| H2 (Ентропія біграм з перетином без пробілів): 4.118281456606397 |
| H2 (Ентропія біграм без перетину з пробілами): 3.919944391193998 |
| H2 (Ентропія біграм без перетину без пробілів): 4.116656741575615 |



Надлишковість джерела відкритого тексту (мови) дорівнює

Надлишковість для нашого файлу

|  |  |  |
| --- | --- | --- |
|  | з пробілами | без пробілів |
| R1 | 0,128 | 0,108 |
| R2(крок1) | 0,2161 | 0,176 |
| R2(крок2) | 0,2160 | 0,177 |

Де крок1- з перетином, крок2 – без перетину

Надлишковість згідно з експериментами у CoolPinkProgram.exe

|  |  |  |
| --- | --- | --- |
| CoolPinkProgram | | |
|  | min | max |
| H(10) | 2,431 | 2,983 |
| H(20) | 2,190 | 2,797 |
| H(30) | 1,587 | 2,321 |

|  |  |  |
| --- | --- | --- |
| CoolPinkProgram | | |
|  | min | max |
| R(10) | 0,514 | 0,403 |
| R(20) | 0,562 | 0,441 |
| R(30) | 0,683 | 0,536 |

Висновки   
У ході виконання лабораторної роботи було обчислено частоти букв, біграм та ентропії ,для випадків тексту відкритого джерела з пробілами та без ,також з перетинами та без. Засвоєно ключові понятя ентропії на символ джерела та його надлишковості, пораховано надлишковість рос.мови у тексті ,який ми знайшли самі та за допомогою програми CoolPinkProgram.exe ,в якій містився інший текст.