Seat No.: _____ Enrolment No.

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER– VI EXAMINATION – SUMMER 2020

Subject Code: 2161603 **Date:28/10/2020**

Subject Name: DATA COMPRESSION AND DATA RETRIVAL

Time: 10:30 AM TO 01:00 PM **Total Marks: 70**

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1	**(a)**	Define following terms: i. Compression Ratio ii. Instantaneous Code iii. Prefix Code	**03**
	(b)	What is Data Compression? Explain types of Data compression and measure of performance of data compression.	**04**
	(c)	For symbol A, B, C, D, E, F, G and H respectively with probability 0.2, 0.18, 0.16, 0.14, 0.10, 0.10, 0.08, 0.04. Find out Huffman code, Source Entropy, Average Length, Code efficiency and Redundancy.	**07**
Q.2	**(a)**	Generate GOLOMB code for m=5 and n=5 to 10.	**03**
	(b)	Explain modeling and coding. Explain how this will help to reduce entropy for following data. 9,11,11,11,14,13,15,17,16,17,20,21.	**04**
	(c)	Encode "aardvark" using Adaptive Huffman code. Derive Output string, Codes and final tree.	**07**
		OR	
	(c)	Explain Rice code with example.	**07**
Q.3	**(a)**	Write a short note on skip pointer.	**03**
	(b)	Generate TUNSTALL code P(A)=0.4, P(B)=0.3, P(C)=0.2 , P(D)=0.1 and n=4 bits.	**04**
	(c)	Encode and decode BILL GATES using arithmetic coding.	**07**
		OR	
Q.3	**(a)**	Write a short note on Prefix Code.	**03**
	(b)	What is Uniquely Decodable Code? Determine whether the following codes are uniquely decodable or not. i. {0,01,11,111} ii. {0,10,110,111} iii. {1,01,010,111}	**04**
	(c)	Encode the sequence this␣bis␣bth using prediction with partial match(PPM).	**07**
Q.4	**(a)**	Explain Vector Quantization in detail.	**03**
	(b)	Encode and decode this␣bis␣bthe using The Burrows Wheeler Transform	**04**
	(c)	Use LZ78 to encode the following string wabba#wabba#wabba#wabba#woo#woo#woo	**07**
		OR	
Q.4	**(a)**	Differentiate between Static and Dynamic Dictionary	**03**
	(b)	Encode L= ssht␣th␣ii␣be using Move to Front Coding.	**04**

(c) Decode the following string using LZW **07**
5 2 3 3 2 1 6 8 10 12 9 11 7 16 5 4 4 11 21 23 4
Consider the following initial dictionary

Initials	Dictionary
1	#
2	a
3	b
4	o
5	w

Q.5 **(a)** Explain Tokenization. **03**

(b) Explain and compare Incident matrix and Inverted index with example. **04**

(c) Explain Lemmatization and Stemming in detail. **07**

OR

Q.5 **(a)** Write a short note on stop word removal. **03**

(b) Explain process generating triple in all three possible cases of LZ77 algorithm. **04**

(c) Explain challenges in XML information retrieval **07**

****