Regular Expressions:

.

- A Regex, or Regular Expression, is a sequence of characters that forms a search pattern.
- Regex can be used to check if a string contains the specified search pattern
- **Regex Module :**Python has a built-in package called ’re’, which can be used to work with Regular Expressions. When you have imported the re module, you can start using regular expressions: **import re**

- Q1.Check if the string starts with "The“

```
import re
 txt = "The Future Institute of Engineering  And
            Management"
x=re.search("^The", txt)
print(x)
```

- **Q2. Check if the string starts with "The" and ends with " Mgmt "**

```
import re
 txt = "The Future Institute of Engineering  And
            Mgmt"
x=re.search("Mgmt $", txt)
print(x)
```

- **Regex Functions**
- **The findall() function: returns a list containing all matches.**
- The list contains the matches in the order they are found. If no matches are found, an empty list is returned

Q3 W.A.Program to find if the word ‘the’ is there in the given string or not

```
import re
txt = "The rain in India"
x = re.findall("China", txt)
print(x)
```

Q4 **Find ‘The’ string in the given string**

```
import re
txt = "The rain in Theater Road"
x = re.findall("The", txt)
print(x)
```

- **The search() Function**
- The search() function searches the string for a match, and returns a Match object if there is a match. If there is more than one match, only the first occurrence of the match will be returned

- Q5. **WAP to find if the word ‘the’ is there in the given string or not**

```
import re
txt = "The rain in Theater Road"
x = re.search("The", txt)
print(x)
```

- Case insensitive regular expression :
- Q6: Find ‘the’ in the given string

```
import re
matches = re.findall('the','the rain Theater road',flags=re.IGNORECASE)
for match in matches:
    print(match)
```

- **Use re.finditer() method in Python regular expression**
- The re.finditer() works exactly the same as the re.findall() method except it **returns an iterator yielding match objects matching the regex pattern** in a string instead of a list.
- It scans the string from left to right, and matches are returned in the iterator form. Later, we can use this iterator object to extract all matches.

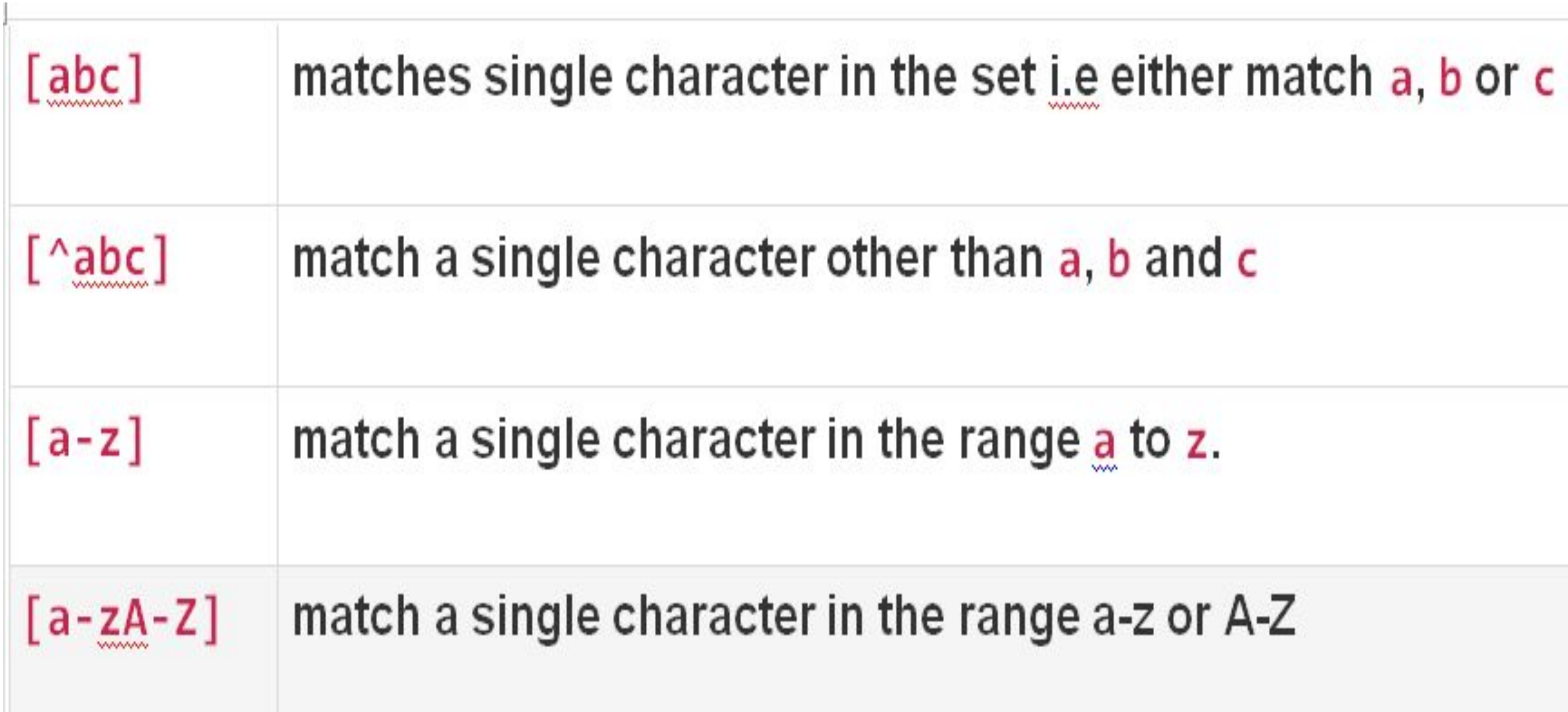
- **Q7 Find the substring ‘the’ using finditer()**

```
import re
matches = re.finditer('the','the rain Theater road')
for i in matches:
    print(i)
```

Output:

```
<re.Match object; span=(0, 3), match='the'>
<re.Match object; span=(9, 12), match='the'>
```

Metacharacters



[abc]	matches single character in the set i.e either match a, b or c
[^abc]	match a single character other than a, b and c
[a-z]	match a single character in the range a to z.
[a-zA-Z]	match a single character in the range a-z or A-Z

- **Q8 Find all lower case characters alphabetically between "a" and "m"**

```
import re
txt = "The rain in Spain“
x = re.findall("[a-m]", txt)
print(x)
```

Output:

```
['h', 'e', 'a', 'i', 'i', 'a', 'i']
```

.(a period) -- matches any single character except newline '\n'

```
import re
txt = "heeeeeeeo planet"
x = re.findall("he*o", txt)
print(x)
```

- **\d Matches any decimal digit; this is equivalent to the class [0-9].**

```
import re
s = "Tim's phone numbers are 12345-41521 and 78963-85214"
match = re.findall(r'\d\d', s)
if match:
   print(match)
```

Output:

```
['12', '34', '41', '52', '78', '96', '85', '21']
```

\d is Used as regular expression to find all digits in the string (one digit at a time by default other wise mention ‘\d{lenth}’

```
import re
s = "Tim's phone numbers are 12345-41521 and 78963-85214"
match = re.findall(r'\d{5}', s)
if match:
   print(match)
```

Output:

```
['12345', '41521', '78963', '85214']
```

- \d+   What does this pattern mean?
- The \d is a special regex sequence that matches any digit from 0 to 9 in a target string.
- The + metacharacter indicates number can contain at minimum one or maximum any number of digits.

```
import re
string = "Emma is a basketball player who was
    born on June 17, 1993. She played 112
    matches with scoring average 26.12 points
    per game. Her weight is 51 kg."
result = re.findall(r"\d+", string)
print(result)
```

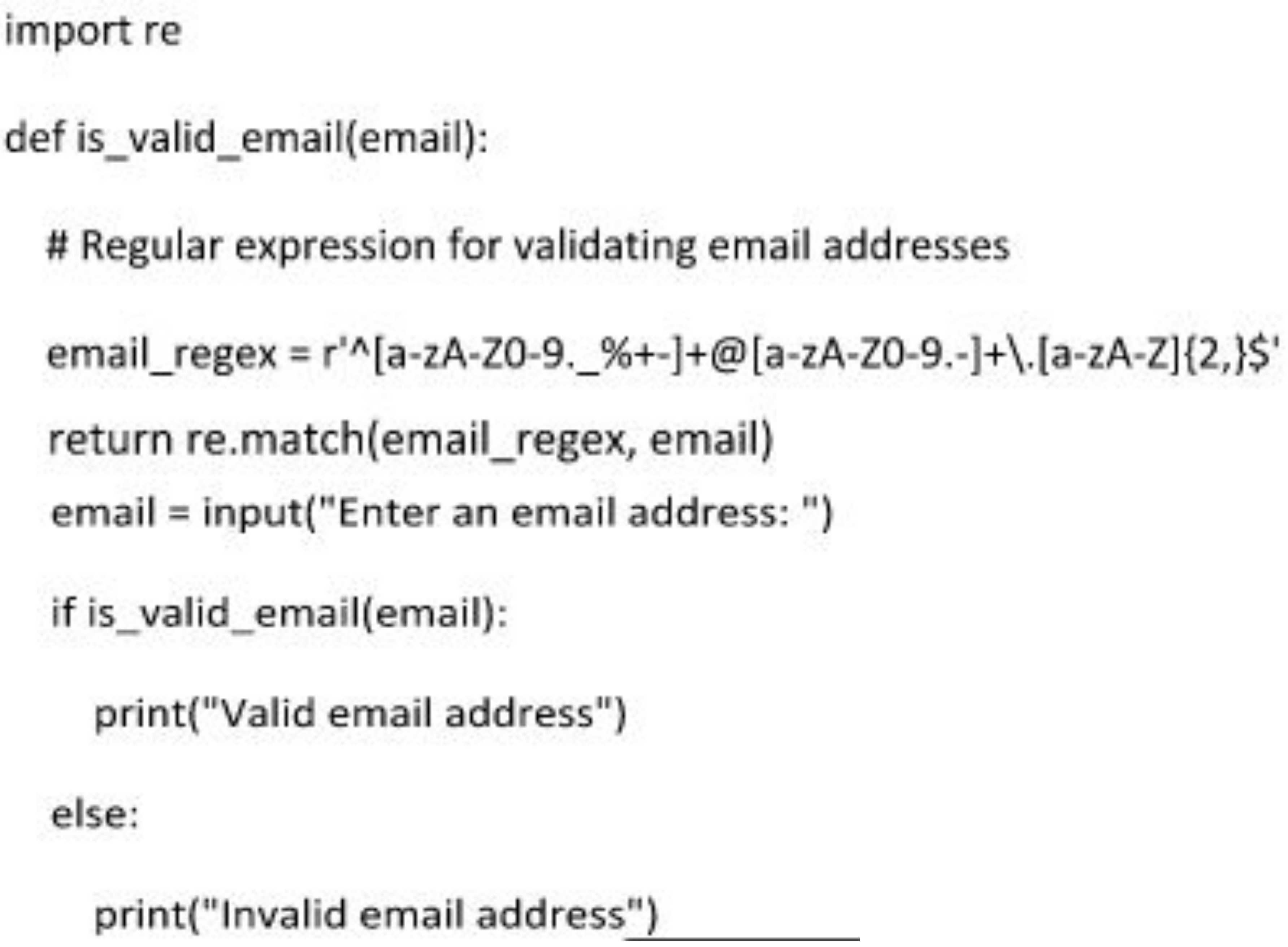
- **\D Matches any non-digit character**
- **\s Matches any whitespace character**
- **\S Matches any non-whitespace character**
- **\w Matches any alphanumeric character**
- **\W Matches any non-alphanumeric character.**
- **. Matches with any single character except newline ‘\n'.**
- **? match 0 or 1 occurrence of the pattern to its left**
- **+ 1 or more occurrences of the pattern to its left**
- *** 0 or more occurrences of the pattern to its left**
- **\b boundary between word and non-word. /B is opposite of /b**

- **\ It is used for special meaning characters like . to match a period or + for plus sign.**
- **{n,m} Matches at least n and at most m occurrences of preceding**
- **a| b Matches either a or b**

Assignments

- 1. Write a Python program using Regular expression to whether an input string is email or not.
- 2. Write a python program using regular expression to search a sequence that starts with ‘He’ and followed by 0 or more characters and an ‘o’.
- 3. Write a RE pattern to print all the words that with start with letters ‘a’, ‘b’, ‘f’ and ends with ‘t’ by ignoring the cases.
- 4. Write a RE pattern to find a particular substring, like “a…#” present in the given string using findall() function and print how many times it appear in this string.

Solution for assignment 1



```
import re

def is_valid_email(email):

    # Regular expression for validating email addresses

    email_regex = r'^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$'
    return re.match(email_regex, email)
    email = input("Enter an email address: ")

    if is_valid_email(email):

        print("Valid email address")

    else:

        print("Invalid email address")
```

Solution for Assignment 2

```
import re

text = "Hello, how are you? Heyo, Heo, Heo! Heo"

pattern = r'He.*o'

matches = re.findall(pattern, text)

print(matches)
```

Solution for Assignment 3

```
import re

    text = "The cat sat on the mat. A boat floated by the
            dock.bats sleep upside down."

pattern = r'\b[aAbBfF]\w*t\b'

matches = re.findall(pattern, text, re.IGNORECASE)

print(matches)
```

- \b asserts a word boundary to ensure that the match occurs at the beginning and end of words.
- [aAbBfF] matches any character 'a', 'A', 'b', 'B', 'f', or 'F'.
- \w* matches zero or more word characters (letters, digits, or underscores).
- t\b matches the character 't' at the end of the word followed by a word boundary.
- The re.IGNORECASE flag makes the pattern matching case-insensitive, so it matches words regardless of whether they are in lowercase or uppercase.

Solution for Assignment 4

- `import re`

- `text = "abc# ax#a# aadf# a123# aa# aaB# a##"`

- `pattern = r'a.{3}#'`

- `matches = re.findall(pattern, text)`

- `print(matches)`
- `print("Number of occurrences:", len(matches))`

- 'a' matches the character 'a'.
- '.{3}' matches any three characters.
- '#' matches the character '#'.
- So, 'a.{3}#' will match any substring that starts with 'a', followed by any three characters, and ends with '#'.

- The re.findall() function returns all non-overlapping matches of the pattern in the string.