Міністерство освіти і науки України Національний технічний університет України “Київський політехнічний інститут ім. Ігоря Сікорського” Фізико-технічний інститут

Лабораторна робота №1 з предмету «Криптографія»

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав студент 3 курсу

Кравець Андрій ФБ-02

Київ - 2022

Мета: 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.

Постановка задачі:

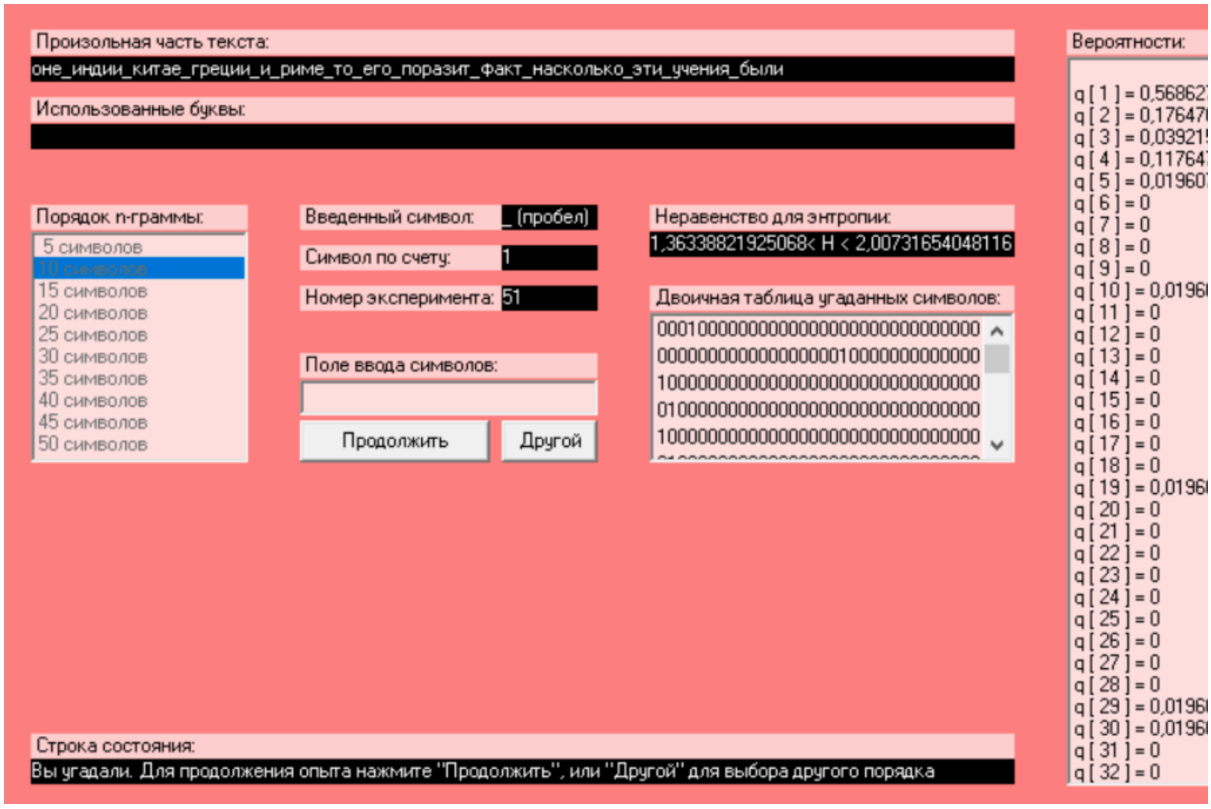
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними 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.

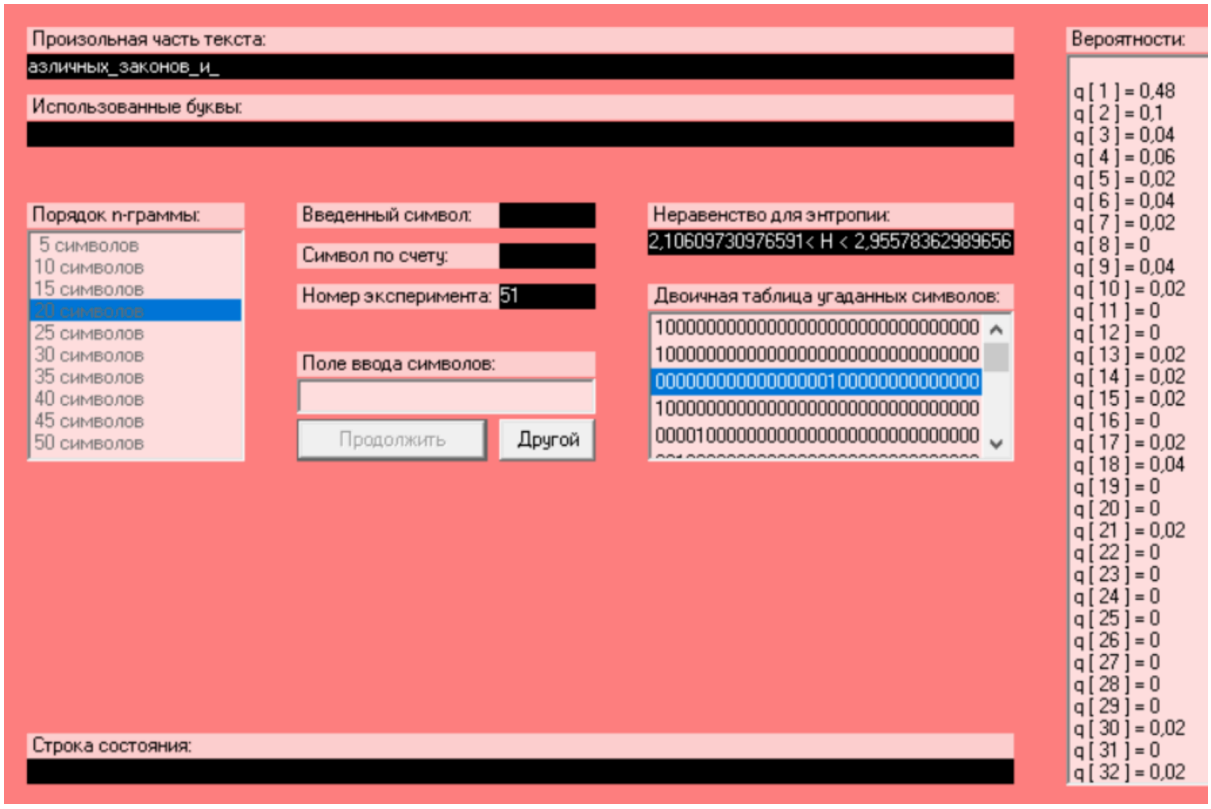
Результати виконання:

```
H1 монограм з пробілом 4.379095399823697
надлишковість монограм з пробілом 0.13923786050740905

H1 монограм без пробілів 4.479179556501572
надлишковість монограм без пробілів 0.11204805760275627


H2 біграм з пробілами та з перетинами 3.9653930266671646
надлишковість біграм з пробілами та з перетинами 0.22055587423364942

H2 біграм з пробілами та без перетинів 3.9639366646633825
надлишковість біграм з пробілами та без перетинів 0.22084213912623474

H2 біграм без пробілів та з перетинами 4.16024639153056
надлишковість біграм без пробілів та з перетинами 0.1752733245871636

H2 біграм без пробілів та без перетинів 4.159935381649082
надлишковість біграм без пробілів та без перетинів 0.17533497914351248
```

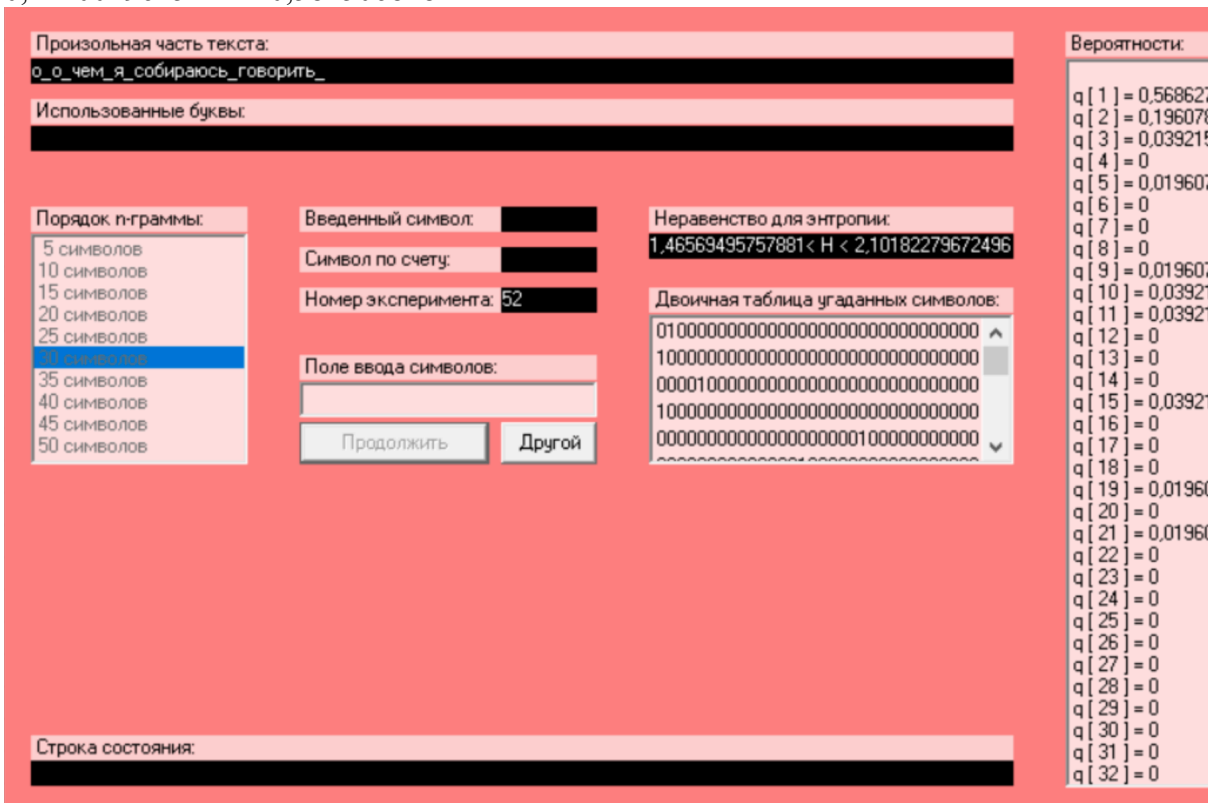
Оцінки $H^{(10)}$, $H^{(20)}$, $H^{(30)}$ у CoolPinkProgram



0,6021325946 < R< 0,7303937861



0,4140029645 <R < 0,5825068482



0,7093803606 < R < 0,601537875

Висновки:

Під час виконання лабораторної роботи я порівняв різні моделі джерела відкритого тексту для наближеного визначення ентропії, освоїв поняття ентропії на символ джерела та його надлишковості.