SARVAJANIK UNIVERSITY
INCLUSIVE | INTEGRATED | INNOVATIVE

SARVAJANIK UNIVERSITY
Sarvajanik College of Engineering and Technology
Bachelor of Technology

Year: B. Tech III (Semester VI)

Subject Name: Data Compression **Subject Code:** BTIT14608
Type of course: Professional Elective Course
Prerequisite (if any): Programming for Problem Solving, Data Structures

Rationale: Information is generated and used in digital form in the form of numbers represented by bytes of data. Number of bytes required to represent multimedia data can be huge. Given the explosive growth of data that needs to be transmitted and stored, compression techniques need to be used . The objective behind this course is to aware students about various algorithms for compression of text files as well as image files.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments / projects / open book tests / closed book tests) CA2: Sincerity in attending classes / class tests / timely submissions of assignments / self-learning attitude / solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work / Quality of work submitted / Active participation in lab sessions / viva on practical skills learned in course.

Contents:

Sr. No.	Contents	Total Hrs
1.	**Introduction:** Basic Compression Techniques- Ad Hoc Text Compression, Run-Length Encoding (RLE) for text compression, RLE for image compression, Move-to-front Coding, Modeling and Coding, Lossy vs. Lossless compression Mathematical Preliminaries for Lossless Compression, Models – Physical Models, Probability Models, Markov Models Coding, Uniquely Decodable Codes, Prefix codes	05
2.	**Huffman coding:** The Huffman Coding Algorithm, Minimum variance Huffman codes, Adaptive Huffman coding – Update Procedure, Encoding Procedure, Decoding Procedure, Golomb Codes, Rice codes, Tunstall Codes, Applications of Huffman Coding – Lossless Image compression, Text compression, Audio Compression	10
3.	**Arithmetic coding:** Coding a sequence, Generating a Tag, Deciphering the Tag, Generating Binary Code, Uniqueness and Efficiency of the Arithmetic code, Algorithm implementation, Integer Implementation, Comparison of Huffman and Arithmetic coding, Applications	06

4.	**Dictionary Techniques:** Static Dictionary, Diagram Coding, Adaptive Dictionary – LZ77, LZ78, LZW, LZMW, LZAP, Applications – Image compression	**10**
5.	**Lossless Image Compression:** The Old JPEG Standard, CALIC, JPEG-LS	**06**
6.	**Lossy Data Compression:** Distortion criteria – The Human Visual System, Auditory Perception, Models – Probability Models, Linear System Models, Physical Models, Scalar Quantization- The Quantization Problem, Uniform Quantizer, Adaptive Quantization – Forward Adaptive , Backward Adaptive, Non uniform Quantization – pdf optimized Quantization, Companded Quantization, Entropy Coded Quantization – Entropy coding of Lloyd, Max Quantizer Outputs Vector Quantization- Advantages of Vector Quantization over Scalar Quantization The Linde-Buzo-Gray Algorithm, Tree structured Vector Quantization, Structured Vector Quantization	**08**

Suggested Specification table with Marks (Theory): (For B. Tech only)

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
15	20	20	5	-	-

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (Revised Bloom's Taxonomy)

Reference Books:

Sr No.	**Title of book /article**	**Author(s)**	**Publisher and details like ISBN**
1	Introduction to Data Compression	Khalid Sayood	Morgan Kaufmann
2	Data Compression The Complete Reference	David Salomon	Springer
3	The Data Compression Book	Mark Nelson	John Wiley & Sons
4	Data Compression: Methods and Theory	James A. Storer	Computer Science Press

Note: Students should refer to the latest editions of books

Course Outcomes (CO):

Sr. No.	**CO statements**	**Marks % weightage**
CO-1	Explain basic compression schemes such as Ad Hoc Text Compression and RLE.	20%

CO-2	Describe Mathematical Preliminaries involved in different compression (Lossy and Lossless) techniques.	20%
CO-3	Explain and differentiate various Lossless compression algorithms as well as Lossy compression algorithms.	30%
CO-4	Use different Loseless or Lossy compression algorithm to compress the given text as well as image data.	30%

Suggested List of Experiments:

Sr. No **Statements**

1. For any given input file (containing sequence of symbol/letters), compute the frequency, probability, and information content of each symbol and present them in tabular format. Also find average length as well as entropy of the code given in the file.
2. Take set of symbols and associated code. Check whether given code is Prefix or not. Also check whether the code is UDC or not.
3. Encode a given input file containing message with the English (lowercase) alphabets using "Huffman algorithm" and show the results as follows.

Letter	Probability	Code

4. Generate the Golomb Codes for input parameter m=5 and for the set of integers n={0,1,2,…,15}. Display the result as follows.

n	q	r	Codeword

5. Encode a given string "BILL GATES" using Arithmetic Encoding/Decoding scheme.
6. Write a program to implement digram coding for given text file.
7. Encode a given input file using dictionary method "LZ77" and store the result in file. Also perform decoding on the same file.
8. Encode a given input file using dictionary method "LZ78" and store the result in file.
9. Encode a given input file using dictionary method "LZW" and store the result in file.