

Άσκηση Πράξης 10

Εξαγωγή πληροφορίας από το διαδίκτυο (*web scraping*)

- 1 ***RSS_Meteo_Regex.py*** **Εξαγωγή ημερήσιου δελτίου καιρού από rss news feed του [meteo.gr](http://www.meteo.gr)**. Μεταβείτε στην ιστοσελίδα www.meteo.gr/rss/news.cfm και κατεβάστε το περιεχόμενό της ως "ακατέργαστο" αρχείο κειμένου (το συγκεκριμένο αρχείο δεν είναι κωδικοποιημένο με το πρωτόκολλο html αλλά το πρωτόκολλο xml – πρακτικά όμως αυτό δεν έχει για εσάς καμία διαφορά στον τρόπο περαιτέρω χειρισμού του). Αφού μελετήσετε την δομή του αρχείου, δημιουργήστε ένα πρόγραμμα σε Python το οποίο θα χρησιμοποιεί κανονικές εκφράσεις για να ταιριάζει και να εμφανίζει *την πρόγνωση του σημερινού καιρού*.
 - 1α. ***RSS_Meteo_BSoup.py*** Κάντε το ίδιο χωρίς χρήση κανονικών εκφράσεων, χρησιμοποιώντας την βιβλιοθήκη BeautifulSoup.
 - 2 ***HMU_Calendar_Regex.py*** **Εξαγωγή ακαδημαϊκού ημερολόγιου από την αντίστοιχη ιστοσελίδα του ΕΛ.ΜΕ.ΠΑ** Μεταβείτε στην ιστοσελίδα <https://hmu.gr/akadimaiko-imerologio-2024-2025> και κατεβάστε το περιεχόμενό της ως "ακατέργαστο" αρχείο κειμένου html. Αφού μελετήσετε την δομή του αρχείου, δημιουργήστε ένα πρόγραμμα σε Python το οποίο θα χρησιμοποιεί κανονικές εκφράσεις για να ταιριάζει και να εμφανίζει *το Ακαδημαϊκό Ημερολόγιο του τρέχοντος έτους*.
 - 2α. ***HMU_Calendar_BSoup.py*** Κάντε το ίδιο χωρίς χρήση κανονικών εκφράσεων, χρησιμοποιώντας την βιβλιοθήκη BeautifulSoup.
 - 3 ***ECE_HMU_News_Regex.py*** **Εξαγωγή των νέων του Τμήματος ΗΜΜΥ / ΕΛΜΕΠΑ από την αντίστοιχη ιστοσελίδα**. Μεταβείτε στην ιστοσελίδα https://ece.hmu.gr/news_gr/ και κατεβάστε το περιεχόμενό της ως "ακατέργαστο" αρχείο κειμένου html. Αφού μελετήσετε την δομή του αρχείου, δημιουργήστε ένα πρόγραμμα σε Python το οποίο θα χρησιμοποιεί κανονικές εκφράσεις για να ταιριάζει και να εμφανίζει *τα τρέχοντα νέα του Τμήματος ΗΜΜΥ / ΕΛΜΕΠΑ*.
 - 3α. ***ECE_HMU_News_BSoup.py*** Κάντε το ίδιο χωρίς χρήση κανονικών εκφράσεων, χρησιμοποιώντας την βιβλιοθήκη BeautifulSoup.

(για το σπίτι – Εργασία 10^η)

- 4 ***RSS_Meteo_ALL.py*** Η ιστοσελίδα www.meteo.gr/rss/news.cfm εκτός από την πρόγνωση του σημερινού καιρού, περιέχει -όπως θα παρατηρήσετε- και τα σημαντικότερα νέα της ημέρας. Επεκτείνετε το πρόγραμμα της Άσκησης 1 ώστε κάθε φορά που θα τρέχει, *μαζί με την πρόγνωση του σημερινού καιρού*, να εμφανίζει ένα-ένα και *τα σημαντικότερα νέα της ημέρας*. Μπορείτε να το κάνετε είτε με χρήση κανονικών εκφράσεων, είτε με BeautifulSoup.

Ανατρέξτε στις σημειώσεις της θεωρίας για να δείτε πως μπορείτε να εγκαταστήσετε τις βιβλιοθήκες requests (και πιθανόν BeautifulSoup) που θα χρειαστείτε.