

COMPILADORES

Prof. Me. Stéfano Borges

1. EXPRESSÕES REGULARES – Usada pelo LEX / FLEX:

[0-9] => Reconhece um dígito
[a-zA-Z] => Reconhece uma letra (comum = sem acentos)
[\ \t\n] => Reconhece um espaço em branco ou um tab ou uma nova linha
xxxxx => Reconhece a sequência de caracteres “xxxxx”
Símbolos especiais: Exemplo:
+ => 1 ou mais ocorrências [0-9]+ => Um número
* => 0 (nenhuma) ou mais ocorrências [0-9][0-9]* => Um número
? => 0 (nenhuma) ou apenas 1 ocorrência -?[0-9]+ => Um número com/sem sinal
\n => Reconhece a marca de fim de linha / nova linha
.<=> Aceita um caracter qualquer de entrada
xxx\$ => Reconhece xxx se for seguido de um fim de linha
^xxx => Reconhece xxx se este estiver imediatamente após o início de uma linha
[^x] => Reconhece qualquer caracter menos “x”
[xyz] => Reconhece um dos caracteres “xyz” indicados
[a-z] => Reconhece um caracter pertencente ao intervalo de “a-z”
x{n} => Reconhece um número exato “n” de ocorrência de “x”
x{n,} => Reconhece a ocorrência de no mínimo “n” vezes de “x”
x{n,m} => Reconhece a ocorrência de “x” entre no mínimo “n” e no máximo “m” vezes
xx|yy => Reconhece a ocorrência de “xx” ou de “yy”
(x|y) => Agrupa (sub)expressões regulares
“x” => Reconhece exatamente o caracter “x” (usado com caracteres especiais). Ex.: “+”

2. Exemplos de expressões regulares simples:

DIGITO [0-9]
LETRA [a-zA-Z]
ESPACO [\ \t\n]
INTEIRO [0-9]+
INTSIGNED -?[0-9]+
DECIMAL [0-9]*\.[0-9]+ => aceita .33 / não aceita números sem casas decimais
INTOUDEC ([0-9]+)|([0-9]*\.[0-9]+)
IOUDSIGNED -?((([0-9]+)|([0-9]*\.[0-9]+))
NOMEVAR [a-z][a-z0-9_]* => usando opção “case insensitive”...

3. Exercícios: Defina as expressões regulares capazes de reconhecer...

- 1) Nros. de Telefones no Brasil
- 2) Placas de Carros Brasileiros
- 3) ISBN de um livro
- 4) Endereços IP válidos
- 5) Prefixos de estações de rádio (e.g. 102.3 MHz)
- 6) Números romanos
- 7) Número de matrícula da Uniube
- 8) Números reais (qualquer notação, incluindo científica)
- 9) Tags HTML (padrão)
- 10) URL de páginas Web
- 11) Palavras da Língua Portuguesa
- 12) Strings de um programa em linguagem “C”