

## Απαλλακτική Εργασία για το Μάθημα

### “ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΜΕ ΡΥΤΗΘΝ”

2024-2025

Ο στόχος της εργασίας είναι να αξιοποιήσετε τις γνώσεις σας στην Python για να δημιουργήσετε ένα πρόγραμμα που:

1. Θα αντλεί δεδομένα από μια δυναμική ιστοσελίδα μέσω web scraping.
2. Θα εκτελεί αιτήματα GET σε REST APIs για τη συλλογή δεδομένων.
3. Θα δημιουργεί και αποθηκεύει δεδομένα σε αρχεία JSON και εικόνες.
4. Θα παρέχει μια υπηρεσία/API μέσω ενός FastAPI wrapper.
5. Θα παράγει ένα γράφημα με ανάλυση δεδομένων.

Η εργασία περιλαμβάνει βασικές δεξιότητες που καλύψαμε στο μάθημα, όπως web scraping, requests, APIs, αποθήκευση δεδομένων και οπτικοποίηση. Χωρίζεται σε τρία διαφορετικά τμήματα με ξεχωριστά παραδοτέα, με διαφορετική συνεισφορά στο τελικό βαθμό. Υποχρεωτικό είναι μόνο το πρώτο τμήμα.

## Τμήμα 1 (70% του βαθμού)

Το βασικό script της εργασίας θα ονομάζεται **P2XXXX.py**, όπου **P2XXXX** είναι ο Α.Μ. σας.

Έχει δημιουργηθεί μια ιστοσελίδα (<https://tma111.netlify.app/netlify/functions/generate>) που δέχεται έναν αριθμό μητρώου (Α.Μ.) ως παράμετρο URL (π.χ. `generate?id=P12345`) και επιστρέφει 50 τυχαίους αριθμούς στο διάστημα 0-1000, οι οποίοι όμως είναι **ντετερμινιστικοί** (παραμένουν ίδιοι για τον ίδιο κωδικό Α.Μ.). Οι αριθμοί εμφανίζονται σε **διάφορες δυναμικές διατάξεις HTML** (π.χ. πλέγμα, πίνακας, παράγραφοι, λίστες κ.λπ.), σύμφωνα με τον Α.Μ. σας.

Τα ζητούμενα είναι τα εξής:

1. **Εξαγωγή quote IDs (25% του βαθμού)**: Με χρήση της βιβλιοθήκης **BeautifulSoup** (σε συνδυασμό με την βιβλιοθήκη requests) πρέπει να εξάγεται τους αριθμούς από τη σελίδα για το Α.Μ. σας, οι οποίοι αντιστοιχούν σε IDs για αποφθέγματα/quotes, να τους αποθηκεύσετε σε μία λίστα και να τους εμφανίσετε στο χρήστη. Ενίοτε, θα είναι απαραίτητο να χρησιμοποιήσετε τις απαραίτητες μεθόδους για συμβολοσειρές ή JSON ώστε να εξάγετε το αριθμητικό τμήμα που αντικατοπτρίζει το quote id.
2. **Ανάκτηση quotes για τα επιλεγμένα IDs (25% του βαθμού)**: Έπειτα, θα πρέπει να χρησιμοποιήσετε το API <https://dummyjson.com/quotes> για να συλλέξετε αποφθέγματα με τα IDs που προέκυψαν στο πρώτο βήμα. Παρουσίαση του API και των παραμέτρων του υπάρχει εδώ: <https://dummyjson.com/docs/quotes>. Κατόπιν, δημιουργήστε ένα αρχείο JSON (**quotes.json**) που θα περιέχει τα αποφθέγματα που

ανακτήσατε (μόνο τα IDs για το A.M. σας), ταξινομημένα ως προς το quote ID (από το μικρότερο στο μεγαλύτερο), με τη μορφή:

```
[
  {
    "id": 15,
    "quote": "I Will Praise Any Man That Will Praise Me.",
    "author": "William Shakespeare"
  },
  {
    "id": 28,
    "quote": "Technological Progress Is Like An Axe In The Hands Of A Pathological Criminal.",
    "author": "Albert Einstein"
  },
  {
    "id": 29,
    "quote": "No One Would Choose A Friendless Existence On Condition Of Having All The Other Things In The World.",
    "author": "Aristotle"
  }
]
```

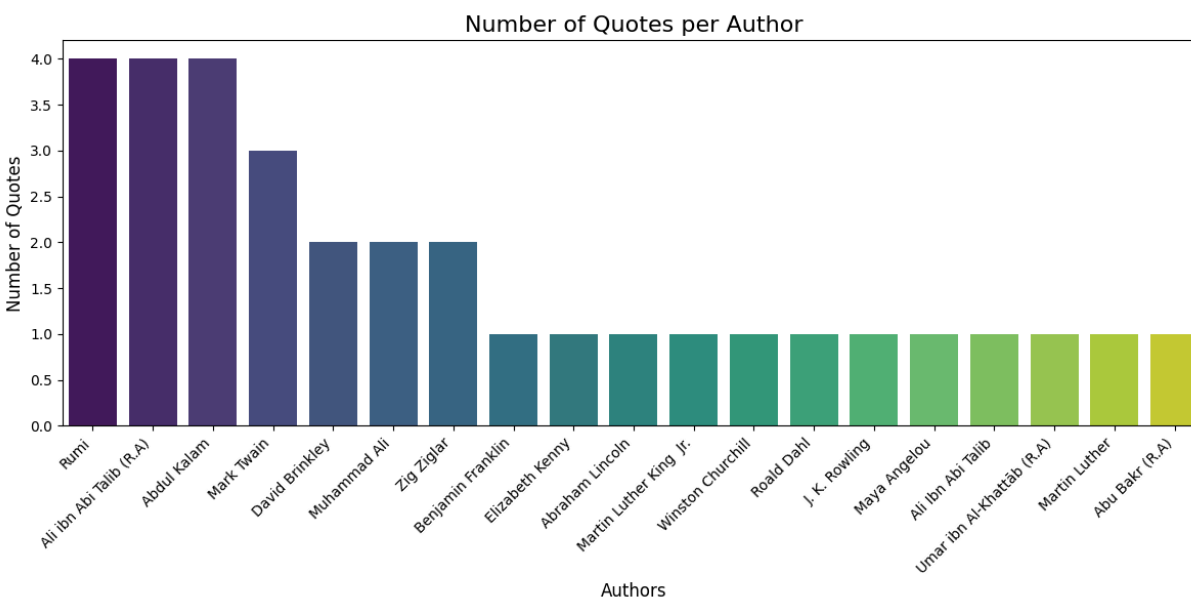
3. **Δημιουργία και αποθήκευση εικονων με τα αποφθέγματα (20% του βαθμού):** Στην δυναμική σελίδα παράγεται και ένα στοιχείο με διαφορετικό χρώμα παρασκήνιου και χρώμα κειμένου (div με id="colors"). Για κάθε απόφθεγμα που θα ανακτήσετε στο προηγούμενο βήμα, χρησιμοποιήστε το API <https://dummyjson.com/docs/image> για να δημιουργήσετε μια εικόνα που θα περιλαμβάνει το κείμενο του αποφθέματος και χρώμα παρασκήνιου/κειμένου ίδια με το στοιχείο div id="colors" στη σελίδα (π.χ. Για background-color (παρασκήνιο) "#1866db" και color (κείμενο) "#fdfff31" προκύπτει: <https://dummyjson.com/image/1200x200/1866db/fdfff31?text=Technological Progress Is Like An Axe In The Hands Of A Pathological Criminal.&fontSize=18>) και αποθηκεύστε τις εικόνες σε έναν φάκελο (/quotes/) στον ίδιο κατάλογο με το κύριο script της εργασίας σας, με όνομα το ID του εκάστοτε quote και κατάληξη ".png", π.χ. "15.png".

## Τμήμα 2 (15% του βαθμού)

Δημιουργήστε ένα API με χρήση της βιβλιοθήκης **FastAPI** που θα δέχεται το όνομα ενός συγγραφέα (author) ως παράμετρο GET, θα διαβάζει το αρχείο **quotes.json** και θα επιστρέφει σε JSON list τα αποφθέγματα του συγκεκριμένου συγγραφέα όπως υπάρχουν στο αποθηκευμένο json αρχείο (ή κενή λίστα αν δεν υπάρχουν δεδομένα για κάποιον author). Το όνομα του αρχείου πρέπει να είναι **api.py**, και θα εκθέτει ένα endpoint, το /quotes. Για παράδειγμα: <http://localhost:8080/quotes?author=Aristotle>

## Τμήμα 3 (15% του βαθμού)

Δημιουργήστε ένα ιστόγραμμα (histogram) χρησιμοποιώντας μία από τις βιβλιοθήκες **seaborn** ή **matplotlib**, το οποίο θα δείχνει τον αριθμό αποφθεγμάτων ανά συγγραφέα από το αρχείο **quotes.json**. Στον άξονα **x** θα πρέπει να υπάρχουν τα ονόματα των συγγραφέων, και στον άξονα **y** τον αριθμό των αποφθεγμάτων για τον αντίστοιχο συγγραφέα. Οι συγγραφείς θα πρέπει να είναι ταξινομημένοι σύμφωνα με τον αριθμό των quotes που τους αντιστοιχούν (από το μεγαλύτερο αριθμό στο μικρότερο). Διαμορφώστε την εμφάνιση του γραφήματος ελεύθερα. Για παράδειγμα:



### Οδηγίες για την υλοποίηση και παράδοση της εργασίας

- ΔΕΝ επιτρέπεται η χρήση της μεθόδου BeautifulSoup `find_all(text=True)` για την εξαγωγή κειμένου από το HTML. Θα πρέπει να χρησιμοποιήσετε αποκλειστικά μεθόδους όπως `find`, `find_all`, ή CSS selectors για την αναζήτηση HTML tags, attributes, και τη δομή DOM προκειμένου να εντοπίσετε και να εξάγετε τα απαραίτητα δεδομένα.
- ΔΕΝ επιτρέπεται η χρήση κανονικών εκφράσεων (regular expressions). Για την εξαγωγή των τιμών των IDs θα πρέπει να χρησιμοποιήσετε αποκλειστικά μεθόδους για strings, τεμαχισμό (slicing), το JSON module κλπ, σύμφωνα με την παραγόμενη δομή HTML για τον A.M. σας.
- Όλο το project θα πρέπει να είναι κατατεθειμένο σε έναν φάκελο που θα περιέχει:
  1. Τον κώδικα Python σε τρία αρχεία (**P2XXXX.py**, **api.py**, **plot.py**), για κάθε τμήμα της εργασίας. **ΜΟΝΟ το P2XXXX.py είναι υποχρεωτικό.**
  2. Το παραγόμενο JSON (**quotes.json**) και τις εικόνες png με τα quotes (50 στον αριθμό).

3. Σε περίπτωση που υλοποιήσατε το 3ο τμήμα της εργασίας, το ιστόγραμμα σε μορφή PNG (**histogram.png**).
- Η εργασία σας θα πρέπει να μπορεί να τρέχει απρόσκοπτα με εντολές από το τερματικό.
  - Φροντίστε να χειρίζεστε τυχόν σφάλματα/εξαιρέσεις στον κώδικα σας.

**Στον παρακάτω σύνδεσμο μπορείτε να βρείτε τα αναμενόμενα & σωστά παραδοτέα (πλην του κώδικα) για τον Α.Μ.: [Π123456](#):**

[https://thales.cs.unipi.gr/modules/document/file.php/TMA111/2024\\_2025/final\\_project/%CE%A0123456\\_example\\_output.zip](https://thales.cs.unipi.gr/modules/document/file.php/TMA111/2024_2025/final_project/%CE%A0123456_example_output.zip)

**Καλή επιτυχία !**