Representing Regular Languages: The Story So Far

- NFAs
- DFAs
- Regular Expressions
Components of a State Machine?
From NFA to DFA: the Subset Construction
The Subset Construction Formalized

✓ Input: An NFA \((\Sigma, Q, \rightarrow, q_0, F)\)

✓ Output: A DFA \((\Sigma, \mathcal{P}(Q), \rightarrow', q_0', F')\)

✓ Definition?
Regular Expressions

✓ What are the “primitive” regular expressions?

✓ What other regular expression forms have you seen?
Mu'ammar Qaddafi
Mo'ammar Gadhafi
Muammar Kaddafi
Muammar Qadhafi
Moammar El Kadhafi
Muammar Gadafi
Mu'ammar al-Qadafi
Moamer El Kazzafi
Moamar al-Gaddafi
Mu'ammar Al Qathafi
Muammar Al Qathafi
Mo'ammar el-Gadhafi
Moamar El Kadhafi
Muammar al-Qadhafi
Mu'ammar al-Qadhafi
Mo'ammar Muhammad Abu Minyar al-Qadhafi

from the RX Library
Exercise

1. Give a regular expression for C identifiers:
   - Can contain letters, digits, and underscores
   - Must begin with a letter or underscore.

   ✓ Main
   ✓ __Z6rotateiii

2. Give a regular expression for Ada identifiers:
   - Can contain letters, digits, and underscores (_)
   - Must begin with a letter
   - Have no consecutive underscores or an underscore at the end

   ✓ woohoo32
   ✓ Last_Nonzero_Row
Exercise

✓ 6
✓ 9
✓ 2u
✓ 2L
✓ 2UL
✓ 2lu
✓ 2uu
✓ 02
✓ 002
✓ 09
✓ 02ul
✓ 0Xff90
✓ 0x0
✓ 0x0F
✓ 0x0F1U
✓ 0x0x3
✓ 0
✓ 0lu
✓ 0x
From Automata to Regexps

- Two approaches
  - Solving equations
  - Generalized NFAs
Let \( L_q \) be the set of strings are accepted when starting from state \( q \).

What is \( L_{q_0}, L_{q_1}, L_{q_2}, \ldots \)?

How is \( L_{q_1} \) related to \( L_{q_2} \)?
Example

\[ L_A = \varepsilon L_B \mid b L_D \]

\[ L_B = \]

\[ L_C = \]

\[ L_D = \]
The equation
\[ L = AL \mid B \]
with \( A \) and \( B \) being languages and \( L \) an unknown
has the solution
\[ L = A^* B \]
This is the smallest solution.
If \( \varepsilon \notin A \), this solution is unique.
Otherwise, \( A^*C \) is a solution for any \( B \subseteq C \).
Generalized NFAs

Just like an NFA, but edges have regular expressions rather than single symbols.

Since regular expressions can be turned into NFAs, we aren’t adding any extra power.
Removing States

✓ Pick a state \( q \) to remove.
✓ For every incoming/outgoing pair
  
  \[
  \begin{array}{c}
  U \\
  q \\
  W \\
  V
  \end{array}
  \]

✓ replace by the direct edge

\[
\begin{array}{c}
S \\
UV^*W
\end{array}
\]
Representing Regular Languages: Conclusion

- NFAs
- DFAs
- Regular Expressions
Closure Properties

✓ A family of languages is a set of languages.
  ✓ The family of all finite languages
  ✓ The family of all languages
  ✓ The family of all regular languages

✓ A family $F$ is closed under an operation if applying the operation to languages in $F$ always produces a result in $F$. 
Example: Finite Languages

Are the finite languages closed under:

- Union? (A ∪ B)
- Intersection? (A ∩ B)
- Concatenation (AB)
- Star? (A*)
- Complement? (A^c)
Closure Properties for Regular Languages

- Regular languages are closed under
  - Concatenation
  - Union
  - Star
  - Complement
  - Intersection

- Proofs: Consider the corresponding automata...
Complement?
Complement
Intersection
Intersection