edesk.pearson.pl zasysacz nagrań

# cel:

pobrać automatycznie i zapisać w plikach dźwiękowych materiał udostępniany do podręcznika(ów)

# opis:

pliki są dostępne tylko do odtworzenia on-line po zalogowaniu na platformie edesk.pearson.pl. Z analizy kodu strony wynika, że nie są przechowywane adresy do plików dźwiękowych tylko są one dostarczane jako usługa po identyfikatorze konkretnego nagrania. Wykorzystywane są obszerne i dosyć skomplikowane mechanizmy JS z wykorzystaniem jQuery i Bootstrap. Również biblioteka „modernizr”. Obsługa nagrań jest w chyba autorskim skrypcie: /Content/audio?v=mGMacGLw3fC3GLp5Pf8Ozj2gr7BIlYtgJFAZonSRAJ41

Wygląda na to, że kwestie autoryzacji są realizowane z wykorzystaniem cookies ważnych dobę, więc nie powinno być problemów z autentykacją.

# koncepcja rozwiązania:

ponieważ dotarcie do adresów plików dźwiękowych wydaje się bardzo skomplikowane a ponadto najprawdopodobniej dostęp do nich będzie chroniony spróbujemy zrobić robota który:

1. pobierze ze strony listę wszystkich nagrań
2. kolejno wywoła każde z nagrań
3. przechwyci dźwięk z odtwarzanego nagrania i zapisze je do pliku (np. wav) odpowiednio nazywając pliki i umieszczając je logicznie w odpowiednich katalogach

Ewentualna konwersja wav do mp3 już jako postprocessing, ponieważ do tego mam już gotowe dobre narzędzia. Dopuszczalne jest wywołanie z poziomu już wybranej konkretnej książki albo pulpitu użytkownika gdzie po kolei byłyby wywoływane kolejne książki/rozdziały.

Przebieg programu będzie wymagał dłuższego czasu jednak ponieważ będzie to czynność realizowana przez komputer w pełni automatycznie to nie stanowi to problemu wobec faktu, że wystarczy jednorazowe wykonanie. Można przyjąć, że proces logowania jest realizowany ręcznie i występuje przed uruchomieniem skryptu i zakładamy, że na czas działania skryptu mamy aktywne dane autentykacyjne.

# szczegóły rozwiązania:

## element pojedynczego nagrania w kodzie strony:

### otoczka wywołania danego nagrania:

<div **class="material col-md-3 col-xs-4 material-4"**>

<a href="**/Audio/Index/564?p=11**" target="\_blank" class="openmodal widget">

<div class="widget-content themed-background-info text-light-op text-center">

<i class="fa fa-music fa-3x"></i><br />

<strong>**Ścieżka (5)**</strong>

</div>

</a>

</div>

Gdzie:

1. numer po Index/ oznacza identyfikator nagrania natomiast parametr „?p={n}”, w którym n jest identyfikatorem pierwszego nagrania danego CD (w przypadku pierwszego nagrania występuje „?p=-1”)
2. zapisana w znacznikach <strong> nazwa / numer ścieżki powinna być wykorzystana przy nazywaniu wynikowego pliku dźwiękowego (numery lepiej dopełniać do określonej długości zerami)

### kod sterujący odtwarzania danego nagrania:

pojawia się po wywołaniu danego nagrania przez naciśnięcie na href z powyższego punktu.

<div class="audiojs " classname="audiojs" id="audiojs\_wrapper0">

<audio src="/Audio/GetAudio?access\_token=11b72a42-6355-4312-a1c7-4bc149f9e065" preload="auto"></audio>

<div class="play-pause">

**<p class="play"></p>**

<p class="pause"></p>

<p class="loading"></p>

<p class="error"></p>

</div>

<div class="scrubber">

<div class="progress" style="width: 0px;"></div>

<div class="loaded" style="width: 35.3767px;"></div>

</div>

<div class="time">

<em class="played">00:00</em>/<strong class="duration">**01:12**</strong>

</div>

<div class="error-message"></div>

</div>

Start odtwarzania uzyskujemy przez naciśnięcie „play”.

Można też pobierać długość nagrania żeby odpowiednio sterować nagrywaniem dźwięku w trakcie jego odtwarzania a również zawrzeć długość w nazwie pliku. Do kontroli nagrania jeśli się da to lepsze byłoby odpytywanie na bieżąco czy „played” jest już równe „duration” (lub ponownie ustawi się na „00:00”) i wtedy kończyć nagrywanie – byłoby to o tyle lepsze, że unika się ryzyka ucinania końcówki nagrania, gdyby rozpoczęło się ono z opóźnieniem (są programy, które potem pozwalają wyczyścić nagrania hurtowo z wiodącej ciszy).

Zamknięcie okienka modalnego z konkretnym nagraniem można uzyskać przez „naciśnięcie”

a)

<div class="modal-header">

<button type="button" class="close" data-dismiss="modal" aria-label="Close"><span aria-hidden="true">×</span></button>

<h4 class="modal-title" id="myModalLabel">

<div class="widget-content themed-background-info text-light-op text-center" style="height: 92px;">

<i class="fa fa-music fa-3x"></i><br>

<strong>**Ścieżka (5)**</strong>

</div>

</h4>

</div>

Gdzie przy okazji można też pobierać informację do nazwy pliku danego nagrania.

b)

<div class="modal-footer">

<button type="button" class="btn btn-default" data-dismiss="modal">Zamknij</button>

</div>

## Nazwę dla CD /publikacji

można wyciągnąć ze struktury:

<div class="**col-sm-2**">

<img src="/Files/Covers/music9.png" alt="**Class CD 1**" style="max-width: 100%">

</div>

Lub jeszcze lepiej z nagłówka dla danej listy nagrań:

<**div class="content-header"**>

<div class="row">

<**div class="col-sm-6"**>

<div class="header-section">

<h1>**Class CD 1**</h1>

</div>

</div>

<div class="col-sm-6 hidden-xs">

<div class="header-section">

<ul class="breadcrumb breadcrumb-top">

<li>

Ścieżka (1)

</li>

<li>

<a href="/Home/Publication/4">**English Class B1**</a>

</li>

<li>

<a href="/Home/Course/1">English Class</a>

</li>

<li>

<a href="/Home/Category/1">Szkoła Podstawowa</a>

</li>

</ul>

</div>

</div>

</div>

</div>

## Potrzebne komponenty

1. kontrolowane wywoływanie poszczególnych elementów (stron/modułów)
2. tworzenie nazw plików i katalogów
3. nagrywanie dźwięku w trakcie jego odtwarzania ze strony/modułu
4. zapisywanie nagranego dźwięku do pliku

## Schemat algorytmu

stan startowy:

jestem zalogowany do strony z aktywną autentykacją

start:

przyjmuję parametry (katalog docelowy, adres strony, stałe elementy nazw, itp.)

zapisuję do listy wszystkie CD jakie mam dostępne na pulpicie (strona pulpitu=<https://edesk.pearson.pl/>)

dla każdego CD z listy: (przykład: [https://edesk.pearson.pl/Home/Show/**11**](https://edesk.pearson.pl/Home/Show/11))

ustalam katalog i prefix nazwy plików dla tego CD

tworzę listę utworów dla tego CD (strona tego CD)

dla każdego utworu z listy utworów dla tego CD:

wywołuję okienko odtwarzania utworu (przykład: [https://edesk.pearson.pl/Audio/Index/**559?p=11**](https://edesk.pearson.pl/Audio/Index/559?p=11)) [można przez podanie adresu, ale z https]

przygotowuję i startuję nagrywanie dźwięku

uruchamiam odtwarzanie dźwięku **[jak???]**

sprawdzam co sekundę(?) czy nagranie się odtwarza

gdy koniec utworu to kończę nagrywanie

określam sufix nazwy pliku

zapisuję nagranie do pliku

raportuję zakończenie działania ☺

# Opis rozwiązania (Python)

## krok 1: utworzenie list z utworami do nagrania

## krok 2: nagrywanie utworów z danej listy