

The Glob and Regular Expressions

Note 1. In this notes packet, you will learn two new *query programming languages*: the Glob and Regular Expressions (regex). These languages are similar to CSS Selectors in that: (1) they are used to "find" information inside another filetype; (2) they are commonly used from within other languages.

1 Glob

Note 2. The glob is used for creating a "list" of files that match a particular pattern. You are responsible for the following syntax:

Pattern	Matches
*	any string (including empty)
?	exactly one character
[abc]	one character in the set
[^abc]	one character not in the set

1.1 With the Shell

Note 3. The glob is always applied to any arguments that are passed into a program.

Note 4. Anytime we write code that uses the glob, we say we are *globbing*.

Problem 5. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.md
3 $ echo 'hola mundo' > README.txt
4 $ echo 'salve munde' > README
5 $ ls RE* | wc -l
```

Problem 6. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.md
3 $ echo 'hola mundo' > README.txt
4 $ echo 'salve munde' > README
5 $ ls *.txt | wc -l
```

Problem 7. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.md
3 $ echo 'hola mundo' > README.txt
4 $ echo 'salve munde' > README
5 $ cat *.* | wc -l
```

Problem 8. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat *e | wc -l
```

Problem 9. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat README.?? | wc -l
```

Problem 10. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat README.?? | wc -l
```

Problem 11. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat README.e? | wc -l
```

Problem 12. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat README.[elf]? | wc -l
```

Problem 13. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat README.[^elf]? | wc -l
```

Problem 14. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat romance/README.[^elf]? | wc -l
```

Problem 15. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat */README.[^elf]? | wc -l
```

Problem 16. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat *a*/README.??? | wc -l
```

1.2 With Python

Note 17. Python has a built-in module `glob` for globbing. The function `glob.glob` returns a list of all paths that match the glob expression.

Note 18. The problems below are identical to problems in the previous section, except that a python file `example.py` is created and the glob calculated inside this file. DO NOT FORGET THAT THE `example.py` FILE MIGHT MATCH THE GLOB PATTERNS!

Problem 19. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.md
3 $ echo 'hola mundo' > README.txt
4 $ echo 'salve munde' > README
5 $ cat > example.py <<EOF
6 import glob
7 paths = glob.glob('*.*)
8 print(len(paths))
9 EOF
10 $ python3 example.py
```

Problem 20. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat > example.py <<EOF
10 import glob
11 paths = glob.glob('README.[elf]?')
12 print(len(paths))
13 EOF
14 $ python3 example.py
```

Problem 21. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat > example.py <<EOF
15 import glob
16 paths = glob.glob('*/*README.[^elf]?')
17 print(len(paths))
18 EOF
19 $ python3 example.py
```

Note 22. Recall that we are not required to use the `glob` module within python in order to match filenames. This happens at the shell-level, and so all programs that accept arguments get this ability.

Problem 23. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.md
3 $ echo 'hola mundo' > README.txt
4 $ echo 'salve munde' > README
5 $ cat > example.py <<EOF
6 import sys
7 print(len(sys.argv))
8 EOF
9 $ python3 example.py *.*
```

Problem 24. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ echo 'hello world' > README.en # english
3 $ echo 'hola mundo' > README.es # spanish
4 $ echo 'salve munde' > README.la # latin
5 $ echo 'hallo welt' > README.de # german
6 $ echo 'ola mundo' > README.pt # portuguese
7 $ echo 'bonjour le monde' > README.fr # french
8 $ echo 'saluton mondo' > README.epo # esperanto
9 $ cat > example.py <<EOF
10 import sys
11 print(len(sys.argv))
12 EOF
13 $ python3 example.py README.[elf]?
```

Problem 25. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat > example.py <<EOF
15 import sys
16 print(len(sys.argv))
17 EOF
18 $ python3 example.py */README.[^elf]?
```


2 Regex

Note 26. Regular expressions (regex) closely related to the glob:

1. The glob is generally used for finding files based on their *name*, and regex is used for finding files based on their *contents*.
2. Regex has a syntax similar to the glob, but the semantics is slightly different. It is easy to get them confused, so pay attention!

You are responsible for the following regex syntax:

Pattern	Matches
.	exactly one character
*	zero or more of preceding element
+	one or more of preceding element
?	zero or one of preceding element
[abc]	one character in the set
[^abc]	one character not in the set
^	start of string
\$	end of string
(a b)	alternation (a or b)

Note 27. For space reasons, I am giving you a very limited set of regex/glob patterns below. You should try creating your own problems with your own regex/glob patterns to check your understanding. This is another example of me trying to get you to learn-how-to-learn.

2.1 In the shell

Note 28. The shell program for using regex is called `grep`. It is so commonly used that programmers often say we are “grepping” for something whenever we are “searching” for something (similar to how muggles will often say they are “googling” something even if they are using another search engine like bing, baidu, or yandex).

Problem 29. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat romance/* | grep 'o' | wc -l
```

Problem 30. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat romance/* | grep '^o' | wc -l
```

Problem 31. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat romance/* | grep 'o$' | wc -l
```

Problem 32. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat romance/* | grep 'o.*d' | wc -l
```

Problem 33. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world'      > romance/README.en # english
5 $ echo 'hola mundo'      > romance/README.es # spanish
6 $ echo 'salve munde'     > romance/README.la # latin
7 $ echo 'hallo welt'      > germanic/README.de # german
8 $ echo 'hallo wereld'    > germanic/README.nl # dutch
9 $ echo 'hej verden'      > germanic/README.da # danish
10 $ echo 'hei verden'     > germanic/README.no # norwegian
11 $ echo 'ola mundo'      > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo'  > romance/README.epo # esperanto
14 $ cat romance/* | grep -E 'hello|salve|hallo' | wc -l
```

2.2 Within Python

Note 34. Python has a built-in module `re` for working with regex. The function `re.search` returns `True` if the input `str` matches the regex, otherwise returns `False`. It is customary (but not required) in python to use *raw-strings* to define regex. (Recall that raw string literals are prefixed with the letter `r` like `r'...'` and that the backslash escape characters like `\n` are not interpreted inside of rawstrings.)

Problem 35. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat > example.py <<EOF
15 import glob
16 import re
17 count = 0
18 for filepath in glob.glob('romance/*'):
19     with open(filepath) as f:
20         for line in f:
21             if re.search(r'o.*d', line):
22                 count += 1
23 print(count)
24 EOF
25 $ python example.py
```

Problem 36. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat > example.py <<EOF
15 import glob
16 import re
17 count = 0
18 for filepath in glob.glob('germanic/*'):
19     with open(filepath) as f:
20         for line in f:
21             if re.search(r'd$', line):
22                 count += 1
23 print(count)
24 EOF
25 $ python example.py
```

Problem 37. Write the output of the final command in the following shell script.

```
1 $ cd; rm -rf quiz; mkdir quiz; cd quiz
2 $ mkdir romance
3 $ mkdir germanic
4 $ echo 'hello world' > romance/README.en # english
5 $ echo 'hola mundo' > romance/README.es # spanish
6 $ echo 'salve munde' > romance/README.la # latin
7 $ echo 'hallo welt' > germanic/README.de # german
8 $ echo 'hallo wereld' > germanic/README.nl # dutch
9 $ echo 'hej verden' > germanic/README.da # danish
10 $ echo 'hei verden' > germanic/README.no # norwegian
11 $ echo 'ola mundo' > romance/README.pt # portuguese
12 $ echo 'bonjour le monde' > romance/README.fr # french
13 $ echo 'saluton mondo' > romance/README.epo # esperanto
14 $ cat > example.py <<EOF
15 import glob
16 import re
17 count = 0
18 for filepath in glob.glob('*/*'):
19     with open(filepath) as f:
20         for line in f:
21             if re.search(r'salve|hallo|hello', line):
22                 count += 1
23 print(count)
24 EOF
25 $ python example.py
```