Lecture 4 - Sep. 20

Lexical Analysis

Strings, Languages
Regular Expressions

Announcements

- Assignment 1 Released
  + Required slides already made available
  + In-class discussion will catch up this or next week
- Programming Test date semi-confirmed:
  + 2:00pm to 3:20pm on Saturday, October 29
  + Venue to be confirmed (LAS)
- Quiz 1 next Tuesday

Is there any reason I need to wait to go through the ANTLR4 tutorial series on YouTube over reading week? Will I need the lecture right before to understand it?

- RE
- CFG
- OOP and Composite & visitor design patterns

Formulating Strings

✓

Set of Strings of Length k

Σ^k alphabet

L^k language

$\Sigma^0 = \emptyset$ $\{\}$ ε $\{\varepsilon\}$

$|\Sigma^k| = ?$

$|\Sigma|^k$

$$\Sigma^k = \{ w \mid |w| = k \wedge w \in \Sigma^* \}$$

$\boxed{\boxed{\{0,1\}}^2}$ alphabet

all strings from $\{0,1\}$ with length 2

Set of Nonempty Strings

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots$$

$$= \bigcup_{k>0} \Sigma^k$$

Set of Strings of All Possible Lengths

alphabet symbol

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

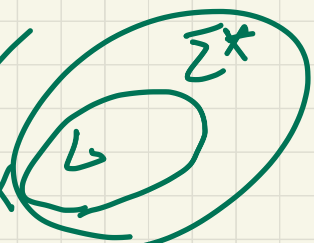
$$\Sigma^1 = \{a, b\}$$

string of length 1

$$L \subseteq \Sigma^*$$

① $w \in L \Rightarrow w \in \Sigma^*$ ✓

② $w \in \Sigma^* \Rightarrow w \in L$ ✗



$$\{xy \mid (x = 0 \land y = 1) \land$$

$|x|$

$$\{w_1 w_2 \mid w_1 \in \{0\}^* \land w_2 \in \{1\}^* \land |w_1| = |w_2|\}$$

union

$$\boxed{01^* + 10^*}$$

denotes some language (set of strings)

0^+

$$\{0x \mid x \in \{1\}^*\} \cup \{1x \mid x \in \{0\}^*\}$$

$$\{yx \mid (x \in \{1\}^* \wedge y = 0) \vee (x \in \{0\}^* \wedge y = 1)\}$$

$$\Sigma = \{0, 1\}$$

simplest "non-empty" RE: 0

1

Σ^k all strings with length k

L^k k concatenations of strings chosen from L.

Regular Language Operations

$L = \{ab, bc, ca\}$
$M = \{ba, cb\}$

$|L^i| = |L|^i$

1. *Union*

$|L \cup M| = \{w \mid w \in L \lor w \in M\}$

$\{ab, bc, ca, ba, cb\}$

2. *Concatenation*

$|LM| = \{xy \mid x \in L \land y \in M\}$

$\{wv \mid w \in L \land v \in M\}$

$\{abba, abcb, bcba, bccb, caba, cacb\}$

3. *Kleene Closure* (or *Kleene Star*)

$|L^*| =$ [redacted]

$L^0 = \{\varepsilon\}$

$L^1 = \{x \mid x \in L\} = L$

$L^2 = \{xy \mid x \in L \land y \in L\}$

$\vdots$

Cardinalities?

$\underline{L} = \{0\}^*$

0 concatenations

$L^* = \underline{\underline{L}^0} \cup \underline{\underline{L}}^1 \cup \underline{\underline{L}}^2 \cup \cdots$

$= \{\varepsilon\} \cup \{x \mid x \in \{0\}^*\}$

$\cup \{xy \mid x \in \{0\}^* \wedge y \in \{0\}^*\}$

$\cup$

$\vdots$

Constructions of REs

Recursive Case: Given that E and F are regular expressions:

- The union $E + F$ is a regular expression.

  $$L(\ E + F\ ) =$$ [redacted]

  (perl, J.S.)

  $E \cup F$ ✗

  $L(E) \cup L(F)$ ✓

- The concatenation EF is a regular expression.

  $$L(\ EF\ ) =$$ [redacted]

  $\boxed{L(E)L(F)}$

  $L(E \cup F)$ ✗

  language concat-

- Kleene closure of E is a regular expression.

  $$L(\ E^*\ ) = (L(E))^*$$

- A parenthesized E is a regular expression.

  $$L(\ (E)\ ) = L(E)$$

given a written RE (e.g. E), $L(E)$ denotes a language

Base Case:

- Constants ϵ and $\varnothing$ are regular expressions.

  RE

  $$\boxed{L(\epsilon)} = \{\epsilon\}$$
  $$L(\varnothing) = \varnothing$$

- An input symbol $a \in \Sigma$ is a regular expression.

  RE

  $$L(a) = \{a\}$$