ALPHABET Σ

STRINGS on Σ

LANGUAGE: set of strings on Σ

$L_\varepsilon = \{\varepsilon\}$

$L_\emptyset = \{\}$

$L_a = \{a\}$ for any $a \in \Sigma$

$L_\Sigma = \{a \mid a \in \Sigma\}$ $\quad \{"a" \mid 'a' \in \Sigma\}$

L_1 & L_2 are languages, $L_1 \cup L_2$ is the normal set-theoretic union.

$\{abc, ab\} \cup \{ca, cd\} = \{abc, ab, ca, cd\}$

$L_1 L_2$ is the <u>concatenation</u> of L_1 and L_2 $\quad \{\alpha\beta \mid \alpha \in L_1 \wedge \beta \in L_2\}$

$\{ab, ac\}\{cd, be\} = \{abcd, abbe, accd, acbe\}$

EXPONENTIATION

L^k $\underbrace{LLLL\ldots L}_{k \text{ times}}$

$\{a,b\}^3$ $\{aaa, aab, aba, abb, baa, bab, bba, bbb\}$

$\{a\}^3$ $\{aaa\}$

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

$L^1 = L$

KLEENE CLOSURE

$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \ldots\ldots$

$\{a\}^* = \{\varepsilon, a, aa, aaa, aaaa, \ldots\}$

REGULAR LANGUAGE:

Start with L_a $a \in \Sigma$

L_ε

$L_\emptyset$

apply union (alternation), concatenation, Kleene closure

→ Regular Language

Regular Expressions

R.E.	Language	
a	L_a	
$\alpha + \beta$	$L_\alpha \cup L_\beta$	$\alpha \mid \beta$
$\alpha\beta$	$L_\alpha L_\beta$	
α^k	L_α^k	
α^*	L_α^*	

— use parentheses for grouping

$a(b+c) = \{ab, ac\}$

$a(b+c)^2 = \{abb, abc, acb, acc\}$

$a(bc+c) = \{abc, ac\}$

$\{a\}(\{b\}\{c\} \cup \{c\})$

$\{a\}(\{bc\} \cup \{c\})$

$\{a\}(\{bc, c\})$

$ab^* = \{a, ab, abb, abbb, \ldots\}$

A Regular Expression α <u>matches</u> a string s if s is in the language denoted by α.

ab^* matches abb can also say s matches α.

Use regular expressions in many places.

- editors
- shells
- lexical analyzers, grammars.

Notational Extensions

$\alpha?$ $(\varepsilon + \alpha)$ "optional α"

α^+ $\alpha\alpha^*$ "one or more α"

$[abdf]$ $(a|b|d|f)$

$[ab...z]$ $(a|b|\ldots\ldots|z)$

$[^{\wedge}abc]$ every character except a, b, c.

} character classes

$[\hat{}abc]$

$\alpha \bowtie \beta$ $\{\varepsilon, \alpha, \alpha\beta\alpha, \alpha\beta\alpha\beta\alpha, \ldots.\}$ "β-seperated list of 0 or more α's"

$(\varepsilon | \alpha(\beta\alpha)^*$

(argument lists) $arg \bowtie ,$

ε, arg, "arg, arg", "arg, arg, arg"...

floating literal:

$$[0..9]^+ \cdot [0..9]^+ ((e|E)(+|-)[0..9]^+)^?$$

one or more digits . one or more digits.

3.4

109.725

2.3e+21

34.22E-7

optional part!

an 'e' or 'E', sign + or -, at least one digit.