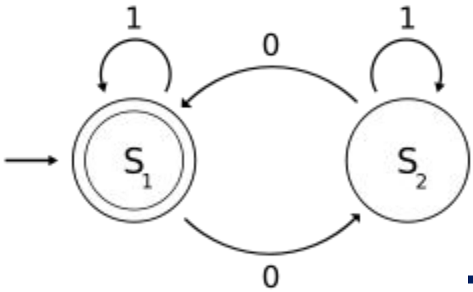
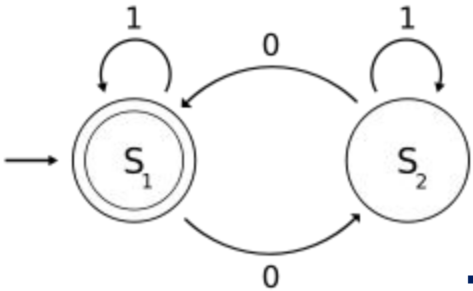


Nonregular Languages

Sipser: Section 1.4 pages 77 - 82



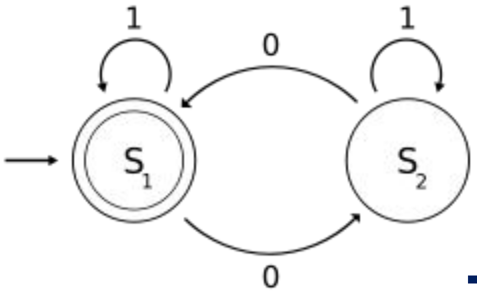
Today's adventures

- How do we show a language is not regular
  - Pigeonhole principle

- Pumping lemma, a property of regular languages

- Our first non-regular languages

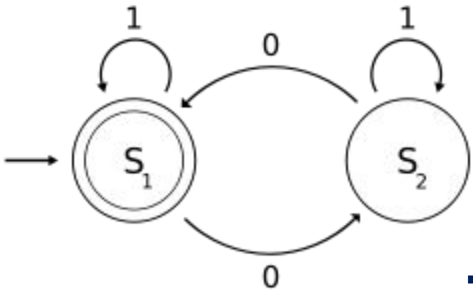
Weekly tip: Being wrong is essential for progress



How Do We Know What We Are Missing?

Regular languages may be specified either by regular expressions or by deterministic or nondeterministic finite automata.

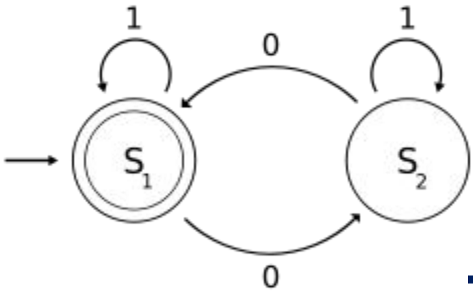
How do we show a language is not regular?



Bounded Memory

Since finite automata are not allowed to back up, the amount of memory required to determine whether or not a string is in the language must be bounded.

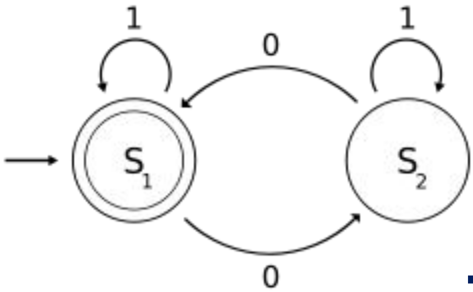
For example, consider $L = \{0^n 1^n : n \geq 0\}$.



That is a Thought, Not a Proof!

$C = \{w \mid w$ has an equal number of 0s and 1s$\}$.

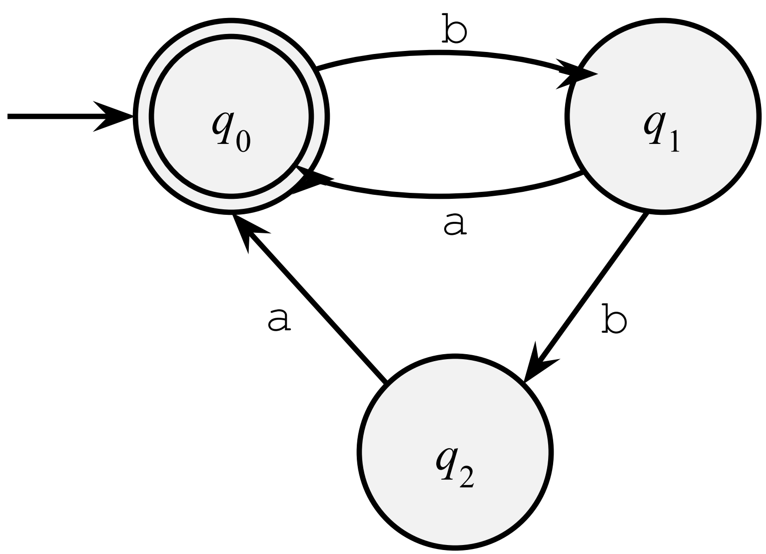
$D = \{w \mid w$ has an equal number of occurrences of 01 and 10 as substrings$\}$.



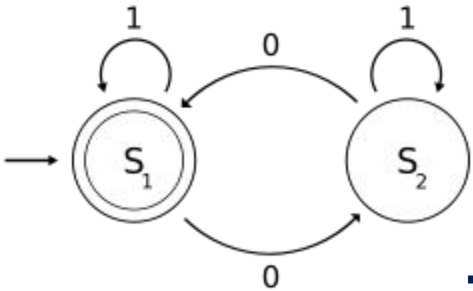
Look at it Another Way

From the specification point of view, a regular language is infinite if and only if its corresponding regular expression contains a Kleene star.

Kleene stars correspond to loops in finite automata.



Both Kleene stars and loops give rise to simple repetitive patterns in the language.

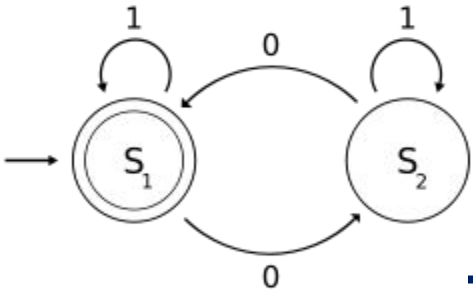


The Pigeonhole Principle

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting an infinite language L. Suppose $|Q| = p$.

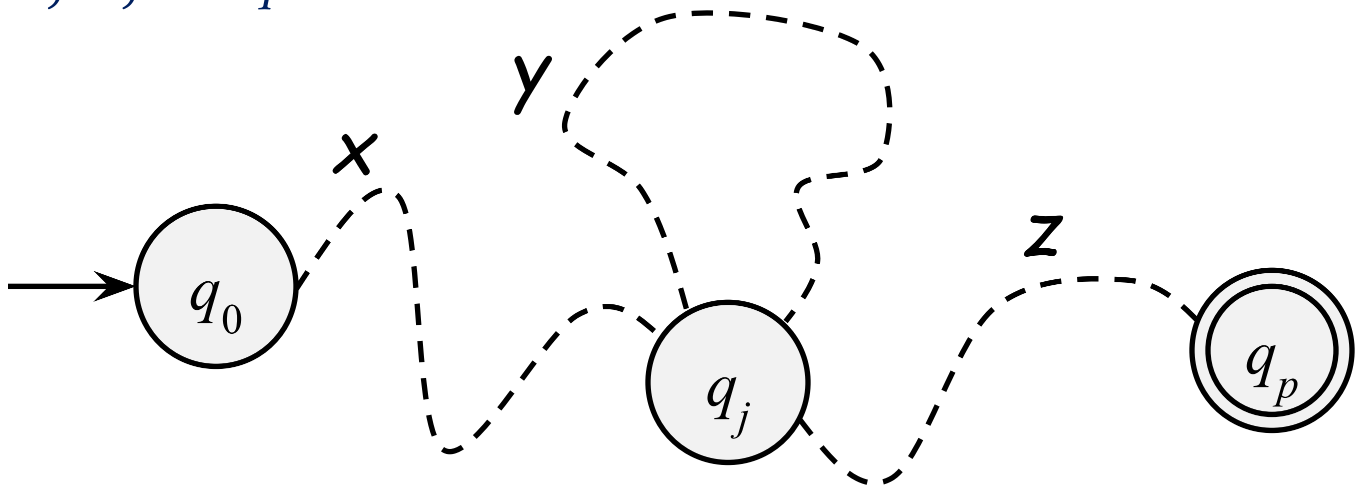
Let $w = w_1 w_2 \cdots w_p \in L$.

$$
\begin{array}{ccccccccccccccccccc}
w = & & w_1 & & w_2 & & w_3 & & w_4 & & w_5 & & w_6 & & w_7 & \dots & & w_p & \\
& \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & \uparrow & & \uparrow \\
& q_0 & & q_1 & & q_2 & & q_3 & & q_4 & & q_5 & & q_6 & & q_7 & q_{p-1} & & q_p
\end{array}
$$



Machine Loops

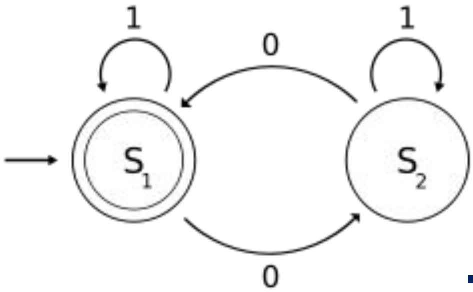
Let q_j be the first repeated state, that is, $q_j = q_{j+k}$ for some k, $0 \leq j < j+k \leq p$.



Where $w = w_1 w_2 \cdots w_p = xyz$,

$x = w_1 w_2 \cdots w_j \qquad y = w_{j+1} \cdots w_{j+k} \qquad z = w_{j+k+1} \cdots w_p$

We conclude that $xy^i z \in L$ for all $i \geq 0$.



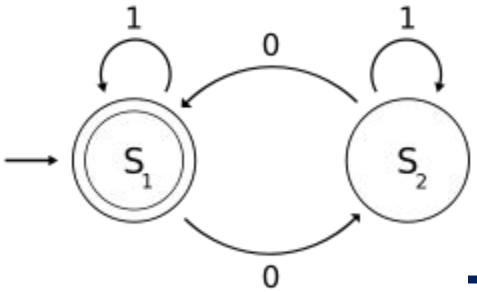
The Pumping Lemma

Theorem. If A is a regular language, then there is a number p (the pumping length) such that for any w $\in$ L with $|w| \geq p$ there exist x,y,z with $w = xyz$, where

1. for each $i \geq 0$, $xy^iz \in A$,

2. $|y| > 0$, and

3. $|xy| \leq p$.



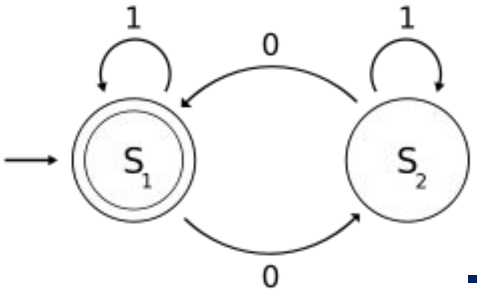
Deciding Regularity

Is $L = \{\ 0^n 1^n : n \geq 0\ \}$ regular? If so, then there are strings x, y, and z such that $xy^iz \in L$ for all $i \geq 0$. What does y look like?

String y consists entirely of 0s?

String y consists entirely of 1s?

String y consists of both 0s and 1s?

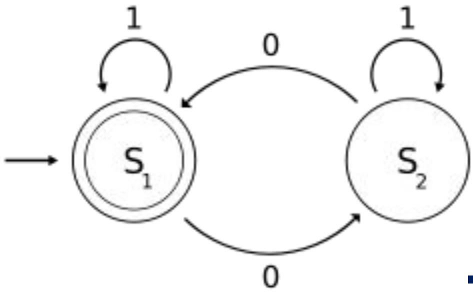


How do we use this?

We will use the pumping lemma to prove a language is NOT regular!

If we have a language that we want to prove is not regular, we will suppose, for a contradiction, that it is regular! Then, the pumping lemma should hold for it!

Then, the fun beings...



Contradicting the pumping lemma

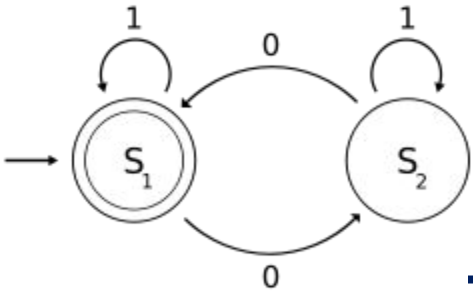
GUIDELINES

Choose a string s in the language of length at least p

Some choices for string s may lead to contradictions to the pumping lemma whereas other choices may not

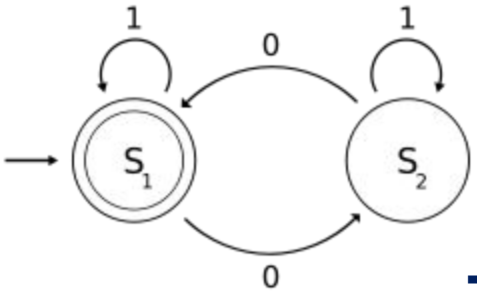
We do not get to choose how s is split into x, y, and z

Usually, try to choose s so that the first p characters of s are all the same, e.g., 0^p or 1^p. Why?
While we don't know exactly what substring y corresponds to, we do know that y resides in the first p characters of s. So if the first p characters of s are, for example, all 0's, then y must be a substring of all 0's. y's length is between 1 and p, i.e., $1 \leq |y| \leq p$



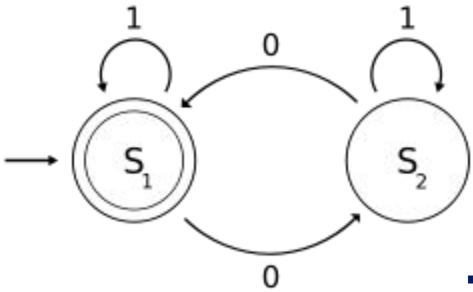
Combining Results

Is $C = \{\ w \mid w$ has an equal number of 0s and 1s $\}$ a regular language?



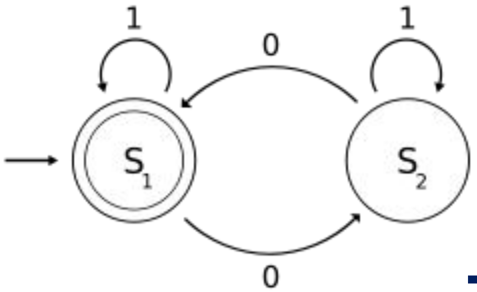
Picking the Right Substring to Pump

Is $PAL = \{\ w \in \{0, 1\}^* \mid w \text{ is a palindrome }\}$ a regular language?



One more pigeon to go...

Can we use this property to prove a language is regular?



Is everything a language?