

CS1501: KMP Notes

Keren Ye

2016-02-26

1. Build the DFA

1.match transitions: if in state j and next char $c == \text{pat.charAt}(j)$, goto $j + 1$

2.mismatch transitions: if in state j and next char $c != \text{pat.charAt}(j)$, then the last $j - 1$ characters of input are $\text{pat}[1..j-1]$

To compute $\text{dfa}[c][j]$: simulate $\text{pat}[1..j-1]$ on DFA and take transition c .

Running time: takes only constant time if we maintain state X

★ **Algorithm (pseudo code) - Try understand the process:**

```
for each state j
  copy dfa[][x] to dfa[][j] for mismatch case
  set dfa[pat.charAt(j)][j] to j + 1 for match case
  update x
```

Algorithm (Java code):

```
public KMP(String pat) {
  this.pat = pat;
  M = pat.length();
  dfa = new int[R][M];
  dfa[pat.charAt(0)][0] = 1;
  for (int X=0, j=1; j < M; ++j) {
    for (int c = 0; c < R; ++c)
      dfa[c][j] = dfa[X][X]; // copy mismatch cases
    dfa[pat.charAt(j)][j] = 1 + j; // set match case
    X = dfa[pat.charAt(j)][X]; // update restart state x
  }
}
```

What is interpretation of DFA after reading $\text{txt}[i]$

State = number of characters in pattern that have been matched

2. Apply string matching:

```
public int search(String txt) {
  int i, j, N = txt.length();
  for (i = 0, j = 0; i < N && j < M; ++i)
    j = dfa[txt.charAt(i)][j];
  if (j == M) return i - M;
  else return N;
}
```

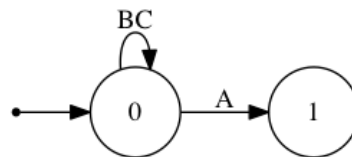
3. Example of Creating the DFA

Build DFA for pattern $ABABAC$

a. Initial DFA

x denotes the tracking state

	A	B	A	B	A	C	
	0	1	2	3	4	5	6(AC)
A	1						
B	0						
C	0						



b. Transitions from state 1

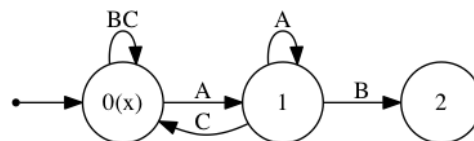
$x = 0$

copy $DFA[][0]$ to $DFA[][1]$

$DFA[B][1] = 2$

$x = DFA[B][x] = 0$

	A	B	A	B	A	C	
	0(x)	1	2	3	4	5	6(AC)
A	1	1					
B	0	2					
C	0	0					



c. Transitions from state 2

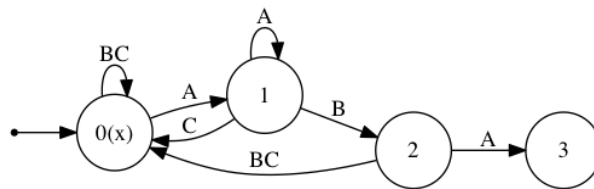
$x = 0$

copy $DFA[] [0]$ to $DFA[] [2]$

$DFA[A][2] = 3$

$x = DFA[A][x] = 1$

	A	B	A	B	A	C	
	0(x)	1	2	3	4	5	6(AC)
A	1	1	3				
B	0	2	0				
C	0	0	0				



d. Transitions from state 3

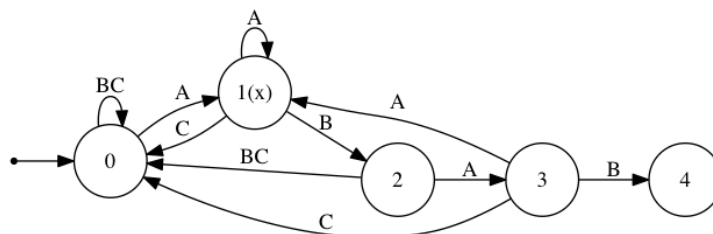
$x = 1$

copy $DFA[] [1]$ to $DFA[] [3]$

$DFA[B][3] = 4$

$x = DFA[B][x] = 2$

	A	B	A	B	A	C	
	0	1(x)	2	3	4	5	6(AC)
A	1	1	3	1			
B	0	2	0	4			
C	0	0	0	0			



e. Transitions from state 4

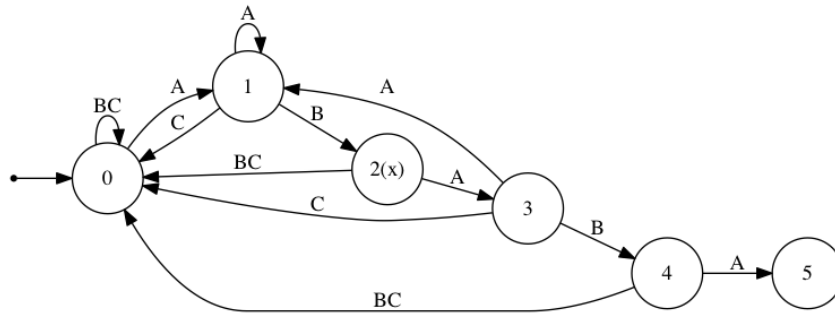
$x = 2$

copy $DFA[] [2]$ to $DFA[] [4]$

$DFA[A][4] = 5$

$x = DFA[A][x] = 3$

	A	B	A	B	A	C	
	0	1	2(x)	3	4	5	6(AC)
A	1	1	3	1	5		
B	0	2	0	4	0		
C	0	0	0	0	0		



f. Transitions from state 5

$x = 3$

copy $DFA[] [3]$ to $DFA[] [5]$

$DFA[C][5] = 6$

$x = DFA[C][x] = 0$

	A	B	A	B	A	C	
	0	1	2(x)	3	4	5	6(AC)
A	1	1	3	1	5	1	
B	0	2	0	4	0	4	
C	0	0	0	0	0	6	

