

ΥΠΟΛΟΓΙΣΤΙΚΗ ΓΛΩΣΣΟΛΟΓΙΑ

ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ

"REGULAR EXPRESSIONS (ΚΑΝΟΝΙΚΕΣ ΕΚΦΡΑΣΕΙΣ)"

Στην άσκηση αυτή καλείστε να χρησιμοποιήσετε regular expressions για να αντλήσετε πληροφορίες από ένα αρχείο κειμένου. Το αρχείο (weather.txt - δίνεται μαζί με την εκφώνηση) αποτελεί τμήμα μετεωρολογικού δελτίου και περιέχει ημερομηνίες, τοπωνύμια και τιμές μετεωρολογικών παραμέτρων. Σκοπίμως είναι μικρό σε μέγεθος, ώστε να μπορείτε να επαληθεύσετε οπτικά αν τα regular expressions που γράφετε δίνουν σωστά αποτελέσματα.

ΠΡΟΕΤΟΙΜΑΣΙΑ

Για την άσκηση αυτή θα χρειαστείτε τη γλώσσα Python, έκδοση 3.x (<https://www.python.org/>). Διαβάστε το "Regular Expression HOWTO" (<https://docs.python.org/3/howto/regex.html>) και το documentation του module re (<https://docs.python.org/3/library/re.html>)

ΔΙΑΔΙΚΑΣΙΑ

ΕΡΩΤΗΜΑ #1

Διαβάστε το αρχείο "weather.txt" με τον ακόλουθο κώδικα Python:

```
import re

f=open("weather.txt", mode="r", encoding="utf8")
content=f.read()
f.close()

pattern=re.compile("[0-9]") # <-- Εδώ βάζουμε το RE
result=pattern.findall(content)
print(result)
print(len(result))
```

Γράψτε τί τοπώνει το πρόγραμμα.

ΕΡΩΤΗΜΑ #2

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τοπώνει **ολόκληρους τους (πραγματικούς ή ακέραιους) αριθμούς που εμφανίζονται στο κείμενο**. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #3

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τις τιμές mm που εμφανίζονται στο κείμενο**. Δεν πειράζει αν τυπώνει και τη μονάδα μέτρησης (mm) ή όχι. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #4

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τα περιεχόμενα των παρενθέσεων που εμφανίζονται στο κείμενο**. Δεν πειράζει αν τυπώνει και τα σύμβολα των παρενθέσεων ή όχι. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #5

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τις ημερομηνίες που εμφανίζονται στο κείμενο**. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #6

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τις ημερομηνίες που εμφανίζονται στο κείμενο, μαζί με την προηγούμενη λέξη κάθε ημερομηνίας**. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #7

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τις λέξεις που περιέχουν τονισμένο φωνήεν**. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΕΡΩΤΗΜΑ #8

Αλλάξτε το regular expression στο πρόγραμμα, έτσι ώστε να τυπώνει **μόνο τις λέξεις που ξεκινούν με κεφαλαίο γράμμα και έχουν μήκος τουλάχιστον 3 χαρακτήρες**. Γράψτε το νέο regular expression και την έξοδο του προγράμματος.

ΠΑΡΑΔΟΤΕΟ

Παραδίδετε ένα αρχείο κειμένου (κατά προτίμηση PDF) στο οποίο θα έχετε γράψει τα στοιχεία σας (Ονοματεπώνυμο, Τμήμα, Κατεύθυνση ή Πρόγραμμα Σπουδών, Έτος Φοίτησης και Αριθμός Μητρώου) και τις απαντήσεις στα ερωτήματα.