Python w BigData

**dr Tomasz Dwojak**

[t.dwojak@amu.edu.pl](mailto:t.dwojak@amu.edu.pl)

Tytuł rozprawy: Efficient Algorithms for Hybrid Neural Machine Translation

**Materiały do zajęć:**

<https://github.com/tomekd/python>

Zadanie domowe do 9.12 (dzień przed zajęciami)

Narzędzia:

Pycharm – narzędzie do programowania w pythonie. Community (darmowe) – nam wystarczy. Professional – roczna licencja edukacyjna.

Anaconda – dystrybucja pythona, menadżer pythonowych środowisk

JupyterLab – jedno z narzędzi w ramach Anacondy dostępne

**Dlaczego Python**:

Obok R najpopularniejszy język programowania danych

Nacisk na czytelność kodu

Pliki źródłowe to: \*.py

**Popularne biblioteki:**

pandas – analiza danych

numpy - obliczenia

matplotlib, plotly, bokeh – wizualizacja

scikit, pytorch,tensorflow – uczenie maszynowe

Python 2.7 już nie jest wspierany (od 2020) – a nadal można trafić na materiały jego dotyczące (w tym szkoleniowe)

## Rozpoczęcie pracy w Pycharm

Get form VSC i wkelić link do github

.idea – folder przechowuje Workspace z Pycharma

Drzewo katalogów: …

Po napisaniu kodu pythona do uruchomienia musimy uruchomić go poprzez interpreter, który wczytuje kolejne instrukcje począwszy od pierwszej linijki.

Wybór interpretera: …

Uruchomienie kodu

Uruchomienie linijki kodu …

Git -> git pull – do pobrania zmian w poniedziałek

Podstawy pythona

Zmienne

* Nazwy muszą być unikatowe
* Wielkość liter ma znaczenie
* Nie trzeba określać typu

Funkcja print

Typy liczbowe:

* Liczby całkowite: int
* Liczby rzeczywiste (zmiennoprzecinkowe): float

Konwersja typów:

int() -konwersja na typ całkowity, liczbę zmiennoprzecinkową obetnie z części dziesiętnej (nie zaokrągla matematycznie)

float() – jest w stanie wyciągnąć liczbę z tekstu reprezentującego liczbę np. x=‘3.14’ -> float(x) -> 3.14

Jeśli wykonamy mnożenie zmiennej tekstowej, to zostanie ona powtórzona określoną liczbę danych

**Wartości logiczne – typ bool:**

* True/False
* Przy konwersji na typ bool – wszystkie liczby różne od 0 mają wartość True,
* Do negacji służy ‘not’

**Operatory porównania:**

* ==
* !=
* >
* <
* >=
* <=

Warunki logiczne

* and
* or

Brak wartości lub pusta: **None - typ NoneType** odpowiedniki w R – **NA,** w SQL – **Null**

**Komentarze**

**#** - jednolinijkowy – linijka zahashowana

**”””** wielolinijkowe oznaczone potrójnym podwójnym cudzysłowem **”””**

**Zmienne tekstowe - string**

Nie ma znaczenia czy w ‘ ‘ czy w ” ”

Do konwersji funkcja **str()**

**Specjalne:**

\’

\”

\n

\t

big\_no = ‘N’ + ‘O’ \* 8 + ‘!’  
print (big\_no) -> NOOOOOOOO!

Operacje na stringach:

len() – zwraca długość łańcucha znaków

lower() – zmienia na małe litery

upper() – zmienia na duże litery

replace() – zamiana liter

strip() – usuwa białe znaki (z początku/końca)

startswith() – sprawdzenie czy dany ciąg rozpoczyna się od określonych liter

Funkcje mogą mieć w sobie wbudowane dodatkowe funkcje – metody np.

Print(user.replace(‘T’, ‘R’))

f-stingi

Nawiasy {} w te nawiasy wstawiamy nazwy zmiennych, funkcji

print(f"Nazywam się {firstname} {surname}.")

print(f"Nazywam się {firstname} {surname}.")

typy złożone

listy(list)

w nawiasach []

list\_0\_9 = list(range(10)) -> zwraca listę [0, 1, … 8, 9]

len() – natypie list zwróci długość listy (liczbę elementów listy)

- w pythonie wiele zaczyna się od 0 – np. indeksowanie

Indeksowanie:

Nazwa zmiennej typu list i w [] indeks elementu – pierwszy element ma indeks 0

[0] – pierwszy element na liście

[:5] – 5 pierwszych

[-5:] – ostatnie pięć

[1:5] – od drugiego do piątego – nieintuicyjne, dlaczego nie do szóstego?

Jest też ujemne indeksowanie [-1] – ostatni element na liście, [-2] – przedostatni element

websites.append('bing')

extend – łączy listy

pop – usuwa ostatni element z listy, a jeśli będzie nr indeksu to ten konkretny element

remove – usuwa pierwsze wystąpienie danej wartości

sort – sortuje listę

count – zlicza wystąpienia elementu

index – zwraca indeks tego elementu (pierwsze jego wystąpienie)

**all**

**any**

**Krotki(tuple)**

**Słowniki (Dict)**

**Przechowuje dane jako para: klucz – wartość**

**keys() – zwraca klucze**

**values() – zwraca wartości**

**items() – zwraca krotki – pary, klucz-wartość**

## zen of python

Beautiful is better than ugly.

Explicit is better than implicit.

Simple is better than complex.

Complex is better than complicated.

Flat is better than nested.

Sparse is better than dense.

Readability counts.

Special cases aren't special enough to break the rules.

Although practicality beats purity.

Errors should never pass silently.

Unless explicitly silenced.

In the face of ambiguity, refuse the temptation to guess.

There should be one-- and preferably only one --obvious way to do it.

Although that way may not be obvious at first unless you're Dutch.

Now is better than never.

Although never is often better than \*right\* now.

If the implementation is hard to explain, it's a bad idea.

If the implementation is easy to explain, it may be a good idea.

Namespaces are one honking great idea -- let's do more of those!

Pandas z Internetu

This python pandas tutorial is taken from the [https://lectures.quantecon.org/py/](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbnNxcVFjUWtJSWs2dURRekhUblRMcFVOdUkxZ3xBQ3Jtc0ttR3dhejh0cktJTjR3X3JRbW9nTFRkaTQwaU9uQ2JpbE93Z2pWYlBSc1RSLTRLWFBIazBQcWpEcm4tOEJndzM2RFM2NU5wT1VZbnk0ZWE3Zmx4bTZRaXg1NG04X1JJbm4xRVdaMDVTektLUzg4QktFUQ&q=https%3A%2F%2Flectures.quantecon.org%2Fpy%2F&v=iGFdh6_FePU) pandas tutorial. Pandas is an excellent python module. It is designed for cleaning and analysing data. In this tutorial I show you how to use it. 3 Data Science Learning Platforms I would recommend 1. Data Quest - [https://bit.ly/3hcLjbY](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbmtKZzZ3dmxHZTZmSDhlbk1GNzRibW9mYk9wd3xBQ3Jtc0ttVUxHaWpSVGxDd2hqOFZfdDBpWmdfVUdnZFhQMnplaXA3ZFlraHZhUV83WDNnLTZfWlA2RUE5X3hlSEJkODgxQXNjZ0g4OEZlUVdqakpDdEktNWJzbUpJVmFjVTBubWxULXNQcEU5b3V3dzhOR0VCMA&q=https%3A%2F%2Fbit.ly%2F3hcLjbY&v=iGFdh6_FePU) (my favourite) 2 Data Camp - [https://bit.ly/39erDk8](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbkNsWE5vMVVjY3dpa29CQlVFYmYxa1N2SnZBUXxBQ3Jtc0tuWVl3Ni1nVWtqeGsyMzF1ZThCMC1TWUhfcDBUSkx1U2dVanhScV9LdWlMOFIxMlB6cUdPMzgxNVE5MFJhUW0xaktJX2hqaTZlYlAzdkIwRm9haHZmU3ZMLTV6eDBWNG9pWWY4TUFzTGd3ODJCakZhZw&q=https%3A%2F%2Fbit.ly%2F39erDk8&v=iGFdh6_FePU) 3 365 Data Science - [https://bit.ly/3hbUHdd](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbUZhSVBDVnpCdUFSam5MRkRCc0ZITGpQZFMtQXxBQ3Jtc0tuTDZNc3FaMWRUdVJqNkEwUXRIYmFITVA1RUViNUpYREtNV2JLRmtxNzEzdEdKVmFuN2xvU0lucHNqeGdZMVduUUpLX1JDWkhNS0VjcHRkNWxfOUNQUUdjcmxOUGxlMW90dWZ3ZzFZOEx1UThQM2E0aw&q=https%3A%2F%2Fbit.ly%2F3hbUHdd&v=iGFdh6_FePU) 2 Recommended Python Courses 1. Exploratory Data Analysis with Python and Pandas - [https://bit.ly/2QXMpxJ](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa0dRb3RueXNfNEpiVFF5aWhzNFlJUXh2aDNTd3xBQ3Jtc0ttS29UY0FNakpWV3M0QXNZZktweHliRldpRmx0YXlnQUNFeVhwUUJaWFo5TERnTEhOOXVxamNUQTNVaEw0VUZTNlpTLThMdlM0Sk4wbjBpbUZIZVJuNEdvVGVwQmRldV9FTTlldEVtMUdLY0FwYklKSQ&q=https%3A%2F%2Fbit.ly%2F2QXMpxJ&v=iGFdh6_FePU) 2. Complete Python Programmer Bootcamp - [http://bit.ly/2OwUA09](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa3BJUzl2S3J6Q3RHS3VZUWZ5d2NONktMbUFjQXxBQ3Jtc0ttYkRXbHVSM1ZfVjF3cHVCSThkN0JBMUZfZUlYU1Ryd2NPMHdDUW43eWM3R2tfQXQ0bGtOQnAydk9XQzFlVFBuNVo5cWEtNkEtNi1Eemg3RmJtSlJHZDN6b2hzUXlmOVF4WHp5Njl1MTJ2aGxTTmh6bw&q=http%3A%2F%2Fbit.ly%2F2OwUA09&v=iGFdh6_FePU) Data Science Interview Preparation StrataScratch [https://bit.ly/30ul0nX](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbWF4cFNQaFhnUE9PeElFY2ZKQXZxbnNkeUdTQXxBQ3Jtc0tsM2lYNlUwSldRckI2cTQyTE0wdE9QbFphVHlTSGJwblMwRDBlX1luOGU3bkF0M3p3QmtRai1FdTdFRXg4U09vMTN0ckFYNi1hcHpnaURnLWlVN3ZiTXFzSzBhZVhaZXRNanVaeTNsd3VQM2RhRnR0VQ&q=https%3A%2F%2Fbit.ly%2F30ul0nX&v=iGFdh6_FePU) (These contain affiliate links, which means I receive a percentage of any sales made. There is no additional cost for anybody clicking on them) For the links from the tutorial go to my facebook page.