Specifica del Progetto di Laboratorio
Architettura degli Elaboratori II

Lavinia Fabrello*
matricola: 907511, Turno: A
lavinia.fabrello@studenti.unimi.it

1 Modello Linguistico

Nella crittoanalisi, l'analisi delle frequenze è lo studio della frequenza di utilizzo delle lettere o gruppi di lettere in un testo cifrato. Questo metodo è utilizzato per violare i cifrari classici. Le indagini quantitative sui testi si servono spesso di qualche forma di analisi delle frequenze.

Questo progetto si propone di utilizzare l'analisi delle frequenze per individuare, attraverso un confronto con un opportuno database, la lingua in cui è scritto un testo di lingua ignota.

2 Struttura del programma

All'avvio dell'esecuzione, vengono proposte all'utente 3 scelte:

1. L'acquisizione del database;
2. Il calcolo delle statistiche fino a quel momento acquisite;
3. La valutazione del testo di lingua ignota.

2.1 Acquisizione del database

1. Viene richiesto il path del file da utilizzare;
2. Viene richiesta la lingua a cui il file fa riferimento. Opzioni di scelta:
    - Inglese;
    - Italiano;
    - Francese;
    - Lingua sconosciuta;

*Progetto approvato il 20/05/2018, consegnato il 24/05/2018

2.2 Creazione delle statistiche

Vengono calcolate e stampate le frequenze di comparizione di ciascun carattere nelle varie lingue e nella lingua ignota. Per ogni lingua, se sono presenti dati nel database, vengono stampate le seguenti informazioni:

1. Numero di caratteri in memoria;
2. Numero di ricorrenze per ciascun carattere;
3. Frequenza di ricorrenza di ciascun carattere.

2.3 Analisi

Valutazione del testo di lingua ignota. Viene calcolata la somma dei quadrati delle differenze delle frequenze di ogni lingua e del testo di lingua ignota. Viene stimata come lingua più probabile quella che presenta un valore inferiore.

3 Flusso del programma

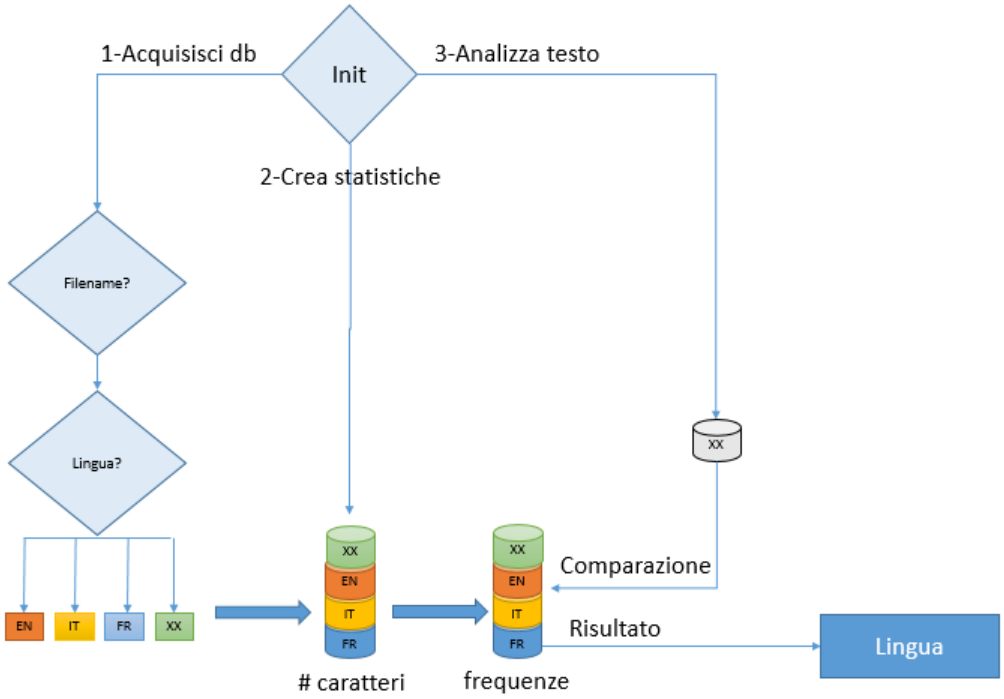

Figura 1: Flusso del programma