МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО" ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

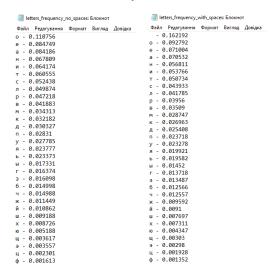
Виконали:

Студенти групи ФБ-03 Митрофанова М.М. та Мец Є.В. **Тема:** «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Мета: 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.

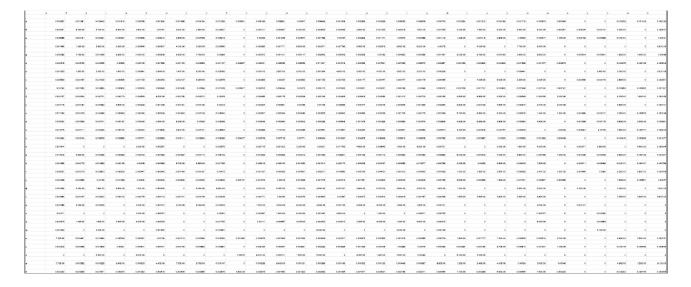
ХІД РОБОТИ

- 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.
 - Частоти букв



• Частоти біграм

1) bigrams_no_spaces



2) bigrams_no_spaces_crossing ***** 610M 3) bigrams_spaces_crossing 2,000 1,000 4 4000 42000 1,000 1,000 1,000 1,000 1,000 1,000 1,000 4) bigrams_spaces_no_crossing Annual Community of the 10000 10000 10000 10000 10000 10000 10000 10000 10000 10000 .

Посилання на папку з табличками:

https://drive.google.com/drive/folders/1pEcjDZcUFTM9kulQN3bzGm1Kw1_mdcdh?usp=sharing

• Ентропія

```
Letters with whitespaces entropy: 4.371059 ; redundancy: 0.125788

Letters without whitespaces entropy: 4.453914 ; redundancy: 0.100982

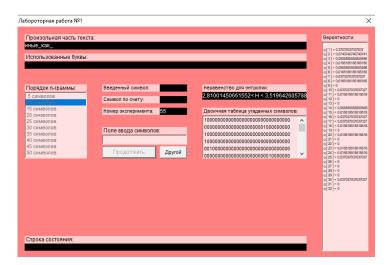
Bigrams with whitespaces with crossing entropy: 3.9558 ; redundancy: 0.20884

Bigrams with whitespaces without crossing entropy: 3.951953 ; redundancy: 0.209609

Bigrams without whitespaces with crossing entropy: 4.126059 ; redundancy: 0.167159

Bigrams without whitespaces without crossing entropy: 4.12453 ; redundancy: 0.167467
```

2. За допомогою програми CoolPinkProgram оцінити значення H^{10} , H^{20} , H^{30} . H^{10} :

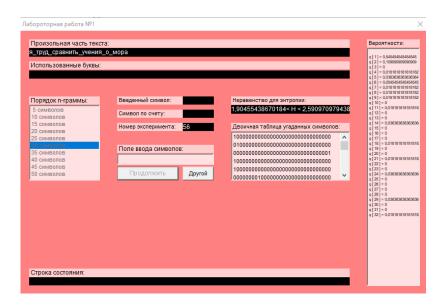


 $2,8100 < H^{10} < 3,5196$

 H^{20} :



 H^{30} :



 $1,90455 < H^{20} < 2,59097$

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

$$R = 1 - \frac{H_{\infty}}{H_0}$$

 H^{10} .

1)
$$R = 1 - \frac{2,81001450661552}{5} \approx 0,4379$$

2)
$$R = 1 - \frac{3,519642605788}{5} \approx 0,2960$$

 H^{20} :

3)
$$R = 1 - \frac{1,83390965865442}{5} \approx 0,6332$$

4)
$$R = 1 - \frac{2,512482675750}{5} \approx 0,4975$$

 H^{30} :

5)
$$R = 1 - \frac{1,90455438670184}{5} \approx 0,62$$

6)
$$R = 1 - \frac{2,590970979438}{5} \approx 0,482$$

Труднощі:

• Вивід матриці біграм

ВИСНОВКИ

Під час виконання комп'ютерного практикуму ми ознайомились з поняттям ентропії на символ джерела та його надлишковості, написали програми для підрахунку частот букв і частот біграм в тексті, попрацювали з CoolPinkProgram, де оцінювали та порівнювали H^{10} , H^{20} , H^{30} . В результаті експериментів ми отримали значення ентропії та надлишковості власного джерела, які наближаються до значень, отриманих з прикладеної програми.