BU CS 332 – Theory of Computation

https://forms.gle/QHsstTW4vZS2CgDV7

Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

"Myhill-Nerode" note

Alexander Poremba & Mark Bun

February 10, 2026

Last Time

- Regular expressions characterize the regular languages
  - Every regex can be converted to an NFA recognizing its language
  - Every NFA can be converted to a regex generating its language

NFA → Regex

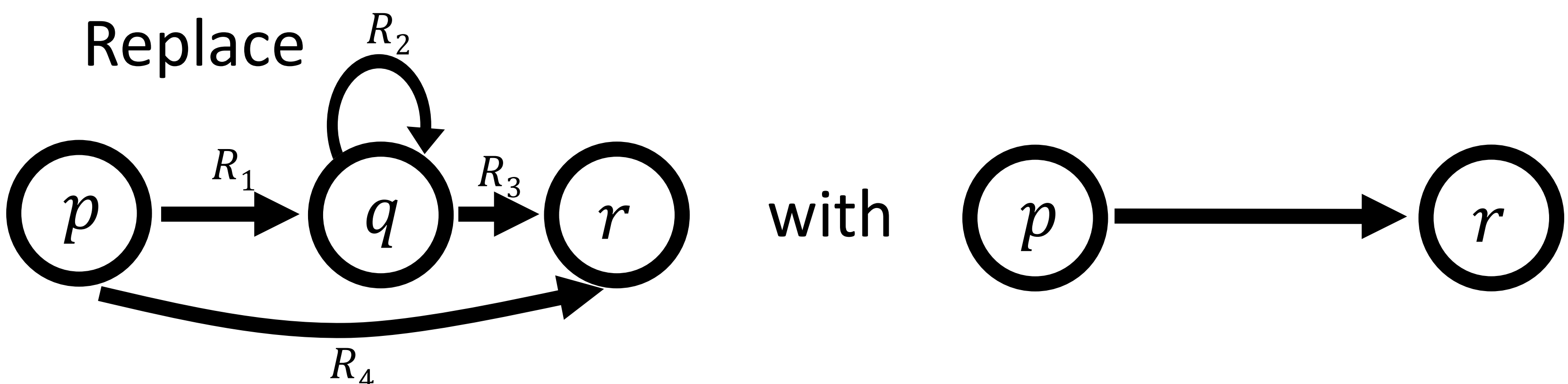
Theorem 2: Every NFA has an equivalent regular expression

Given NFA N:

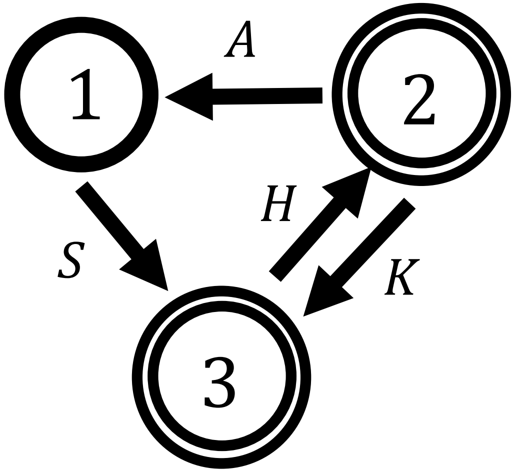
1. Make N into a GNFA by adding new start and accept state
2. For each non-start and non-accept state q:

   For each pair of states p, r:

   Replace



3. Output regex labeling transition from start to accept state


1
2
3
A
S
H
K

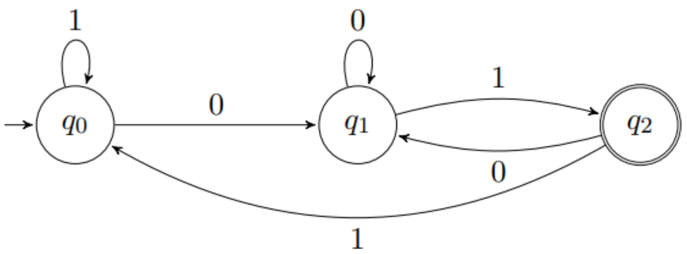
Limits of Finite Automata

Motivating Questions

- We’ve seen several techniques for showing languages are regular

- How can we tell if we’ve found the smallest DFA recognizing a given language?
- Which languages are not regular? How can we prove so?

An Example



$A = \{\, w \in \{0,1\}^* \mid w \text{ ends with } 01 \,\}$

Claim: *Every* DFA recognizing A needs at least 3 states

Proof: Let M be any DFA recognizing A. Consider running M on each of $x = \varepsilon, y = 0, w = 01$

$A = \{ w \in \{0,1\}^* \mid w \text{ ends with } 01 \}$

A General Technique

Definition: Strings x and y are **distinguishable** by L if there exists a "distinguishing extension" $z \in \Sigma^*$ such that exactly one of xz or yz is in L.

Ex. $x = \varepsilon, \quad y = 0$

Definition: A set of strings S is **pairwise distinguishable** by L if every pair of distinct strings $x, y \in S$ is distinguishable by L.

Ex. $S = \{\varepsilon, 0, 01\}$

A General Technique

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L must have at least $|S|$ states

Proof: Let M be a DFA with $< |S|$ states.

Claim: There are distinct strings $x, y \in S$ such that M ends up in same state on x and y

Another Example

$B = \{\, w \in \{0,1\}^* \mid |w| = 2 \,\}$

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L must have at least $|S|$ states

$S = \{ \qquad\qquad\qquad\qquad \}$

Distinguishing Extension

Which of the following is a distinguishing extension for $x = 0$ and $y = 00$ for language $B = \{ w \in \{0, 1\}^* \mid |w| = 2 \}$?

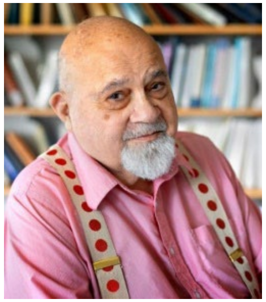
a) $z = \varepsilon$
b) $z = 0$
c) $z = 1$
d) $z = 00$

Historical Note

Converse to the distinguishing set method:

If L has **no** distinguishing set of size $> k$, then L is recognized by a DFA with k states

Myhill-Nerode Theorem (1958): L is recognized by a DFA with $\leq k$ states **if and only if** L does not have a distinguishing set of size $> k$



Non-Regularity

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L must have at least $|S|$ states

Corollary: If S is an **infinite** set that is pairwise distinguishable by L, then no DFA recognizes L

The Classic Example

Theorem: $A = \{0^n 1^n \mid n \geq 0\}$ is not regular

Proof: We construct an infinite pairwise distinguishable set

Palindromes

Theorem: $L = \{w \in \{0,1\}^* \mid w = w^R\}$ is not regular

Proof: We construct an infinite pairwise distinguishable set

Now you try!

Use the distinguishing set method to show that the following languages are not regular

$L_1 = \{0^i 1^j \mid i > j \geq 0\}$

<u>Your job:</u> Build an infinite set S such that for all $x \neq y \in S$, there exists a z such that exactly one of xz and yz is in L

Now you try!

Use the distinguishing set method to show that the following languages are not regular

$$L_2 = \{1^{n^2} \mid n \geq 0\}$$

Reusing a Proof

Finding a distinguishing set can take some work…

Let's try to reuse that work!

How might we show that

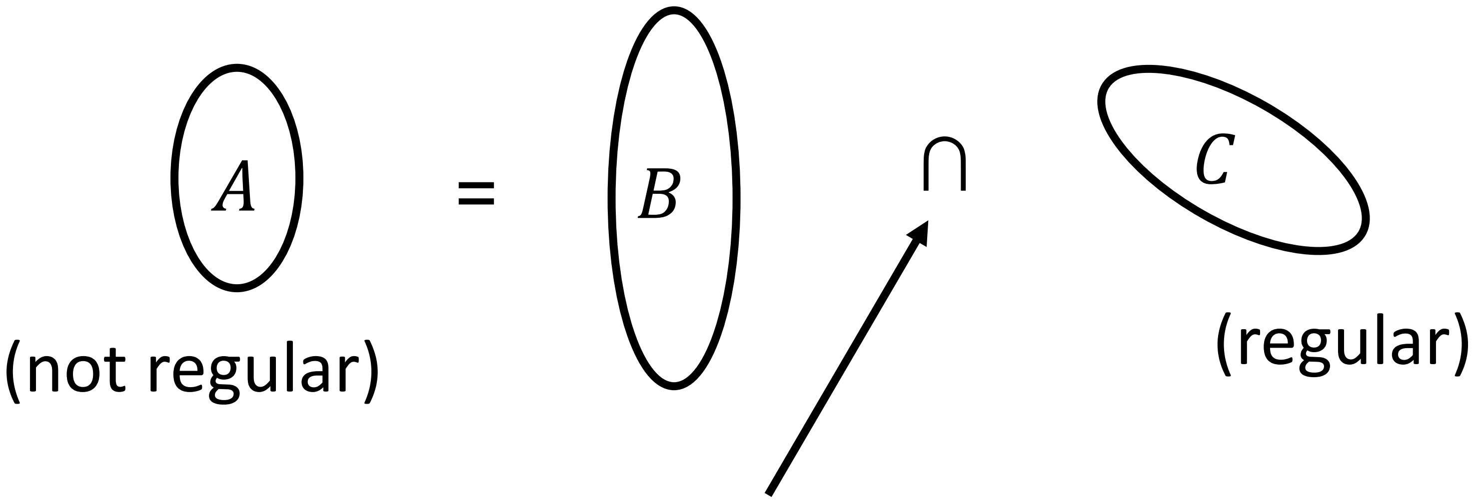
$BALANCED$ = $\{w \mid w$ has an equal # of 0s and 1s$\}$

is not regular?

$\{0^n 1^n \mid n \geq 0\}$ = $BALANCED \cap \{w \mid$ all 0s in w appear before all 1s$\}$

Using Closure Properties

If A is not regular, we can show a related language B is not regular



any of $\{\circ, \cup, \cap\}$ or, for one language, $\{\neg, {}^{R}, {}^{*}\}$

<u>By contradiction:</u> If B is regular, then $B \cap C$ $(= A)$ is regular.

But A is not regular so neither is B!

Example

Prove $B = \{0^i 1^j | i \neq j\}$ is not regular using

- Nonregular language

  $A = \{0^n 1^n | n \geq 0\}$ and

- Regular language

  $C = \{w \mid$ all 0s in w appear before all 1s$\}$

Which of the following expresses A in terms of B and C?

a) $A = B \cap C$

b) $A = \bar{B} \cap C$

c) $A = B \cup C$

d) $A = \bar{B} \cup C$

Proof that B is nonregular

Assume for the sake of contradiction that B is regular

We know: $A = \bar{B} \cap C$

!DANGER!

Let $B = \{0^i 1^j | i \neq j\}$ and write $B = A \cup C$ where

- Nonregular language

$$A = \left\{0^i 1^j \middle| i > j \geq 0\right\}$$ and

- Nonregular language

$$C = \left\{0^i 1^j \middle| j > i \geq 0\right\}$$ and

Does this let us conclude B is nonregular?