**CS 240 Tutorial - LZW Example**

Given the string ABBCBBABBCCCCCAB and the original dictionary perform an LZW encoding and a decoding. Also calculate the compression achieved.

A,B, and C are in the original dictionary.

FINAL DICTIONARY

| Character | Code | Character | Code |
|---|---|---|---|
| A | 0000 | ABB | 1000 |
| B | 0001 | BCC | 1001 |
| C | 0010 | CC | 1010 |
| AB | 0011 | CCC | 1011 |
| BB | 0100 | CA | 1100 |
| BC | 0101 | | |
| CB | 0110 | | |
| BBA | 0111 | | |

**Encoding:**
**Take longest string W in Dictionary, output code, then ADD W+1 to dictionary**

| **Step** | Character | Code |
|---|---|---|
| 1 | W=A, add AB to dictionary | 0000 |
| 2 | W=B, add BB to dictionary | 0001 |
| 3 | W=B, add BC to dictionary | 0001 |
| 4 | W=C, add CB to dictionary | 0010 |
| 5 | W=BB, add BBA to dictionary | 0100 |
| 6 | W=AB, add ABB to dictionary | 0011 |
| 7 | W=BC, add BCC to dictionary | 0101 |
| 8 | W=C, add CC to dictionary | 0010 |
| 9 | W=CC, add CCC to dictionary | 1010 |
| 10 | W=C, add CA to dictionary | 0010 |
| 11 | W=AB, done. | 0011 |

Final Encoded String: 00000001000100100100001101010010101000100011

Bits in original string are 16*4 = 64
Bits in compressed string are 11*4 = 44
44/64 = 0.69 or a 31% compression.
Now to do the decoding remember the dictionary only starts with A,B, and C in it.

**LZW Decompression**

Decode: 0000 0001 0001 0010 0100 0011 0101 0010 1010 0010 0011
With the initial dictionary:

| | |
|---|---|
| A | 0000 |
| B | 0001 |
| C | 0010 |

Decoding
- add to the dictionary the longest suffix in the dictionary of the previous code, plus the current code's first letter

| Step | Procedure | Dictionary Action | Decoded String So Far |
|---|---|---|---|
| 1 | Decode 0000 | none | A |
| 2 | Decode 0001 | add AB = 0011 | B |
| 3 | Decode 0001 | add BB = 0100 | B |
| 4 | Decode 0010 | add BC = 0101 | C |
| 5 | Decode 0100 | add CB = 0110 | BB |
| 6 | Decode 0011 | add BBA = 0111 | AB |
| 7 | Decode 0101 | add ABB = 1000 | BC |
| 8 | Decode 0010 | add BCC = 1001 | C |
| 9 | Decode 1010 | add CC = 1010 | CC |
| 10 | Decode 0010 | add CCC = 1011 | C |
| 11 | Decode 0011 | add CA = 1100 | AB |

For the step 9, you get a dictionary miss. So you must append the last character of your last dictionary addition to the last thing you decoded and add that to the dictionary.