DFA minimization

Regexes

DFA min:

Def: Let A be a DFA. A is sane if every state of A is reachable from its start state, i.e., for every state q of A, there exists a string w (over input alphabet of A) such that $A(w)=q$.
"q is reachable"

Def: DFA A is minimal if no DFA with fewer states is equivalent to A.

Obvious: Every minimal DFA is sane.

DFA minimization:
1) Remove unreachable states making it sane [BFS]
2) Compute an indistinguishability relation on the state set (equivalence relation on (pairs of states)
3) Merge equivalent states into a single state.

Def: Let $A := \langle Q, \Sigma, \delta, q_0, F\rangle$ be a DFA. Two states $q, r \in Q$ are distinguishable if there exists a string $w \in \Sigma^*$ such that reading w from state q leads to acceptance/rejection behavior opposite that of reading the same string starting from r.

Def: A as above and $q \in Q$. Define the DFA

$$A_q := \langle Q, \Sigma, \delta, \underbrace{q}_{\text{only difference}}, F\rangle$$

Then q and r are distinguishable iff $\exists w \in \Sigma^*$, one of

$A_q(w)$ and $A_r(w)$

is accepting & the other rejecting.
Equivalently, q & r are distinguishable iff $L(A_q) \neq L(A_r)$.

Def: q & r are indistinguishable if they are not distinguishable, i.e., $L(A_q) = L(A_r)$. {thus an equiv. relation.

Step 2: compute this equiv. relation
Idea: Find all pairs of distinguishable states then pairs left over are indist.

Prop: $A := \langle Q, \Sigma, \delta, q_0, F\rangle$ a DFA. Let $q, r \in Q$.
1) If one of q & r is accepting and the other rejecting, then q & r are distinguishable (let $w := \varepsilon$).
2) If there exists $a \in \Sigma$ such that $\delta(q,a)$ and $\delta(r,a)$ are distinguishable, then q and r are distinguishable.
3) Every pair of distinguishable state is obtained by first applying (1) above, then (2) zero or more times.

[Proof: by induction on the length of the shortest distinguishing string]

Ex:

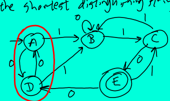
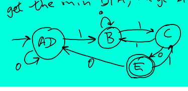


Table of distinguishabilities
Put "X" in an entry corresp to a known dist. pair of states.

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

(A,D) is the only indistinguishable pair of states.

To get the min DFA, merge A & D:



<u>Detour</u>: regexes.

Σ alphabet,

$L, M \subseteq \Sigma^*$ define

$$LM := \{xy : x \in L \ \& \ y \in M\}$$

$$L^* := \{\varepsilon\} \cup L \cup LL \cup LLL \cup \cdots \cup L^n \cup \cdots$$

$$= \{x_1 \cdots x_n : n \geq 0 \ \& \ x_1, \ldots, x_n \in L\}$$

$a \in \Sigma$: shorthand:

- $a^* := \{a\}^* = \{\varepsilon, a, aa, \ldots, a^n, \ldots\}$

$a \cup b$ $\quad (b \in \Sigma)$

$:= \{a\} \cup \{b\} = \{a, b\}$

$aL = \{a\}L$,

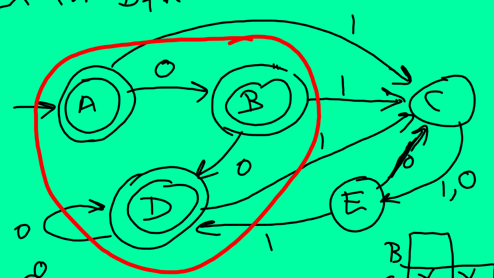
$a \cup L = \{a\} \cup L$, etc.

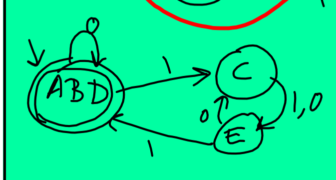
Expressions of this form denote languages and are called <u>regular expressions</u> (<u>regexes</u>).

---

end of detour.

Ex for DFA min





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