CS435/Spring 2014/Handout: LZW

LZW encoding

This handout describes a simplified version of LZW (Lempel-Ziv-Welch) encoding, a common data compression method.

The input consists of a sequence of integers in the range 0–255, representing a stream of 8-bit characters. The output consists of a sequence of integers in the range 0-4095, corresponding to codes. The encoder maintains a table with slots indexed 0–4095. Each entry is a list of integers. Initially slots 0–257 contain a list containing only the index. Indices 0–255 correspond to single 8-bit characters, 256 is reserved for the code `CLEAR_TABLE`, and 257 for the code `EOD` (end of data).

The encoder works as follows:

1. Accumulate the longest sequence of one or more input characters matching a sequence already present in the table.
2. Output the code corresponding to that sequence.
3. Construct a new table entry for the first unused code. Its value is the sequence found in step 1 followed by the next input character.

When the table is full (and possibly before), the encoder issues a `CLEAR_TABLE` code, which resets the encoding table. When input is complete the encoder issues an `EOD` code.

Example

Suppose the input sequence is

```
45 45 45 45 45 65 45 45 45 66
```

The output sequence will be

```
256 45 258 258 65 259 66 257
```

generated as follows:

Input sequence	**Output code**	**Code added**	**Sequence represented by new code**
-	256 (`CLEAR_TABLE`)		
45	45	258	45 45
45 45	258	259	45 45 45
45 45	258	260	45 45 65
65	65	261	65 45
45 45 45	259	262	45 45 45 66
66	66		
	257 (`EOD`)		

LZW decoding

The decoder works similarly:

1. If the current free position is greater than 258, append the first element of the current sequence to the last table entry.
2. Output the sequence corresponding to the current code.
3. Make a new table entry consisting of a copy of the current sequence.

Here is a trace of decoding the previous example:

Input	**Code modified**	**Sequence**	**Output**
256			
45	258	45	45
258	258	45 45	45 45
	259	45 45	
258	259	45 45 45	45 45
	260	45 45	
65	260	45 45 65	65
	261	65	
259	261	65 45	45 45 45
	262	45 45 45	
66	262	45 45 45 66	66
	263	66	
257			