### МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ 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

Виконав: ФБ-14 Фролов Павло

> Перевірила: Селюх П. В.

#### 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:

- 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:

Використовував текст "Мертві душі". За допомогою функцій text\_format, text\_format\_spaceless форматую текст, створюю словники з частотами за допомогою char\_freq та bigram\_freq. Ентропія та надлишковість отримується у функції get\_entropy\_r.

Таблиці частот букв і біграм: (повні в output.txt)

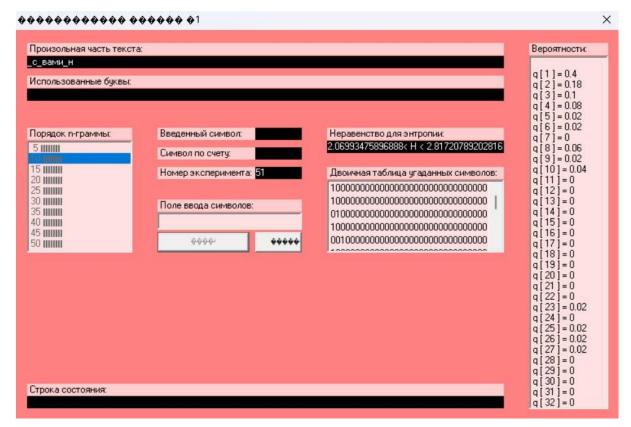
```
Частота літер/біграм в тексті:
m: 2.4853%
e: 7.1373%
p: 3.4690%
т: 5.1375%
в: 3.9196%
ы: 1.5864%
: 16.3038%
д: 2.6153%
y: 2.5323%
ш: 0.8026%
и: 5.7357%
н: 5.1806%
к: 3.4570%
a: 6.4257%
л: 3.7493%
```

```
Частота літер/біграм в тексті:
ме: 0.3759%
ep: 0.5390%
рт: 0.0798%
тв: 0.2328%
вы: 0.2644%
ые: 0.1255%
e: 1.8357%
д: 0.8372%
ду: 0.1968%
уш: 0.0911%
ши: 0.2243%
и: 1.8372%
 : 0.3201%
 y: 0.4137%
ун: 0.0197%
ни: 0.7267%
```

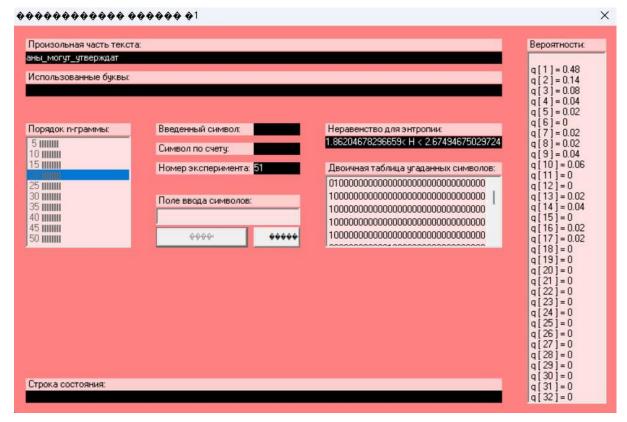
#### Оцінки Н1, Н2 та надлишковість:

```
Ентропія і надлишковість Н1: 4.471864048360621 0.11349828293564157
Ентропія і надлишковість Н1 з пробілами: 4.384307832031327 0.13821329632477442
Ентропія і надлишковість Н2: 4.135533940513343 0.18017231749542584
Ентропія і надлишковість Н2 з пробілами: 3.97206315944395 0.21924478204783482
Ентропія і надлишковість Н2 без перетинів: 4.135608894901897 0.18015745854769494
Ентропія і надлишковість Н2 з пробілами без перетинів і : 3.9707420433767355 0.21950446277837554
```

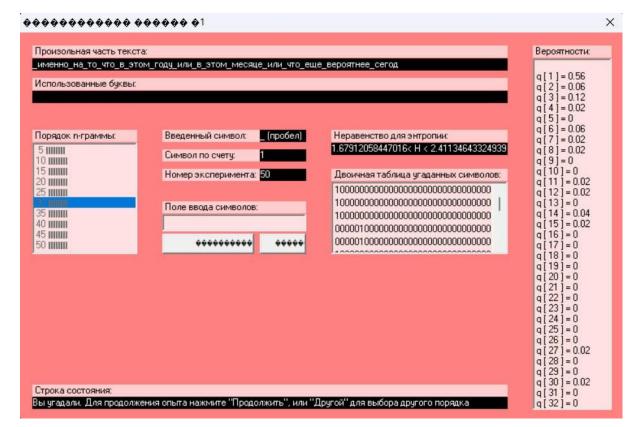
H10, H20, H30 y CoolPinkProgram



2.0699347 < H10 < 2.8172078 0.43655844 < R < 0.58601306



1.8620467 < H20 < 2.6749467 0.46501066 < R < 0.62759066



1.6791205 < H30 < 2.4113464 0.5177307 < R < 0.6641759

#### Висновки:

Виконавши комп'ютерний практикум я отримав практичні знання у аналізі частот монограм та біграм. Використовуючи Python реалізував формули ентропії та надлишковості. Отримані дані використав при роботі з CoolPinkProgram.