UNIT-3

Data Compression

Memory Management.

Compression Techniques: (Zipping)

1. Huffman Encoding.
2. Run - Length Encoding.
3. Ziv - Lempel Encoding
   (ZV77)

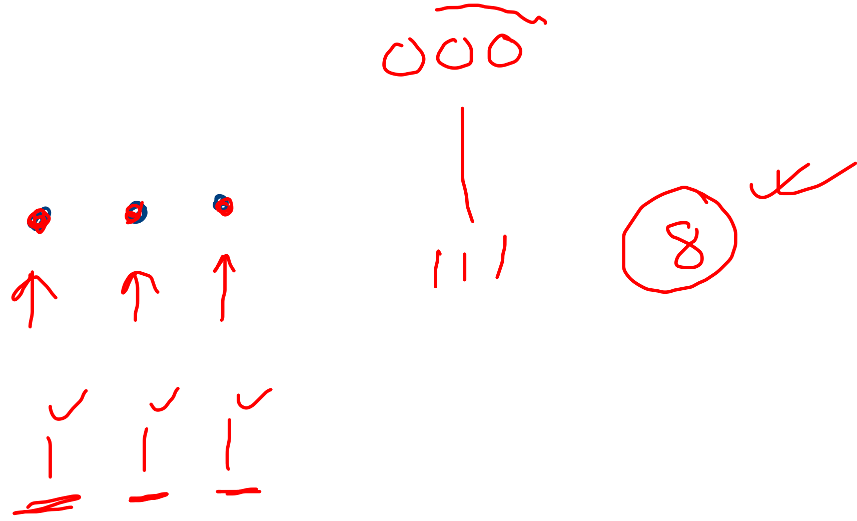
① Huffman Coding.

000 010 010 001 - - - - - - - -

A I I E S I P I A E P T . . . .

Not an example of Huffman Encoding

Char	Code	Fixed Length	Freq.	Total Bits
A	000	3	10	30
E	001	3	15	45
I	010	3	12	36
S	011	3	3	9
T	100	3	4	12
P	101	3	13	39
\n	110	3	1	3
				174



Huffman

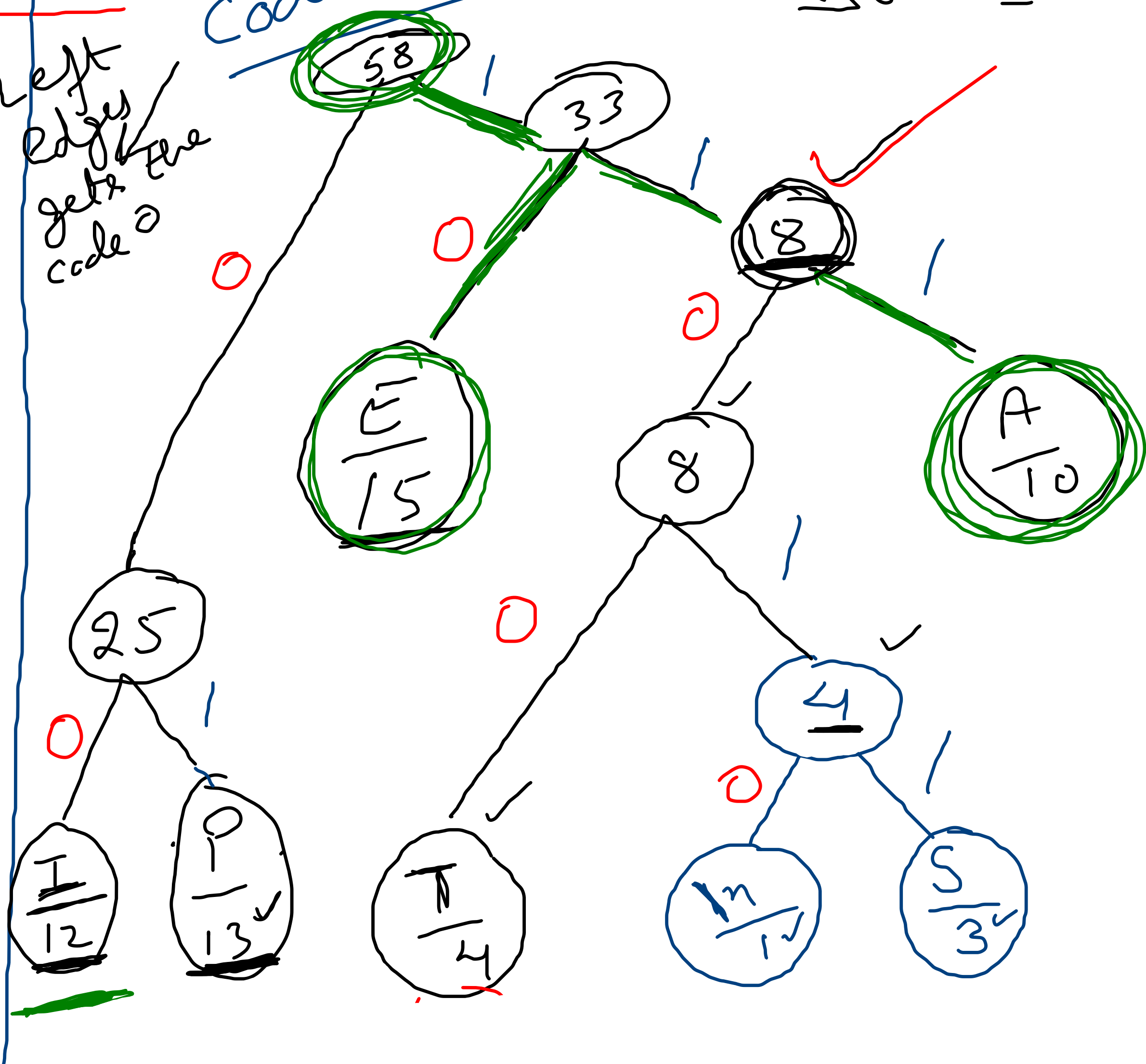
↳ Variable length code

Chars	Var. length code	Length	Freq.	Total Bits-
A	111	3	10 ✓	30
E	10	2	15	30
I	00	2	12	24
S	11011	5	3 ✓	15
T	1100	4	4 ✓	16
P	01	2	13	26
\n	11010	5	1	5 ✓
				146 bits

Left edges gets the code 0

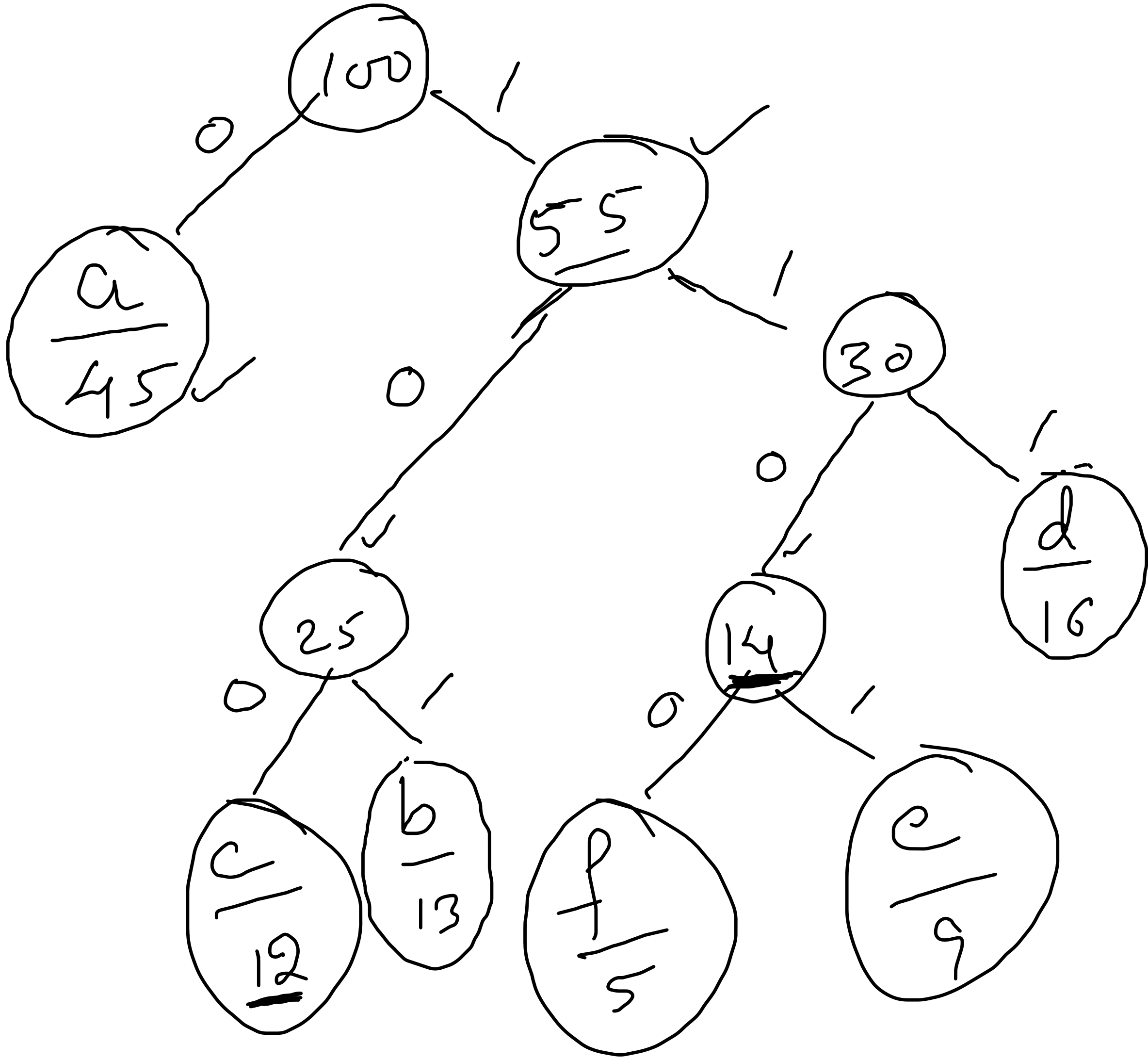
Code Tree

Right-edges gets 1



Example - 2 :

Chars.	Freq	Var. Len. code	Len.	Total Bits.
a	45 ✓	0	1 ✓	45
b	13 ✓	101	3	39
c	12 ✓	100	3	36
d	16 ✓	111	3	48
e	9 ✓	1101	4	36
f	5 ✓	1100	4	20
				224



② Run Length Encoding :

A run is defined as a sequence of identical characters.

eg:- $s = \underline{aaa}\,\underline{bb}\,\underline{a}$ → 3 runs

encoded form is a pair. (n, ch)

n → no. of consective same chars.

ch → character.

~~en~~ decoded. ~~exco~~

$3a2b1a$ ✓

aabba

eg:-2. $s = nnn***r\%\%\%\%\%\%\%$ → encoded form. $3n3*1r7\%$

Problem: eg:- $s = 11111111111544444$

encoded form → 1111554

Soln: In place of n we can use ASCII char. ✓✓

eg:- 43 consective 1's $\xrightarrow{\text{ASCII code}}$ +1

Quick Quiz

① Huffman coding uses ____ length codes:

✓ A) Variable

B) Fixed.

② Generate a Run Length code for the folling message:

aaaa##!??bbb

A) 4b2#1!3?2b

B) 4a2#1!2?3b ✓

C) ab3AbB

③ ZIV-LEMPEL Code

✓ LZ77 (version)

✓ LZW (another version]

In earlier methods:

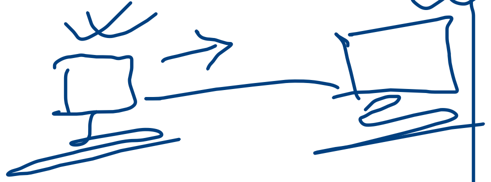
- Prior knowledge is required.
- Whole message needs to be encoded.

Universal compression.

LZ77 - Buffer of symbols.

Input

l_1 l_2



Codeword

1 ___ 2 ___ 3 ___

1 → Position in which match is found

2 → Length of the match.

3 → First mismatching symbol.

eg: 2 2 b

Input/Message	Buffer l_1 l_2 4 4	Code Transmitted
aababacbaacbaadaaa...	aaaa 0123 4567	a
aababacbaacbaadaaa...	aaaa aaba	2 2 b
abacbaacbaadaaa...	aaab abac	2 3 c
baacbaadaaa...	abacbaac	1 2 a
cbaadaaa...	cbaacbaa	0 3 a
daaa....	cbaadaaa 3	3 0 d
aaa...		

For Memory Management part of Unit-3, Separate slides have already been shared.