COMS 3261 – Computer Science Theory

Lecture 3, Part 2: Regular Expressions

Idea: We can use the regular operations $(\cup, \circ, *)$ to build up expressions describing languages.

Example: $(0 \cup 1)0^*$ $(\Sigma = \{0, 1\})$

- $(0 \cup 1)$: the language $\{0\}$ $\{1\}$
- 0^*: zero or more concatenated 0's.
- (hidden concatenation)
- $(0 \cup 1)$: $\{0, 1\}$

Read this: 'either 0 or 1, followed by some number of zeroes.

This language contains: 00, 10, 0, 1, 10000, 000...

not: 101

Def: We say that a regular expression evaluates to the language of strings it describes.

$$(0 \cup 1)0^* = \{w \mid w \text{ is a string starting with 0 or 1, followed by some number of 0's}\}.$$

Example: $(0 \cup 1)^*$?

$= \{$language of all binary strings$\}$

Def (Regular Expressions, formally.)

(Idea: We'll define inductively – because regular expressions can be long and complicated, we'll recursively describe how to build one.)

We say that R is a regular expression if:

- $R = a$, for some $a \in \Sigma$. $(\{a\})$
- $R = \varepsilon$ $(\{\varepsilon\})$

- $R = \emptyset$ ($\{\}$)
- $R = R_1 \cup R_2$, where R_1 and R_2 are regular expressions.
- $R = R_1 \circ R_2$
- $R = R_1^*$.

$R: (\emptyset \cup 1)\emptyset^*$

$R_1: (\emptyset \cup 1) \circ R_2: \emptyset^*$

$R_4: \emptyset$ $R_5: 1$ | $(R_3: \emptyset)^*$

Some shorthand for regular expressions:

- Σ := 'any symbol in the alphabet Σ'

(Example: if $\Sigma = \{0, 1\}$, then $\Sigma := (0 \cup 1)$)

- R^+ := 'one or more copies of R, concatenated'

$R^+ = RR^*$

- For $k \in \mathbb{N}$, we write R^k := 'k copies of R, concatenated together.

(Example: $(01)^3 = 010101$.)

Order of operations:

1) Kleene star
2) Concatenation
3) Union.

$0 \cup 10^*$

0 10000

10000

10000

When in doubt, use parentheses!

(Lots of) Examples.

$\emptyset^* 1 \emptyset^* = \{w \mid w \text{ contains a single } 1\}$

$1^*(\emptyset 1^+)^* = \{w \mid w$ consists of zero or more 1's, followed by zero or more substrings consisting of one zero and one or more 1's.$\}$

$= \{w \mid$ every 0 in w is followed by at least one $1.\}$

$(\Sigma\Sigma\Sigma)^* = \{w \mid \underline{\text{length of } w} \text{ is a multiple of } 3\}$

$|w|$

$0\Sigma^*0 \cup 1\Sigma^*1 \cup 0 \cup 1$

$= \{w \mid w$ starts and ends with a 1, or starts and ends with a $0\}$.

$(0 \cup \varepsilon)(1 \cup \varepsilon) = \{01, 1, 0, \varepsilon\}$

$1^*\emptyset := \emptyset$

$\emptyset^* := \{\varepsilon\}$

Let $\Sigma = \{-, ., 0, 1, \cdots 9\}$, $D := \{0, \cdots 9\}$

$(\varepsilon \cup -)(D^+ \cup D^+.D^+) =$

numbers like 103, -92.9, 48.666, 0.1, ...

Next: Converting regular expressions to NFAs.