**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**ФАКУЛЬТЕТ РАДІОФІЗИКИ, ЕЛЕКТРОНІКИ ТА КОМП’ЮТЕРНИХ СИСТЕМ**

**КАФЕДРА КОМП’ЮТЕРНОЇ ІНЖЕНЕРІЇ**

Лабораторна робота №4

З дисципліни «Захист інформації у комп’ютерних системах»

«Дослідження частотних характеристик української мови»

Виконав   
Юрченко П. А.   
Студент 4 курсу СА

Київ 2024

**Мета**: дослідити вірогіднісні параметри появи літер української мови для різних типів текстів. Аналіз найбільш імовірних літер, біграм та триграм для використання в частотному криптоаналізі.

**Хід роботи**

## 1. Знаходження відносної частоти появи літер українського алфавіту

1. Створіть програму (будь-якою зручною для вас мовою), яка в якості аргументів приймає перелік текстових файлів, та аналізуючи їх вміст обраховує частоти появи літер українського алфавіту. Для спрощення аналізу з тексту виключаються всі знаки пунктуації окрім “пробілу”.
2. Для різних достатньо довгих текстів (оберіть принаймні 2, причому тексти не повинні повторюватись в групі), наприклад, художній твір, вірш, технічна документація, блог, тощо, проаналізуйте отримані результати та в звіті наведіть:
   1. діаграму відсортовану в алфавітному порядку
   2. діаграму відсортовану по частотам появи літер
   3. послідовність літер по мірі спадання частоти появи

## 2. Знаходження відносної частоти появи біграм українського алфавіту

1. Аналогічно до першого завдання, створіть програму яка підрахує відносні частоти появи біграм української мови аналізуючи вміст текстів, що задаються.
2. Для текстів, що були використані в першому завданні проаналізуйте отримані результаті та в звіті наведіть:
   1. таблицю з відносними частотами біграм, відсортовану за спаданням частоти (використовуйте декілька колонок, щоб результат був максимально компактним)
   2. послідовність з 30-ти найбільш імовірних біграм
   3. діаграму відсортовану по частотам появи 30-ти найбільш імовірних біграм
   4. матрицю частот появи біграм (імовірність позначена кольором)

## 3. Знаходження відносної частоти появи триграм українського алфавіту

1. Аналогічно до першого завдання, створіть програму яка підрахує відносні частоти появи триграм української мови аналізуючи вміст текстів, що задаються.
2. Для тексті, що були використані в першому завданні проаналізуйте отримані результаті та в звіті наведіть:
   1. таблицю з відносними частотами триграм, відсортовану за спаданням частоти
   2. послідовність з 30-ти найбільш імовірних триграм
   3. діаграму відсортовану по частотам появи 30-ти найбільш імовірних триграм

## 4. Криптоаналіз

1. Зашифруйте за допомогою афінного шифру (a та b оберіть випадковим чином) 3 фрагменти тексту та надайте отримані шифротексти колегам.
2. Отримайте 3 шифротексти від колег, виконайте їх криптоаналіз та проілюструйте в звіті процес його проведення за допомогою частотного аналізу.