Міністерство освіти і науки України
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
Фізико-технічний інститут

СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериметальна оцінка ентропії на символ джерела відкритого тексту

Виконав: студент гр. ФІ-94, Кріпака І.А.

Перевірив: Чорний О.М.

Київ – 2021

1) **Мета комп’терного практикуму**

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

2) **Постановка задачі**

Створити програму для експериментальної оцінки ентропії на символ джерела відкритого тексту, порівняти різні моделі джерела відкритого тексту для наближеного визначення ентропії.

3) **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.
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

4) **Опис труднощів**

Реалізуючи програмний код у мене виникла проблема із чищенням файлу від інших “непотрібних” символів, вирішилася досить легко за допомогою регулярних виразів.

5) **Практична частина**
Посилання на код

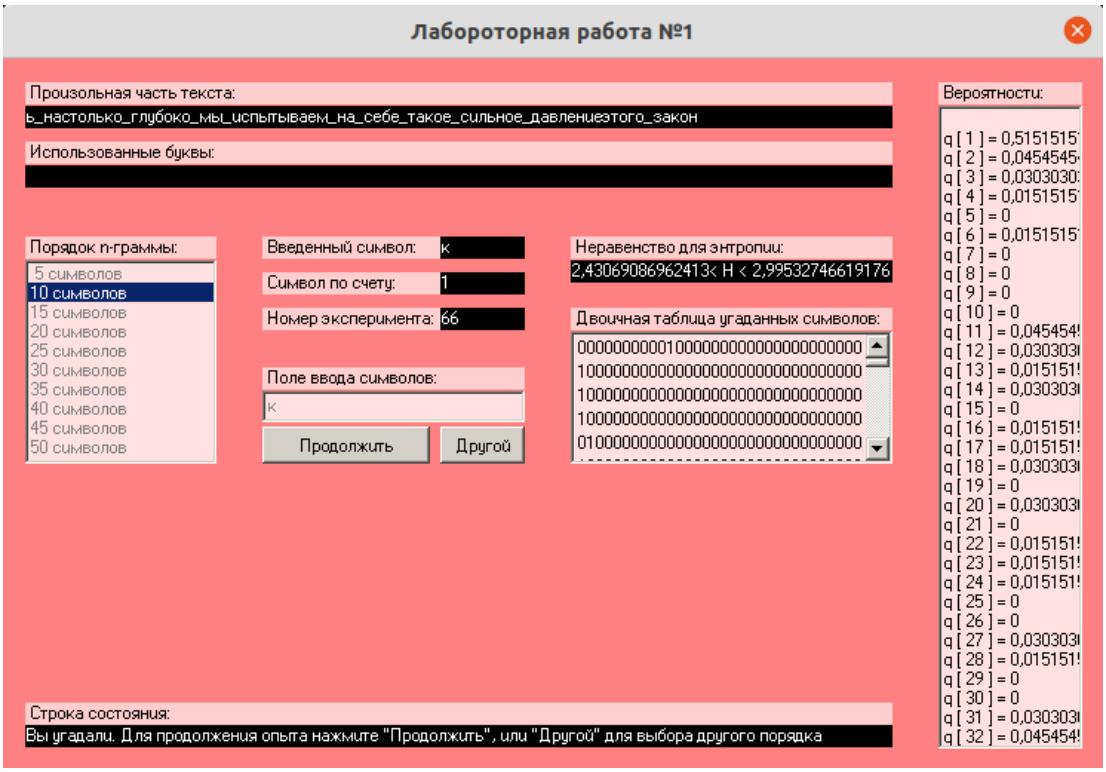
```
Total letters sum: 5696814.0
Total letters sum: 4793739.0
Total letters sum: 5681234.0
Total letters sum: 4793738.0
Entropy for:
One symbol with gap (H1): 4.3840365490920545
One symbol without gaps (H1): 4.460337963530066
Bigram with gaps (H2): 4.000823175252397
Bigram without gaps (H2): 4.153176280546273
```

6) **Результати підрахунку частот** у файлі cp1_probability_tables.ods

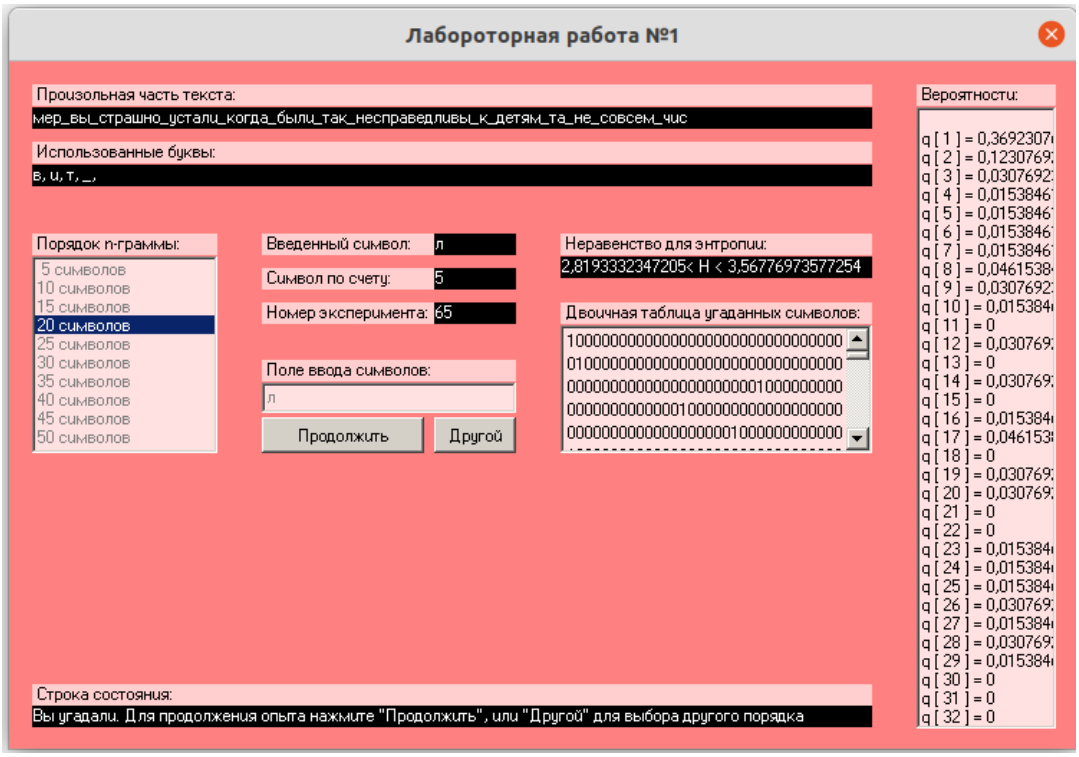
7) **Значення ентропії**

	Алфавіт із пробілом	Алфавіт без пробілу
H0	5	4.954196
H1	4.3840365490920545	4.460337963530066
H2	4.000823175252397	4.153176280546273

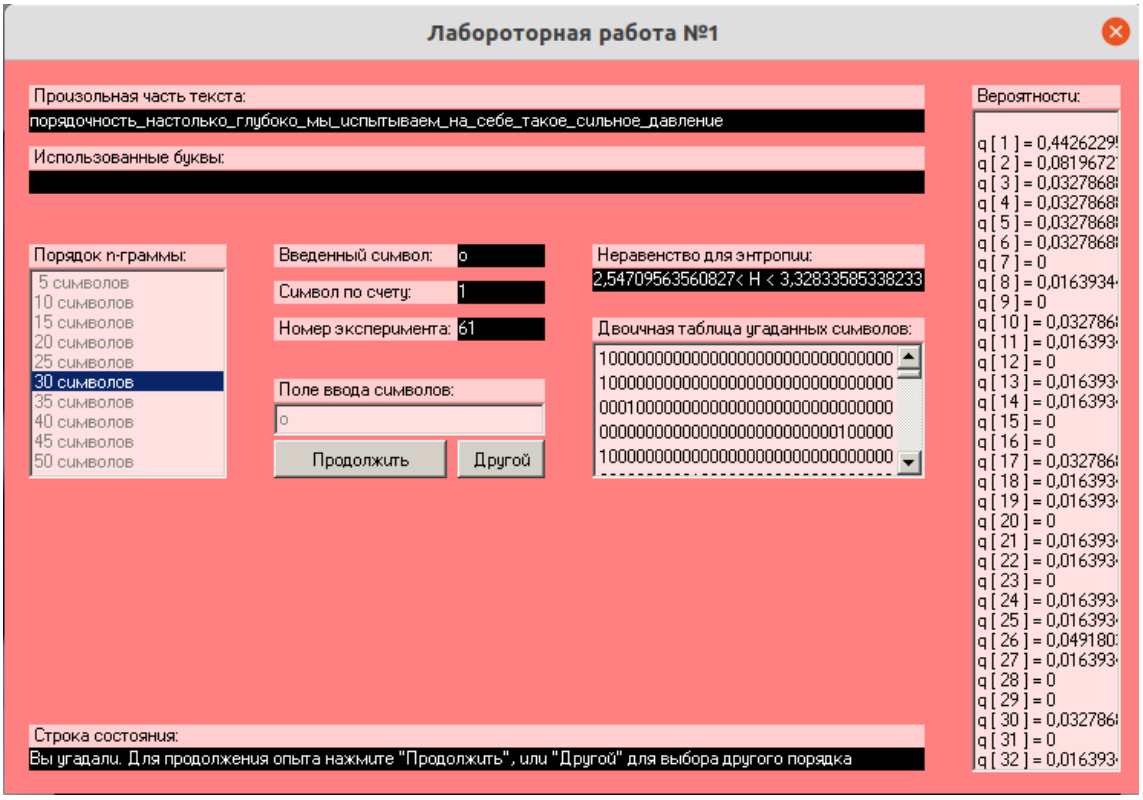
Надлишковість для H(10)
2.43069086962413 < H(10) < 2.99532746619176
H(10) = 2.9, H0 = 5
R = 1 - 2.9/5 = 1 – 0.58 = **0.42**



Надлишковість для H(20)
2.8193332347205 < H(20) < 3.5677693577254
H(20) = 3.2, H0 = 5
R(20) = 1 - 3.2/5 = 1 – 0.64 = **0.36**



Надлишковість для H(30)
2.54709563560827 < H(30) < 3.32833585338233
H(30) = 3, H0 = 5
R(30) = 1 - 3/5 = 1 – 0.6 = **0.4**



	Для H(10)	**Для H(20)**	**Для H(30)**
R(*)	0.42	0.36	0.4