**Міністерство освіти і науки України**

**Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"**

**Фізико-технічний інститут**

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали студенти групи ФБ-23:

Кушнарьов Данііл та Присєвок Оксана

Київ - 2024

**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.

**Хід роботи:**

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.

|  |  |  |
| --- | --- | --- |
| Частота букв | | |
|  | з пробілом | без пробіла |
|  | 0.24167689107009496 |  |
| о | 0.08219220100288062 | 0.108386781353485 |
| а | 0.06587405313133468 | 0.08686805446861794 |
| е | 0.06327749919982929 | 0.08344398114033777 |
| и | 0.05397684839432412 | 0.07117922130909692 |
| н | 0.047931558732529606 | 0.06320730328285448 |
| т | 0.04560172836871866 | 0.06013495808280032 |
| с | 0.04270911127707244 | 0.05632046653218928 |
| л | 0.039800490771364556 | 0.052484871293884676 |
| р | 0.036421103168675986 | 0.048028475909261174 |
| в | 0.031976154913048116 | 0.04216692665237478 |
| к | 0.02902885949002454 | 0.03828033083547741 |
| м | 0.02333297770190974 | 0.03076917665721104 |
| д | 0.021629947722180732 | 0.02852339255848799 |
| п | 0.020132294889576442 | 0.026548439118498964 |
| у | 0.01902939293715993 | 0.02509404330828443 |
| я | 0.014756481382694975 | 0.019459358694103805 |
| з | 0.013897631494718874 | 0.018326794121216065 |
| ь | 0.013749599914648459 | 0.018131585010050633 |
| ы | 0.013704256908140403 | 0.01807179122825221 |
| г | 0.012726715032540276 | 0.016782707638303897 |
| ч | 0.012057238877627227 | 0.01589987003645662 |
| б | 0.011362423983783208 | 0.014983618262427873 |
| й | 0.008849887976101568 | 0.011670339294538893 |
| х | 0.006569401472314094 | 0.008663063798206538 |
| ж | 0.006520057612290622 | 0.008597994094484727 |
| ш | 0.005874586578470073 | 0.007746812024177791 |
| ю | 0.004507628294036061 | 0.005944205367019509 |
| э | 0.004070201643017177 | 0.005367371236728858 |
| ц | 0.0026992425050677477 | 0.003559488657647185 |
| щ | 0.0021617945161634482 | 0.0028507564792717823 |
| ф | 0.0015923396991358156 | 0.002099816925509258 |
| ъ | 0.00027872612824069137 | 0.0003675558939961767 |
| ё | 3.067321028486077e-05 | 4.044873474599074e-05 |

Матриця частот біграм з пробілами і перетинами

Зображення, що містить текст, знімок екрана, ряд, візерунок

Автоматично згенерований опис

Матриця частот біграм без пробілів з перетинами

Зображення, що містить текст, знімок екрана, ряд, число

Автоматично згенерований опис

Матриця частот біграм з пробілами без перетинів

Зображення, що містить текст, знімок екрана, ряд, число

Автоматично згенерований опис

Матриця частот біграм без пробілів і перетинів

Зображення, що містить текст, знімок екрана, ряд, число

Автоматично згенерований опис

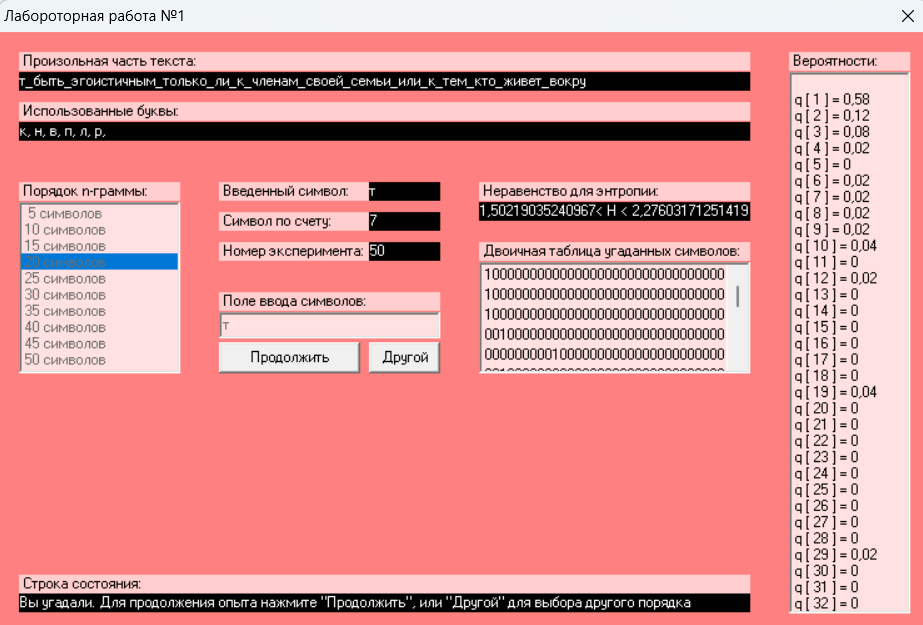
|  |  |  |
| --- | --- | --- |
|  | Ентропія | Надлишковість |
| H1 з пробілами | 4.174091254429877 | 0.1795338099405135 |
| H1 без пробілів | 4.4522887764954024 | 0.11737888215172909 |
| H2 з пробілами і перетинами | 1.9163815358245144 | 0.6233129173374901 |
| H2 без пробілів з перетинами | 2.077326218603828 | 0.5881911346633593 |
| H2 з пробілами без перетинів | 1.9163808165517653 | 0.6233130587189157 |
| H2 без пробілів і перетинів | 2.0773270059012603 | 0.588190978589624 |

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

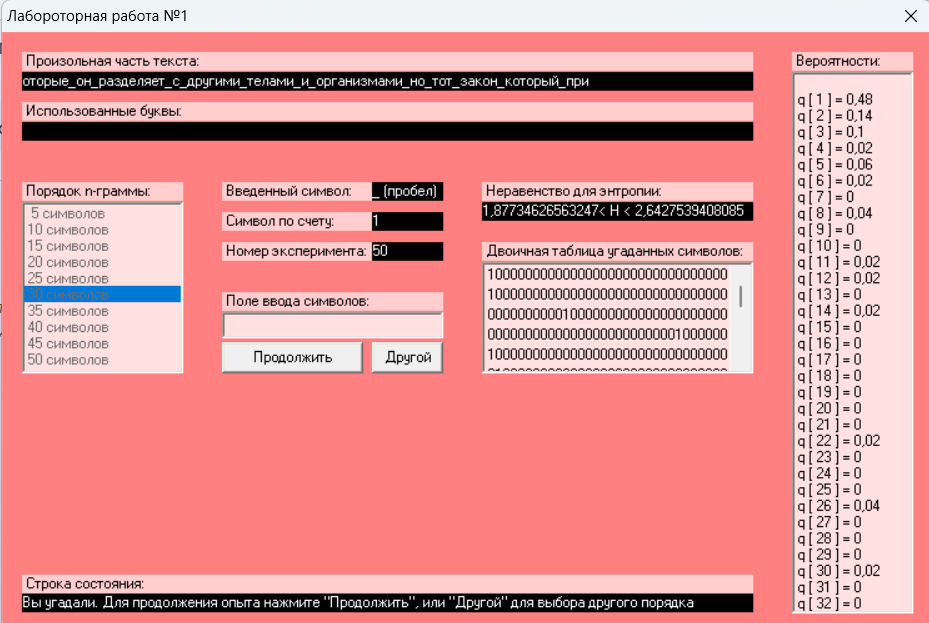
H⁽¹⁰⁾Зображення, що містить текст, знімок екрана, програмне забезпечення, Веб-сайт

Автоматично згенерований опис

H⁽²⁰⁾



H⁽³⁰⁾



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

|  |  |  |
| --- | --- | --- |
|  | Ентропія | Надлишковість |
| H⁽¹⁰⁾ | 1,71743121272368 < H < 2,5366396833792 | 0,49267206332416 < R < 0,656513757455264 |
| H⁽²⁰⁾ | 1,50219035240967 < H < 2,27603171251419 | 0,544793657497162 < R < 0,699561929518066 |
| H⁽³⁰⁾ | 1,87734626563247 < H < 2,6427539408085 | 0,4714492118383 < R < 0,624530746873506 |

**Висновок:**

Під час виконання даної лабораторної роботи, ми отримали практичні навички з підрахунку ентропії та надлишковості, ознайомились з програмою CoolPinkProgram і завдяки ній обрахували значення H⁽¹⁰⁾, H⁽²⁰⁾, H⁽³⁰⁾.