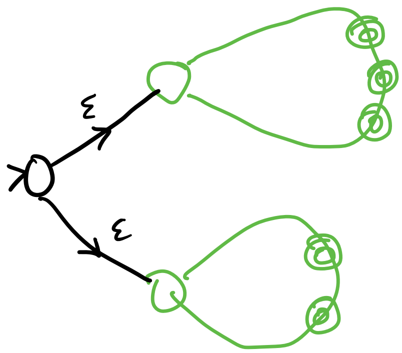
The class RL is closed under Union, concat, $*$

Jan 16 2025

Theorem: RL is closed under $\cup$

Proof:



$\leftarrow M_1$ FA for L_1

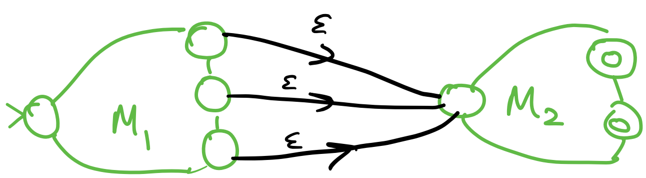
$\leftarrow M_2$ FA for L_2

∎

Theorem: RL is closed under $\circ$

Proof:

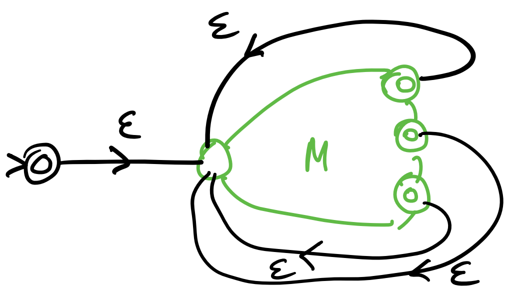
$L_1 = \{dog, cat\}$ $L_2 = \{bowl, fur\}$ $L_1 \circ L_2 = \{dogbowl, catbowl, dogfur, catfur\}$



Want: FA for $L(M_1) \circ L(M_2)$ ∎

Theorem: The class RL is closed under $*$.

Proof:



want: FA that accept $(L(M))^*$

Aside: $(L(\emptyset))^* = \{\varepsilon\}$.

Section 1.3 Regular Expressions

Eg of r.e.'s over $\Sigma = \{0,1\}$ $\cup$ $+$ $\equiv$ "or" "union"

$(0 \cup 1)0^*$ $1(1 \cup 10)^*$

$(0+1)0^*$ $1(1+10)^*$ ← we will use '+' for '$\cup$'

$(0+1)^*1$

awk, grep, Perl, REGEX, Excel – all use some form.

Syntax of r.e's over Σ.

Defn: R is a r.e. over Σ if R is one of:

1. σ, $\sigma \in \Sigma$
2. ε
3. $\emptyset$

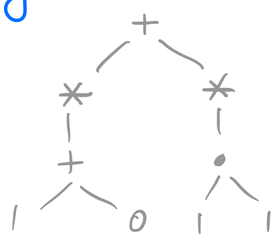
these are the base cases

4. $(R_1 + R_2)$ where R_1 and R_2 are r.e.s
5. $(R_1 \cdot R_2)$ " " " " " "
6. $(R_1{}^*)$ where R_1 is an r.e.

Nothing else is a r.e.

Note: This defn is recursive.

eg: $\Sigma = \{0,1\}$ $\to$ $\{$ 1, $(1+0)$, $((1\cdot1)^*)$, $(((1+0)^*) + ((1\cdot1)^*))$ $\ldots\}$



Technically, for every $+, *, \circ$ $\exists$ a pair of parens.

Semantics of a r.e:

A re denotes a language $L(r)$ where

1. $L(\sigma)$ $\sigma \in \Sigma$ is $\{\sigma\}$
2. $L(\varepsilon)$ is $\{\varepsilon\}$
3. $L(\phi)$ is $\{\}$
4. $L((R_1 + R_2))$ is $L(R_1) \cup L(R_2)$
5. $L((R_1 \cdot R_2))$ is $L(R_1) \cdot L(R_2)$
6. $L((R_1{}^*))$ is $(L(R_1))^*$

Conventions:

- replace $\cdot$ with juxtaposition
- precedence of operators will allow us to drop (,). except when we want to enforce an eval order.

1. $*$ 2. $\cdot$ 3. $+$

Hence $a + ba$ $<$ $(a+b)a$ / $a+(ba)$?

ab^* ← $(ab)^*$ or $a(b^*)$?

Also Σ is allowed as a r.e.

$\Sigma = \{a,b,c\}$ then $\Sigma a = \{aa, ba, ca\}$

Σ^* = all strings over Σ.

Let $\Sigma = \{a, b\}$

"ends in a" $\Sigma^* a$ $\underline{(a+b)^*}a$

"∀ b is immediately followed by a" $(a + ba)^*$

"contains aab as substring" $\Sigma^* aab \Sigma^*$

"does not contain aab as substring"

Aside: Recursive Definitions.

Def$\underline{n}$ A string w over Σ ("$w \in \Sigma^*$") is

1. ε
2. $w'\sigma$ where $\sigma \in \Sigma$, $w' \in \Sigma^*$.

Nothing else is a string over Σ.

Def$\underline{n}$: for $w \in \Sigma^*$ $|w|$ ("length of w") is given by:

1. $|\varepsilon| = 0$
2. $|\underline{\sigma w'}| = 1 + |w'|$

Other notational conveniences and notes

- Complement $\bar{L} = \Sigma^* \setminus L$
- $\{\varepsilon\}$ is a language; $|\{\varepsilon\}| = 1$ whereas $|\emptyset| = 0$

Ex. Give a r.e. for $\overline{L(\varepsilon)}$ $\qquad \Sigma \cdot \Sigma^* = \Sigma^+$

- α^+ is a notation we will use in r.e.'s to mean "at least one α (or string generated from α), concatenated together"

$$L(\alpha^+) = \begin{cases} L(\alpha^*) & \text{if } \varepsilon \in L(\alpha) \\ L(\alpha^*) & \text{otherwise} \end{cases}$$

- precedence of $^+$ is same as *

$abb^* \approx ab^+$ $\qquad\qquad$ ab^*

"a followed by at least on b" $\qquad$ "a followed by 0 or more b's"

r.e.'s

Informal: "a number of a's followed by at least one b"

Formal description: $L_1 = \{w \in \{a,b\}^* \mid$ w consists of a (possibly empty) block of a's followed by a non-empty block of b's $\}$

an r.e. for L_1: a^*b^+

$L_2 = \{w \in \{0,1\}^*$: all odd positions in w are $0\}$

$(10)^+$ 0101   0001   . . .

1111 $((0+1)+0)^*$ $(0(0+1))^*(\varepsilon+0)$

$L_3 = \{w \in \{0,1\}^* : \#_0(w)$ is even $\}$

$L_4 = \{ w \in \{a, b, c\}^* : \text{at least one of } a, b, c \text{ is not in } w \}$