

Python – Regular Expressions, Matplot

Αρχές Γλωσσών Προγραμματισμού και Μεταφραστών

Γ. Γαροφαλάκης, Σ. Σιούτας, Π. Χατζηδούκας



Python

- Regular Expressions
- Matplotlib
- Ανάκτηση δεδομένων από το διαδίκτυο



•Regular Expressions

- Αποτελούν μια ακολουθία χαρακτήρων για να δημιουργήσουμε μια μέθοδο αναζήτησης
- Η Python διαθέτει built-in το πακέτο re μέσω του οποίου μπορούμε να δημιουργήσουμε Regular Expressions
- Εισάγουμε τη βιβλιοθήκη με την εντολή **import re**

•Regular Expressions

Function	Description
findall	Returns a list containing all matches
search	Returns a Match object if there is a match anywhere in the string
split	Returns a list where the string has been split at each match
sub	Replaces one or many matches with a string

•Regular Expressions

Character	Description	Example
[]	A set of characters	"[a-m]"
\	Signals a special sequence (can also be used to escape special characters)	"\d"
.	Any character (except newline character)	"he..o"
^	Starts with	"^hello"
\$	Ends with	"world\$"
*	Zero or more occurrences	"aix*"
+	One or more occurrences	"aix+"
{}	Exactly the specified number of occurrences	"al{2}"
	Either or	"falls stays"

•Regular Expressions

Character	Description
\A	Returns a match if the specified characters are at the beginning of the string
\b	Returns a match where the specified characters are at the beginning or at the end of a word (the "r" in the beginning is making sure that the string is being treated as a "raw string")
\B	Returns a match where the specified characters are present, but NOT at the beginning (or at the end) of a word (the "r" in the beginning is making sure that the string is being treated as a "raw string")
\d	Returns a match where the string contains digits (numbers from 0-9)
\D	Returns a match where the string DOES NOT contain digits
\s	Returns a match where the string contains a white space character
\S	Returns a match where the string DOES NOT contain a white space character
\w	Returns a match where the string contains any word characters (characters from a to Z, digits from 0-9, and the underscore _ character)
\W	Returns a match where the string DOES NOT contain any word characters
\Z	Returns a match if the specified characters are at the end of the string

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται ο έλεγχος αν το string **txt** αρχίζει με “**The**” και τελειώνει με “**Italy**”

```
import re

txt = "The weather in Italy"
x = re.search("^The.*Italy$", txt)
```

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται split στο string txt σε κάθε white space χαρακτήρα

```
import re

txt = "The weather in Italy"
x = re.split("\s", txt)
```

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται split στο string txt μόνο στο πρώτο white space χαρακτήρα

```
import re

txt = "The weather in Italy"
x = re.split("\s", txt, 1)
```

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται αντικατάσταση στο string txt σε κάθε χαρακτήρα white space με το νούμερο 5

```
import re

txt = "The weather in Italy"
x = re.sub("\s", "5" txt)
```

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται αντικατάσταση στο string txt μόνο στις δυο πρώτες εμφανίσεις του χαρακτήρα white space με το νούμερο 5

```
import re

txt = "The weather in Italy"
x = re.sub("\s", "5" txt, 2)
```

•Regular Expressions - Παραδείγματα

Στο παρακάτω παράδειγμα γίνεται αναζήτηση στο string txt για το τμήμα που περιέχει ένα χαρακτήρα I με κεφαλαία στην αρχή μιας λέξης. Χρησιμοποιούμε το property group στην εκτύπωση για να μας επιστρέψει ολόκληρο το τμήμα που έγινε match

```
import re

txt = "The weather in Italy"
x = re.search(r"\bI\w+", txt)
Print(x.group())
```

Ανάκτηση δεδομένων από το διαδίκτυο (1/2)

Βιβλιοθήκη **urllib.request**

Η βιβλιοθήκη **urllib.request** επιτρέπει την ανάκτηση δεδομένων από διαδικτυακούς πόρους όπως ιστοσελίδες, XML, JSON και άλλους διαδικτυακούς πόρους.

req_object = urllib.request.Request(Δημιουργεί ένα αντικείμενο Request που χρησιμοποιείται για να προσδιορίσει τη διεύθυνση URL του διαδικτυακού πόρου που θέλουμε να αποκτήσουμε.)

mypage = urllib.request.urlopen(Χρησιμοποιείται για να ανοίξει το αντικείμενο Request και να λάβει την απάντηση από τον διαδικτυακό πόρο.)

Σημαντικό: χρειάζεται αποκωδικοποίηση η html:

html = mypage.read().decode()

Τα δεδομένα που επιστρέφονται συνήθως είναι σε δυαδική μορφή (bytes) και χρειάζονται αποκωδικοποίηση σε κείμενο.

Ανάκτηση δεδομένων από το διαδίκτυο (2/2)

```
import urllib.request
```

```
# Δημιουργία του αντικειμένου Request για τον πόρο  
req_object = urllib.request.Request('https://example.com')
```

```
# Ανάκτηση της σελίδας  
mypage = urllib.request.urlopen(req_object)
```

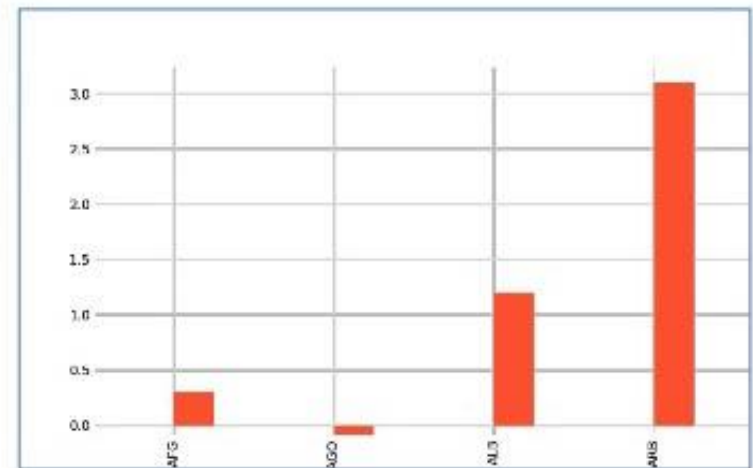
```
# Ανάγνωση και αποκωδικοποίηση του HTML περιεχομένου  
html = mypage.read().decode()
```

```
# Εκτύπωση των πρώτων 500 χαρακτήρων της HTML  
print(html[:500])
```

Matplot

- Χρησιμοποιείται για απεικόνιση γραφικών 2D

Country Name	Country Code	2010	2011	2012	2013	2014
Afghanistan	AFG	20.6	20.9	19.7	21.1	20.8
Angola	AGO	10.8	10.7	10.7	10.6	10.5
Albania	ALB	25.799999	27	28.3	28.7	29.2
Arab World	ARB	25.022214	28.11752	29.11321	29.33531	29.70457
United Arab Emirates	ARE	9.8000002	9.8	9.8	9.9	10
Argentina	ARG	19.5	18.8	18.4	19.7	21.3
Armenia	ARM	38.299999	38.7	35	32.5	35.1
Australia	AUS	11.4	11.4	11.7	12.2	13.1
Austria	AUT	8.8000002	8.2	8.7	9.1	9.2
Azerbaijan	AZE	14.6	14.5	14.3	13.4	13.6
Burundi	BDI	10.8	10.8	10.8	10.8	10.7
Belgium	BEL	22.5	18.6	19.7	23.1	23.6
Benin	BEN	2	2	2	1.8	1.7
Burkina Faso	BFA	5.1999998	5.3	5.2	5.2	5
Bangladesh	BGD	8.1999998	8.2	8.2	8.9	9.1
Bulgaria	BGR	22.9	25.2	28.2	29.7	25.9
Bahrain	BHR	10.2	11.4	10.5	10.6	10.9
Bahamas, The	BHS	36	27.2	30.4	30.8	30.1
Bosnia and Herzegovina	BIH	57.200001	57.1	61.7	57.4	57.5
Belarus	BLR	13.2	12.5	11.8	12	12
Belize	BLZ	20.9	24.3	26	22.4	22



Matplot

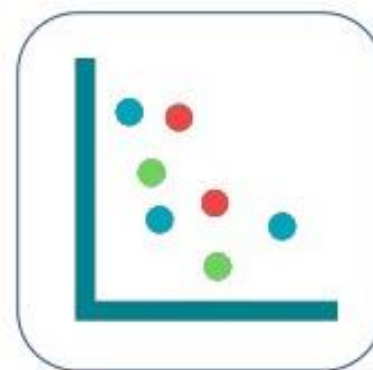
- Types of plots



Bar graph



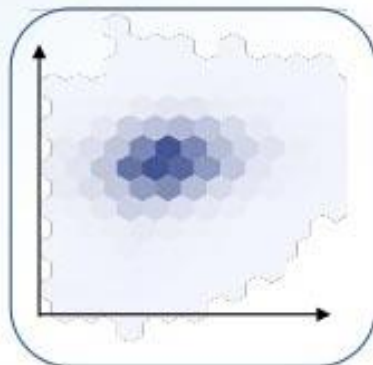
Histograms



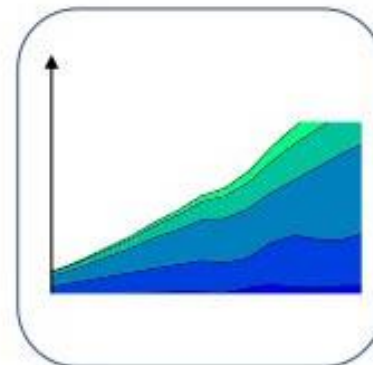
Scatter Plot



Pie Plot



Hexagonal Bin Plot



Area Plot

Matplot

- Παράδειγμα Plot

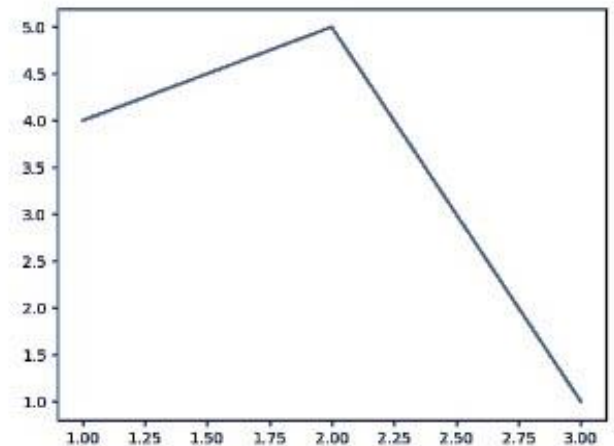
```
from matplotlib import pyplot as plt
```

```
#Plotting to our canvas
```

```
plt.plot([1,2,3],[4,5,1])
```

```
#Showing what we plotted
```

```
plt.show()
```



Matplot

- Προσθήκη τίτλων και ετικετών

```
from matplotlib import pyplot as plt
```

```
x = [5,8,10]
```

```
y = [12,16,6]
```

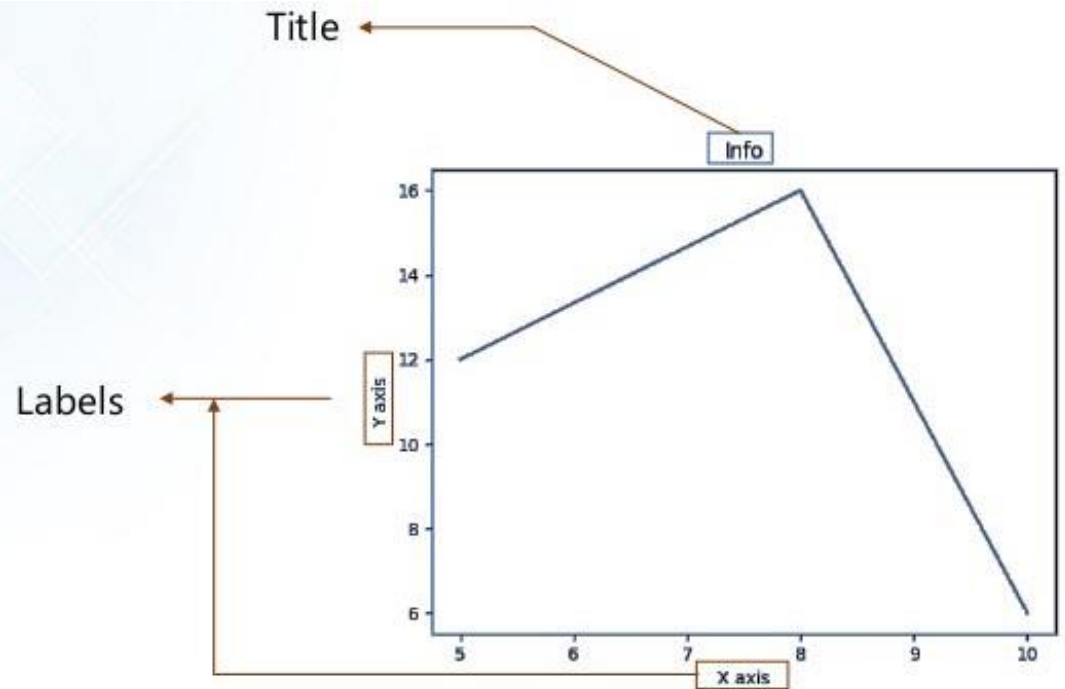
```
plt.plot(x,y)
```

```
plt.title('Info')
```

```
plt.ylabel('Y axis')
```

```
plt.xlabel('X axis')
```

```
plt.show()
```



Matplot

- Παράδειγμα Graph Bar

```
import matplotlib.pyplot as plt

plt.bar([1,3,5,7,9],[5,2,7,8,2], label="Example one")

plt.bar([2,4,6,8,10],[8,6,2,5,6], label="Example two", color='g')
plt.legend()
plt.xlabel('bar number')
plt.ylabel('bar height')

plt.title('Info')

plt.show()
```

