

AllenCompiler

Marissa Allen

Overview:

The AllenCompiler will read the input characters from a plain text file or a String, scan the file using the rules defined in the jflex generated Scanner class of the java code, and output a sequence of tokens for the parser. Every token will consist of the lexeme, or characters in input stream, and the token type. The scanner will distinguish which characters are keywords and which are unexpected tokens. The Scanner and its classes are stored in the scanner package.

Scanner:

The scanner program is going to scan through the input file and return a lexeme. The lexeme is going to be the string containing the actual characters that were read in that make up one particular token in the language. The token type is the individual token type descriptor of the keywords, identifiers, and symbols. The keywords and symbols are listed after the design section, anything else is an identifier.

Scanners have a state machine that does pattern matching, a diagram of the state machine can be seen below. As long as the scanner sees characters that are not spaces, newline characters, or tabs, it will keep adding them to the lexeme. If the scanner does see those, and it already has a lexeme, it's going to push any spaces, newline characters, or tabs back into the input stream and return the lexeme because

it's done. If the scanner sees a space, newline characters, or tab, and it doesn't have a lexeme yet it keeps going to find a lexeme and throws the whitespace character away.

This program contains four java files – Token, TokenType, ScannerTest, and Scanner. It also contains one jflex file called AllenScanner. The AllenScanner file contains the instructions on how each Token should be handled. The Token class is a constructor class for a token object. All the tokens will consist of the lexeme, or characters in the input stream, and the token type. The TokenType class consists of all of the different types a token can be. These types are then passed into the scanner class. There is a type for every keyword and symbol as well as two types for ID and number. ScannerTest takes an input file or a string and feeds the characters into the scanner class. As long as the scanner doesn't hit a null token, the scanner prints out any tokens that are not whitespace tokens or null tokens. The Scanner file is the java file that was created when the AllenScanner file was run. The scanner for this class uses the keywords and symbols defined for the token and prints out the tokens that are not whitespace tokens or null tokens along with any errors.

Recognizer:

The beginning of the parser is created by building a recognizer implemented as a top-down recursive descent parser. The recursive descent parser will act as a recognizer, answering the yes-no question “is the input a pascal program?”. A recognizer is a parser that doesn't perform any syntax-directed translation. The recognizer will only say whether the input string is from the language described by the

grammar or not. In all of the non-terminal methods listed in the grammar, it will do this by looking at a stream of tokens from the scanner in a particular order, interpreting the token types, and comparing the current lookahead token to the token type. If the current lookahead token matches all of the expected token types listed in the method and any other methods called; then the method will execute the rule for the non-terminal symbol in the micro pascal grammar.

Change Log:

02/15/2019	Marissa Allen	Added the Recognizer section of the compiler.
------------	---------------	-----------------------------------------------

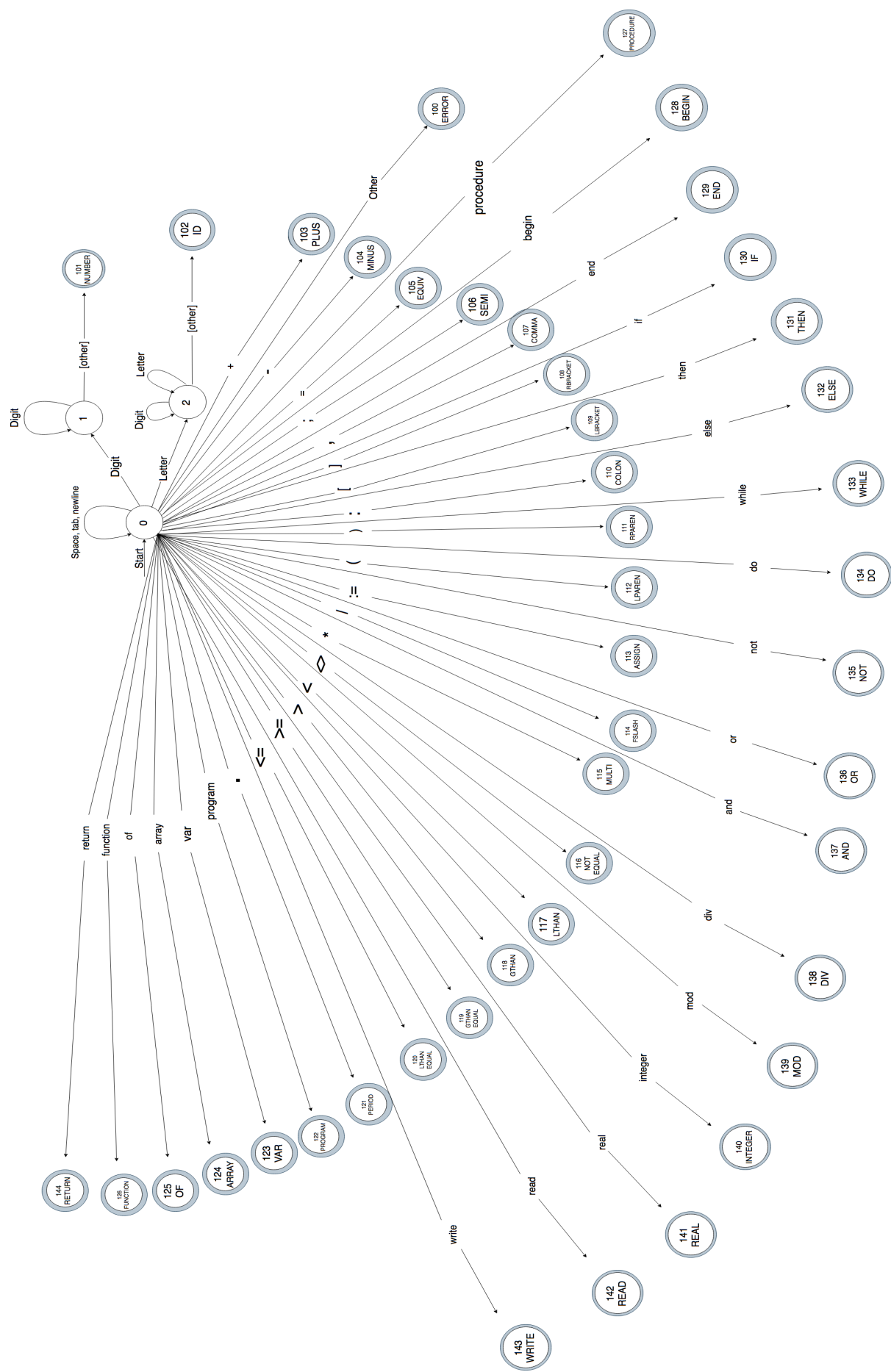
List of Keywords/Reserved Words:

program
var
array
of
function
procedure
begin
end
if
then
else
while
do
not
or
and
div
mod

integer
real
read
write
return

List of Symbols:

;
,
[
]
:
)
(
+
-
*
=
◇
<
≤
≥
>
/
:=
.



FSM Design: Designed by Marissa Allen and Cohl Dorsey

