ADVANCED DATA COMPRESSION TECHNIQUES

SLOT: E2+TE2

TEAM MEMBERS:

MYTHILI ANUKEERTHANA KS-22MID0011

VARSHITHA V-22MID0105

LEKKALA TEJASWI-21MIC0113

TOPIC: ZSTANDARD BASED SHORT MESSAGE COMPRESSION TECHNIQUE

ABSTRACT:

With the exponential growth of digital communication, the volume of text exchanged through SMS, chat applications, and online messaging platforms has increased significantly. This surge highlights the need for efficient methods to store, transmit, and manage textual data without compromising speed or reliability. Data compression provides a practical solution by reducing redundancy and minimizing storage and bandwidth usage.

This project presents an optimized SMS text compression system based on the **Zstandard (Zstd)** algorithm—a modern, high-performance compression method developed by Facebook. Zstandard is known for its excellent trade-off between compression ratio, speed, and low memory consumption, making it ideal for short-text scenarios like SMS.

At the algorithmic level, Zstandard combines two powerful techniques:

- **LZ77 Pattern Matching:** Identifies repeated substrings in the SMS text and replaces them with compact *match-length* (ML) and *match-offset* (MO) tokens. This step captures redundancy efficiently.
- **FSE (Finite State Entropy) Coding:** A modern entropy coding method used in place of Huffman coding. FSE compresses the token stream into a highly compact bit sequence using state-driven probability modeling, achieving superior efficiency and speed.

Our implementation integrates:

- A **Flask backend**, responsible for performing compression and decompression operations using Zstd's LZ77 + FSE pipeline.
- A **React-based frontend**, which provides a clean interface for entering SMS text, viewing compressed output, and analyzing performance metrics such as processing time, size reduction, and compression ratio.

Experimental evaluation shows that Zstandard can achieve **60–70% compression efficiency** for typical SMS messages, outperforming legacy compressors like Gzip and Bzip2 in both speed and ratio—particularly in short, repetitive text environments.

This project demonstrates how advanced modern algorithms like LZ77 and FSE can be effectively applied to real-world communication systems, offering a powerful solution for optimized SMS storage and transmission.