Theory of Computation, CSCI 438 spring 2022
Regular languages describe regular expressions, pg. 69-76, Jan. 31

Theorem: A language is regular if and only if some regular expression describes it. Proved "if" direction (⇐) last time. This time prove "only if" direction (⇒). Given a regular language prove that it is described by a regular expression.

To convert an NFA to a regular expression:

1. Construct a GNFA (Generalized NFA)
    a. Transitions are regular expressions and strings can be read rather than a single alphabet symbol.
    b. Create a new start state which has a transition to every other state (and none going into it). ε-transition from this new state into the start that of the original NFA.
    c. Create a new final state which has a transition from every other state (and none going out of it). ε-transition from each accept state in the original NFA (these will no longer be accept states) into this newly created accept state.
    d. For every internal state there is a transition to every other state.

    (Note that the graph is very cluttered if all of the Φ-transitions are drawn, that is the transitions where there is no direct path. Thus, know they are there, but don't draw them.)

2. Repeat until no more internal states, i.e. there is only a start and final state:
    - Select an internal state (not the start or final state) to remove from the GNFA.
    - Remove that state but update each transition in the remaining GNFA so that nothing was lost with the removal. That is, look at each pair of nodes q_i, q_j where q_i is any state except the final state and q_j is any state except the start state. It helps to list these out. Notice that you need to consider $q_i=q_j$ as well as the others, that is consider states going to themselves. Determine the new transition that is needed between q_i and q_j. Given that R_1 goes from q_i to q_{rip}, R_2 goes from q_{rip} to q_{rip}, R_3 goes from q_{rip} to q_j, and R_4 goes from $q_i=q_j$ (see Figure 1.63, page 72) , the new transition is $R_1R_2^*R_3 \cup R_4$.

      Note: Only consider direct paths q_i to q_{rip} to q_j, that is, paths with only two arcs. There are frequently longer loops that pass from q_i to q_{rip} and several transitions later return to q_j. These will automatically be dealt with when dealing with the two-step paths.