## 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.

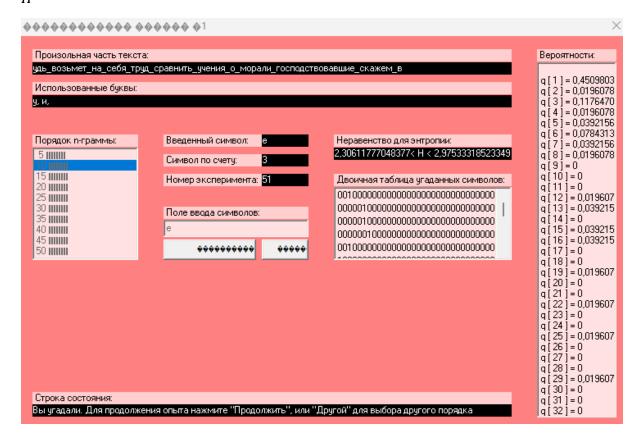
Від самого початку планувалося виконати лабораторну роботу на мові C++, але виникла проблема з кодуванням символів, тому роботу було виконанено мовою Python.

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

Ентропія літер з пробілом: 4.417851 Ентропія літер без пробіла: 4.488255 Ентропія біграм з пробілом: 4.030182 Ентропія біграм без пробіла: 4.184205

Ентропія біграм(не перетинаються) з пробілом: 4.029912 Ентропія біграм(не перетинаються) без пробіла: 4.18347

Частота та ентропія всіх букв і біграм наведені у файлі lab1\_table. Значення  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$  були розраховані програмою CoolPinkProgram:  $H^{(10)}$ 





## $H^{(30)}$



 $H^{(10)}$ : 2,03611777048377 < H < 2,97533318523349

 $H^{(20)}$ : 1,75286696036016 < H < 2,41095528071352

 $H^{(30)}$ : 1,73083777411109 < H < 2,60092288971848

Надлишковість джерела відкритого тексту:

 $H_0 = log_2$ 34 (для алфавіту з пробілом)

 $H_0 = log_2$ 33 (для алфавіту без пробіла)

Для букв з пробілом:

$$R_1 = 1 - \frac{H_1}{H_0} = 1 - \frac{4.4178519973618835}{5.087} = 0.13154$$

Для біграм з пробілом:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.030182621075429}{5.087} = 0,20774$$

Для букв без пробіла:

$$R_1 = 1 - \frac{H_1}{H_0} = 1 - \frac{4.4882552037331696}{5.044} == 0,11018$$

Для біграм без пробіла:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.184205702319153}{5.044} = 0,17045$$

Для біграм(не перетинаються) з пробілом:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.0299126670791425}{5.087} = 0,20780$$

Для біграм(не перетинаються) без пробіла:

$$R_2 = 1 - \frac{H_2}{H_0} = 1 - \frac{4.1834700614518585}{5.044} = 0,17060$$

$$H^{(10)}: 1 - \frac{2,03611777048377}{5} < R < 1 - \frac{2,97533318523349}{5}$$
$$0,59277 > R > 0,40493$$

$$H^{(20)}: 1 - \frac{1,752866960360167}{5} < R < 1 - \frac{2,41095528071352}{5}$$
$$0,64942 > R > 0,51780$$

$$H^{(30)}: 1 - \frac{1,73083777411109}{5} < R < 1 - \frac{2,60092288971848}{5}$$
$$0,65383 > R > 0,47981$$

## Висновок

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