**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 |
|  | 0.14817289384313237 |  |
| о | 0.09232679830599627 | 0.108386781353485 |
| а | 0.07399656345547997 | 0.08686805446861794 |
| е | 0.07107984498098216 | 0.08344398114033777 |
| и | 0.06063239010622728 | 0.07117922130909692 |
| н | 0.05384169424341342 | 0.06320730328285448 |
| т | 0.05122458732253633 | 0.06013495808280032 |
| с | 0.047975300023519504 | 0.05632046653218928 |
| л | 0.04470803603128544 | 0.052484871293884676 |
| р | 0.04091195764691078 | 0.048028475909261174 |
| в | 0.035918931105821304 | 0.04216692665237478 |
| к | 0.032608223438312225 | 0.03828033083547741 |
| м | 0.026210018710741523 | 0.03076917665721104 |
| д | 0.024296998940873157 | 0.02852339255848799 |
| п | 0.022614680067292756 | 0.026548439118498964 |
| у | 0.021375786293071034 | 0.02509404330828443 |
| я | 0.016576009204066926 | 0.019459358694103805 |
| з | 0.015611260001408174 | 0.018326794121216065 |
| ь | 0.015444975589148669 | 0.018131585010050633 |
| ы | 0.015394041625033144 | 0.01807179122825221 |
| г | 0.014295965281013167 | 0.016782707638303897 |
| ч | 0.013543940281425132 | 0.01589987003645662 |
| б | 0.01276345218424313 | 0.014983618262427873 |
| й | 0.009941111349135846 | 0.011670339294538893 |
| х | 0.007379432565678598 | 0.008663063798206538 |
| ж | 0.007324004428258763 | 0.008597994094484727 |
| ш | 0.006598944468496594 | 0.007746812024177791 |
| ю | 0.005063435256190349 | 0.005944205367019509 |
| э | 0.004572072308252351 | 0.005367371236728858 |
| ц | 0.0030320689226417947 | 0.003559488657647185 |
| щ | 0.0024283516420960227 | 0.0028507564792717823 |
| ф | 0.0017886809751157624 | 0.002099816925509258 |
| ъ | 0.0003130940735336636 | 0.0003675558939961767 |
| ё | 3.4455328666384034e-05 | 4.044873474599074e-05 |

Матриця частот біграм з пробілами і перетинами

Зображення, що містить текст, знімок екрана, Шрифт, документ

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, Шрифт, ряд

Автоматично згенерований опис

Матриця частот біграм без пробілів і перетинів

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

Автоматично згенерований опис

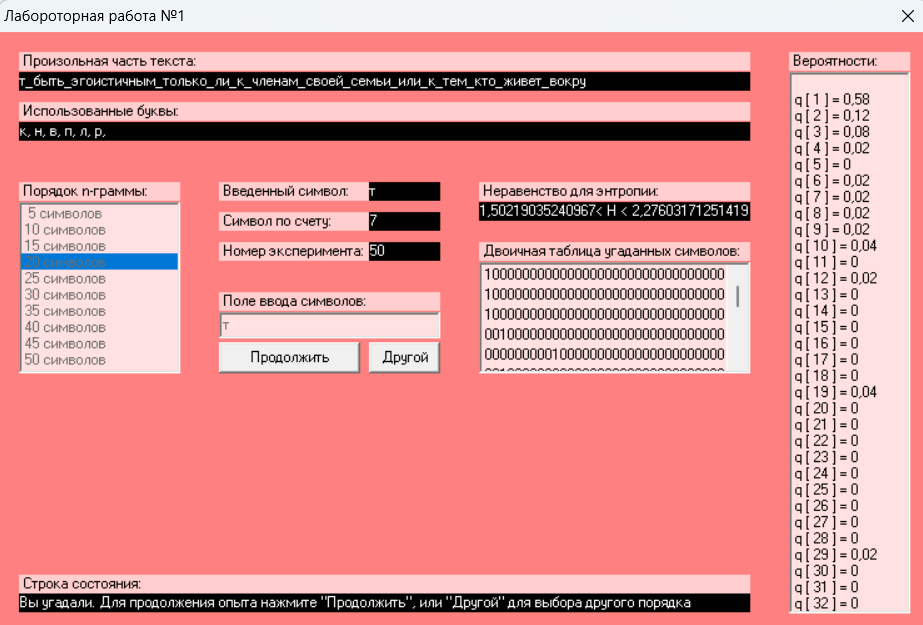
|  |  |  |
| --- | --- | --- |
|  | Ентропія | Надлишковість |
| H1 з пробілами | 4.397829284684674 | 0.1795338099405135 |
| H1 без пробілів | 4.4522887764954024 | 0.11737888215172909 |
| H2 з пробілами і перетинами | 4.027285872666266 | 0.20839011540053143 |
| H2 без пробілів з перетинами | 4.154652437207656 | 0.1763822693267184 |
| H2 з пробілами без перетинів | 4.027285125159497 | 0.20839026233168212 |
| H2 без пробілів і перетинів | 4.154654011802521 | 0.17638195717924798 |

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

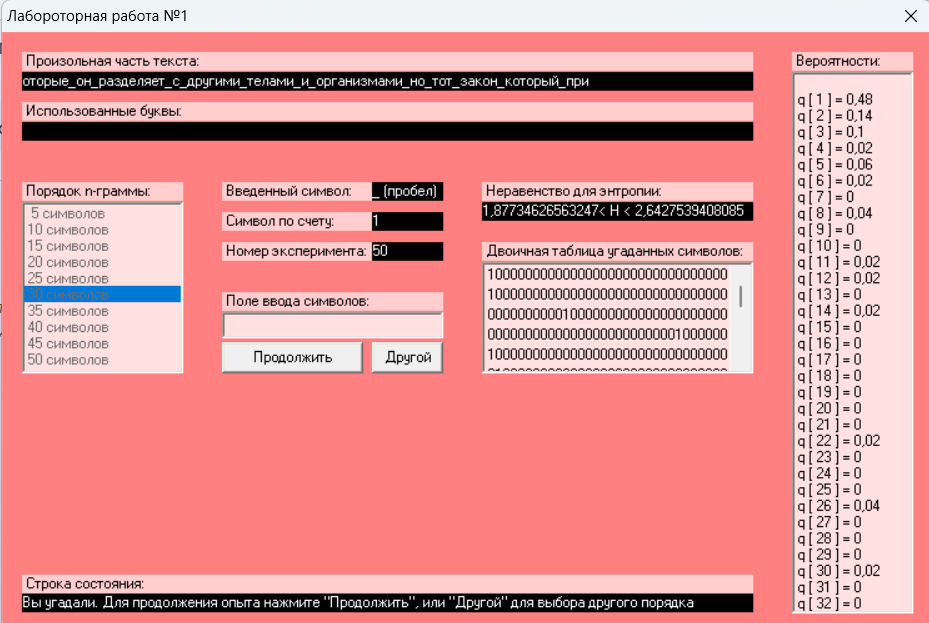
H⁽¹⁰⁾Зображення, що містить текст, знімок екрана, програмне забезпечення, Веб-сайт

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⁾.