Reading: Sipser, §1.3 and “The Diagonalization Method,” pages 174–178 (from just before Definition 4.12 until just before Corollary 4.18).
Converting Finite Automata to Regular Expressions

**Theorem**: For every regular language \( L \), there is a regular expression \( R \) such that \( L(R) = L \).

**Proof**: Define **generalized NFAs** (GNFAs) (of interest only for this proof)

- Transitions labelled by regular expressions (rather than symbols).
- One start state \( q_{\text{start}} \) and only one accept state \( q_{\text{accept}} \).
- Exactly one transition from \( q_i \) to \( q_j \) for every two states \( q_i \neq q_{\text{accept}} \) and \( q_j \neq q_{\text{start}} \) (including self-loops).
**NFAs to GNFAs**

**Lemma:** For every NFA $N$, there is an equivalent GNFA $G$.

- Add new start state, new accept state. Transitions?

- If multiple transitions between two states, combine. How?

- If no transition between two states, add one. With what label?
GNFAs to REs

**Lemma:** For every GNFA $G$, there is an equivalent RE $R$.

- By induction on the number of states $k$ of $G$.
- **Base case:** $k = 2$. Set $R$ to be the label of the transition from $q_{\text{start}}$ to $q_{\text{accept}}$.
- **Inductive Hypothesis:** Suppose every GNFA $G$ of $k$ or fewer states has an equivalent RE (where $k \geq 2$).
- **Induction Step:** Given a $(k + 1)$-state GNFA $G$, we will construct an equivalent $k$-state GNFA $G'$.

**Rip:** Remove a state $q_r$ (other than $q_{\text{start}}, q_{\text{accept}}$).

**Repair:** Augment labels on all transitions $q_i \rightarrow q_j$ to also include strings that could have followed the transitions $q_i \rightarrow q_r \rightarrow q_j$. 
Ripping and repairing GNFA: details

Given a \((k + 1)\)-state GNFA \(G\), we construct an equivalent \(k\)-state GNFA \(G'\) as follows:

**Rip:** Remove a state \(q_r\) (other than \(q_{\text{start}}, q_{\text{accept}}\)).

**Repair:** For every two states \(q_i \notin \{q_{\text{accept}}, q_r\}, q_j \notin \{q_{\text{start}}, q_r\}\), let \(R_{i,r}, R_{r,r}, R_{r,j}\) be REs on transitions \(q_i \rightarrow q_j, q_i \rightarrow q_r, q_r \rightarrow q_r\) and \(q_r \rightarrow q_j\) in \(G\), respectively.

In \(G'\), put RE \(R_{ij} \cup R_{i,r} R_{r,r} R_{r,j}\) on transition \(q_i \rightarrow q_j\).

Argue that \(L(G') = L(G)\), which generated by a regular expression by IH.

Note that this proof is also constructive.
Example conversion of an NFA to a RE
Example conversion of an NFA to a RE (cont.)
Examples of Regular Languages

• \( \{ w \in \{a, b\}^* : |w| \text{ even & every 3rd symbol is an } a \} \)

• \( \{ w \in \{a, b\}^* : \text{There are not 7 } a\text{’s or 7 } b\text{’s in a row} \} \)

• \( \{ w \in \{a, b\}^* : w \text{ has both an even number of } a\text{’s and an even number of } b\text{’s} \} \)

• Are there non-regular languages???
Goal: Existence of Non-Regular Languages

Intuition:

• Every regular language can be described by a finite string (namely a regular expression).

• To specify an arbitrary language requires an infinite amount of information.

  For example, an infinite sequence of bits would suffice: $\Sigma^*$ has a lexicographic ordering, and the $i$’th bit of an infinite sequence specifying a language would say whether or not the $i$’th string is in the language.

$\Rightarrow$ Some language must not be regular.

How to formalize?
Cardinality

A set $S$ is

- **finite** if there is a bijection $\{1, \ldots, n\} \leftrightarrow S$ for some $n \geq 0$

  In that case, we say $|S| = n$

  ($|S|$ is the **size** or **cardinality** of $S$)

- **infinite** if it is not finite

  So $\mathbb{N} = \{0,1,2,\ldots\}$ is infinite

$\Rightarrow$ What about $\{\mathbb{N}\}$?
Countability

A set \( S \) is

- countably infinite if there is a bijection \( f : \mathbb{N} \leftrightarrow S \)

  This means that \( S \) can be “enumerated,” i.e. listed as \( \{s_0, s_1, s_2, \ldots \} \) where \( s_i = f(i) \) for \( i = 0, 1, 2, 3, \ldots \)

So \( \mathbb{N} \) itself is countably infinite

So is \( \mathbb{Z} \) (integers) since \( \mathbb{Z} = \{0, -1, 1, -2, 2, \ldots \} \)

**Q:** What is \( f \)?

- countable if \( S \) is finite or countably infinite

- uncountable if it is not countable
Facts about Infinite Sets

• **Proposition:** The union of 2 countably infinite sets is countably infinite.

If $A = \{a_0, a_1, \ldots\}$, $B = \{b_0, b_1, \ldots\}$

Then $A \cup B = C = \{c_0, c_1, \ldots\}$

where $c_i = \begin{cases} a_{i/2} & \text{if } i \text{ is even} \\ b_{(i-1)/2} & \text{if } i \text{ is odd} \end{cases}$

“Hilbert’s Grand Hotel Paradox”

Q: If we are being fussy, there is a small problem with this argument. What is it?

• **Proposition:** If there is an **onto** function $f : \mathcal{N} \rightarrow S$, then $S$ is countable.