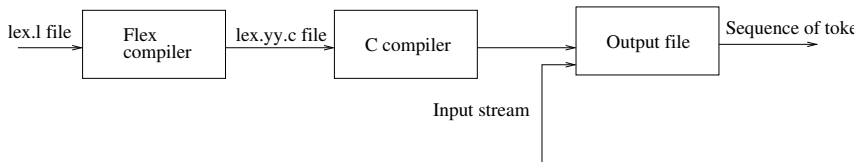


Towards automating lexical analysis

Sudakshina Dutta

Indian Institute of Technology Goa

16th February, 2022



- ▶ The token description that flex uses are regular expressions
- ▶ As the input is divided into tokens, a program often needs to find relationship among the token
- ▶ It is done by grammar rules
- ▶ Bison takes the grammar and produces parsing routines written in C

Conditional activation

- ▶ Flex provides a mechanism for conditionally activating rules

```
⟨STRING⟩[^"]* {  
/* Eats up the entire string */ }
```

- ▶ The following rule will be active when start condition is 'INITIAL' or 'STRING'

```
⟨INITIAL, STRING⟩. {  
/* Eats up the next non-newline character */ }
```

Declaring start condition

- ▶ Start conditions are declared in the definitions (first) section of the input using unindented lines beginning with either '%s' or '%x' followed by a list of names
- ▶ A start condition is activated using the BEGIN action
- ▶ Until the next BEGIN action is executed, rules with the given start condition will be active and rules with other start conditions will be inactive

%s example

%%

<example>"abcd" { // Do something }

Start condition

- ▶ 'BEGIN(0)' returns to the original state where only the rules with no start conditions are active.
 - ▶ This state can also be referred to as the start-condition 'INITIAL', so 'BEGIN(INITIAL)' is equivalent to 'BEGIN(0)'

- ▶ Whenever `yylex()` is called, it scans tokens from the global input file `y Yin` (which defaults to `stdin`)
FILE *y Yin;
- ▶ It is up to the developer to open a file for reading and store it into `y Yin` before you call `yylex`
- ▶ If the scanner reaches an end-of-file, subsequent calls are undefined unless either `y Yin` is pointed at a new input file

yywrap

- ▶ When yylex is finished, it call function yywrap()
- ▶ If yywrap() returns 1, then yylex returns 0 to its caller. It means end-of-file
- ▶ If yywrap returns 0, then yylex assumes that you have stored a different file into yyin, and it starts reading that file
- ▶ No change of start condition. It does not go back to INITIAL

References

- ▶ Lex and Yacc by John R. Levine, Tony Mason and Doug Brown (O'Reilly)
- ▶ <https://www.gnu.org/software/bison/manual/bison.pdf>