

Урок 2. Двійкове кодування

Вивчення нового матеріалу

Слайд № 1

Подання чисел у двійковій системі називають також **двійковим кодом**, оскільки, як ви вже знаєте, числами кодується різна інформація.

0 1 0 0 1 0 1 1 1 0 0 1

1 - наявність електричного сигналу

0 - відсутність

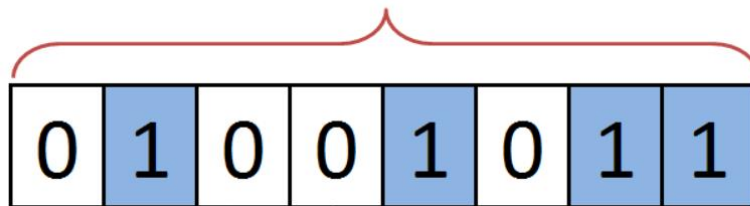
Ось приклад двійкового коду:

010010111001

Слайд № 2

Кожна цифра у двійковому коді, називається **бітом**, а послідовність із восьми бітів – **байтом**.

1 байт=8 бітів

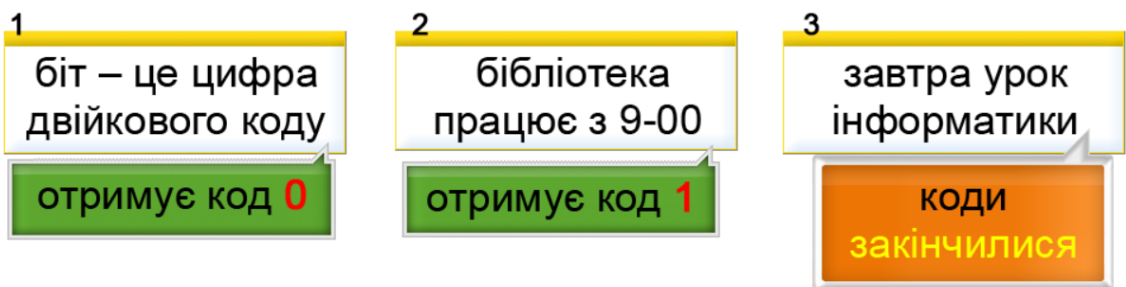


1 біт

1 **біт** – це цифра
(0 або 1)

1 **байт** = 8 бітів

Спробуємо присвоїти різним повідомленням два коди **0** та **1**, які можна отримати з 1 біта.



Отже, два повідомлення кодуються двома значеннями біта, а третє повідомлення вже нема чим закодувати.

Розглянемо, яким чином утворюються двійкові коди різної довжини

Щоб утворити з них двобітні повідомлення, спочатку дописуємо зліва нулі, а потім - одиниці

0	0
0	1
1	0
1	1

Розглянемо, яким чином утворюються двійкові коди різної довжини

Таким чином отримуємо чотири двобітні повідомлення

1	0	0
2	0	1
3	1	0
4	1	1

Утворення трьохбітних повідомлень

3 біти

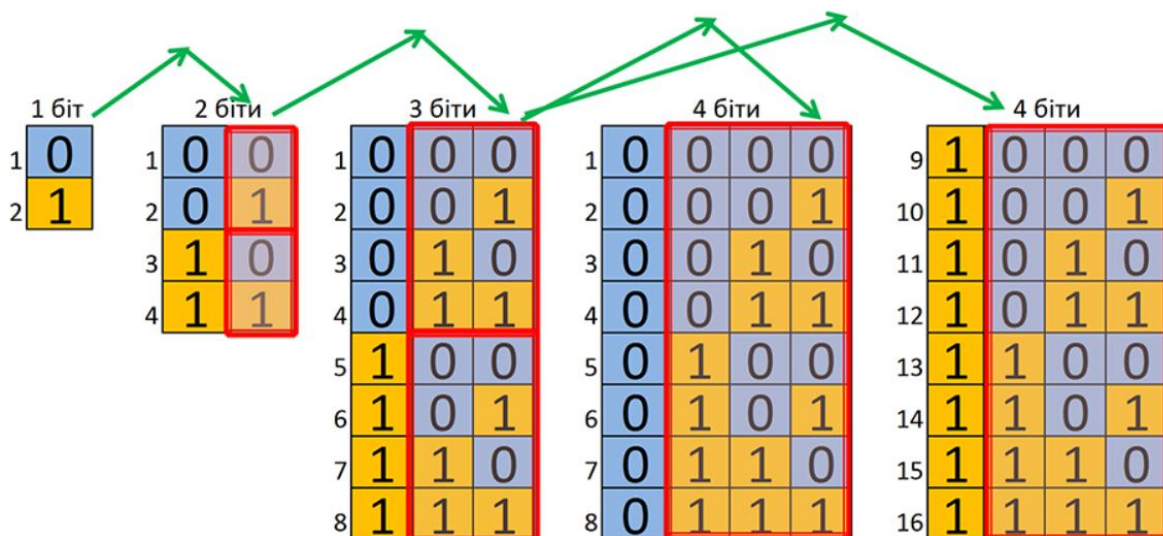
2 біти

1	0	0
2	0	1
3	1	0
4	1	1

1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Трьохбітні повідомлення утворюються аналогічно: до двобітних спочатку дописуємо зліва нулі, потім – одиниці. У результаті маємо 8 трьохбітних повідомлень.

Схема утворення 4-бітних повідомлень



Закономірність

	4 біти					4 біти			
1	0	0	0	0	9	1	0	0	0
2	0	0	0	1	10	1	0	0	1
3	0	0	1	0	11	1	0	1	0
4	0	0	1	1	12	1	0	1	1
5	0	1	0	0	13	1	1	0	0
6	0	1	0	1	14	1	1	0	1
7	0	1	1	0	15	1	1	1	0
8	0	1	1	1	16	1	1	1	1

кількість бітів	кількість повідомлень, що можна закодувати
1	2
2	$2*2=2^2=4$
3	$2*2*2=2^3=8$
4	$2*2*2*2=2^4=16$
5	$2*2*2*2*2=2^5=32$
6	$2*2*2*2*2*2=2^6=64$
7	$2*2*2*2*2*2*2=2^7=128$
8	$2*2*2*2*2*2*2*2=2^8=256$

З додаванням кожного біта кількість повідомлень, які можна закодувати, зростає вдвічі, оскільки до попереднього набору спочатку дописуємо зліва "0", потім "1".

3 байтів складаються
кілобайти, гігабайти, мегабайти тощо.

1 кілобайт (1 Кбайт) = 1024 байти;

1 мегабайт (1 Мбайт) = 1024 Кбайт;

1 гігабайт (1 Гбайт) = 1024 Мбайт;

1 терабайт (1 Тбайт) = 1024 Гбайт.

**Зауважте, що 1 кілометр дорівнює 1000 метрів,
1 кілограм — 1000 грамів,
а 1 кілобайт — 1024 байтів.**

$$2^{10} = 1024$$

Те, що кожна наступна одиниця вимірювання інформації в 1024 рази більша від попередньої, пов'язано з тим, що звичною для нас є десяткова система числення, де всі «круглі» величини є степенями 10,

а у двійковому кодуванні застосовують
двійкову систему числення,
де «круглі» величини – степені двійки.

Давайте підрахуємо, скільки повідомлень можна закодувати за допомогою 1Кб.

$$1 \text{ Кб} = 1024 \text{ байти} = 1024 * 8 \text{ бітів} = 8192 \text{ біти.}$$

Отже, така кількість бітів дозволяє закодувати

$$2^{8192} \text{ повідомлень}$$

Це число надзвичайно велике.

Уявіть собі: двійку множимо саму на себе 8192 рази!

Насправді це незрівнянно більше, ніж кількість найелементарніших часток у Всесвіті.

На початку 20 століття завдяки появі радіозв'язку популярною стала азбука, яка була запропонована ще в 1838 і названа на честь американського винахідника Самуеля **Морзе**.

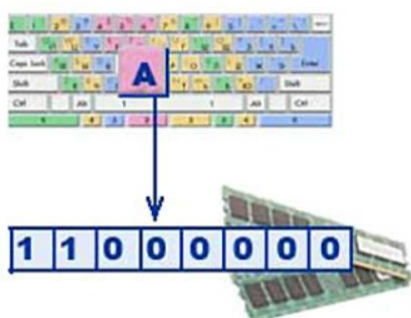


буква	код
А	. -
Б	- ...
В	. - -
Г	
Ґ	- - .
Д	- ..
Е	.
Є	.. - ..
Ж	... -
З	- - ..

Її можна вважати двійковим кодом, оскільки для передачі повідомлень використовувалися два види звукових сигналів: короткий (на папері його позначали крапкою) і довгий (тире).

Символи алфавіту кодувалися певним набором крапок та тире.

Натискання клавіші на клавіатурі надсилає на *контролер* (спеціальну мікросхему) електричний сигнал, який перетворюється на двійковий код і зіставляється з певним символом **таблиці кодів**.



Ви вже знаєте, що за допомогою **1 байта** можна закодувати 256 повідомлень, тобто створити 256 двійкових кодів.

Така кількість дає змогу закодувати латинські символи, символи кирилиці, цифри та спеціальні знаки.

Система кодування ASCII

(American Standard Code for Information Interchange – стандартний код інформаційного обміну)

Створена інститутом стандартизації США.
Для запису одного символу виділено 1 байт.
Символам присвоєно **256** кодів від 00000000 до 11111111.

Таблиця кодів поділена на дві частини.

Структура таблиці ASCII

	номери	коди	призначення
Стандартна (інтернаціональна) частина. (128 символів)	0-31	00000000 - 00011111	Керуючі символи.
	32-127	00100000 - 01111111	Цифри, букви латинського алфавіту і специфічні символи, що використовуються найчастіше.
Альтернативна частина. (128 символів)	128-255	10000000 - 11111111	Національні алфавіти, написання яких відмінно від латинського. (Може мати різні варіанти).

Фрагмент таблиці ASCII-
кодів для кодування цифр

Код у десятковій системі	Код у двійковій системі	Символ
48	110000	0
49	110001	1
50	110010	2
51	110011	3
52	110100	4
53	110101	5
54	110110	6
55	110111	7
56	111000	8
57	111001	9

Фрагмент таблиці ASCII-
кодів для кодування
символів латиниці

Код у десятковій системі	Код у двійковій системі	Символ
65	1000001	A
66	1000010	B
67	1000011	C
68	1000100	D
69	1000101	E
70	1000110	F
71	1000111	G
72	1001000	H
73	1001001	I
74	1001010	J
75	1001011	K
76	1001100	L
77	1001101	M
78	1001110	N
79	1001111	O

Фрагмент таблиці ASCII-
кодів для кодування
символів кирилиці

Код у десятковій системі	Код у двійковій системі	Символ
192	11000000	A
193	11000001	Б
194	11000010	В
195	11000011	Г
196	11000100	Д
197	11000101	Е
198	11000110	Ж
199	11000111	З
200	11001000	И
201	11001001	Й
202	11001010	К
203	11001011	Л
204	11001100	М
205	11001101	Н
206	11001110	О

Система кодування Unicode (уніфіковане кодування)



- З поширенням Інтернету та розвитком інформаційних технологій виникла необхідність розробки **єдиної системи** кодування текстової інформації, яка б охоплювала всі алфавіти та галузі знань.
- На зміну системі ASCII, у якій виділено для коду код 1 байт і можна кодувати лише 256 різних символів, в 1991 році було запропоновано систему багатобайтового кодування – **Unicode (Юнікод)**.

	A64	A65	A66	A67	A68	A69
0	З	Ы		Д	Т	
1	з	ы		д	т	
2	З	Ь	Д	С	Ч	
3	з	ь	д	с	ч	
4	З	О	Л	Ж	Н	
5	з	о	л	ж	н	
6	І	А	М	Ч	Ш	
7	і	а	м	ч	ш	
8	Ї	А	О	Д		
9	ї	а	о	д		
A	У	Х	О	Б		
B	у	х	о	б		
C	С	І	О	Т		
D	с	і	о	т		
E	Ъ	А	В	Ц		
F	ъ	а	в	ц		

Що представлено в Unicode?

- Практично всі сучасні писемності.
- Багато історичних писемностей, таких як давньогрецька, єгипетські ієрогліфи, клинопис, писемність майя та інші.
- Широкий набір піктограм, математичних, технічних, комерційних і музичних символів.

Проблеми Unicode

Очевидно, що використання в кодових таблицях 4 байтів дає можливість закодувати багато варіантів символів, але це призведе до збільшення обсягу текстів (однобайтові тексти збільшаться в 4 рази).

Компроміс?

- Як компроміс, на етапі розробки Юнікоду було запропоновано кілька форматів змінної довжини.
- **Кодування змінної довжини** використовує для різних груп символів різну довжину (1, 2 або 4 байти).

Найпоширеніші формати перетворення Unicode

UTF-8

- загальна кількість кодів **65 536**;
- сумісна з 8-бітовим (ASCII) кодуванням тексту;
- використовується в операційних системах та веб-просторі.

UTF-16

- загальна кількість кодів **1 112 064**;
- сумісна з UTF-8;
- використовується в сучасних операційних системах, веб-просторі.

UTF-32

- загальна кількість кодів **2 147 483 648**;
- сумісна з UTF-16;
- використовується для зберігання рядків в ОС UNIX.

Unicode transformation format (UTF)