Regular Expressions

In this lecture we introduce regular expressions and show how they relate to the specification of languages.

Languages Represented by Regular Expressions

Language	Definition
L(∅)	∅
L(ε)	{ε}
L(a)	{a}
L(ST)	{XY | X ∈ L(S) and Y ∈ L(T)}
L(S|T)	{X | X ∈ L(S) or X ∈ L(T)}
L(S) *	$\varepsilon \cup \{X \mid X \in L(S)\} \cup \{X_1X_2 \mid X_1 \in L(S) \text{ and } X_2 \in L(S)\} \cup \ldots$

Priority of Operators

- The Kleene star operator binds most tightly
- The concatenation operator is next
- Last, the choice operator
- Example: 1|01* is the same as 1|(0(1)*)

Example

L((a|b)c(d|e))

$= \{XY \mid X \in L(a|b) \text{ and } Y \in L(c(d|e))\}$

$= \{XY \mid X \in \{X \mid X \in L(a) \text{ or } X \in L(b)\} \text{ and } Y \in L(c(d|e))\}$

$= \{XY \mid X \in \{X \mid X \in L(a) \text{ or } X \in L(b)\} \text{ and } Y \in \{X_1Y_1 \mid X_1 \in L(c) \text{ and } Y_1 \in L(d|e)\}\}$

$= \{XY \mid X \in \{X \mid X \in L(a) \text{ or } X \in L(b)\} \text{ and } Y \in \{X_1Y_1 \mid X_1 \in L(c) \text{ and } Y_1 \in \{X_2 \mid X_2 \in L(d) \text{ or } X_2 \in L(e)\}\}$

$= \{XY \mid X \in \{X \mid X \in \{a\} \text{ or } X \in \{b\}\} \text{ and } Y \in \{X_1Y_1 \mid X_1 \in \{c\} \text{ and } Y_1 \in \{X_2 \mid X_2 \in \{d\} \text{ or } X_2 \in \{e\}\}\}$

$= \{ acd, bcd, ace, bce \}$

Laws for Choice

$$L(R \mid R) = L(R)$$
$$L(R \mid S) = L(S \mid R)$$
$$L((R \mid S) \mid T) = L(R \mid (S \mid T))$$

Laws for Sequence

$$L(\varepsilon R) = L(R) = L(R\varepsilon)$$
$$L(\emptyset R) = L(\emptyset) = L(R\emptyset)$$
$$L((RS)T) = L(R(ST))$$

Laws

$$L(R(S \mid T)) = L(RS \mid RT)$$
$$L((R \mid S)T) = L(RT \mid ST)$$
$$L(RR^*) = L(R^*R)$$
$$L(RR^* \mid \varepsilon) = L(R^*)$$
$$L((R \mid S)^*) = L((R^*S^*)^*)$$
$$L((RS)^*R) = L(R(SR)^*)$$

Example: L(0(10)*1 | (01)*) = L((01)*)

R(SR)* = (RS)*R
0(10)* = (01)*0

R*R = RR*
01*01 = 01(01)*

R* = RR* | ε
01* = 01(01)* | ε

R|R = R
01(01)*|01(01)* = 01(01)*

RR*| ε = R*
01(01)* | ε = 01*

L(0(10)*1 | (01)*)

L((01)*01 | (01)*)

L(01(01)* | (01)*)

L(01(01)* | 01(01)* | ε)

L(01(01)* | ε)

L((01)*)