

Thonny - C:\python\scraper.py @ 14:1

File Edit View Run Tools Help

scraper.py x

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 url = "https://news.google.com"
5
6 response = requests.get(url)
7 soup = BeautifulSoup(response.text, "html.parser")
8
9 titles = soup.find_all("h3")
10
11 print("\nΒρέθηκαν τίτλοι:\n")
12 for i, t in enumerate(titles, start=1):
13     print(f"{i}. {t.get_text(strip=True)}")
14
```

Γραμμή εντολών

```
C:\python>python scraper.py

Βρέθηκαν τίτλοι:

1. Top stories
2. Local news
3. Picks for you

C:\python>
```

Thonny - C:\python\Web Title Extractor.py @ 13:1

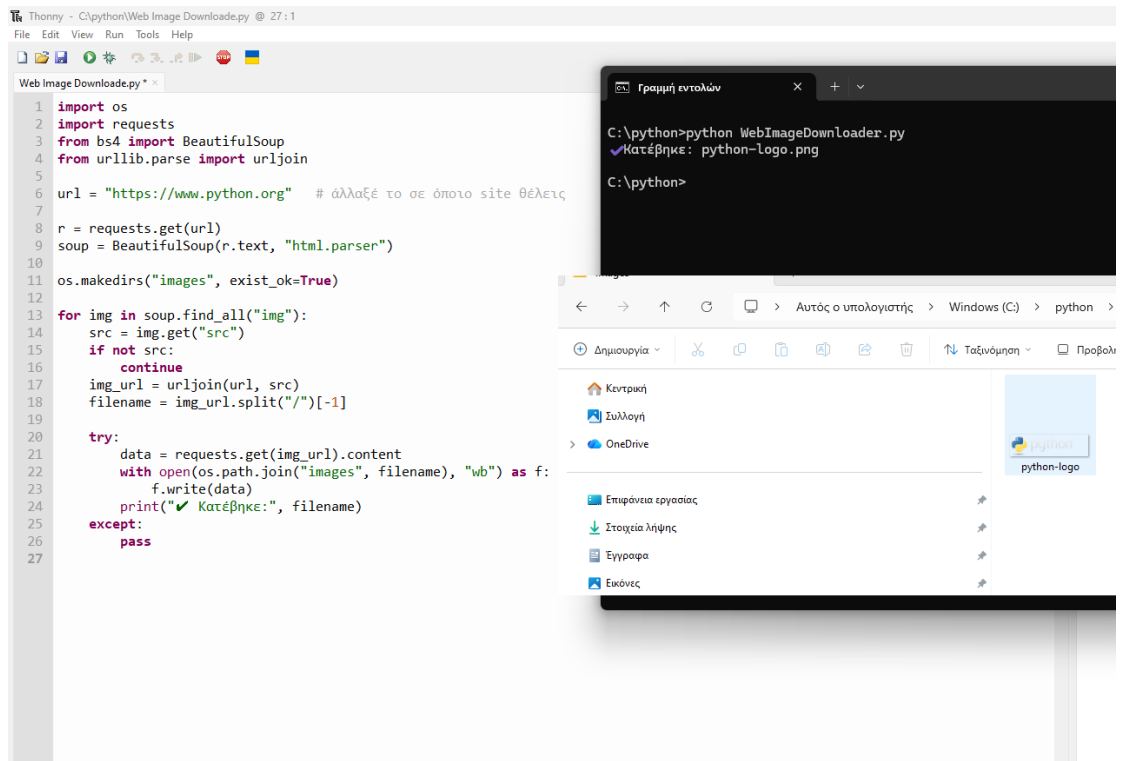
File Edit View Run Tools Help

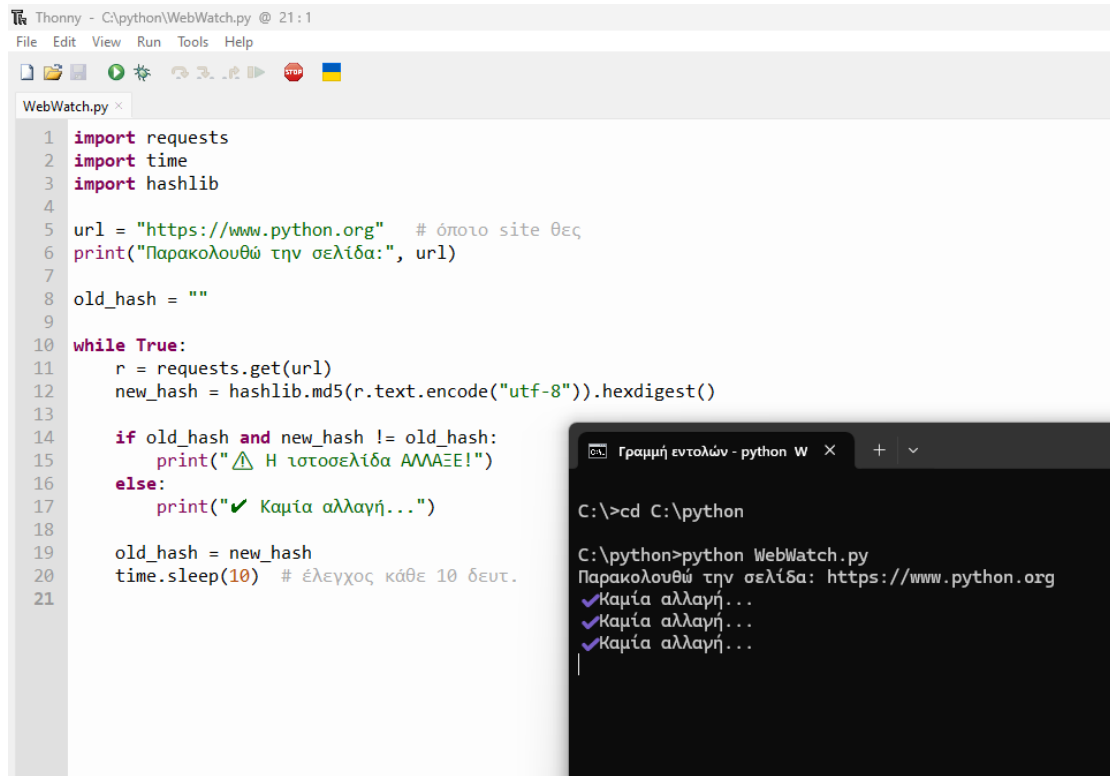
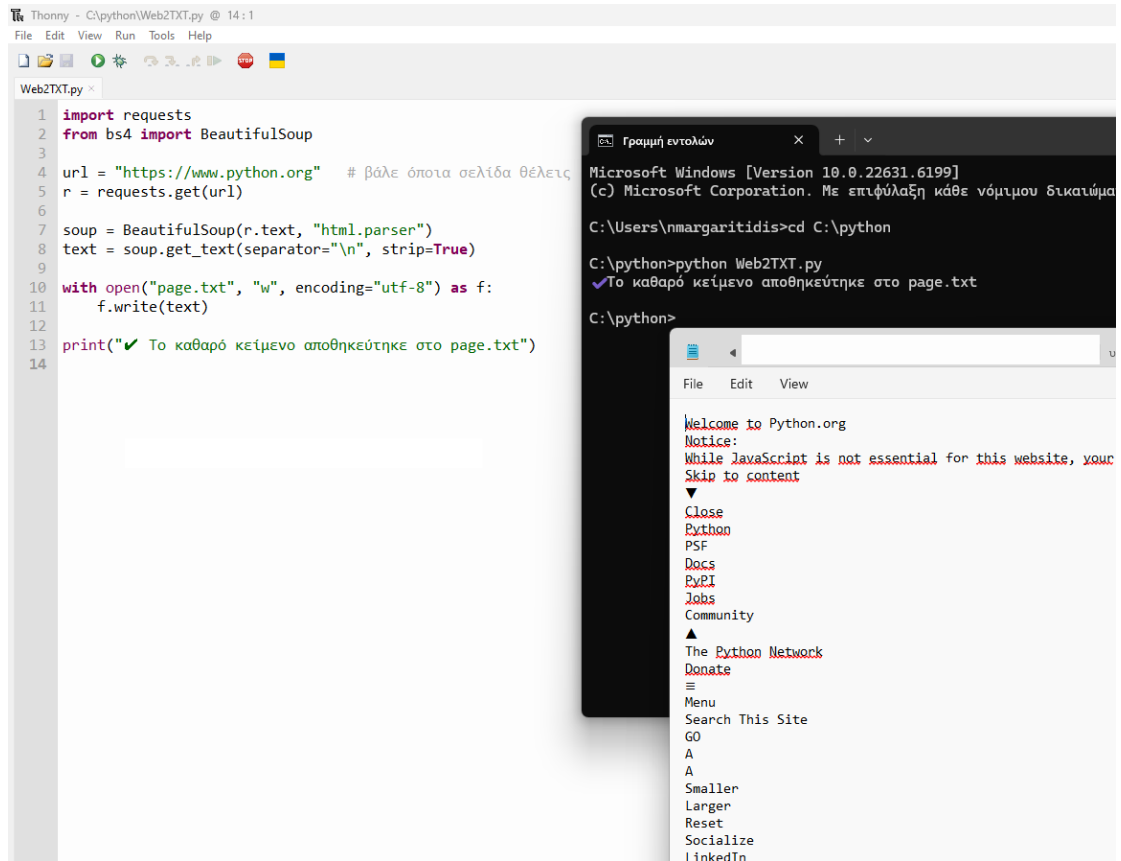
scraper.py x Web Title Extractor.py x

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 url = "https://news.google.com" # άλλαξε το σε όποιο site θέλεις
5
6 r = requests.get(url)
7 soup = BeautifulSoup(r.text, "html.parser")
8
9 titles = soup.find_all(["h1", "h2", "h3"])
10
11 for t in titles:
12     print("•", t.get_text(strip=True))
13
```

Windows PowerShell

```
PS C:\python> python "Web Title Extractor.py"
• Your briefing
• Top stories
• Local news
• Picks for you
PS C:\python> S|
```





ΑΝΑΛΥΣΗ ΕΝΤΟΛΩΝ ΚΑΙ ΠΡΟΓΡΑΜΜΑΤΩΝ PYTHON WEB SCRAPING

1. ΑΝΑΛΥΣΗ ΕΝΤΟΛΩΝ CMD / POWERSHELL

`cd C:` Αλλάζει φάκελο. Μετακινεί το CMD στον φάκελο C:.

`python scraper.py` Εκτελεί ένα αρχείο Python με όνομα `scraper.py`.

`python --version` Εμφανίζει την έκδοση της Python.

`pip --version` Εμφανίζει την έκδοση του pip.

`pip install requests beautifulsoup4` Εγκαθιστά τις βιβλιοθήκες `requests` και `beautifulsoup4`.

`python -m pip install --upgrade pip` Αναβαθμίζει το pip στην τελευταία έκδοση.

`where python` Εντοπίζει το `python.exe` στο σύστημα.

`setx PATH ...` Προσθέτει διαδρομή στο PATH των Windows.

`start %USERPROFILE%` Ανοίγει τον φάκελο `AppData`, ακόμη κι αν είναι κρυφός.

2. ΑΝΑΛΥΣΗ ΕΝΤΟΛΩΝ PYTHON ΜΕΣΑ ΣΤΑ ΠΡΟΓΡΑΜΜΑΤΑ

`import requests` Φορτώνει τη βιβλιοθήκη για HTTP αιτήματα.

`import bs4 / BeautifulSoup` Φορτώνει τη βιβλιοθήκη για διάβασμα HTML.

`requests.get(url)` Κατεβάζει το HTML της σελίδας.

`BeautifulSoup(r.text, "html.parser")` Μετατρέπει το HTML σε αντικείμενο για αναζήτηση.

`soup.find_all(...)` Βρίσκει όλα τα tags συγκεκριμένου τύπου (h1, h2, img, links κλπ).

`urljoin(url, src)` Μετατρέπει relative URLs σε absolute URLs.

`with open(..., "wb")` Αποθηκεύει εικόνες ή binary δεδομένα.

`soup.get_text(...)` Εξάγει μόνο το καθαρό κείμενο μιας ιστοσελίδας.

`hashlib.md5(...)` Υπολογίζει hash για ανίχνευση αλλαγών σελίδας.

`time.sleep(10)` Περιμένει 10 δευτερόλεπτα πριν τον επόμενο έλεγχο.

3. ΑΝΑΛΥΣΗ ΤΩΝ 5 ΠΡΟΓΡΑΜΜΑΤΩΝ WEB SCRAPING

1) Web Title Extractor — Mini Scraper 10 γραμμών

- Παίρνει όλους τους τίτλους (h1, h2, h3) από ιστοσελίδα.
- Δείχνει τη δομή της σελίδας.

2) Web Image Downloader

- Κατεβάζει όλες τις εικόνες από site.

- Δημιουργεί φάκελο images και αποθηκεύει μέσα τα αρχεία.

3) Web Link Extractor

- Σκανάρει όλα τα links της σελίδας.
- Τα μετατρέπει σε πλήρη URLs.
- Χρήσιμο για crawlers.

4) Web2TXT

- Μετατρέπει ιστοσελίδα σε καθαρό TXT.
- Αφαιρεί HTML, scripts, menus και κρατά μόνο κείμενο.

5) WebWatch — Website Change Detector

- Ανιχνεύει αλλαγές σε ιστοσελίδα συγκρίνοντας hashes.
- Ιδανικό για monitoring ανακοινώσεων.

ΤΕΛΙΚΗ ΕΙΚΟΝΑ

Τα 5 προγράμματα καλύπτουν: - scraping δεδομένων - κατέβασμα εικόνων - κατέβασμα links - μετατροπή σε κείμενο - monitoring αλλαγών

Αυτή είναι μια πλήρης μίνι εργαλειοθήκη Web Scraping σε Python.