CSC 226
Lab 10
Summer 2018

1.a)

	A	A	C	A	A	A	B
	0	1	2	3	4	5	6
A	1	2	2	4	5	6	2
B	0	0	0	0	0	0	7
C	0	0	3	0	0	3	3

1.b)

	A	B	A	B	A	B	A	B
	0	1	2	3	4	5	6	7
A	1	1	3	1	5	1	7	1
B	0	2	0	4	0	6	0	8

2.a) M is the length of the pattern and R is the size of the alphabet.

For the pattern “AACAAAB”

M = 7,

R = 3 (or 128, or 256, or other larger character sets) The alphabet contains only 3 characters, namely, A, B, and, C.

Explanation for “R = 3”: This will be space efficient, but the code given in Figure 1 of Lab 10 will have to be changed. We cannot use “dfa[pat.charAt(j)][.]” anymore. We will have to use “dfa[pat.charAt(j)-65][.]”. Hopefully, the reason is clear. However, if we set R=256 (The alphabet contains the Extended ASCII characters) no such modifications are required. In that case we will have 253 redundant rows.

For the pattern “ABABABAB”

M = 8

R = 2 (Same reasoning will apply)

2.b) All other values are set to 0.

3. Code is given in the file DFA_Construction.java.

4. Code is given in the file KMP_Search.java. The first pattern is found in test1.txt at index 373. And the second pattern is found in test2.txt at index 28.