**КРИПТОГРАФІЯ**

**КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1**

**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.

**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.

**Хід Роботи**

1. **Розрахунок частот та ентропії для літер**

Маємо два випадки: з пробілом та без. Далі наведені таблиці з отриманими значеннями:

**З пробілом**

H1 = 4.354029047, R = 0.129194190592



**Без пробілу**

H1 = 4.44876864953, R = 0.12919419



1. **Розрахунок частот та ентропій для біграм**

Загалом маємо 4 випадки , а саме: біграми з перетином з пробілом, біграми з перетином без пробілу, біграми без перетину з пробілом, біграми без перетину без пробілу. Далі будуть наведені невеликі частини таблиць з частотами біграми, бо тих біграм багато і не хочеться засоряти і так кривий протокол. Усі частоти можна буде переглянути у csv файлах, як мають бути у цій же папці)

**З перетином біграм:**

**З пробілом**

H2 = 3.9432864, R = 0.1749084



**Без пробілу**

H2 = 4.126525, R = 0.1508674



**Без перетину біграм:**

**З пробілом**

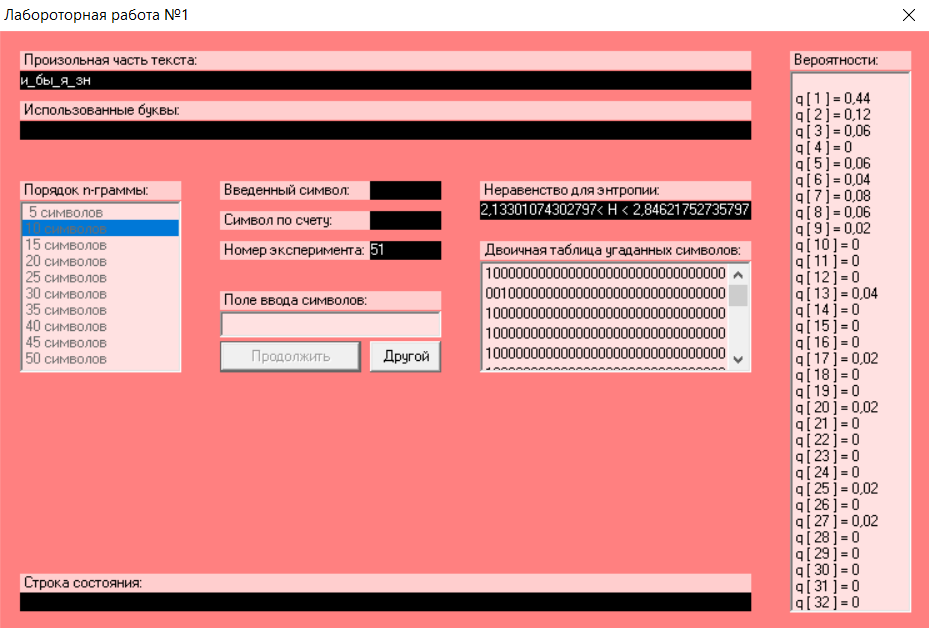
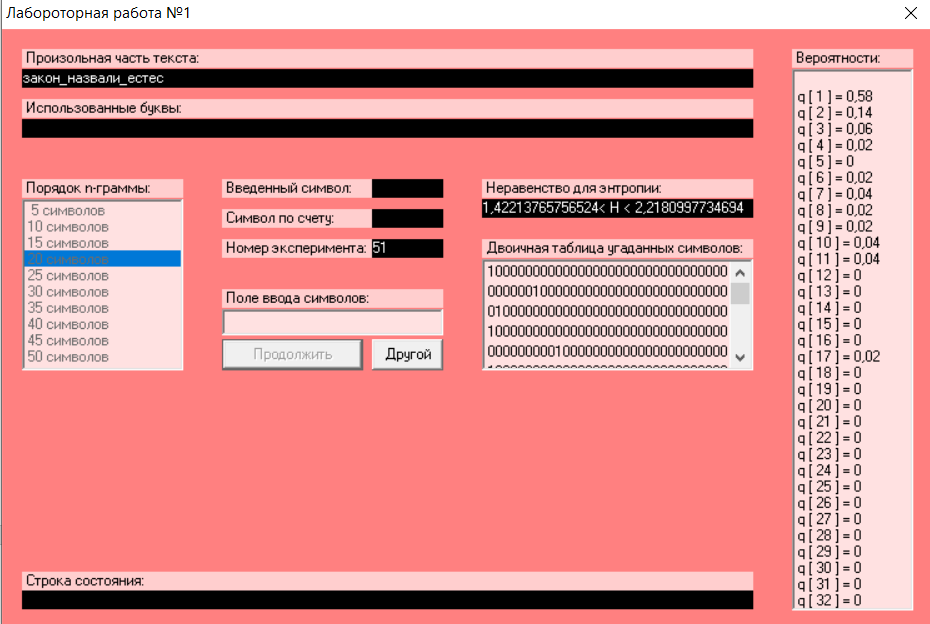
H2 = 3.94394, R = 0.1703768

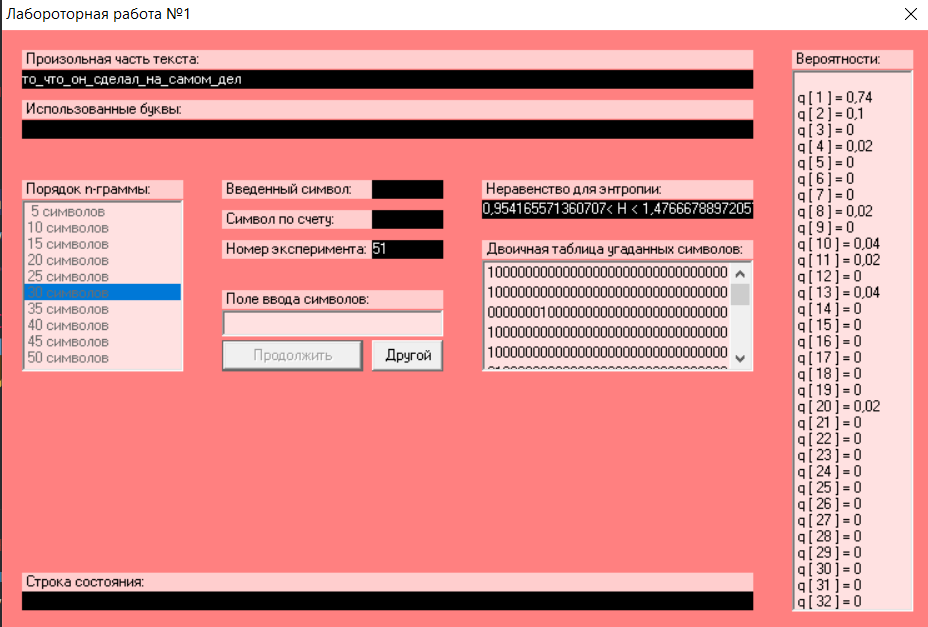
****

**Без пробілу**

H2 = 4.126635, R = 0.1465657



1. **Оцінка значення Н(10), Н(20), Н(30) за допомогою програми CoolPinkProgram**
2. 2,13301074302797 < Н(10) < 2,84621752735797
3. 1,42213765756524 < Н(20) < 2,2180997734694
4. 0,954165571360707 < Н(30) < 1,4766678897205



Оцінимо надлишковість російської мови, використовуючи отримані дані:

1. Для Н(10) надлишковість 0.430756494529 < R < 0.573397851394
2. Для Н(20) надлишковість 0.556380045306 < R < 0.715572468487
3. Для Н(30) надлишковість 0.704666422056 < R < 0.809166885728

**Труднощі**

Під час виконання даної лабораторної роботи ми зіткнулися з наступними труднощами:

1. На етапі розробки коду для виконання підрахунків ентропії для літер та біграм виникла проблема підрахунку для випадків, коли якась літера (в нас це була літера “ё”), та деякі біграми не були наявними у тексті, тому під час підрахунку за формулою log₂0 отримували помилку виконання коду. Після виявлення причини помилки, ми прийшли до висновку: необхідно виконати об’єднання літер “ё” з “е”, та “ъ” з “ь”, а також трохи змінити код так, що перед виконанням обрахунків перевірялося умова не рівності частоти нулю.
2. Після виконання програми та аналізу отриманих результатів, довелося визначитися з оформленням даних для біграм, оскільки досить складно надати дані для всіх біграм (враховуючи їх загальну кількість), тому ми вирішили надати як приклад 20 довільно вибраних біграм й прикріпити csv-файл з повним списком отриманих результатів.

**Висновок**

В результаті виконання даної лабораторної роботи ми ознайомилися з поняттям ентропії, отримали навички її підрахунку для подальшого розгляду надлишковості вибраної мови. Також провели експерименти з наданою програмою CoolPinkProgram, де спробували вгадати зашифровану літеру по попереднім n-грамам. Загалом поставлене завдання допомогло нам розібратися у значенні криптографії для захисту інформації, розвити навички роботи у команді та автоматизації роботи з даними за допомогою програмування.