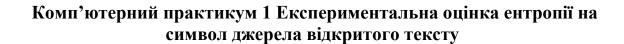
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ СІКОРСЬКОГО» ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ



Виконали:

Толмачов Євгеній

Котович Анна

Перевірив: Деркач О.Г

Мета роботи

Засвоєння понть 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 дерела

Постановка задачі

- 1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення $H_{(10)}$, $H_{(20)}$, $H_{(30)}$.
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

Під час виконання лабораторної роботи особливих проблем не виникло. Єдиними труднощами були ознайомлення з модулями для роботи з файлами (що не було чимось важким саме по собі), уникання повторених поспіль пробілів та підрахунок біграмм що не перетинаються

Результати

1. Код та таблиці біграмм наведені у додатку до листа та за посиланням https://github.com/ZheZheDoshka/labs_2021_5/tree/main/%D0%BA%D1%80%D0%B8%D0%BF%D1%82%D0%B01

```
a: 122891 6: 25734 в: 69076 г: 26756 д: 43934 е: 129487 е: 0 ж: 16739 э: 24341 ж: 98484 й: 15473 ж: 58935 л: 75058 м: 4

ж215 ж: 102532 о: 169391 л: 36461 р: 59259 с: 78473 т: 88755 у: 40448 ф: 1563 ж: 11561 ц: 4068 ч: 24661 ш: 12884 щ: 4418

ж: 413 ж: 27750 ж: 29474 »: 5233 ж: 9400 ж: 32503 : 297579

total without probell.77895e+06

H = 4.4334

M = 4.35151

H2 - neperesek = 4.12613

d2 - neperesek s prob = 3.9464

H21 - peresek s prob = 3.94651

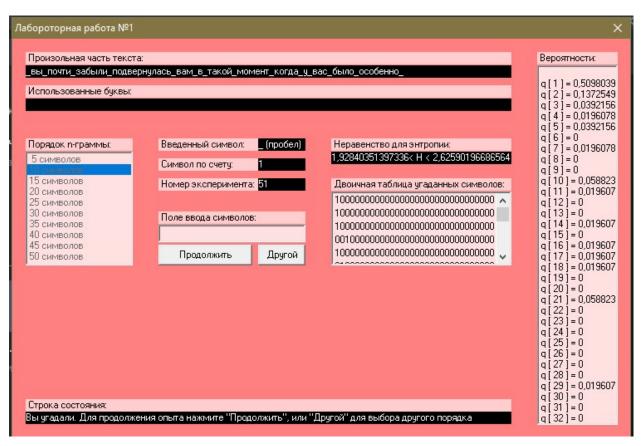
R = 0.433704

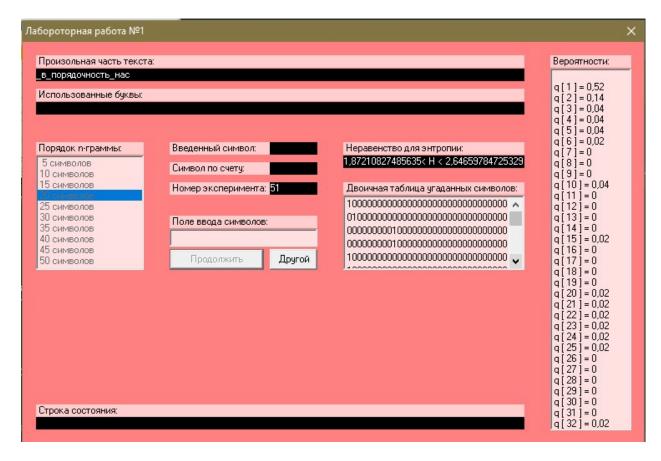
W:\lab1\крыпта1\Debug\крыпта1.exe (процесс 4132) завершает работу с кодом 0.

жНобы автоматически закрывать консоль при остановке отладки, установите параметр "Сервис" -> "Параметры" -> "Отладка" -
"Автоматически закрыть консоль при остановке отладки".
```

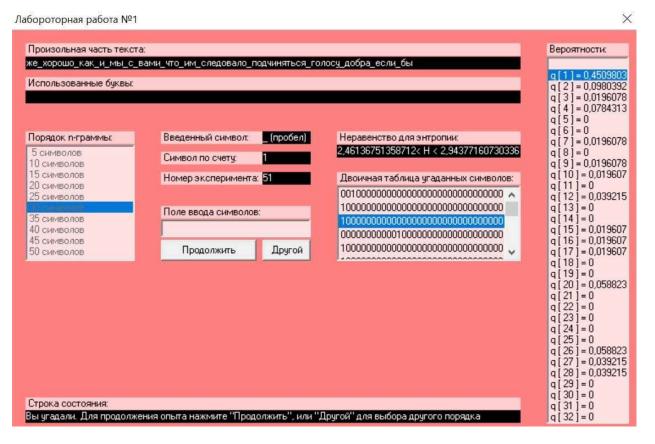
2. Результати роботи CoolPinkProgram

Оцінка для $H(10): 1.96 < H_{(10)} < 2.81 \Rightarrow H_{(10)} \approx 2.387$





Оцінка для $H(30): 0.787 < H(30) < 1.34 \Rightarrow H(30) \approx 1.1$



Значення ентропій

Ентропія монограм для тексту без пробілів: 4.44334

Ентропія монограм для тексту з пробілами: 4.35151

Ентропія біграм для тексту без пробілів, біграми перетинаються: 4.12613

Ентропія біграм для тексту без пробілів, біграми не перетинаються: 4.12688

Ентропія біграм для тексту з пробілами, біграми перетинаються: 3.94651

Ентропія біграм для тексту з пробілами, біграми не перетинаються: 3.94651.

Оцінки надлишковості

Надлишковість російської мови: $R = 1 - H\infty/H0 = 0.4337$

Висновок:

В ході цієї лабораторної було підраховно частоти букв та біграм, а також оцінено відповідні значення ентропій та надлишковость джерела. За допомогою програми CoolPinkProgram практичним шляхом оцінили значення ентропій $H_{(10)}$, $H_{(20)}$, та $H_{(30)}$