A Second Introduction to Lex

Review: What is Lex/Flex?

- Recap from lecture
  - Software tool
  - Takes a lexical specification (typically .l file) and generates a lexical analyzer implemented in C (lex.yy.c)

Review: Useful Lex Regular Expressions

- Should already be in your notes
  - c, \c
  - [a-z], [^A-Z]
  - r*, r+, r?, r{m,n}, (r)
  - r_1r_2, $r_1 | r_2$, r_1 / r_2
  - ^r, r$

Review: Format of a Lex file

```
header
%%
body
%%
helper functions
```

Lex File: Header Section

- Place header files here
  - Example

```
%{
  #include “helper.h”
%}
```

  - Everything between %{ and %} is copied verbatim
- Place definitions here
  - Syntax: <name> <definition>
  - Example:

```
DIGIT   [0-9]
```

  - Helps with readability

Lex File: Body Section

- Place lexical rules here
  - Syntax: <rule> <action>
  - Examples

```
[0-9]+                processInt(yytext);
{DIGIT}+“.”{DIGIT}*     { printf(“float”);
                          return FLOATTOK; }
```

  - Why are there braces around DIGIT?
  - Why are there quotes around the period?

Lex File: Helper Functions

- Place C function here
  - Example function definition

```
void processInt(char *str) {
  printf(“int value is %d\n”,atoi(str));
}
```

- Note: declare helper function in header
  - Otherwise gcc will complain that function is implicitly declared

Lex Functions and Variables

- `char * yytext`: current matched text
- `int yyleng`: the length of the matched text
- `int yylex()`: scanner function
- `char input()`: retrieve the next character in the input stream
- `int yyterminate()` : terminate scanner and return 0

Example 1

- Simple calculator scanner

Example 1: Observations

- Multiple instructions should be enclosed with braces
    - Can also enclose single instructions
- The ‘-’ character needs to be escaped
- Actions consisted of output to screen

Example 2

- How does the scanner interact with the parser?
  - Scanner needs to return values
- What is the precedence of matched patterns?
- How can the scanner ignore things like comments?
- How can the scanner count lines?
- How are unrecognized tokens handled?

Further Reading

- Flex manual
  - flex.sourgeforge.net/manual/index.html
- Dragon book
  - Aho et al., Compilers: Principles, Techniques and Tool: Pages 140-144