Regular expressions over $\Sigma$

- $\emptyset$ is an r.e.
- $\varepsilon$ is an r.e.
- $a$ is an r.e. for each $a \in \Sigma$

If $R_1$ and $R_2$ are r.e., then so are:

- $(R_1 \cup R_2)$
- $(R_1 \cdot R_2)$
- $(R_1^*)$

The language denoted by $R$, $L(R)$ is:

- $L(\emptyset) = \emptyset$
- $L(\varepsilon) = \{\varepsilon\}$
- $L(R_1 \cup R_2) = L(R_1) \cup L(R_2)$
\[ \mathcal{L}(\phi^*) = \mathcal{L}(\phi)^* \]

Short hands
\[ \Sigma = \{a, b, c\} \]
\[ \mathcal{L}(((a \cup b) \cup c)) = \Sigma \]
\[ ((\Sigma^* \cup \Sigma).a) \]
\[ (((a \cup b) \cup c)^* \cup \Sigma).a) \]
precedence & associativity
\[ (a \cup b \cup c)* \]
\[ a \cup b \cup c^* \]
"words ending with ".TXT" +

*$ .TXT$

(au bu ... uz) * (au ... uz vua ... p)*

2 * (2ud) *

(ΣΣ) *

0 * 10 *

(ΣuΣ)(ΣuΣ)

Σ =

00 * 0 * (10*10+)*

00 * (0 * (0*10*)*) *

(0 * 10*1) * 0 *

(d * d+ u d+ . d*) (ΣuΣ Σ u +u -) d+"
Theorem:
A regular expression $R \exists$ an NFA $M_R$ s.t. $L(R) = L(CMR)$

Proof:
By induction on $k$, the # of $\cup, \cdot, \star$ operators in $R$

Base cases ($k = 0$):
Then $R$ is "$\emptyset$", "$\varepsilon$", or "$a" for $a \in \Sigma$
Explicitly give simple NFA's recognizing $\emptyset$, $\{\varepsilon\}$, and $\{a\}$ for each $a \in \Sigma$
(details omitted)

Induction step ($R$ has $k > 0$ operators)
I.H.: assume that for all regular expressions $R'$ with $\leq k$ operators,
$\exists$ NFA $M_{R'}$ recognizing $L(CR')$

$R$ has $k > 0$ operators. So
$R = (R_1 \cup R_2)$ or $(R_1 \cdot R_2)$ or $(R_1)^*$
where $R_1, R_2$ if any) have $\leq k - 1$
operators. By I.H., $\exists M_{R_1}, (CMR_1)$ s.t.
$L(R_1) = L(CMR_1)$, $i = 1, 2$. Modify/join
it/them as in previous proofs of closure
under $\cup, \cdot, \star$ to get $M_R$ s.t. $L(R) = L(M_R)$.
Regular language

Every **can be described by a regular expression.

GNFA

Note: No loss in assuming no edges into $q_0$ / out of $F$ / only one $q_f \epsilon F$
GNFA

\[ G = (Q, \Sigma, S, q_0, \beta) \]
\[ \Sigma, Q, \beta, q_0, q_\epsilon \in \mathbb{Q} \text{ as usual} \]
\[ S: (Q - \{|q_\epsilon\}|) \times (Q - \{|q_\epsilon\}|) \rightarrow R_\Sigma \]

**Defn**
- \( G \) can be in state \( \beta \in Q \) after reading
  - \( x \in \Sigma^* \) if \( \exists k \geq 0, \exists r_0, r_1, \ldots, r_k \in \mathbb{Q} \)
  - \( \exists x_1, \ldots, x_k \in \Sigma^* \)
  - Such that
    (i) \( x = x_1 \cdot x_2 \cdot \ldots \cdot x_k \)
    (ii) \( r_0 = q_0 \)
    (iii) \( r_k = q_\beta \)
    (iv) \( \forall i \leq k, x_i \in L(\delta(r_{i-1}, r_i)) \)
- \( L(G) = \{ x \mid G \text{ can be in state } \beta \ldots \} \)

**Note:** \( \delta \) syntax a little different;
maps state pair to label (reg. expr.)
rather than state x symbol = new state.