

Інформатика. Перший рік

Розділ 1

Інформація. Інформаційні процеси та системи

9 клас

Тема 01-02

Кодування інформації

Кодування інформації

Як відомо, інформація може передаватися і зберігатися. Але виникає питання: у якому вигляді?

Для людини найпоширенішим засобом обміну інформацією є розмова.

Сигналами передачі такої інформації є звуки.

Деякі набори звуків утворюють слова, а групи слів – речення, які несуть інформаційний зміст.

В такому випадку кажуть, що інформація закодована звуками і передається звуками.

Етапи передачі інформації

Розглянемо етапи передачі інформації від однієї людини (джерела) до іншої (споживача):

1. Джерело формує інформацію, яку планує передати споживачеві.
2. За допомогою органів мовлення джерело кодує інформацію у вигляді звуків. Оскільки звуки мають властивість поширюватися у речовинному середовищі, вони доходять до органу слуху споживача.
3. Мозок споживача декодує отриману інформацію звукової форми і відкладає її у власній пам'яті.

Отже інформація передається у закодованому вигляді.

Причому, щоб споживач розумів отриману інформацію, він повинен володіти тією ж системою кодування, що й джерело.

Мова

Система засобів і правил кодування інформації називається **мовою**.

Мова є набором правил кодування інформації.

Код є формою представлення елементарного сигналу.

Зверніть увагу!

Ми розглянули приклад усної мови. В ній коди є звуковими сигналами і основним недоліком її є те, що ці коди, будучи придатними для передачі інформації, зовсім непридатні для її зберігання. Бо звук – це коливання середовища, яке з часом затухає. Для того, щоб зберегти інформацію передану звуковим способом, необхідно перетворити її в іншу форму. З цією метою людина винайшла письмові мови, в яких кожен звук кодується буквою, рідше – двома. Це дало можливість зберігати інформацію протягом тривалого часу у вигляді текстів. Завдяки чому можна не тільки передавати інформацію у закодованому вигляді але й зберігати її у вигляді кодів-букв.

Різнovidи мов

За час своєї історії людина винайшла багато різноманітних мов.

Багато із них придатні для кодування будь-якої інформації.

Але є і спеціалізовані з обмеженими можливостями кодування.

Сьогодні існують розмовні та відповідні їм письмові мови різних народів, мова жестів, міміки, мова креслення, математики та багато інших.

Всі вони є системами кодування інформації, мають певну будову та правила кодоутворення.

Переклад

Часто виникає задача перекодування інформації з однієї мови на іншу. Наприклад, під час написання диктанту учень виконує перекодування інформації сформованої засобами усної мови на коди письмової.

Також нерідко виникає потреба перекладу газет, журналів, книжок, фільмів з однієї мови на іншу.

Переклад – це процес перекодування інформації з однієї мови на іншу.

Якість перекладу. Перекладач

Слід мати на увазі, що під час перекладу інформація частково спотворюється. Тому існує поняття якості перекладу.

Якість перекладу – це ступінь тотожності вихідної з процесу перекладу інформації до вхідної.

На якість перекладу впливають як характеристики обох мов – вхідної та вихідної, так і можливості перекладача.

Перекладач – об'єкт, здатний виконати перекодування інформації з однієї мови на іншу.

Комп'ютери та машинні мови

З виникненням комп'ютерів виникло питання створення спеціальної мови для спілкування з ними.

Мову, якою можна спілкуватися з комп'ютером називають **машинною мовою**.

Комп'ютер – це машина для виконання інформаційних процесів: отримання інформації, її опрацювання, утворення нової інформації, зберігання та виведення.

Зверніть увагу!

В основу машинної мови покладені вказівки комп'ютеру на виконання певних дій. Ця мова значно простіша за будь-яку із людських, але вимагає більшої точності, однозначності та строгої послідовності. Через це перед перекладом на машинну мову спочатку необхідно однією із людських мов підготувати інформацію у вигляді алгоритму – послідовності вказівок, які повинен виконати комп'ютер.

Алгоритмісти та програмісти

Перед перекладом на машинну мову спочатку необхідно однією із людських мов підготувати інформацію у вигляді алгоритму – послідовності вказівок, які повинен виконати комп'ютер.

Людина, що створює алгоритм називається **алгоритмістом**.

Алгоритми створюються на спеціальних алгоритмічних мовах.

Наступним кроком підготовки інформації для введення в комп'ютер є переклад алгоритму на машинну мову.

Перекладач з алгоритмічної мови на машинну називається **програмістом**.

Дані та програма

Інформація, яку отримує комп'ютер, ділиться на дві групи:

- **дані** (інформація для зберігання, опрацювання і виведення)
- **програма** (машинний код, який є алгоритмом опрацювання даних).

Зверніть увагу!

Для роботи обидва види інформації в закодованому вигляді поміщаються в пам'ять комп'ютера. Далі виконується програма.

Але виникає питання: в якій формі виконати кодування інформації, щоб її можна було розмістити та опрацювати в комп'ютері?

Типи комп'ютерних сигналів

За конструкцією сучасний комп'ютер є електронною машиною, в якій використовується електрична енергія, яка може породжувати також магнітну.

Тому сигналами кодування в комп'ютері можуть бути або **електричний**, або **магнітний**.

Як же такими сигналами можна закодувати інформацію?
Все залежить від типу інформації, яка кодується.

Зверніть увагу!

Такі електронні пристрої як оптичні дисководи здатні для передачі інформації використовувати оптичні сигнали для зчитування інформаційних кодів з оптичних дисків, або запису на них. Але в роботі комп'ютера оптичні сигнали є не основними, а вторинними.

Ненадійний варіант кодування інформації машинними сигналами

Для числової інформації можна, наприклад, розробити сигнали для 10 цифр, а потім у певних точках електронної пам'яті розмістити їх, створивши там електричну напругу.

Скажімо, цифрі 5 відповідає електричний сигнал 5 Вольт, а цифрі 9 – 9 Вольт.

В такому випадку ми матимемо найзручнішу систему кодування числової інформації. Але у неї є суттєвий недолік. Щоб підтримувати, наприклад, строго 5 Вольт у певній точці пам'яті, потрібна досить складна електроніка.

Сучасні можливості електроніки допускають значні відхилення від бажаних значень і тому в певний час 5 може перетворитися на 4 або 6, в залежності від того, наскільки і в який бік зміниться сигнал.

Тому така система кодування виявляється ненадійною і може привести до спотворення інформації. У зв'язку з цим була запропонована система двійкового кодування.

Двійкове кодування

Виявилося, що елементарна точка пам'яті комп'ютера може надійно містити лише два види сигналу: **низький** – приблизно 1,5 Вольт, та **високий** – приблизно 5 Вольт.

Різниця в 3,5 Вольти гарантує, що при коливаннях рівня напруги в межах комп'ютерних пристроїв при умові їх технічної справності, не призведе до незапланованої зміни сигналу.

Для зручності низький рівень сигналу позначають **0**, а високий **1**.

При цьому виникає ситуація, коли в одній числовій позиції можна закодувати лише два числа: **0** та **1**.

Ця ситуація стала причиною створення системи **двійкового кодування чисел**.

Двійкові числа

В системі двійкового кодування використовуються не десяткові, а двійкові числа.

Двійкове число – це число записане за правилами двійкової позиційної числової системи, в якій використовується лише дві цифри: **0** та **1**.

Наприклад, десятковому числу **25** відповідає двійкове число **11001**

Існують алгоритми перетворення чисел із десяткової системи в двійкову та навпаки, але їх вивчення не входить до складу курсу інформатики в 9 класі.

Двійковий код символу

Якщо перші комп'ютери були створені виключно для задач обчислення (їх так і називали – електронна обчислювальна машина, скорочено ЕОМ), то згодом область задач комп'ютерів розширилася.

Виникла необхідність у введенні та **опрацюванні текстової інформації**. Це поставило задачу розробки кодів для символів тексту.

Принципи кодування символів.

Принцип перший

Для початку необхідно було визначитися із принципами кодування. Першим принципом була вимога **використання коду, що містить лише два види сигналів.**

Як розглядалося раніше, це обумовлено технічними особливостями електронних пристроїв.

Як і у випадку з кодуванням чисел для цього підходило двійкове кодування, але за іншими правилами .

Другий принцип кодування символів

Другий принцип – **однакова довжина коду для всіх символів**.

Це дозволяло зручно розділити пам'ять на окремі комірки та пронумерувати їх, надавши їм так звані адреси.

Адреса комірки дає можливість швидкого та зручного доступу до її вмісту. В кожній такій комірці в певний момент часу може міститися лише один символ.

Було обрано довжину комірки із **8 позицій** для елементарних кодових сигналів. Така система називається **8-розрядною**.

Одна така позиція отримала назву «**біт**».

Комірка отримала назву «**байт**».

Отже **1 байт=8 біт**.

Зверніть увагу!

Важливо розуміти, що 1 байт пам'яті – це 8 біт, але 8 довільних біт пам'яті – це не байт. **Байт – це 8 біт однієї комірки пам'яті.**

Третій принцип кодування символів

Третім принципом є **необхідність постановити у відповідність кожному символу, який може зустрітися в тексті, двійковий код.**

Для цього розробниками, а вони були американцями, була створена кодова таблиця символів ASCII - американський стандартний код для інформаційного обміну.

Вона містить коди 128 символів, із яких 32 перших мають управляюче значення і не виводяться ні на екран, ні на принтер, а свого часу в основному використовувалися для управління процесом виведення на матричні принтери.

Інші 96 символів – це літери латинського (англійського) алфавіту, розділові знаки, спеціальні знаки та цифри.

Таблиця символів ASCII (перша частина)

Порядковий номер	Код	Символ	Порядковий номер	Код	Символ	Порядковий номер	Код	Символ	Порядковий номер	Код	Символ
0	00000000	код управління	32	00100000	пробіл	64	01000000	@	96	01100000	`
1	00000001	код управління	33	00100001	!	65	01000001	A	97	01100001	a
2	00000010	код управління	34	00100010	"	66	01000010	B	98	01100010	b
3	00000011	код управління	35	00100011	#	67	01000011	C	99	01100011	c
4	00000100	код управління	36	00100100	\$	68	01000100	D	100	01100100	d
5	00000101	код управління	37	00100101	%	69	01000101	E	101	01100101	e
6	00000110	код управління	38	00100110	&	70	01000110	F	102	01100110	f
7	00000111	код управління	39	00100111	'	71	01000111	G	103	01100111	g
8	00001000	Backspace	40	00101000	(	72	01001000	H	104	01101000	h
9	00001001	код управління	41	00101001	)	73	01001001	I	105	01101001	i
10	00001010	код управління	42	00101010	*	74	01001010	J	106	01101010	j
11	00001011	код управління	43	00101011	+	75	01001011	K	107	01101011	k
12	00001100	код управління	44	00101100	,	76	01001100	L	108	01101100	l
13	00001101	код управління	45	00101101	-	77	01001101	M	109	01101101	m
14	00001110	код управління	46	00101110	.	78	01001110	N	110	01101110	n
15	00001111	код управління	47	00101111	/	79	01001111	O	111	01101111	o
16	00010000	код управління	48	00110000	0	80	01010000	P	112	01110000	p
17	00010001	код управління	49	00110001	1	81	01010001	Q	113	01110001	q
18	00010010	код управління	50	00110010	2	82	01010010	R	114	01110010	r
19	00010011	код управління	51	00110011	3	83	01010011	S	115	01110011	s
20	00010100	код управління	52	00110100	4	84	01010100	T	116	01110100	t
21	00010101	код управління	53	00110101	5	85	01010101	U	117	01110101	u
22	00010110	код управління	54	00110110	6	86	01010110	V	118	01110110	v
23	00010111	код управління	55	00110111	7	87	01010111	W	119	01110111	w
24	00011000	код управління	56	00111000	8	88	01011000	X	120	01111000	x
25	00011001	код управління	57	00111001	9	89	01011001	Y	121	01111001	y
26	00011010	код управління	58	00111010	:	90	01011010	Z	122	01111010	z
27	00011011	код управління	59	00111011	;	91	01011011	[	123	01111011	{
28	00011100	код управління	60	00111100	<	92	01011100	\	124	01111100	
29	00011101	код управління	61	00111101	=	93	01011101	]	125	01111101	}
30	00011110	код управління	62	00111110	>	94	01011110	^	126	01111110	~
31	00011111	код управління	63	00111111	?	95	01011111	_	127	01111111	Delete

Таблиця кодів символів ASCII (друга частина)

Неважко побачити, що 8-й біт (відлік бітів іде справа наліво) кожного коду розглянутої таблиці нульовий.

Тому існує можливість додати до цієї таблиці ще одну, в якій також можна закодувати ще 128 символів.

Всі їх коди будуть починатися з 1.

Ця можливість була використана для створення **таблиці кодів національних символів різних країн**.

Наприклад, для країн, в яких використовується **Кирилиця** – Росії, України та Білорусії, була створена так звана **cp866** таблиця, щоправда вона була зроблена для операційної системи MS DOS і не підходить для MS Windows.

В **MS Windows** використовується кодова таблиця **win1251**. Наведена далі.

Таблиця символів ASCII (друга частина) win1251

Примітка.

Ця таблиця також містить декілька управляючих символів (порожні комірки в стовбцях «Символ»).

Порядковий номер	Код	Символ	Порядковий номер	Код	Символ	Порядковий номер	Код	Символ	Порядковий номер	Код	Символ
128	10000000	Ђ	160	10100000		192	11000000	А	224	11100000	а
129	10000001	Ѓ	161	10100001	Ў	193	11000001	Б	225	11100001	б
130	10000010	„	162	10100010	ў	194	11000010	В	226	11100010	в
131	10000011	ѓ	163	10100011	Ј	195	11000011	Г	227	11100011	г
132	10000100	„	164	10100100	□	196	11000100	Д	228	11100100	д
133	10000101	...	165	10100101	ѓ	197	11000101	Е	229	11100101	е
134	10000110	†	166	10100110	!	198	11000110	Ж	230	11100110	ж
135	10000111	‡	167	10100111	§	199	11000111	З	231	11100111	з
136	10001000	€	168	10101000	Ё	200	11001000	И	232	11101000	и
137	10001001	‰	169	10101001	©	201	11001001	Й	233	11101001	й
138	10001010	Љ	170	10101010	€	202	11001010	К	234	11101010	к
139	10001011	«	171	10101011	«	203	11001011	Л	235	11101011	л
140	10001100	Њ	172	10101100	¬	204	11001100	М	236	11101100	м
141	10001101	Ќ	173	10101101	-	205	11001101	Н	237	11101101	н
142	10001110	Ћ	174	10101110	®	206	11001110	О	238	11101110	о
143	10001111	Ў	175	10101111	Ї	207	11001111	П	239	11101111	п
144	10010000	ђ	176	10110000	°	208	11010000	Р	240	11110000	р
145	10010001	·	177	10110001	±	209	11010001	С	241	11110001	с
146	10010010	’	178	10110010	!	210	11010010	Т	242	11110010	т
147	10010011	“	179	10110011	!	211	11010011	У	243	11110011	у
148	10010100	”	180	10110100	ѓ	212	11010100	Ф	244	11110100	ф
149	10010101	•	181	10110101	µ	213	11010101	Х	245	11110101	х
150	10010110	–	182	10110110	¶	214	11010110	Ц	246	11110110	ц
151	10010111	—	183	10110111	·	215	11010111	Ч	247	11110111	ч
152	10011000	□	184	10111000	ё	216	11011000	Ш	248	11111000	ш
153	10011001	™	185	10111001	№	217	11011001	Щ	249	11111001	щ
154	10011010	Љ	186	10111010	ё	218	11011010	Ъ	250	11111010	ъ
155	10011011	»	187	10111011	»	219	11011011	Ы	251	11111011	ы
156	10011100	Њ	188	10111100	ј	220	11011100	Ь	252	11111100	ь
157	10011101	ќ	189	10111101	ѕ	221	11011101	Э	253	11111101	э
158	10011110	ћ	190	10111110	ѕ	222	11011110	Ю	254	11111110	ю
159	10011111	и	191	10111111	ї	223	11011111	Я	255	11111111	я

Таблиця кодів символів UNICODE

Для різних операційних систем та різних народів було створено багато кодових одnobайтних таблиць.

Кожна із них має своє кодове ім'я і вказується при установці операційної системи.

Щоб спростити проблему кодових таблиць була створена **двобайтна кодова таблиця UNICODE**, яка може містити до **65536** символів, а значить, містити алфавіти всіх народів світу та спеціальні символні набори. Така система називається **16-розрядною**.

Але поки що «старі» таблиці застосовуються значно ширше.

Вимірювання обсягу даних

Обсяг даних вимірюється довжиною двійкового коду, тобто кількістю байтів у повідомленні.

Наприклад, текст «**Андрій грає у футбол**» має 20 байтів з урахуванням пропусків.

Об'єм інформації вимірюється кількістю байтів у коді повідомлення (текстового, графічного тощо).

Зверніть увагу!

У тексті розділові знаки і «пропуски» - теж символи.

Нехай на аркуші 56 рядків по 64 символів у рядку. Підрахуємо кількість інформації на аркуші: $56 \times 64 = 3584$ байт.

Вимірювання великих обсягу даних

Для вимірювання великих обсягів даних використовують такі позначення: **кілобайти (КБ)**, **мегабайти (МБ)**, **гігабайти (ГБ)** і **терабайти (ТБ)**.

В обчислювальній техніці префікси *кіло-*, *мега-*, *гіга-*, *тера-* мають дещо інший зміст, ніж в інших науках.

Кіло- - префікс, що позначає коефіцієнт $2^{10} = 1024$.

Таким чином:

$1 \text{ КБ (кілобайт)} = 2^{10} \text{ байтам} = 1024 \text{ байтам}$

$1 \text{ МБ (мегабайт)} = 2^{10} \text{ КБ} = 1024 \text{ КБ}$

$1 \text{ ГБ (гігабайт)} = 2^{10} \text{ МБ} = 1024 \text{ МБ}$

$1 \text{ ТБ (терабайт)} = 2^{10} \text{ ГБ} = 1024 \text{ ГБ}$

Щоб приблизно підрахувати об'єм інформації у книзі, необхідно кількість символів на одній сторінці помножити на кількість сторінок.

Нехай книжка має **256** сторінок по **3584** байти на сторінці. Об'єм інформації такої книги:

$3584 \times 256 = 917504 \text{ байт} = 896 \text{ КБ} = 0,875 \text{ МБ}$

Підсумки

Інформація зберігається і передається у закодованому вигляді.

Кодами можуть бути звукові сигнали, зображення, в тому числі значки, електричні та магнітні сигнали тощо.

Для кодування інформації користуються певними мовами.

Процес перекодування інформації з однієї мови на іншу називається перекладом.

Для спілкування з комп'ютером використовують спеціальні машинні мови.

Оскільки в комп'ютері для кодування доступні лише електричні та магнітні сигнали, для кодування інформації використовують двійкову систему кодування.

Кодування чисел виконується за певними математичними правилами.

Для кодування символів були створені спеціальні таблиці.

Обсяг даних в інформатиці вимірюється в байтах, кілобайтах, мегабайтах, гігабайтах та терабайтах.