Міністерство освіти і науки, молоді та спорту України
Національний технічний університет України
“Київський політехнічний інститут”
Фізико-Технічний інститут

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1
за семестровий курс предмету
«Симетрична криптографія»

Роботу виконав:
Студент групи ФІ-04
Бєш Радомир

Приймав:
Чорний Олег Миколайович

Київ-2023

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 H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 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) Пошук частот букв і біграм в тексті та H1 , H2

Результати:

Text without spaces

'H1': 4.385129362944809,

'H2_with_intersection': 3.9881215496418245,

'H2_without_intersection': 3.988124884466209

Text with spaces

'H1': 4.504101687378125,

'H2_with_intersection': 4.175756131511824,

'H2_without_intersection': 4.173557907692821

2) Таблиця результатів частот букв

Частоти букв			
Літера	Частота	Літера	Частота
А	'1.78e-03'	Р	'1.08e-03'
Б	'2.61e-04'	С	'2.89e-03'
В	'7.68e-03'	Т	'6.72e-03'
Г	'4.39e-04'	У	'5.00e-03'
Д	'1.15e-03'	Ф	'6.11e-04'
Е	'1.87e-02'	Х	'5.54e-03'
Ж	'1.20e-04'	Ц	'1.58e-04'
З	'1.35e-03'	Ч	'1.75e-04'
И	'1.83e-02'	Ш	'1.06e-04'
Й	'7.51e-03'	Щ	'6.87e-06'
К	'3.59e-03'	Ъ	'6.87e-06'
Л	'1.69e-03'	Ы	'4.44e-03'
М	'7.68e-03'	Ъ	'7.17e-03'
Н	'1.84e-03'	Э	'4.46e-05'
О	'1.90e-02'	Ю	'2.74e-03'
п	'3.67e-04'	Я	'9.57e-03'

3) Таблиця результатів частот біграм (найвідоміші)

Частоти біграм	
‘ст’	'1.19e-02'
‘но’	'1.16e-02'
‘ен’	'1.01e-02'
‘то’	'1.13e-02'
‘на’	'1.06e-02'
‘ов’	'8.25e-03'
‘ни’	'9.95e-03'
‘ра’	'8.94e-03'
‘во’	'7.24e-03'

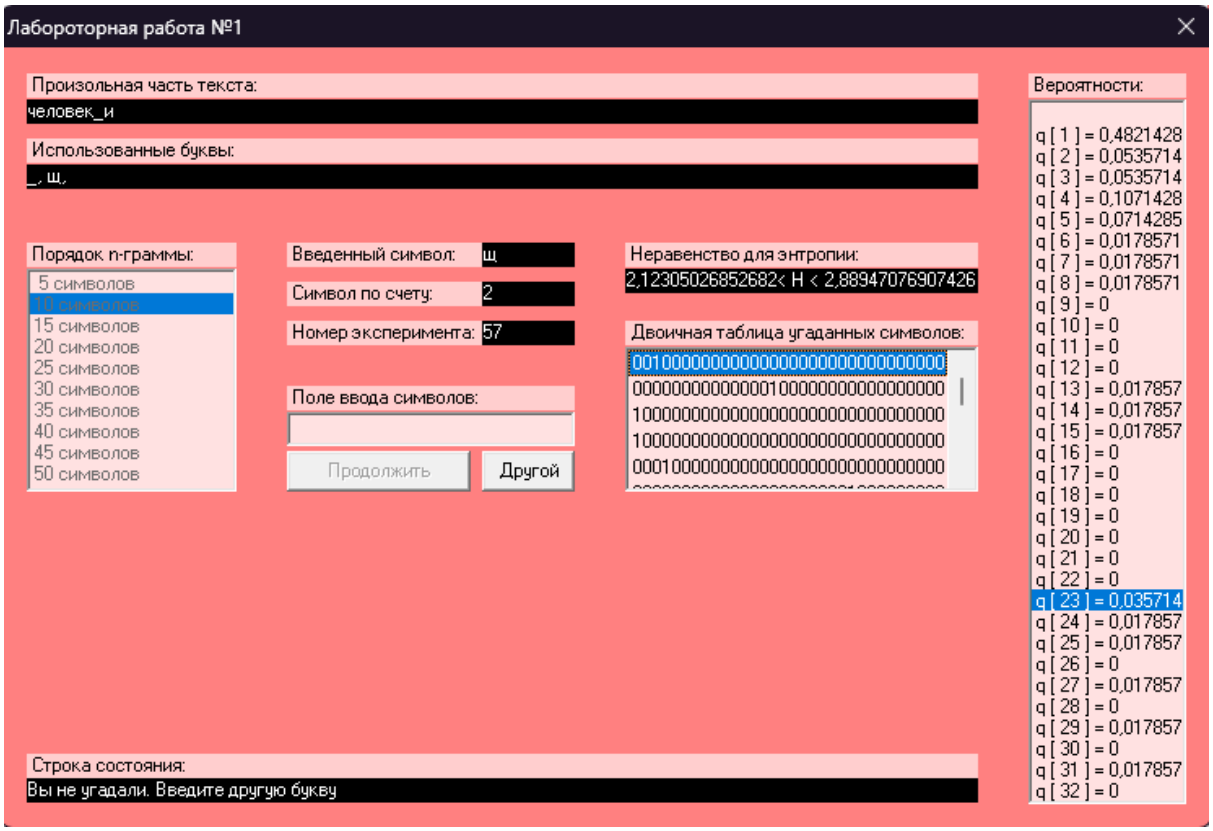
4) 2. За допомогою програми CoolPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

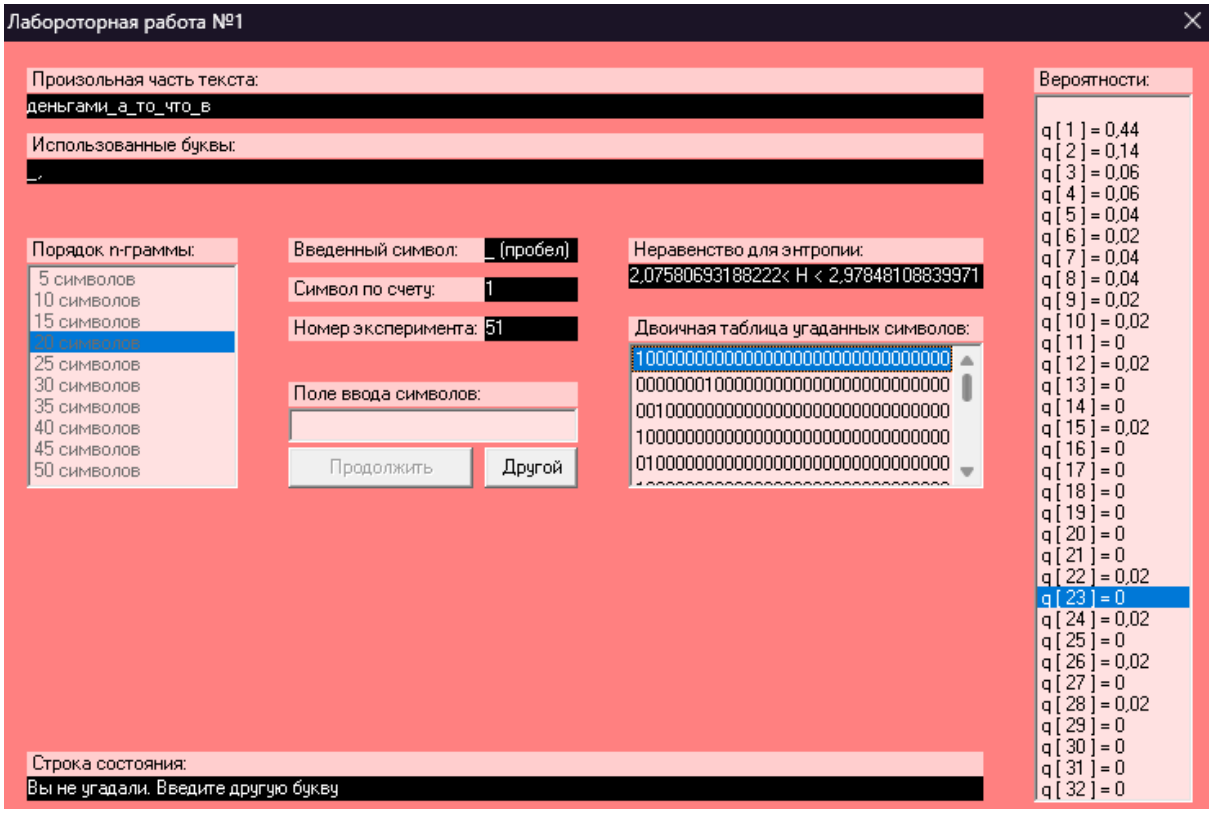
Приклад для H1:

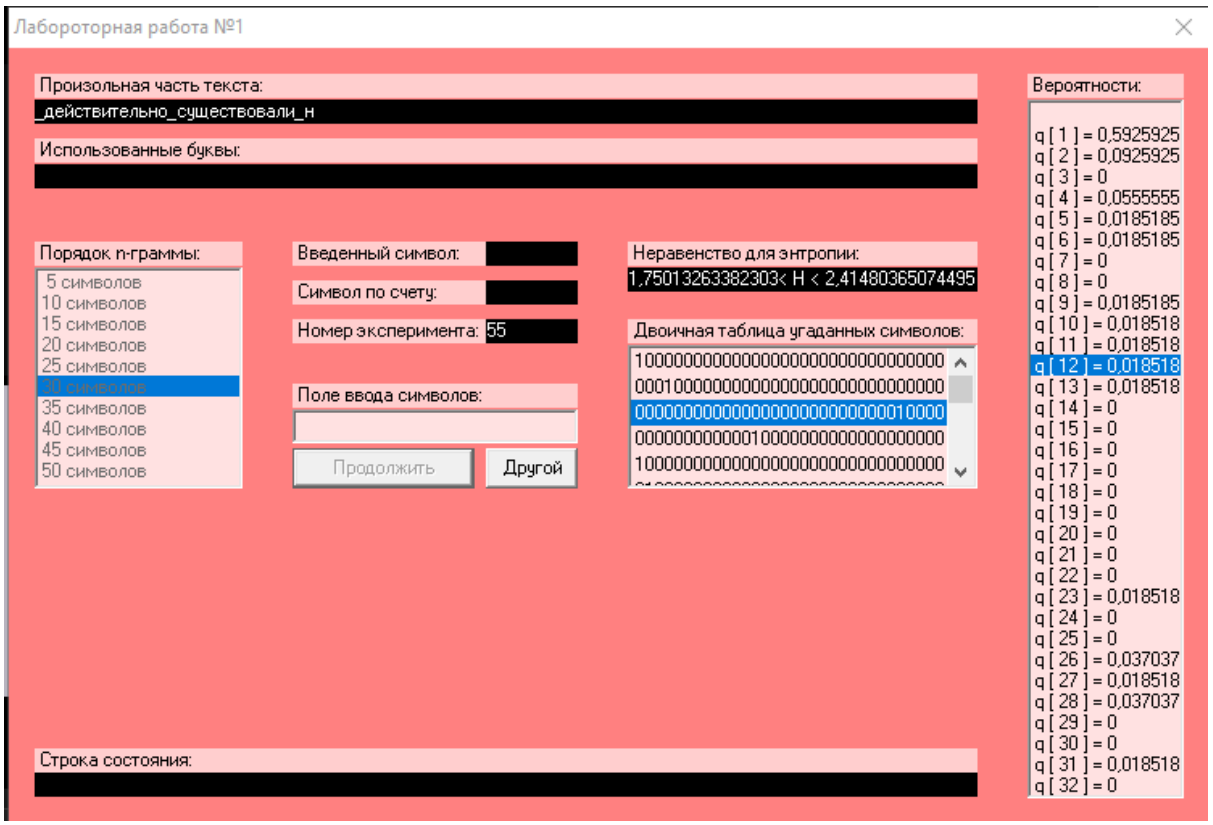
'H1': 4.385129362944809

m=32

$R=1-\frac{4.385129362944809}{\log_2 32}$=0,1229741274110382







1. $H^{(10)}$ $0{,}422106 \le R \le 0{,}57539$
2. $H^{(20)}$ $0{,}4043038 \le R \le 0{,}5848388$
3. $H^{(30)}$ $0{,}517039 \le R \le 0{,}6499736$

5) Висновки:

Під час виконання комп’ютерного практикуму ознайомився з поняттям ентропії та надлишковості. Навчився обраховувати частоту літер , біграм ,ентропію та надлишковості на прикладі обраного тексту російською мовою. За допомогою програми CoolPinkProgram знайшов межі умовної ентропії джерела та оцінил надлишковість.