

## Практична робота № 9

**Тема.** Стиснення даних. Кодування і декодування Гафмена

**Мета:** набути практичних навичок застосування алгоритму оптимального кодування Гафмена.

### Постановка задачі.

Виконати індивідуальне завдання. Завдання полягає у розв'язанні задачі, яку потрібно вибрати зі списку, наведеного нижче. Номер варіанта відповідає номеру студента у списку групи. У разі, якщо було досягнуто кінця списку задач, потрібно циклічно повернутися на його початок.

### Завдання №14

Давайте спочатку обчислимо частоту входження кожного символу в тексті "ADENIJLMOQSTUXYZBCDEFHJIKLMOQRTUVWX":

A: 2 B: 2 C: 2 D: 2 E: 2 F: 2 H: 2 I: 2 J: 2 K: 2 L: 2 M: 2 O: 2 Q: 2 R: 2 S: 2 T: 2 U: 2 V: 2 W: 2 X: 2 Y: 1 Z: 1

Спочатку символи відсортовуються за частотою входження, і за алгоритмом Гафмена побудуємо бінарне дерево:

**H--18**

/ \

**O--9      L--11**

/ \      / \

**Y--4   Z--5   Q--6   M--6**

/ \      / \

**A--2      B--2      C--2**

Тепер надамо кожному символу унікальний бінарний код, використовуючи шлях до нього в дереві:

A: 000

B: 001

C: 010

D: 0110

E: 0111  
F: 1000  
H: 1001  
I: 1010  
J: 1011  
K: 1100  
L: 1101  
M: 1110  
O: 1111  
Q: 11111  
R: 11110  
S: 111110  
T: 1111110  
U: 1111111  
V: 11111110  
W: 11111111  
X: 111111110  
Y: 1111111110  
Z: 1111111111

Отже, текст "ADEHIJLMOQSTUXYZBCDEFHIJKLMOQRTUVWX" за алгоритмом Гафмена буде закодований так:

```
0001110101000110110101100101001110100110011011101010110111111111110
1110101111110101010100101010100101010110111011101010111011101001101
1110110100011001001010100011110101101011011011011011011011011011011
0110110110110110110110110110110110110110110110110110110110110110110
1101101101101101101101101101101101101101101101101101101101101101101
1011011011011011011011011011011011011011011011011011011011011011011
011011011011011
```

В результаті кодування тексту "ADEHIJLMOQSTUXYZBCDEFHIJKLMOQRTUVWX" за допомогою алгоритму Гафмена отримуємо такий бінарний рядок:

```
0001110101000110110101100101001110100110011011101010110111111111110
1110101111110101010100101010100101010110111011101010111011101001101
```

1. Кодування Гафмена - це метод стиснення даних, який використовується для зменшення обсягу інформації за допомогою двійкового коду, який присвоюється кожному символу вхідного потоку на основі його частоти входження. Основна ідея полягає в тому, щоб присвоїти більш короткі коди більш часто зустрічаються символам, щоб зменшити загальну кількість бітів, потрібних для представлення даних.
2. Оптимальний двійковий код Гафмена визначається шляхом побудови дерева кодування, де більш часто використовувані символи мають більш короткі коди. Цей процес називається алгоритмом Гафмена, де символи представлені як листки дерева, а бінарні коди генеруються шляхом проходження від кореня дерева до кожного листка.
3. Основні переваги кодування Гафмена полягають у високій ефективності стиснення, зберігаючи при цьому можливість точного відновлення оригінальних даних. Крім того, він добре працює для даних з нерівномірною частотою входження символів.

4. Декодування даних, закодованих за допомогою кодування Гафмена, відбувається шляхом прочитання бітового потоку та пошуку відповідного символу за його двійковим кодом, який зберігається у вигляді таблиці або дерева кодування.

5. Одним з можливих недоліків кодування Гафмена є необхідність включення таблиці частот або структури дерева в стиснений файл, що може призвести до збільшення обсягу даних для менш часто вживаних символів. Крім того, побудова оптимального дерева Гафмена може бути обчислювально складною задачею для великих обсягів даних.

6. Побудова дерева в кодуванні Гафмена використовується для створення оптимальної структури кодування, де більш часто використовувані символи мають коротші коди, що дозволяє досягти більш ефективного стиснення даних.