

# DOKUMENTACJA TECHNICZNA PROJEKTU „PROCES ETL NA PRZYKŁADZIE SERWISU PYSZNE.PL”

## 1. Użyte technologie

- Python v.3.7.6
- MySQL v. 10.1.38-MariaDB
- PHP v. 7.3.4
- Java Script v. 1.6

## 2. Środowisko

Minimalne wymagania sprzętowe:

- Aplikacja działa na urządzeniach z systemem operacyjnym Windows, minimalna wersja: Windows 7

Biblioteki:

- urlopen
- BeautifulSoup
- request
- mysql.connector
- intertools
- sys
- json
- os

## 3. Linki do oprogramowania

- Link do repozytorium aplikacji:

<https://github.com/JosephSodomized/pysznyScraping>

- Link do Python v.3.7.6:

<https://www.python.org/downloads/release/python-376/>

- Link do XAMPP:

<https://www.apachefriends.org/index.html>

#### 4. Instalacja aplikacji

Aby aplikacja działa poprawnie należy przed jej uruchomieniem zainstalować program XAMPP. W trakcie instalacji programu postępować zgodnie z instrukcją. Po instalacji programu należy uruchomić XAMPP control, następnie uruchomić Serwer Apache oraz bazę danych MYSQL wciskając odpowiednie przyciski "START", co zostało przedstawione na rysunku 1.



Rysunek 1 Zaznaczone odpowiednie przyciski w programie XAMPP

Następnie po uruchomieniu serwera należy pobrać repozytorium projektu ze strony github.

Projekt należy rozpakować w ścieżce xampp/htdocs.

W celu uruchomienia aplikacji należy wprowadzić w pasku przeglądarki internetowej:

<http://localhost/pysznyScraping/index.php>

#### 5. Model danych

Dane zostały zawarte w modelu relacyjnym składającym się z jednej tabeli, posiadającej dziewięć kolumn: ID, Nazwa restauracji, Rodzaj kuchni, Liczba recenzji, Czas dostawy, Koszt dostawy, Zamówienie minimalne, Ocena, Ostatnia recenzja.

**6. Tabela z listą i opisem zmiennych, tabela z listą i opisem funkcji z podziałem na pliki**

ZMIENNA	OPIS
<b>webscraper.py</b>	
namesList	Tablica zawierająca nazwy wyszukiwanych restauracji
kitchensList	Tablica zawierająca rodzaje kuchni wyszukiwanych restauracji
reviewCount	Tablica zawierająca liczby recenzji wyszukiwanych restauracji
averageDeliveryTime	Tablica zawierająca czasy dostaw w wyszukiwanych restauracjach
deliveryCost	Tablica zawierająca koszty dostaw wyszukiwanych restauracji
minimumOrder	Tablica zawierająca minimalną cenę zamówienia w wyszukiwanych restauracjach
hrefLinks	Tablica zawierająca nazwy linków przekierowujących na strony wyszukiwanych restauracji
ratingNumbers	Tablica zawierająca oceny wyszukiwanych restauracji
lastWrittenReviews	Tablica zawierająca ostatnie opinie wyszukiwanych restauracji
postcode	Kod pocztowy wprowadzony przez użytkownika
<b>index.php</b>	
process	Zmienna typu string. Nazwa aktualnie wykonywanego procesu.
postcode	Zmienna typu string. Kod pocztowy wprowadzony przez użytkownika.
error	Zmienna typu string. Informacja o nieistniejącym kodzie pocztowym.
countOfExtractRecords	Zmienna typu intiger, zawiera liczbę rekordów, które zostały pobrane ze strony pyszne.pl

countOfLoadRecords	Zmienna typu intiger, zawiera liczbę rekordów, które zostały załadowane do bazy danych
<b>result.php</b>	
conn	Połączenie z bazą danych
sql	Rezultat zapytania do bazy
<b>index.js</b>	
form	Formularz, w którym zawarty został kliknięty przycisk
formData	Tablica z danymi formularza po procesie serializacji
data	Tablica zawierająca kod pocztowy nazwę procesu
url	Url na który zostanie wysłane zapytanie.
errorMesage	Informacja o nieistniejącym kodzie pocztowym.
messageClass	Nazwa klasy, w której pojawić ma się informacja o pobranych/załadowanych rekordach
countOfRecords	Liczba pobranych/załadowanych rekordów

NAZWA	PRZYJMOWANE PARAMETRY	OPIS DZIAŁANIA	WARTOŚĆ ZWRAcana
<b>webscraper.py</b>			
main	postcode	Wywoływana gdy jeszcze żaden z przycisków ETL nie zostanie wciśnięty	

procesExstract	postcode	Wywoływana poprzez przycisk Ekstract. Używając wprowadzonego przez użytkownika kodu pocztowego wysyła zapytanie na stronę pyszne.pl otrzymując w odpowiedzi html zawierający listę restauracji. Następnie wywoływana jest ekstrakcja potrzebnych danych na temat znalezionych restauracji i przypisanie ich do konkretnych tablic, po czym wywoływana jest funkcja saveFiles	-
saveFiles	-	Wywołuje funkcję saveFile dla każdego rodzaju wyekstrahowanych danych	-
saveFile	name, content	Zapisuje każdy rodzaj wyekstrahowanych danych do pliku	-
procesTransform	-	Wywołuje funkcję loadFiles a następnie wykonuje transformację danych, w celu przekształcenia ich do formy odpowiedniej do danych. Następnie wywołuje funkcję saveFiles	-
loadFiles	-	Powoduje otwarcie plików zawierających wyekstrahowane dane na temat wyszukanych informacji	

procesLoad	-	Używając funkcji loadFiles otwiera pliki z zapisanymi danymi na temat wyszukanych restauracji, następnie tworzy połączenie do bazy danych, zapisuje informację do danej tabeli w bazie danych, wywołując na końcu funkcję usuwającą pliki z danymi dotyczącymi wyszukanych restauracji.	
deleteFiles	-	Usuwa pliki z danymi dotyczącymi wyszukanych restauracji.	
<b>index.js</b>			
clearClass	Nazwa klasy	Usuwa daną klasę	-