

Compiler

Lab 2: Jflex with scanner (Lexical analysis)

1. Given the following regular definitions related to a filename description using the JFLEX program, write a scanner which recognizes a correct pathname.

a. Regular Definitions

PathFileName \rightarrow (Drive :)? (\)? (PathName \)^{*} FileName (. FileType)?

Drive \rightarrow letter

PathName \rightarrow id

FileName \rightarrow id

FileType \rightarrow id

id \rightarrow (letter | digit) (letter | digit)^{*}

letter: all the alphabet letters and the special characters except: slash, backslash, colon, star, question mark, double quotes, minus, greater, OR symbol.

digit: digits from 0 to 9

b. Example

Input file:

A:\folder\subfolder1\file.exe

E:\file2.bat

folder1\image\file4

\test\sum.c

A:readme.txt

Documents

Outputs:

Correct: A:\folder\subfolder1\file.exe

Correct: E:\file2.bat

Correct: folder1\image\file4

Correct: \test\sum.c

Correct: A:readme.txt

Correct: Documents

2. Write the regular definitions and the correspondent JFLEX program for recognizing a subset of the URL. This subset must include:

- The recognition of more than one scheme (like http:, ftp:, gopher: and similar ones).
- The domains expressed as names (e.g. www.itc.edu.kh) or as IP addresses (e.g. 192.168.4.4).
- The use of ports (domain followed by “:” and by port number).
- The use of anchors in the HTML file (file name followed by a “#” and by the anchor name).
- Sequences of escape (i.e. the character “%” followed by a pair of hexadecimal digits) in any position.

a. Example

http://www.mysite.com/file.html

ftp://10.9.9.71/file.zip

http://another.site.com/%7Euler/index.html#rif33

nntp://news.site.ch:8181/data/