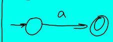
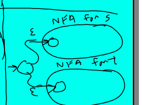
- Regex → NFA (finish)
- DFA minimization

Recall: Σ alphabet.

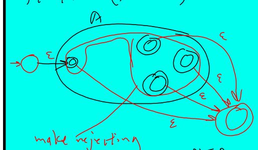
r regex over Σ	equiv NFA
$\varnothing$	
$a \in \Sigma$	
$s \cup t$ (s, t regexes over Σ)	

Def: An NFA is clean if:

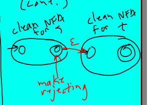
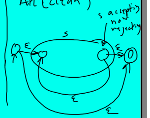
- it has a unique accepting state, and this state is not the start state.
- no transitions into the start state.
- no transitions out of the accepting state.

Easy Lemma: For every NFA there is an equivalent clean NFA.

Picture proof: Given NFA A, we modify it thus (in red):



rules for regex → NFA conversion (cont.)

st	
s^*	 Alt (clean) 

DFA minimization:

Given a DFA $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ want to find an equivalent DFA with the fewest possible states. (Happens to be unique.)

Def: A DFA is sane if every state is reachable from the start state (via some input string).

Observe: Removing unreachable states (from start state) results in an equivalent DFA.

Def: Let A be as above. For every $q \in Q$, we define

$$A_q := \langle Q, \Sigma, \delta, q, F \rangle$$

Def: Let $q, r \in Q$. Say that q & r are distinguishable if

$$L(A_q) \neq L(A_r)$$

(q, r are indistinguishable otherwise).

Minimize A:
1. Remove unreachable states
2. Merge indist. states into the same state.

1. is easy (BFS from start state)

2. Find all pairs of dist states. Pairs left over are indist and can be merged.

Rules for distinguishability

Let $q, r \in Q$.

q and r are dist. iff

1. one of q & r is accepting and the other is rejecting.

   Then $\varepsilon \in L(A_q) \triangle L(A_r)$

   (q & r are distinguished by ε)

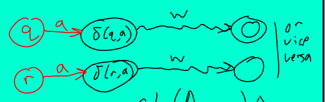
2. There exists an $a \in \Sigma$ such that

   $\delta(q,a)$ and $\delta(r,a)$

   are distinguishable.

[If string w distinguishes $\delta(q,a)$ from $\delta(r,a)$

then, say,



$w \in L(A_{\delta(q,a)}) \triangle L(A_{\delta(r,a)})$

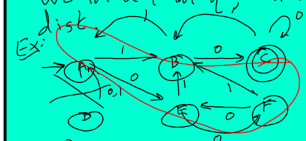
$\Rightarrow$ aw distinguishes q from r

i.e., $aw \in L(A_q) \triangle L(A_r)$

Nothing else — applying these rules a finite number of times yields all dist. pairs.

Initialize a table $T[q,r]$ with all blank. Put "X" into an entry $T[q,r]$ when we notice that q, r are dist.

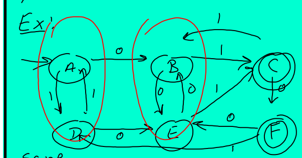
Ex:



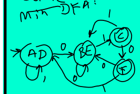
Min DFA:

	A	B	C	D	E
B	X				
C	X	X			
D	X	X	X		
E		X	X	X	
F		X	X	X	

Ex:



same Min DFA:



	A	B	C	D	E
B	X				
C	X	X			
D		X	X		
E	X		X	X	
F	X	X	X	X	X