Suppose that L, M are languages.

Union: $L \cup M = \{w | w \in L \text{ or } w \in M\}$
Concatenation: $LM = \{xy | x \in L \text{ and } y \in M\}$
Star (or closure): $L^* = \{w | w = x_1 x_2 \ldots x_n \text{ and } x_i \in L\}$

Regular Expressions

Basis Case: $a \in \Sigma, \varepsilon, \emptyset$

Induction Step: If E, F are regular expressions, the so are $E + F, EF, (E)^*, (E)$.

What are $L(a), L(\varepsilon), L(\emptyset), L(E + F), L(EF), L(E^*)$?

Ex. 3.1.3(a) Give a reg exp for the set of strings of 0s and 1s not containing 101 as a substring:

$$(\varepsilon + 0)(1^* + 00^*0)^*(\varepsilon + 0)$$

Theorem: A language is regular iff it is given by some regular expression.

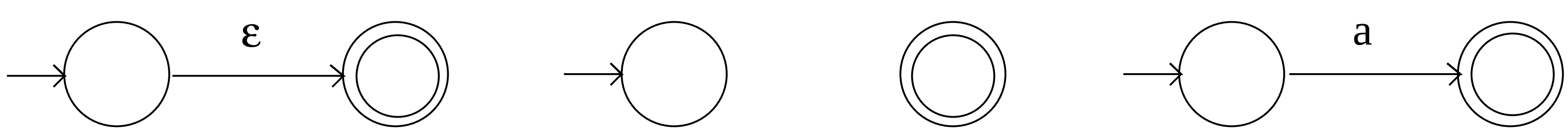
Proof: reg exp $\Longrightarrow$ ε-NFA & DFA $\Longrightarrow$ reg exp

[$\Longrightarrow$]

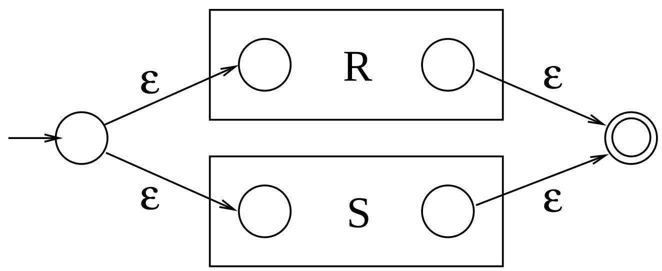
Use structural induction to convert R to an ε-NFA with 3 properties:

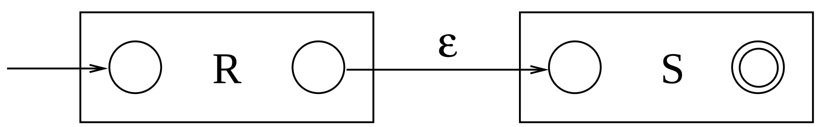
1. Exactly one accepting state
2. No arrow into the initial state
3. No arrow out of the accepting state

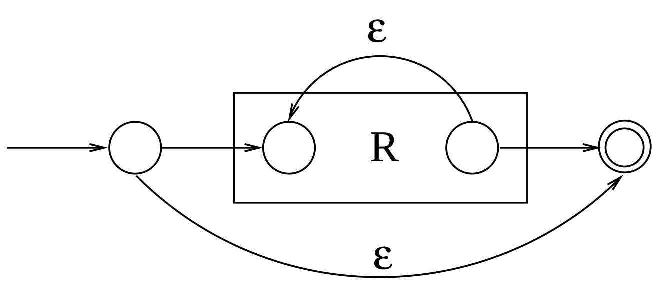
Basis Case: $\varepsilon, \emptyset, a \in \Sigma$



Induction Step: $R + S, RS, R^*, (R)$







[$\Longleftarrow$] Convert DFA to reg exp.

Method 1

Suppose A has n states. $R_{ij}^{(k)}$ denotes the reg exp whose language is the set of strings w such that:

> w takes A from state i to state j with all intermediate states $\leq k$

What is R such that $L(R) = L(A)$?

$R = R_{1j_1}^{(n)} + R_{1j_2}^{(n)} + \cdots + R_{1j_k}^{(n)}$ where $F = \{j_1, j_2, \ldots, j_k\}$

Build $R_{ij}^{(k)}$ by induction on k.

Basis Case: $k = 0$, $R_{ij}^{(0)} = x + a_1 + a_2 + \cdots + a_k$ where $i \xrightarrow{a_l} j$ and $x = \emptyset$ if $i \neq j$ and $x = \varepsilon$ if $i = j$

Induction Step: $k > 0$

$$R_{ij}^{(k)} = \underbrace{R_{ij}^{(k-1)}}_{\text{path does not visit } k} + \underbrace{R_{ik}^{(k-1)} \left(R_{kk}^{(k-1)}\right)^* R_{kj}^{(k-1)}}_{\text{visits } k \text{ at least once}}$$

Method 2

DFA $\Longrightarrow$ Gε-NFA $\Longrightarrow$ Reg Exp

Generalized ε-NFA:

$$\delta : (Q - \{q_{\text{accept}}\}) \times (Q - \{q_{\text{start}}\}) \longrightarrow \mathcal{R}$$

where the start and accept states are unique.

G accepts $w = w_1 w_2 \dots w_n$, $\underline{w_i \in \Sigma^*}$, if there exists a sequence of states

$$q_0 = q_{\text{start}}, q_1, \dots, q_n = q_{\text{accept}}$$

such that for all i, $w_i \in L(R_i)$ where $R_i = \delta(q_{i-1}, q_i)$.

When translating from DFA to Gε-NFA, if there is no arrow $i \longrightarrow j$, we label it with $\emptyset$.

For each i, we label the self-loop with ε.

Eliminate states from G until left with just $q_{\text{start}} \xrightarrow{R} q_{\text{accept}}$: