

Lab Sheet 2

Task 1.

- Launch Python in Gnome terminal,
- Use the command 'python3',
- We want use RegEx under Python,
- We need to the RegEx packages which is 're' '
- 'import re'
- Open the text file of our corpus,
- file_path = '/home/ndoumi/teaching/iapa/practical/Samsung_confidentiality_statement.txt'
- f = open(file_path, 'r')
- Create a text object and assign to it the content of your corpus,
- text = f.read()
- Set a pattern
- pattern = 'Samsung'
- Compile the pattern
- pattern = re.compile(pattern)
- Use match method: find only the occurrences at the beginning of text
- match_result = re.match(pattern, text)
- print(match_result.group())
- Use search method: find the remaining occurrences of text
- search_result = re.search(pattern, text)
- print(search_result)
- Use findall method: find all occurrences into the text and returns them as list
- findall_result = re.findall(pattern, text)
- print(findall_result)

Task 2.