

Specyfikacja oprogramowania

„Web Scraper”

Spis Treści

Spis Treści	2
1. Wstęp	
1.1 Cel	3
1.2 Definicje	3
2. Krótki opis	3
3. Użytkownicy systemu	3
3. Funkcje	3,4
4. Interfejs graficzny	4
6. Narzędzia produkcji	4
7. Harmonogram prac nad projektem	5

1. Wstęp

1.1 Cel

Głównym celem tworzenia tego oprogramowania jest zaliczenie przedmiotu „Podstawy języków programowania II” oraz rozwijanie umiejętności programowania.

1.2 Definicje

Wy tłumaczenie skrótów i definicji występujących w specyfikacji:

- Adres lub link - Adres strony internetowej

2. Krótki opis

Program ma na celu wyszukiwania danych wybranych przez użytkownika, na stronach internetowych podanych w postaci adresów. Dodatkowo oprogramowanie formatuje dane i przedstawia je w wybrany przez użytkownika sposób np. plik tekstowy.

3. Użytkownicy systemu

Oprogramowanie kierowane jest do osób, które potrzebują narzędzia umożliwiającego automatyczne przetworzenie i przedstawienie w odpowiednim formacie, wybranej zawartości wielu stron internetowych w krótkim czasie.

4. Funkcje

Funkcje w które powinno być wyposażone oprogramowanie:

- Pobieranie plików graficznych z podanych adresów stron internetowych.
- Wyszukiwanie zdań i/lub akapitów na podstawie wprowadzonych słów lub wyrażeń .
- Wyszukiwanie numerów telefonów przedstawionych w różnych formatach. Możliwość wyboru formatów numerów.
- Wyszukiwanie odnośników (adresów) do innych stron internetowych.
- Możliwość załadowania adresów z pliku tekstowego.
- Pobieranie tabel z stron internetowych i zapisywanie ich w formacie CSV.
- Zapisywanie uzyskanych danych do pliku tekstowego.
- (Dodatkowo) Możliwość wykonywania wyżej wymienionych funkcji na znalezionych odnośnikach z możliwością wyboru „głębokości” przeprowadzania operacji. *(tzn: Na podanej stronie internetowej wyszukujemy np. numery telefonów. Wykonujemy tą samą operację dla adresów znajdujących się na tej stronie. Czynność powtarzamy N razy w zależności od wybranej „głębokości”).*

- (Dodatkowo) Przedstawienie statystyk dotyczących uzyskanych danych w formie wykresów.

5. Interfejs graficzny

- Menu
 - Konfiguracja - przycisk otwierający okno konfiguracji.
 - Start - przycisk rozpoczynający wybrane operacje.
 - Stop - przycisk zatrzymujący działanie operacji przed ich zakończeniem.
- Okno konfiguracji
 - Wybór sposobu wprowadzenia adresów stron (pole tekstowe lub plik tekstowy).
 - Wybór operacji które mają być przeprowadzone.
 - Konfiguracja wybranych operacji.
- Pole tekstowe do wprowadzania adresów. Zablokowane gdy w konfiguracji sposób wprowadzenia adresów został zdefiniowany jako plik tekstowy.
- Okno prezentujące uzyskane dane
 - Pole tekstowe z uzyskanymi danymi.
 - Zapisz - przycisk umożliwiający zapisanie danych do pliku tekstowego.
 - Zamknij - przycisk zamykający okno.

6. Narzędzia produkcji

Oprogramowanie będzie napisane za pomocą języka Python przy wykorzystaniu biblioteki PyQT oraz chrome driver (selenium) w zintegrowanym środowisku PyCharm

7. Harmonogram prac nad projektem

Tydzień	Data	Cel
1	03.12.2018	Prototyp interfejsu graficznego oraz pobieranie plików graficznych
2	10.12.2018	Wyszukiwanie zdań na podstawie wprowadzonych wyrażeń oraz wyszukiwanie adresów stron.
3	17.12.2018	Możliwość wczytywania adresów z pliku tekstowego oraz zapisywanie uzyskanych danych do pliku tekstowego Wyszukiwanie akapitów na podstawie wprowadzonych wyrażeń.
4	07.01.2018	Pobieranie tabel i zapisywanie w formacie CSV oraz usprawnienie interfejsu graficznego.
5	14.01.2018	Testowanie oraz poprawa błędów