

# Sliding Window Compressor

By Zijie Wu

This program compresses the input on a byte by byte basis.

## How to run

---

Download/clone the repo, `cd` to project's root directory.

### In terminal

Size of sliding window is optional, 9 means 4K, 13 means 64K

```
# Compress file
python cli.py [--size {9, 13}] zip path/to/file
```

```
# Decompress file
python cli.py unzip path/to/file
```

### In Python scripts

```
from Compressor.compressor import Compressor

c = Compressor()
c.compress_to_file(in_file, out_file)
c.decompress_to_file(in_file, out_file)
```

## How it works

---

This program compresses data on a byte by byte basis. For each substring found in sliding window, a pointer will be encoded to represent the substring.

Each pointer is represented by variable number of bytes (set by parameter passed in, and don't change during the compression).

The structure of pointer is as follows:

```
# For example
# For a pointer with size of 3 bytes:
# The first byte is an escape character, to identify the pointer
# The remaining bytes are used to encode offset and length data
# If the number of bits for offset is 10 (this is set by user), then the number of bits for length is  $(2 * 8 - 10) = 6$ .

# Offset: 11111111 11 = 2047 (bytes)
# Length: 1111 = 15 (bytes)

# In binary:
# 11001100 11111111 11111111
# In hex:
bytes(b"\xCC\xff\xff")
```

For literal with the same value as escape character (0xCC), it will be encoded as `0xCC 0x00 0x00 ...`.

It uses bytearray to act as sliding window and buffer.