

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ  
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ СІКОРСЬКОГО» ФІЗИКО-  
ТЕХНІЧНИЙ ІНСТИТУТ

**Комп'ютерний практикум 1 Експериментальна оцінка ентропії на  
символ джерела відкритого тексту**

Виконали:

Толмачов Євгеній

Котович Анна

Перевірів: Деркач О.Г

## 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

## Постановка задачі

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

Під час виконання лабораторної роботи особливих проблем не виникло. Єдиними труднощами були ознайомлення з модулями для роботи з файлами (що не було чимось важким саме по собі), уникання повторених поспіль пробілів та підрахунок біграмм що не перетинаються

## Результати

1. Код та таблиці біграмм наведені у додатку до листа та за посиланням [https://github.com/ZheZheDoshka/labs\\_2021\\_5/tree/main/%D0%BA%D1%80%D0%B8%D0%BF%D1%82%D0%B01](https://github.com/ZheZheDoshka/labs_2021_5/tree/main/%D0%BA%D1%80%D0%B8%D0%BF%D1%82%D0%B01)





Ентропія монограм для тексту без пробілів: 4.44334

Ентропія монограм для тексту з пробілами: 4.35151

Ентропія біграм для тексту без пробілів, біграми перетинаються: 4.12613

Ентропія біграм для тексту без пробілів, біграми не перетинаються:  
4.12688

Ентропія біграм для тексту з пробілами, біграми перетинаються: 3.94651

Ентропія біграм для тексту з пробілами, біграми не перетинаються:  
3.94651.

### **Оцінки надлишковості**

Надлишковість російської мови:  $R = 1 - H_{\infty} / H_0 = 0.4337$

### **Висновок:**

В ході цієї лабораторної було підраховано частоти букв та біграм, а також оцінено відповідні значення ентропій та надлишковість джерела. За допомогою програми CoolPinkProgram практичним шляхом оцінили значення ентропій  $H_{(10)}$ ,  $H_{(20)}$ , та  $H_{(30)}$