Using Lex

Regular Expressions (1 of 4)

- Special characters are:
 - − \ (back slash)
 - " (double quote)
 - . (period)
 - ^ (caret or up arrow)
 - \$ (dollar sign)
 - [(open bracket)
 -] (close bracket)
 - * (asterisk)

- + (plus sign)
- ? (question mark)
- { (open brace)
- } (close brace)
- | (vertical bar)
- / (slash)
- - (dash or hyphen)
- ((open parenthesis)
-) (close parenthesis)

Regular Expressions (2 of 4)

- c matches the single non-operator char c
- \c matches the character c
- "s" matches the string s
- matches any character except newline
- ^ matches beginning of line
- \$ matches end of line
- [s] matches any one character in s
- [^s] matches any one character not in s

Regular Expressions (3 of 4)

- r* matches zero or more strings matching r
- r+ matches one or more strings matching r
- r? matches zero or one strings matching r
- r{m, n} matches between m and n occurrences of r
- r₁r₂ matches r₁ followed by r₂
- r₁|r₂ matches either r₁ or r₂
- (r) matches r
- r₁/r₂ matches r₁ when followed by r₂
- {name} matches the regex defined by name

Regular Expressions (4 of 4)

Within square brackets, ranges of characters are allowed

— a-zA-Z

File Format (1 of 3)

- Extension is .lex
- <definitions>%%<rules>%%<user functions>
- Anything in the <definitions> sections that is delimited by a line with "%{" to a line with "%}" is copied directly to the output C file

File Format (2 of 3)

 Each line in the <definitions> section (other than those between "%{" and "%}") has the format:

```
<name> <regex> "%}"
```

- When used in a rule with the name enclosed within braces, the <regex> is substituted
 - This does not happen when the name within braces is quoted

File Format (3 of 3)

- The kind of token gets returned
- The value of the token gets placed in yylval

Compiling a Lex file

- lex lexer-standalone.lex
- gcc lex.yy.c –c
 - -c means to create an object file, but do not link
 - object file will have the extension ".o"
- gcc -pedantic -Wall lex.yy.o lexer.c -lfl -o lexer
 - pedantic means to ensure standard compliance
 - Wall means to issue all warnings
 - If I means to link with the flex libraries (on some systems, -II may be needed to link with lex libraries)
 - o is used to specify the name of the executable file

Special Symbols

- yytext the matched string as a null terminated string
- yyleng the length of the matched string
- yylex() the name of the generated lexer function
- yylval the value of the token matched