



**ΠΕΡΙΕΧΟΜΕΝΑ:**

1. Διάβασμα και Γράψιμο αρχείων csv
2. Διαχείριση ως Λεξικό
3. Διαχωριστές και Τύλιγμα Πεδίων
4. Διάλεκτοι και Sniffers

**ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ:**

1. Python Advanced: Μάθημα 2 - Iterators

## ΜΑΘΗΜΑ 7.1: csv

### 1. Διάβασμα και Γράψιμο αρχείων csv

- Το **module csv** διαχειρίζεται αρχεία csv.
- Τα αρχεία csv (comma separated values) είναι συνηθισμένος τύπος αρχείου για μεταφορά δεδομένων, από προγράμματα όπως το excel ή οι βάσεις δεδομένων.

#### Διάβασμα από αρχείο csv:

- Έχοντας ανοίξει ένα αρχείο (κειμένου) για διάβασμα (βλ. και μάθημα 15)
- Χρησιμοποιούμε την κλάση reader:

Κατασκευαστής	Επεξήγηση
reader(f)	f: Χειριστής αρχείου

- Το αντικείμενο είναι iterable και κάνοντας επανάληψη επί αυτού, παίρνουμε τις γραμμές του csv, ως λίστες με τις τιμές που περιέχουν.

#### Παράδειγμα 1: read csv.py

```
from csv import reader
with open("countries.csv", "r", encoding="utf-8") as file:
    csv_reader = reader(file)
    print(type(csv_reader))
    for row in csv_reader:
        print(row)
```

- Παρατηρήστε ότι κάθε γραμμή είναι λίστα από συμβολοσειρές και δεν γίνονται μετατροπές τύπων.

#### Γράψιμο σε αρχείο csv:

- Έχοντας ανοίξει ένα αρχείο (κειμένου) για γράψιμο (βλ. και μάθημα 15)
- Χρησιμοποιούμε την κλάση writer:

Κατασκευαστής	Επεξήγηση
writer(f)	f: Χειριστής αρχείου

- και τη μέθοδό της:

Κατασκευαστής	Επεξήγηση
writerow(line)	Γράφει τη γραμμή line

#### Παράδειγμα 2: write csv.py

```
from csv import reader, writer

with open("countries.csv", "r", encoding="utf-8") as fread, \
    open("countries_new.csv", "w", encoding="utf-8", newline="") as fwrite:

    csv_reader = reader(fread)
    csv_writer = writer(fwrite)
    for line in csv_reader:
        newline = line[:3] + [line[14]]
        csv_writer.writerow(newline)
```

#### Παρατήρηση:

- Το όρισμα με λέξη-κλειδι newline="", απαιτείται μόνο στα Windows

**Διάβασμα csv ως λεξικό:**

- Χρησιμοποιούμε την κλάση DictReader:

Κατασκευαστής	Επεξήγηση
DictReader(f)	f: Χειριστής αρχείου

- Επιστρέφεται ένα αντικείμενο που είναι iterable.
- Κάθε στοιχείο είναι μία γραμμή δεδομένων, εκφρασμένη ως λεξικό. Κλειδιά είναι τα ονόματα των στηλών και τιμές είναι τα αντίστοιχα κελιά της γραμμής.

**Παρατήρηση:**

- Δίνεται η δυνατότητα να δώσουμε δικά μας κλειδιά, μέσω της παραμέτρου με λέξη-κλειδί fieldnames.

**Παράδειγμα 3: read\_csv\_dictionary.py**

```
from csv import DictReader, DictWriter

with open("countries.csv", "r", encoding="utf-8") as file:
    csv_reader = DictReader(file)
    data = list(csv_reader)
    print(data)

print("="*25)
row = data[0]
print(row)
print(row["Name"], row["Continent"])
```

**Γράψιμο λεξικού σε αρχείο csv:**

- Μέσω της κλάσης DictWriter

Κατασκευαστής	Επεξήγηση
DictWriter(f, fieldnames)	f: Χειριστής αρχείου fieldnames: καθορίζει τη σειρά με την οποία θα γραφούν τα κλειδιά

- Προετοιμάζουμε λεξικά τα οποία περιέχουν τα δεδομένα μας (πρέπει τα κλειδιά να περιέχονται στα fieldnames) και έπειτα καλούμε την μέθοδο του αντικειμένου:

Κατασκευαστής	Επεξήγηση
writerow(dict)	Γράφει το λεξικό dict

**Παράδειγμα 4: write\_csv\_dictionary.py**

```
from csv import DictReader, DictWriter

with open("countries.csv", "r", encoding="utf-8") as file:
    csv_reader = DictReader(file)
    data = list(csv_reader)

with open("countries_new3.csv", "w", encoding="utf-8", newline="") as file:
    csv_writer = DictWriter(file, fieldnames=["Name", "Continent"])
    csv_writer.writeheader()
    for row in data:
        new_row = {"Name": row["Name"],
                   "Continent": row["Continent"]}
        csv_writer.writerow(new_row)
```

**Σημείωση:**

- Το CSV δεν είναι αυστηρό πρότυπο.
- Ο διαχωρισμός δεν είναι ανάγκη να γίνεται με κόμμα, αλλά συχνά υπάρχουν csv αρχεία που ο διαχωρισμός γίνεται με ;, :, \t (ή με άλλο σύμβολο).

**Χρήση διαχωριστή:**

- Μπορούμε να ορίσουμε ποιος διαχωριστής χρησιμοποιείται στο αρχείο, όταν καλούμε την reader() και τη writer() αντίστοιχα, ορίζοντας το όρισμα με λέξη-κλειδί delimiter ίσο με το σύμβολο.

**Παράδειγμα 5: delimiter.py**

```
from csv import reader, writer

with open("excel.csv", "r") as file:
    csv_reader = reader(file, delimiter=";")
    data = list(csv_reader)

with open("excel_new.csv", "w", newline="") as file:
    csv_writer = writer(file, delimiter=",")
    for row in data:
        csv_writer.writerow(row)
```

**Σημείωση:**

- Ενδέχεται ο διαχωριστής να πρέπει να χρησιμοποιείται μέσα στα πεδία.
- Τότε μπορούμε να ενθέσουμε τα επίμαχα πεδία μέσα σε ειδικό χαρακτήρα, ώστε να μη δημιουργηθεί πρόβλημα.

**«Τύλιγμα» πεδίων:**

- Ορίζουμε το σύμβολο που θα τυλίσσει (μπρος και πίσω) τα πεδία, όταν καλούμε την reader() και τη writer() αντίστοιχα, ορίζοντας το όρισμα με λέξη-κλειδί quotechar ίσο με το σύμβολο.

**Παράδειγμα 6: quotechar.py**

```
with open("excel_new2.csv", "w", newline="") as file:
    csv_writer = writer(file, delimiter=",", quotechar="")
    for row in data:
        csv_writer.writerow(row)
```

**Άλλες παράμετροι:**

- (μεταξύ άλλων) προσφέρονται επίσης οι εξής παράμετροι με λέξεις-κλειδιά:

παράμετρος	επεξήγηση
lineterminator	χαρακτήρας που χρησιμοποιείται για το διαχωρισμό των γραμμών
escapechar	εναλλακτικός τρόπος για την αποφυγή του διαχωριστή που χρησιμοποιείται σε πεδίο

**Διάλεκτοι:**

- Κάθε reader ή writer δημιουργεί εσωτερικά ένα αντικείμενο της κλάσης **dialect**.
  - Περιέχει τις παραμέτρους με λέξεις-κλειδιά που χρησιμοποιούνται για την μετατροπή του csv από/σε δεδομένα.

**Παράδειγμα 7: dialect.py**

```
with open("excel.csv", "r") as file:
    csv_reader = reader(file, delimiter=";")
    dialect = csv_reader.dialect
    print("delimiter: " + dialect.delimiter)
    print("quotechar: " + dialect.quotechar)
    print("lineterminator: " + dialect.lineterminator)
```

**Σημείωση:**

- Ορίζονται και προκαθορισμένες διάλεκτοι για συνηθισμένους τύπους csv αρχείων:
  - **excel**: Αρχεία Excel
  - **excel\_tab**: Αρχεία Excel διαχωρισμένα με tab
  - **unix\_dialect**: Αρχεία UNIX
- Μπορούν να χρησιμοποιηθούν κατά το άνοιγμα ενός αρχείου csv στην παράμετρο με λέξη-κλειδί **dialect**.

**Sniffers:**

- Ενδιαφέρουσα δυνατότητα:
  - Να “μυριστούμε” τη διάλεκτο ενός αρχείου csv.
- Κατασκευάζουμε ένα αντικείμενο της κλάσης **Sniffer**.
- Καλούμε τη μέθοδό του sniff() με όρισμα ένα μέρος του αρχείου (π.χ. 1024 bytes)
- Επιστρέφεται ένα αντικείμενο Dialect που μπορούμε να το θέσουμε ως παράμετρο με λέξη-κλειδί στον κατασκευαστή του αναγνώστη.

**Παράδειγμα 8: sniffer.py**

```
from csv import reader, Sniffer

with open("countries.csv", "r", encoding="utf-8") as file:
    sniff = Sniffer().sniff(file.read(1024))
    file.seek(0)
    csv_reader = reader(file, dialect=sniff)
    dialect = csv_reader.dialect
    print("delimiter: " + dialect.delimiter)
    print("quotechar: " + dialect.quotechar)
    print("lineterminator: " + dialect.lineterminator)
    data = list(csv_reader)
    print(data)
```