

CMPS340 File Processing
HW #3 Fall 2003
Ziv-Lempel-Welch Text Compression/Decompression
Sample Solutions

1. With respect to the three-character alphabet $\{\$, a, b\}$ (in which $\$$ is used strictly to signal end-of-string), indicate the sequence of nonnegative integers produced by applying the ZLW compression algorithm to the string

ababaaabbbbaa $\$$

Show how this sequence would be encoded as a bit string, according to the compression algorithm. (Each integer in the sequence is encoded using the fewest number of bits necessary for encoding the largest number that could possibly occupy its position in the sequence, which is $\lceil \lg k \rceil$ bits, where k is the number of symbols in the dictionary up to that point of execution.

Assume that, initially, the *dictionary* (one possible representation of which is a *digital search tree* (or *trie*), as demonstrated in class) specifies that $\$, a$, and b are to be encoded as 0, 1, and 2, respectively.

Show the final contents of the dictionary (either as a trie or as a table).

Solution: Expressed as a sequence of nonnegative integers, the compressed form of the string is

1 2 3 1 6 2 8 6 0

The corresponding bit string (including gaps to show the boundaries between the codewords for different symbols) is

01 10 011 001 110 010 1000 0110 0000

The final dictionary is

$\$$	0
a	1
b	2
ab	3
ba	4
aba	5
aa	6
aab	7
bb	8
bba	9
aa $\$$	10

2. Working under the scenario described in the previous problem (i.e., same alphabet, same initial dictionary), apply the ZLW decompression algorithm to the sequence

2 1 3 5 5 4 8 7 4 0

in order to recover the original string of a 's, b 's, and $\$$'s.

Solution Inserting gaps between adjacent symbols, the answer is

b a ba bab bab ab aba baba ab \$

The final dictionary is

\$	0
a	1
b	2
ba	3
ab	4
bab	5
babb	6
baba	7
aba	8
abab	9
babaa	10
ab\$	11