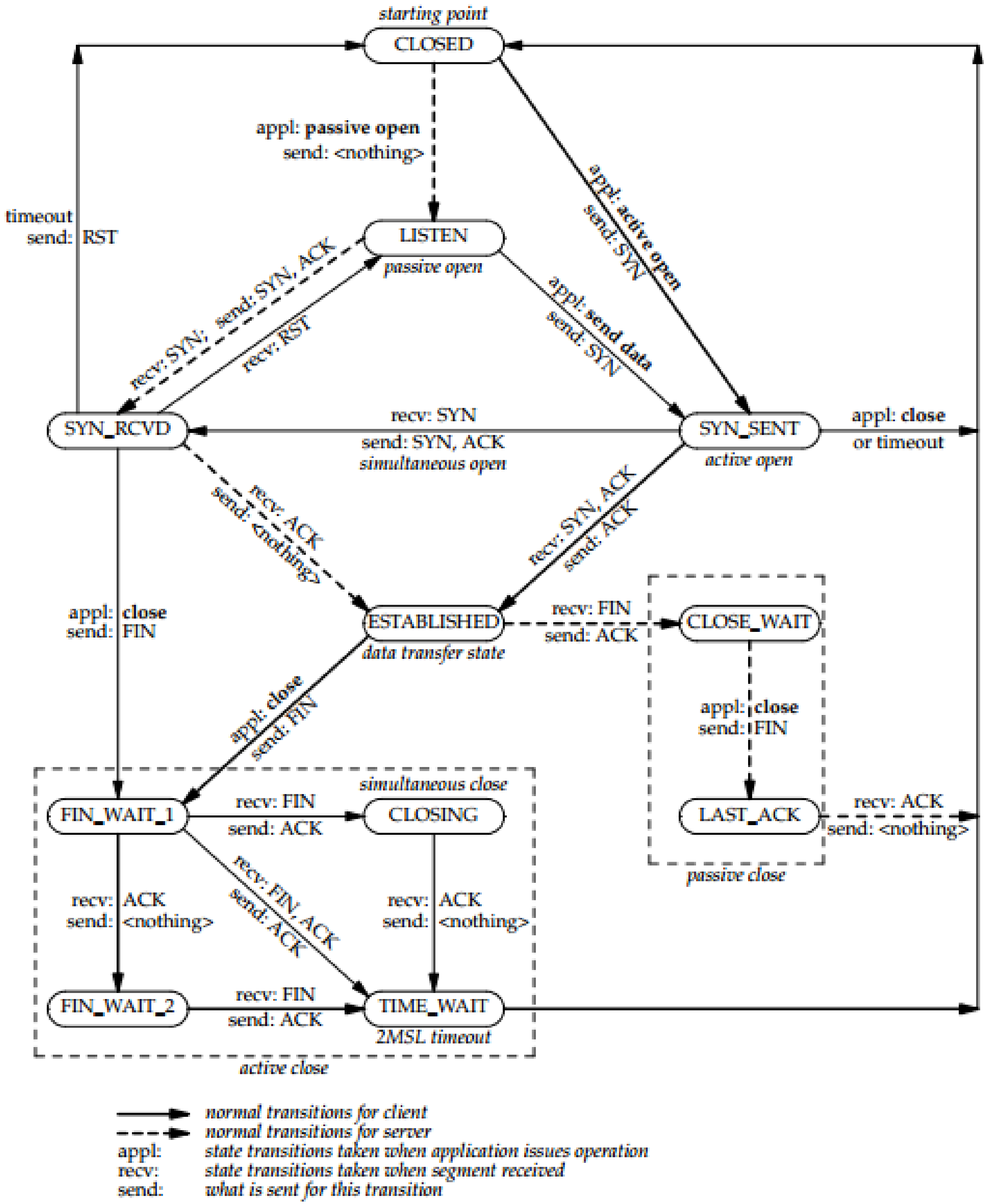
Nonregular Languages

Theorem: The following are all equivalent:

- L is a regular language.
- There is a DFA D such that $\mathscr{L}(D) = L$.
- There is an NFA N such that $\mathscr{L}(N) = L$.
- There is a regular expression R such that $\mathscr{L}(R) = L$.

Buttons as Finite-State Machines:

http://cs103.stanford.edu/button-fsm/



http://www.tti.unipa.it/~gneglia/ip_networks06/slides/TCPIP_State_Transition_Diagram.pdf

Computers as Finite Automata

- My computer has 8GB of RAM and 750GB of hard disk space.
- That's a total of 758GB of memory, which is 6,511,170,420,736 bits.
- There are “only” $2^{6,511,170,420,736}$ possible configurations of my computer.
- Could in principle build a DFA representing my computer, where there's one symbol per type of input the computer can receive.

A Powerful Intuition

- **Regular languages correspond to problems that can be solved with finite memory.**
  - Only need to remember one of finitely many things.
- Nonregular languages correspond to problems that cannot be solved with finite memory.
  - May need to remember one of infinitely many different things.

A Sample Language

- Let $\Sigma = \{$**`a`**, **`b`**$\}$ and consider the following language:

$$L = \{ \mathtt{a}^n \mathtt{b}^n \mid n \in \mathbb{N} \}$$

- That is, L is the language of all strings of n **`a`**'s followed by n **`b`**'s:

{ ε, **`ab`**, **`aabb`**, **`aaabbb`**, **`aaaabbbb`**, … }

- Is this language regular?

$$L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$$

- **Claim:** When any DFA for L is run on any two of the strings ε, **a**, **aa**, **aaa**, **aaaa**, etc., the DFA must end in different states.
- Suppose $\mathtt{a}^n$ and $\mathtt{a}^m$ end up in the same state, where $n \neq m$.
- Then $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$ will end up in the same state. *(Why?)*
- The DFA will either accept a string not in the language or reject a string in the language, which it shouldn't be able to do.
- ***Can't place all these strings into different states; there are only finitely many states!***

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong. Thus L is not regular.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\mathtt{b}^n$ and $\mathtt{a}^m\mathtt{b}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\mathtt{b}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\mathtt{b}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D must end in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong. Thus L is not regular. ■

Why This Matters

- We knew that not all languages are regular, and now we have a concrete example of a nonregular language!
- Intuition behind the proof:
  - Find infinitely many strings that need to be in their own states.
  - Use the pigeonhole principle to show that at least two of them must be in the same state.
  - Conclude the language is not regular.

Practical Concerns

- Webpages are specified using HTML, a markup language where text is decorated with tags.
- Tags can nest arbitrarily but must be balanced:

  <div><div>...<div> </div></div>...</div>

- Using similar logic to the previous proof, can prove that the language

$$\{ \texttt{<div>}^n\texttt{</div>}^n \mid n \in \mathbb{N} \}$$

  is not regular.
- There is no regular expression that can parse HTML documents!

Another Language

- Consider the following language L over the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \stackrel{?}{=}\}$:

$$L = \{\ w\stackrel{?}{=}w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

- L is the language all strings consisting of the same string of **a**'s and **b**'s twice, with a $\stackrel{?}{=}$ symbol in-between.
- Examples:

$$\mathtt{ab}\stackrel{?}{=}\mathtt{ab} \in L \qquad \mathtt{bbb}\stackrel{?}{=}\mathtt{bbb} \in L \qquad \stackrel{?}{=} \in L$$

$$\mathtt{ab}\stackrel{?}{=}\mathtt{ba} \notin L \qquad \mathtt{bbb}\stackrel{?}{=}\mathtt{aaa} \notin L \qquad \mathtt{b}\stackrel{?}{=} \notin L$$

Another Language

$$L = \{\ w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

- This language corresponds to the following problem:

Given strings *x* and *y*, does *x* = *y*?

- Justification: $x = y$ iff $x \stackrel{?}{=} y \in L$.
- **Question:** Is this language regular?

$$L = \{\ w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

- **Claim**: Any DFA for L must place the strings ε, `a`, `aa`, `aaa`, `aaaa`, etc. into separate states.
- Suppose $\mathtt{a}^n$ and $\mathtt{a}^m$ end up in the same state, where $n \neq m$.
- Then $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$ will end up in the same state.
- The DFA will either accept a string not in the language or reject a string in the language, which it shouldn't be able to do.
- ***But that's impossible: we only have finitely many states!***

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction.

Theorem: The language $L = \{\ w \overset{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$.

Theorem: The language $L = \{\, w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$.

Theorem: The language $L = \{\, w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\texttt{a}^n$ and $\texttt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\texttt{a}^n \stackrel{?}{=} \texttt{a}^n$ and $\texttt{a}^m \stackrel{?}{=} \texttt{a}^n$. If this state is accepting, then D accepts $\texttt{a}^m \stackrel{?}{=} \texttt{a}^n$, which is not in L.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$, which is in L.

Theorem: The language $L = \{\ w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^*\}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\texttt{a}^n$ and $\texttt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\texttt{a}^n \stackrel{?}{=} \texttt{a}^n$ and $\texttt{a}^m \stackrel{?}{=} \texttt{a}^n$. If this state is accepting, then D accepts $\texttt{a}^m \stackrel{?}{=} \texttt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\texttt{a}^n \stackrel{?}{=} \texttt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong.

Theorem: The language $L = \{ w\overset{?}{=}w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\texttt{a}^n$ and $\texttt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\texttt{a}^n\overset{?}{=}\texttt{a}^n$ and $\texttt{a}^m\overset{?}{=}\texttt{a}^n$. If this state is accepting, then D accepts $\texttt{a}^m\overset{?}{=}\texttt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\texttt{a}^n\overset{?}{=}\texttt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Theorem: The language $L = \{ w\overset{?}{=}w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\overset{?}{=}\mathtt{a}^n$ and $\mathtt{a}^m\overset{?}{=}\mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\overset{?}{=}\mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\overset{?}{=}\mathtt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular.

Theorem: The language $L = \{\, w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L.

Theorem: The language $L = \{\, w\overset{?}{=}w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\texttt{a}^n$ and $\texttt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\texttt{a}^n\overset{?}{=}\texttt{a}^n$ and $\texttt{a}^m\overset{?}{=}\texttt{a}^n$. If this state is accepting, then D accepts $\texttt{a}^m\overset{?}{=}\texttt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\texttt{a}^n\overset{?}{=}\texttt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ ..., and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$ and $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m \stackrel{?}{=} \mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n \stackrel{?}{=} \mathtt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{a, b\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings a^n and a^m, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings a^n and a^m. Since D is deterministic, D must end in the same state when run on strings $a^n \stackrel{?}{=} a^n$ and $a^m \stackrel{?}{=} a^n$. If this state is accepting, then D accepts $a^m \stackrel{?}{=} a^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $a^n \stackrel{?}{=} a^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings a^0, a^1 …, and a^n, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong.

Theorem: The language $L = \{ w\stackrel{?}{=}w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\texttt{a}^n$ and $\texttt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\texttt{a}^n$ and $\texttt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\texttt{a}^n\stackrel{?}{=}\texttt{a}^n$ and $\texttt{a}^m\stackrel{?}{=}\texttt{a}^n$. If this state is accepting, then D accepts $\texttt{a}^m\stackrel{?}{=}\texttt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\texttt{a}^n\stackrel{?}{=}\texttt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\texttt{a}^0$, $\texttt{a}^1$ …, and $\texttt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong. Thus L is not regular.

Theorem: The language $L = \{\, w \stackrel{?}{=} w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$ is not regular.

Proof: First, we'll prove that if D is a DFA for L, then when D is run on any two different strings $\mathtt{a}^n$ and $\mathtt{a}^m$, the DFA D must end in different states. We proceed by contradiction. Suppose D is a DFA for L where D ends in the same state when run on two distinct strings $\mathtt{a}^n$ and $\mathtt{a}^m$. Since D is deterministic, D must end in the same state when run on strings $\mathtt{a}^n\stackrel{?}{=}\mathtt{a}^n$ and $\mathtt{a}^m\stackrel{?}{=}\mathtt{a}^n$. If this state is accepting, then D accepts $\mathtt{a}^m\stackrel{?}{=}\mathtt{a}^n$, which is not in L. Otherwise, the state is rejecting, so D rejects $\mathtt{a}^n\stackrel{?}{=}\mathtt{a}^n$, which is in L. Both cases contradict that D is a DFA for L, so our assumption was wrong. Thus D ends in different states.

Now we'll prove the theorem. Assume for the sake of contradiction that L is regular. Since L is regular, there must be a DFA D for L. Let n be the number of states in D. When D in run on the strings $\mathtt{a}^0$, $\mathtt{a}^1$ …, and $\mathtt{a}^n$, by the pigeonhole principle since there are $n + 1$ strings and n states, at least two of these strings must end in the same state. Because of the result proven above, we know this is impossible.

We have reached a contradiction, so our assumption was wrong. Thus L is not regular. ■

The General Pattern

- These previous two proofs have the following shape:
  - Find an infinite collection of strings that cannot end up in the same state in any DFA for a language L.
  - Conclude that since any DFA for L has only finitely many states, that L cannot be regular.
- Two questions:
  - What makes the strings unable to end in the same state?
  - Is there a bigger picture here?

Distinguishability

- Let L be a language over Σ.
- Two strings $x, y \in \Sigma^*$ are called **distinguishable relative to L** iff there is some string $w \in \Sigma^*$ where $xw \in L$ and $yw \notin L$.
- In other words, there is some (possibly empty) string w you can append to x and to y where one resulting string is in L and one is not.
- Intuitively: x and y can't end up in the same state in any DFA for L; otherwise, the DFA will be wrong on at least one of xw and yw.

Theorem (Myhill-Nerode): Let L be a language over Σ. If there is a set $S \subseteq \Sigma^*$ with the following properties:

- S is infinite (i.e. it contains infinitely many strings).
- if $x, y \in S$ and $x \neq y$, then x and y are distinguishable relative to L.

then L is not regular.

Proof:

Proof: Let L be an arbitrary language over Σ.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw. If that state is accepting, then D accepts yw, but $yw \notin L$.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw. If that state is accepting, then D accepts yw, but $yw \notin L$. If that state is rejecting, then D rejects xw, but $xw \in L$.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw. If that state is accepting, then D accepts yw, but $yw \notin L$. If that state is rejecting, then D rejects xw, but $xw \in L$. This is impossible, since D is a DFA for L.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw. If that state is accepting, then D accepts yw, but $yw \notin L$. If that state is rejecting, then D rejects xw, but $xw \in L$. This is impossible, since D is a DFA for L. Our assumption was wrong, so L must be nonregular.

Proof: Let L be an arbitrary language over Σ. Let $S \subseteq \Sigma^*$ be an infinite set where for any distinct $x, y \in S$, the strings x and y are distinguishable relative to L. We will prove that L cannot be regular.

We proceed by contradiction; assume L is regular. Since L is regular, there is some DFA D whose language is L. Let the number of states in D be n. Choose any $n + 1$ different strings from S; since S is infinite, such strings must exist. By the pigeonhole principle, since D has n states and there are $n + 1$ strings, at least two of these strings must end in the same state when run through D; call them x and y.

Since x and y are distinguishable relative to L, there must be some string w such that $xw \in L$ and $yw \notin L$. Because D is deterministic and D ends in the same state when run on x and y, the DFA D must end in the same state when run on xw and yw. If that state is accepting, then D accepts yw, but $yw \notin L$. If that state is rejecting, then D rejects xw, but $xw \in L$. This is impossible, since D is a DFA for L. Our assumption was wrong, so L must be nonregular. ■

Using Myhill-Nerode

- To prove that a language L is not regular using the Myhill-Nerode theorem, do the following:
  - Find an infinite set of strings.
  - Prove that any two distinct strings in that set are distinguishable relative to L.
- The tricky part is picking the right strings, but these proofs can be *very* short.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any strings $\mathtt{a}^n, \mathtt{a}^m \in S$ where $\mathtt{a}^n \neq \mathtt{a}^m$.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any strings $\mathtt{a}^n$, $\mathtt{a}^m \in S$ where $\mathtt{a}^n \neq \mathtt{a}^m$. Then $\mathtt{a}^n\mathtt{b}^n \in L$ and $\mathtt{a}^m\mathtt{b}^n \notin L$, so $\mathtt{a}^n$ and $\mathtt{a}^m$ are distinguishable relative to L.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any strings $\mathtt{a}^n, \mathtt{a}^m \in S$ where $\mathtt{a}^n \neq \mathtt{a}^m$. Then $\mathtt{a}^n\mathtt{b}^n \in L$ and $\mathtt{a}^m\mathtt{b}^n \notin L$, so $\mathtt{a}^n$ and $\mathtt{a}^m$ are distinguishable relative to L. Thus S is an infinite set of strings distinguishable relative to L.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any strings $\mathtt{a}^n, \mathtt{a}^m \in S$ where $\mathtt{a}^n \neq \mathtt{a}^m$. Then $\mathtt{a}^n\mathtt{b}^n \in L$ and $\mathtt{a}^m\mathtt{b}^n \notin L$, so $\mathtt{a}^n$ and $\mathtt{a}^m$ are distinguishable relative to L. Thus S is an infinite set of strings distinguishable relative to L. Therefore, by the Myhill-Nerode Theorem, L is not regular.

Theorem: The language $L = \{ \mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any strings $\mathtt{a}^n, \mathtt{a}^m \in S$ where $\mathtt{a}^n \neq \mathtt{a}^m$. Then $\mathtt{a}^n\mathtt{b}^n \in L$ and $\mathtt{a}^m\mathtt{b}^n \notin L$, so $\mathtt{a}^n$ and $\mathtt{a}^m$ are distinguishable relative to L. Thus S is an infinite set of strings distinguishable relative to L. Therefore, by the Myhill-Nerode Theorem, L is not regular. ■

Theorem: The language $L = \{ w \stackrel{?}{=} w \mid w \in \{\texttt{a}, \texttt{b}\}^* \}$ is not regular.

Proof: Let $S = \{ \texttt{a}^n \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each natural number. Now, consider any $\texttt{a}^n, \texttt{a}^m \in S$ where $\texttt{a}^n \neq \texttt{a}^m$. Then $\texttt{a}^n \stackrel{?}{=} \texttt{a}^n \in L$ and $\texttt{a}^m \stackrel{?}{=} \texttt{a}^n \notin L$, so $\texttt{a}^n$ and $\texttt{a}^m$ are distinguishable relative to L. Thus S is an infinite set of strings that are all distinguishable relative to L. Therefore, by the Myhill-Nerode Theorem, L is not regular. ■

Why it Works

- The Myhill-Nerode Theorem is, essentially, a generalized version of the argument from before.
  - If there are infinitely many distinguishable strings and only finitely many states, two distinguishable strings must end up in the same state.
  - Therefore, two strings that cannot be in the same state must end in the same state.
- Proof focuses on the infinite set of strings, not the DFA mechanics.

Announcements!

Midterm Grading

- The TAs and I are going to try to get the midterms graded and returned at the end of Monday's lecture.
- We'll try to get Problem Set 4 graded and returned sometime next week.
- We'll release midterm solutions on Monday along with statistics and common mistakes. If you're curious to learn the answer to any of the problems, please feel free to email us or stop by office hours!

Back to CS103!

Another Language

- Consider the following language L over the alphabet $\Sigma = \{\texttt{a}\}$:

$$L = \{ \texttt{a}^n \mid n \text{ is a power of two } \}$$

- L is the language of all strings of a's whose lengths are powers of two:

$$L = \{ \texttt{a}, \texttt{aa}, \texttt{aaaa}, \texttt{aaaaaaaa}, \dots \}$$

- Question: Is L regular?

Some Math

- Consider any two powers of two 2^n and 2^m where $2 \leq 2^n < 2^m$.
- Then
  - $2^n + 2^n = 2^{n+1}$ is a power of two.
  - $2^m + 2^n = 2^n(2^{m-n} + 1)$ is not a power of two, because $2^{m-n} + 1$ is an odd divisor of $2^m + 2^n$.
- **Idea:** Take our infinite set of strings to be the set of all strings whose length is a power of two greater than or equal to 2.
- Show any pair of strings $\mathtt{a}^{2^n}$, $\mathtt{a}^{2^m}$ in the set are distinguishable by showing $\mathtt{a}^{2^n}$ distinguishes them.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two} \}$ is not regular.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two} \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$.

Theorem: $L = \{ \mathtt{a}^n \mid n$ is a power of two $\}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$.

Theorem: $L = \{ \texttt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \texttt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\texttt{a}^{2^n}, \texttt{a}^{2^m} \in S$ be any two strings in S where $\texttt{a}^{2^n} \neq \texttt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\texttt{a}^{2^n}\texttt{a}^{2^n}$ and $\texttt{a}^{2^m}\texttt{a}^{2^n}$. The string $\texttt{a}^{2^n}\texttt{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\texttt{a}^{2^n}\texttt{a}^{2^n} \in L$.

Theorem: $L = \{ \mathbf{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathbf{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathbf{a}^{2^n}, \mathbf{a}^{2^m} \in S$ be any two strings in S where $\mathbf{a}^{2^n} \neq \mathbf{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathbf{a}^{2^n}\mathbf{a}^{2^n}$ and $\mathbf{a}^{2^m}\mathbf{a}^{2^n}$. The string $\mathbf{a}^{2^n}\mathbf{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\mathbf{a}^{2^n}\mathbf{a}^{2^n} \in L$. However, string $\mathbf{a}^{2^m}\mathbf{a}^{2^n}$ has length $2^m + 2^n = 2^n(2^{m-n} + 1)$.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$. The string $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\mathtt{a}^{2^n}\mathtt{a}^{2^n} \in L$. However, string $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$ has length $2^m + 2^n = 2^n(2^{m-n} + 1)$. Since $m > n$, we know $2^{m-n} + 1$ is odd, and so $2^m + 2^n$ is not a power of two because it has an odd factor.

Theorem: $L = \{ \mathtt{a}^n \mid n \text{ is a power of two } \}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$. The string $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\mathtt{a}^{2^n}\mathtt{a}^{2^n} \in L$. However, string $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$ has length $2^m + 2^n = 2^n(2^{m\text{-}n} + 1)$. Since $m > n$, we know $2^{m\text{-}n} + 1$ is odd, and so $2^m + 2^n$ is not a power of two because it has an odd factor. Thus $\mathtt{a}^{2^m}\mathtt{a}^{2^n} \notin L$, so $\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}$ are distinguishable relative to L.

Theorem: $L = \{ \mathtt{a}^n \mid n$ is a power of two $\}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$. The string $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\mathtt{a}^{2^n}\mathtt{a}^{2^n} \in L$. However, string $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$ has length $2^m + 2^n = 2^n(2^{m-n} + 1)$. Since $m > n$, we know $2^{m-n} + 1$ is odd, and so $2^m + 2^n$ is not a power of two because it has an odd factor. Thus $\mathtt{a}^{2^m}\mathtt{a}^{2^n} \notin L$, so $\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}$ are distinguishable relative to L.

Since S is an infinite set of strings distinguishable relative to L, by the Myhill-Nerode Theorem L is not regular.

Theorem: $L = \{ \mathtt{a}^n \mid n$ is a power of two $\}$ is not regular.

Proof: Let $S = \{ \mathtt{a}^{2^n} \mid n \in \mathbb{N} \}$. This set is infinite because it contains one string for each positive natural number. Let $\mathtt{a}^{2^n}, \mathtt{a}^{2^m} \in S$ be any two strings in S where $\mathtt{a}^{2^n} \neq \mathtt{a}^{2^m}$. Assume without loss of generality that $n < m$, so $1 \leq n < m$.

Consider the strings $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$. The string $\mathtt{a}^{2^n}\mathtt{a}^{2^n}$ has length 2^{n+1}, which is a power of two, so $\mathtt{a}^{2^n}\mathtt{a}^{2^n} \in L$. However, string $\mathtt{a}^{2^m}\mathtt{a}^{2^n}$ has length $2^m + 2^n = 2^n(2^{m-n} + 1)$. Since $m > n$, we know $2^{m-n} + 1$ is odd, and so $2^m + 2^n$ is not a power of two because it has an odd factor. Thus $\mathtt{a}^{2^m}\mathtt{a}^{2^n} \notin L$, so $\mathtt{a}^{2^n}$ and $\mathtt{a}^{2^m}$ are distinguishable relative to L.

Since S is an infinite set of strings distinguishable relative to L, by the Myhill-Nerode Theorem L is not regular. ■

What Comes Next

- What does it mean to compute with infinite memory?
- What classes of languages lie beyond the regular languages?
- And how will we reason about them?
- We shall see!

Next Time

- **Context-Free Languages**
  - Context-Free Grammars
  - Generating Languages