Burrows Wheeler Transform

1. Let $BW(S) = aipooon\$mno$ a be the Burrows Wheeler transform of a string S. Write out the first column F and show how to use it to find the original string. Using this procedure, find S.

2. A given string T is 40 characters long (including the ending \$ character). Part of its FM index (the "M" array and the "occ" table, but not the BWT) is included on the next page. Recall the recursive procedure for finding a pattern $P = c\sigma$, where c is a single character and σ is a previously matched pattern:

$$\text{sp}(c\sigma) = \text{M}[c] + \text{occ}(c, \text{sp}(\sigma) - 1)$$
$$\text{ep}(c\sigma) = \text{M}[c] + \text{occ}(c, \text{ep}(\sigma)) - 1$$

   where "sp" is the start point of the pattern and "ep" is the end point of the patter (inclusive). M[c] is the index of the first c in the F column of the sorted cyclic permutations (π^{sorted}), and occ(c,i) is the number of occurrences of character c in the BWT (L column), up through index i.

   Use this recursive procedure to find the start point and end point (with respect to the F column) of the pattern $P = gca$. Then explain clearly the asymmetric -1's in the recursive formula.

3. Suppose the the cyclic permutations (π) of a string R have alphabetical order: 1, 6, 3, 4, 5, 2. (So R has length 6, including the \$ character.) Also suppose that for a pattern P, you know sp(P) = 5 and ep(P) = 6.

   (a) Use this information to recover the positions of P in the *original* string R.

   (b) Come up with a 6-character string R (using a, c, g, and t) and a 2-character pattern P that matches these properties (alphabetical order of permutations and sp/ep of P).

	a	c	g	t
M	2	15	20	28
i	occ			
1	0	0	0	1
2	0	1	0	1
3	0	1	0	2
4	1	1	0	2
5	1	1	0	3
6	2	1	0	3
7	3	1	0	3
8	3	1	0	4
9	3	1	1	4
10	4	1	1	4
11	5	1	1	4
12	5	2	1	4
13	5	3	1	4
14	5	3	1	5
15	5	3	2	5
16	5	3	2	6
17	5	3	3	6
18	5	3	4	6
19	6	3	4	6
20	6	3	4	7
21	7	3	4	7
22	8	3	4	7
23	9	3	4	7
24	9	3	4	7
25	9	3	4	8
26	9	3	4	9
27	9	3	4	10
28	10	3	4	10
29	10	4	4	10
30	10	4	4	11
31	11	4	4	11
32	11	4	5	11
33	11	4	5	12
34	11	5	5	12
35	11	5	5	13
36	11	5	6	13
37	12	5	6	13
38	12	5	7	13
39	13	5	7	13
40	13	5	8	13