Data Structures and Algorithms -- M. Richards

p4 q5

MR

p11 q5

2 Outline the mechanism used in the Burrows-Wheeler block compression algorithm, illustrating your description by applying it to the string ALFALFA. [14]

Briefly discuss the advantages and disadvantages of the Burrows-Wheeler algorithm compared to other commonly used compression methods. [6]

Answer:

```
ALFALFA.   .ALFALFA
LFALFA.A   A.ALFALF
FALFA.AL   ALFA.ALF
ALFA.ALF   ALFALFA.
LFA.ALFA   FA.ALFAL
FA.ALFAL   FALFA.AL
A.ALFALF   LFA.ALFA
.ALFALFA   LFALFA.A
```

Last column: AFF.LLAAAA

Use Move-to-front buffering then Huffman

Send new Huffman table every 100k bytes or so to make it adaptive.

Decoding can be done in linear time:

```
l[i]            .  A  F  L    p[i]
               0  0  0  0
A                 1             0
F                    1          0
F                    2          1
.               1               0
L                       1       0
L                       2       1
A                 2             1
A                 3             2
               1  3  2  2

          .0 A0 A1 A2 F0 F1 L0 L1
          A0 F0 F1 .0 L0 L1 A1 A2
                .0                .
          A0                      A
             F0                   F
                   L0             L
                         A1       A
                F1                F
                      L1          L
                            A2    A
                   .0
```

Needs larger blocks (100k to 4M) to be effective.
Needs a lot of work space for sorting and decoding.
Sorting of suffixes is tricky to avoid bad cases such as a big file concatenated with itself.
But it typically compresses 20% better than LZW for larger text files.