

# UNICODE

leo

March 5, 2023

Un carattere UNICODE viene identificato dal suo codice numerico detto **code point** solitamente rappresentato con 8 cifre esadecimali.

Ogni carattere occupa 32bit di conseguenza un file di testo occupa 4 volte lo spazio rispetto ad una codifica ASCII. Per ovviare a questo problema UNICODE definisce vari formati di codifica più compatti.

## 1 UTF-8 | 8bit Unicode Transformation Format

È una codifica a lunghezza variabile tra 8bit e 32bit.

*Non ho capito come funziona.*