Міністерство освіти і науки України

Національний Технічний Університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

Криптографія

Комп'ютерний практикум №1

Виконав:  
 Студент групи ФБ-81

Шмалій Григорій Григорович

Перевірив:

Чорний О.М.

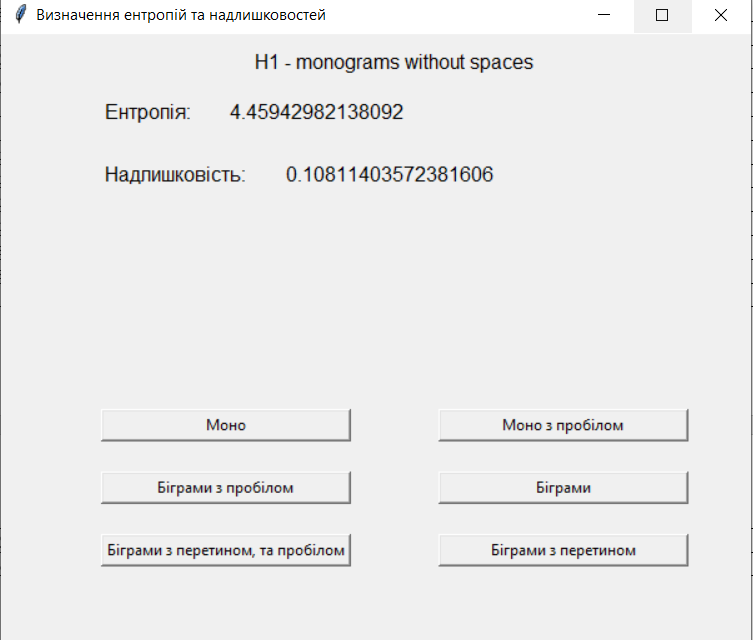
**Мета роботи:** Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

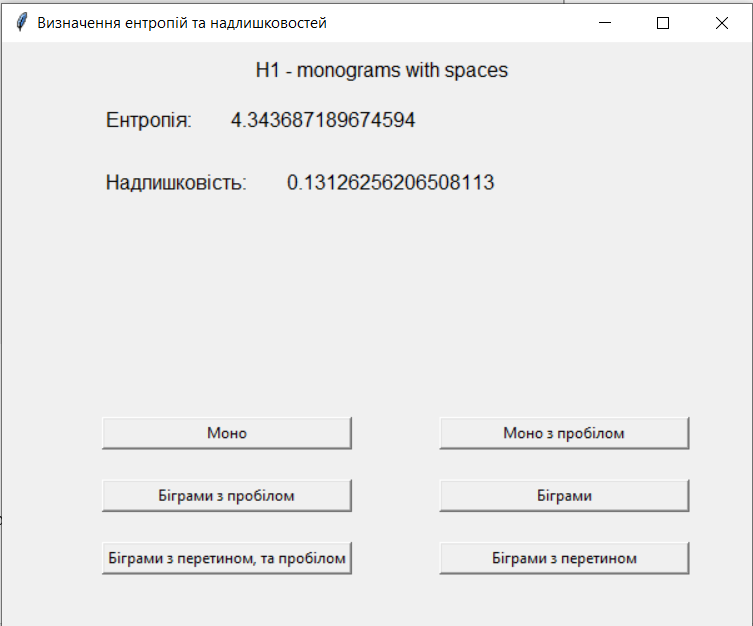
**Порядок виконання роботи:**

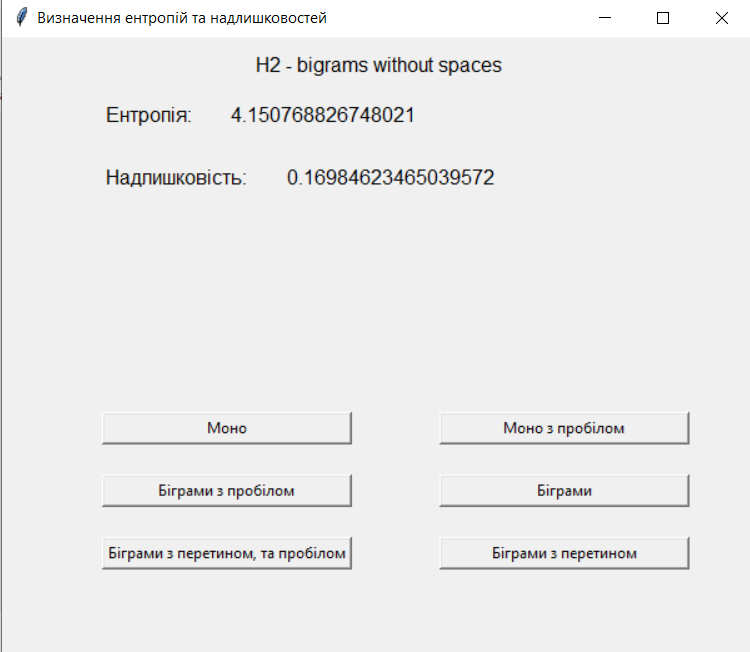
0. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму. 1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де ймовірності 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.

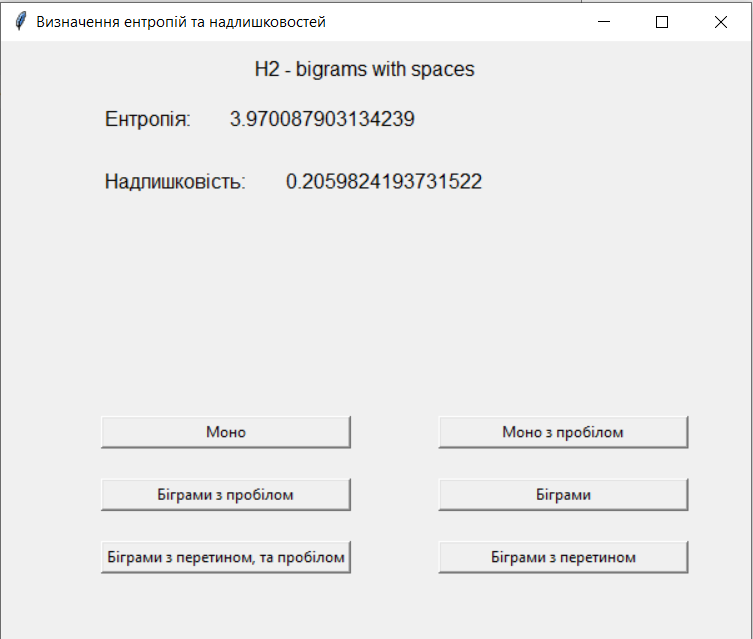
**Хід роботи:**

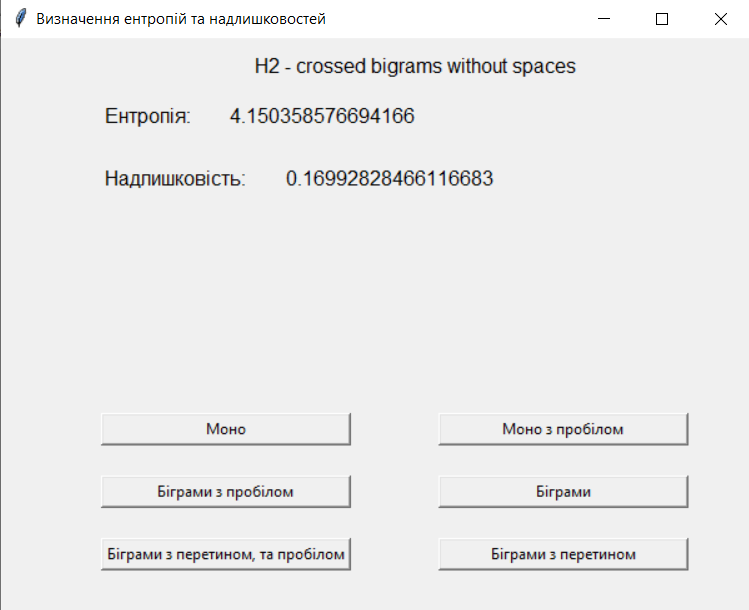
**Скріни з комп’ютерної програми:**

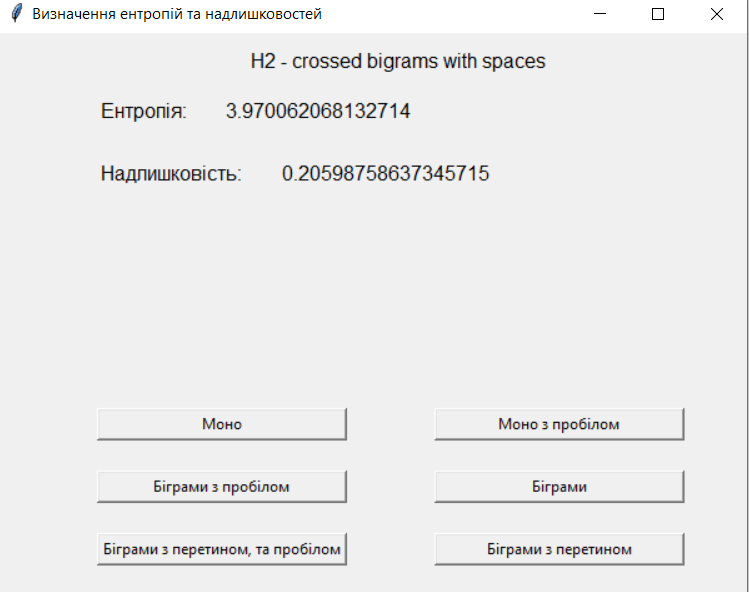




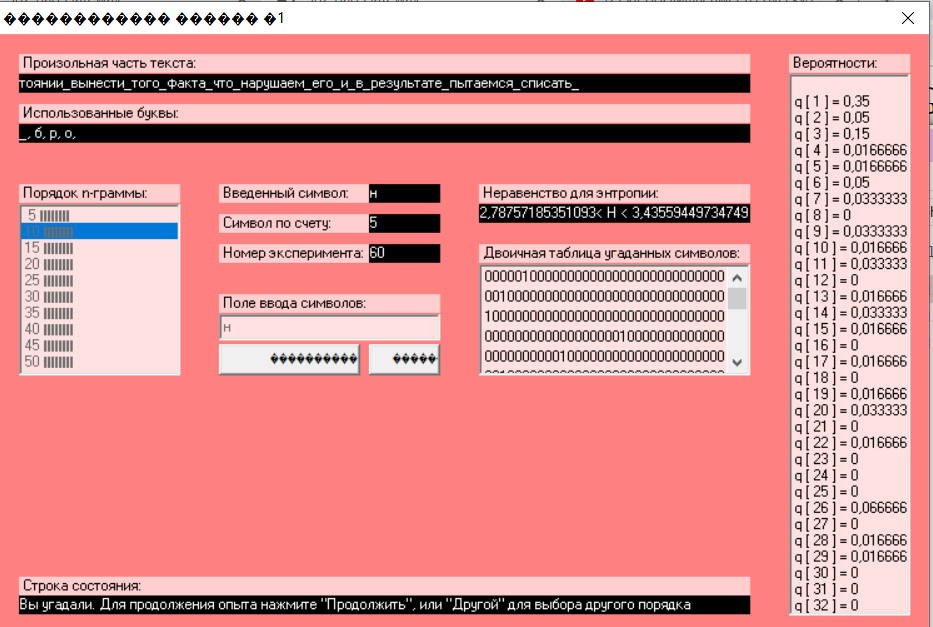


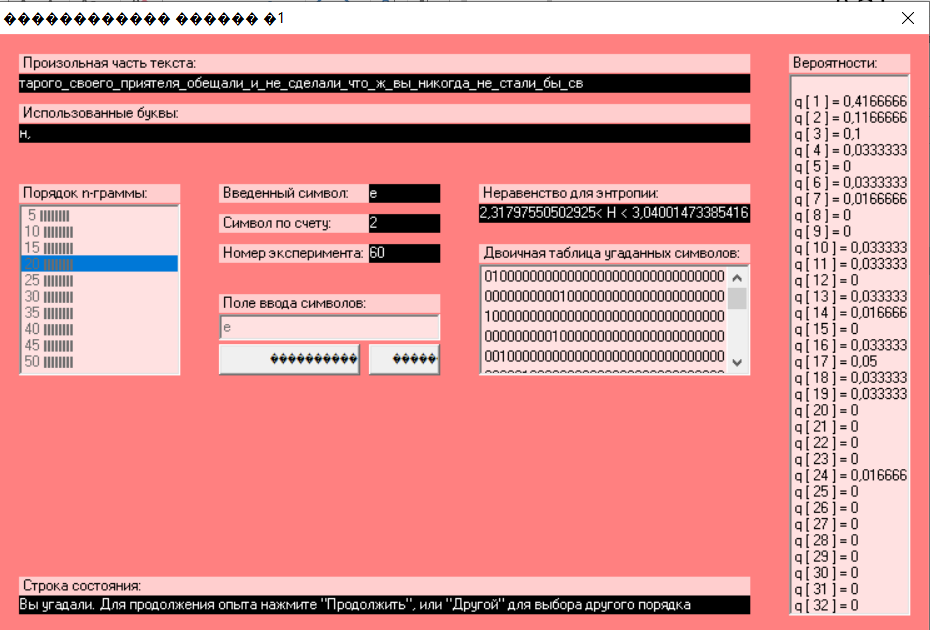


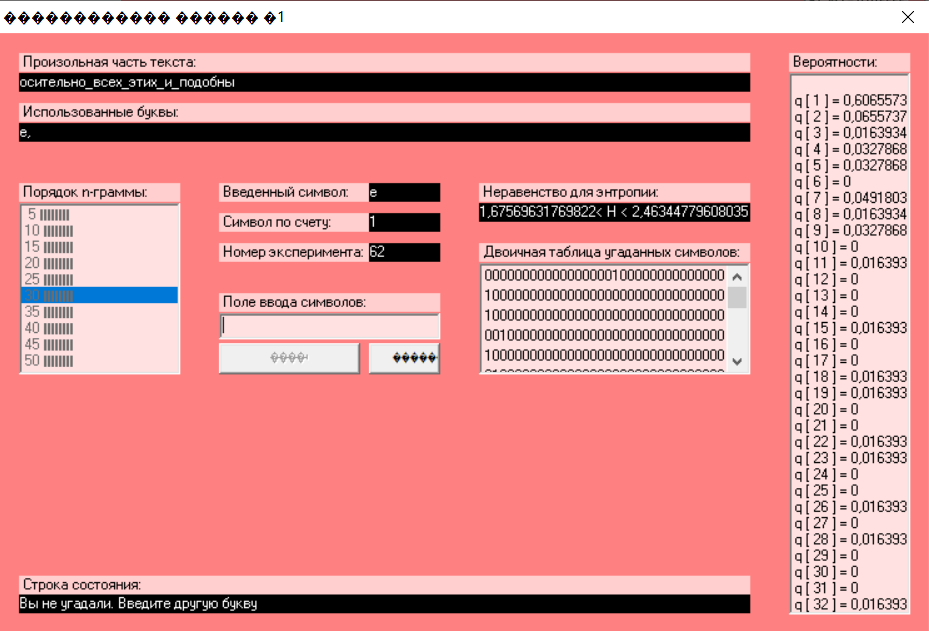




Робота з CoolPinkProgram







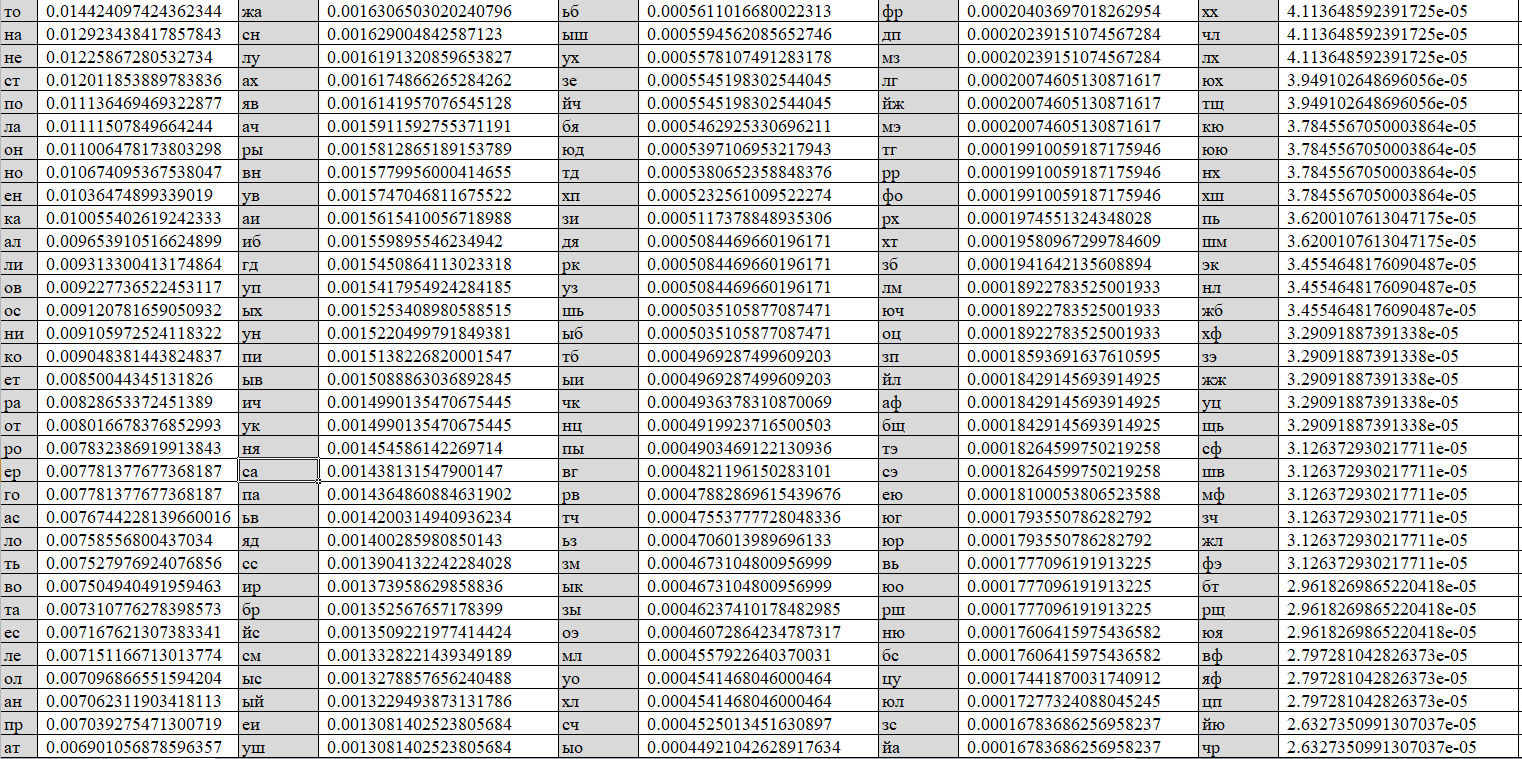
**Таблиці частот:**

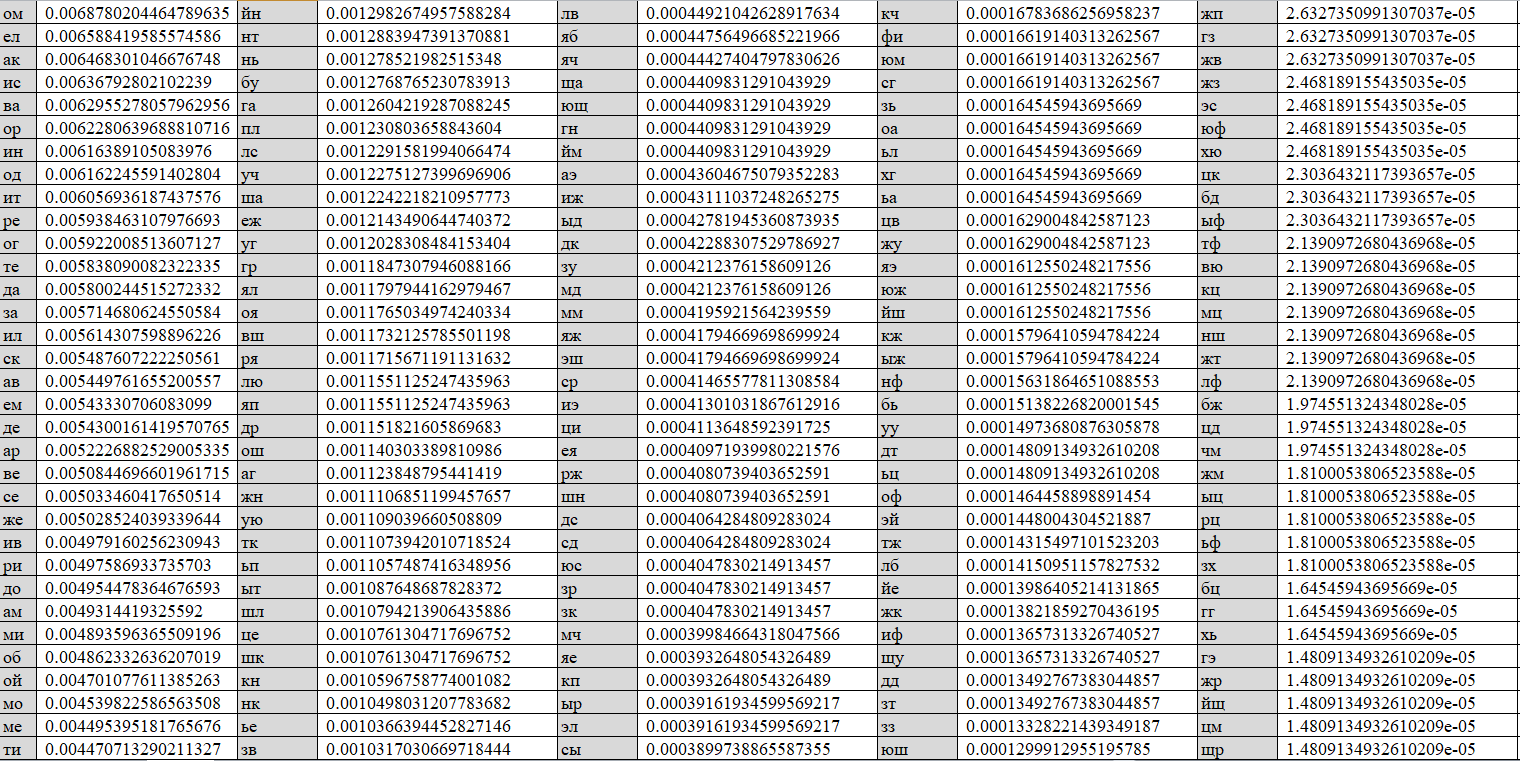
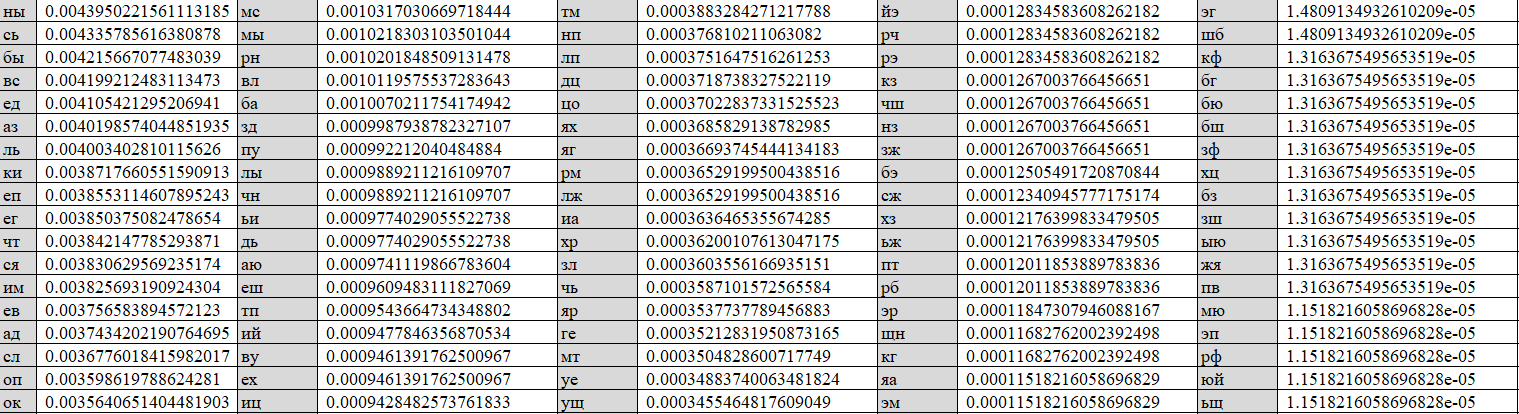
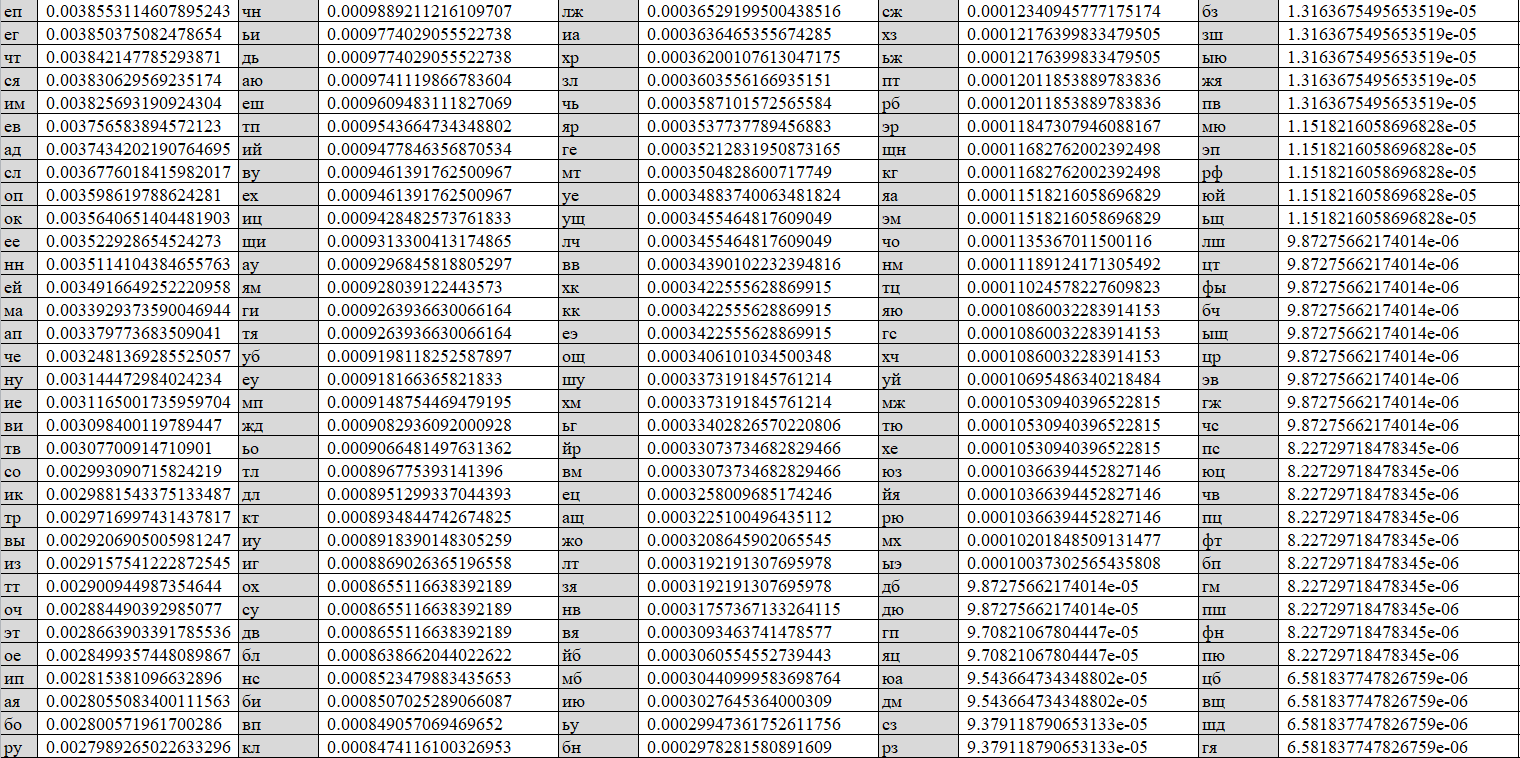
**1. Монограми без пробілу**

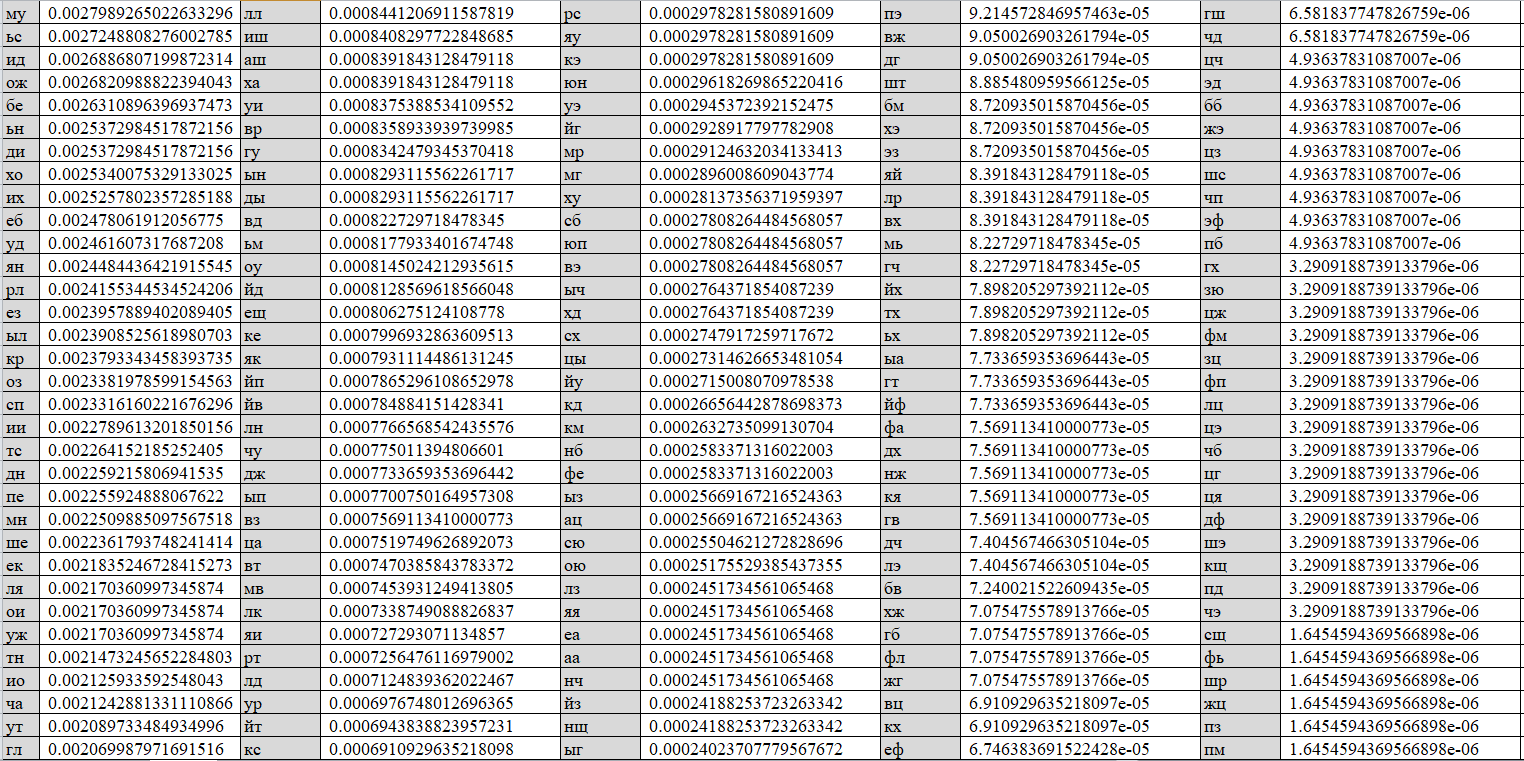
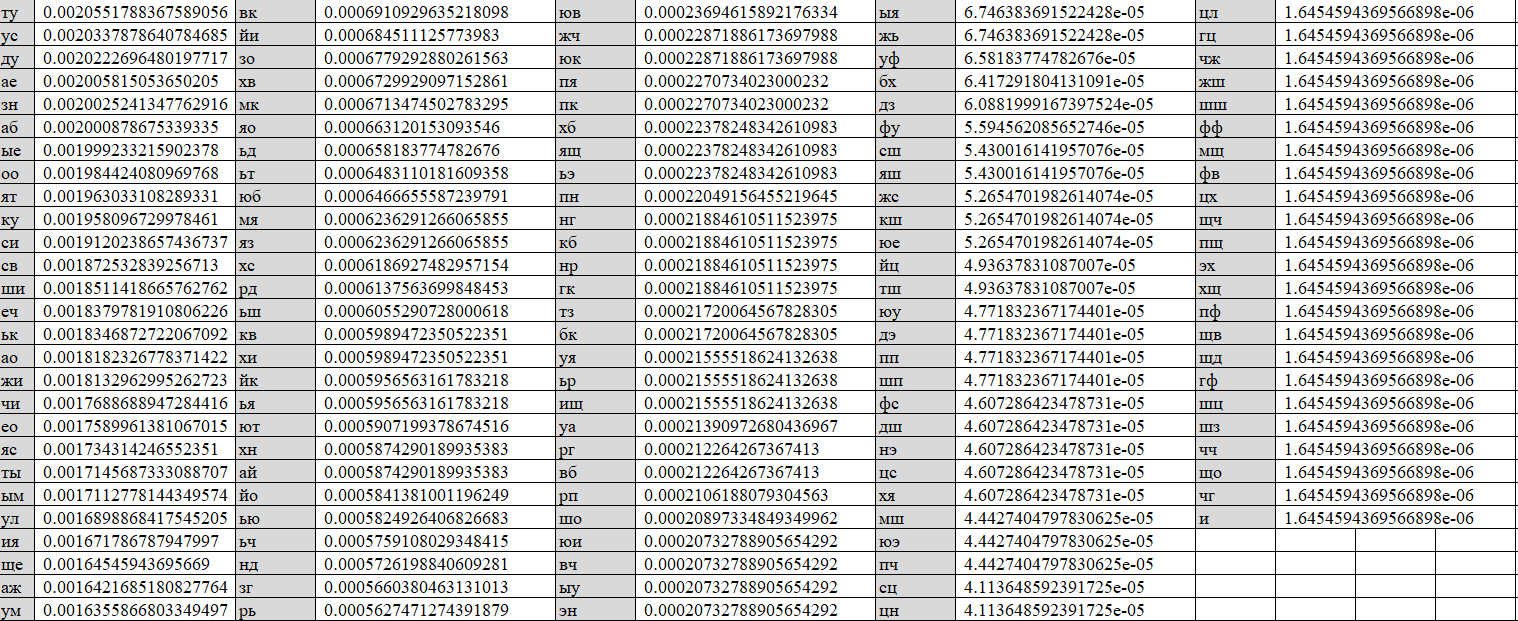
|  |  |
| --- | --- |
| а | 0.08521265523894148 |
| б | 0.017028873723225268 |
| в | 0.041282965778529204 |
| г | 0.01722386082692632 |
| д | 0.031332864377007975 |
| е | 0.08684248415215576 |
| ж | 0.011746121854598857 |
| з | 0.016166652268884748 |
| и | 0.0668575401183909 |
| й | 0.011380829559057644 |
| к | 0.03482041852295212 |
| л | 0.05233634864023234 |
| м | 0.031244832224704127 |
| н | 0.06592867750202597 |
| о | 0.10673445965124459 |
| п | 0.02702504802688683 |
| р | 0.04183337241302711 |
| с | 0.0534207073013209 |
| т | 0.06348105457582078 |
| у | 0.028649940557728935 |
| ф | 0.0011452407103454234 |
| х | 0.009707395934889116 |
| ц | 0.003354271821895324 |
| ч | 0.014029198701731437 |
| ш | 0.009010543290016578 |
| щ | 0.003284339738289461 |
| ы | 0.019255182173077794 |
| ь | 0.020058989769347533 |
| э | 0.004448503247728236 |
| ю | 0.005661207850493432 |
| я | 0.019495419448523815 |

**2. Монограми з пробілом:**

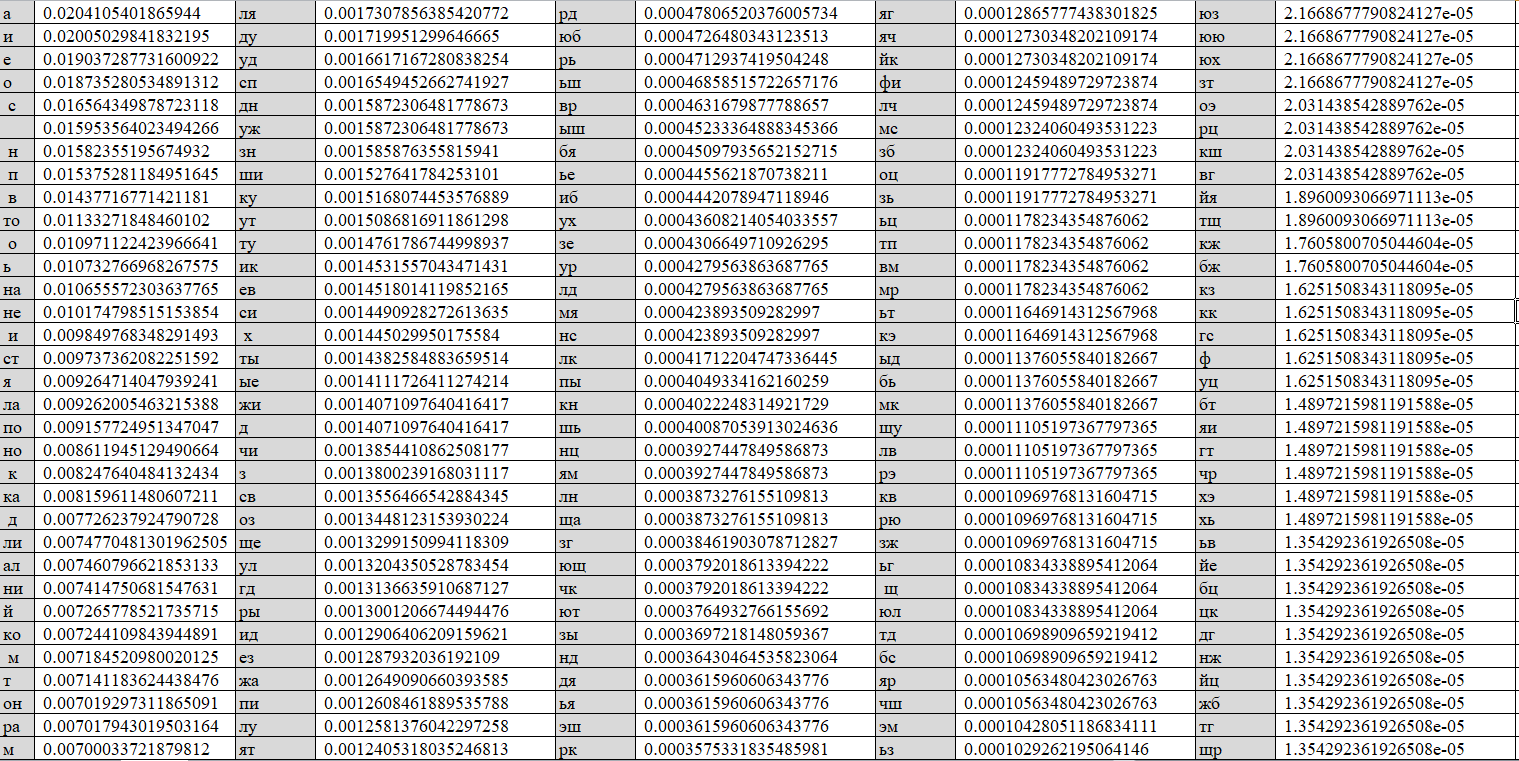
|  |  |
| --- | --- |
| а | 0.07013406140090711 |
| б | 0.014015571653577431 |
| в | 0.033977841068374155 |
| г | 0.014176055298465723 |
| д | 0.025788435155804566 |
| е | 0.07147548798539531 |
| ж | 0.009667616025612377 |
| з | 0.01330592245592794 |
| и | 0.05502693010361691 |
| й | 0.009366963121264692 |
| к | 0.028658857816907798 |
| л | 0.04307530000961548 |
| м | 0.025715980514441496 |
| н | 0.054262432065309396 |
| о | 0.08784752834872486 |
| п | 0.022242897752280968 |
| р | 0.034430851863438576 |
| с | 0.043967778676125044 |
| т | 0.05224792217694371 |
| у | 0.023580261459683394 |
| ф | 0.0009425874839008496 |
| х | 0.007989647789185433 |
| ц | 0.0027607249797871867 |
| ч | 0.011546696677785406 |
| ш | 0.007416104973909558 |
| щ | 0.00270316755440531 |
| ы | 0.015847929219263996 |
| ь | 0.016509501038065096 |
| э | 0.0036613294004683144 |
| ю | 0.0046594428712081505 |
| я | 0.016045655904105265 |
|  | 0.1769525171554985 |

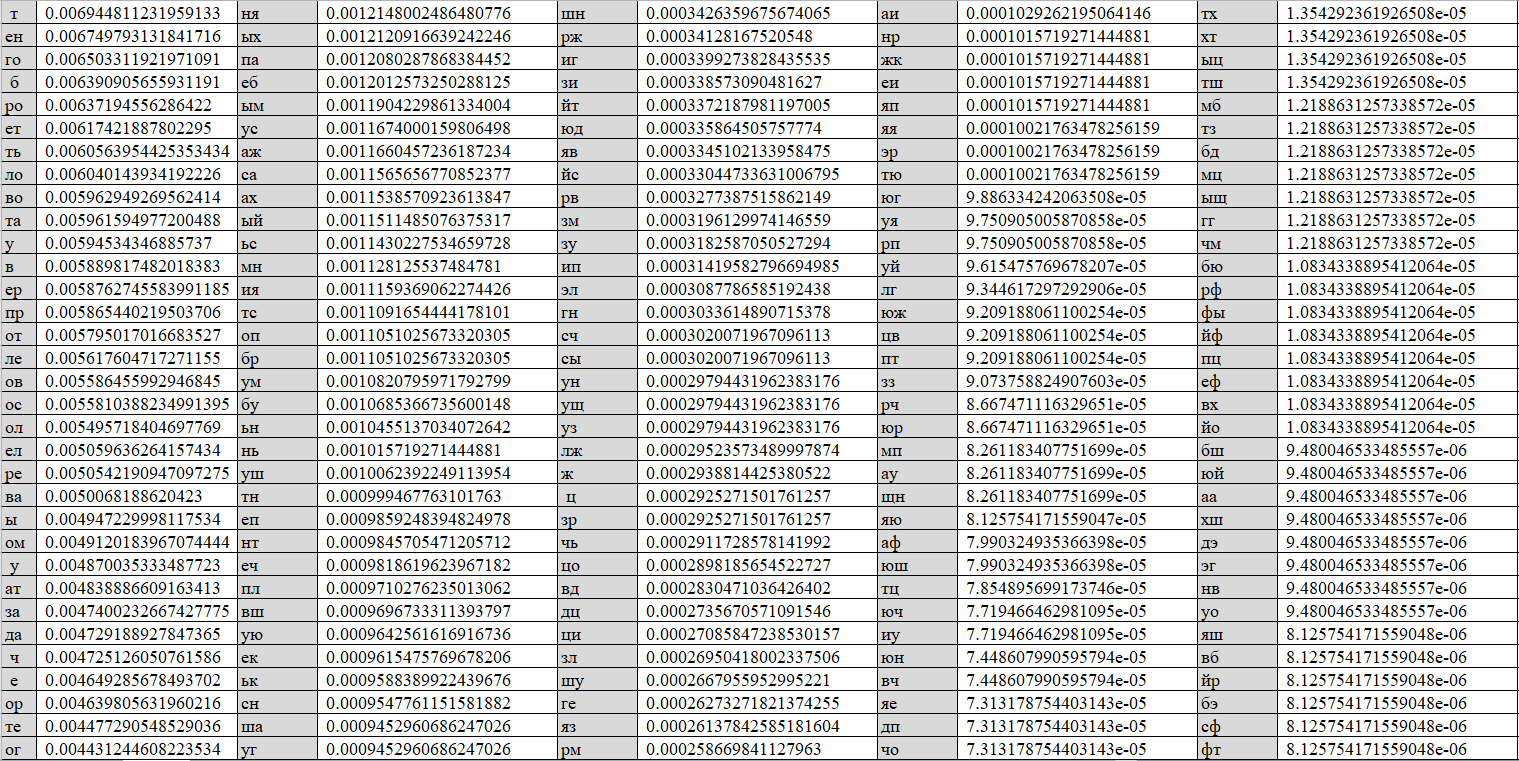
**3. Біграми без пробілів:**

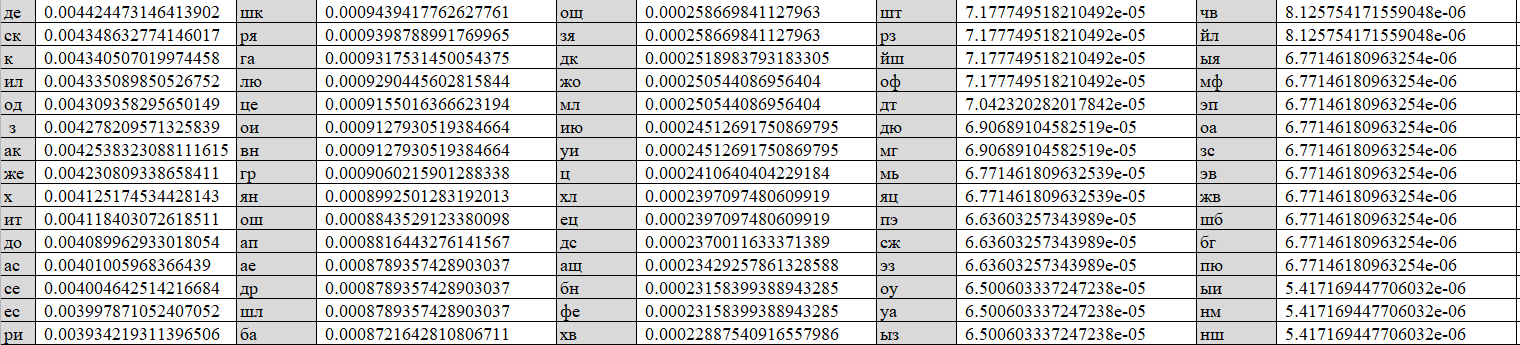


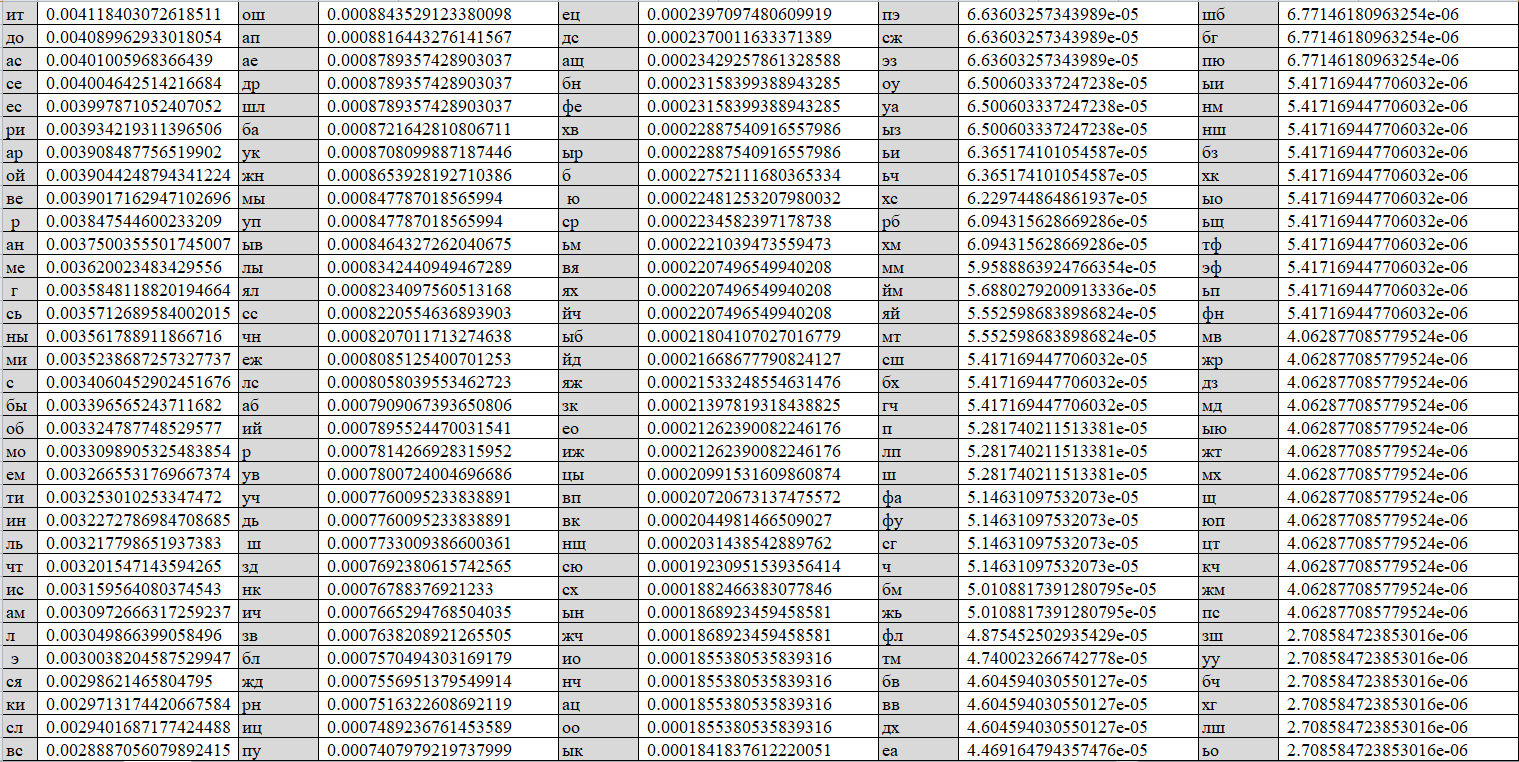


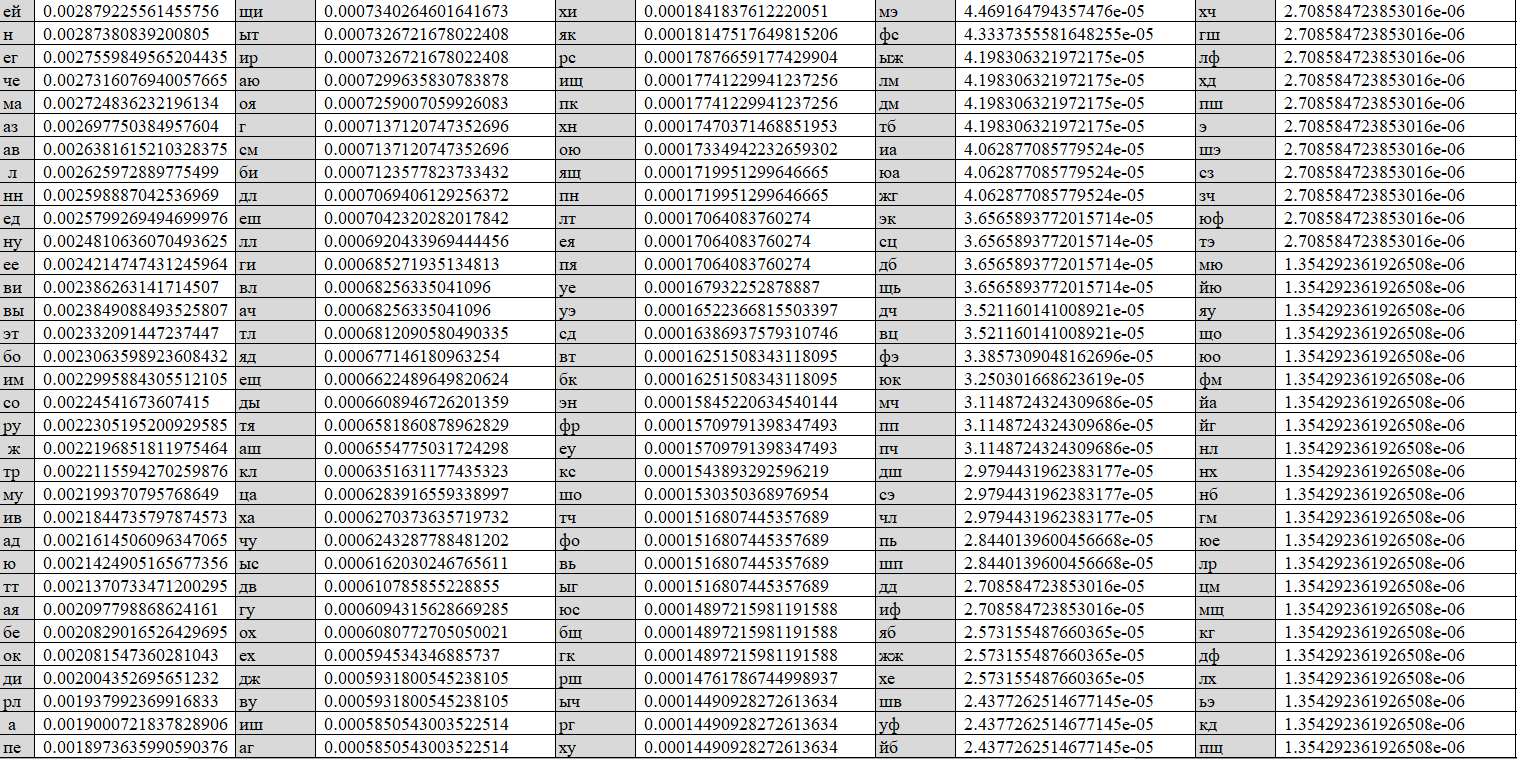
**4. Біграми з пробілом**

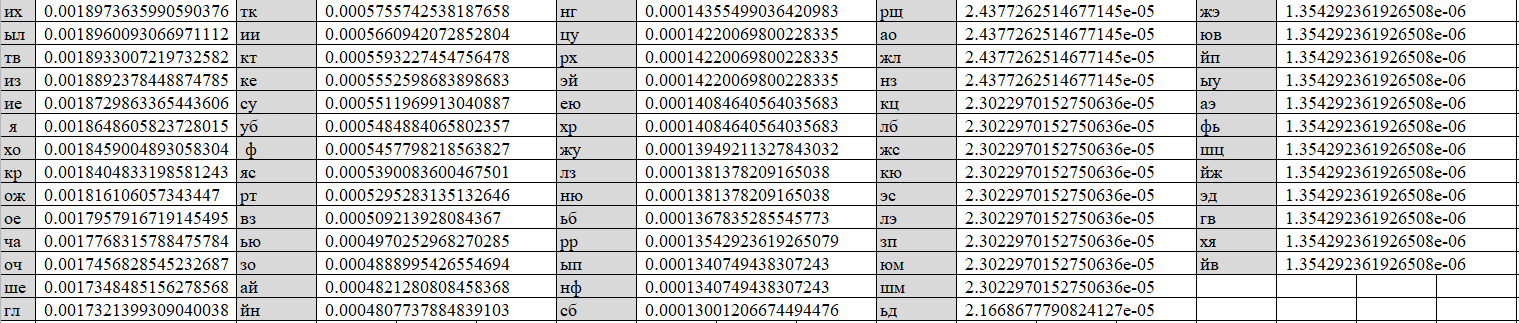




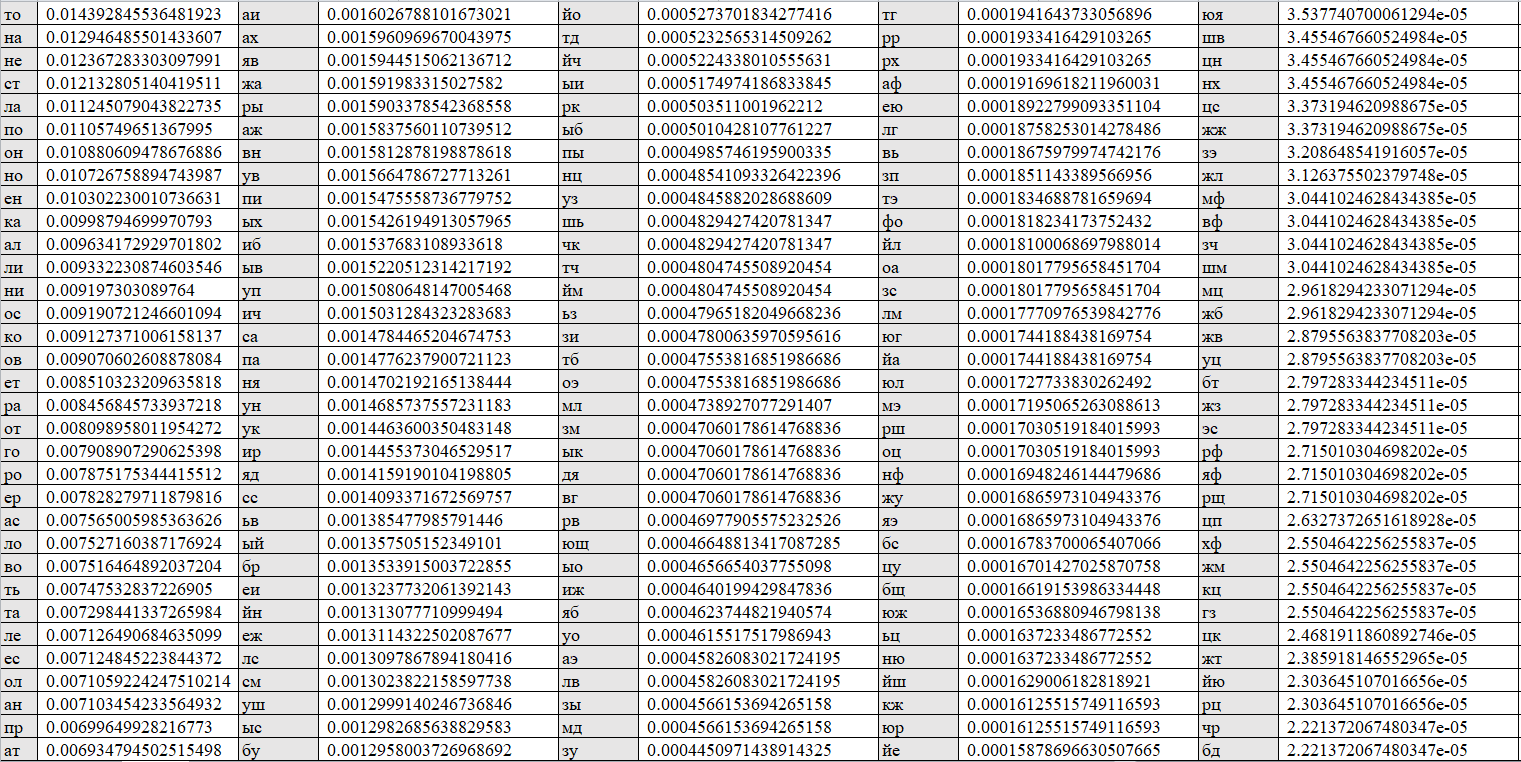


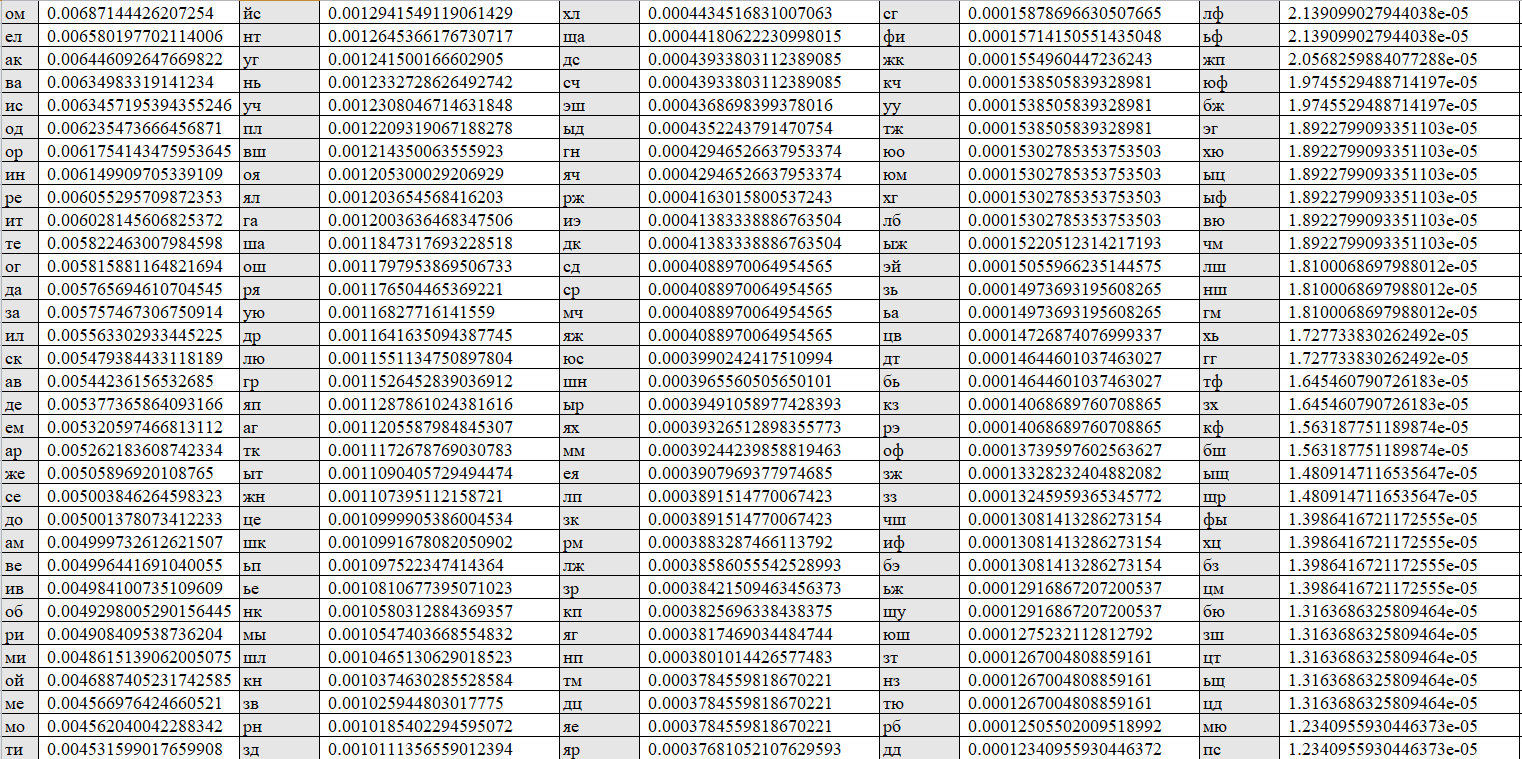


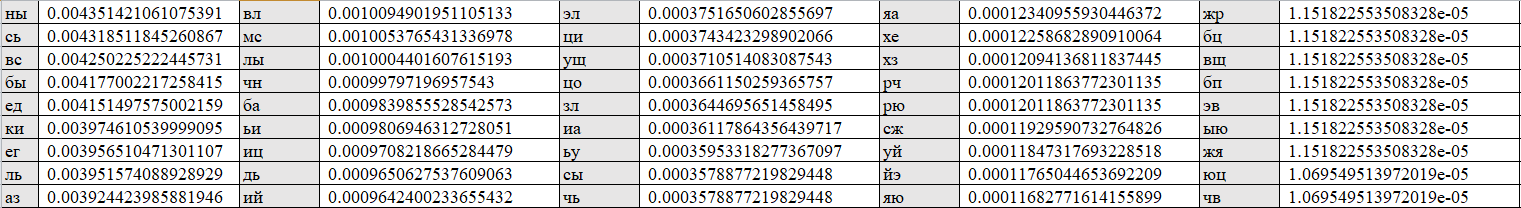


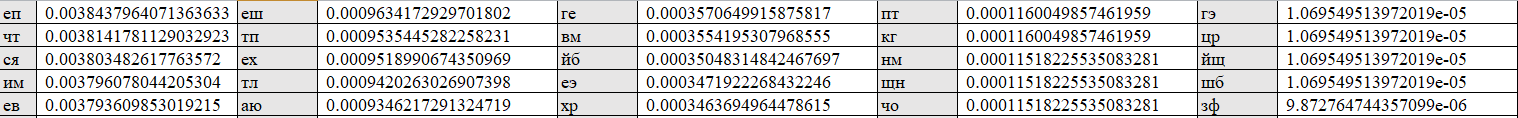


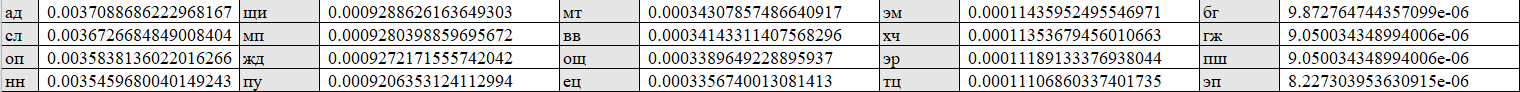
**5. Біграми з перетином без пробілів**

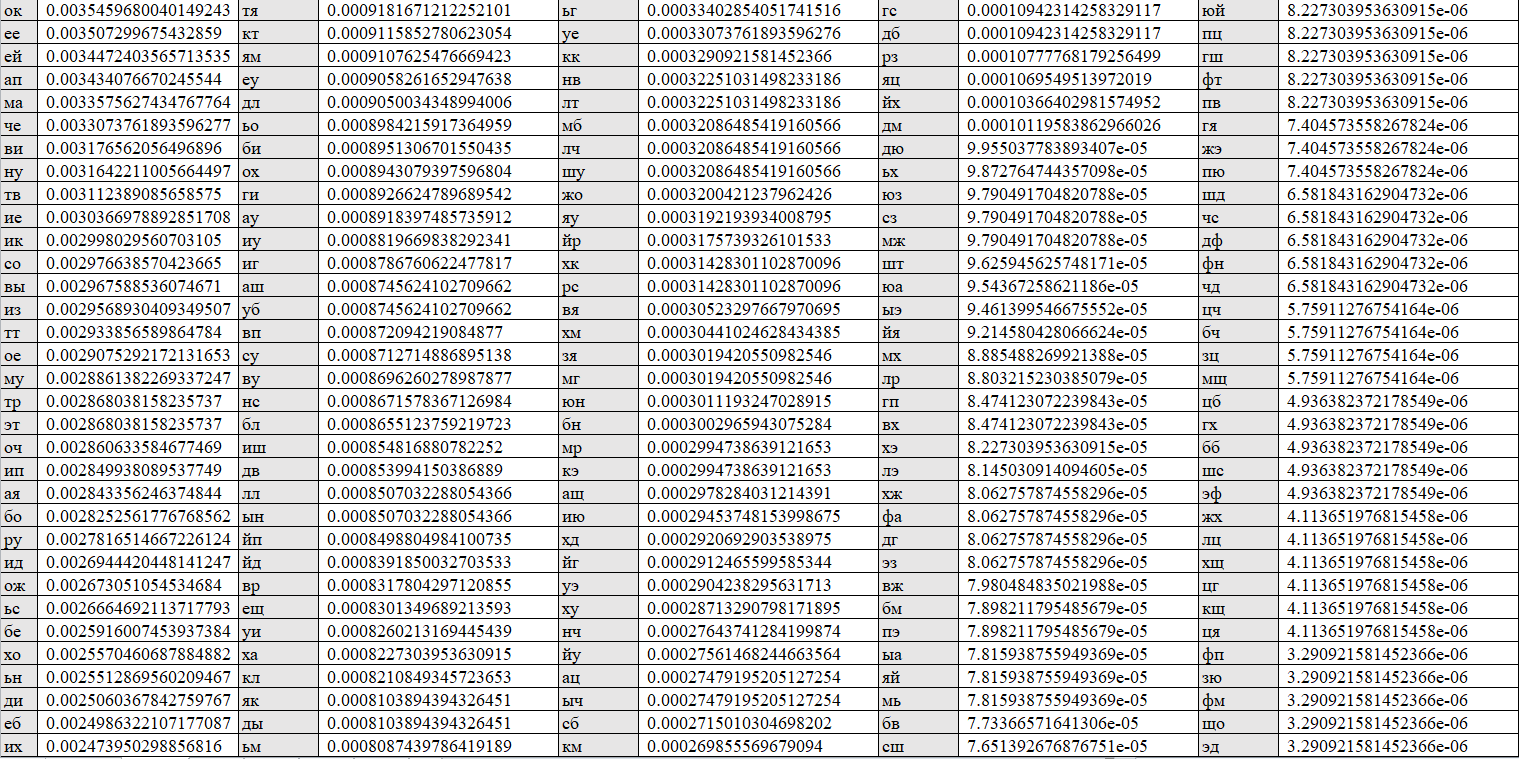


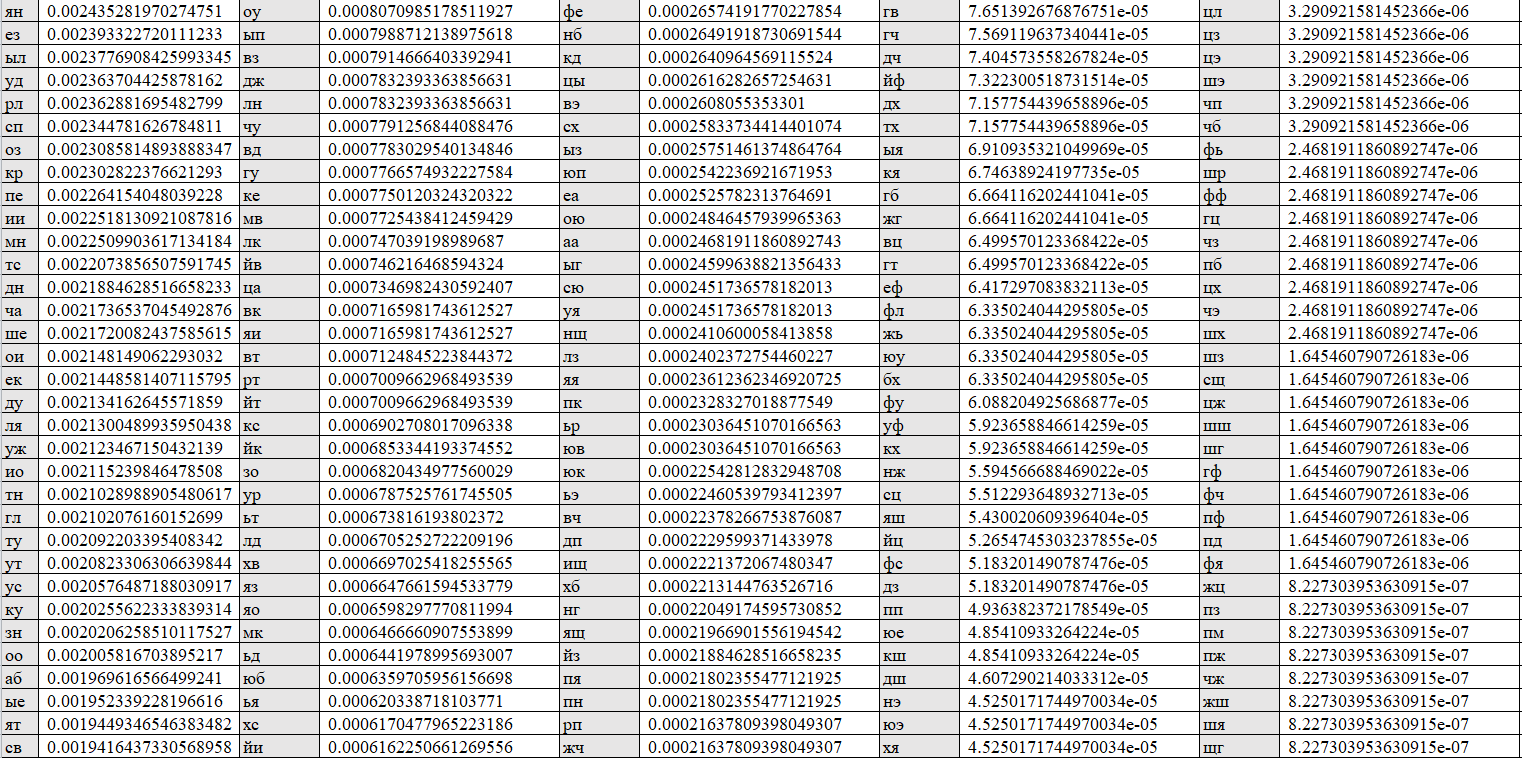


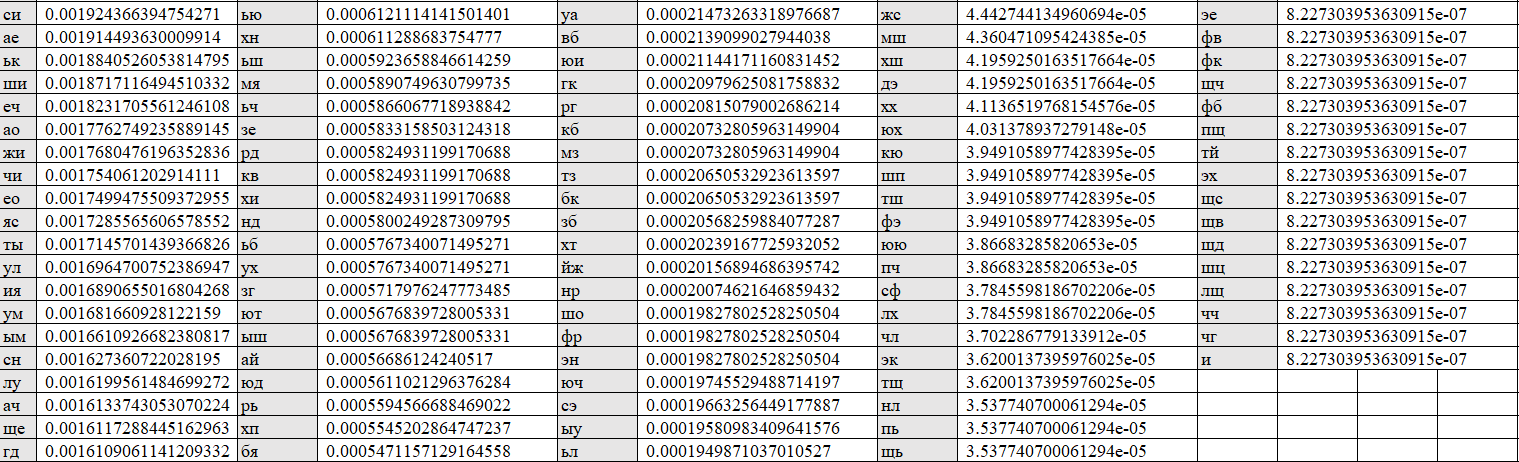




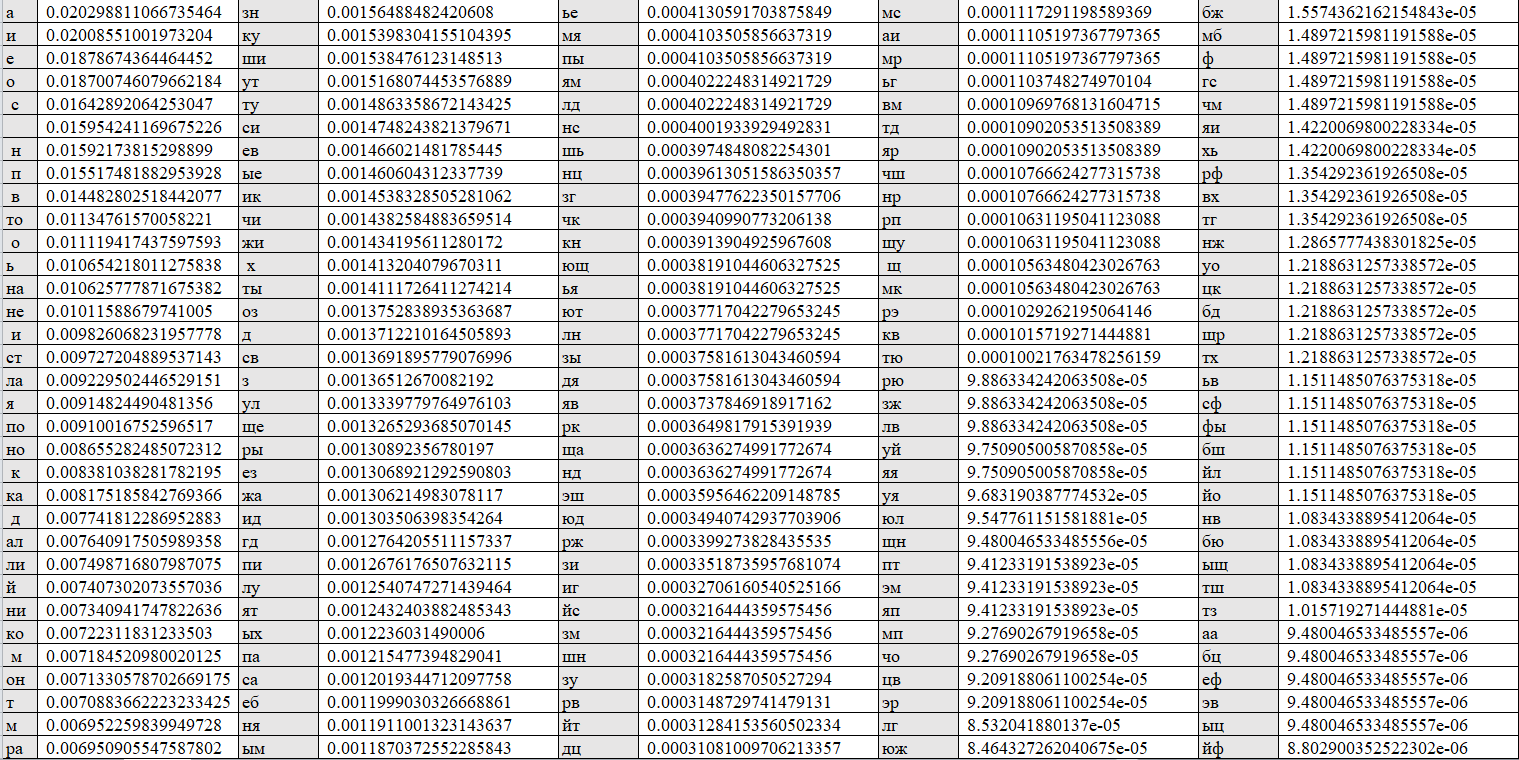


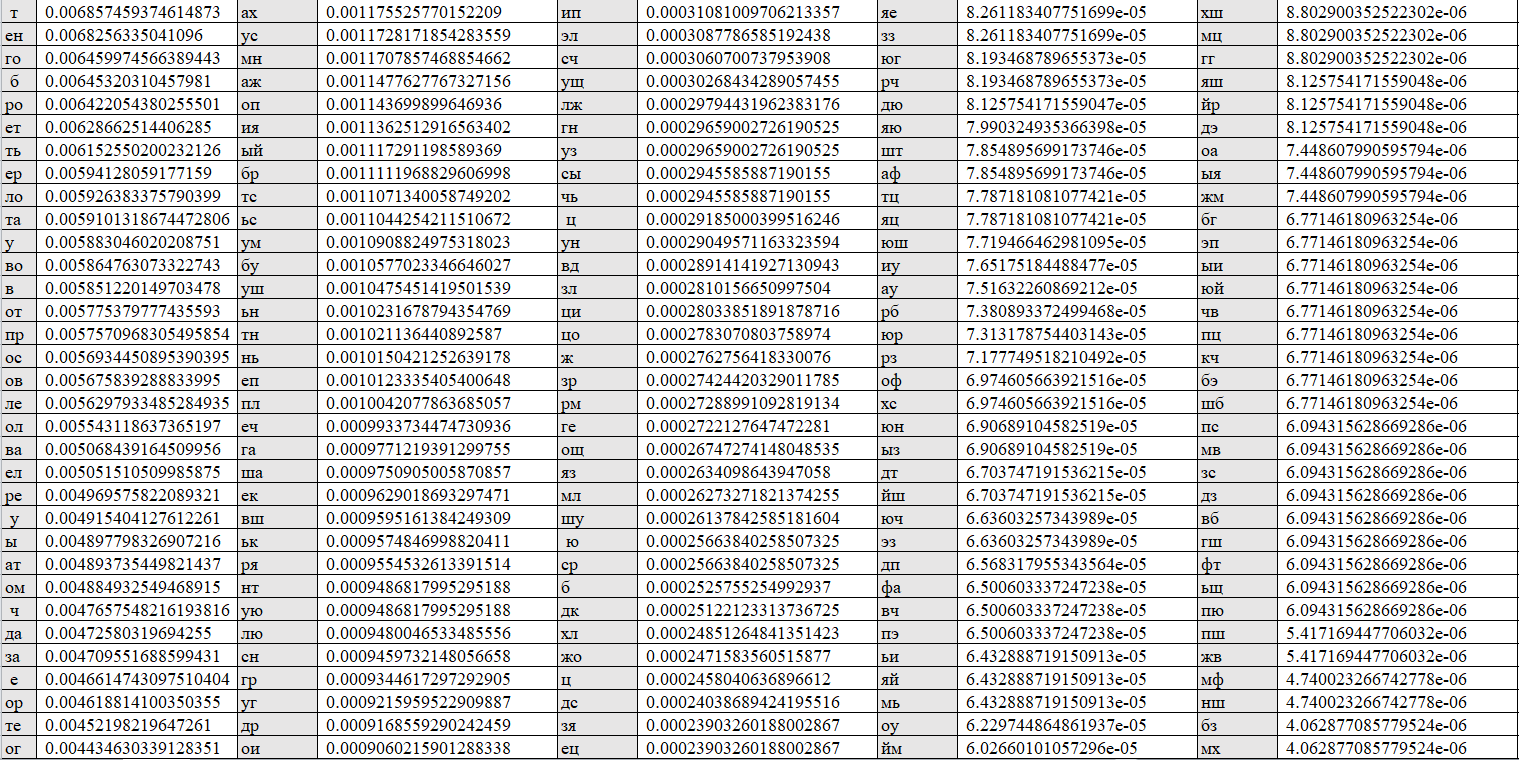


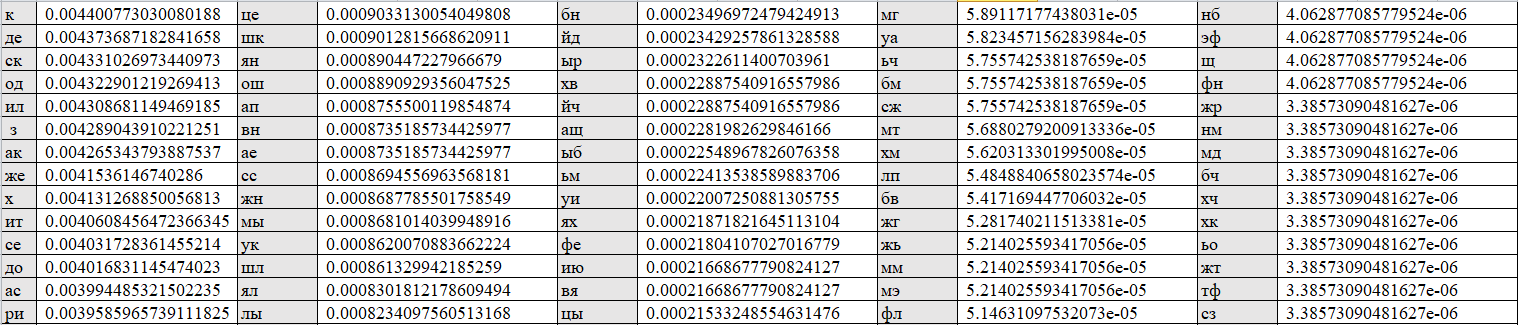


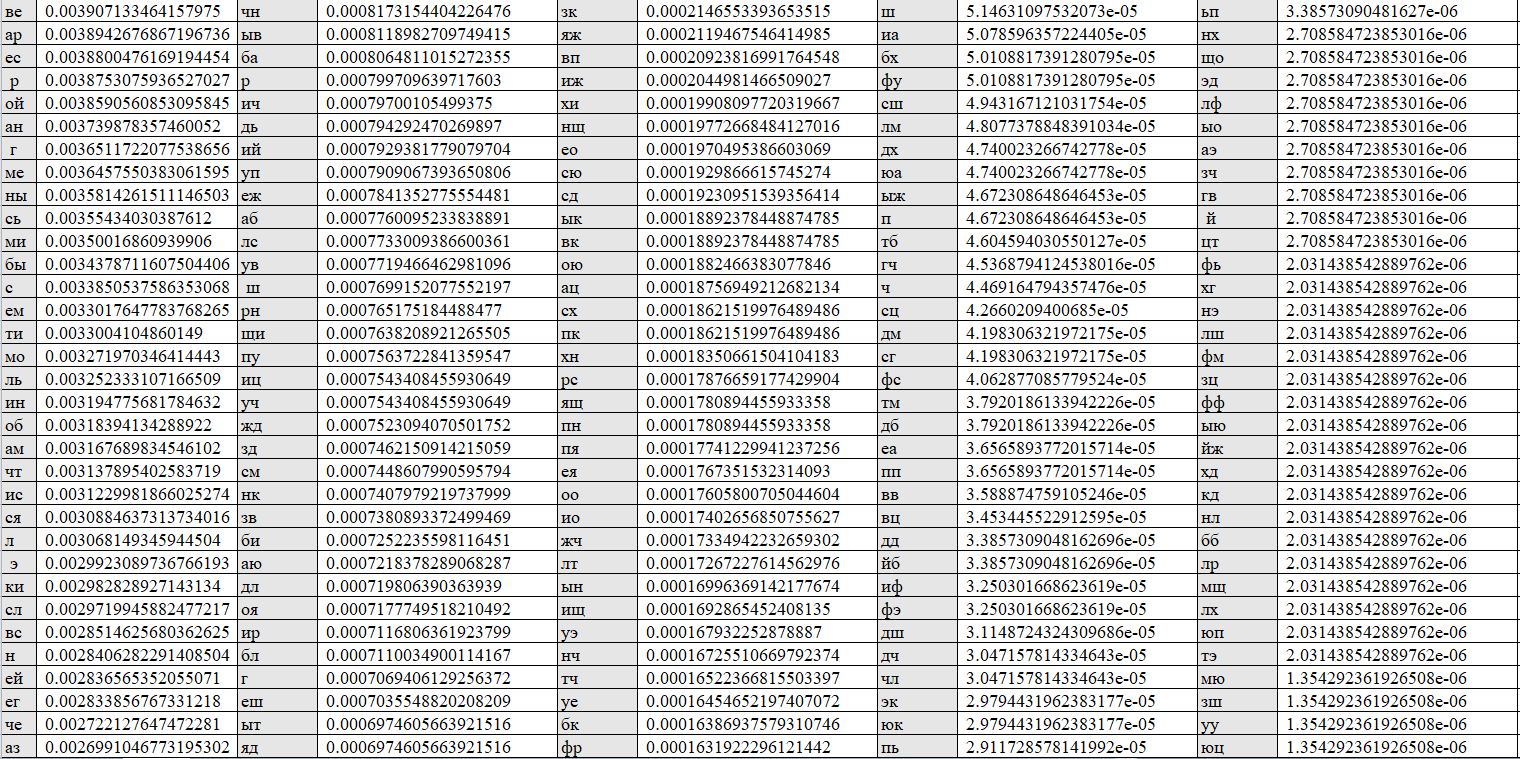


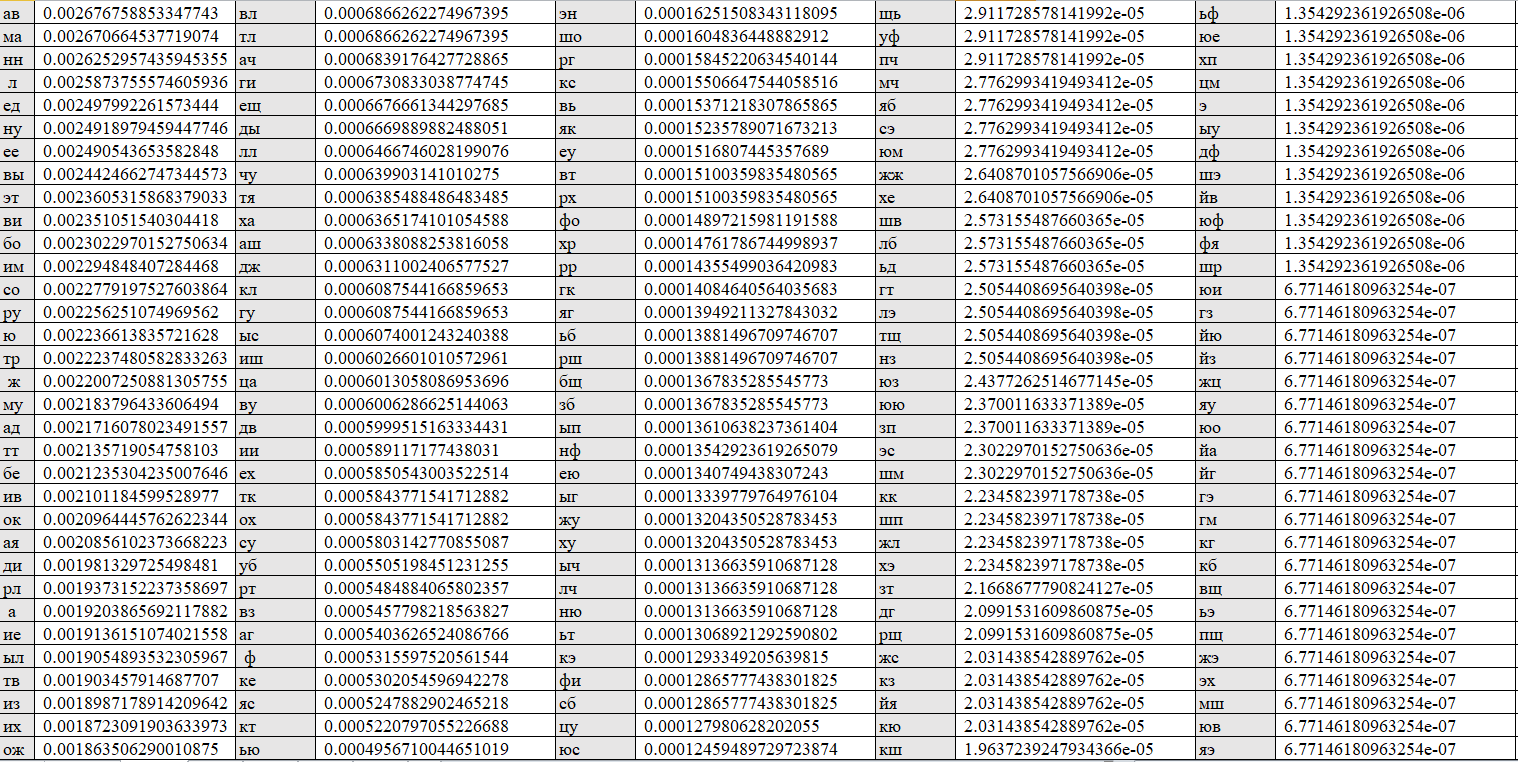
**6. Біграми з перетином та пробілами**

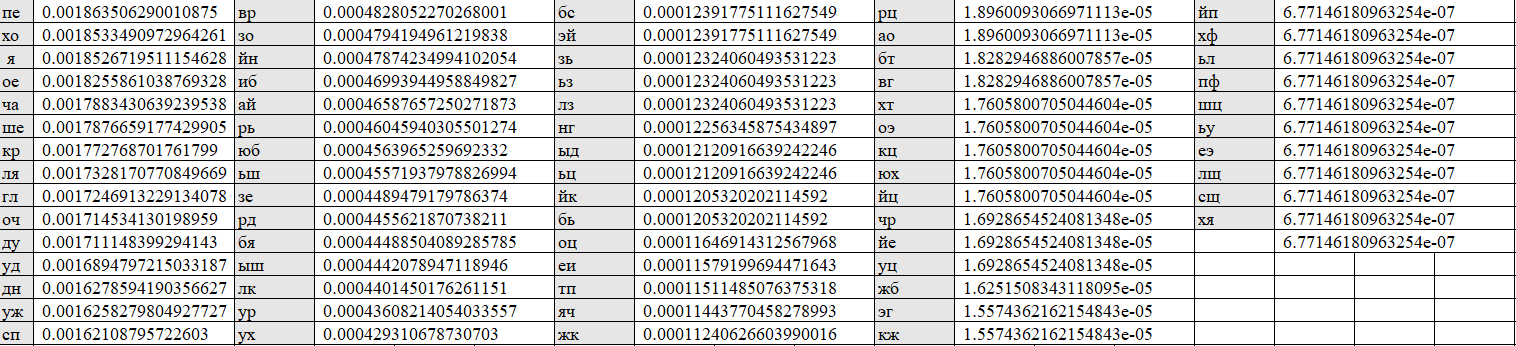












**Таблиці ентропій та надлишковостей**

|  |  |
| --- | --- |
| n-grama | Entropy |
| monograms without spaces | 4.45942982138092 |
| monograms with spaces | 4.343687189674594 |
| bigrams without spaces | 4.150768826748021 |
| bigrams with spaces | 3.970087903134239 |
| crossed bigrams without spaces | 4.150358576694166 |
| crossed bigrams with spaces | 3.970062068132714 |
| n=10 | 2.7875185351093<H<3.43559449734749 |
| n=20 | 2.31797550502925<H<3.04001473385416 |
| n=30 | 1.67569631769822<H<2.46344779608035 |

|  |  |
| --- | --- |
| n-grama | Надлишковість |
| monograms without spaces | 0.10811403572381606 |
| monograms with spaces | 0.13126256206508113 |
| bigrams without spaces | 0.16984623465039572 |
| bigrams with spaces | 0.2059824193731522 |
| crossed bigrams without spaces | 0.16992828466116683 |
| crossed bigrams with spaces | 0.20598758637345715 |
| n=10 | 0.312881100530502<R<0.44249629297813997 |
| n=20 | 0.391997053229168<R<0.53640489899415 |
| n=30 | 0.50731044078393<R<0.664860736460356 |

Висновок:

Під час виконання комп'ютерного практикуму я здобув навички у пошуку частот та ентропій для монограм, біграм, монограм та біграм з пробілом, біграм з перетином та біграм з перетином та пробілами. В якості тексту був взятий фрагмент роману 'Унесенные ветром'. Також я вдосконалив свої навички роботи з Python в тому числі бібліотеками numpy(для підрахунку ентропій) та tkinter(для побудови користувацького інтерфейсу). Вдосконалив роботу з засобом контролю версій. Також отримав досвід роботи з CoolPinkProgram, яка на основі відгаданих символів уривків тексту рахує ентропії для різних n-грам.