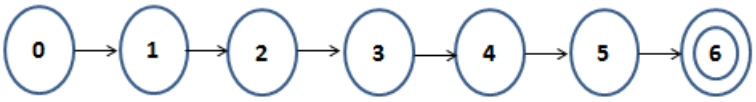
COS 226 Data Structures and Algorithms
Computer Science Department
Princeton University
Spring 2016

Week 11 handout

1. **Substring search**

(a) Construct the Knuth-Morris-Pratt DFA for the string ABAABA over the alphabet $\{A, B\}$. Complete the transition diagram and the corresponding DFA table. State 6 is the accept state.



	0	1	2	3	4	5
A						
B						

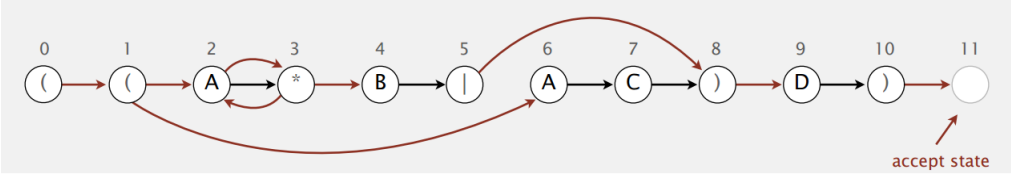
(b) Below is a partially-completed Knuth-Morris-Pratt DFA for a string s of length 8 over the alphabet $\{A, B\}$. State 8 is the accept state. Reconstruct the DFA and s in the space below.

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

2. **Regular expressions and Non-Deterministic Finite Automata (NFA)**

(a) Convert the regular expression $(a \mid (b* \mid cd)*)$ into an equivalent NFA using the algorithm described in lecture by adding black edges and ϵ transition edges to the diagram below. The final accept state is highlighted.

(b) Consider the NFA given below and determine which of the strings given in (i) and (ii) are accepted or rejected by the NFA. For a string that is accepted show how the machine transitions get to the accept state. When a string is not accepted, show that machine transitions never get to the accept state. You can take epsilon transitions and can be in multiple states before character A is scanned.



i. $AAAB$

char	states
$\emptyset$	
A	
A	
A	
B	

String Accepted by NFA?

YES NO

ii. ACD

char	states
$\emptyset$	
A	
C	
D	

String Accepted by NFA?

YES NO

3. **Boyer-Moore algorithm (Bonus Problem)**

   Suppose that you run the Boyer-Moore algorithm (using only the mismatched character heuristic) to search for the pattern

   N E E D L E

   in the text

   L O O K I N G F O R X S A D D L E A N D A N E E D L E

   What is the sequence of characters in the text that is compared with the last character in the pattern?

   Show your work below.

   L O O K I N G F O R X S A D D L E A N D A N E E D L E