Sardar Vallabhabhai Patel Institute of Technology ,Vasad
IT department
Subject: Data compression and Data Retrieval
Subject code: 2161603
TYIT

Assignment No. 1

1. Define Data Compression.
2. Data Compression = Modeling + Coding Explain.
3. What is Entropy? How it affects the compression rate?
4. Check whether following code are Uniquely decodable ?

   a) { 0,11,01,111 } b) { 1,10,110,111 }

Sardar Vallabhbhai Patel Institute of Technology ,Vasad
IT department
Subject: Data compression and Data Retrieval
Subject code: 2161603
TYIT

Assignment No. 2

1. Define Self Information
2. Define average length of code
3. Define Redundancy
4. Given the probability model in Table 4.9, find the real valued tag for the sequence a1 a1 a3 a2 a3 a1.

TABLE 4.9 Probability model for Problems 5 and 6.

Letter	Probability
a_1	.2
a_2	.3
a_3	.5

5. A sequence is encoded using the LZ77 algorithm. Given that C(a) = 1, C(b) = 2, C(r) = 3, and C(t)= 4 .
   (i) decode the following sequence of triples:
   <0, 0, 3>, < 0, 0, 1>, <0, 0, 4>, < 2, 8, 2>, < 3, 1, 2>, <0, 0, 3>,
   <6, 4, 4>, <9,5, 4>
   Assume that the size of the window is 20 and the size of the look- ahead buffer is 10.

6. A sequence is encoded using the LZW algorithm and the initial dictionary is shown in table below:

Index	entry
1	B
2	a
3	h
4	i
5	s
6	t

The output of LZW encoder is the following sequence:

6 , 3 , 4 , 5 , 2 , 3 , 1 , 6 , 2 , 9 , 11 , 16 , 12 , 4 , 20 , 10 , 8 , 23 ,13

Decode this sequence.

Sardar Vallabhabhai Patel Institute of Technology ,Vasad
IT department
Subject: Data compression and Data Retrieval
Subject code: 2161603
TYIT

Assignment – 3

1. Explain Huffman Coding in detail. Define minimum variance Huffman Codes.

2. List out rules of Huffman Coding with an example of

Symbol	Count
A	14
B	7
C	5
D	5
E	4

3. Encode “PPQRXPQY” using Adaptive Huffman Code. Derive Output string, Codes and final tree.

4. Generate GOLOMB code for m=9 and n=8 to 13.

5. Write procedure to generate TUNSTALL code. Generate TUNSTALL code with probability of P(A) = 0.3, P(B) = 0.1, P(C) = 0.6 and n = 3 bits.

6. Explain Burrows – Wheeler Transform algorithm with suitable example.

Sardar Vallabhabhai Patel Institute of Technology ,Vasad
IT department
Subject: Data compression and Data Retrieval
Subject code: 2161603
TYIT

Assignment No. 4

1. Explain challenges in XML retrieval
2. Explain Tokenization
3. Write a short note on scalar quantization
4. Write a short note on Vector Quantization