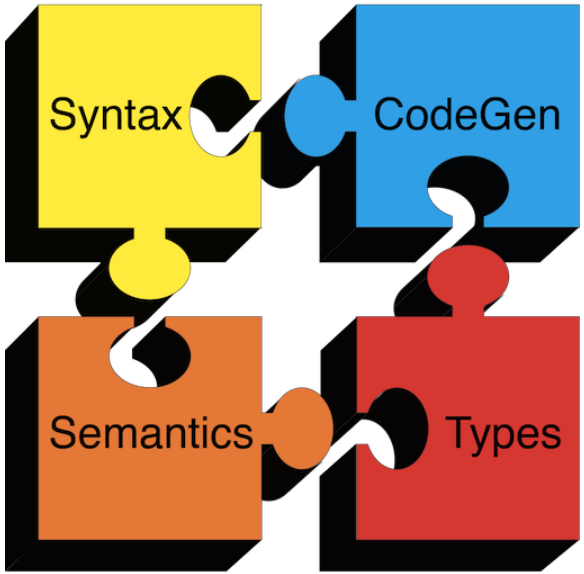

Compilers

Lexical Specifications

Lexical Specifications

Keyword: “if” or “else” or “then” or …

'i' 'f' + 'e' 'l' 's' 'e'

'if' + 'else' + 'then' + ...

Lexical Specifications

Integer: *a non-empty string of digits*

digit = '0' + '1' + '2' + '3' + '4' + '5' + '6' + '7' + '8' + '9'

digit digit*

||

digit+

AA*

||

A+

Lexical Specifications

Identifier: *strings of letters or digits, starting with a letter*

letter = [a-zA-Z]

letter (letter + digit)*

Lexical Specifications

Whitespace: *a non-empty sequence of blanks, newlines, and tabs*

$(' ' + '\n' + '\t')^+$

Lexical Specifications

letter+ '@' letter+ '.' letter+ '.' letter+

Lexical Specifications

PASCAL

digit = '0' +'1'+'2'+'3'+'4'+'5'+'6'+'7'+'8'+'9'

digits = digit$^+$

opt_fraction = ('.' digits) + ε = ('.' digits)?

opt_exponent = ('E' ('+' + '-' + ε) digits) + ε

num = digits opt_fraction opt_exponent

('E' ('+'+'-')? digits)?

Lexical Specifications

Choose the regular languages that are correct specifications of the English-language description given below:

Twelve-hour times of the form "04:13PM". Minutes should always be a two digit number, but hours may be a single digit.

- ☐ (0 + 1)?[0-9]:[0-5][0-9](AM + PM)
- ☐ ((0 + ε)[0-9] + 1[0-2]):[0-5][0-9](AM + PM)
- ☐ (0*[0-9] + 1[0-2]):[0-5][0-9](AM + PM)
- ☐ (0?[0-9] + 1(0 + 1 + 2):[0-5][0-9](A + P)M

Lexical Specifications

- Regular expressions describe many useful languages
  - email
  - phone numbers
  - file names
- Regular languages are a language specification
  - We still need an implementation
- Next time: Given a string s and a rexp R, is

$$s \in L(R)\,?$$