Homework 8

Caesar's Cipher

Caesar's cipher is one of the oldest and simplest encryption methods. It works by replacing each letter of the original text with another letter located a fixed number of positions further down the alphabet. This fixed number is known as a **shift** or **key**.

Mathematically, the transformation can be expressed as:

$$C_i = (P_i + k) \mod 26$$

where:

- C_i: cipher letter;
- P_i: unencrypted letter;
- k: shift (key);
- $mod(26)$: ensures that the result remains within the 26 letters of the alphabet.

In the code, when the user selects Caesar's cipher, a random shift between 1 and 26 is generated. The entered text is encrypted by applying this transformation to each letter, rendering the message unintelligible without knowing the shift used.

RSA Encryption

RSA encryption is an asymmetric encryption algorithm that uses a pair of keys: a **public key** to encrypt data and a **private key** to decrypt it. The security of RSA is based on the difficulty of factoring large prime numbers.

The fundamental steps of **RSA** are:

1. **Key generation**:
    - Choose two large prime numbers, p and q.
    - Calculate $n = p \times q$ and $\phi(n) = (p-1)(q-1)$.
    - Choose a number e such that $1 < e < \phi(n)$ and that is coprime with $\phi(n)$, i.e. $\gcd(e, \phi(n)) = 1$.
    - Calculate the multiplicative inverse modulo e modulo $\phi(n)$, obtaining d such that $e \times d \equiv 1 \mod \phi(n)$.
2. **Crypting**:
    - Convert the message to a number m such that $0 \leq m < n$.
    - Calculate the ciphertext c using:

$$c = m^e \mod n$$

In the code, default values are used for p, q, e and d. The text entered by the user is encrypted character by character using the RSA algorithm, transforming each character into a number and applying the encryption function.

Frequency Analysis

A key component of the code is **frequency analysis**, a cryptanalysis technique used to decipher substitution ciphers.

- **Letter Frequency**: The percentage occurrence of each letter in the cipher text is calculated. These frequencies are compared with typical letter frequencies in English and Italian to help identify the language of the text.
- **Coincidence Index (CI)**: This is a statistical measure that indicates the probability that two letters chosen at random from the text are the same. It is calculated as:

  $$IC = \frac{\sum_{i=1}^{n} f_i(f_i - 1)}{N(N-1)}$$

  where:
  - f_i: frequency of the i-th letter in the text;
  - N: total number of letters in the text.
- **Chi Square Test**: Used to compare the frequencies of letters observed in the cipher text with those expected in a specific language. The statistic is calculated as:

  $$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

  where:
  - O_i: observed frequency of the i-th letter;
  - E_i: expected frequency of the i-th letter in the language under consideration.
- **Bigrams and Trigrams**: The frequency of letter pairs (bigrams) and letter triplets (trigrams) in the text is analysed and compared with known frequencies in English and Italian languages. This helps to refine language identification and improve decipherment.

Automatic Decryption

To decrypt a text ciphered with Caesar's cipher without knowing the shift, the code implements a brute force attack:

- **Exploration of All Possible Shifts**: The code tries all shifts from 0 to 25.
- **Score Calculation**: For each shift, it calculates the coincidence index, chi square and scores based on bigrams and trigrams.
- **Best Shift Selection**: The shift with the best total score (i.e. the lowest chi square) is selected as the most likely shift.
- **Text Decryption**: Once the shift has been identified, the text is decrypted by applying the inverse shift.

This method exploits the statistical properties of natural languages to recover the original text without knowing the encryption key.

Graphical Visualisation

The code uses graphics to help the user better understand the encryption and decryption process:

- **Letter Frequency Graph**: Shows the distribution of letters in the cipher text, facilitating visual analysis of frequencies.
- **Language Frequency Comparison**: A second graph shows typical letter frequencies in the identified language (English or Italian), allowing direct comparison with the cipher text.