CSE431 – Translation of Computer Languages

- Scanner Generators

Doug Shook

Scanner Generators

- We have seen that DFAs, Regular Expressions, and RLGs are all capable of working with regular languages

- Which of the above would be the most appropriate for
  - Conceptually discussing a language?
  - Specifying a language?
  - Determining whether a string is in a language?

Scanner Generators

- Idea: we want to have a FSA to find tokens for us
  - But coding up FSAs is tedious
- Let's use regular expressions instead!
  - Convert to FSA as necessary
- JFlex is one such scanner generator (or lexer)
  - Specify tokens by regex
  - Scanner generated from regex

JFlex Specifications

- User Code
  - Allows you to specify Java code that is copied into the scanner
- Options / Declarations
  - Options allow you to customize the scanner
    - Name, scope, additional methods
  - Declarations allow you to specify macros
- Matching rules
  - Regex and action blocks

JFlex Specifications

- Regex in Jflex include the three basic operations plus many additional extensions:

?	0 or 1 occurrences
!	NOT
a{n}	repetition (a{4} = aaaa)
/	If followed by
~	match up to
^	character class negation

Transducers

- What is the output of an FSA?
  - Is it desirable to have more output than that?
  - When and why?

Transducers

- Transducers are FSAs that allow you to specify (more) output
  - Where have you seen these before?
  - What are the two types of transducers?

- Compilers often use transducers in conjunction with the symbol table
  - This is your task for homework #1

Symbol Table

- Examine the following piece of code:

```
int a = 7;
int b = 2;
if a > b
   bool c = false;
```

- What tokens will be generated by this code?
  - What do we know about the variables a, b, and c?

Symbol Table

- Symbol tables are used to store *identifiers*
  - What is the definition of an identifier?
- Information is stored with each identifier so that it can be used in later stages of compilation
  - What information might we want to track?
- More on symbol tables to come!

Programming Assignment #1

- Due two weeks (2/4)
- Next Monday's studio will help you – a lot!
- Let’s take a look at the supplied code