

Lempel-Ziv-Welch current implementation

(In this project)

Encoding :

Lets say we have the following string: AABABCCABC

In order to encode the string here is how we must proceed:

Then from our string: AABABCCABC

Find the longest codeword in the dictionary:

Instantiate a dictionary with all the different unique char in order that they can be found:

- AABABCCABC, index: 0

Concatene the codeword and the next value to the dictionary.

Repeat for the remaining symbols:

- AABABCCABC, index: 0
- AABABCCABC, index: 1
- AABABCCABC, index: 4
- AABABCCABC, index: 2
- AABABCCABC, index: 2
- AABABCCABC, index: 6

index	codeword
0	A
1	B
2	C
3	AA
4	AB
5	BA
6	ABC
7	CC
8	CA

Done!

We now have the encoded value 0014226.

$$\text{compression ratio} = \frac{\text{uncompressed size}}{\text{compressed size}} = \frac{10}{7} \approx 1.43 : 1$$

Decoding

Let's decode `0014226` back to its original value.

Then from: `0014226`

Write the value in the index:

- `0014226`, codeword: `A`

If we have a previous codeword, we concatenate it with the first symbol of the current codeword and add it to the dictionary.

Repeat for the remaining indexes:

- `0014226`, codeword: `A`
- `0014226`, codeword: `B`
- `0014226`, codeword: `AB`
- `0014226`, codeword: `C`
- `0014226`, codeword: `C`
- `0014226`, codeword: `ABC`

Done!

Instantiate a dictionary from the same previous unique symbol:

NOTE: Symbols in `red` indicate the previous codeword to which the current symbol is added.

index	codeword
0	A
1	B
2	C
3	AA
4	AB
5	BA
6	ABC
7	CC

We now have the decoded value `AABABCCABC`.

The end 🎊🎊