Міністерство освіти і науки України

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

Фізико-технічний інститут

Криптографія

Лабораторна робота №1

Виконав студент групи ФБ-13

Лагно Костянтин

**Київ 2023**

**Експериментальна оцінка ентропії на символ джерела відкритого тексту**

**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.

**Опис труднощів**: серед труднощів, які виникли під час виконання даного практикуму, можу виділити такі:

* Проблеми з форматуванням тексту – прийняв рішення написати окрему програму для підготовки тексту для подальшої роботи, оскільки початкова версія пропускала певні спеціальні символи, які були в тексті, як-от спецсимвол перенесення рядка, тощо, і потм вони вспливали під час обрахунку біаграм.
* Недостатній рівень знання мови програмування – довелось переписувати програму декількома версіями.
* Криві ручки – невиправимо ;)

**Хід роботи:**

Відформатувавши текст так, як мені потрібно, приступив до самого підрахунку частот. Усі вихідні дані додатково додам окремою таблицею, тут наведу лише частину для загального розуміння.

1. Власна програма
2. Текст з пробілами, частота букв, їх ентропія H1 та надлишковість R1:





1. Текст без пробілів, частота букв, ентропія H1 та надлишковість R1:





1. Текст з пробілами, біграми з перетином, частота, ентропія H2 та надлишковість R2:





1. Текст без пробілів, біграми з перетином, частота, ентропія H2 та надлишковість R2:





1. Текст з пробілами, біграми без перетину, частота, ентропія H3 та надлишковість R3:



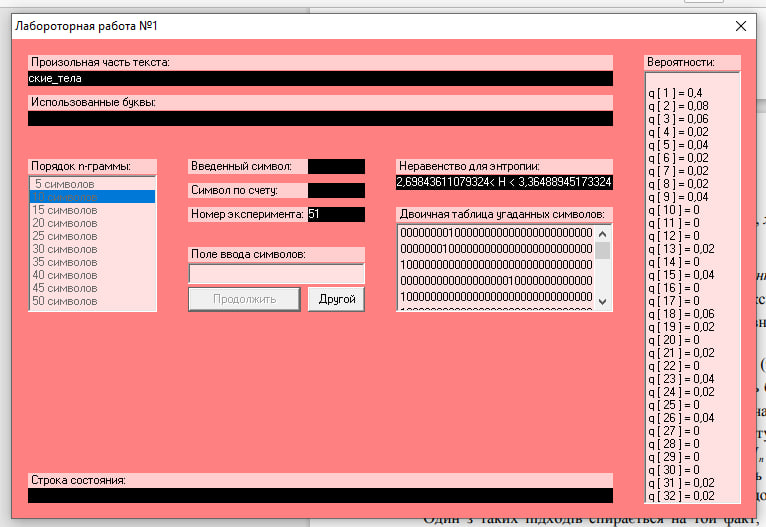


1. Текст без пробілів, біграми без перетину, частота, ентропія H3 та надлишковість R3:

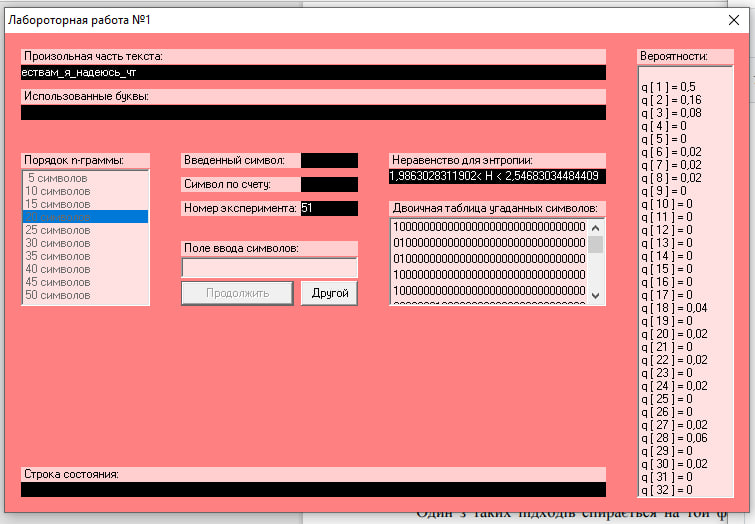




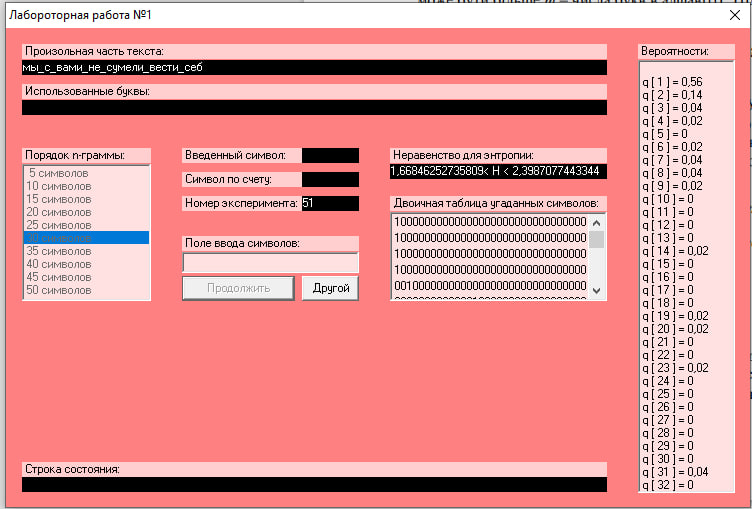
1. Оцінки H10, H20, H30 програмою CoolPinkProgram



H10



H20



H30

Таблиця ентропій для CPP:



1. Оцінка надлишковості російської мови.

Формула:



Порівнюючи ці дані з даними, отриманими моєю програмою, можна зробити висновок, що при меншому значення n надлишковість зростає швидше, ніж при більших значеннях n. Це пов’язано з зниженням ентропії, оскільки в моїй програмі ентропія в середньому була в районі 4, при n = 2, а при використанні CPP і n > 10, ентропія зменшувалась.

Також можна сказати, що надлишковість російського тексту прямує до значення приблизно 0.74.

**Висновки:** у ході виконання лабораторної роботи я дізнався детальніше про ентропію та надлишіковість в теорії інформації, використав ці знання під час практичних дослідів.