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

**КРИПТОГРАФІЯ**

**Комп’ютерний практикум №1  
Експериментальна оцінка ентропії на символ джерела відкритого тексту**

Роботу виконали:  
Касаб О.Р.  
Косигін О.С.  
Групи ФБ-06

Київ – 2022

**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.

**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 (10) H, (20) H, (30) H. 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

**Хід Роботи**Для виконання даної роботи потрібно було обрати текст російською мовою. Ми обрали текст ранобе

“No game no life”, коротко – “NGNL”

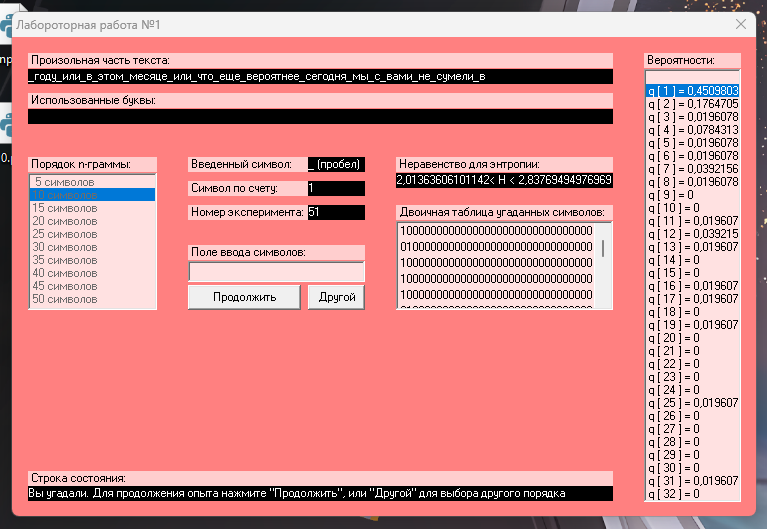
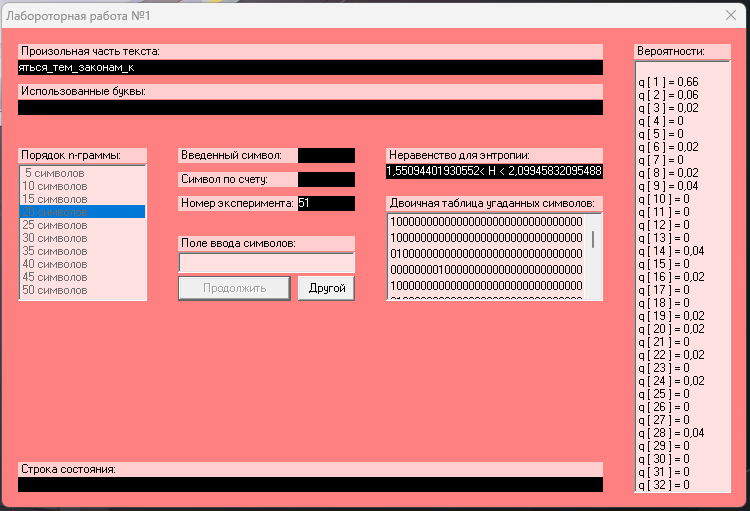
Найбільшими труднощами особисто для нас стали розбір та зрозуміння завдань лабораторної роботи, підрахунок необхідних n-грам та вивід результатів

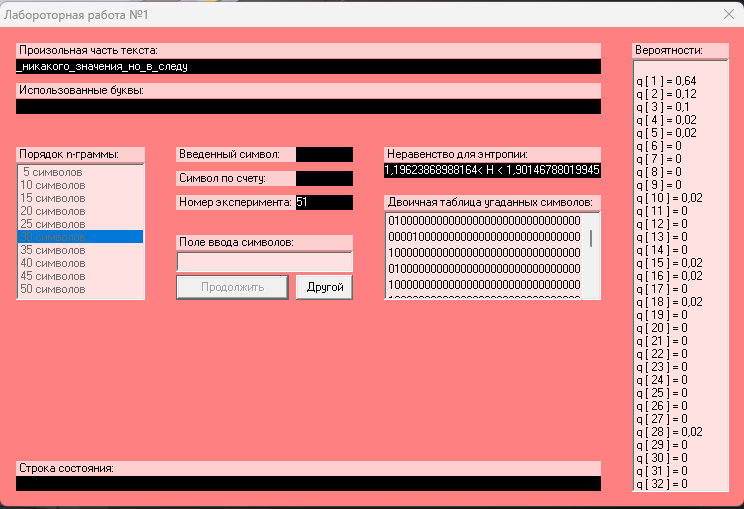
**частоти нграм з пробілами і без**  


**частоти біграм з пробілами**на скріншоті вказані дані з кінця таблиці, для перегляду повних таблиць значень зверніть увагу на **csv** файли прикріплені до лаби:  


---H1 entropy---  
H1: 4.401484367688899  
H1 without spaces: 4.460162425027035  
H1 with redundancy: 0.13483704843981703  
H1 without spaces with redundancy: 0.12330319371318177

---H2 entropy---  
H2 bigrams: 4.003559323053849  
H2 bigrams without spaces: 4.133327222431862  
H2 bigrams with redundancy: 0.2130538447196012  
H2 bigrams without spaces with redundancy: 0.1806097769859547

**CoolPinkProgram**H(10): ****  
H(20):****

H(30):

**Результати:**  
2.0136 < H(10) < 2.8376  
1.5509 < H(20) < 2.0994  
1.1962 < H(30) < 1.9014

**Оцінка надлишковості R російської мови у різних моделях відкритого тексту:**  


**Висновки:**

Протягом даної лабораторної роботи нам необхідно було навчитися працювати з великими масивами інформації , використовуючи математичні функції та формули, а саме ентропія, надлишковість тексту, порівняння різних текстів відносно ентропії та зрозуміння явища ентропії