CS445 02/28/2019

Test 1 topics:
1) Knuth-Morris-Pratt (Ch 32 from Cormen et al.)
2) Aho-Corasick (see paper + Exercises in Ch 3 Dragon1)
3) RegExp 2 DFA (Ch 3 Dragon1, §3.9)
(no NFA)

RE → DFA
RE → NFA (Thompson construction) → DFA (Subset construction (exponential blowup in size))

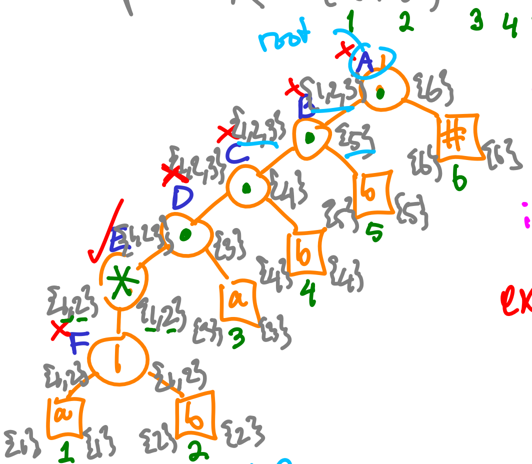
CS345 Automata Theory

4) A proof of why $L = \{ww : w \in \{0,1\}^*\}$ is not regular using Communication Complexity

RegExp to DFA: §3.9, p135

Example: $R = (a|b)^* abb\#$ → DFA?

$\quad\quad\quad\;\; 1 \;\; 2 \quad\; 3\;4\;5\;6$



i) firstpos(n) = set of positions that can match the 1st symbol of a string generated by the subexpression rooted at n (node)

ii) lastpos(n) = ditto but for last symbol

ex: firstpos(F) = {1,2}
firstpos(D) = {1,2,3}
firstpos(A) = {1,2,3}

ex: lastpos(F) = {1,2}
lastpos(D) = {3}
lastpos(A) = {6}

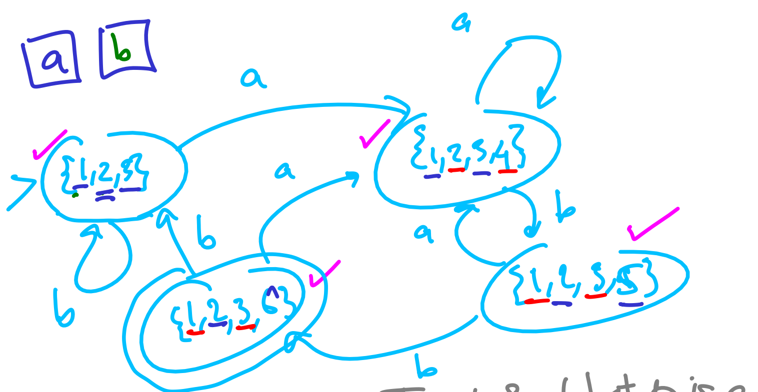
iii) nullable(n) = $\begin{cases} \text{True} & \text{if } \varepsilon \in L(RegExp_n) \\ \text{False} & \text{else} \end{cases}$

iv) followpos(i) = set of positions that can follow i in syntax tree

ex:

i	followpos
1	1,2,3
2	1,2,3
3	4
4	5
5	6

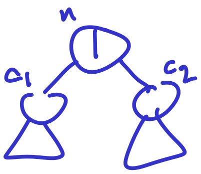
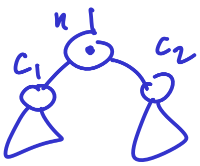
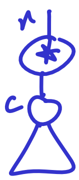
$(a|b)^* abb$



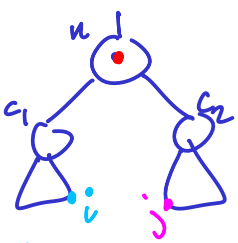
DFA construction: p141 Algorithm 3.5

DStates ← {firstpos(root)}
while ∃ unmarked state T in DStates:
  mark T
  for each $a \in \Sigma$:
    U is the set of positions in followpos(p) for some $p \in T$ s.t. symbol at p is a
    if $U \neq \phi$ and $U \notin$ DStates: add U to DStates
    DTrans[T, a] = U

P138

Node n	nullable(n)	firstpos(n)	lastpos(n)
n: $\boxed{\varepsilon}$	true	ϕ	ϕ
$a \in \Sigma$ — n: $\boxed{a}$, i	false	$\{i\}$	$\{i\}$
	nullable(c_1) or nullable(c_2)	firstpos(c_1) $\cup$ firstpos(c_2)	lastpos(c_1) $\cup$ lastpos(c_2)
	nullable(c_1) and nullable(c_2)	$\begin{cases} \text{firstpos}(c_1) & \text{if } \overline{\text{nullable}(c_1)} \\ \text{firstpos}(c_1) \cup \text{firstpos}(c_2) & \text{if nullable}(c_1) \end{cases}$	$\begin{cases} \text{lastpos}(c_2) & \text{if } \overline{\text{nullable}(c_2)} \\ \text{lastpos}(c_1) \cup \text{lastpos}(c_2) & \text{if nullable}(c_2) \end{cases}$
	true	firstpos(c)	lastpos(c)

followpos(i) :

1) 

$i \in \text{lastpos}(c_1) \Rightarrow$

$\text{firstpos}(c_2) \subseteq \text{followpos}(i)$

2) 

$i \in \text{lastpos}(n) \Rightarrow$

$\text{firstpos}(n) \subseteq \text{followpos}(i)$