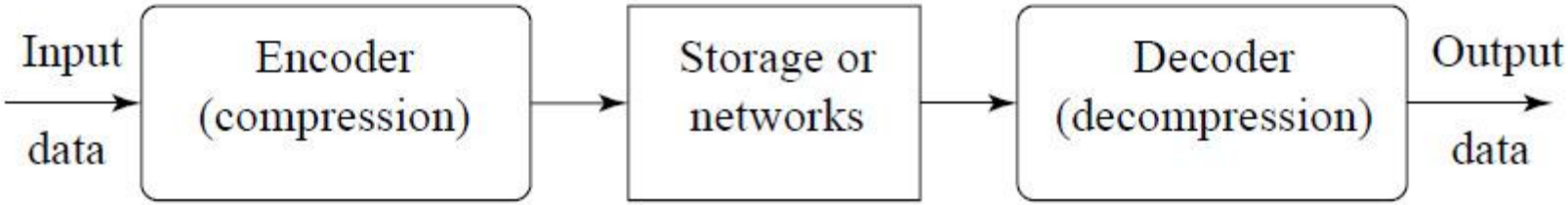
Loseless Compression

Definition

- Compression: the process of coding that will effectively reduce the total number of bits needed to represent certain information.



Compression ratio

- **Compression ratio:**

$$Compression\ Ratio = \frac{B0}{B1}$$

- *B*0 { number of bits before compression
- *B*1 { number of bits after compression

File Sizes

- In a text document that uses ASCII code to represent text characters, each byte stores an ASCII code that corresponds to a character.
- The more characters in a text document, the more bytes are required to store the file.
- Digital media files (image, sound, and especially video files) can be very large.

Disadvantages of Large File Size

- take longer to copy the file from one computer to another
- take longer to send the file over the Internet
- take longer to process (such as during opening and saving) the file

Strategies to Reduce Digital Media File Size

- Reduce the sampling rate
- Reduce the bit depth
- Apply file compression

Reduce Sampling Rate

- Recall the weighing puppy scenario
- If you weigh the puppy more frequently, it will take more paper.
- For digital media files, higher sampling rate means more data to store.
- In other words, lower sample rate will produce less data, i.e. smaller file size.

Reduce Bit Depth

- Bit depth refers to the number of allowable levels you can map the data
- For digital media files, lower bit depth means less data to store.

Compression

- File compression means techniques to reduce file size
- Two categories in terms of whether the data get lost during the compression:
  - lossy compression
  - lossless compression

Lossy Compression

- Some data will be lost and cannot be recovered

- Examples:
  - JPEG compression for images
  - MP3 for audio
  - most compressors for videos

Lossy Compression

- Avoid using lossy compression (if possible) when you want to keep the file for further editing.
- Generally, you can do so with images and audio.
- Video files are generally so large that it is inevitable to save them with lossy compression.

Trade-offs of Reducing File Size

Data will be lost or altered when you apply these strategies:

- reduce sampling rate
- reduce bit depth
- apply lossy compression

When data is lost or altered, you sacrifice the exactness of the media original information. This affects the quality of the media.

Weighing the Trade-offs

- Depend on projects and intended use of the files
- Weigh the file size (storage requirement and speed of transfer and processing of the file) against the quality of the digital media files
- Losing data vs. "perceivable" quality
  - Sometimes it may be acceptable if losing data does not cause "perceivable" deterioration in quality

Example: MP3

- MP3 audio uses a lossy compression.
- It reduces the file size by selectively removing and altering the audio data (such as certain ranges of audio frequencies) that are not very perceivable by human.

Review Question

Our everyday decimal numbering system is base-_____.
Computers use base-_____ , which is also known as the _____ numbering system.

Review Question

This is a quote from a T-shirt:

> "There are only 10 types of people in the world: Those who understand binary and those who don't."

How many types does this quote actually mean?

(Hint: It is talking in binary!)

Review Question

(i) Name three general strategies to reduce the size of a digital media file.

(ii) Which of these strategies does not necessarily sacrifice the quality of the media file?

Review Question

In general, if you want to keep a digital media file for further editing, you should _____.

A. avoid applying lossy compression

B. avoid applying lossless compression

Dictionary-based Coding (LZW)

- LZW uses fixed-length codewords to represent variable-length strings of symbols/characters that commonly occur together, e.g., words in English text.
- the LZW encoder and decoder build up the same dictionary dynamically while receiving the data.
- LZW places longer and longer repeated entries into a dictionary, and then emits the *code* for an element, rather than the string itself, if the element has already been placed in the dictionary.

LZW Compression Algorithm

```
BEGIN
    s = next input character;
    while not EOF
        { c = next input character;
            if s + c exists in the dictionary
                s = s + c;
            else
                { output the code for s;
                add string s + c to the dictionary with a new code;
                s = c;
                }
        }
    output the code for s;
END
```

Example of LZW

- LZW compression for string “ABABBABCABABBA“
- Let's start with a very simple dictionary (also referred to as a “string table"), initially containing only 3 characters, with codes as follows:

Code	String
1	A
2	B
3	C

- Now if the input string is “ABABBABCABABBA", the LZW compression algorithm works as follows:

LZW Compression

s	c	output	code	string
			1	A
			2	B
			3	C
A	B	1	4	AB
B	A	2	5	BA
A	B			
AB	B	4	6	ABB
B	A			
BA	B	5	7	BAB
B	C	2	8	BC
C	A	3	9	CA
A	B			
AB	A	4	10	ABA
A	B			
AB	B			
ABB	A	6	11	ABBA
A	EOF	1		

- The output codes are: 1 2 4 5 2 3 4 6 1. Instead of sending 14 characters,
- only 9 codes need to be sent (compression ratio = 14/9 = 1.56).

LZW Decompression Algorithm

```
BEGIN
    s = NIL;
    while not EOF
    {
        k = next input code;
        entry = dictionary entry for k;
        output entry;
        if (s != NIL)
          add string s + entry[0] to dictionary with a new code;
        s = entry;
    }
END
```

LZW Decompression

s	k	entry/output	code	string
			1	A
			2	B
			3	C
NIL	1	A		
A	2	B	4	AB
B	4	AB	5	BA
AB	5	BA	6	ABB
BA	2	B	7	BAB
B	3	C	8	BC
C	4	AB	9	CA
AB	6	ABB	10	ABA
ABB	1	A	11	ABBA
A	EOF			

Task to do

- By using the LZW compression algorithm, please compress the following string:
  1. >>>>>*>>*>>*>>>
  2. QQQ QQ QQAA AQQA
  3. UQAUQAUQAUQA
  4. 123K321K231K123K

- You are required to produce:
  1. String table
  2. Output string
  3. Compression ratio

Task to do 2

1. Create a string of 16 characters and post it.
2. A peer will compress the string for you.
3. You are required to evaluate the answer.

Reference

- Fundamentals of Multimedia, Chapter 7, Li & Drew c, Prentice Hall 2003