Pre-lecture brain teaser

Prove at the following languages are regular:

- All strings that end in $1011$
  
  \[ \text{RE: } (0+1)^*1011 \]

- All strings that contain $101$ or $010$ as a substring.
  
  \[ \text{RE: } (0+1)^* (101 + 010) (0+1)^* \]

- All strings that do not contain $111$ as a substring.
  
  \[ \text{RE: } \checkmark \]
ECE-374-B: Lecture 4 - RegExp-DFA-NFA Equivalence

Instructor: Abhishek Kumar Umrawal
January 30, 2024

University of Illinois at Urbana-Champaign
Pre-lecture brain teaser

Prove at the following languages are regular:

• All strings that end in 1011

• All strings that contain 101 or 010 as a substring.

• All strings that do not contain 111 as a substring.
Regular Languages, DFAs, NFAs

Theorem
Languages accepted by **DFAs**, **NFA**s, and regular expressions are the same.
Theorem
Languages accepted by **DFAs**, **NFA**s, and regular expressions are the same.

- DFAs are special cases of **NFA**s (easy)
- **NFA**s accept regular expressions (seen)
- DFAs accept languages accepted by **NFA**s (shortly)
- Regular expressions for languages accepted by **DFA**s (shown previously)
Theorem
Languages accepted by DFAs, NFAs, and regular expressions are the same.
Theorem
Languages accepted by DFAs, NFAs, and regular expressions are the same.
Regular Expression to NFA
Proving equivalence

- Regular expressions
- NFAs
- DFAs
Thompson’s algorithm

Given two NFAs $s$ and $t$:

$L = L_s \cdot L_t$

$L = L_s \cup L_t$

$L = (L_s)^*$
Let’s take a regular expression and convert it to a DFA.

Example: \((0 + 1)^* (101 + 010)(0 + 1)^*\)

\((0+1)^* = L_1^* \quad \text{where} \quad L_1 = 0 + 1\)

\[ L_1 = L_1^{(1)} + L_1^{(2)} \]

\[ = 0 + 1 \]
Let's take a regular expression and convert it to a DFA.
Example: \((0 + 1)^* (101 + 010)(0 + 1)^*\)
Find NFA for \((0 + 1)^*\)
Regular expression to NFA example

Find NFA for \((0 + 1)^*\)
Find NFA for \((0 + 1)^*\)
Find NFA for \((101 + 010)\)
Find NFA for \((101 + 010)\)
Find NFA for \((101 + 010)\)
Let’s take a regular expression and convert it to a NFA.

Example: \((0 + 1)^* (101 + 010)(0 + 1)^*\)
Let's take a regular expression and convert it to a NFA.

Example: \((0 + 1)^* (101 + 010)(0 + 1)^*\)

Using the concatenation rule:
Let’s take a regular expression and convert it to a NFA.

Example: \((0 + 1)^*(101 + 010)(0 + 1)^*\)

Using the concatenation rule:

What does Thompson’s algorithm mean?!
Equivalence of NFAs and DFAs

DFA $\rightarrow^? NFA$

NFA $\rightarrow^? DFA$
Another Way to look at NFAs

Is 010110 accepted?
Is $010110$ accepted?
Another Way to look at NFAs

Is 010110 accepted?
Another Way to look at NFAs

Is 010110 accepted?

0

1
Another Way to look at NFAs

Is 010110 accepted?
Another Way to look at NFAs

Is 010110 accepted?

```
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
```

```
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
start
q0 0,1 1 q1 0 q2 1 q3
```
Another Way to look at NFAs

Is 010110 accepted?
Another Way to look at NFAs

Is 010110 accepted?
Conversion of NFA to DFA
Proving equivalence

Regular expressions

Thompson’s Alg.

NFAs

DFAs
Theorem
For every NFA $N$ there is a DFA $M$ such that $L(M) = L(N)$. 
DFAs are memoryless...

- DFA knows only its current state.
- The state is the memory.
- To design a DFA, answer the question: What minimal info needed to solve problem.
Simulating NFA

NFAs know many states at once on input 010110.
It is easy to state that the state of the automata is the states that it might be situated at.
The state of the NFA

It is easy to state that the state of the automata is the states that it might be situated at.

configuration: A set of states the automata might be in.
The state of the NFA

It is easy to state that the state of the automata is the states that it might be situated at.

configuration: A set of states the automata might be in.

Possible configurations: \( \mathcal{P}(q) = \emptyset, \{q_0\}, \{q_0, q_1\}, \ldots \)
It is easy to state that the state of the automata is the states that it might be situated at.

configuration: A set of states the automata might be in.

Possible configurations: $\mathcal{P}(q) = \emptyset$, \{q$_0$\}, \{q$_0$, q$_1$\}...

Big idea: Build a DFA on the configurations.
Example

If receives 0:

If receives 1:
Example

If receives 0:

If receives 1:
Simulating an NFA by a DFA

• Think of a program with fixed memory that needs to simulate NFA $N$ on input $w$.
• What does it need to store after seeing a prefix $x$ of $w$?
• It needs to know at least $\delta^*(s, x)$, the set of states that $N$ could be in after reading $x$
• Is it sufficient?
Simulating an NFA by a DFA

- Think of a program with fixed memory that needs to simulate NFA $N$ on input $w$.
- What does it need to store after seeing a prefix $x$ of $w$?
  - It needs to know at least $\delta^*(s, x)$, the set of states that $N$ could be in after reading $x$.
- Is it sufficient? Yes, if it can compute $\delta^*(s, xa)$ after seeing another symbol $a$ in the input.
- When should the program accept a string $w$? If $\delta^*(s, w) \cap A \neq \emptyset$.

**Key Observation:** DFA $M$ simulating $N$ should know current configuration of $N$.

State space of the DFA is $\mathcal{P}(Q)$. 
Definition
A non-deterministic finite automata (NFA) \( N = (Q, \Sigma, \delta, s, A) \) is a five tuple where

- \( Q \) is a finite set whose elements are called states,
- \( \Sigma \) is a finite set called the input alphabet,
- \( \delta : Q \times \Sigma \cup \{\epsilon\} \rightarrow \mathcal{P}(Q) \) is the transition function (here \( \mathcal{P}(Q) \) is the power set of \( Q \)),
- \( s \in Q \) is the start state,
- \( A \subseteq Q \) is the set of accepting/final states.

\( \delta(q, a) \) for \( a \in \Sigma \cup \{\epsilon\} \) is a subset of \( Q \) — a set of states.
NFA $N = (Q, \Sigma, \emptyset, \delta, A)$. We create a DFA $D = (Q', \Sigma, \delta', s', A')$ as follows:

- $Q' = \mathcal{P}(Q)$
- $s' = \text{reach}(s)$
- $A' = \{ X \subseteq Q \mid X \cap A \neq \emptyset \}$
- $\delta'(X, a) = \bigcup_{q \in X} (\delta^*(a, a))$ for each $X \subseteq Q$, $a \in \Sigma$
NFA $\rightarrow$ DFA

$\langle Q, \Sigma, s, \delta, A \rangle \rightarrow \langle Q', \Sigma', s', \delta', A' \rangle$

1. $Q' = P(Q)$
2. $\Sigma' = \Sigma$

$n = 2^{\{q_0, q_1, s\}} \Rightarrow \{\emptyset, \{q_0\}, \{q_1\}, \{s\}, \{q_0, q_1\}, \{q_0, s\}, \{q_1, s\}, \{q_0, q_1, s\}\}$

. Evaluate $s'$
   using the formula.

Observation: We don't have worry about the states that cannot be visited from the $\text{reach}(s)$
All bitstrings where the 2nd last bit is 1.

\[
\begin{array}{c|c|c}
q & 0 & 1 \\
\hline
q_0 & \{q_0\} & \{q_0, q_1\} \\
q_1 & \{q_2\} & \{q_2\} \\
q_2 & \emptyset & \emptyset \\
\end{array}
\]

\[s: Q \times \Sigma \cup \varepsilon \rightarrow P(Q)\]

"INCREMENTAL CONSTRUCTION"

RIY: What happens when the given NFA has \(\varepsilon\) transitions?
DFAs to Regular expressions
Proving equivalence

- Regular expressions
- NFAs
- DFAs
- Thompson's Algorithm
- Subset Construction
State Removal method

If $q_1 = \delta(q_0, x)$ and $q_2 = \delta(q_1, y)$

then $q_2 = \delta(q_1, y) = \delta(\delta(q_0, x), y) = \delta(q_0, xy)$
We want to remove $q_1$:
State Removal method - Example
State Removal method - Example

\[

text
\]

\[
\begin{align*}
q_0 & \xrightarrow{01} q_0 \quad \text{start} \\
q_0 & \xrightarrow{1+00} q_2 \\
q_0 & \xrightarrow{0+11} q_2 \\
q_2 & \xrightarrow{10} q_2
\end{align*}
\]
State Removal method - Example

(01)* + (1 + 00)(10)* (0 + 11)

Start from the start state and remove all except the accept states.

RE: \((01)^* + (1 + 00)(10)^* (0 + 11)\)^*
Multiple accept states:

- Find expressions for each accept state and take a union!
State Removal method - Example

\[(01)^* + (1 + 00)(10)^*(0 + 11)\]
Algebraic method

Transition functions are themselves algebraic expressions!

Demarcate states as variables.

Can rewrite $q_1 = \delta(q_0, x)$ as $q_1 = q_0x$

Solve for accepting state.
Algebraic method - Example

\[ q_0 = q_1 + q_2 \]
\[ q_1 = q_0 \]
\[ q_2 = q_3 \]
\[ q_3 = q_0 + q_2 \]
Algebraic method - Example

- $q_0 = \epsilon + q_11 + q_20$
- $q_1 = q_00$
- $q_2 = q_01$
- $q_3 = q_10 + q_21 + q_3(0+1)$
Algebraic method - Example

- $q_0 = \varepsilon + q_11 + q_20$
- $q_1 = q_00$
- $q_2 = q_01$
- $q_3 = q_10 + q_21 + q_3(0 + 1)$

Now we simply solve the system of equations for $q_0$:

- $q_0 = \varepsilon + q_11 + q_20$
- $q_0 = \varepsilon + q_001 + q_010$
- $q_0 = \varepsilon + q_0(01 + 10)$

Theorem (Arden’s Theorem)

$R = Q + RP = QP^*$
Algebraic method - Example

\[
\begin{align*}
q_0 &= \epsilon + q_1 1 + q_2 0 \\
q_1 &= q_0 0 \\
q_2 &= q_0 1 \\
q_3 &= q_1 0 + q_2 1 + q_3 (0 + 1)
\end{align*}
\]

Now we simply solve the system of equations for \( q_0 \):

\[
\begin{align*}
q_0 &= \epsilon + q_1 1 + q_2 0 \\
q_0 &= \epsilon + q_0 01 + q_0 10 \\
q_0 &= \epsilon + q_0 (01 + 10) \\
q_0 &= \epsilon (01 + 10)^* = (01 + 10)^*
\end{align*}
\]
Converting NFAs to Regular Expression
Proving equivalence

Regular expressions

- NFAs
  - Thompson’s Alg.
  - Algebraic Method
  - Subset Construction

- DFAs
Stage 0: Input
Stage 1: Normalizing

A
\[ a, b \]

B
\[ b, a \]

C
\[ a, b \]

init
\[ \epsilon \]

A
\[ a \]

B
\[ b \]

C
\[ \epsilon \]

AC
\[ a + b \]
Stage 2: Remove state A
Stage 4: Redrawn without old edges
Stage 4: Removing B
Stage 5: Redraw

\[ ab^*a + b \]

\[ C \xrightarrow{\epsilon} AC \]

\[ a + b \]
Stage 6: Removing C

\[(ab^*a + b)(a + b)^* \epsilon\]
Stage 7: Redraw

\[(ab \cdot a + b)(a + b)\]
Thus, this automata is equivalent to the regular expression

$$(ab^*a + b)(a + b)^*.$$

Regular expressions to DFAs
Proving equivalence

- Regular expressions
- NFAs
- DFAs
- Algebraic Method
- Thompson's Alg.
- Subset Construction
- State removal
Difficulty going from RegEx’s to DFAs

Lemma
Many regular expressions cannot be easily converted to DFAs.
Lemma
Many regular expressions cannot be easily converted to DFAs.

Consider \( w \in \Sigma^* | w \) has a substring 010 or 101
Lemma
Many regular expressions cannot be easily converted to DFAs.

Consider \( \{ w \in \Sigma^* | w \) has a substring 010 or 101}\)

- Is possible using Brzozowski\(^1\) algorithm. Not needed for this course.
But here’s the idea anyway....

Draw the DFA for \( w = \{ w \in \Sigma^* | w \text{ has a substring } 010 \} \). What does each state represent?
Brzozowski Method

Brings us to the **Brzozowski derivative** where \((u^{-1}S)\) of a set \(S\) of strings and a string \(u\) is the set of strings obtainable from a string in \(S\) by cutting off the prefixing \(u\).

Consider the language \(R = (ab + c)^*\)
Brzozowski Method

Brings us to the **Brzozowski derivative** where \((u^{-1}S)\) of a set \(S\) of strings and a string \(u\) is the set of strings obtainable from a string in \(S\) by cutting of the prefixing \(u\).

Consider the language \(R = (ab + c)^*\)

<table>
<thead>
<tr>
<th>(R)</th>
<th>(a^{-1}R)</th>
<th>(b^{-1}R)</th>
<th>(c^{-1}R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q_0 = \varepsilon^{-1}R = (ab + c)^*)</td>
<td>(b(ab + c)^*)</td>
<td>(\emptyset)</td>
<td>((ab + c)^*)</td>
</tr>
<tr>
<td>(q_1 = b(ab + c)^*)</td>
<td>(\emptyset)</td>
<td>((ab + c)^*)</td>
<td>(\emptyset)</td>
</tr>
<tr>
<td>(q_2 = \emptyset)</td>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
</tr>
</tbody>
</table>
Brzozowski Method

Brings us to the Brzozowski derivative where \((u^{-1}S)\) of a set \(S\) of strings and a string \(u\) is the set of strings obtainable from a string in \(S\) by cutting of the prefixing \(u\).

Consider the language \(R = (ab + c)^*\)

<table>
<thead>
<tr>
<th></th>
<th>(a^{-1}R)</th>
<th>(b^{-1}R)</th>
<th>(c^{-1}R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q_0 = \varepsilon^{-1}R = (ab + c)^*)</td>
<td>(b (ab + c)^*)</td>
<td>(\emptyset)</td>
<td>((ab + c)^*)</td>
</tr>
<tr>
<td>(q_1 = b (ab + c)^*)</td>
<td>(\emptyset)</td>
<td>((ab + c)^*)</td>
<td>(\emptyset)</td>
</tr>
<tr>
<td>(q_2 = \emptyset)</td>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
</tr>
</tbody>
</table>

```
R = (ab + c)^* 
q_0 = \varepsilon^{-1}R = (ab + c)^* 
q_1 = b (ab + c)^* 
q_2 = \emptyset 
```

![Diagram of a finite automaton](image)
Difficulty going from RegEx’s to DFAs

Lemma
Many regular expressions cannot be easily converted to DFAs.

Consider \( \{ w \in \Sigma^* | w \text{ has a substring } 010 \text{ or } 010 \} \)

- Is possible using Brzozowski\(^2\) algorithm. *Not needed for this course.*
- Easier to just convert RegEx \( \rightarrow \) NFA \( \rightarrow \) DFA.
Conclusion
Proving equivalence

- Regular expressions
  - DFA
  - NFA
- Algebraic Method
- Subset Construction
- State removal
- Thompson’s Alg.
  - NFA → DFA

But what about the expressions aren’t regular?!
Proving equivalence

But what about the expressions at aren’t regular?! See on Thursday