

# PODSTAWY PROGRAMOWANIA W PYTHON

Dzień 14



# AGENDA

## DAY 14

- egzamin
- praca z excel
- webscrapping

# | Praca z plikami excel

# OpenPyXL

<http://openpyxl.readthedocs.io/en/default/index.html>

```
pip install openpyxl  
import openpyxl
```

```
plik = openpyxl.load_workbook("example.xlsx")  
arkusze = plik.get_sheet_names()  
arkusz = plik.get_sheet_by_name('Owoce')
```

# | webscrapping

# HTML, CSS. JS

HTML – znacznikowy język tworzący strukturę stron internetowych

CSS – arkusz stylów – odpowiada za wygląd strony

JavaScript (nie mylić z Java) – język kliencki (wykonywany przez przeglądarkę) – odpowiada za animację, wykonywanie skryptów

<https://www.w3schools.com/>

# Developer tools

## F12

Narzędzia do inspekcji stron internetowych  
F12

# beautiful soup 4

## bs4

- pakiet do parsowania i wyszukiwania elementów html/xml

- requests

moduł do wysyłania żądań do serwerów www

```
response = requests.get('trojmiasto.pl')
```



# bs4

<code>soup.select('#author')</code>	element z id = author
<code>soup.select('.notice')</code>	element o klasie = notice
<code>soup.select('div span')</code>	elementy <span> wewnątrz elementu <div>
<code>soup.select('div &gt; span')</code>	elementy <span> bezpośrednio wewnątrz elementu <div>, żaden inny element nie może wystąpić pomiędzy
<code>soup.select('input[name]')</code>	elementy <input>, które mają atrybut name z dowolną wartością
<code>soup.select('input[type="buton"]')</code>	elementy <input>, które mają atrybut type z wartością "buton"

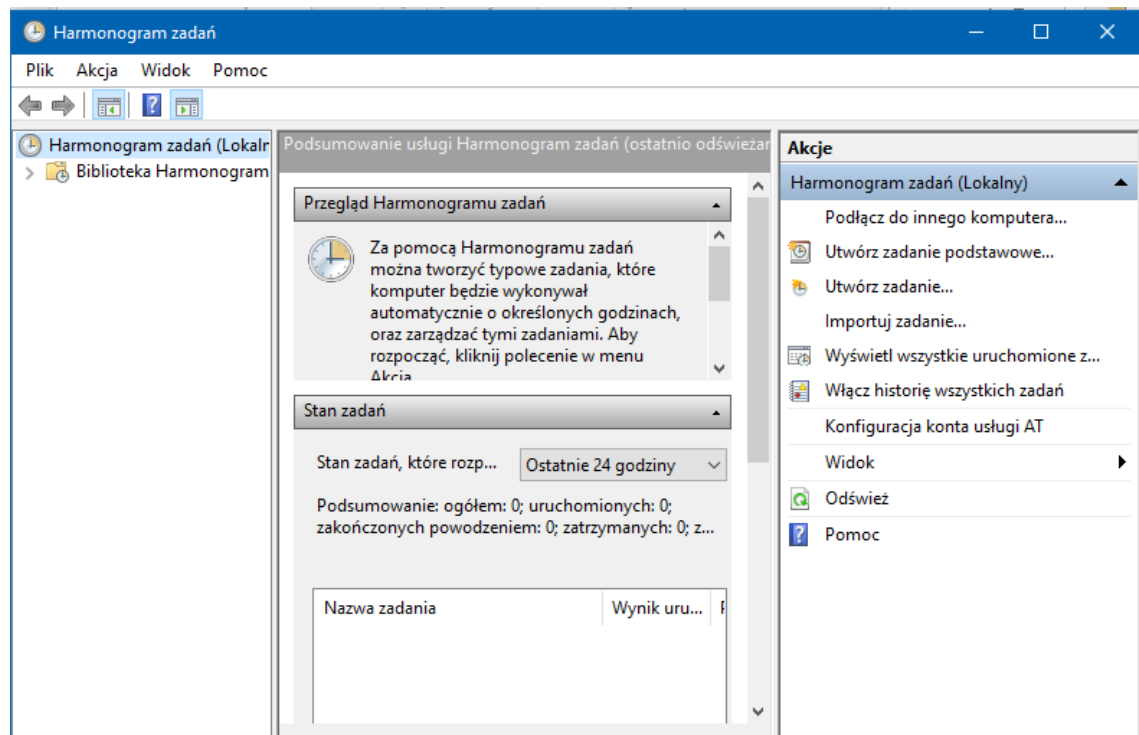
`soup.select('p #author')`

# TASK SCHEDULING (WIN)

## CRON (Linux, osX)

## HARMONOGRAM ZADAŃ

Pozwala na  
wykonywanie poleceń  
w określonym czasie





# Thanks!!