Lecture 8: Announcements later.

Last time: Let L_1, L_2 be regular languages

$$L_1 \cdot L_2 = \{ xy \mid x \in L_1, y \in L_2 \}$$

$\{0, 11\} \cdot \{00\} = \{000, 1100\}$

$$L_1^k = \{ w \mid w = x_1 \cdots x_k, \text{ each } x_i \in L \}$$

$\{0,11\}^0 = \{\varepsilon\}$ $\quad \{0,11\}^2 = \{00, 110, 1111, 011\}$

$$L_1^* = \{ w \mid w = x_1 \cdots x_k, \text{ for some } k \geq 0 \text{ each } x_i \in L \} = \bigcup_{k=0}^{\infty} L_1^k$$

$\{0,11\}^* = \{\varepsilon\} \cup \{0,11\} \cup \{00, 110, 1111, 0111\} \cup \ldots$

Regular languages are closed under $\cdot$, power, $*$

We saw a new proof that if L_1, L_2 are regular

then $L_1 \cup L_2$ is regular.

Def: the regular operators are $\cup, *$,

Today: Regular expressions.

DFAs, NFAs they are "machines",
they have a syntax (5-tuple) and define
a language

Def: a Regular Expression (regexp) over
alphabet Σ is R such that:

base cases
$$\begin{cases} R = \phi \\ R = \varepsilon \\ R = a \quad \text{for some } a \in \Sigma \end{cases}$$

inductive cases
$$\begin{cases} R = (R_1 \cup R_2) \\ R = (R_1 \circ R_2) \\ R = (R_1)^* \end{cases}$$
where R_1, R_2 are regular expressions

Each regexp has an associated language $L(R)$
defined as follows:

* $L(\emptyset) = \{\}$ ← the regexp $\emptyset$

* $L(\varepsilon) = \{\varepsilon\}$

* $L(a) = \{a\} \quad \forall\, a \in \Sigma$

} 3 base cases

* $L((R_1 \cup R_2)) = L(R_1) \cup L(R_2)$
* $L((R_1 \cdot R_2)) = L(R_1) \cdot L(R_2)$
* $L((R_1)^*) = (L(R_1))^*$

Note: we sometimes omit parentheses
* order of operations: $*$, $\cdot$, $\cup$
* concatenation symbol is sometimes omitted

eg: $ab \cup ab^*a \rightarrow ((a \cdot b) \cup ((a \cdot (b)^*) \cdot a))$

Let R be some regexp:

$R \cdot \varepsilon \overset{L?}{\rightsquigarrow} L(R)$

$L(R \cdot \varepsilon) = L(R) \cdot L(\varepsilon)$

$= L(R) \cdot \{\varepsilon\}$

$= \{xy \mid x \in L(R), y \in \{\varepsilon\}\}$

$R \cdot \emptyset \overset{L?}{\rightsquigarrow} L(R)?$

$\emptyset$?

$L(R \cdot \phi) = L(R) \cdot L(\phi)$

$= L(R) \cdot \{\}$

$= \{xy \mid x \in L(R), y \in \{\}\}$

$= \{\}$ there's no y to pick.

$= \{x \cdot \varepsilon \mid x \in L(R)\}$

$= \{x \mid x \in L(R)\}$

$L(R^*) = L(R)^*$

$\phi^* \Rightarrow^{L?} L(\phi^*) = L(\phi)^* = \{\}^*$

$= \{w \mid w$ is the concatenation of $k \geq 0$ strings from $\{\}\}$

$= \{\varepsilon\}$

$\varepsilon^* \overset{L?}{\Rightarrow} L(\varepsilon^*) = L(\varepsilon)^*$

$= \{\varepsilon\}^*$

$= \{\varepsilon\}$

Given R, what's $L(R)$?

$(a \cup b)^* \rightarrow (L(a) \cup L(b))^* = (\{a\} \cup \{b\})^*$

$= \{a, b\}^*$

$\underbrace{(0 \cup 1)^*}_{\text{any } x \in \{0,1\}^*} \underbrace{00}_{} \underbrace{(0 \cup 1)^*}_{\text{any } y \in \{0,1\}^*} = \{0,1\}^* \cdot \{00\} \cdot \{0,1\}^*$

$= \{w \mid w$ contains $00\}$

$x\,00\,y$

$((0 \cup 1)(0 \cup 1)(0 \cup 1))^* = \{w \in \{0,1\}^* \mid |w| \text{ is } 0 \bmod 3\}$

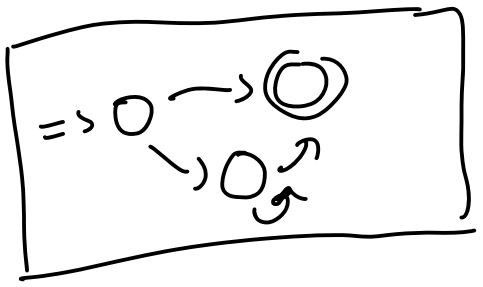
$\{0,1\}^3 = \{$ all strings over $\{0,1\}$ of length $3\}$

Thm: L is regular $\iff \exists$ a regular expression R with $L(R) = L$.

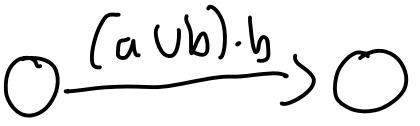
(1) If you have an NFA N with $L(N) = L$, then you can build R with $L(R) = L$.

It's in the book

Idea:



we can define a generalized NFA where transitions are labelled with regexp



There is an algorithm that removes states 1 by 1 (of the generalized NFA), at the end you get



(2) If there is a regular expression R with

$L(R) = L$ then L is regular.

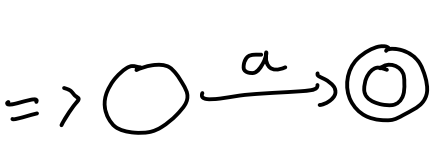
Proof: by (strong) induction on the number of operators $\cup, *, \circ$ in R

Base case: R has $k = 0$ operators

$R = \emptyset \Rightarrow L(R) = \emptyset$ $\Rightarrow$ 

$R = \varepsilon \Rightarrow L(R) = \{\varepsilon\}$ $\Rightarrow$ 

If $R = a$ for some $a \in \Sigma$, $L(R) = \{a\}$

$\Rightarrow$ 

So if $k = 0$, $L(R)$ is regular.

Inductive step: Assume that for all $i < k$ if R has i operators $\cup, *, \circ$ then $L(R_1), L(R_2)$ are regular

Now consider R with k operators:

* $R = (R_1 \cup R_2)$ then $L(R) = L(R_1) \cup L(R_2)$

R_1, R_2 has $\leq k - 1$ operators so $L(R_1), L(R_2)$ are

regular. By closure under union $L(R)$ is regular.

* $R = (R_1 \cdot R_2)$. By IH, $L(R_1), L(R_2)$ are regular (they have at most $k-1$ operators)
By closure under concatenation $L(R_1) \cdot L(R_2)$ is regular. So $L(R)$ is regular.

* $R = (R_1)^*$. R_1 has $k-1$ operators.
By IH, $L(R_1)$ is regular. By closure under $*$ $L(R) = L(R_1)^*$ is regular.

That completes the proof!

$R = (\underbrace{a}_{R_1} \overset{0}{\cdot} \underbrace{(a \cdot b^* \cdot a)}_{R_2})$ $\overset{3}{\cdot}$ R has 4 operators

So regexp are another characterization of regular languages.

L is regular $\iff$ there $\exists$ a DFA D with $L(D) = L$
$\iff$ there $\exists$ a NFA N with $L(N) = L$
$\iff$ there $\exists$ a regexp R with $L(R) = L$

$L = \{w \in \{0,1\}^* \mid w$ does not contain $00\}$

$1^*(011^*)^*$

↓ every 0 is followed by a 1

any number of 1's

$L(011^*) = \{01, 011, 0111, \ldots\}$

we can never have 00

all strings end with something in $(L(011^*))^*$

$1^*(011^*)^* \cup 1^*(011^*)^*0$

all strings that do not contain 00 and the last symbol is 1

all strings that do not contain 00 and end in 0

$L = \{w \mid w$ doesn't have length divisible by 3 $\}$

$((0 \cup 1)(0 \cup 1))^*$ all strings of even length

=> idea: if length is not 0 mod 3, it's

$0+1 \bmod 3$ or $0+2 \bmod 3$.

$((0 \cup 1)(0 \cup 1)(0 \cup 1))^*.(0 \cup 1)$ ← length $1 \bmod 3$

$\cup$

$((0 \cup 1)(0 \cup 1)(0 \cup 1))^* (0 \cup 1)(0 \cup 1)$

↑ length $2 \bmod 3$

$(aa)^* \stackrel{?}{=} aa^*$ not equal

$(aa)^*$ we have ε in $L((aa)^*)$

aa^* ↑ strings must contain a

$(0 \cup 1)^* = (0^* \cup 1^*)^*$

$(0 \cup 1)^*$ is all strings over $\{0, 1\}$

so $L((0^* \cup 1^*)^*) \subseteq L((0 \cup 1)^*)$

$L((0^* \cup 1^*)^*) = (L(0^* \cup 1^*))^*$

$L(0^* \cup 1^*) = \{\varepsilon, 0, 1, 00, 11, 000, 111, \ldots\}$

so $(L(0^* \cup 1^*))^* = \{w \mid w = x_1, \cdots, x_k, k \geq 0$

and each $x_i \in \{\varepsilon, 0, 1, 00, \ldots\}$

$= \{0,1\}^*$

so $L((0^* \cup 1^*)^*)$ is all strings over $\{0,1\}$

$L(0^* \cup 1^*) = L(0^*) \cup L(1^*) = \{\varepsilon, 0, 00, 000, \ldots\}$

$\cup \{\varepsilon, 1, 11, 111, \ldots\}$

Announcements: This is the end of the material for Quizz 2, Hw2

DO NOT USE NFAs on Quiz 1

HW1 solutions out soon