# Протокол лабораторної роботи №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав

Студент 3 курсу

Групи ФБ-13

Короткевич Іван

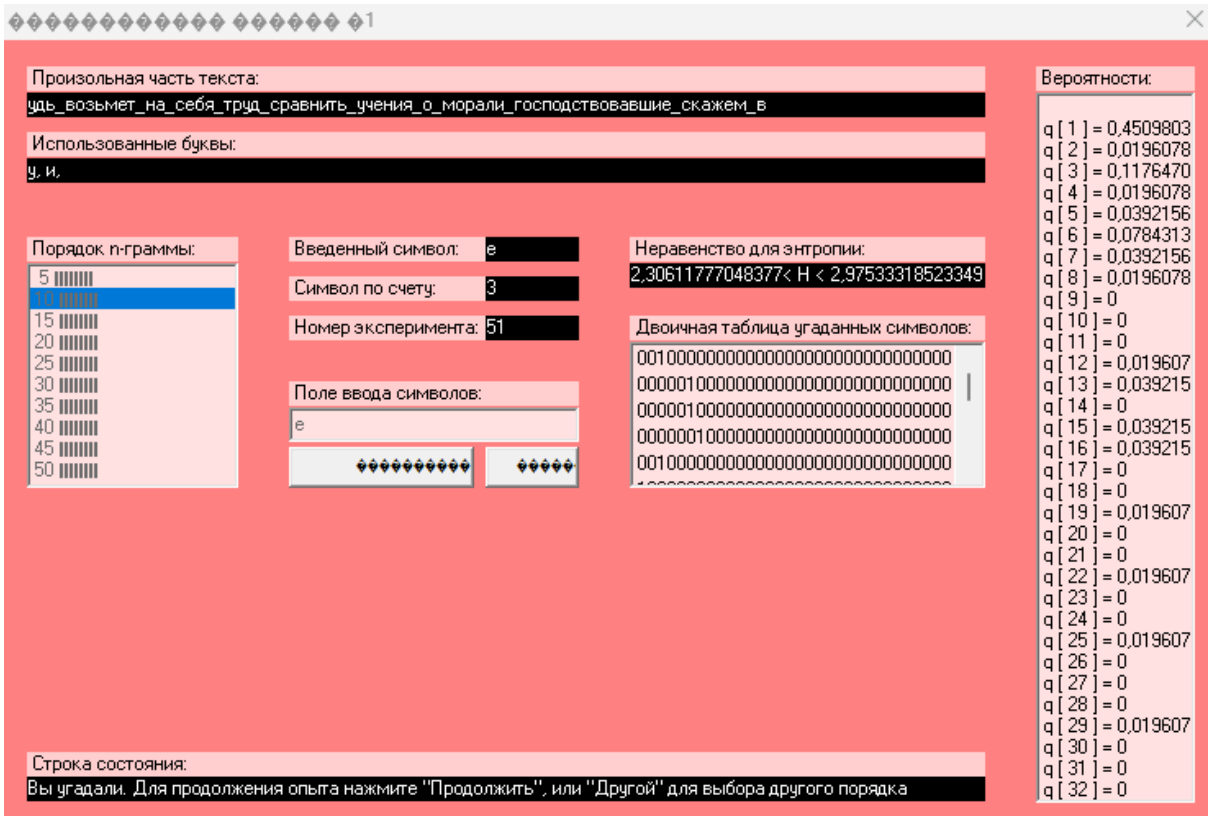
**Мета роботи**: засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

Від самого початку планувалося виконати лабораторну роботу на мові C++, але виникла проблема з кодуванням символів, тому роботу було виконанено мовою Python.

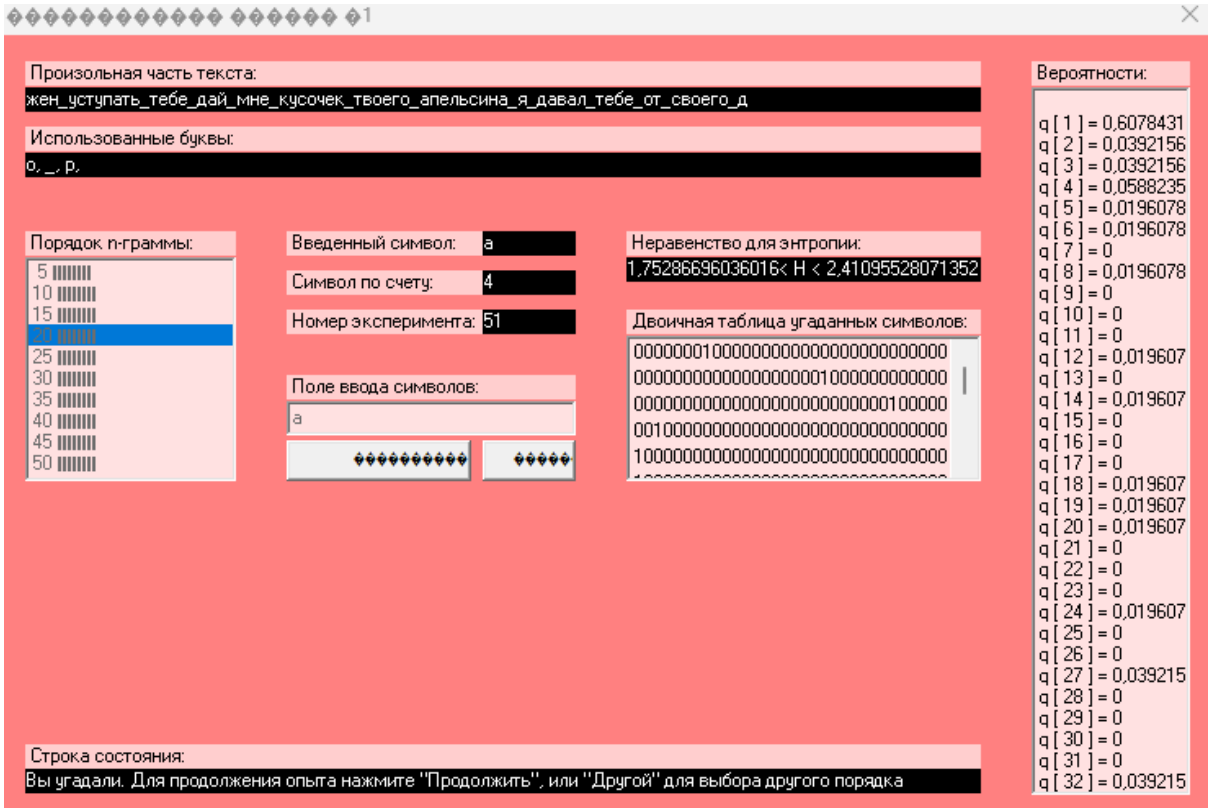
Також чомусь програма зараховує якесь порожнє місце(не пробіл) як символ(перший рядок у таблицях) і зараховує літеру б два рази.

Частота  та ентропія букв і біграм наведені у файлі __.
Значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$ були розраховані програмою CoolPinkProgram:

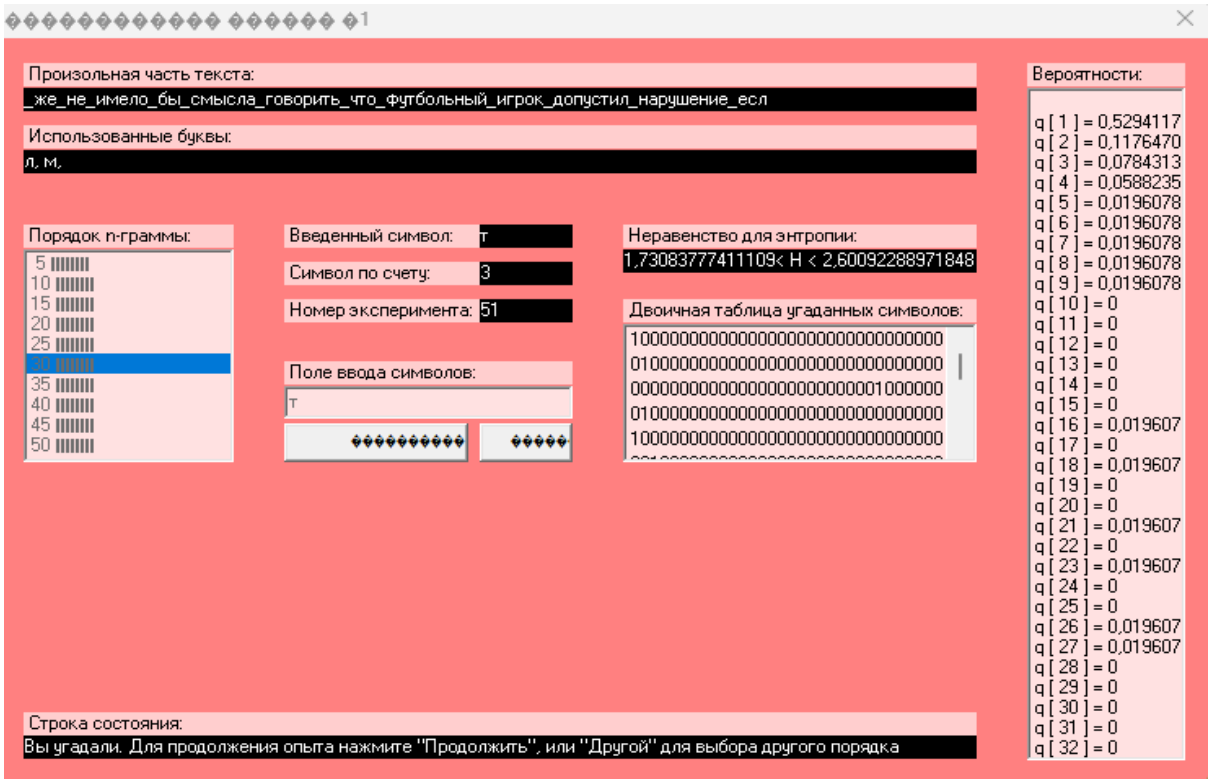
$H^{(10)}$



$H^{(20)}$


Произольная часть текста:
жен_уступать_тебе_дай_мне_кусочек_твоего_апельсина_я_давал_тебе_от_своего_д
Использованные буквы:
о, _, р,
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ: а
Символ по счету: 4
Номер эксперимента: 51
Поле ввода символов:
а
Неравенство для энтропии:
1,75286696036016< H < 2,41095528071352
Двоичная таблица угаданных символов:
Вероятности:
q [ 1 ] = 0,6078431
q [ 2 ] = 0,0392156
q [ 3 ] = 0,0392156
q [ 4 ] = 0,0588235
q [ 5 ] = 0,0196078
q [ 6 ] = 0,0196078
q [ 7 ] = 0
q [ 8 ] = 0,0196078
q [ 9 ] = 0
q [ 10 ] = 0
q [ 11 ] = 0
q [ 12 ] = 0,019607
q [ 13 ] = 0
q [ 14 ] = 0,019607
q [ 15 ] = 0
q [ 16 ] = 0
q [ 17 ] = 0
q [ 18 ] = 0,019607
q [ 19 ] = 0,019607
q [ 20 ] = 0,019607
q [ 21 ] = 0
q [ 22 ] = 0
q [ 23 ] = 0
q [ 24 ] = 0,019607
q [ 25 ] = 0
q [ 26 ] = 0
q [ 27 ] = 0,039215
q [ 28 ] = 0
q [ 29 ] = 0
q [ 30 ] = 0
q [ 31 ] = 0
q [ 32 ] = 0,039215
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$H^{(30)}$


Произольная часть текста:
_же_не_имело_бы_смысла_говорить_что_футбольный_игрок_допустил_нарушение_есл
Использованные буквы:
л, м,
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ: т
Символ по счету: 3
Номер эксперимента: 51
Поле ввода символов:
т
Неравенство для энтропии:
1,73083777411109< H < 2,60092288971848
Двоичная таблица угаданных символов:
Вероятности:
q [ 1 ] = 0,5294117
q [ 2 ] = 0,1176470
q [ 3 ] = 0,0784313
q [ 4 ] = 0,0588235
q [ 5 ] = 0,0196078
q [ 6 ] = 0,0196078
q [ 7 ] = 0,0196078
q [ 8 ] = 0,0196078
q [ 9 ] = 0,0196078
q [ 10 ] = 0
q [ 11 ] = 0
q [ 12 ] = 0
q [ 13 ] = 0
q [ 14 ] = 0
q [ 15 ] = 0
q [ 16 ] = 0,019607
q [ 17 ] = 0
q [ 18 ] = 0,019607
q [ 19 ] = 0
q [ 20 ] = 0
q [ 21 ] = 0,019607
q [ 22 ] = 0
q [ 23 ] = 0,019607
q [ 24 ] = 0
q [ 25 ] = 0
q [ 26 ] = 0,019607
q [ 27 ] = 0,019607
q [ 28 ] = 0
q [ 29 ] = 0
q [ 30 ] = 0
q [ 31 ] = 0
q [ 32 ] = 0
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

Надлишковість джерела відкритого тексту:

Для букв з пробілом:

$$R_1 = 1 - \frac{H_1}{H_0} = 1 - \frac{4.4178519973618835}{5.087} =$$
$$= 0{,}131540790768255651661096913701 59$$

Для біграм з пробілом:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.030182621075429}{10.175} =$$
$$= 0{,}603913255913962751842751842751 84$$

Для букв без пробіла:

$$R_1 = 1 - \frac{H_1}{H_0} = 1 - \frac{4.4882552037331696}{5.044} =$$
$$= 0{,}110179380703178112609040444091 99$$

Для біграм без пробіла:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.184205702319153}{10.089} =$$
$$= 0{,}585270522121205966894637724254 14$$

З наведених вище даних можна зробити висновок, що при збільшені n, R буде збільшуватись.