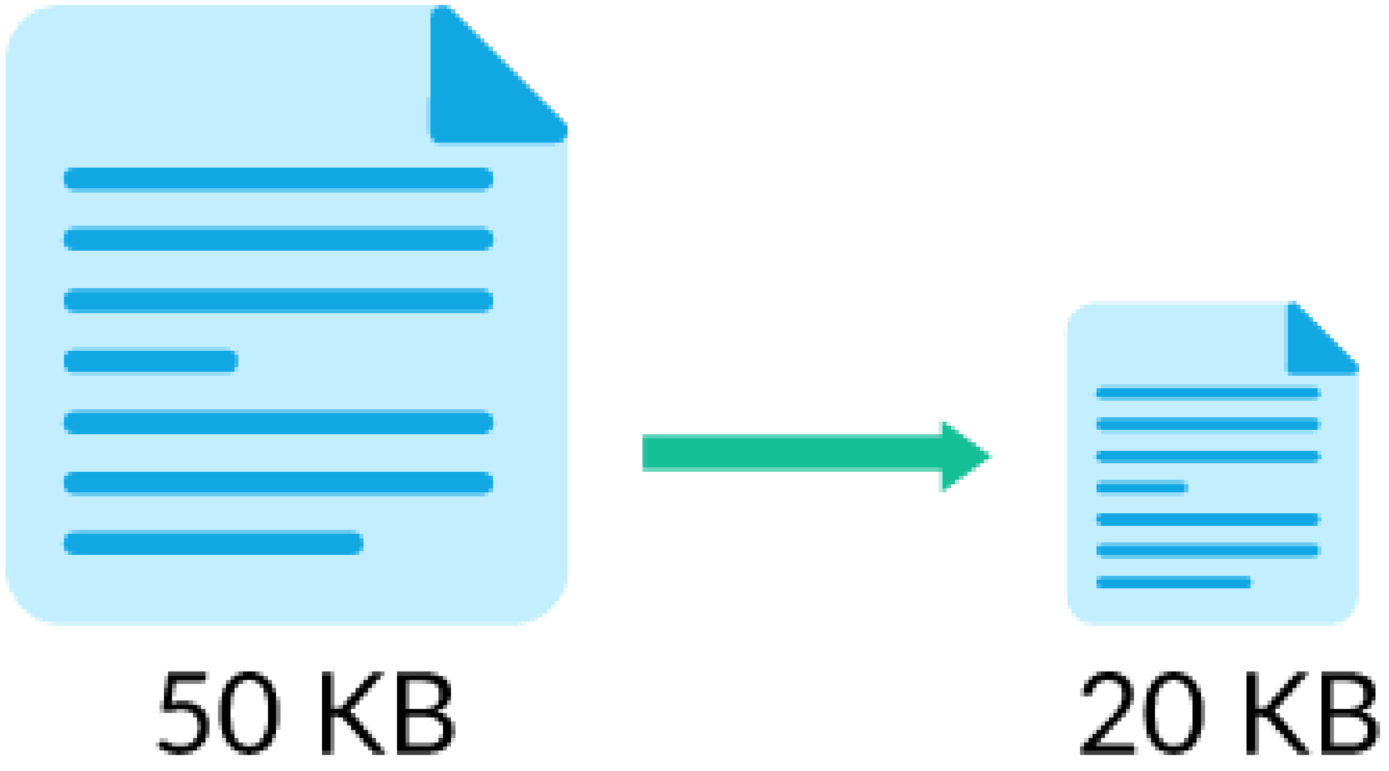
TEXT COMPRESSOR

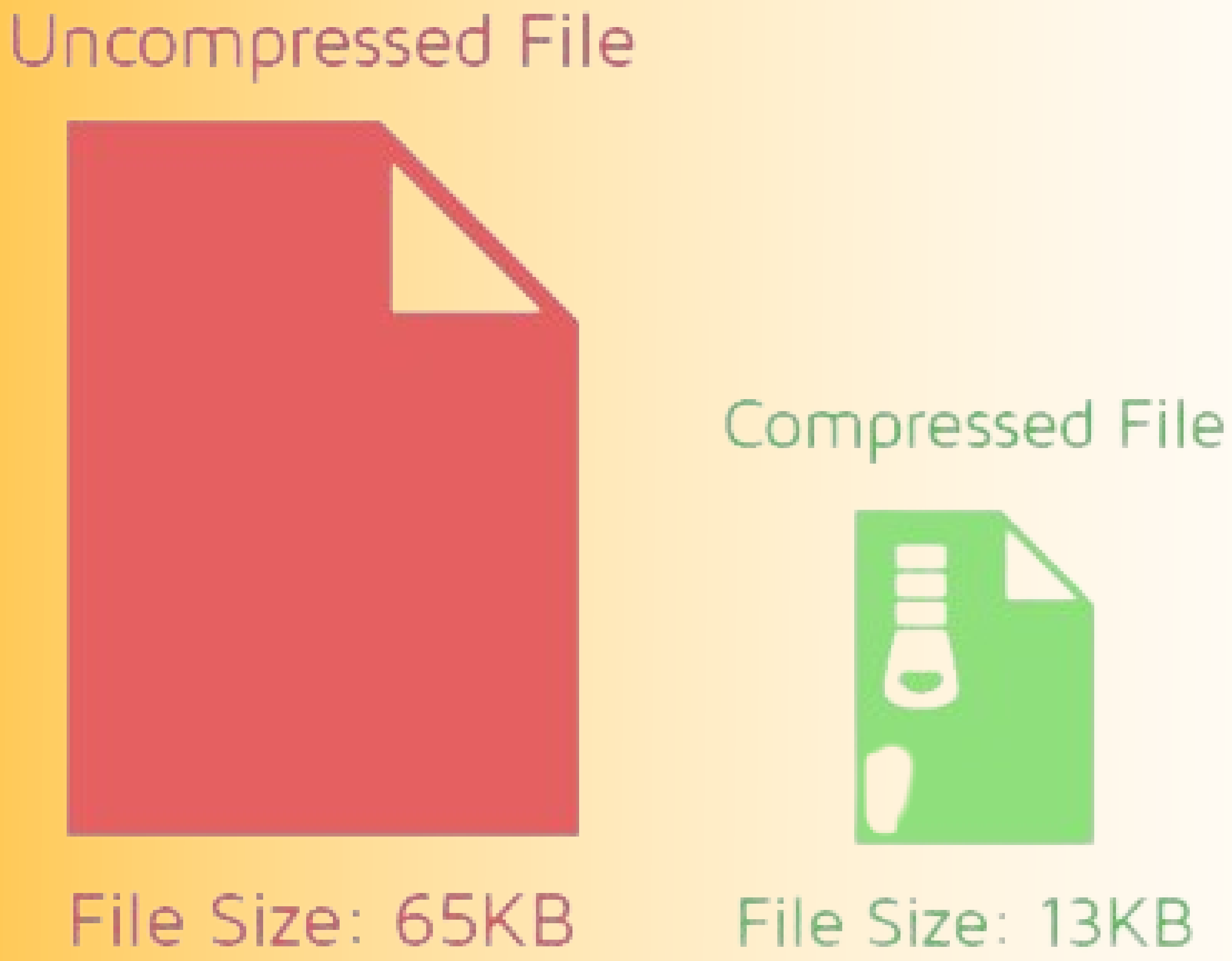
- Types of Compression:
    - Lossless
    - Lossy

- Types of Text Encoding:
    - Fixed length encoding
    - Variable length encoding (Based on frequency)

- Methods used for compression:
    - Run-length Encoding
    - Huffman Coding
    - Shannon-FANO Coding



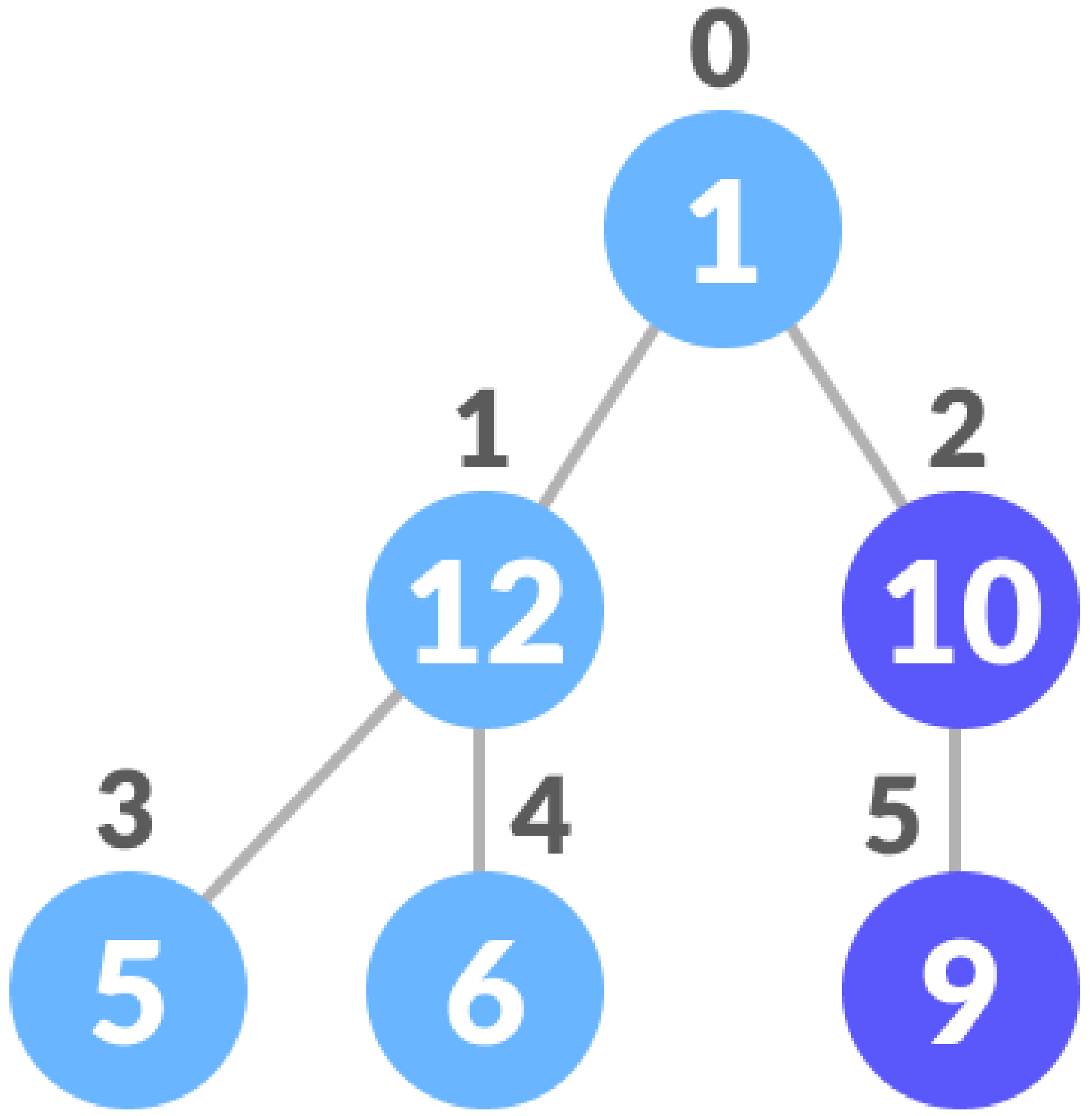
Huffman Coding Algorithm



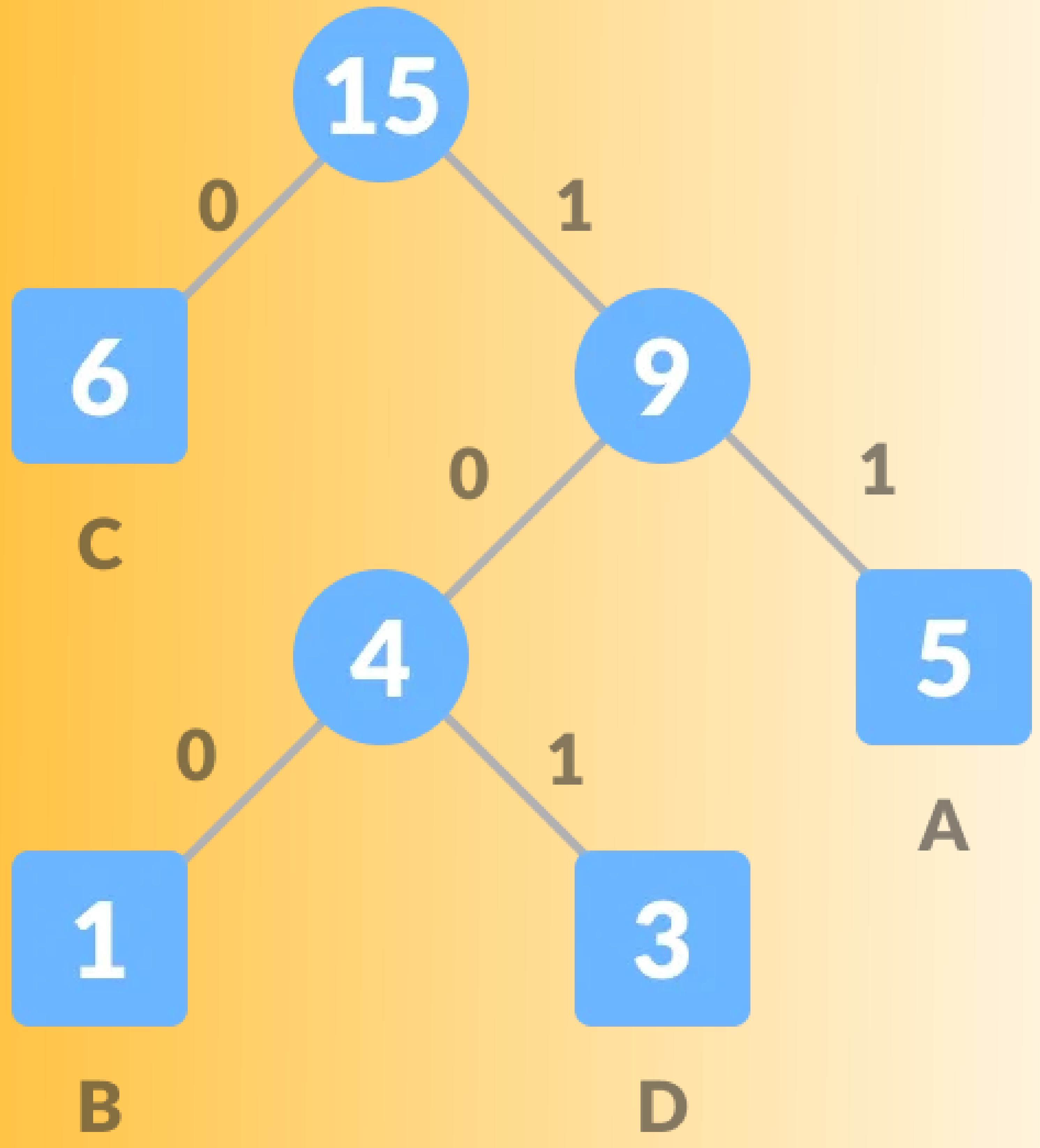
- Huffman code is a particular type of optimal prefix code.
- Uses **Lossless Compression**.
- Formulates a **variable-length** code table.
- It follows a **Greedy** approach; deals with generating minimum length prefix-free binary codes.
- Most frequent character allotted shortest code, while least frequent is given longest code.
- Complexity : **O(n log n)**

What Have We Implemented?

- Counting Sort (Modified)
- Linked List
- Stack
- Priority Queue
- Arrays
- Full Binary Tree
- Min Heap
- File Handling

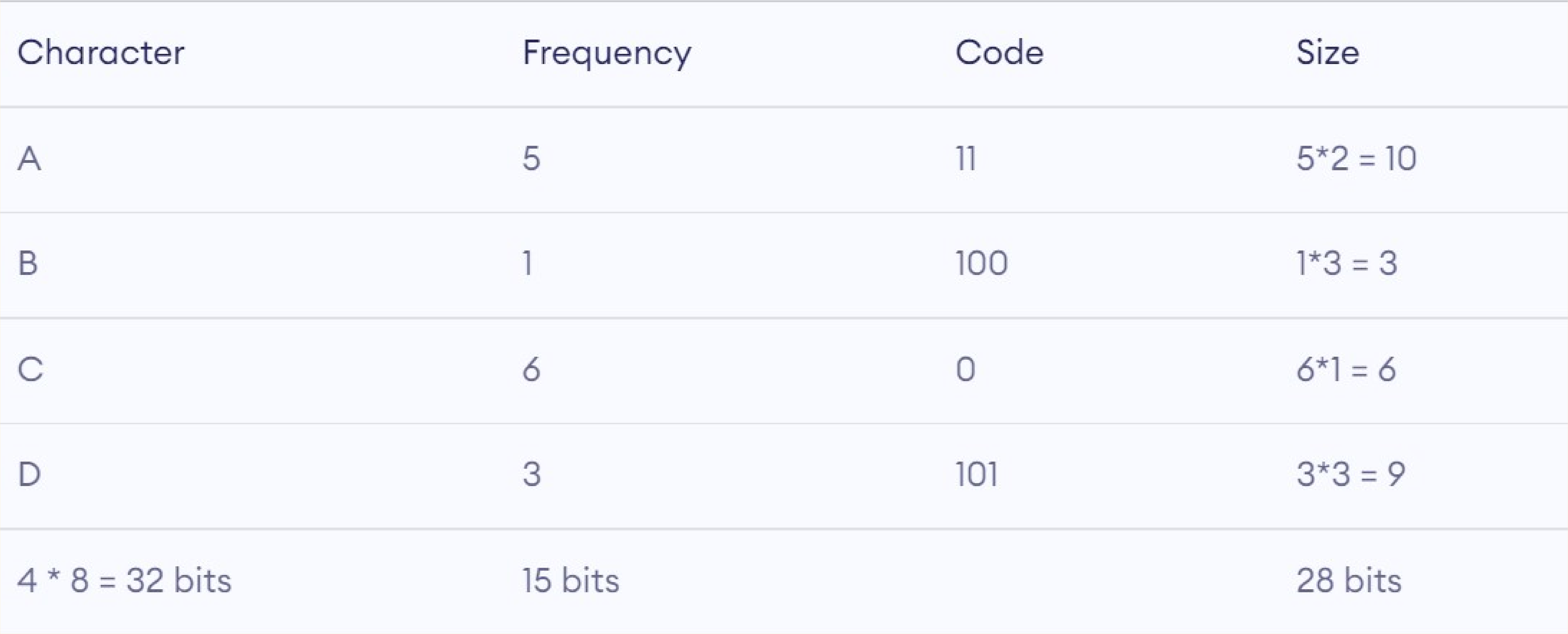


0	1	2	3	4	5
1	12	10	5	6	9



Analysing the Algo

- All of the file's unique characters and their frequencies are calculated.
- The characters and frequencies are then added to a Min-heap.
- 2 minimum frequency characters are extracted and added to a dummy root.
- Value of this dummy root is the sum of frequencies of its nodes.
- This root node is added back to the Min-heap.
- Process is repeated until there is only one element left in the Min-heap.



Character	Frequency	Code	Size
A	5	11	5*2 = 10
B	1	100	1*3 = 3
C	6	0	6*1 = 6
D	3	101	3*3 = 9
4 * 8 = 32 bits	15 bits		28 bits