

Άσκηση

Εξοικείωση με τα εργαλεία του μαθήματος

Για τις ασκήσεις που θα πραγματοποιηθούν στο πλαίσιο του εργαστηρίου, θα χρησιμοποιήσουμε τη γλώσσα Python (έκδοση 3.7 ή μεγαλύτερη) που είναι πολύ δημοφιλής στην ανάπτυξη έξυπνων εφαρμογών με χρήση μηχανικής μάθησης και έχει αρχίσει να γίνεται η de facto γλώσσα στην επιστήμη των δεδομένων καθώς διαθέτει ένα ευρύ φάσμα πακέτων που διευκολύνουν στη δημιουργία εφαρμογών, από τις απλούστερες έως τις πιο σύνθετες.

Βήμα 1 Εγκατάσταση Anaconda

Κατεβάστε το Anaconda <https://www.anaconda.com/download/> και εγκαταστήστε το. Την προτροπή για την εγκατάσταση του VSCode μπορείτε να την απορρίψετε.

Βήμα 2 Δημιουργία λογαριασμού

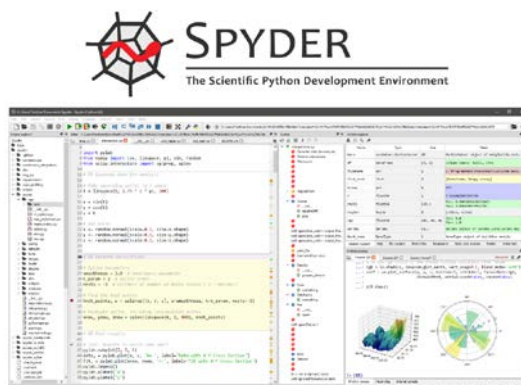
Δημιουργείτε λογαριασμό στο Anaconda cloud για να μπορέσετε να έχετε εύκολη πρόσβαση στα πακέτα.

Βήμα 3 Εξοικείωση με το Anaconda

Εξοικειωθείτε με το περιβάλλον του Navigator: [Getting started with Navigator](#) (10 λεπτά) και [video](#)

Μπορείτε εναλλακτικά (στην περίπτωση που δεν θέλετε να χρησιμοποιήσετε το Navigator) ή επιπρόσθετα, να εξοικειωθείτε με το conda: [Getting started with conda](#) (20 λεπτά). Επίσης, στο [Conda cheat sheet](#) θα βρείτε ένα γρήγορο οδηγό pdf για το conda.

Μέσα από το Navigator έχετε τη δυνατότητα να χρησιμοποιήσετε διάφορα περιβάλλοντα-διεπαφές ανάπτυξης. Ένα προτεινόμενο είναι το SPYDER.



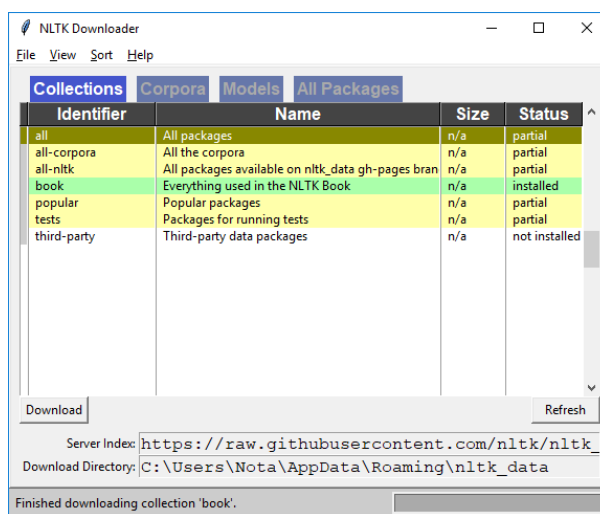
Βήμα 4 Εγκατάσταση του NLTK-εξοικείωση

Θα εγκαταστήσουμε το Natural Language Toolkit (NLTK) με το οποίο μπορούμε να δημιουργήσουμε προγράμματα για Επεξεργασία Φυσικής Γλώσσας σε Python. Πληροφορίες και εκτεταμένη βοήθεια για το συγκεκριμένο πακέτο μπορούμε να βρούμε τόσο στο <http://www.nltk.org> όσο και στο [1].

Για να εγκαταστήσουμε το NLTK package τρέχουμε τις παρακάτω εντολές

```
import nltk
nltk.download()
```

Στο παράθυρο που εμφανίζει, στην καρτέλα (tab) Collection επιλέγουμε το book



Μόλις φορτωθούν όλα τα δεδομένα από τον NLTK Downloader, ανοίγουμε ένα terminal ή SPYDER και τρέχουμε την παρακάτω εντολή

```
from nltk.book import *
```

που σημαίνει: “from NLTK's book module, load all items”. Τα αποτελέσματα φαίνονται παρακάτω:

```
In [1]: from nltk.book import *
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

Πειραματιστείτε μόνοι σας με τις εντολές `text1`, `text4`, `texts()`, `sents()`

Βήμα 5 Εύρεση κειμένου

Η εντολή `concordance`, μας δείχνει τις εμφανίσεις της ζητούμενης λέξης μαζί με κάποιο από το κείμενο μέσα στο οποίο βρίσκεται (`context`). Τρέχοντας την παρακάτω εντολή, ψάχνουμε τις εμφανίσεις της λέξης `monstrous` στο βιβλίο `Moby Dick` (`text1`).

```
text1.concordance("monstrous")
```

Τα αποτελέσματα φαίνονται παρακάτω:

```
In [8]: text1.concordance("monstrous")
Displaying 11 of 11 matches:
ong the former , one was of a most monstrous size . ... This came towards us ,
ON OF THE PSALMS . " Touching that monstrous bulk of the whale or ork we have r
ll over with a heathenish array of monstrous clubs and spears . Some were thick
d as you gazed , and wondered what monstrous cannibal and savage could ever hav
that has survived the flood ; most monstrous and most mountainous ! That Himmal
they might scout at Moby Dick as a monstrous fable , or still worse and more de
th of Radney .'" CHAPTER 55 Of the Monstrous Pictures of Whales . I shall ere l
ing Scenes . In connexion with the monstrous pictures of whales , I am strongly
ere to enter upon those still more monstrous stories of them which are to be fo
ght have been rummaged out of this monstrous cabinet there is no telling . But
of Whale - Bones ; for Whales of a monstrous size are oftentimes cast up dead u
```

Πειραματιστείτε μόνοι σας:

1. Ψάξτε για άλλες πιθανές λέξεις στο ίδιο βιβλίο
2. Ψάξτε στο βιβλίο `Sense and Sensibility` για τη λέξη `affection`
3. Ψάξτε στο βιβλίο `Genesis` πότε ή πόσο έζησαν κάποιοι άνθρωποι (π.χ. `lived`)

Αν θέλουμε σε κάποιο ή κάποια κείμενα (π.χ. `text1` και `text2`) να βρούμε άλλες λέξεις που χρησιμοποιούνται με παρόμοιο τρόπο με την λέξη `monstrous` τότε τρέχουμε την εντολή `similar`

```
text1.similar("monstrous")
text2.similar("monstrous")
```

Στα αποτελέσματα που βλέπουμε παρακάτω, παρατηρούμε ότι ο κάθε συγγραφέας χρησιμοποιεί πολύ διαφορετικά την λέξη `monstrous`.

```
In [7]: text1.similar("monstrous")
true contemptible christian abundant few part mean careful puzzled
mystifying passing curious loving wise doleful gamesome singular
delightfully perilous fearless

In [8]: text2.similar("monstrous")
very so exceedingly heartily a as good great extremely remarkably
sweet vast amazingly
```

Με παρόμοιο τρόπο μπορούμε να βρούμε τα κοινά εννοιολογικά πλαίσια (`context`) στα οποία μπορεί να βρίσκονται δύο ή περισσότερες λέξεις π.χ.

```
text2.common_contexts(["monstrous", "very"])
```

Βήμα 6 Εγκατάσταση πρόσθετων πακέτων

Απαραίτητα πακέτα που θα χρειαστούν στην πορεία των εργαστηριακών μαθημάτων είναι τα παρακάτω:

NLTK-Data: This contains the linguistic corpora that are analyzed and processed in the book.

NumPy: This is a scientific computing library with support for multidimensional arrays and linear algebra, required for certain probability, tagging, clustering, and classification tasks.

Matplotlib: This is a 2D plotting library for data visualization, and is used in some of the book's code samples that produce line graphs and bar charts.

Στην πορεία των μαθημάτων θα χρειαστεί να εγκαταστήσετε πρόσθετα πακέτα. Εάν δεν θυμάστε πως εγκαθιστούμε ένα πακέτο, γυρίστε στο [Βήμα 3](#).

Αναφορές

- [1] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing With Python*. O'REILLY. DOI:<https://doi.org/10.1088/1751-8113/44/8/085201>