СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали студенти групи ФІ-94
Костюк Кирило і Панасюк Єгор
Варінт-4

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.

Хід роботи

Написали програму для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2. Підрахували частоти букв і біграм, а також значення H1та H2 на тексті Гоголя "Мертві душі". За допомогою програми CoolPinkProgram оцінили значення H(10), H(20), H(30)

Output

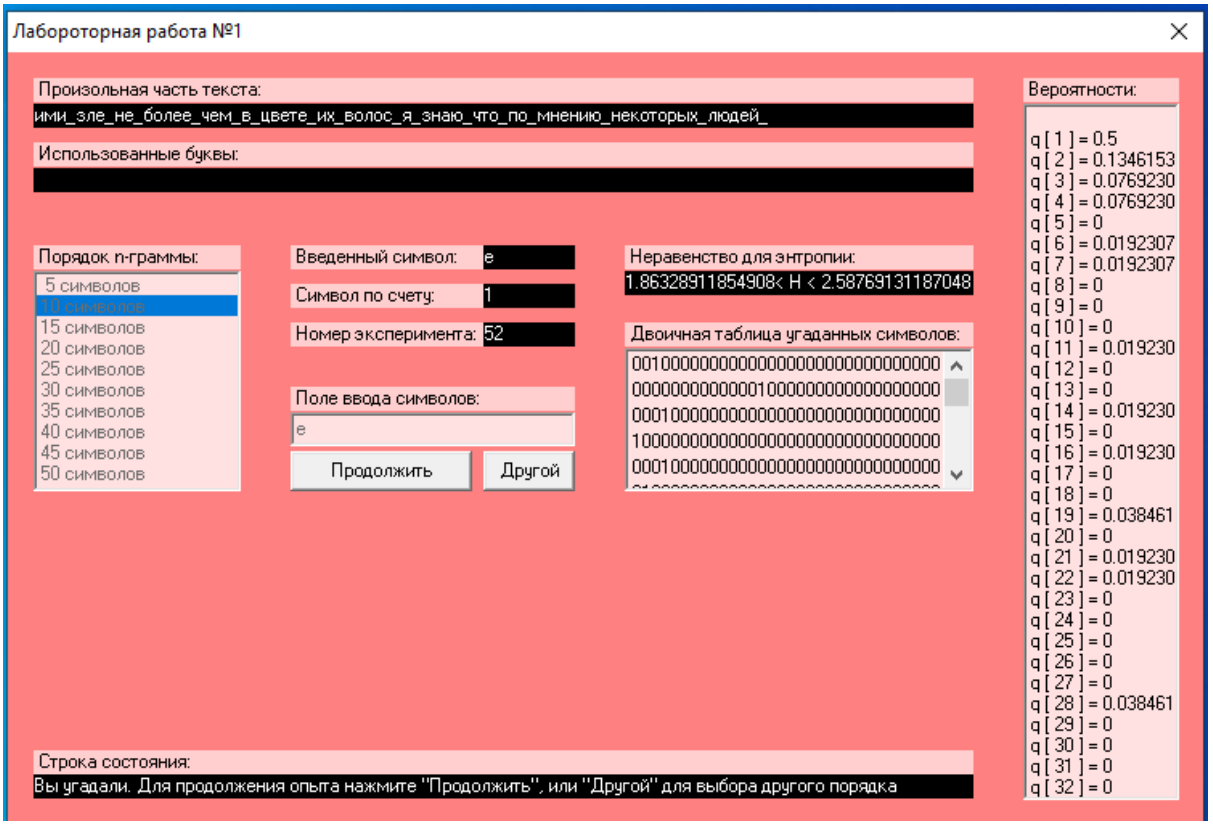
Монограми з пробілом	**Монограми без пробілу**
letter: frequency: 0.158805	letter: а frequency: 0.0767927
letter: а frequency: 0.0645977	letter: б frequency: 0.0191969
letter: б frequency: 0.0161484	letter: в frequency: 0.0467936
letter: в frequency: 0.0393625	letter: г frequency: 0.0175203
letter: г frequency: 0.014738	letter: д frequency: 0.031274
letter: д frequency: 0.0263075	letter: е frequency: 0.0855975
letter: е frequency: 0.0720042	letter: ж frequency: 0.0111428
letter: ж frequency: 0.00937329	letter: з frequency: 0.0168903
letter: з frequency: 0.0142081	letter: и frequency: 0.0683831
letter: и frequency: 0.0575235	letter: й frequency: 0.010541
letter: й frequency: 0.008867	letter: к frequency: 0.0413734
letter: к frequency: 0.0348031	letter: л frequency: 0.0447019
letter: л frequency: 0.037603	letter: м frequency: 0.0297693
letter: м frequency: 0.0250418	letter: н frequency: 0.061827

letter: н frequency: 0.0520086 letter: о frequency: 0.0948859 letter: п frequency: 0.0234339 letter: р frequency: 0.0347837 letter: с frequency: 0.0429483 letter: т frequency: 0.0518932 letter: у frequency: 0.025498 letter: ф frequency: 0.00100979 letter: х frequency: 0.00820632 letter: ц frequency: 0.00266497 letter: ч frequency: 0.0159856 letter: ш frequency: 0.00805193 letter: щ frequency: 0.0024925 letter: ы frequency: 0.0158994 letter: ь frequency: 0.0179065 letter: э frequency: 0.00189302 letter: ю frequency: 0.00505036 letter: я frequency: 0.0160051	letter: о frequency: 0.112799 letter: п frequency: 0.0278579 letter: р frequency: 0.0413503 letter: с frequency: 0.0510562 letter: т frequency: 0.0616898 letter: у frequency: 0.0303116 letter: ф frequency: 0.00120043 letter: х frequency: 0.00975555 letter: ц frequency: 0.00316807 letter: ч frequency: 0.0190035 letter: ш frequency: 0.00957201 letter: щ frequency: 0.00296304 letter: ы frequency: 0.018901 letter: ь frequency: 0.0212869 letter: э frequency: 0.00225039 letter: ю frequency: 0.0060038 letter: я frequency: 0.0190266

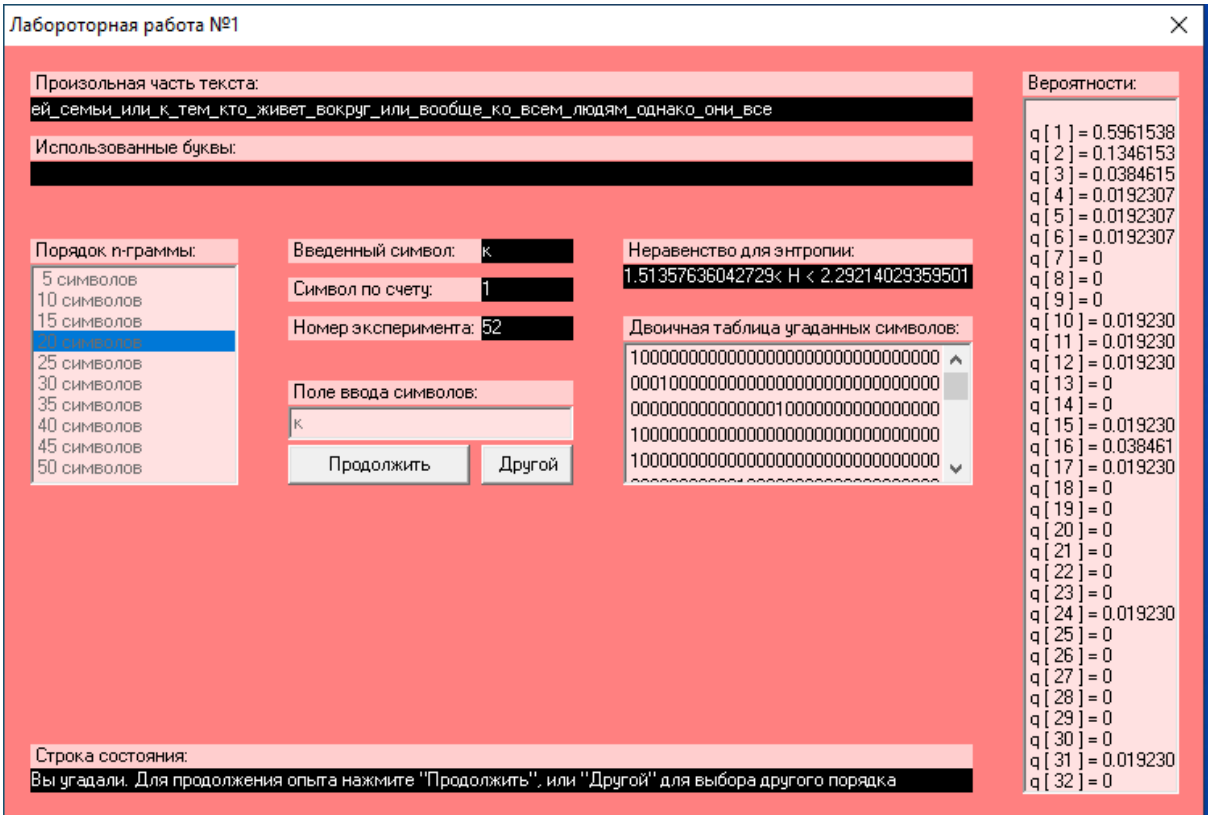
Частоти біграм можна подивитися за даним посиланням:
https://docs.google.com/document/d/1oz12Z8hmy2ekJH4YoBJ72HYR3jsBvWS58CaBcTp2UJk/edit?usp=sharing

H0=5	З пробілом	Без пробілу
H1	4.38921	4.46717
H2(перетин)	3.97137	4.13174
H2(не перетин)	3.97131	4.13155

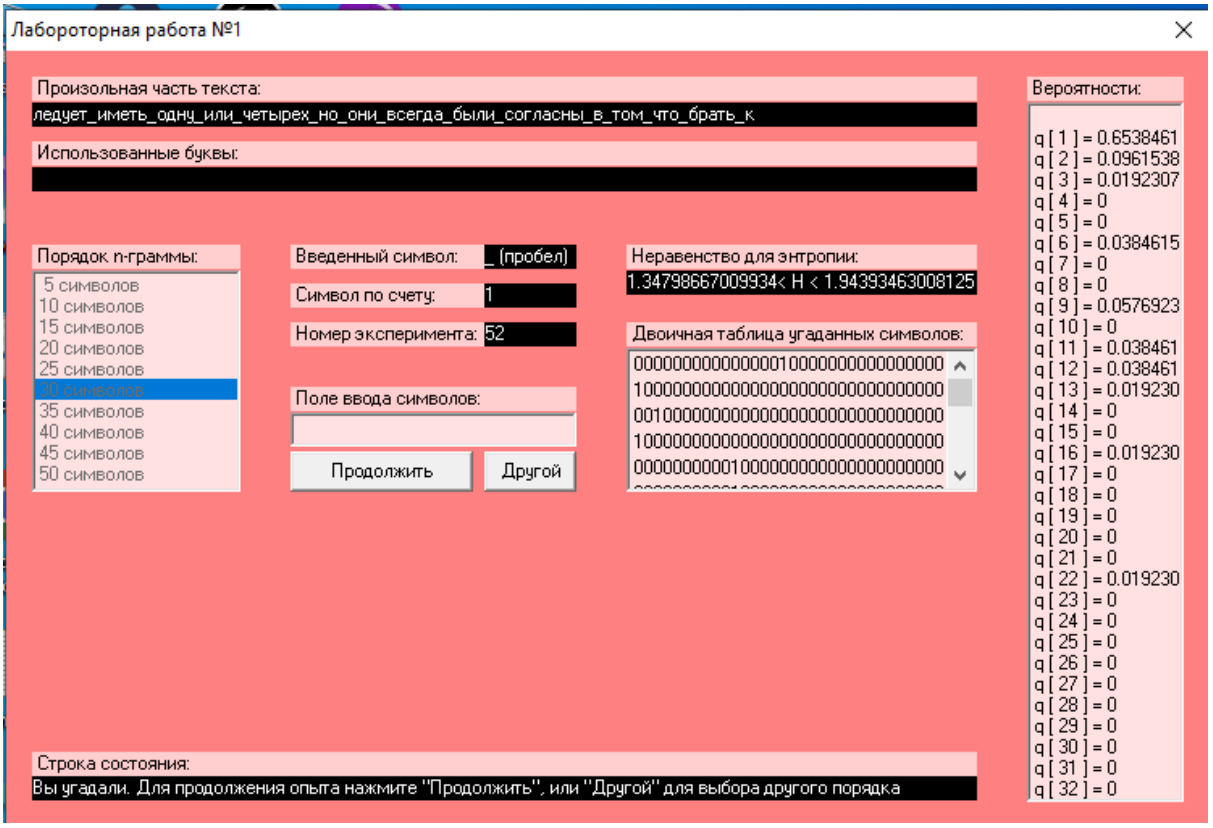
З програмки:



$1.86 < H(10) < 2.58$



$1.51 < H(20) < 2.29$



1.34 < H(30) < 2.94

Всі результати з програмки:

1.86 < H(10) < 2.58

1.51 < H(20) < 2.29

1.34 < H(30) < 2.94

Оцінка надлишковості

При H1 = 4.38921(з пробілом):

R = 1 - 0.877842 = 0.122158

При H1 = 4.467171(без пробілу):

R = 1 - 0.8934342 = 0.1065658

При H2 = 3.97137(перетин з пробілом):

R = 1 - 0.794274 = 0.205726

При H2 = 4.13174(перетин без пробілу):

R = 1 - 0.826348 = 0.173652

При H2 = 3.97131(не перетин з пробілом):

R = 1 - 0.794262 = 0.205738

При H2 = 4.13155(не перетин без пробілу):

$R = 1 - 0.82631 = 0.17369$

Нерівності з даних з програмки:
При H(10):
$0.484 < R < 0.628$
При H(20):
$0.542 < R < 0.698$
При H(30):
$0.412 < R < 0.732$

Висновок
Ми набули практичних навичок, щодо оцінки ентропії на символ джерела. Написали програмний засіб, який допомагає в цьому. Лаба весела, але наступні веселіші.