



Πολυτεχνείο Κρήτης

Σχολή Μηχανικών Παραγωγής και Διοίκησης

**Ανάπτυξη λογισμικού σε Python για την
συλλογή δεδομένων από το ScienceDirect
και δημιουργία βάσεων δεδομένων με
άρθρα περιοδικών**

Απαλλακτική Εργασία

**Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης
Παναγιώτης – Χαράλαμπος Σκουλάξινος**

Καθηγητής Μαθήματος
Ματσατσίνης Νικόλαος, Καθηγητής

Χανιά, Φεβρουάριος 2022

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

Πίνακας περιεχομένων

1 Εισαγωγή.....	3
1.1 Σκοπός εργασίας.....	3
1.2 Σχετικά με το web crawling.....	3
1.3 Βιβλιοθήκες.....	3
2 Υλοποίηση.....	4
2.1 Ανάλυση στοιχείων HTML του ScienceDirect.....	4
2.2 Ο κώδικας.....	6
2.2.1 Παραμετροποίηση.....	6
2.2.2 Περιγραφή.....	7
3 Αποτελέσματα.....	10
3.1 Τρόπος λειτουργίας.....	10
3.2 Παράδειγμα – Αναζήτηση “decision support systems”.....	10

1 Εισαγωγή

1.1 Σκοπός εργασίας

Σε αυτή την εργασία, επιχειρήθηκε η υλοποίηση ενός web crawler σε γλώσσα προγραμματισμού Python, ο οποίος συλλέγει δεδομένα από την ιστοσελίδα [ScienceDirect](https://www.sciencedirect.com) του εκδότη επιστημονικών περιοδικών Elsevier. Συγκεκριμένα, σκοπός ήταν η συλλογή όλων των αποτελεσμάτων από μια συγκεκριμένη αναζήτηση στην εν λόγω ιστοσελίδα και η δημιουργία μιας βάσης δεδομένων με όλα τα συλλεγμένα δεδομένα. Τα δεδομένα που συλλέγει το λογισμικό είναι: ο τίτλος του άρθρου, οι συγγραφείς αυτού, τα ιδρύματα/φορείς στα οποία ανήκουν οι συγγραφείς, ο εκδότης, το τεύχος και ο μήνας δημοσίευσης, οι λέξεις – κλειδιά και ο σύνδεσμος που οδηγεί στη σελίδα του στο ScienceDirect. Η βάση δεδομένων με τα παραπάνω δεδομένα εξάγεται σε μορφή αρχείου Excel.

1.2 Σχετικά με το web crawling

Ένα λογισμικό ονομάζεται web crawler (ή spider, spiderbot, bot) και συχνά για λόγους συντομίας crawler, στα ελληνικά ανιχνευτής ιστού, όταν είναι υπεύθυνο για τη συλλογή δεδομένων από τις ιστοσελίδες τις οποίες είναι προγραμματισμένο να επισκέπτεται. Ένα τέτοιο πρόγραμμα ακολουθεί κάθε σύνδεσμο μέσα σε μια ιστοσελίδα (ή κάθε σύνδεσμο που ενδιαφέρει τον προγραμματιστή), καταχωρεί τα δεδομένα ενδιαφέροντος, μέχρις ότου να φτάσει σε αδιέξοδο. [1]

1.3 Βιβλιοθήκες

Για την υλοποίηση του λογισμικού χρησιμοποιήθηκαν κάποιες βιβλιοθήκες Python. Αυτές, μαζί με την έκδοση για την οποία δοκιμάστηκε η λειτουργία του, φαίνονται στον Πίνακα 1. Η έκδοση της Python που δοκιμάστηκε το λογισμικό ήταν η 3.8.10, σε λειτουργικό σύστημα Ubuntu 20.04 LTS.

Πίνακας 1: Οι βιβλιοθήκες που χρησιμοποιήθηκαν.

Όνομα βιβλιοθήκης	Χρήση
selenium 4.1.0	Φόρτωση της ιστοσελίδας και πλοήγηση σε αυτή, εξαγωγή των HTML elements με τα ζητούμενα δεδομένα.
pandas 1.0.5	Δημιουργία της βάσης δεδομένων και εξαγωγή αυτής σε μορφή αρχείου Excel.
time	Υπεύθυνη για τα απαραίτητα delays που προβλέπει το λογισμικό.
datetime	Εξαγωγή της τρέχουσας ημερομηνίας και ώρας για το όνομα του αρχείου.

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

2 Υλοποίηση

Σε αυτό το κεφάλαιο περιγράφεται ο τρόπος που υλοποιήθηκε ο εν λόγω web crawler, συγκεκριμένα για την περίπτωση του ScienceDirect. Περιγράφεται αρχικά η ανάλυση των στοιχείων HTML της ιστοσελίδας, έπειτα εξηγείται περιληπτικά ο κώδικας και τελικά παρουσιάζεται η βάση δεδομένων που εξάγεται.

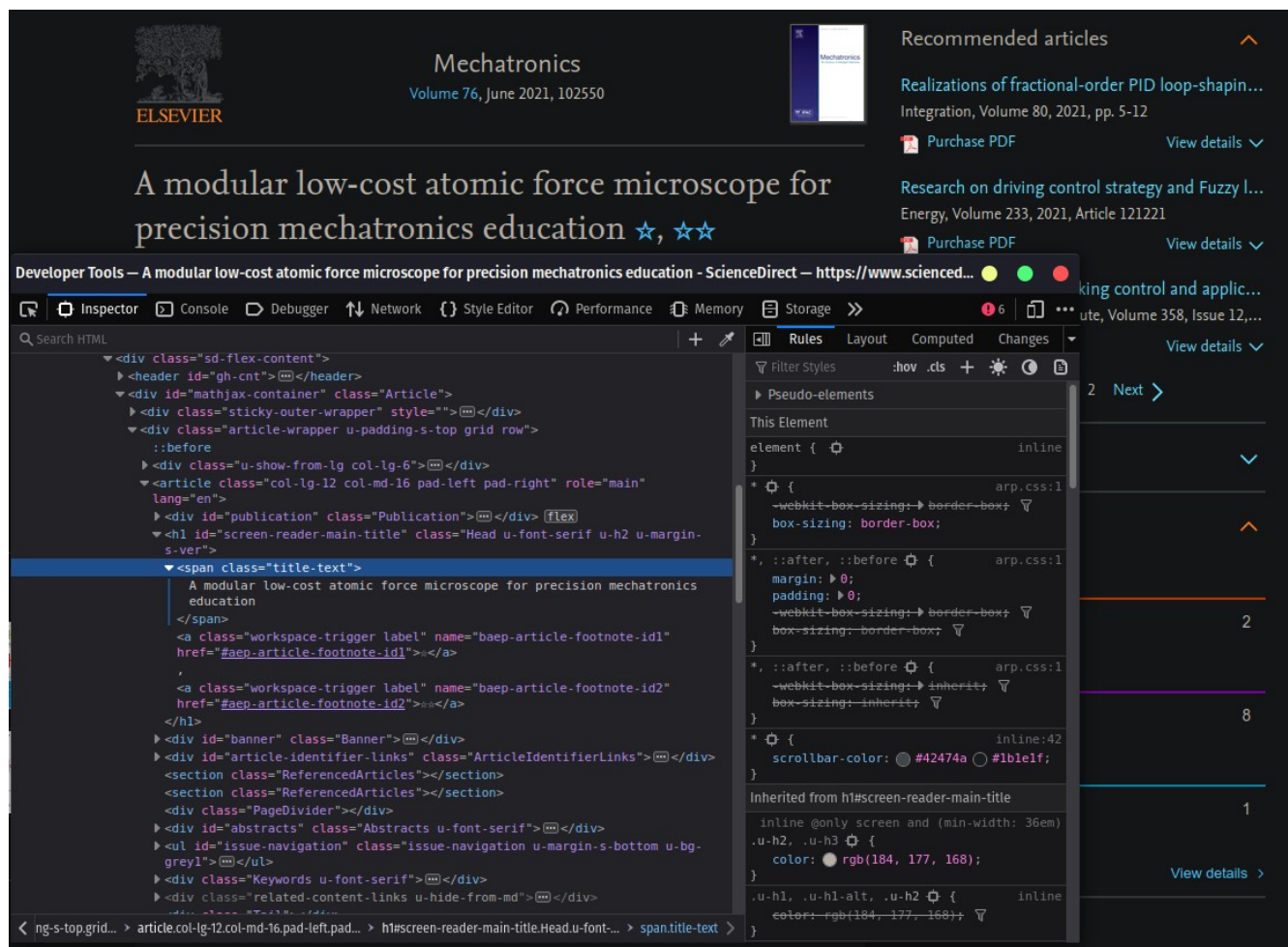
2.1 Ανάλυση στοιχείων HTML του ScienceDirect

Προκειμένου να μπορέσουμε να εξάγουμε πληροφορίες από την εν λόγω ιστοσελίδα, θα πρέπει πρώτα να ανακαλύψουμε σε ποια “δοχεία” βρίσκονται τα δεδομένα ενδιαφέροντος. Η διαδικασία που ακολουθήθηκε περιλαμβάνει την ανάγνωση του HTML κώδικα της ιστοσελίδας και τη σημείωση των ονομάτων των classes των elements που περιέχουν τα δεδομένα με χρήση της λειτουργίας inspect element όλων των μοντέρνων web browser, για τη δημιουργία του αλγορίθμου συλλογής των δεδομένων έπειτα, με χρήση της Selenium. Στον Πίνακα 2 φαίνονται τα δεδομένα που συλλέγουμε μαζί με το αντίστοιχο class με το οποίο τα βρίσκουμε στον κώδικα της σελίδας, με τη σειρά που υλοποιούνται στο λογισμικό. Στην Εικόνα 1 φαίνεται ένα παράδειγμα της παραπάνω διαδικασίας, συγκεκριμένα πώς διαβάζουμε τον τίτλο του άρθρου.

Πίνακας 2: Τα δεδομένα που συλλέγουμε και το αντίστοιχο όνομα του class στα οποία ανήκουν.

Δεδομένο προς συλλογή	Class στοιχείου HTML
Σύνδεσμος αποτελέσματος αναζήτησης	result-list-title-link
Τίτλος άρθρου	title-text
Όνομα συγγραφέα	given-name
Επώνυμο συγγραφέα	surname
Συσχετίσεις συγγραφέα (αρίθμηση)	author-ref
Συσχετίσεις συγγραφέα	affiliation
Λέξεις - κλειδιά	keyword
Τίτλος περιοδικού/δημοσιευτή	publication-title-link
Τεύχος και ημερομηνία έκδοσης	text-xs
Περίληψη και σημαντικά σημεία	Abstracts

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)



Εικόνα 1: Στιγμιότυπο με τη σελίδα ενός άρθρου και του πηγαίου κώδικα αυτής, με μαρκαρισμένο το element που αντιστοιχεί στον τίτλο του άρθρου και το class του.

2.2 Ο κώδικας

2.2.1 Παραμετροποίηση

Πριν την εκτέλεση του λογισμικού, ο χρήστης καλείται να παραμετροποιήσει κατάλληλα τον κώδικα, ώστε να εκτελείται σωστά στο σύστημά του. Οι γραμμές του κώδικα που αφορούν την παραμετροποίηση του λογισμικού επισημαίνονται με το σχόλιο “Configuration”, όπως φαίνεται στην Εικόνα 2. Συγκεκριμένα, πρέπει να ορισθεί η αλφαριθμητική μεταβλητή *browser* ίση με ‘chrome’ ή ‘firefox’, ανάλογα με τον πλοηγό της επιλογής του. Επιπλέον, πρέπει να δοθεί στην αλφαριθμητική μεταβλητή *pathToDriver* η τοποθεσία στο file system του υπολογιστή του, στην οποία βρίσκεται ο browser driver της επιλογής του. Σύμφωνα με το documentation της βιβλιοθήκης selenium, η τοποθεσία του driver πρέπει να βρίσκεται στην μεταβλητή περιβάλλοντος PATH στο λειτουργικό σύστημα του χρήστη. Σε αυτή την περίπτωση, ο χρήστης αρκεί να ορίσει το όνομα του αρχείου του driver (‘geckodriver.exe’ για τον Firefox και ‘chromedriver.exe’ για τον Chrome, αν τρέχει σε Windows ή το ίδιο χωρίς την κατάληξη ‘.exe’ αν τρέχει σε Linux). Περισσότερες πληροφορίες σχετικά με την ορθή εγκατάσταση των drivers βρίσκονται στην ιστοσελίδα https://www.selenium.dev/documentation/webdriver/getting_started/install_drivers/. [2]

Επιπλέον παραμετροποίηση που επιτρέπει το λογισμικό είναι η αλλαγή των ονομάτων των στηλών της βάσης δεδομένων, μέσω της αλφαριθμητικής λίστας *dbCols*, όπως επίσης και της δεκαδικής μεταβλητής *sleepTime*, η οποία είναι υπεύθυνη για τη μεταβολή των delays που έχει το λογισμικό. Εν δυνάμει, μείωση της τιμής αυτής της μεταβλητής παρέχει γρηγορότερη εκτέλεση του προγράμματος, όμως παρατηρήθηκε μετά από δοκιμές ότι πολύ χαμηλές τιμές έχουν ως αποτέλεσμα την παράλειψη δεδομένων, καθώς δεν προλαβαίνει να φορτώσει πλήρως η ιστοσελίδα. Τέλος, η αλφαριθμητική μεταβλητή *searchTerm* ορίζει τη λέξη – κλειδί για την αναζήτηση στο ScienceDirect. Αν περιέχει κενά, αυτά πρέπει να ορισθούν με τον χαρακτήρα ‘+’ και όχι με whitespace.

```
10  ### Configuration
11  dbCols = ['Title', 'Authors', 'Affiliations', 'Publisher', 'Date & Volume',
12          'Keywords', 'Abstract', 'Link']
13  sleepTime= 0.5
14  searchTerm = "decision+support+systems"
15  browser = 'firefox' # 'firefox' || 'chrome'
16  pathToDriver = 'geckodriver'
17  url_ = "https://www.sciencedirect.com/search?q=" + searchTerm
18  ### end Configuration
```

Εικόνα 2: Οι μεταβλητές παραμετροποίησης του λογισμικού.

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

2.2.2 Περιγραφή

Στην αρχή του κώδικα, εισάγονται όλα τα στοιχεία των βιβλιοθηκών (κεφάλαιο 1.3) που είναι απαραίτητα για την υλοποίηση του αλγορίθμου (Εικόνα 3) και έπειτα ορίζονται όλες οι μεταβλητές παραμετροποίησης (Εικόνα 2). Μετά ρυθμίζεται ο web driver ανάλογα με την επιλογή του χρήστη, ορίζεται το κύριο dataframe, ονόματι *df*, όπου θα αποθηκευτούν όλα τα δεδομένα, φορτώνεται το αρχικό url της αναζήτησης και αποθηκεύεται η τρέχουσα ημερομηνία και ώρα (timestamp) για την ονομασία του αρχείου της βάσης δεδομένων (Εικόνα 4).

```
1 from selenium.webdriver.chrome.service import Service
2 from selenium import webdriver
3 from selenium.webdriver.common.by import By
4 from selenium.common.exceptions import NoSuchElementException
5 import pandas as pd
6 import time
7 from datetime import datetime
8 import sys
```

Εικόνα 3: Import βιβλιοθηκών

```
19 df = pd.DataFrame(columns = dbCols)
20
21 browser_driver = Service(pathToDriver)
22 if browser == 'firefox':
23     driver = webdriver.Firefox(service=browser_driver)
24 elif browser == 'chrome':
25     driver = webdriver.Chrome(service=browser_driver)
26 else:
27     print("Invalid browser setting; please recheck.")
28     sys.exit(100)
29
30 url_ = "https://www.sciencedirect.com/search?q=" + searchTerm
31 driver.get(url_)
32 driver.maximize_window()
33 #driver.minimize_window()
34
35 dateTimeObj = datetime.now()
36 timestampStr = dateTimeObj.strftime("%d-%m-%Y_%H:%M:%S")
37 fileName = searchTerm + "_" + timestampStr + '.xlsx'
```

Εικόνα 4: Ρύθμιση web driver και φόρτωση σελίδας, αποθήκευση ημερομηνίας και ώρας

Μετά, το λογισμικό μπαίνει στον κύριο (ατέρμονο) βρόχο. Εκεί, δημιουργείται ένα νέο, προσωρινό dataframe στο οποίο θα αποθηκεύεται κάθε φορά η νέα γραμμή με τα νεοσύλλεκτα δεδομένα. Με τη συνάρτηση *find_elements(...)* της βιβλιοθήκης Selenium αποθηκεύουμε τα HTML elements όλων των άρθρων στη λίστα *results*, και από αυτή εξαγάγουμε τους συνδέσμους στη λίστα *links* (Εικόνα 5). Το πρόγραμμα τώρα μπαίνει στον δεύτερο βρόχο (*for*), όπου επισκέπτεται κάθε ένα *link* στη λίστα *links*. Εκεί, χρησιμοποιώντας πάλι τη *find_elements(...)*, σε συνεργασία με τη μέθοδο *By.CLASS_NAME* για να βρούμε το συγκεκριμένο class που αναζητούμε. Από εκεί και πέρα η λογική αναζήτησης επαναλαμβάνεται, ψάχνοντας τα classes του Πίνακα 2. Επειδή κάποια στοιχεία μπορεί να μην υπάρχουν σε κάποιες σελίδες, φαινόμενο το οποίο αν δεν το διαχειριστούμε θα οδηγήσει σε *unhandled exceptions* και κατά συνέπεια σε “κρασάρισμα”, τα τοποθετούμε σε ένα *try statement*, προσπαθώντας να “πιάσουμε” το *NoSuchElementException* της Selenium, το οποίο προκαλείται αν δεν

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

υπάρχει το στοιχείο που ζητήσαμε (Εικόνα 6). Σε αυτή την περίπτωση, σε αυτό το κελί τοποθετείται η φράση ‘N/A’. Στην περίπτωση των *keywords* και των *affiliations*, λόγω της φύσης των κελιών, στην πρώτη περίπτωση χωρίζουμε κάθε λέξη με κόμμα, ενώ στη δεύτερη χωρίζουμε κάθε στοιχείο με νέα γραμμή, ώστε τα δεδομένα να είναι πιο εύκολα προς ανάγνωση και επεξεργασία. Όσον αφορά τους συγγραφείς, επειδή λαμβάνουμε ξεχωριστά τα ονόματα, τα επώνυμα και τον συμβολισμό για το affiliation τους, τα συνδέουμε όλα μεταξύ τους με κατάλληλο αλγόριθμο (Εικόνα 7).

```
39 while True:
40     time.sleep(sleepTime*2)
41     df2 = pd.DataFrame()
42     results = driver.find_elements(By.CLASS_NAME, "result-list-title-link")
43     links = [result.get_attribute('href') for result in results]
```

Εικόνα 5: Ο κύριος βρόχος επανάληψης, με την αποθήκευση των συνδέσμων.

```
52     try:
53         author_ref = driver.find_elements(By.CLASS_NAME, "author-ref")
54         if author_ref == []:
55             _ref = False
56     except NoSuchElementException:
57         _ref = False
```

Εικόνα 6: Η λογική εύρεσης στοιχείων, με δοκιμή.

```
63     try:
64         keywords = driver.find_elements(By.CLASS_NAME, "keyword")
65         for keyword in keywords:
66             kw += keyword.text + ", "
67     except NoSuchElementException:
68         kw = "N/A"
69
70     try:
71         driver.find_element(By.ID, "show-more-btn").click()
72         affiliations = driver.find_elements(By.CLASS_NAME, "affiliation")
73         for aff in affiliations:
74             all_affs += aff.text + "\r"
75     except NoSuchElementException:
76         all_affs = "N/A"
77     all_affs = all_affs.strip()
78
79     if _ref:
80         for name, surname, ref in zip(author_name, author_surname, author_ref):
81             all_names += name.text + " " + surname.text + " (" + ref.text + "), "
82     else:
83         for name, surname in zip(author_name, author_surname):
84             all_names += name.text + " " + surname.text + ", "
```

Εικόνα 7: Η ένωση των αλφαριθμητικών των λέξεων - κλειδιών και των συγγραφέων.

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

Ο δεύτερος βρόχος κλείνει αποθηκεύοντας όλα τα παραπάνω δεδομένα σε μια λίστα με όνομα *thisArticle*, μετά η λίστα αυτή περνιέται στο *df2*, το οποίο με τη σειρά του γίνεται *append* στο *df*, το οποίο αποτελεί και τη βάση δεδομένων μας. Με την εντολή *driver.back()* επιστρέφουμε πίσω στη σελίδα με τα αποτελέσματα αναζήτησης, ώστε ο αλγόριθμος να προχωρήσει στο επόμενο άρθρο στη λίστα (Εικόνα 8).

```
102     thisArticle = (title, all_names, all_affs, mag, mag2, kw, abstract, link)
103     df2= pd.DataFrame(thisArticle, index = dbCols)
104     df = df.append(df2.T, ignore_index=True)
105     #print(df)
106     time.sleep(sleepTime)
107     driver.back()
108
109     time.sleep(2)
```

Εικόνα 8: Η αποθήκευση των δεδομένων που βρέθηκαν.

Μόλις τελειώσει ο δεύτερος βρόχος, τότε καταρχάς αποθηκεύονται όλα τα δεδομένα σε αρχείο Excel με μορφή ονόματος *'searchTerm_dd-mm-yyyy_hh:mm:ss.xlsx'* και γίνεται απόπειρα (με *try statement*) να πατηθεί το κουμπί *next* (το ψάχνουμε με το όνομά του, δηλαδή *By.LINK_TEXT*) και να επαναληφθεί η παραπάνω διαδικασία. Αν δε βρεθεί σημαίνει ότι ολοκληρώθηκε η διαδικασία και δεν υπάρχουν άλλες σελίδες να επισκευθούμε. Ο αλγόριθμος σε αυτό το σημείο αποθηκεύει τα δεδομένα και τερματίζει (Εικόνα 9).

```
110     try:
111         print("Writing content to excel!")
112         df.to_excel(fileName, sheet_name=searchTerm)
113         print("Done!")
114         driver.find_element(By.LINK_TEXT, "next").click()
115     except NoSuchElementException:
116         print("No more pages to crawl!")
117         break
```

Εικόνα 9: Αποθήκευση δεδομένων στο αρχείο, επόμενη σελίδα ή τερματισμός.

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

3 Αποτελέσματα

3.1 Τρόπος λειτουργίας

Εκτελώντας το λογισμικό, αυτό αρχικά αναζητά τη ζητούμενη λέξη – κλειδί στη μηχανή αναζήτησης του ScienceDirect. Κατόπιν, επισκέπτεται μία – μία τις σελίδες που επιστρέφει η αναζήτηση και συλλέγει τα δεδομένα ενδιαφέροντος. Μόλις τελειώσει με τη συλλογή δεδομένων που φαίνονται στη σελίδα, προσπαθεί να μεταβεί στην επόμενη σελίδα των αποτελεσμάτων αναζήτησης και αποθηκεύει τη βάση δεδομένων στο αρχείο Excel. Αν δεν υπάρχει επόμενη σελίδα, το λογισμικό κλείνει. Από προεπιλογή, τα αποτελέσματα που επιστρέφει ανά σελίδα το ScienceDirect είναι 25, αριθμός ο οποίος αλλάζει αν ο χρήστης τροποποιήσει το url κατάλληλα. Γενικά ο χρήστης δύναται να τροποποιήσει το url αναζήτησης ανάλογα με την προτίμησή του, όπως για παράδειγμα την επιβολή επιπλέον φίλτρων αναζήτησης.

3.2 Παράδειγμα – Αναζήτηση “decision support systems”

Χάριν παραδείγματος, το λογισμικό εκτελέστηκε με λέξη αναζήτησης “decision support systems” και σύλλεξε 325 άρθρα που βρίσκονται στο επισυναπτόμενο αρχείο. Στην Εικόνα 10 φαίνεται μέρος του αρχείου Excel με τη βάση δεδομένων που προέκυψε με την εκτέλεση για την παραπάνω αναζήτηση.

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022

Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)

Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

	A	B	C	D	E	F	G	H	I	J	K	L
1		Title	Authors	Affiliations	Publisher	Date & Volume	Keywords	Abstract	Link			
2	0	Application, adoption and opportunities of a decision support system for agricultural water	Iffat Ara (a), Lydia Tur	Queensland A	Agricultural Water	Volume 257, 1 November	Irrigation, Software	Download : Download	https://www.sciencedirect.com/sci			
3	1	A collaborative decision support system for sustainable urban drainage	Mona Jabbari (a), Shay	Department of	Decision Support S	Volume 153, February 2	Automatic cluster	Automatic clu	https://www.sciencedirect.com/sci			
4	2	A decision support system for sustainable urban drainage	Mohsen Tayebi (a), Al	Department of	Journal of Natural	Volume 99, March 2022	Air pollution contr	The environm	https://www.sciencedirect.com/sci			
5	3	Sustainable Urban Drainage System	Pascual Ferrans (a), M	Centro de Inve	Science of The Tot	Volume 806, Part 2, 1 F	Urban drainage m	Download : Download	https://www.sciencedirect.com/sci			
6	4	The internet of things-based decision support system for intelligent decision making	Yuan Guo (a), Nan W	School of Mech	Mechanical System	Volume 142, August 20	Intelligent decisio	To comprehen	https://www.sciencedirect.com/sci			
7	5	Model of the geospatial situation of decision support system for urban drainage	Ireneusz Miciula (a), P	University of S	Procedia Comput	Volume 176, 2020, Page	Decision support	The article pre	https://www.sciencedirect.com/sci			
8	6	Decision support system in management of warship combat	Anna Sobotka, Joanna	AGH University	Automation in Con	Volume 128, August 20	Decision support	Processes of o	https://www.sciencedirect.com/sci			
9	7	Warship Combat Decision Support System	Hong Wen Lin, Jian Ch	Dalian Naval A	Procedia Comput	Volume 166, 2020, Page	data mining, deci	Taking the mo	https://www.sciencedirect.com/sci			
10	8	Decision support for medical disaster response	Fabio Cibella (a), Simo	National Resea	International Journ	Volume 50, November 2	Validation of Deci	In medical dis	https://www.sciencedirect.com/sci			
11	9	A bibliometric analysis and systematic review of decision support systems for maritime transport	Mateusz Gil (a), Krzys	Dalhousie Uni	Safety Science	Volume 128, August 20	AHPAnalytic Hiera	Maritime trans	https://www.sciencedirect.com/sci			
12	10	Decision support system based planning for renewable energy	Dipal Baruah (a), D.C.	Department of	Renewable Energy	Volume 38, September 2021, Pages 22-35	Download : Download	https://www.sciencedirect.com/sci				
13	11	A multi-criteria evolutionary-based algorithm for nitrogen fertilizer management	Benyamin Khoshnevis	Key Laboratory	Journal of Cleaner	Volume 256, 20 May 20	Nitrogen fertilizer	Nitrogen misr	https://www.sciencedirect.com/sci			
14	12	A fuzzy decision support system for falls and delirium in older inpatients	Ihsan Erozan,	Dumlupinar U	Journal of Manufac	Volume 52, Part A, July	Critical componen	Management	https://www.sciencedirect.com/sci			
15	13	Falls and delirium in older inpatients	Birgit A. Damoiseaux	Amsterdam U	Safety Science	Volume 142, October 20	Fall, Delirium, Ho	In our study,	https://www.sciencedirect.com/sci			
16	14	How to elicit and cease herding behaviour in a decision support system for emergency response	Boukje Compen (a), F	UNU-MERIT, M	Decision Support S	Volume 152, January 20	Behavioural econ	Behavioural e	https://www.sciencedirect.com/sci			
17	15	A new emergency decision support system for rapid response	Audrey Fertier (a), An	Laboratoire de	Decision Support S	Volume 133, June 2020	Information syste	This paper stu	https://www.sciencedirect.com/sci			
18	16	A Decision Support System for rapid response	Stefanos Doltsinis (a),	Intelligent Aut	Computers in Indu	Volume 116, April 2020	Cyber-Physical Sy	Production ran	https://www.sciencedirect.com/sci			
19	17	Measures of success of computerized clinical decision support systems	Mengting Ji (a), Guang	Ren Ji Hospital	Health Policy and	Volume 10, Issue 1, Mar	Clinical decision	The findings o	https://www.sciencedirect.com/sci			
20	18	A decision support system for territorial decision making	Vanessa Assumma (a),	Centre of Math	Science of The Tot	Volume 756, 20 Februar	Territorial resilie	Download : Download	https://www.sciencedirect.com/sci			
21	19	A decision support system to design resilient water distribution networks	Cristóbal Ramos-Salg	Organization B	Reliability Enginee	Volume 216, December	Water distribution	Asset manage	https://www.sciencedirect.com/sci			
22	20	Clinical decision support systems for chronic disease management	Leonice Souza-Pereira	Computer Scie	Computer Method	Volume 195, October 20	Chronic disease,	A Clinical Deci	https://www.sciencedirect.com/sci			
23	21	Capital shortfall: A multicriteria decision support system for financial stability	Minas-Polyvios Tsagka	Montpellier Bu	Decision Support S	Volume 145, June 2021	Financial stability	Following the	https://www.sciencedirect.com/sci			
24	22	Decision support systems for sustainable development	Maximilian Zarte (a), A	Faculdade de	Journal of Cleaner	Volume 219, 10 May 20	Decision support	Achieving sus	https://www.sciencedirect.com/sci			
25	23	COVID-19 pandemic decision support system	Costas A. Varotsos (a),	Department of	Safety Science	Volume 142, October 20	COVID-19, Decisio	The year 2020	https://www.sciencedirect.com/sci			
26	24	A deep reinforcement learning based decision support system for textile manufacturing	Zhenglei He (a), Kim-P	National Local	Computers in Indu	Volume 125, February 2	Deep reinforcement	Textile manuf	https://www.sciencedirect.com/sci			
27	25	Do computerized clinical decision support systems improve patient safety?	Anne-Laure Sennesae	Université cath	Thrombosis Resea	Volume 187, March 202	Clinical decision	CDSS might p	https://www.sciencedirect.com/sci			
28	26	Declarative and Mathematical Programming for Food Recycling Decision Support	Federico Chesani (a),	Dipartimento	Engineering Applic	Volume 95, October 202	Food recycling, D	a Prolog syste	https://www.sciencedirect.com/sci			
29	27	Effects of using the MyFood decision support system on food waste reduction	Mari Mohn Paulsen (a),	Department of	Clinical Nutrition	Volume 39, Issue 12, De	Disease-related m	The interventi	https://www.sciencedirect.com/sci			
30	28	Interpretable confidence measures for decision support systems in machine learning	Jasper van der Waa (a),	Technical Univ	International Journ	Volume 144, December	Machine learning	Decision supp	https://www.sciencedirect.com/sci			
31	29	RV-DSS: Towards a resilience and vulnerability assessment of decision support systems	Donya Hajializadeh (a),	School of Engi	Computers & Indu	Volume 156, June 2021	Systems interdep	This study pre	https://www.sciencedirect.com/sci			
32	30	Sustainability assessment of sugarcane production	Katia Regina Evaristo	Federal Univer	Environmental Dev	Volume 32, December 2	Agricultural produ	The concept o	https://www.sciencedirect.com/sci			
33	31	The framework of an agricultural land decision support system	Hongqing Li, Yaoyang	Department of	Science of The Tot	Volume 717, 15 May 20	Decision support	Agricultural la	https://www.sciencedirect.com/sci			
34	32	A decision support system for grain storage	Elham Mardaneh (a),	Global Grain	Knowledge-Based	Volume 223, 8 July 202	90B90, 90B50, O	This paper pre	https://www.sciencedirect.com/sci			
35	33	A decision support system for waste management	Yang Luo (a), John An	Green Data Ce	Applied Energy	Volume 250, 15 Septem	Data centre, Ener	Data centre se	https://www.sciencedirect.com/sci			
36	34	A real-time fuzzy decision support system for crop production	Maona Li (a), Ruixiu S	Crop Productio	Computers and Ele	Volume 163, August 20	Decision support	An irrigation o	https://www.sciencedirect.com/sci			
37	35	Missing care: A framework to address decision support systems in electronic health records	Saeed Piri,	Department of	Decision Support S	Volume 136, September	Electronic health	In recent deca	https://www.sciencedirect.com/sci			
38	36	Intelligent embedded system for decision support in nondestructive measurement	Manoel M. Silva (a), E	Nondestructive	Measurement	Volume 185, November	Pulsed eddy curre	Download : Download	https://www.sciencedirect.com/sci			
39	37	Development of a decision support system for big data analytics	Abaid Ullah (a), Salm	Department of	Science of The Tot	Volume 731, 20 August	Decision support	Download : Download	https://www.sciencedirect.com/sci			
40	38	A real-time Decision Support System for big data analytics	Ines Sbair (a), Saousse	Université de	Procedia Comput	Volume 176, 2020, Page	Big data analytic	Recently, the	https://www.sciencedirect.com/sci			
41	39	A decision support system for home healthcare	Ahmet Kandakoglu (a),	Division of Nep	Decision Support S	Volume 130, March 202	Healthcare, Home	Working toget	https://www.sciencedirect.com/sci			

Εικόνα 10: Η βάση δεδομένων που εξάγει το λογισμικό, μετά την εκτέλεσή του με λέξη αναζήτησης "decision support systems".[3]

“Ανάπτυξη λογισμικού σε Python για την συλλογή δεδομένων από το ScienceDirect και δημιουργία βάσεων δεδομένων με άρθρα περιοδικών” – Φεβρουάριος 2022
Χρήστος – Κωνσταντίνος Μπιρμπουτσάκης (2017010036)
Παναγιώτης – Χαράλαμπος Σκουλάξινος (2017010035)

Βιβλιογραφία

- [1] Web crawler. (2022). Wikipedia. https://en.wikipedia.org/wiki/Web_crawler
- [2] Selenium documentation. (2022). Selenium. <https://www.selenium.dev/documentation/>
- [3] ScienceDirect search for "decision support systems". (2022). ScienceDirect. <https://www.sciencedirect.com/search?q=decision+support+systems>