MATH 2301 — 23 Sep 2022

* Regexes & language of a regex

r a regex, then $L(r)$ = set of all words that match r.

** Shorthand notation

In a regex, we will start using the additional symbol Σ to denote any single letter of Σ.

E.g. If $\Sigma = \{0, 1\}$

- $0 \Sigma 1 = 0(0|1)1$

i.e. $\Sigma = (0|1)$

- $0 \Sigma^* 1 = 0(0|1)^* 1$

E.g. If $\Sigma = \{a, b, c\}$ then the symbol Σ in a regex is shortcut for $(a|b|c)$.

---

* Today: The language of a regex

Table of regex constructors vs corresponding language

(1) $r = \emptyset$, then $L(r) = \emptyset$

(2) $r = \varepsilon$, then $L(r) = \{\varepsilon\}$

(3) $r = a$ for $a \in \Sigma$, then $L(r) = \{a\}$

(4) $r = r_1 r_2$, then $L(r) = L(r_1) \circ L(r_2)$

(5) $r = r_1 | r_2$ then $L(r) = L(r_1) \cup L(r_2)$

(6) $r = (r_1)^*$ then $L(r) = L(r_1)^*$

** Rmk: The table tells us how to go from regexes to languages.

But we don't know how to go backwards!

Q1: Given a language $L \subseteq \Sigma^*$, is there a regex r such that $L(r) = L$?

Q2: If yes to Q1, then is there only one such regex?

We haven't yet answered Q1 and Q2.

In fact, the answer to Q2 is NO.

E.g. Let $L = L(r)$, then
$$L = L(r|\emptyset) = L(r) \cup L(\emptyset) = L \cup \emptyset = L.$$

$\Rightarrow$ there usually are multiple regexes that have the same language.

** Examples

(1) ~~E~~ $r = (\Sigma\Sigma 0)^* = ((0|1)(0|1)0)^*$

$L(r)$ contains; e.g: ε, 000, 110, 010, 000000, 010110, 000110, ...

= Words in which every third letter is a zero, and the length of the word is a multiple of 3

(~~auto~~ description automatically includes ε.)

(2) $r = (0\Sigma^*0 \mid 1\Sigma^*1 \mid 0 \mid 1)$.

$L(r)$ = Non-empty strings that start & end with the same letter.

(3) $L = \{ w \in \Sigma^* \mid w$ contains exactly $2k$ "1"s, for some $k \geq 0 \}$.

Attempt #1 : $(11)^*$, $(101)^*$ ×

Attempt #2 : $(11)^*\Sigma^* \mid \Sigma^*(11)^*$ ×

$(11)^* \mid (11)^*0^* \mid 0^*(11)^* \mid (10^*1)^* \mid 0^*$

[Does not cover, e.g. 0101]

Attempt #3 : $(0^*10^*10^*)^* \mid 0^*$ ✓ [Check!]

Attempt #4 : $0^*(10^*1)^*0^*$ ×

$(000\cdots0)\,(\underbrace{1}0\underbrace{1\,1}00\underbrace{1\,1}\underbrace{1\,1}0000\,1)\,000$

(4) $L = \Sigma^*$

$r = \Sigma^* = (0|1)^*$

$r_1 = \Sigma^* | 0$, $r_2 = \Sigma^* | 0 | 110$ all recognise the same language.

** Non-example

$L = \{ 0^n 1^n \mid n \geq 0 \}$

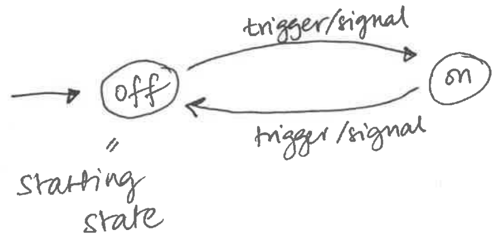
L contains ε, 01, 0011, 000111, ...

Claim: There is no regular expression r such that $L(r) = L$. !

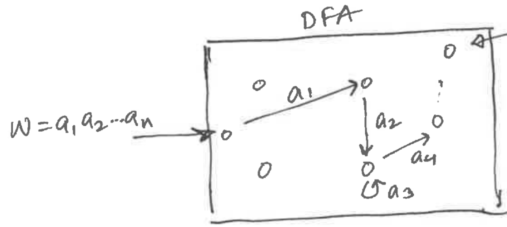
(Prove this later...)

* Deterministic finite automata (DFAs)

Informal example: sensor tap



DFAs are machines with a finite number of "states", and "transition arrows" between states, labelled by possible inputs.



final state will either be "accepting" or "rejecting"