Natural Language Processing

Lab #2
Regular Expressions in Python

An Introduction to Regex in Python:

A regular expression is simply a sequence of characters that define a pattern.

When you want to match a string to perhaps validate an email or password, or even extract some data, a regex is an indispensable tool.

While unicode characters can be used to match any international text, most patterns use normal ASCII (letters, digits, punctuation and keyboard symbols like $@%#!.)

Why should I learn regex?

Regular expressions are everywhere. Here's some of the reasons why you should learn them:

- **They do a lot with less** — You can write a few characters to do something that could have taken dozens of lines of code to implement
- **They are super fast** — Regex patterns wrote with performance in mind takes a very short time to execute. Backtracking might take some time, but even that has optimal variations that run super fast
- **They are portable** — The majority of regex syntax works the same way in a variety of programming languages

Common applications of regex are:

- Input validation (emails, usernames, passwords)
- Web scraping
- Data wrangling
- Simple parsing
-

Also, regex is used for text matching in spreadsheets, text editors, IDEs and Google Analytics.

Our first regex pattern:

Python uses raw string notations to write regular expressions – r"write-expression-here"
First, we'll import the re module.
Then write out the regex pattern.

Solution:

```
import re
pattern = re.compile(r"")
```

<u>Questions for lab:</u>

1) Write a Python program to check that a string contains only a certain set of characters (in this case a-z, A-Z and 0-9).

2) Write a Python program that matches a string that has an a followed by zero or more b's

3) Write a Python program that matches a string that has an a followed by one or more b's

4) Write a Python program that matches a string that has an a followed by zero or one 'b'

5) Write a Python program that matches a string that has an a followed by three 'b'

6) Write a Python program that matches a string that has an a followed by two to three 'b'.

7) Write a Python program to find sequences of lowercase letters joined with a underscore.

8) Write a Python program to find the sequences of one upper case letter followed by lower case letters.

9) Write a Python program that matches a string that has an 'a' followed by anything, ending in 'b'.

10) Write a Python program that matches a word at the beginning of a string.

11) Write a Python program that matches a word at the end of string, with optional punctuation.

12) Write a Python program that matches a word containing 'z'

13) Write a Python program that matches a word containing 'z', not at the start or end of the word.

14) Write a Python program to match a string that contains only upper and lowercase letters, numbers, and underscores.

15) Write a Python program where a string will start with a specific number.

16) Write a Python program to remove leading zeros from an IP address

17) Write a Python program to check for a number at the end of a string.

18) Write a Python program to search the numbers (0-9) of length between 1 to 3 in a given string.

19) Write a Python program to search some literals strings in a string. Go to the editor Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox', 'dog', 'horse'

20) Write a Python program to search a literals string in a string and also find the location within the original string where the pattern occurs

Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox'

21) Write a Python program to find the substrings within a string.

Sample text :

'Python exercises, PHP exercises, C# exercises'

Pattern :

'exercises'

Note: There are two instances of exercises in the input string.

22) Write a Python program to find the occurrence and position of the substrings within a string.

23) Write a Python program to replace whitespaces with an underscore and vice versa.

24) Write a Python program to extract year, month and date from a an url.

25) Write a Python program to convert a date of yyyy-mm-dd format to dd-mm-yyyy format.

26) Write a Python program to match if two words from a list of words starting with letter 'P'.

27) Write a Python program to separate and print the numbers of a given string.

28) Write a Python program to find all words starting with 'a' or 'e' in a given string.

29) Write a Python program to separate and print the numbers and their position of a given string.

30) Write a Python program to abbreviate 'Road' as 'Rd.' in a given string.