

Branch Name:	Information Technology
Semester/Year:	Semester VI / Third Year
Subject Title:	Data Compression and Data Retrieval
Subject Code:	1ET1050601
Pre-requisite:	Basic knowledge of Programming.

Course Objective:

- To understand working of different compression techniques.
- To study about information theory.
- To understand working of various quantizers.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
03	-	02	04	70	30	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weightage (%)
1	Introduction to Data Compression: Definition, Compression Techniques: Lossless Compression, Lossy Compression, Measures of Performance, Data Compression = Modeling + Coding.	04	5
2	Introduction to Information Theory: Probability, Entropy, Average Information, Average Length, Efficiency, Redundancy, Uniquely Decodable Codes, Prefix Codes, Instantaneous Codes	04	10
3	Huffman Coding: The Shannon-Fano Algorithm, The Huffman Coding Algorithm, Minimum Variance Huffman Codes, Adaptive Huffman Coding, Update Procedure, Encoding Procedure, Decoding Procedure, Building the Tree, Golomb Codes, Rice Codes, Tunstall Codes, Applications of Huffman Coding	07	20
4	Arithmetic Coding: Introduction, Coding a Sequence, Generating a Tag, Deciphering the Tag. Generating a Binary Code.	03	15
5	Dictionary-Based Compression: Static Dictionary, Digram Coding, Adaptive Dictionary, The LZ77 Approach, The LZ78 Approach, The LZW Approach, Applications. File Compression, Image Compression- GIF, PNG	06	20
6	Predictive Coding: Prediction with Partial match (ppm): The basic algorithm, The ESCAPE SYMBOL, Length of context, The Exclusion Principle, The Burrows-Wheeler Transform: Move-to-front coding, Facsimile Encoding.	05	10
7	Scalar Quantization: The Quantization Problem, Uniform Quantizer, Adaptive Quantization, Forward Adaptive Quantization, Backward		

	Adaptive Quantization, Nonuniform Quantization, pdf-Optimized Quantization, Companded Quantization.	05	10
8	Vector Quantization: Advantages of Vector Quantization over Scalar Quantization, The Linde-Buzo-Gray Algorithm, Tree-Structured Vector Quantizers, Structured Vector Quantizers.	05	10

Course Outcome:

After learning the course, the students will be able to:

1. Understand Encoding and Decoding.
2. Understand Compression and decompression.
3. Differentiate between Lossy and Lossless compression techniques.
4. Understand basic concept of information retrieval.

List of Text Books:

1. Introduction to Data Compression, Khalid Sayood, Morgan Kaufmann Publishers .
2. The Data Compression book, Mark Nelson, Jean Loup Gailly.
3. Data Compression : The Complete Reference”, David Saloman, Springer

List of Suggested titles of Experiments (If Any):

1. Implement a program entropy which takes text message as the argument and print frequency of letter, probability & entropy of symbol and the average length of the code. Also calculate Redundancy of code.
2. Write a program to check Prefix Property of given set of codes.
For example:
Input codes: (0, 10, 011, 110, 1111) then output: true
Input codes: (9, 59, 95, 555) then output: false, because of code 9 is a prefix of 95.
3. Write a program to check whether set of codes are uniquely decodable or not.
4. Write a program to generate binary code in case of arithmetic coding.
5. Design and implement to Encode a message using Shannon-Fano Code.
6. Implement Huffman Code to generate binary code when symbol and probabilities are given.
7. Design and Implement Diagram Coder for text file.
8. Design and Implement LZ77 Sliding Window Coder.
9. Design and Implement LZ78 Sliding Window algorithm.
10. Design and Implement LZW Sliding Window algorithm.

e-Resources (If any):

1. <http://nptel.ac.in/>
2. http://ocw.usu.edu/Electrical_and_Computer_Engineering/Information_Theory/