



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά
μαθήματα ΠΠ

Εισαγωγή

στους Η/Υ και τις Εφαρμογές

Ενότητα 5: Επεξεργασία δεδομένων με τη γλώσσα
προγραμματισμού python

Υπο-ενότητα 5.4: Αρχεία και επεξεργασία
δεδομένων

Μανώλης Τζαγκαράκης, Βικτωρία Δασκάλου
Σχολή Οργάνωσης και Διοίκησης Επιχειρήσεων
Τμήμα Οικονομικών Επιστημών

Σκοποί ενότητας

- Να πραγματοποιηθεί μία εισαγωγή στις δυνατότητες της γλώσσας `pytho`
- Να παρουσιαστούν οι βασικές δυνατότητες των γλωσσών προγραμματισμού για την επεξεργασία δεδομένων



Περιεχόμενα ενότητας

1. Χαρακτηριστικά της γλώσσας `python`
2. Προγραμματισμός με `python`
 - Υπολογισμοί και μεταβλητές
 - Συμβολοσειρές
 - Έλεγχος ροής
 - Λίστες
 - Αρχεία
 - Συναρτήσεις
 - Γραφικές παραστάσεις



Αρχεία

Αρχεία δεδομένων σε μορφή κειμένου

Σε διάφορα δημόσια αποθετήρια υπάρχουν δεδομένα σε μορφή πίνακα αποθηκευμένα ως αρχεία κειμένου τα οποία θέλουμε να επεξεργαστούμε με `rython`:

- Αρχεία κειμένου, δεδομένα σε γραμμές ή σε στήλες και χωρισμένα με τον χαρακτήρα `tab`
 - Αρχεία συνήθως με κατάληξη `.txt`
- Αρχεία κειμένων με δομημένη μορφή CSV (Comma Separated Value), χωρισμένα με χαρακτήρα κόμμα `“,”` ή άλλους χαρακτήρες όπως το `“;”` (ελληνικό ερωτηματικό)
 - Αρχεία συνήθως με κατάληξη `.csv`

Όλα τα αρχεία κειμένου επεξεργάζονται με απλούς επεξεργαστές κειμένου όπως το MS notepad

Διάβασμα από αρχείο κειμένου (1)

```
with open("animals.txt", mode="r", encoding="utf-8") as my_file:
    for line in my_file:
        print(line)
```

Βελτίωση: `print(line.rstrip("\n"))`

Μορφή αρχείου animals.txt

```
antelope
bear
cat
dog
elephant
fox
```

`open(file, mode='r', buffering=-1, encoding=None, errors=None, newline=None, closefd=True, opener=None)`

- `file`: όνομα αρχείου
- `mode`:
 - **w** - ανοίγει το αρχείο για εγγραφή (διαγράφει τα υπάρχοντα περιεχόμενα)
 - **r** - ανοίγει το αρχείο για ανάγνωση
 - **a** - ανοίγει το αρχείο και γράφει στο τέλος του διατηρώντας τα υπάρχοντα περιεχόμενα
- `encoding`: για αρχεία text, το προκαθορισμένο εξαρτάται από την πλατφόρμα & χρησιμοποιούμε το UTF8



Διάβασμα από αρχείο κειμένου(2)

- `myfile.read()` -> διαβάζει όλες τις γραμμές σε μία συμβολοσειρά `'antelope\nbear\ncat\ndog\nelephant\nfox'`
- `myfile.readlines()` -> διαβάζει όλες τις γραμμές σε μία λίστα
`['antelope', 'bear', 'cat', 'dog', 'elephant', 'fox']`
- `myfile.read().splitlines()` -> διαβάζει όλες τις γραμμές σε μία συμβολοσειρά και τη χωρίζει σε γραμμές
`['antelope\n', 'bear\n', 'cat\n', 'dog\n', 'elephant\n', 'fox']`
- `for line in myfile` -> διαβάζει μία-μία τις γραμμές του αρχείου ως συμβολοσειρές στη μεταβλητή `line`



Διάβασμα από αρχείο CSV

Μορφή αρχείου kalesmenoi.csv

```
"Οικ. Παπαδόπουλος";"Γαμπρού";2;3  
"Οικ. Φίδις";"Νύφης";1;1  
"Οικ. Ακρίδας";"Κουμπάρου";2;2  
"Οικ. Σαλούστρος";"Κουμπάρου";2;0  
"Οικ. Παπαγιάννη";"Νύφης";2;0
```

```
import csv  
with open('D:\TOE\mathima\kalesmenoi.csv', mode='r', newline='', encoding='utf-8',) as f:  
    reader=csv.reader(f, delimiter=';', quotechar='"', quoting=csv.QUOTE_NONNUMERIC)  
    for row in reader:  
        print(row)
```

1. Χρήση βιβλιοθήκης csv για αρχεία τύπου CSV (Comma Separated Values)
2. Άνοιγμα του αρχείου με την `open()`
3. Χρήση συνάρτησης `csv.reader()` για διάβασμα από αρχείο. Ορισμός του `delimiter`, `quotechar` και μετατροπή `non-quoted` σε `float`
4. Χρήση του `for in` για πρόσβαση σε μία-μία τις γραμμές



Διάβασμα από αρχείο CSV : εύρεση sum στήλης

```
import csv
names=[]
poiou=[]
megaloi=[]
paidia=[]
with open('D:\TOE\mathima\kalesmenoi.csv', mode='r', newline='',encoding='utf-8',) as f:
    reader=csv.reader(f,delimiter=';',quotechar='"', quoting=csv.QUOTE_NONNUMERIC)
    for row in reader:
        name,poianou,adult,child=row
        names.append(name)
        poiou.append(poianou)
        megaloi.append(adult)
        paidia.append(child)
print(sum(megaloi))
print(sum(paidia))
```



Διάβασμα από αρχείο CSV: εύρεση sum στήλης (2ος) τρόπος

```
import csv
with open('D:\TOE\mathima\kalesmenoi.csv', mode='r', newline='', encoding='utf-8',) as f:
    reader=csv.reader(f, delimiter=';', quotechar='"', quoting=csv.QUOTE_NONNUMERIC)
    names, poiou, megaloi, paidia=zip(*reader)
print(sum(megaloi))
print(sum(paidia))
```

- Χρήση συνάρτησης `zip` (*πίνακας) και χωρίζει σε στήλες



Βιβλιοθήκη Στατιστικής

```
from statistics import *
```

mean() ->Αριθμητικός μέσος

median() ->Διάμεσος

mode() ->Επικρατούσα τιμή διακριτών μεταβλητών

pstdev() ->Τυπική απόκλιση πληθυσμού

pvariance() ->Διακύμανση πληθυσμού

stdev() ->Τυπική απόκλιση δείγματος

variance() ->Διακύμανση δείγματος

+ Βασικές
συναρτήσεις
βαθμωτών
τύπων

- `max()`
- `min()`
- `sum()`



Έξοδος σε αρχείο csv

Μορφή αρχείου εξόδου

```
"name";"poianou";"megaloi";"paidia"  
"Οικ. Παπαδόπουλος";"Γαμπρού";2;3  
"Οικ. Φίδις";"Νύφης";1;1  
"Οικ. Ακρίδας";"Κουμπάρου";2;2
```

```
import csv  
data = [['name', 'poianou', 'megaloi', 'paidia'], ['Οικ. Παπαδόπουλος', 'Γαμπρού', 2.0, 3.0]]  
with open('myOutput.csv', mode='w', newline='') as myWFile:  
    writer = csv.writer(myWFile,  
                        delimiter=';', quotechar='"', quoting=csv.QUOTE_NONNUMERIC)  
    for line in data:  
        writer.writerow(line)
```

1. Άνοιγμα του αρχείου με την `open()` με `mode=w` (γράψιμο)
2. Χρήση συνάρτησης `csv.writer()` για γράψιμο στο αρχείο. Ορισμός του `delimiter`, `quotechar` και μετατροπή `non-quoted` σε `float`
3. Χρήση του `for in` για πρόσβαση σε μία-μία τις γραμμές
4. Χρήση `writerow` για γράψιμο γραμμής στο αρχείο



Παράδειγμα δημόσιων δεδομένων: Απογραφή πληθυσμού

- Έστω αρχείο με πληροφορίες για τον πληθυσμό της Ελλάδας, όπως αυτές προέκυψαν κατά την απογραφή του 2011 και δημοσιεύτηκαν από την Ελληνική Στατιστική Αρχή (<http://www.statistics.gr/>*)
- Το αρχείο περιέχει στοιχεία του πληθυσμού για κάθε διαμέρισμα, περιφέρεια, δήμο και δημοτική ενότητα της Ελλάδας κατά φύλλο και ομάδες ηλικιών

**Το φύλλο εργασίας είναι διαθέσιμο στη σελίδα*

<http://www.statistics.gr/portal/page/portal/ESYE/PAGE-census2011tables> (Πίνακες Απογραφής 201, Πίνακας 4α: Απογραφή Πληθυσμού 2011. Μόνιμος Πληθυσμός κατά φύλλο και ομάδες ηλικιών Μεγάλες Γεωγραφικές Ενότητες (NUTS 1), Αποκεντρωμένες Διοικήσεις, Περιφέρειες (NUTS 2), Περιφερειακές Ενότητες, Δήμοι, Δημοτικές Ενότητες)

Τμήμα αρχείου απογραφής ως CSV

Δημιουργία Αρχείου apograph.csv

```
"level";"code";"descr";"population"  
0;"          ";"ΣΥΝΟΛΟ ΧΩΡΑΣ";10815197  
1;"1  ";"ΒΟΡΕΙΑ ΕΛΛΑΔΑ";3110596  
2;"11  ";"ΑΠΟΚΕΝΤΡΩΜΕΝΗ ΔΙΟΙΚΗΣΗ ΜΑΚΕΔΟΝΙΑΣ - ΘΡΑΚΗΣ";2490051  
3;"111  ";"ΠΕΡΙΦΕΡΕΙΑ ΑΝΑΤΟΛΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ ΚΑΙ ΘΡΑΚΗΣ";608182  
4;"11101";"ΠΕΡΙΦΕΡΕΙΑΚΗ ΕΝΟΤΗΤΑ ΡΟΔΟΠΗΣ";112039  
5;"1110101";"ΔΗΜΟΣ ΚΟΜΟΤΗΝΗΣ";66919
```

- Κάθε γραμμή του αρχείου αφορά μία γεωγραφική περιοχή
- Οι στήλες της κάθε γραμμής του αρχείου:
 - Επίπεδο γεωγραφικής περιοχής (π.χ. Οι Δήμοι χαρακτηρίζονται στο επίπεδο 5)
 - Κωδικός
 - Περιγραφή
 - Πληθυσμός



Ερωτήματα

- Ποιός είναι ο μικρότερος πληθυσμός δήμου;
- Ποιός είναι ο μέσος όρος του πληθυσμού των δημοτικών ενοτήτων;
- Ποιά είναι η τυπική απόκλιση του πληθυσμού όλων των δήμων;
- Ποιά είναι η διακύμανση και η τυπική απόκλιση του πληθυσμού των ατόμων με ηλικία μικρότερη από ή ίση από 59 ετών ανεξαρτήτου φύλου των δημοτικών ενοτήτων;

Διάβασμα από csv σε λεξικό (1)

```
# Βρίσκει το μέσο όρο & τυπική απόκλιση του πληθυσμού δήμου του αρχείου apograph.csv
import csv,sys
from statistics import *
dhmoi=[]
filename='apograph.csv'
with open(filename, mode='r', newline='',encoding='utf-8',) as f:
    reader=csv.DictReader(f,delimiter=';',quotechar='', quoting=csv.QUOTE_NONNUMERIC)
    try:
        for row in reader:
            if row['level']==5:
                dhmoi.append(row['population'])
        print('Μέσος δήμων: '+'{:0.2f}'.format(mean(dhmoi)))
        print('Τυπική απόκλιση δήμων: '+'{:0.2f}'.format(stdev(dhmoi)))
    except csv.Error as e:
        sys.exit('file %s, line %d: %s' % (filename, reader.line_num, e))
```



Διάβασμα από csv σε λεξικό (2)

Νέα στοιχεία:

- Χρήση `csv.DictReader`: Διαβάζει σε δομή λεξικού για κάθε γραμμή:

```
{'level': 6.0, 'population': 6350.0, 'descr':  
'ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΜΑΡΩΝΕΙΑΣ', 'code': '111010402'}
```

- Χρήση `try` για χειρισμό των λαθών
- Χρήση `format()` για μορφοποιημένες εκτυπώσεις.

Δες `Format Cookbook`

<https://mkaz.com/2012/10/10/python-string-format/>



Εργασίες σε αρχεία και λίστες

Στο eclass στο χώρο του Μαθήματος Έγγραφα->Σημειώσεις->Python υπάρχει το αρχείο kalesmenoι.csv

Γράψτε σταδιακά πρόγραμμα σε python που να επεξεργάζεται το αρχείο CSV, το οποίο επεκτεινόμενο σε κάθε νέο ερώτημα να κάνει τα ακόλουθα:

1. Να διαβάζει το αρχείο γραμμή-γραμμή και να την εκτυπώνει στην οθόνη
2. Να αγνοεί την πρώτη γραμμή και να εκτυπώνει τις υπόλοιπες στην οθόνη
3. Να εκχωρεί τις τιμές των στηλών του αρχείου csv σε κατάλληλες δομές τύπου λίστας
4. Να εκχωρεί μόνο τις τιμές των καλεσμένων της νύφης σε κατάλληλες δομές τύπου λίστας
5. Να βρίσκει το όνομα της οικογένειας με το μεγαλύτερο αριθμό παιδιών που είναι καλεσμένοι της νύφης



Βιβλιογραφία

1. Εισαγωγή στους Υπολογιστές με τη Γλώσσα Python, Ν. Αβούρης - Κ. Σγάρμπας - Σ. Καξίρας - Μ. Κουκιάς - Β. Παλιουράς, 2013
2. A byte of Python (Ελληνικά)
[http://dide.flo.sch.gr/Plinet/Meetings/Meeting23/A Byte of Python-el.pdf](http://dide.flo.sch.gr/Plinet/Meetings/Meeting23/A%20Byte%20of%20Python-el.pdf)
3. Python Schools, <http://www.pythonschool.net/>
4. Non-Programmer's Tutorial for Python 3,
[http://en.wikibooks.org/wiki/Non-Programmer%27s Tutorial for Python 3/Print version](http://en.wikibooks.org/wiki/Non-Programmer%27s_Tutorial_for_Python_3/Print_version)
5. Hands-on Python Tutorial, Dr. A. N. Harrington, Loyola Uni. Chicago
<http://anh.cs.luc.edu/python/hands-on/3.1/handsonHtml/index.html>
6. Python Programming [http://en.wikibooks.org/wiki/Python Programming](http://en.wikibooks.org/wiki/Python_Programming)
7. Python Documentation, <https://docs.python.org/3.4/contents.html>



Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0



Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρών, Μανώλης Τζαγκαράκης, Βικτωρία Δασκάλου, Σχολή Οργάνωσης και Διοίκησης Επιχειρήσεων, Τμήμα Οικονομικών Επιστημών. «Εισαγωγή στους Η/Υ και Εφαρμογές. Επεξεργασία δεδομένων με τη γλώσσα προγραμματισμού python». Έκδοση: 1.0. Πάτρα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:

<https://eclass.upatras.gr/courses/ECON1242/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.



Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνα 1: Guido van Rossum, By Doc Searls (2006oscon_203.JPG) [CC BY-SA 2.0 (<http://creativecommons.org/licenses/by-sa/2.0>)], via Wikimedia Commons, Πηγή : https://en.wikipedia.org/wiki/Guido_van_Rossum

Εικόνα 2: Οι 10 πιο διαδεδομένες γλώσσες προγραμματισμού, By TIOBE Software B.V. [CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0>)], via Wikimedia Commons, Πηγή: <https://commons.wikimedia.org/wiki/File:Tiobeindex.png>

Εικόνα 3: Γράφημα που δημιουργήθηκε μέσω της βιβλιοθήκης Matplotlib, By <http://matplotlib.org/> (<http://matplotlib.org/>) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons, Πηγή : <https://commons.wikimedia.org/wiki/File:Matplotlib1.hires.png>

