# Cel:

Na stronie [www.zapytajfizyka.fuw.edu.pl](http://www.zapytajfizyka.fuw.edu.pl) znajduje się seria ciekawych wykładów popularyzujących współczesną fizykę. Zakres ich jest bardzo zróżnicowany i w zależności od indywidualnych preferencji, osoba zainteresowana, będzie z pewnością chciała w pierwszej kolejności obejrzeć prelekcje najbardziej ją interesujące. Zadaniem aplikacji jest opracowanie zwartego streszczenia dla każdego z wykładów aby zainteresowany był w stanie w ciągu relatywnie krótkiego czasu wybrać najbardziej interesujące go treści.

# Sposób wykonania

W pierwszej kolejności należy uruchomić skrypt *downloading.py*, za pomocą polecenia:

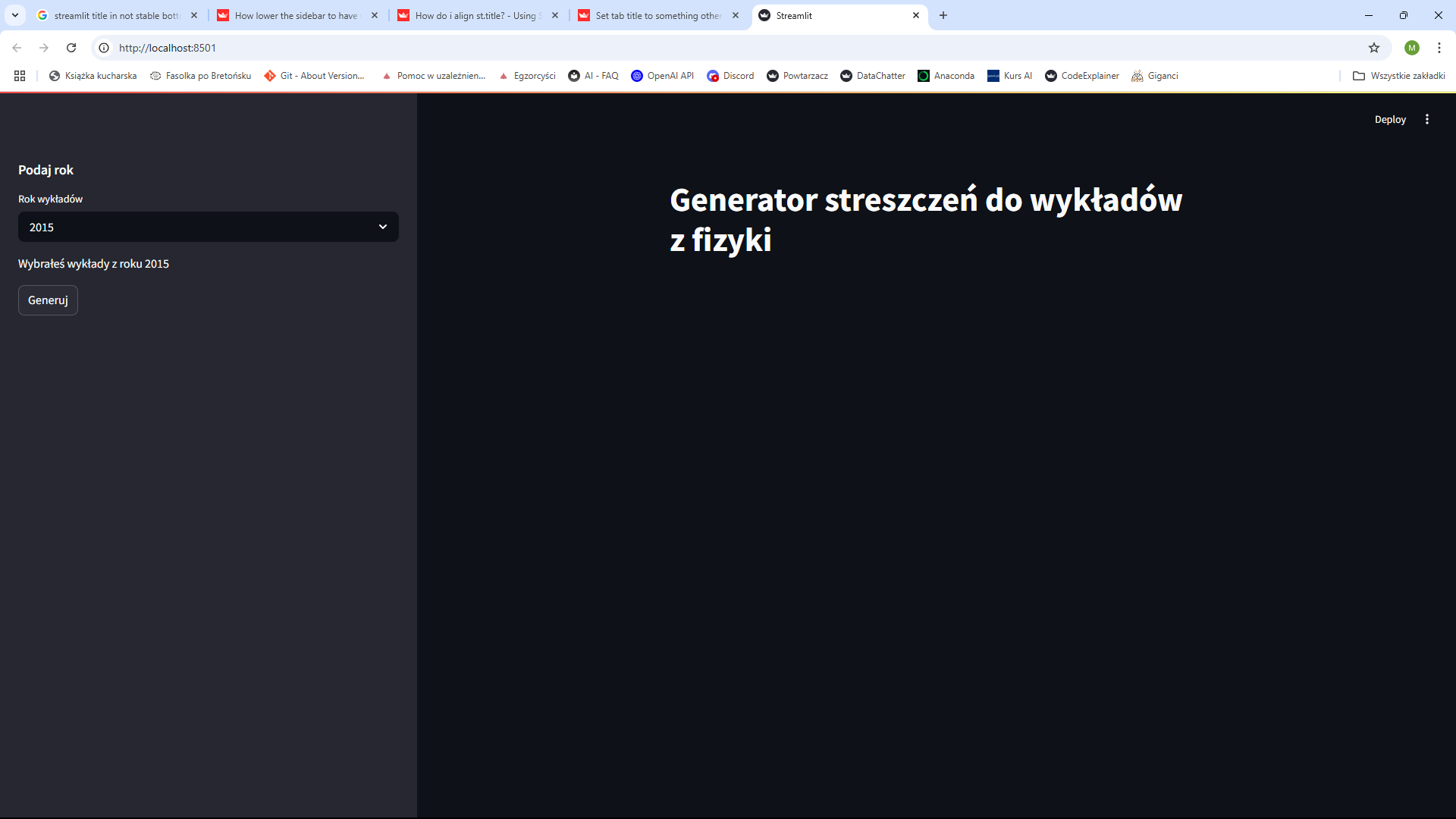
$ **python** **downloading.py**

Zadaniem tego skryptu jest pobranie z portalu *YouTube* wszystkich filmów z wykładami i umieszczenie ich w odpowiednim folderze **lectures***rok* gdzie *rok* to rok opublikowania prelekcji.

Po tym dość długotrwałym procesie można już uruchomić skrypt *summarize.py* za pomocą polecenia:

$ **streamlit run summarize.py**

Na ekranie monitora powinien się pojawić nast. obraz:



W sidebare użytkownik powinien wybrać rok w którym zostały wygłoszone wykłady (w jednym roku jest ich od 3 do 9).

Naciśnięcie klawisza Generuj zainicjuje przetwarzanie wg nast. algorytmu:

1. Sprawdzenie czy istnieją foldery dla plików audio, transkrybowanego tekstu i podsumowania (odpowiednio: *audio*, *text* i *summary*) jeśli ich nie ma to stworzenie takowych.
2. Pobranie pierwszego filmu z zadanego katalogu (np. ***lectures2024***/), wydzielenie z niego ścieżki dźwiękowej i zapisanie do folderu **/audio**.
3. Wydzielanie z pliku fragmentu do dalszego przetwarzania. Ucinane jest 10% początkowego czasu (raczej mało istotne wprowadzenie ) i ucinana jest druga połowa pliku (zostaje więc ok 40 % wyjściowego materiału). Wydzielanie jest podyktowane tym, że obecnie zastosowane metody umożliwiają przetwarzanie plików o wielkości do 25 MB.
4. Sprawdzenie czy rozmiar pliku nie przekracza 25 MB – jeśli tak to pomijanie przetwarzania i przejście do kolejnego filmu.
5. Dokonanie transkrypcji z audio do pliku tekstowego i zapisanie jej do pliku w folderze **text/**
6. Wykonanie streszczenia z pełnego tekstu i zapisanie go do folderu **summary/**

Przetwarzanie jest długotrwałe w związku z tym, po każdym kroku wyświetlony jest komunikat dot. obecnego stanu przetwarzania a poza tym został zobrazowany progressbar graficznie pokazujący jaki procent plików został już przetworzony

# Problemy

1. Na początku skryptu został umieszczony kod obrazujący tytuł aplikacji (*st.title* ..). O ile na początku pojawia się na górze strony to po kolejnych komunikatach „schodzi” na dół – nie wiedziałem jak go „ustabilizować”.
2. Drugi to wzmiankowany wcześniej problem to wydzielanie plików audio do zbyt dużych plików, które nie mogą zostać przetwarzane.

# Podsumowanie

Aplikacja została przetestowana na jednym folderze (czas i koszty 😊). Pomimo ograniczenia pliku audio do 40% jego wyjściowego rozmiaru podsumowania są dobrej jakości i wystarczająco streszczają treść prelekcji.

Problemem pozostają pominięte pliki audio. Jest to do zrobienia z kolejnej wersji poprzez iteracyjne zmniejszanie (np. o 2%) rozmiaru audio aż do uzyskania pożądanego rozmiaru.

Do rozwiązania zadania załączyłem jedno przykładowe streszczenie.