

## Assignment No. 4

In this assignment you will write a Java program that implements a modified version of the Unix `fgrep` tool, which searches for a given string in files. The purpose is to practice Java I/O programming.

Your program will get three arguments:

1. A **simple fixed string to look for**. For simplicity assume that the case of the alphabetic characters is significant (e.g., 'A' is not the same as 'a').
2. A **pattern identifying the pathnames of the files to be searched**. This pattern is defined as follows:
  - The file and directory names in the pattern may include asterisks<sup>1</sup> (\*) as wild card characters, *i.e.* an asterisk will match any sequence of adjacent characters of length 0 or more. Any other character will only match itself. For instance: "a\*.txt" will match names that start with an "a" and end with ".txt", like "a.txt", "a1.txt", "a12.txt", and "a123.txt". Note that "\*a\*" will match any file and directory name that includes an "a" anywhere and "a\*\*.txt" will match what "a\*.txt" matches.
  - An asterisk wild card cannot match a directory separator (e.g. '\' on windows, '/' on linux). That is, searching for "\*.\*.txt" will not match "\foo\data.txt", but will match "\data.txt". To match "\foo\data.txt" use "\foo\*.txt", "\\*\*.\*.txt", "\f\*\*.\*", "\\*\*.\*" etc.
3. **The name of the character encoding of the files to be searched**. The following three encoding schemes will be supported: US-ASCII, UTF-16 and UTF-8.

The program will search for the given string in files recursively, *i.e.* once the pathname pattern matches a directory, the program will search *all* files under this directory and its sub-directories. For example, for the pathname pattern "D:\java\\*", the program will search all files under the "java" directory and under all its subdirectories, meaning that the following files will be searched: "D:\java\foo.txt", "D:\java\doc\data.txt", "D:\java\doc\exercises\ex1.txt" etc. If the pattern is "D:\java\a\*", the program will search the file "D:\java\abc.txt", and all files under the directory "D:\java\arch\\*", but not the file "D:\java\foo.txt", or files under "D:\java\doc\\*".

The program will search files that match the pathname pattern (as described above) for lines containing the given string. These lines will be printed to the standard output.

### Usage example:

Input: "שלום" c:\\* UTF-16

Output: שלום עולם  
 עוד יבוא שלום עלינו  
 יונת שלום

For your convenience you can find on the course website a zip file containing examples for different encoding files. Also, you can use the `java.util.regex.Pattern` class for defining regular expressions and checking that a word matches the regular expression

Carefully define the contract and implement the program. Use `javadoc` for documentation.

<sup>1</sup> Due to a problem of passing arguments containing asterisks in Eclipse, please use '#' instead '\*' in your input.