ΠΑΡΑΔΕΙΓΜΑ ΠΡΟΣΩΠΙΚΗΣ ΤΕΚΜΗΡΙΩΣΗΣ / ΕΡΓΑΣΙΑ Python (ΕΚΔΔΑ) / ΑΝΑΛΥΤΙΚΑ #1:

Να δημιουργήσετε ένα σενάριο σε Python με το όνομα parse.py, το οποίο θα διαβάζει μια ιστοσελίδα που έχετε αποθηκεύσει στο δίσκο σας και θα εμφανίζει στην οθόνη όλους τους συνδέσμους και το κείμενο που τους συνοδεύουν. Προσοχή, το σενάριο δεν θα κατεβάζει την ιστοσελίδα από το διαδίκτυο, αυτό θα το κάνουμε σε επόμενη εβδομάδα.

Για εφαρμογή θα χρησιμοποιήσετε το αρχείο seminaria.html το οποίο περιλαμβάνει όλα τα σεμινάρια πληροφορικής του ΙΝΕΠ για το μήνα Δεκέμβριο 2013 (για να το κατεβάσετε κάντε δεξί κλικ και επιλέξτε "Αποθήκευση δεσμού ως...").


Πρόγραμμα το οποίο απευθύνεται σε στελέχη περισσότερων του ενός φορέων σύμφωνα με τη σχετική εγκύκλιο
Πρόγραμμα το οποίο απευθύνεται σε στελέχη ενός φορέα ή κανονοποιείται με παλαιότερες απήσεις
Αναζήτηση για
• Περιφέρεια: ΑΤΤΙΚΗΣ
• Τόπος Διεξαγωγής: Όλοι
• Θεματικός Κύκλος: ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΕΦΑΡΜ. ΤΕΧΝ.
• Τίτλος: Όλοι
• Ημερομηνία: Εναρξη μετά τις 04/12/2013
Συνολικά βρέθηκαν 9 αποτελέσματα.
Σειρά Ταξινόμησης Τόπος Διεξαγωγή
Α/Α	Τόπος Διεξαγωγής	Τίτλος	Εναρξη	Λήξη	Διάρκεια (Ωρες)	Κωδ (Επικοι
1	ΑΘΗΝΑ	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑΦΩΝ	05/12/2013	06/12/2013	14	8023
2	ΑΘΗΝΑ	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑΦΩΝ	05/12/2013	06/12/2013	14	8023
3	ΑΘΗΝΑ	ΤΑ ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ UNIX ΚΑΙ LINUX	09/12/2013	13/12/2013	35	8023
4	ΑΘΗΝΑ	ΣΧΕΔΙΑΣΜΟΣ ΕΦΑΡΜΟΓΩΝ ΔΙΑΔΙΚΤΥΟΥ ΜΕ ΕΡΓΑΛΕΙΑ ΑΝΟΙΚΤΟΥ ΛΟΓΙΣΜΙΚΟΥ	09/12/2013	13/12/2013	35	8023
5	ΑΘΗΝΑ	ΔΙΑΧΕΙΡΙΣΗ WINDOWS SERVER 2008	09/12/2013	20/12/2013	70	8023
6	ΑΘΗΝΑ	ΟΡΓΑΝΩΣΗ ΚΑΙ ΧΡΟΝΟΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΕΡΓΩΝ	09/12/2013	20/12/2013	70	8023
7	ΑΘΗΝΑ	ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ	16/12/2013	18/12/2013	21	8023
		ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ				

Εικόνα 1: Το αρχείο seminaria.html όπως φαίνεται στον Browser

Για παράδειγμα, αν εκτελέσετε το σενάριο, θα πρέπει να ζητήσει το μονοπάτι του αρχείου και εν συνεχεία θα εκτυπώσει:

```
>>>
Type filepath : c:\seminaria.html


http://www.ekdd.gr/ekdda/custom/seminars/pdf/10627_2_201
ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10627_2_201
ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10394_2_201
ΤΑ  ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ UNIX ΚΑΙ LINUX

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10315_2_201
ΣΧΕΔΙΑΣΜΟΣ ΕΦΑΡΜΟΓΩΝ ΔΙΑΔΙΚΤΥΟΥ ΜΕ ΕΡΓΑΛΕΙΑ ΑΝΟΙΚΤΟΥ ΛΟΓ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10419_2_201
ΔΙΑΧΕΙΡΙΣΗ WINDOWS SERVER 2008

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10379_2_201
ΟΡΓΑΝΩΣΗ ΚΑΙ ΧΡΟΝΟΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΕΡΓΩΝ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10324_2_201
ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10321_2_201
ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10319_2_201
ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ
```

Εικόνα 2: Η έξοδος της εκτέλεσης του σεναρίου

ΟΔΗΓΙΕΣ

Το αρχείο seminaria.html, όπως και όλες οι ιστοσελίδες, είναι αρχεία κειμένου, που σημαίνει ότι μπορείτε να το ανοίξετε με έναν απλό επεξεργαστή κειμένου, για παράδειγμα το Notepad (Σημειωματάριο). Το πρώτο πράγμα που πρέπει να κάνετε, αφού ανοίξετε το αρχείο, είναι να εντοπίσετε την πληροφορία που εμφανίζεται στην Εικόνα 2 και εν συνεχεία να καταλάβετε ότι αυτή η πληροφορία συνοδεύεται πριν και μετά, από συγκεκριμένες λέξεις. Για να ολοκληρώσετε με επιτυχία την εργασία σας θα πρέπει:

1. να εντοπίσετε τις συγκεκριμένες λέξεις που περικλείουν είτε τον υπερσύνδεσμο είτε το κείμενο του και να καταλάβετε το πρότυπο μορφοποίησης που χρησιμοποιεί επαναληπτικά ο συγγραφέας της ιστοσελίδας,

2. εν συνεχεία να δημιουργήσετε μια συνάρτηση με το όνομα subtext, η οποία θα έχει τουλάχιστον ως παραμέτρους, ένα αλφαριθμητικό μέσα στο οποίο θα ψάξουμε, το αρχικό και τελικό αλφαριθμητικό της μορφοποίησης. Για παράδειγμα, αν εκτελέσουμε την συνάρτηση όπως παρακάτω:

Παράδειγμα 1
text = subtext("Greece is a wonderful country", "ee", "erf")

θα πρέπει να μας επιστρέψει το κείμενο "ce is a wond"

Προσοχή, δεν υπάρχει κανένα πρόβλημα αν θέλετε να βάλετε περισσότερες παραμέτρους στην subtext. Εκτός από τις 3 πρώτες παραμέτρους που θα πρέπει να υπάρχουν στη συνάρτηση, έχετε την ελευθερία -αν εσείς το θέλετε - να σχεδιάσετε όλες τις υπόλοιπες παραμέτρους.

Π.χ. να δηλώνετε μια θέση από την οποία θα ξεκινήσει η έρευνα:

Παράδειγμα 2
text = subtext("Greece is a wonderful country", "ee", "erf", 5)

Θα πρέπει να επιστρέψει "" γιατί δεν υπάρχει το αλφαριθμητικό "ee" μετά το 5 χαρακτήρα.

Πιθανόν να μπορεί να θέλετε να δηλώσετε με λογικές μεταβλητές αν η συνάρτηση θα πρέπει να επιστρέψει μαζί με το κείμενο και τις λέξεις της μορφοποίησης.

Παράδειγμα 3
text = subtext("Greece is a wonderful country", "ee", "erf", 0, True, True)

θα πρέπει να σας επιστρέψει το κείμενο "eece is a wonderf". Η πρώτη λογική τιμή αναφέρεται στη προσθήκη της αρχικής λέξης μορφοποίησης και η δεύτερη λογική τιμή στη τελική λέξη.

Επίσης και η λογική των παραμέτρων είναι στην ευθύνη σας. Για παράδειγμα μπορείτε να ενσωματώσετε τη λογική, ότι αν η τελική λέξη μορφοποίησης είναι κενή κατά τη κλήση της συνάρτησης, τότε η συνάρτηση, αν υπάρχει η αρχική λέξη μορφοποίησης να επιστρέφει όλο το κείμενο από το σημείο της εύρεσης της ως το τέλος.

Παράδειγμα 4
text = subtext("Greece is a wonderful country", "ee", "")

θα πρέπει να σας επιστρέψει το κείμενο "ce is a wonderful country"

Το σενάριο parse.py, αφού ανοίξει το αρχείο seminaria.html και διαβάσει το περιεχόμενο του, θα χρησιμοποιήσει τη subtext για να εξαγάγει τον πρώτο υπερσύνδεσμο, με την επόμενη κλήση της το κείμενο που τον συνοδεύει και να τα εκτυπώσει στην οθόνη. Με επαναληπτικές κλήσεις της subtext θα πρέπει εκτυπωθούν όλοι οι σύνδεσμοι με τα κείμενα τους, όπως είδαμε στη Εικόνα 2.

Καλή επιτυχία!

ΑΠΑΝΤΗΣΕΙΣ:

Κρίσιμες παραδοχές: έχουμε αποθηκεύσει το αρχείο γλώσσας υπερκειμένου seminaria.html στον φάκελο «ρίζας» του τοπικού μας σκληρού δίσκου (C) καθώς και ότι το αρχείο πως είναι υπαρκτό και με «ζεύγη» τιμών URL/ΤΙΤΛΟΥ.

«Σκέψεις»-Ανάλυση: ανοίγουμε το html αρχείο με έναν σύγχρονο επεξεργαστή κειμένου (εδώ το Notepad++) και επιλέγουμε στην Προβολή (View) > Αναδίπλωση γραμμών (Word wrap), προκειμένου να ανιχνεύσουμε οπτικά το (κοινό) «μοτίβο» κώδικα που «περικλείει» τις επιθυμητές πληροφορίες που πρέπει να «απομονώσουμε» και να απεικονίσουμε στην οθόνη μας, σαρώνοντας σειριακά όλες τις γραμμές του αρχείου αυτού.

Με βάση λοιπόν την προβολή αυτή, παρατηρούμε πως ΓΙΑ ΤΟ ΣΥΓΚΕΚΡΙΜΕΝΟ ΠΕΡΙΕΧΟΜΕΝΟ ΤΟΥ ΑΡΧΕΙΟΥ (χρωματικός κώδικας στην απεικόνιση):



- Η εκάστοτε πληροφορία που επιθυμούμε να εστιάσουμε, περιέχεται κάθε φορά ΑΥΤΟΤΕΛΩΣ σε κάποια ξεχωριστή γραμμή του
- Πριν τον εκάστοτε σχετιζόμενο υπερσύνδεσμο (link) διατηρείται σταθερό το μοτίβο με περιεχόμενο:
  ΑΘΗΝΑ</td><td style="font-size:11px;"><a href=
- Ακολούθως, ο υπερσύνδεσμος εκκινεί με " (ΣΥΝΟΛΟ 47 ΧΑΡΑΚΤΗΡΕΣ) και ολοκληρώνεται με " παρομοίως
- Πριν το εκάστοτε σχετιζόμενο φιλικό λεκτικό Τίτλου/περιγραφής του υπερσυνδέσμου, διατηρείται σταθερό το μοτίβο με περιεχόμενο:
  target="_blank" title="Σχετική Εγκύκλιος"> (ΣΥΝΟΛΟ 42 ΧΑΡΑΚΤΗΡΕΣ)
- Ακολούθως, ο «Τίτλος» εκκινεί άμεσα και ολοκληρώνεται στη συνέχεια με την πρώτη αναφορά της html ετικέτας (tag)

Κώδικας:

```
#(c) Paraskevas G. Papadopoulos / Computer Systems Engineer, H.I. (MSc)
#email: pdesper@gmail.com

#ΚΡΙΣΙΜΕΣ ΠΑΡΑΔΟΧΕΣ: έχουμε αποθηκεύσει το αρχείο γλώσσας υπερκειμένου
#seminaria.html στον φάκελο «ρίζας» του τοπικού μας σκληρού δίσκου (C)
#καθώς και ότι το αρχείο πως είναι υπαρκτό και με "ζεύγη" τιμών (URL/ΤΙΤΛΟΥ)


#PROTA orizoume SYNARTHSH subtext me TESSERA (4) ORISMATA
#EVRESHS PERIEXOMENOU (EITE URL, EITE TITLOU)
#sto keimeno tis ekastote kaloumenis grammis-string (currenttextline)
#(vasei VOITHITIKHS METAVLHTHS helpproc)
#metaksy 2 motivon sta akra tis (priorareacontent KAI afterareacontent)
def subtext(currenttextline,priorareacontent,afterareacontent,helpproc):
    #STO KYRIOS PROGRAMMA PREPEI NA:
    #kaloume ti synartisi me helpproc=47 GIA EKSETASH URL,
    #               eno me helpproc=42 GIA EKSETASH TITLOU/PERIGRAFHS
    indexstart=currenttextline.find(priorareacontent)
    if indexstart==-1:
        pass
        #DEN VRETHIKE TO MOTIVO ENARKSHS STHN EKSETAZOMENH GRAMMH
    else:
        #VRETHIKE TO MOTIVO ENARKSHS STHN EKSETAZOMENH GRAMMH
        #OPOTE PREPEI STON EPISTREFOMENO DEIKTH THESHS
        #NA PROSTETHOUN OI EPIPLEON helproc THESEIS TOY MOTIVOU ENARKSHS
        #GIA NA EXOUME ETOIMH TH THESH DEIKTH ENARKSHS APEIKONISHS XARAKTHRON
        indexstart=indexstart+helpproc
        #kai SYNEXIZOUME ME TON PROSDIORISMO TOU MOTIVOU TELOUS
        #STHN EKSETAZOMENH GRAMMH
        indexend=currenttextline.find(afterareacontent,indexstart+1)
        if indexend==-1:
            #DEN VRETHIKE TO MOTIVO TELOUS STHN EKSETAZOMENH GRAMMH
            #Note: APITHANO SENARIO GIA SOSTO-PLHRES ARXEIO
            #POU PERIEXEI "ZEYGARIA" (URL-TITLOU)
            #STO PROVLEPOMENO SYMFONHTHEN MOTIVO...
            print("Βοηθητική πληροφόρηση: ΑΤΕΛΗΣ ΠΛΗΡΗΣ εύρεση πληροφορίας στη τρέχουσα γραμμή του
αρχείου-ΕΣΦΑΛΜΕΝΟ περιεχόμενο αρχείου")
        else:
            #VRETHIKE TO MOTIVO TELOUS STHN EKSETAZOMENH GRAMMH
            #OPOTE EXOUME ETOIMH TH THESH DEIKTH TELOUS APEIKONISHS XARAKTHRON
            #kalontas tin stin print apeikonizetai mexri ton xaraktira (EKEI-1)
            #PAME NA APEIKONISOUME TO PERIEXOMENO METAKSY (kai mono)
            #TON DEIKTON ARXHS KAI TELOUS pleon...
            print(currenttextline[indexstart:indexend])
            #elegxos gia paremvoli kenis grammhs (peras i oxi eksetasis zeugariou timon)
            if helpproc==42:
                print()
```

```
#KYRIOS PROGRAMMA
filepath=input("Παρακαλούμε, δώστε το ακριβές (τοπικό) μονοπάτι με το όνομα του (υπαρκτού) αρχείου υπερκειμένου (html) που θα επεξεργαστεί σχετικά το Πρόγραμμα: ")
#ANOIGMA ARXEIOU GIA ANAGNOSH, ORIZONTAS UTF8 KODIKOPOIHSH XARAKTHRON!!!
#(yparxon file, orthou periexomenoy zeygarion URL/TITLOU kai sto dothen local directory)
filerd=open(filepath,"r",encoding="utf-8")
print()

#epeksergazomaste ta stoixeia toy arxeioy GRAMMH-GRAMMH... mexri EOF=KENO alfarithmitiko readline()
line=filerd.readline()
#ean i trexousa GRAMMH den einai keni, pame na eksetasoume to periexomeno tis...
while line!='':
    #KLHSH SYNARTHSHS gia URL (xrisi escape xaraktira \ gia tous eidikous xaraktires)
    subtext(line,"ΑΘΗΝΑ</td><td style=\"font-size:11px;\"><a href=\"","\"",47)
    #KLHSH SYNARTHSHS gia TITLO-PERIGRAFH (xrisi escape xaraktira \ gia tous eidikous xaraktires)
    subtext(line,"target=\"_blank\" title=\"Σχετική Εγκύκλιος\">","</a>",42)
    #proetoimazomaste gia ton epomeno "kyklo"
    #allazoume grammi... mexri EOF=KENO alfarithmitiko readline()
    line=filerd.readline()
#oloklirosame me to arxeio - KLEISIMO ARXEIOU
filerd.close()
#TELOS PROGRAMMATOS
```

Αποτέλεσμα (π.χ. επί του δοθέντος αρχείου seminaria.html):

```
>>>
= RESTART: C:\Users\HY\Desktop\EKDDA-PP-PYTHON-2020\PP-PYTHON-EXAMPLESandVERY-USEFUL\ERGASIA 1-2hs DID
AKTIKHS ENOTHTAS (THEMATIKES 1-10)-parse.py
Παρακαλούμε, δώστε το ακριβές (τοπικό) μονοπάτι με το όνομα του (υπαρκτού) αρχείου υπερκειμένου (html)
που θα επεξεργαστεί σχετικά το Πρόγραμμα: c:\seminaria.html

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10627_2_2013.pdf
ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑΦΩΝ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10627_2_2013.pdf
ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑΦΩΝ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10394_2_2013.pdf
ΤΑ  ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ UNIX ΚΑΙ LINUX

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10315_2_2013.pdf
ΣΧΕΔΙΑΣΜΟΣ ΕΦΑΡΜΟΓΩΝ ΔΙΑΔΙΚΤΥΟΥ ΜΕ ΕΡΓΑΛΕΙΑ ΑΝΟΙΚΤΟΥ ΛΟΓΙΣΜΙΚΟΥ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10419_2_2013.pdf
ΔΙΑΧΕΙΡΙΣΗ WINDOWS SERVER 2008

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10379_2_2013.pdf
ΟΡΓΑΝΩΣΗ ΚΑΙ ΧΡΟΝΟΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΕΡΓΩΝ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10324_2_2013.pdf
ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10321_2_2013.pdf
ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ

http://www.ekdd.gr/ekdda/custom/seminars/pdf/10319_2_2013.pdf
ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ

>>>
```

ΤΕΛΟΣ ΑΝΑΠΤΥΞΗΣ

Τεκμηρίωση (Νοέμβριος 2020):
© Παρασκευάς Γ. Παπαδόπουλος
Μηχανικός Η/Υ Συστημάτων, Η.Ι. (MSc)
Email : pdesper@gmail.com