AllenScanner

Marissa Allen

Overview:

The AllenScanner program will read the input characters from a plain text file 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. Then the scanner is going to distinguish which characters are keywords and which are user made variables.

Design:

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, TestMyScanner, and MyScanner. 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. TestMyScanner takes an input file from the command line and feeds the characters into a 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 MyScanner file is the java file that was created when the AllenScanner jflex file was run. The scanner for this class uses the key words and symbols defined for the token and prints out the tokens that are not whitespace tokens or null tokens along with any errors.

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

List of Symbols:
;

,

[

]

:

)

(

+

-

{

}

*

=

◇

<

<=

>=

>

/

:=

.

Start

Space, tab, newline

Digit

Digit

0

1

Digit

[other]

101
NUM

Letter

2

Digit

Letter

[other]

102
ID

+

103
PLUS

-

Other

104
MINUS

100
ERROR

.

<=

>=

>

<

<>

*

/

:=

{

}

(

)

:

[

]

,

;

=

105
EQUIVALENT

106
SEMI

107
COMMA

108
RIGHT_
BRACKET

109
LEFT_
BRACKET

110
COLON

111
RIGHT_
PARAN

112
LEFT_
PARAN

113
RIGHT_
CURLY

114
LEFT_
CURLY

115
ASSIGN

116
REAL_
DIVISION

117
MULTI

118
NOT_
EQUAL

119
LESS

120
GREATER

121
GREATER_
EQUAL

122
LESS_
EQUAL

123
DOT