Previously on RLFA...
REGULAR EXPRESSIONS \( r \) (e.g. \((a|b)^*aaa(a|b)^*\))

FINITE AUTOMATA \( M \) (e.g. 

\[
\begin{array}{cccc}
q_0 & \rightarrow q_1 & \rightarrow q_2 & \rightarrow q_3 \\
\downarrow & & \downarrow & \downarrow \\
b & b & b & b \\
\end{array}
\]

REGULAR LANGUAGE = set of strings of the form \( L(M) \) (all strings accepted by \( M \)) for some f.a.
REGULAR EXPRESSIONS  \( r \)  

FINITE AUTOMATA  \( M \)  

KLEENE: 
(a) For all \( r \), can construct \( M \) with \( L(M) = L(r) \) 
(b) For all \( M \), can construct \( r \) with \( L(r) = L(M) \)
Typical application: **lexical analysis at start of compilation**

- PL definition specifies legal tokens (keywords, identifiers, etc.) using a reg. exp.
- Lexical analyzer splits a character stream into a stream of tokens by constructing a F.A. from the reg. exp.

""
Examples of non-regular languages

- The set of strings over \( \{ (, ), a, b, \ldots, z \} \) in which the parentheses ‘(’ and ‘)’ occur well-nested.

- The set of strings over \( \{ a, b, \ldots, z \} \) which are palindromes, i.e. which read the same backwards as forwards.

- \( \{ a^n b^n \mid n \geq 0 \} \)
The Pumping Lemma

For every regular language $L$, there is a number $\ell \geq 1$ satisfying the pumping lemma property:

all $w \in L$ with $\text{length}(w) \geq \ell$ can be expressed as a concatenation of three strings, $w = u_1vvu_2$, where $u_1$, $v$ and $u_2$ satisfy:

- $\text{length}(v) \geq 1$
  (i.e. $v \neq \varepsilon$)

- $\text{length}(u_1v) \leq \ell$

- for all $n \geq 0$, $u_1v^nu_2 \in L$
  (i.e. $u_1u_2 \in L$, $u_1vu_2 \in L$ [but we knew that anyway], $u_1vvu_2 \in L$, $u_1vvvu_2 \in L$, etc).
Suppose \( L \) is \( L(M) \) for a DFA \( M \).

If \( n \geq \ell = \) number of states of \( M \), then in

\[
\begin{align*}
\sigma_M &= q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots \xrightarrow{a_\ell} q_\ell \cdots \xrightarrow{a_n} q_n \in \text{Accept}_M \\
\text{\( \ell+1 \) states}
\end{align*}
\]

\( q_0, \ldots, q_\ell \) can’t all be distinct states. So \( q_i = q_j \) for some \( 0 \leq i < j \leq \ell \). So the above transition sequence looks like

\[
\sigma_M = q_0 \xrightarrow{u_1}^{*} q_i = q_j \xrightarrow{v}^{*} q_n \in \text{Accept}_M
\]

where

\[
u_1 \overset{\text{def}}{=} a_1 \ldots a_i \quad v \overset{\text{def}}{=} a_{i+1} \ldots a_j \quad u_2 \overset{\text{def}}{=} a_{j+1} \ldots a_n.
\]
How to use the Pumping Lemma to prove that a language $L$ is not regular

For each $\ell \geq 1$, find some $w \in L$ of length $\geq \ell$ so that

$(†)$

no matter how $w$ is split into three, $w = u_1vu_2$,

with $\text{length}(u_1v) \leq \ell$ and $\text{length}(v) \geq 1$,

there is some $n \geq 0$ for which $u_1v^n u_2$ is not in $L$. 
Examples

(i) \( L_1 \overset{\text{def}}{=} \{ a^n b^n | n \geq 0 \} \) is not regular.

[For each \( \ell \geq 1 \), \( a^\ell b^\ell \in L_1 \) is of length \( \geq \ell \) and has property (†) on Slide 31.]

(ii) \( L_2 \overset{\text{def}}{=} \{ w \in \{a, b\}^* | w \text{ a palindrome} \} \) is not regular.

[For each \( \ell \geq 1 \), \( a^\ell ba^\ell \in L_1 \) is of length \( \geq \ell \) and has property (†).]

(iii) \( L_3 \overset{\text{def}}{=} \{ a^p | p \text{ prime} \} \) is not regular.

[For each \( \ell \geq 1 \), we can find a prime \( p \) with \( p > 2\ell \) and then \( a^p \in L_3 \) has length \( \geq \ell \) and has property (†).]
Example of a non-regular language that satisfies the ‘pumping lemma property’

\[ L \overset{\text{def}}{=} \{c^m a^n b^n \mid m \geq 1 \text{ and } n \geq 0\} \]
\[ \cup \]
\[ \{a^m b^n \mid m, n \geq 0\} \]

satisfies the pumping lemma property on Slide 29 with \( \ell = 1 \).

[For any \( w \in L \) of length \( \geq 1 \), can take \( u_1 = \varepsilon, v = \) first letter of \( w \), \( u_2 = \) rest of \( w \).]

But \( L \) is not regular.

Assume \( L = L(M) \) for some DFA \( M \) and obtain a contradiction...
Given $M$, define a new NFA $N$ as follows:

$\text{States}_N = \text{States}_M$

$\text{Accept}_N = \text{Accept}_M$

$N$ starts here

$N$ has only the $a$ & $b$ transitions of $M$
Given $M$, define a new NFA $N$ as follows:

$\text{States}_N = \text{States}_M$

$\text{Accept}_N = \text{Accept}_M$

So $L(N) = \{a^n b^n \mid n \geq 0\}$, contradiction to Pumping Lemma.