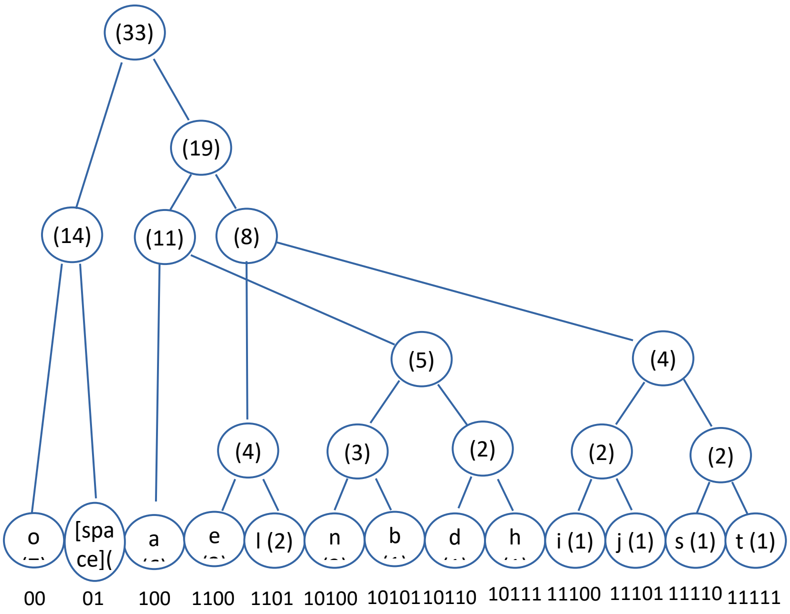
11th Practical Class – text compression

Exercises

1. Consider the phrase "olha o balão na noite de São João" (you should ignore diacritics/accents and consider capital letters as equal to small letters, but do not ignore spaces).
If you use a Huffman encoding, how many bits do you need?
Show the construction steps of the tree you used and which codes associated with each symbol.

Symbol	Freq.
o	7
[space]	7
a	6
e	2
l	2
n	2
b	1
d	1
h	1
i	1
j	1
s	1
t	1



2. Consider the text "pimpampumcadabolamataum":

a) Define a constant coding system for the text above. What is the minimum code size and the cost of encoding for the given text?

There are 11 different symbols, which must be represented using at least *ceil(log2(11))=ceil(3.46)=4 bits* per symbol.

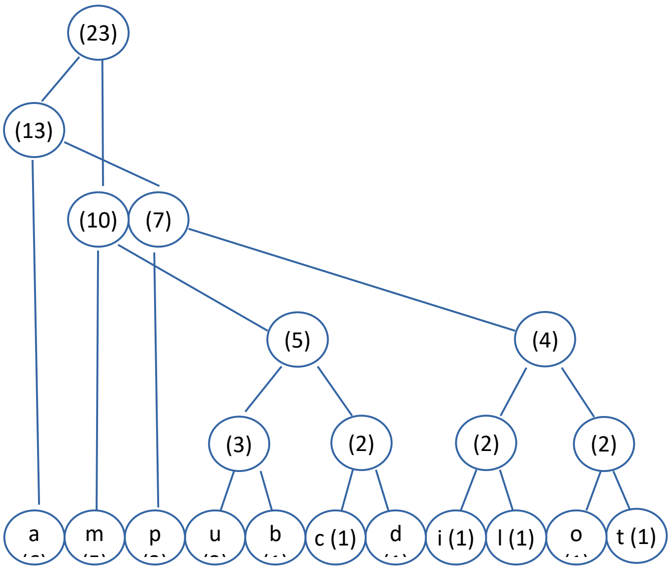
The cost is 4 bit * 23 = 92.

Symbol	Freq.	CODE
p	3	0000
i	1	0001
m	5	0010
a	6	0011

u	2	0100
c	1	0101
d	1	0110
b	1	0111
o	1	1000
l	1	1001
t	1	1010

b) Determine the Huffman coding tree for this text, explaining in detail the whole process. What is the cost of coding in this case?

Symbol	Freq.	Code	Cost
a	6	00	12
m	5	10	10
p	3	010	9
u	2	1100	8
b	1	1101	4
c	1	1110	4
d	1	1111	4
i	1	01100	5
l	1	01101	5
o	1	01110	5
t	1	01111	5



The total cost is 71 bits.

c) Using the Huffman tree calculated in the previous paragraph, present the codification of the phrase "pimpampum" and its cost. Also display the character encoding individually.

p	i	m	p	a	m	p	u	m
010	01100	10	010	00	10	010	1100	10

01001100100100010010110010