Повідомлення стає числом

Мета

Дослідити, як текст стає числами й байтами; порівняти ASCII, Unicode та UTF-8.

Біт і байт

Біт має значення 0 або 1. Вісім бітів утворюють байт. Байт має 256 можливих значень.

Системи числення

Двійкова: 01000001. Десяткова: 65. Шістнадцяткова: 41. Це три записи одного числа.

ASCII та Unicode

Латинська M має ASCII-код 77 (01001101). Українські літери М, Я, У не належать ASCII. Їхні десяткові коди Unicode: 1052, 1071, 1059.

UTF-8

Українське «МЯУ» у UTF-8: D0 9C D0 AF D0 A3. Кожна літера займає два байти.

Практика 1

Складіть число 65 із ваг бітів 128, 64, 32, 16, 8, 4, 2, 1.

Практика 2

Порівняйте байти UTF-8 для латинської M, української М та emoji кота.

Практика 3

Відновіть текст із HEX-послідовності 48 69 21.

Мініпроєкт

Створіть коротке повідомлення, перетворіть на HEX, передайте другу та попросіть відновити текст.

Висновок

Кодування дозволяє зберігати й передавати символи. Кодування не є шифруванням.