Dokumentacja Aplikacji Indeksowania Stron

Spis Treści

**Wprowadzenie................................................................................................... 2**

**Main.py.............................................................................................................. 3**

**Data\_download.py............................................................................................ 6**

**Data\_pars.py..................................................................................................... 8**

**Data\_display.py................................................................................................. 9**

**Źródła............................................................................................................... 10**

**Strony Testowe................................................................................................ 10**

# **Wprowadzenie**

Aplikacja została napisana wykorzystując język Python w wersji 3.12.0. W folderze aplikacji znajdują się następujące pliki:

**Main.Py** - Główny plik odpowiedzialny za podstawowe okno aplikacji wraz z całym jego wyglądem i parametrami okna.

**Download\_data.json** - Plik JSON będący domyślnie pustym, zapisywane są w nim konkretne pliki pobrane przez funkcję zawartą w Data\_download.py. Plik można wyczyścić ręcznie lub jest automatycznie czyszczony w momencie zamykania aplikacji.

**Data\_download.py** - W tym pliku znajduje się funkcja odpowiedzialna za pobranie konkretnych plików z strony, a następnie umieszczenie ich w pliku Download\_data.json.

**Data\_pars.py** - Plik, w którym znajduje się funkcja odpowiedzialna za analizowanie i porządkowanie danych zawartych w pliku Download\_data.json.

**Data\_display.py** - Plik z funkcją odpowiedzialną za zebranie uporządkowanych danych i wyświetlenie ich w nowym oknie aplikacji.

Aplikacja została utrzymana w kolorystyce podobnej do aplikacji wcześniej opisywanej w ramach tego samego zadania, czyli biała z szarymi napisami i pomarańczowe akcenty.

# **Main.Py**

Główny plik aplikacji mający w sobie zaimportowane Biblioteki i pliki z funkcjami oraz danymi takimi jak:

* Tkinter
* Data\_download
* Data\_pars
* Data\_display

Cały kod podzielony jest na osiem odpowiednio podpisanych regionów w celu łatwiejszego poruszania się po nim i odczytu fizycznego.

Podstawowa zmienna aplikacji nazywa się “app” i pod nią są podpisane ustawienia takie jak:

* Stały rozmiar okna aplikacji
* Tytuł Okna
* Rozmiar aplikacji podany w pikselach
* Tło głównego okna

Po uruchomieniu aplikacji ukazuje nam się prostokątne okno z tytułem “Prosta Aplikacja Indeksowania Stron” podpisana w kodzie pod zmienną “title\_label” zawierającą parametry kolorystyczne i wyglądu tytułu. Ustawienie elementów okna głównego zostało oparte o siatkę (“grid”) i ustawione w jednej kolumnie (“column”) z podziałem na osobne rzędy (“row”).

W następnej sekcji umieszczone są funkcje odpowiedzialne za pokazywanie i ukrywanie komunikatów w aplikacji dotyczących wypełnionych przez przyciski funkcji, wraz z wywoływaniem z innych plików funkcji odpowiedzialnych za adekwatne działania.

Następne 4 regiony poświęcone są przyciskom odpowiedzialnym za wykonywanie wyżej wymienionych funkcji za pomocą parametru “command” dodanego do parametrów przycisku, gdzie:

**“Pobieranie Danych”** aktywuje funkcję “dwn\_show” odpowiedzialną za wykonanie funkcji “data\_scrap” z pliku “Data\_download”, a także wyświetlenie napisu “Operacja Przeprowadzana w nowym oknie” przy czym ukrywa komunikaty podpisane pod dwa następne przyciski.

**“Parsowanie Danych”** ma przypisaną funkcję “pars\_show” która to wczytuje funkcję “pars\_file” z pliku “Data\_pars” wraz z wyświetleniem napisu pod przyciskiem “Dane Zostały Przeanalizowane” jednocześnie ukrywając napisy pod pozostałymi przyciskami.

**“Wyświetl Dane”** wykonuje funkcję “display\_show” mającą za zadanie wykonać funkcję “display\_data” z pliku “Data\_display”, następnie wypisuje pod przyciskiem komunikat “Dane Są Wyświetlane W Nowym Oknie” w tym samym czasie ukrywając napisy pod przyciskami powyżej.

**“Usuń Dane”** odnosi się do funkcji “dump\_show” odpowiedzialną za uruchomienie funkcji “data\_dump” w pliku “Data\_Download” wraz z wyświetleniem komunikatu “Dane Zostały Usunięte” oraz ukrycie komunikatów od pozostałych przycisków.

Ostatnie dwa regiony są poświęcone przycisku do zamykania aplikacji i czyszczenia danych oraz aktywacji aplikacji.

Zamknięcie aplikacji dzieje się za pośrednictwem parametru “command” odnoszącego się do funkcji “close\_app” która to niszczy proces za pomocą parametru “destroy”, a także również wykonuje funkcję “data\_dump” z pliku “Data\_Download” w celu wyczyszczenia danych.

Natomiast uruchomienie aplikacji dzieje się za pośrednictwem komendy “mainloop” przypisanej do głównej zmiennej “app”.

# **Data\_download.py**

Plik ten jest podzielony na trzy regiony oddzielające od siebie import bibliotek od funkcji tego pliku.

Biblioteki wykorzystane w tej części aplikacji to;

* **BeautifulSoup**
* **Requests**
* **tkinter**
* **Json**

Funkcja wykonawcza pliku nazywa się “data\_scrap” i za pośrednictwem biblioteki request pobiera ona dane z strony podanej przez użytkownika. Pobrane pliki następnie są przetwarzane przez bibliotekę “BeautifulSoup” wykorzystując “html.parser”.

Następnie w pobranym kodzie HTML wyszukiwana jest najbliższy element oznaczony jako tabela, po czym z tych wierszy pobierane są wiersze oznaczone “<tr>” nagłówki oznaczone “<th>” i dane do tabeli oznaczone “<td>”. Po wyciągnięciu danych są one czyszczone z zbędnych białych znaków.

Pobrane i przetworzone dane są dodawane do listy “data” po czym otwierany jest plik “Download\_data.json” do którego wpisywane są uporządkowane dane, na koniec funkcja wypisuje w terminalu informacje o udanym zapisaniu danych.

W dalszej części kodu zaimplementowana została funkcja odpowiedzialna za otwieranie nowego okna aplikacji, w którym użytkownik może podać adres strony internetowej i pobrać z niej dane. Parametry okna aplikacji zostały nadane takie same jak dla okna głównego.

W pliku znajduje się też funkcja odpowiedzialna za czyszczenie pliku JSON z zawartości poprzez wstawienie w miejsce danych pustej tablicy. Następuje to w przypadku ręcznego wywołania jej przyciskiem do tego stworzonym lub wyjścia z aplikacji za pomocą guzika w oknie głównym.

# **Data\_pars.py**

W pliku tym została zaimportowana tylko jedna biblioteka pozwalająca na odczytywanie i dostęp do pliku JSON z pobranymi danymi.

Funkcją wykonawczą tego pliku jest “pars\_file” mający zaimplementowaną ścieżkę dostępu do pliku JSON zawierającego uporządkowane informacje, które następnie analizuje i układa w celu zwiększenia przejrzystości.

Program zaczyna od otworzenia i odczytania pliku danych, jednocześnie sprawdzając czy dane te mają odpowiednie kodowanie w postaci “utf-8” i w przypadku błędu formatu danych bądź ich braku zwraca wiadomość o błędzie.

Następnie kod pobiera pierwszy wiersz danych, w którym zawarte są nazwy kolumn i przypisuje go do zmiennej “headers”, po czym wykonuje to samo dla reszty wierszy, w których są powiązane dane i przypisuje je do zmiennej “data\_rows”.

W kolejnym etapie tworzona jest lista słowników, w których jako klucz ustawiane są nagłówki, a jako wartości przypisywane są pozostałe dane.

Następnie z wykorzystaniem “textwrap” wypisywane są dane w terminalu układając je numerycznie tak jak były prezentowane na stronie.

# **Data\_display.py**

Plik odpowiedzialny za wyświetlanie tabeli danych w nowym oknie aplikacji. W tym celu potrzebne były biblioteki tkinter i JSON.

W kodzie jest zaimplementowana pusta zmienna app2 która później zostanie zmieniona na zmienną globalną w celu utworzenia instrukcji warunkowej sprawdzającej czy nowe okno aplikacji już istnieje, tak by się nie podwajały. W razie wykrycia takiego okna wyświetla stosowne okno błędu “messagebox”.

Funkcją wykonawczą w tym kodzie jest “display\_data” mająca zaimplementowaną ścieżkę pliku JSON. Poza instrukcją warunkową opisaną wyżej w kodzie znajduje się ten sam sposób weryfikacji poprawności i wczytywania danych do zmiennych jak w pliku “Data\_pars”.

Następnie w kodzie wykorzystane są elementy pliku “Main” w celu utworzenia nowego okna aplikacji “app3” oraz nadaniu mu odpowiednich parametrów wraz z tytułem znajdującym się nad tabelą z danymi.

W końcowej części kodu za pomocą parametry “ttk” biblioteki “tkinter” jest tworzona tabela “treeview danych wyświetlana w aplikacji przypisana pod zmienną “tree”, a także wykorzystując pętlę do ustawienia kolumn i rzędów w pożądanych sposób. Wraz z przyciskiem służącym do zamknięcia okna z tabelą.

# **Źródła użyte w trakcie prac nad projektem:**

* <https://docs.python.org/pl/3>
* <https://docs.python.org/3/library/tk.html>
* <https://www.youtube.com/watch?v=LC9yE7T93cs>
* <https://tedboy.github.io/bs4_doc>
* <https://sites.google.com/view/paztronomer/blog/basic/python-colors>
* <https://stackoverflow.com>

**Przykładowe strony użyte w testowaniu kodu:**

* <https://www.scrapethissite.com/pages/forms/>
* [https://pl.wikipedia.org/wiki/Lista\_największych\_przedsiębiorstw](https://pl.wikipedia.org/wiki/Lista_najwi%C4%99kszych_przedsi%C4%99biorstw)