|  |  |
| --- | --- |
|  |  |
| LZ78 Compression Technique |  |
|  |  |
|  | DSAI 325Information Theory |
|  | Mohamed Ehab Yousri202201236 |

.

LZ78 Compression Technique

LZ78 is a **lossless data compression algorithm** that compresses data in such a manner that no information is lost which replaces repeating sequences within the data with references to their initial occurrence. This process creates a dictionary of previously seen phrases, which in turn helps reduce the overall size of the data.

**Implementation Details:**

**Compression Algorithm:**

• Maintains a registry listing all previously encountered substrings.

• Reads the character one at a time to form a substring that hasn’t been encountered before.

• Records all substrings as pairs (Index, Next Character): o The Index denotes the entry in the dictionary of the longest matched phrase. The Next Character is the subsequent character that follows the prefix.

• Inserts the new substrings into the dictionary.

• Keeps doing so until the entire text has been compressed.

**Decompression Algorithm:**

The decompression goes about as follows:

• Reads the pair of values (Index, Next Character) from the files containing compressed data.

• Makes use of the dictionary to piece the text back together.

• Appends the Next Character to the previously referenced entry in the dictionary.

• Modifies the dictionary while the text is put back together.

• Continues until the full original text has been reconstructed

**The Main Components:**

**• Dictionary - set to memorize substrings that have already been referenced to.**

**• Index References – serves to give new substrings a set index for retrieval.**

The Algorithm Implementation:

Compression Algorithm:

A screen shot of a computer code

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

Decompression Algorithm:

A screen shot of a computer program

AI-generated content may be incorrect.

Main :

A screen shot of a computer program

AI-generated content may be incorrect.

Test Cases:

Test Case 1: "ABCDABCABCDAABCABCE"

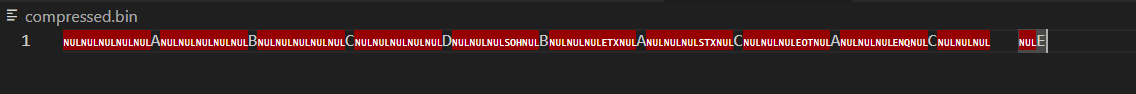
Original:

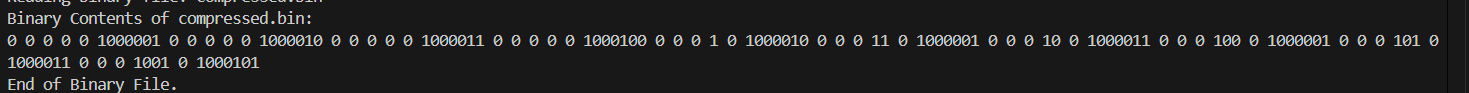
A black background with white letters

AI-generated content may be incorrect.

**Original Size:** 19 characters × 8 bits = 152 bits

Compressed:





**Compressed Size:** 10 tags × 12 bits = 120 bits

Extra: the tags:

A screenshot of a computer

AI-generated content may be incorrect.

**Compression Ratio:** 152 / 120 = 1.27

Test Case 2: "ABABABA":

Original:

A black background with white text

AI-generated content may be incorrect.

**Original Size:** 7 characters × 8 bits = 56 bits

Compressed:

A screenshot of a computer

AI-generated content may be incorrect.

A number on a black background

AI-generated content may be incorrect.

**Compressed Size:** 4 tags × 12 bits = 48 bits

Extra: the tags:

A screenshot of a computer

AI-generated content may be incorrect.

**Compression Ratio:** 56 / 48 = 1.17

The Full Running Process:

A screenshot of a computer

AI-generated content may be incorrect.

References

<https://youtu.be/serGhL1ZMP8?feature=shared>

<https://www.stringology.org/DataCompression/lz78/index_en.html>

<https://hackernoon.com/how-lz78-compression-algorithm-works-x7103tlm>

<https://www.youtube.com/watch?v=Gv6WsNIogbE>