

Assignment : 3

1 Explain Arithmetic Algorithm for Encoding and Decoding.

=> Arithmetic Encoding - Decoding is a technique for lossless data compression.

=> Arithmetic Encoding:

```
low = 0.0, high = 1.0;
while not EOF do:
    range = high - low;
    read(c);
    high = low + range * high_range(c);
    low = low + range * low_range(c);
end do
output(low);
```

⇒ Arithmetic Coding:

$r = \text{input-number}$
repeat

Search c such that r falls in its range

Output c ;

$r = r - \text{low-range}(c)$;

$r = r / (\text{high-range}(c) - \text{low-range}(c))$;

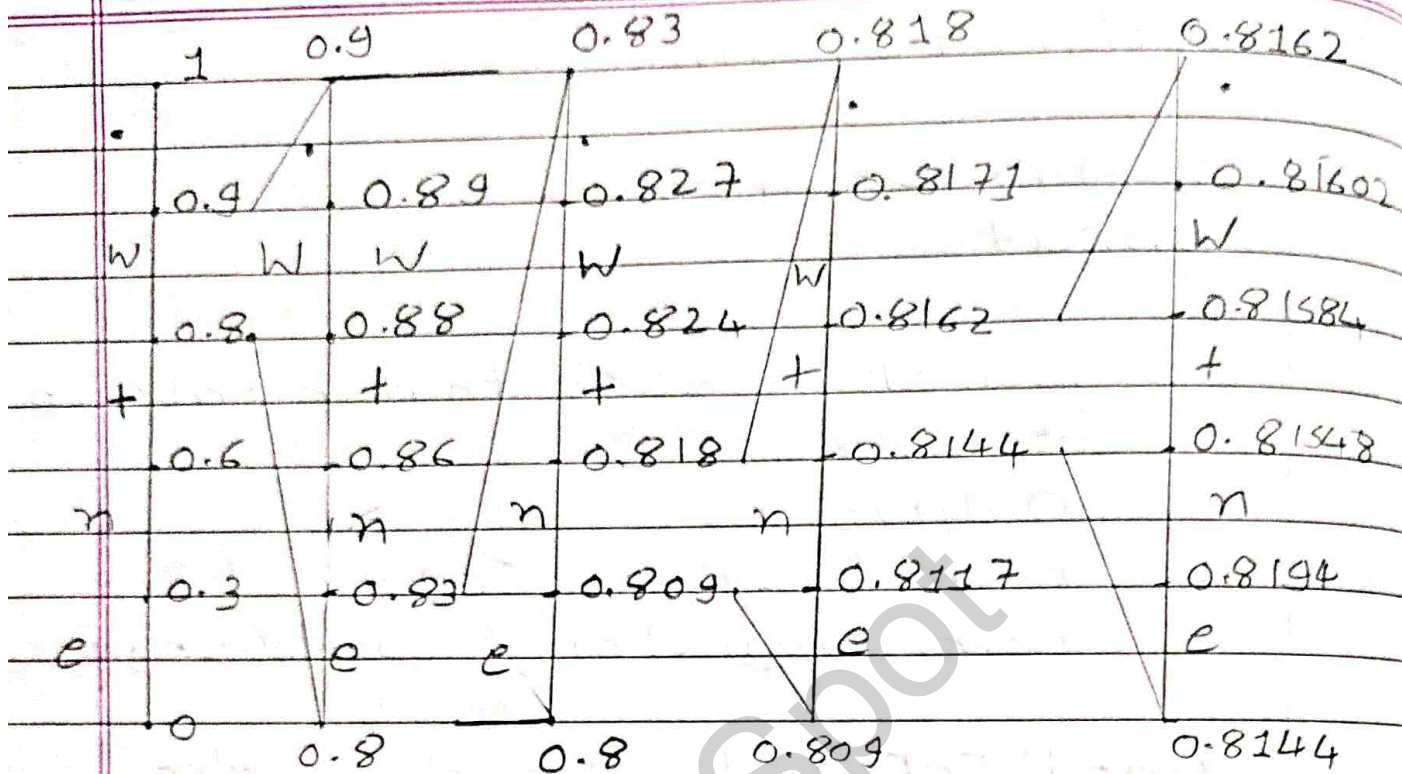
until EOF or the length of the message is reached,

2 Find the tag value using Arithmetic coding of the given sequence "went." with the probability model $P(e) = 0.3$, $P(n) = 0.3$, $P(t) = 0.2$, $P(w) = 0.1$, $P(.) = 0.1$

⇒ First we have to arrange the string into the alphabetical order - entw.

$d = \text{Range} = \text{Upper limit} - \text{Lower limit}$
Range

of alphabet = Lower + d (Probability Limit of alphabet)



⇒ For W: $d = \text{Upper Limit} - \text{Lower Limit}$
 $d = 0.9 - 0.8 = 0.1$

$$\text{Range } e = \text{Lower Limit} + d(CP)$$

$$= 0.8 + 0.1(0.3)$$

$$0.8 = 0.83$$

$$\text{Range } n = \text{Lower Limit} + d(CP)$$

$$= 0.83 + 0.1(0.2)$$

$$0.83 = 0.86$$

$$\text{Range } + = \text{Lower Limit} + d(CP)$$

$$0.86 = 0.86 + 0.1(0.1) = 0.88$$

નિકળતા ગુણ શક્તિઓને સતેજ કરે છે.

$$\text{Range } W = \text{Lower} + d(\text{CP})$$

$$0.88 := 0.88 + 0.1(0.1)$$

$$0.88 := 0.89$$

$$\text{Range } \cdot = \text{Lower} + d(\text{CP})$$

$$= 0.89 + 0.1(0.1)$$

$$0.89 := 0.9$$

$$\Rightarrow \text{For } e : d = \text{Upper Limit} - \text{Lower Limit} = 0.83 - 0.8$$

$$= 0.03$$

$$\text{For } e, \text{ Range } e = \text{lower limit} + d(\text{CP})$$

$$0.8 := 0.8 + 0.03(0.3)$$

$$0.8 := 0.809$$

$$\text{Range } n = \text{lower limit} + d(\text{CP})$$

$$0.809 := 0.809 + 0.03(0.3)$$

$$0.809 := 0.818$$

$$\text{Range } + = 0.818 + 0.03(0.2)$$

$$0.809 := 0.824$$

$$\text{Range } W = 0.824 + 0.03(0.1)$$

$$0.824 := 0.827$$

$$\text{Range } \cdot = 0.827 + 0.03(0.1)$$

$$0.827 : = 0.83$$

\Rightarrow For n : $d = \text{Upper} - \text{lower}$

Limit Limit

$$= 0.818 - 0.809 = 0.009$$

$$\text{Range } e = 0.809 + 0.009(0.3)$$

$$0.809 : = 0.8117$$

$$\text{Range } n = 0.8117 + 0.009(0.3)$$

$$0.8117 : = 0.8144$$

$$\text{Range } t = 0.8144 + 0.009(0.2)$$

$$0.8144 : = 0.8162$$

$$\text{Range } w = 0.8162 + 0.009(0.1)$$

$$0.8162 : = 0.8171$$

$$\text{Range } \cdot = 0.8171 + 0.009(0.1)$$

$$0.8171 : = 0.818$$

$$\Rightarrow \text{For } t : d = \text{Upper Limit} - \text{Lower Limit} \\ = 0.8162 - 0.8144 = 0.0018$$

$$\text{Range } e = 0.8144 + 0.0018(0.3) \\ 0.8144 : = 0.8194$$

$$\text{Range } n = 0.8194 + 0.0018(0.3) \\ 0.8194 : = 0.82548$$

$$\text{Range } t = 0.82548 + 0.0018(0.2) \\ 0.82548 = 0.82584$$

$$\text{Range } w = 0.82584 + 0.0018(0.1) \\ 0.82584 : = 0.82602$$

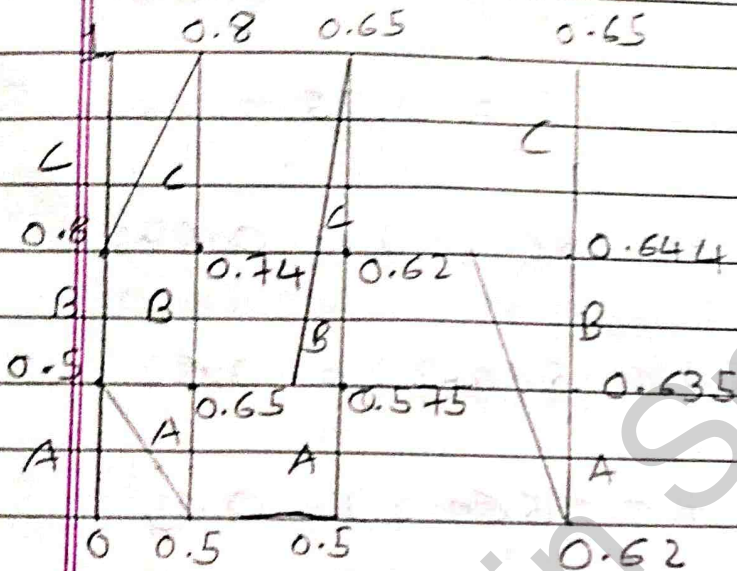
$$\text{Range } \cdot = 0.82602 + 0.0018(0.1) \\ 0.82602 : = 0.8262$$

Hence, Range of cordword is

$$0.82602 < \text{went.} < 0.8262$$

$$\text{Tag} = \frac{0.82602 + 0.8262}{2} = 0.82611$$

6 Decode the given tag value = 0.63725 Using Arithmetic Coding for the alphabetic sequence with the given probabilities $P(A) = 0.5$, $P(B) = 0.3$ and $P(C) = 0.2$



Here, Tag values lies between,
 $0.5 < 0.63725 < 0.8$

So, We have to take B as First character letter.

-> For B, $b = 0.8 - 0.5 = 0.3$

For A - Range = $0.5 + 0.3(0.5) = 0.65$

Range B = $0.65 + 0.3(0.3) = 0.74$

Range C = $0.74 + 0.3(0.2) = 0.8$

After that, Tag Value lies between,

$$0.5 < 0.63725 < 0.65$$

So, we have to take A letter.

$$\rightarrow \text{For A : } d = 0.65 - 0.5 = 0.15$$

$$\text{Range A : } 0.5 + 0.15(0.5) = 0.575$$

$$\text{Range B : } 0.575 + 0.15(0.3) = 0.62$$

$$\text{Range C : } 0.62 + 0.15(0.2) = 0.65$$

After that, Tag value lies between, $0.62 < 0.63725 < 0.65$

So, we have to add C letter

$$\rightarrow \text{For C : } d = 0.65 - 0.62 = 0.03$$

$$\text{Range A} = 0.62 + 0.03(0.5) = 0.635$$

$$\text{Range B} = 0.635 + 0.03(0.3) = 0.644$$

$$\text{Range C} = 0.644 + 0.03(0.2) = 0.65$$

So, Between, 0.644 and 0.65
we get Tag value.

So, Sequence: BAC

6 Define The Tunstall Code.

=> Tunstall Coding is a variable-length entropy coding algorithm used for Lossless data compression.

Here, $N \neq kCN - 1) \leq 2^n$

Where, N = Unique Character
of Given String

n = Total Number
of Tunstall Code
Given String

Here, Using k value we have
to show the Number of Step
of Tunstall Code.

If $k = 2$, than we have to show
Two Step Tunstall Code.

\Rightarrow Generate The Tunstall Code For
 $P(A) = 0.6$, $P(B) = 0.3$, $P(C) = 0.1$
 and $n = 3$ -bit.

Here, We have to take $N = 3$ and
 $n = 3$

$$\therefore N + K(N-1) \leq 2^n$$

$$\therefore 3 + K(3-1) \leq 2^3$$

$$\therefore 3 + 2K \leq 8$$

$$\therefore 2K \leq 5$$

$$\therefore K \leq 2$$

Here, We have to Perform Two
 Step.

\Rightarrow $K = 1$ Step: $K = 2$ Step:

$$B = 0.3$$

$$C = 0.1$$

$$AB = 0.18$$

$$AA = 0.36$$

$$AC = 0.06$$

$$B = 0.3 \quad 000$$

$$C = 0.1 \quad 001$$

$$AB = 0.18 \quad 010$$

$$AAA = 0.036 \quad 011$$

$$AC = 0.06 \quad 100$$

$$AAB = 0.108 \quad 101$$

$$AAC = 0.006 \quad 110$$

7 Generate Tunstall code for $\{A, B, C\}$ with the probabilities $P(A) = 0.7$, $P(B) = 0.2$, $P(C) = 0.1$ and $n = 3$ bits.

Here, we have to take $N = 3$, $n = 3$

$$\therefore N + K(N-1) \leq 2^n$$

$$\therefore 3 + K(3-1) \leq 2^3$$

$$\therefore 3 + 2K \leq 8$$

$$\therefore 2K \leq 5$$

$$\therefore K \leq 2$$

Here, we have to perform two steps.

\Rightarrow $K = 1$ Step:

$$B = 0.2$$

$$C = 0.1$$

$$AA = 0.49$$

$$AB = 0.14$$

$$AC = 0.07$$

$K = 2$ Step:

$$B = 0.2$$

$$C = 0.1$$

$$AA = 0.49$$

$$AB = 0.14$$

$$AAC = 0.049$$

$$ACB = 0.014$$

$$ACC = 0.007$$

8 Define Golomb Code:

⇒ Steps of Create Golomb Code:

1 Find Unary Code
of $n = \lfloor n/m \rfloor \lfloor n/m \rfloor$

Where $n, m =$ Positive Integer

2 Find, $k = \lceil \log_2 m \rceil$

$$c = 2^k - m$$

$$r = n \bmod m$$

IF, $r' = \begin{cases} \text{if } r \leq c \rightarrow k-1 \text{ bit} \\ \text{else,} \\ r+c \text{ to } k \text{-bit} \end{cases}$

3 Concat Step 1 and Step 2.

=> Find Golomb Code For $m=5$ to $n=0$ to 15.

n	m	n/m	q	r	r'	Code
0	5	0	0	0	2-bit	000
1	5	0	0	1	2-bit	001
2	5	0	0	2	2-bit	010
3	5	0	0	3	3-bit $\rightarrow r+c=3$	0110
4	5	0	0	4	3-bit $\rightarrow r+c=7$	0111
5	5	1	10	0	2-bit	1000
6	5	1	10	1	2-bit	1001
7	5	1	10	2	2-bit	1010
8	5	1	10	3	3-bit	10110
9	5	1	10	4	3-bit $\rightarrow r+c=7$	10111
10	5	2	110	0	2-bit	11000
11	5	2	110	1	2-bit	11001
12	5	2	110	2	2-bit	11010
13	5	2	110	3	3-bit	110110
14	5	2	110	4	3-bit $\rightarrow r+c=7$	110111
15	5	3	1110	0	2-bit	111000

$$\text{For, } k = \lceil \log_2 m \rceil = \lceil \log_2 5 \rceil = 3$$

$$c = 2^k - m = 2^3 - 5 = 3$$

Q What are the difficulties in Arithmetic Coding?

=> This are the difficulties in Arithmetic Coding.

1 Precision Issues:

- Arithmetic Coding involves manipulating real numbers, which can lead to precision issues in computer.

2 Probability Modeling:

If Probability distribution is not properly estimated than compression efficiency can suffer.

3 Adaptability:

Arithmetic Coding is less adaptive compared to some other compression Algorithm.

4 Complexity:

The encoding and decoding algorithm for arithmetic coding are more complex to other coding method.

5 Decoding Speed:

Speed of Arithmetic decoding is crucial compare to other coding method.