

Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών**Τμήμα Μηχανικών Η/Υ & Πληροφορικής****Πανεπιστήμιο Πατρών****Εαρινό Εξάμηνο 2022****Διδάσκοντες: Ιωάννης Γαροφαλάκης, Σπυρίδων Σιούτας, Παναγιώτης Χατζηδούκας****Προαιρετική Εργαστηριακή Άσκηση**

Το οικοσύστημα Python που περιλαμβάνει τη γλώσσα προγραμματισμού καθώς και μια σειρά από εξωτερικά πακέτα και βιβλιοθήκες λογισμικού είναι ένα ισχυρό εργαλείο ανοικτού κώδικα, το οποίο έχει μεγάλη χρησιμότητα εκτός από την αυστηρά οριοθετημένη περιοχή της επιστήμης των υπολογιστών, και στην έρευνα κάθε ειδικότητας αναφορικά με την ανάλυση και την επεξεργασία των δεδομένων.

Η γλώσσα Python είναι στην ουσία μια γλώσσα προγραμματισμού γενικού σκοπού, αντίθετα με άλλες γλώσσες οι οποίες επικεντρώνονται αυστηρά σε συγκεκριμένους τομείς όπως π.χ. η στατιστική ανάλυση. Το μεγάλο της πλεονέκτημα αφορά τη δυνατότητα της συγγραφής κώδικα στον οποίο μπορεί να γίνεται κάθε φορά χρήση των βιβλιοθηκών που είναι σχετικές με τις ζητούμενες επιστημονικές περιοχές του εκάστοτε project.

Ειδικότερα τώρα για τη συγκεκριμένη εργαστηριακή άσκηση θα χρησιμοποιηθεί ένα σύνολο δεδομένων με στοιχεία όπως η ποσότητα και το είδος των απορριμάτων που συλλέγονται προς ανακύκλωση στη πόλη του Μπάφαλο. Τα συγκεκριμένα δεδομένα παρέχονται από την ιστοσελίδα <https://data.buffalony.gov/Quality-of-Life/Monthly-Recycling-and-Waste-Collection-Statistics/2cjd-uvx7>

Ζητούμενα:

Αρχικά να υλοποιηθεί python script που θα συγκεντρώνει τα δεδομένα από την παραπάνω ιστοσελίδα του OpenData Buffalo. Στη συνέχεια να κατασκευαστεί πρόγραμμα σε python που θα επεξεργάζεται μόνο τα απαραίτητα δεδομένα σχετικά με τα ζητούμενα γραφήματα της άσκησης.

Συγκεκριμένα θα χρειαστεί να εξαχθούν τα γραφήματα για τις παρακάτω περιπτώσεις:

- Συνολικές ποσότητες ανακυκλώσιμων ειδών ανά έτος
- Συνολική παρουσίαση ανακυκλώσιμων ειδών και των αντίστοιχων ποσοτήτων τους
- Παρουσίαση των 5 μηνών με τη μεγαλύτερη ποσότητα ανακυκλώσιμων ειδών, ανεξαρτήτως έτους και είδους ανακυκλώσιμων ειδών

Στο τέλος τα συγκεκριμένα δεδομένα να φορτωθούν σε αντίστοιχους πίνακες μιας βάσης MySQL (ή όποιας βάσης επιθυμείτε), και να εξαχθούν σε αντίστοιχα αρχεία .csv (τα .csv θα αποτελούν επίσης παραδοτέα της άσκησης).

Προσοχή όλες οι απαραίτητες ενέργειες μετασχηματισμού των δεδομένων θα πρέπει να γίνουν αποκλειστικά σε γλώσσα Python. Η βάση δεδομένων θα χρησιμοποιηθεί **ΜΟΝΟ** για αποθήκευση.

Υπόδειξη: Χρησιμοποιείτε τις απαραίτητες βιβλιοθήκες της Python όπως pandas, xlrd, matplotlib ή όποιες άλλες κρίνετε εσείς απαραίτητο.

Παραδοτέα

1. **Γραπτή Αναφορά** (σε αρχείο pdf ή word) που θα περιλαμβάνει:
 - Τον κώδικα σε γλώσσα python εμπλουτισμένο με σχόλια
 - Screenshots παραδειγμάτων της εφαρμογής (και του σχήματος της βάσης δεδομένων)
 - Τα ζητούμενα γραφήματα (με τίτλους, υπόμνημα)
 - Σχόλια - Παραδοχές που έγιναν για την ανάπτυξη της εργασίας
2. Συμπίεσμένα σε **ένα αρχείο zip**:
 - Την πιο πάνω γραπτή αναφορά
 - Τον **ΤΕΛΙΚΟ** κώδικα σε python.
 - Export του σχήματος, data της βάσης δεδομένων
 - Τα αρχεία .csv που θα περιέχουν τα τελικά δεδομένα

Το αρχείο zip πρέπει να έχει όνομα τον αριθμό μητρώου του φοιτητή (π.χ. **3972.zip**), και να ανεβεί (**ΥΠΟΧΡΕΩΤΙΚΑ**) στο e-class. Σε ξεχωριστό αρχείο .txt μέσα στο zip να αναφέρεται το ονοματεπώνυμο, το έτος, ο αριθμός μητρώου και η e-mail διεύθυνση του φοιτητή.

Διευκρινήσεις

1. Η άσκηση είναι **ατομική**
2. Η άσκηση είναι **προαιρετική με bonus έως 1 μονάδα στον τελικό βαθμό, εφόσον ο βαθμός της είναι ≥ 5**
3. Τελική ημερομηνία παράδοσης είναι η ημερομηνία της γραπτής εξέτασης Ιουνίου 2022 **ΜΟΝΟ!** Αναλόγως θα καθοριστεί και η ημερομηνία της προφορικής εξέτασης της εργασίας.
4. Η άσκηση **ΔΕΝ ΘΑ ΚΡΑΤΗΘΕΙ** για το επόμενο ΑΚΑΔΗΜΑΙΚΟ ΕΤΟΣ!

5. Για τυχόν απορίες ή υποδείξεις μπορείτε να απευθύνεστε με e-mail στο mnvonitsanos@ceid.upatras.gr ή στο χώρο «Συζητήσεις» στη σελίδα του μαθήματος στο eclass

(<https://eclass.upatras.gr/modules/forum/?course=CEID1091>)