Huffman Coding

Course: Algorithmic Problem Solving

Course Code: 17ECSE309

Submitted By:

Pooja Harish Pandit

01FE15BCS131

What is Huffman Coding?

- Huffman coding is a lossless data compression algorithm.

- The idea is to assign variable-length codes to input characters, lengths of the assigned codes are based on the frequencies of corresponding characters.

- The most frequent character gets the smallest code and the least frequent character gets the largest code.

How it works?

SL. No	Letter	Weights
1.	A	17
2.	B	3
3.	C	12
4.	D	5
5.	E	10

1
Sort with respect to weights

B	D	E	C	A
3	5	10	12	17

2
Combine the first two and sort

B & D	E	C	A
8	10	12	17

C	A	B & D & E
12	17	18

..5
Combine until one node remains

B & D & E	C & A
18	29

B & D & E & C & A
47

Construct the tree based on the table generated and assign the bits
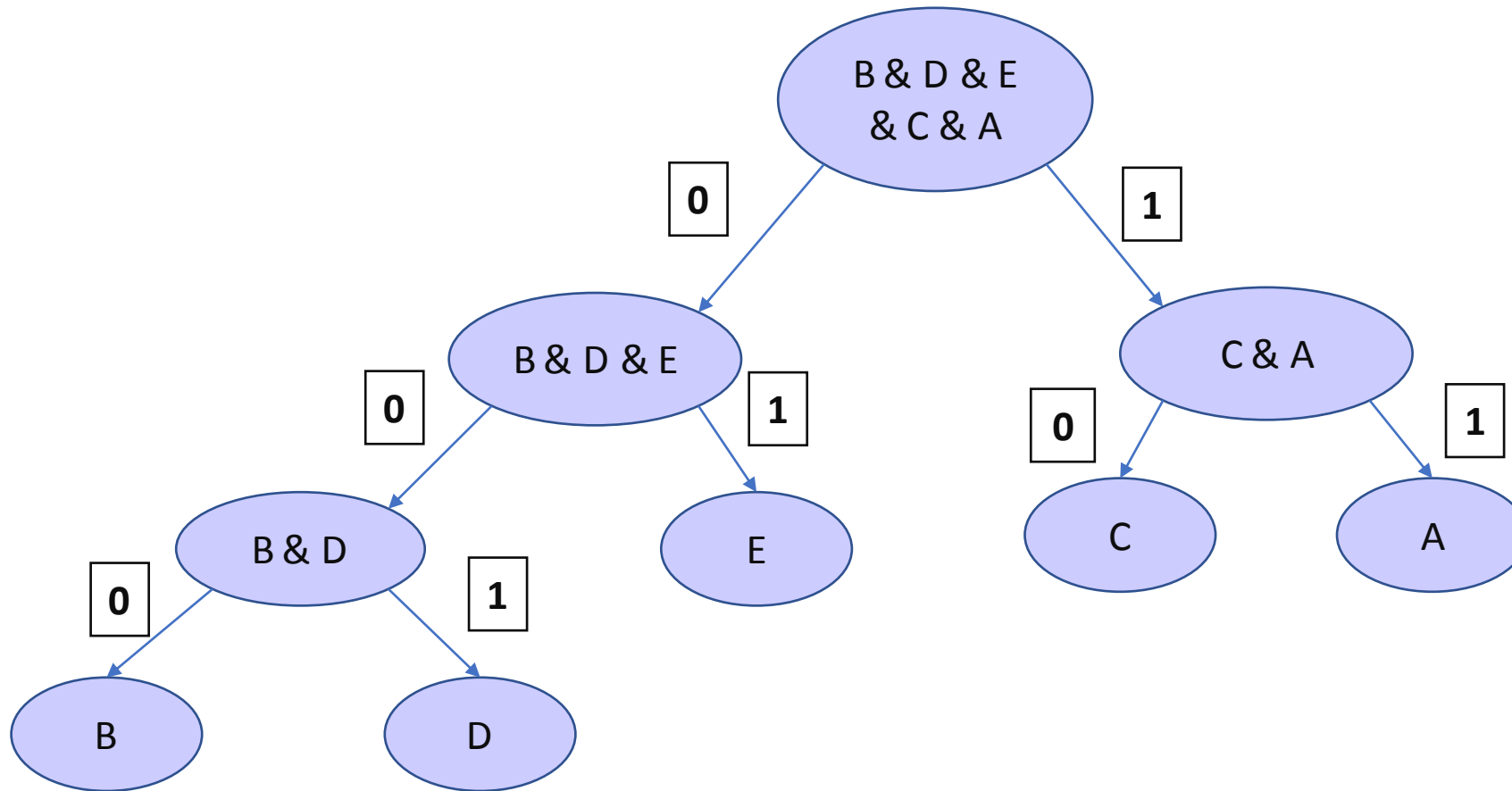
SL. No	Letter	Code
1.	A	11
2.	B	000
3.	C	10
4.	D	001
5.	E	01

Why it works?

The variable-length codes assigned to input characters are Prefix codes, means the codes (bit sequences) are assigned in such a way that the code assigned to one character is not prefix of code assigned to any other character.

This is how Huffman Coding makes sure that there is no ambiguity when decoding the generated bit stream.

Applications

➢Huffman is widely used in all the mainstream compression formats
- Data compression: GZIP, PKZIP (winzip etc) and BZIP2
- Image Compression: JPEG and PNG.

Bibliography

- https://brilliant.org/wiki/huffman-encoding

- https://www.geeksforgeeks.org/greedy-algorithms-set-3-huffman-coding

- https://en.wikipedia.org/wiki/Huffman_coding

- https://stackoverflow.com/questions/2199383/what-are-the-real-world-applications-of-huffman-coding

- https://users.cs.cf.ac.uk/Dave.Marshall/Multimedia/node210.html