CPSC 421/501 Oct 19, 2021

Today: C_k, regex, Myhill-Nerode instead of §1.3 (the useful half)

Q: How many natural languages (e.g. English, French, Twe, Farsi, ...) are first written, then spoken?

[via Prof. Rose-Marie Déchaine, UBC Linguistics, at TAG ISW, 1994]

Answer as per class consesus!

- Most languages first spoken, then written

- Maybe some "created"

- Maybe simultaneously, in some sense

Me (1994): Esperanto ???
(IAL = auxlang)

Q: How many formal languages ($e.g.$ $\{a^3, a^4\}^*$, DIV-BY-3, C_k) does one first describe by a DFA, then by a regex?

regex = regular expression,

e.g.,

$$C_5 = \{ w \in \{a,b\}^* \mid \text{5th last symbol is an "a"} \}$$

$$= \{a,b\}^* \circ \{a\} \circ \{a,b\}^4$$

regex $\Sigma^* a \Sigma^4$

Answer:

In [Sip], §1.3:

regex (regular expression):

$\emptyset$ empty set

ε empty string

and symbol/letter from Σ = some fixed alphabet

then operations:

$\circ$, $*$, $\cup$

where is R_1, R_2 are reg exp:

$$(R_1)\circ(R_2), \quad (R_1)^*, \quad (R_1)\cup(R_2)$$

(Note [Sip], 1.3 we do not allow negation $\neg$) !!!

C_5 as $\underbrace{\Sigma^*}_{\text{any word}} \; a \; \underbrace{\Sigma^4}_{\text{any word of length 4}}$

numerous notation conventions

really, in §1.3 [Sip] refers to

$$\Big(\big(\underbrace{(a \cup b)}_{\Sigma}{}^* \circ a \big) \circ (a \cup b) \circ (a \cup b) \circ (a \cup b) \circ (a \cup b) \Big)$$

but...

$(L_1 \circ L_2) \circ L_3$

$= L_1 \circ (L_2 \circ L_3)$

abbreviate $L_1 \circ L_2 \circ L_3$ or $L_1 L_2 L_3$

true

~~Thm: Any regular language, L is described by some regular expression,~~

Thm: Any regular expressions desribes a regular language

true

This year

this year we don't cover

Sipser 1.3: floating point

Say $\mathbb{N} = \{1, 2, 3, 4, \ldots, 9, 10, 11, \ldots\}$

e.g., $[0..9]^+$

$(0 \cup 1 \cup \dots \cup 9)\circ(0 \cup 1 \cup \dots \cup 9)^*$

$\Sigma = \{0, \dots, 9\}$

$\Sigma^* = \{0, 1, \dots, 9, 00, 01, \dots\}$

$$\left.\begin{array}{l} (\cancel{0 \cup}\ 1 \cup \dots \cup 9)\,\Sigma^* \\ (\Sigma \setminus 0)\,\Sigma^* \end{array}\right\} \{1, 2, 3, \dots 9, 10, 11, \dots\}$$

don't have this

DIV-BY-3 $= \{3, 6, 9, 12, 15, 18, 21, \dots\}$

1 1 1, 1 1 1 1 div by 3

1 1, 1 1 1 is not

1, 1 1 1 is not

1 1 1 is

base 10

(not base 2)

A REGULAR EXPRESSION FOR DIV-BY-3

JOEL FRIEDMAN

Disclaimer: The material may sketchy and/or contain errors, which I will elaborate upon and/or correct in class. For those not in CPSC 421/501: use this material at your own risk. . .

In class we derived a regular expression for DIV-BY-3, defined as

$$L = \{0, 3, 6, 9, 12, 15, \dots\}$$

(which is the language over $\Sigma = \{0, 1, \dots, 9\}$, and we do not allow leading zeros in elements of L and we do not consider the empty string to be part of L.

In class of October 2, 2019 (see class notes), began with a five state NFA, added an ending state, and then eliminated the intermediate states. The expression we got was

$$0 \cup T_0 T_1^*,$$

where

$$T_0 = S_2 \cup S_3 S_4^* S_5, \quad T_1 = S_0 \cup S_1 S_4^* S_5,$$

where

$$S_0 = R_0 \cup R_2 R_0^* R_1, \quad S_1 = S_3 = R_1 \cup R_2 R_0^* R_2, \quad S_2 = R_0' \cup R_2 R_0^* R_1,$$

$$S_4 = R_0 \cup R_1 R_0^* R_2, \quad S_5 = R_2 \cup R_1 R_0^* R_1,$$

where

$$R_0' = 3 \cup 6 \cup 9, \quad R_0 = 0 \cup 3 \cup 6 \cup 9, \quad R_1 = 1 \cup 4 \cup 7, \quad R_2 = 2 \cup 5 \cup 8.$$

In the above expressions we have omitted some necessary parentheses.

Repeated substitution yields the following regualar expression (after adding needed parentheses):

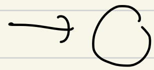
$$0 \cup \Big(\big((3\cup 6\cup 9) \cup (2\cup 5\cup 8)(0\cup 3\cup 6\cup 9)^*(1\cup 4\cup 7) \big) \cup \big((0\cup 3\cup 6\cup 9) \cup (2\cup 5\cup 8)(0\cup 3\cup 6\cup 9)^*(2\cup 5\cup 8) \big) \big((0\cup 3\cup 6\cup 9) \cup (1\cup 4\cup 7)(0\cup 3\cup 6\cup 9)^*(2\cup 5\cup 8) \big)^* \big((2\cup 5\cup 8) \cup (1\cup 4\cup 7)(0\cup 3\cup 6\cup 9)^*(1\cup 4\cup 7) \big) \Big) \Big(\big((0\cup 3\cup 6\cup 9) \cup (2\cup 5\cup 8)(0\cup 3\cup 6\cup 9)^*(1\cup 4\cup 7) \big) \cup \big((0\cup 3\cup 6\cup 9) \cup (2\cup 5\cup 8)(0\cup 3\cup 6\cup 9)^*(2\cup 5\cup 8) \big) \big((0\cup 3\cup 6\cup 9) \cup (1\cup 4\cup 7)(0\cup 3\cup 6\cup 9)^*(2\cup 5\cup 8) \big)^* \big((2\cup 5\cup 8) \cup (1\cup 4\cup 7)(0\cup 3\cup 6\cup 9)^*(1\cup 4\cup 7) \big) \Big)^*$$

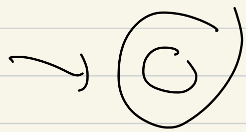
Research supported in part by an NSERC grant.

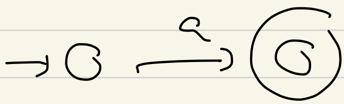
This is theoretical

Thm: Any regular expression describes a regular language:

Pf: NFA

$\emptyset \longrightarrow$ 

$\varepsilon \longrightarrow$ 

$a \longrightarrow$ 

$R_1 \circ R_2$ §1.2

R_1^* §1.2

$R_1 \cup R_2$ §1.2

i.e.

R_1, R_2 build NFA's

M_1, M_2, recognize L_1, L_2

then use §1.2 Thm

L_1^* NFA built from

M_1 NFA

similarly

$L_1 \cup L_2$ NFA from M_1, M_2

$L_1 \circ L_2$ " " M_1, M_2

break 10:13 — 10:18

Thm: If L is regular, then so is

$$L^{rev} = \{ \sigma_n \sigma_{n-1} \cdots \sigma_1 \mid \sigma_1, \ldots, \sigma_n \in \Sigma \text{ and } \sigma_1 \sigma_2 \cdots \sigma_n \in L \}$$

"reverse language"

E.g. $\Sigma = \{a, b\}$

$$C_3 = \Sigma^* a \Sigma^2$$

$$C_3^{rev} = \Sigma^2 a \Sigma^*$$

NFA

$10+1$ states

e.g.

$$C_{10} = \Sigma^* a \Sigma^9$$

how many states

$2^{10} = 1024$

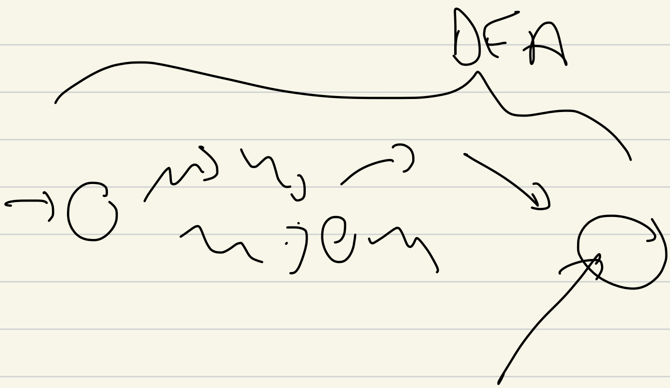
$$C_{10}^{rev} = \Sigma^9 a \Sigma^*$$

how many states in a min DFA

$10+1$

Myhill-Nerode Thm:

Idea: Say you have DFA for L, say $w_1, w_2 \in \Sigma^*$



w_1 takes you to $q \in Q$

w_2 " " " $q \in Q$

Then if $u \in \Sigma^*$

$w_1 u$ is accepted, so is $w_2 u$

$w_1 u$ is rejected, " " $w_2 u$

For any L (regular or not) over some alphabet, Σ, any $w \in \Sigma^*$

Accepting futures of w, accepting means w.r.t. L

$$\boxed{\mathrm{AccFut}_L(w) = \{u \in \Sigma^* \mid wu \in L\}}$$

Then:

① If w_1, w_2 are brought to the same state of a DFA recognizes L, then $\mathrm{AccFut}_L(w_1) = \mathrm{AccFut}(w_2)$

② If the number of different sets $Accfut_L(w)$ as w varies over $w \in \Sigma^*$ is a finite number, m, then there is a DFA recognizing L that has m states.

But if this number is not finite, then L is not regular.

Example:

C_3 ...

Let's start $\Sigma = \{a\}$

Let

$\Sigma = \{a\}, \quad L = \{a^3, a^6, a^9, \ldots\}$

$= \{a^n \mid n \text{ is } \geq 3$
n is not 0,
and n is
divisible by 3$\}$

$L' = \{a^0, a^3, a^6, a^9, \ldots\}$

$$L = \{a^3, a^6, a^9, \ldots\}$$

$$\mathrm{AccFut}_L(\varepsilon) = L = \{a^3, a^6, a^9, \ldots\}$$

$$\mathrm{AccFut}_L(a) =$$

$$\{au \mid u \in \{a\}^* \text{ s.t. } au \in L\}$$

$$= \{a^2, a^5, a^8, \ldots\}$$

$$\mathrm{AccFut}_L(aa) = \{a^1, a^4, a^7, \ldots\}$$

Summary $L = \{a^3, a^6, a^9, \ldots\}$

$$\boxed{\begin{aligned}
\text{AccFut}_L(\varepsilon) &= \{a^3, a^6, a^9, a^{12}, \ldots\} \\
\text{AccFut}_L(a) &= \{a^2, a^5, a^8, \ldots\} \\
\text{AccFut}_L(a^2) &= \{a^1, a^4, a^7, \ldots\} \\
\text{AccFut}_L(a^3) &= \{a^0 = \varepsilon, a^3, a^6, \ldots\}
\end{aligned}}$$

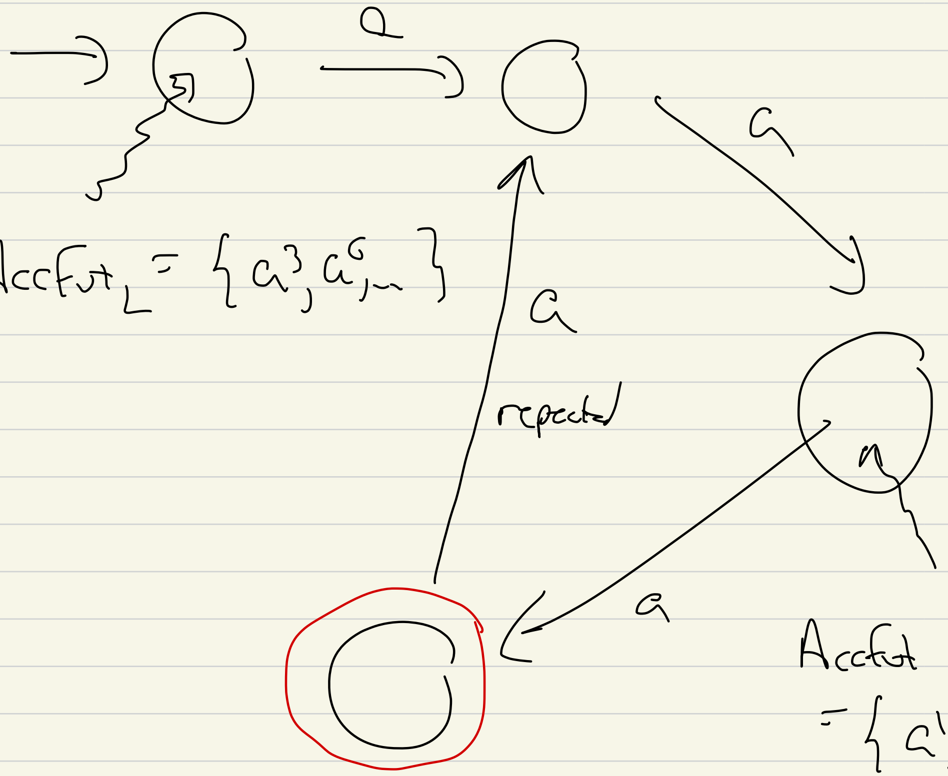
N.B. $\text{AccFut}_L(\varepsilon) \neq \text{AccFut}_L(a^3)$

Beyond This

$\text{AccFut}_L(a^4) = \text{AccFut}(a) = \text{AccFut}_L(a^7), \ldots$

etc.

$\mathrm{AccFut}_L = \{a^2, a^5, a^8, \ldots\}$



$\mathrm{AccFut}_L = \{a^3, a^6, \ldots\}$

AccFut $= \{a^1, a^4, \ldots\}$

AccFut

$\{\varepsilon, a^3, a^6, a^9, \ldots\}$

Accepting States = F