LZW code

Named after the proposers Abraham Lempel, Jacob Ziv, and Terry Welch, the Lempel-Ziv and Welch code is popularly known as the LZW code. It was published by Welch in 1984 as an improved version of LZ78 algorithm that was published by Lempel and Ziv in 1978. LZW is a lossless or error-free compression technique and is the core of the widely used Unix file compression utility 'compress'. This algorithm is also used in the GIF image format. The encoding and decoding steps of this algorithm are as given below:

LZW Encoding

Inputs:

I_s= *the sequence to be encoded*

$D(S_k, C(S_k))$= *Dictionary with (symbol, code) pairs*

Output:

C_s= *LZW code sequence*

Initialization:

$D(S_k, C(S_k))$= *all possible symbols with their corresponding codes (generally the ASCII characters and their corresponding values from 0-255)*

C_s=*Null*

I_s= *the sequence to be encoded*

$C=max(\{C(S_k)\})$

Step1: Read the first symbol from the input sequence and assign to S_w i.e.

S_w =the first symbol from I_s

Step2: Read the next symbol and assign to S_N i.e.

S_N =the Next symbol from I_s

Step3: Is S_N =Null ?

If Yes, goto step8 else step4

Step4: Temp = S_w + S_N

Step5: Is Temp available in dictionary?

If yes, S_w =Temp. go to step2

Else step6

Step6: C=C+1.

Add (Temp, C) to the dictionary

Add $C(S_w)$ to the output sequence C_s *i.e* $C_s = C_s + C(S_w)$

$S_w = S_N$

Go to step2

Step7. $C_s = C_s + C(S_w)$

Output C_s *as the coded sequence and* Stop

LZW Decoding

The beauty of LZW algorithm is though an extended dictionary is created during the compression step; this extended dictionary needs not to be stored or transmitted for decompression purpose. During decompression, the dictionary is once again reconstructed following the steps given below.

Initializations:

C_s*= the coded sequence*

D= Dictionary with all possible symbols (generally the ASCII characters) and their corresponding codes (from 0-255)

C_d= the Decoded character sequence=NULL

Step1: Read the first code word C_w

Step2: Add the string corresponding to the code word C_w to the decoded sequence C_d

Step3: $P_w = C_w$

Step4: C_w = Next code word

Step5: Is C_w present in dictionary?

If Yes,

i) Add the string corresponding to the code word C_w to the decoded sequence C_d

ii) $P = P_w$

iii) C=first character of C_w

iv) Add P+C to the dictionary and assign the next higher code to it.

If No,

i) $P = P_w$

ii) C=first character of P_w

iii) Add P+C to the dictionary and assign the next higher code to it.

iv) Add the string corresponding to the newly entered code to the decoded sequence C_d

Step6: Are there more code words in C_s?

If yes, go to step 3

Else Stop and output C_d as the decoded sequence.

Example 10: Let us encode/ decode the sequence ABAABABABABAB using LZW coding technique

Encoding:

Original Sequence: ABAABABABABAB

Initial dictionary entries: A=0; B=1

Current Symbol	Next symbol	Processed code	Dictionary entry	Code
NULL	A	-	-	-
A	B	0	AB	2
B	A	1	BA	3
A	A	0	AA	4
A	B	-	-	-
AB	A	2	ABA	5
A	B	-		
AB	A	-		
ABA	B	5	ABAB	6
B	A	-		
BA	B	3	BAB	7
B	A	-		
BA	B	-	-	-
BAB	NULL	7	-	-

The Processed LZW code is 0102537

Length of the original sequence = 13

Total number of bits in original code= 13*1=13

Length of the LZW code sequence =7

Total number of bits in LZW coded sequence =7*3=21

Decoding:

Coded sequence C_d=0102537
Dictionary d={(A,0), B(1)}

Current code	Available in Dictionary?	Output Symbol	Dictionary entry	Code
0	Y	A	-	-
1	Y	B	AB	2
0	Y	A	BA	3
2	Y	AB	AA	4
5	N	- ABA	ABA	5

3	Y	BA	ABAB	6
7	N	- BAB	BAB	7

The decoded sequence is: ABAABABABABAB

Example 11: Let us encode/ decode the sequence
BABARDARBABARDABARARAB using LZW coding technique

Encoding:

Original Sequence: BABARDARBARDABARARAB

Initial dictionary entries: A=0; B=1; D=2; R=3

Current Symbol	Next symbol	Processed code	Dictionary entry	Code
NULL	B	-	-	-
B	A	1	BA	4
A	B	0	AB	5
B	A	-	-	-
BA	R	4	BAR	6
R	D	3	RD	7
D	A	2	DA	8
A	R	0	AR	9
R	B	3	RB	10
B	A			
BA	B	4	BAB	11
B	A	-		
BA	R	-	-	-
BAR	D	6	BARD	12
D	A			
DA	B	8	DAB	13
B	A			
BA	R			
BAR	A	6	BARA	14
A	R			
AR	A	9	ARA	15

A	B			
AB	NULL	5		

The Processed LZW code is: 1043203468695

Decoding:

Coded sequence C_d=1043203468695
Dictionary d={(A,0), B(1), (D,2), (R,3)}

Current code	Available in Dictionary?	Output Symbol	Dictionary entry	Code
1	Y	B		
0	Y	A	BA	4
4	Y	BA	AB	5
3	Y	R	BAR	6
2	Y	D	RD	7
0	Y	A	DA	8
3	Y	R	AR	9
4	Y	BA	RB	10
6	Y	BAR	BAB	11
8	Y	DA	BARD	12
6	Y	BAR	DAB	13
9	Y	AR	BARA	14
5	Y	AB	ARA	15
NULL				

The decoded sequence is: BABARDARBABARDABARARAB

Review Questions:

1. Encode the following sequences using LZW code
   ABRAKADABRABABARDARBARDABARDARBAR
   BACHAN_CHHABAN_NANHABAHAN
   i) Decode back the coded sequences
   ii) Find the compression ratio and redundancy