

ArchiveWeb.page - instrukcja

1. Informacje wstępne

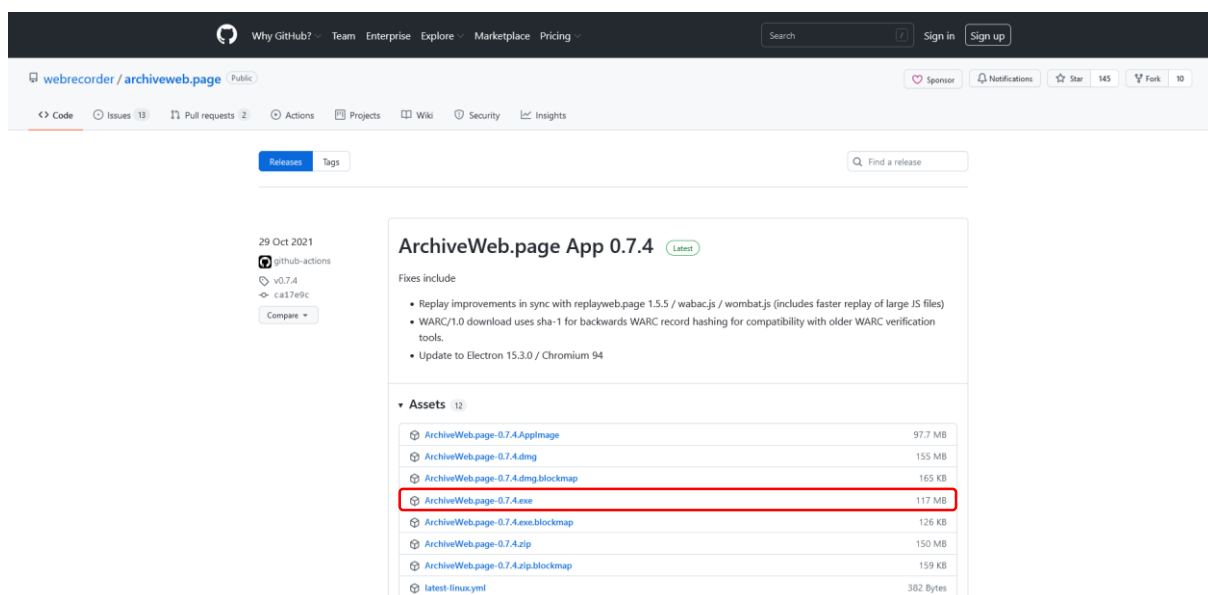
ArchiveWeb.page to oprogramowanie przygotowane w ramach projektu Webrecorder, które umożliwia wykonywanie archiwalnych kopii stron WWW, w taki sposób aby jak najpełniej odwały to, jak wyglądały oraz zachowanie ich funkcjonalności (np. przemieszczanie się za pomocą linków). Omawiane narzędzie przypomina tradycyjną przeglądarkę internetową, która archiwizuje zawartość poszczególnych stron w trakcie ich przeglądania. Zarchiwizowane zasoby można zapisać w formacie WARC, który jest ogólnie przyjętym formatem plików do przechowywania archiwalnych kopii zasobów Webowych. Oprogramowanie jest dostępne do wykorzystania za darmo niezależnie od celu na zasadzie open source. Jest ono wykorzystywane jako dodatkowe narzędzie przez niektóre z archiwów webowych, zwłaszcza w przypadku bardziej zaawansowanych technologicznie witryn WWW lub mediów społecznościowych.

ArchiveWeb.page jest dostępne jako samodzielna aplikacja desktopowa oraz rozszerzenie do przeglądarek opartych o silnik Chromium (np. Google Chrome i MS Edge). W instrukcji zostanie przedstawiona wersja desktopowa, jednakże funkcjonowanie rozszerzenia do przeglądarki wygląda identycznie. Podobnym narzędziem, bazującym na tych samych rozwiązaniach jest Conifer (dawny Webrecorder.io). Można z niego korzystać za pomocą przeglądarki, wymaga jednak założenie darmowego konta w serwisie. Dostępny jest pod adresem <https://conifer.rhizome.org/>.

Instrukcja w języku angielskim, przygotowana przez twórców oprogramowania jest dostępna pod adresem: <https://archiveweb.page/guide>.

2. Pobieranie i instalacja oprogramowania

Oprogramowanie w wersji desktopowej możliwe jest z portalu GitHub pod adresem: <https://github.com/webrecorder/archiveweb.page/releases>. Należy pobrać najnowszą wersję z rozszerzeniem .exe. (Rys. 1) oraz zainstalować (system Windows może ostrzegać przed niesprawdzonym źródłem oprogramowania)



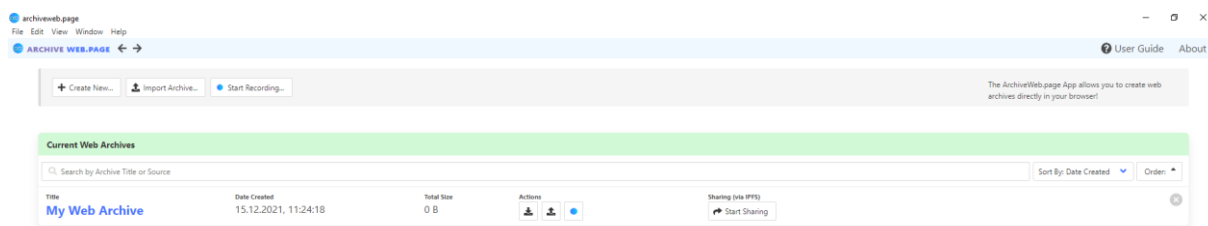
Rys. 1

3. Uruchomienie i okno główne programu

Po instalacji na pulpicie komputera pokaże się ikona ArchiveWeb.page, którą należy uruchomić, po czym pokaże się okno główne programu (Rys. 2). Na szarym pasku widoczne są następujące przyciski:

- Create New... - pozwala na utworzenie nowego zbioru archiwalnych stron WWW.
- Import Archive... - pozwala na zaimportowanie pliku WARC
- Start Recording... - rozpoczyna proces archiwizacji

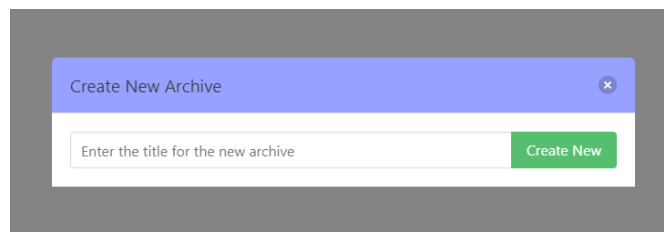
Poniżej znajduje się lista kolekcji stron WWW z informacją o jej nazwie, dacie utworzenia, rozmiarze, przyciski odpowiedzialne za import, eksport, rozpoczęcie archiwizacji, udostępnienie oraz usunięcie zbioru.



Rys. 2

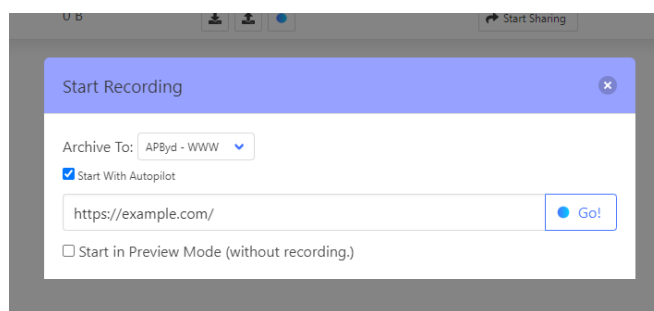
4. Rozpoczęcie archiwizacji

Archiwizację witryny należy od utworzenia nowej kolekcji po kliknięciu w przycisk „Create New..” i nadanie jej nazwy (Rys. 3). Zaleca się utworzenia osobnej kolekcji dla jednej witryny i jej wszystkich podstron (np. osobno dla www.bydgoszcz.ap.gov.pl i www.torun.ap.gov.pl).



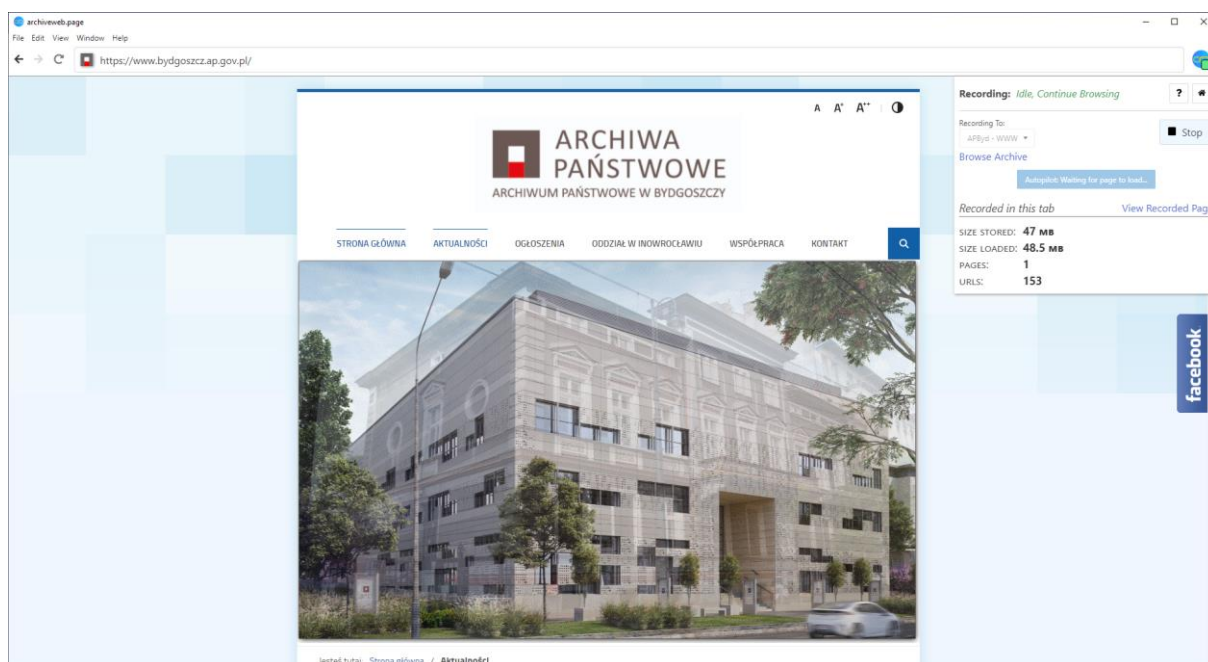
Rys. 3

Następnie można przejść do archiwizowania wybranych stron internetowych poprzez kliknięcie przycisku „Star recording...” znajdujący się na górnym pasku lub przy konkretnej kolekcji. Pojawi się wówczas okno (Rys. 4). Możemy w nim wskazać, do której z kolekcji chcemy dodawać zarchiwizowane strony oraz czy chcemy rozpocząć archiwizację z Autopilotem. Jest to funkcja, która automatycznie przewija stronę na jej dół i uruchamia interaktywne elementy (np. nagrania audio lub wideo). Opcja ta może nie uruchomić się w przypadku niektórych stron. Poniżej należy podać adres URL, od którego zaczynamy archiwizację. Ważne jest aby zachować początek adresu, tj. „https://”! Możliwe jest również przeglądanie strony bez jej archiwizacji poprzez zaznaczenie opcji „Start in Preview Mode”. Aby rozpocząć archiwizację należy kliknąć przycisk „Go!”.



Rys. 4

Po kliknięciu w przycisk „Go!” pojawi się nam nowe okno, w którym wyświetli się wskazana wcześniej strona WWW (Rys. 5). Na górnym pasku widzimy przyciski znane ze standardowych przeglądarek internetowych (wstecz, do przodu, odśwież) oraz pole do wpisywania adresu URL. Obok niego po prawej stronie znajduje się ikona z logiem ArchiveWeb.page oraz małym kwadratem, który zmienia kolor w zależności od statusu archiwizacji (żółty: archiwizacja w trakcie, niebieski: działa Autopilot, zielony: kopia wykonana). Po kliknięciu w nią widoczne będą szczegółowe informacje na temat archiwizacji: opisowy status archiwizacji, do jakiej kolekcji dodawane są kopie, ilość danych zapisanych i załadowanych (te wartości mogą się różnić, ponieważ oprogramowanie nie pobiera dodatkowych kopii posiadanych plików, np. banneru, który znajduje się na wszystkich podstronach), ilość zarchiwizowanych witryn oraz URL (liczba ta odnosi się do wszelkich elementów strony np. grafik, skryptów, aplikacji lub arkuszy CSS). Ponadto znajdują się tam przyciski zatrzymujące lub wznowiające archiwizację („Stop”/”Start”), wyświetlające wykaz zarchiwizowanych stron („Browse Archive”) lub aktualnie archiwizowaną stronę („View Recorded Page”).

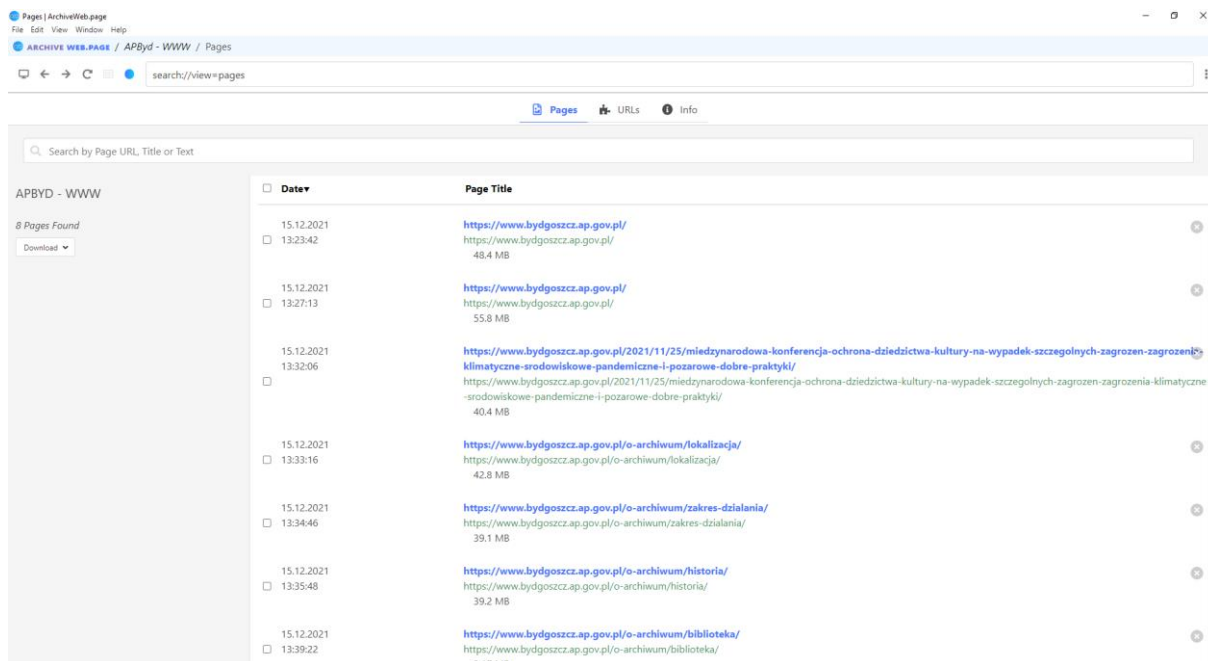


Rys. 5

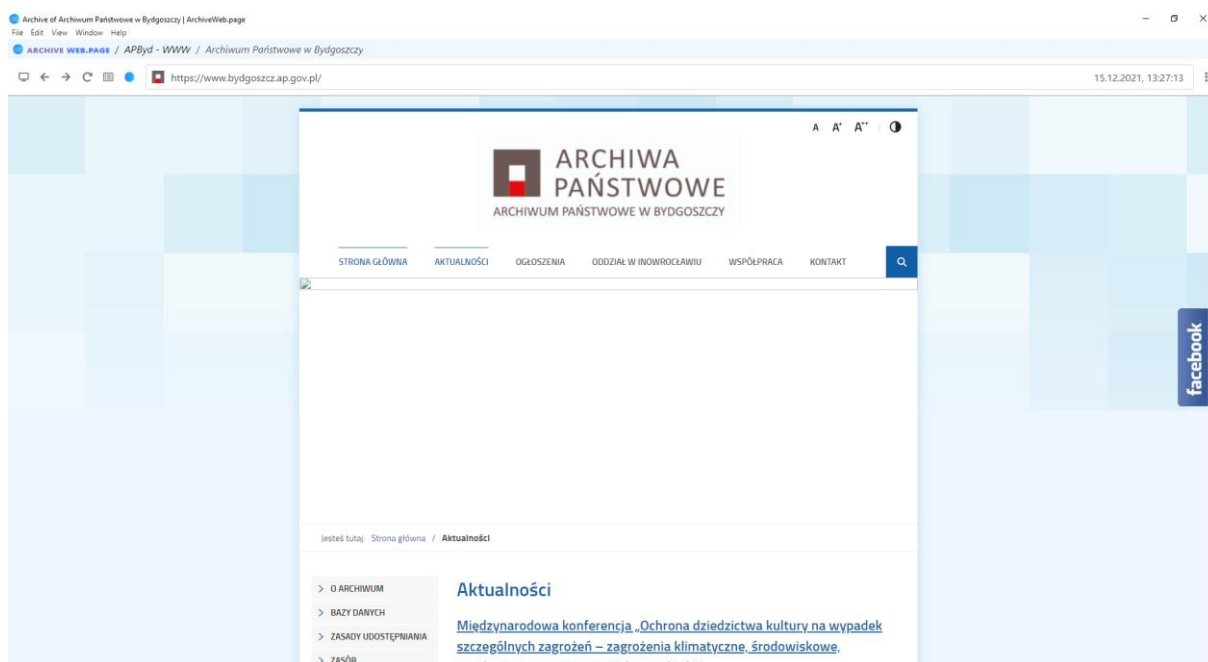
Przed przystąpieniem do archiwizacji zaleca się przygotowania schematu postępowania, np. wykazy podstron, które mają zostać zarchiwizowane oraz kolejności, w której będą przeglądane. Ma to na celu ułatwienie prac i kontrolę ich kompletności. Archiwizacja może zostać przerwana w dowolnym momencie i kontynuowana innego dnia.

5. Wyniki archiwizacji

Strony zarchiwizowane w danej kolekcji możemy przeglądać poprzez kliknięcie w jej nazwę w oknie głównym lub przycisk „Browse archive” widoczny menu widocznym przy przeglądaniu stron. Wyświetli się wówczas lista stron zarchiwizowanych (Rys. 6). Po lewej stronie widoczny jest nazwa kolekcji, liczba stron oraz przycisk „Download” (zostanie omówiony w następnym akapicie). Po prawej stronie widoczna jest lista stron z informacją o dacie i godzinie wykonania jej kopii, tytułem i adresem URL oraz rozmiarem, a także przycisk umożliwiający usunięcie wybranej kopii (przydatne np. w przypadku duplikatów, jednakże akcji tej nie można cofnąć). Możliwe jest przeglądanie też poszczególnych elementów storn w zakładce „URLs” oraz podstawowych informacji o kolekcji w zakładce „Info”. Po kliknięciu w wybraną możemy ją przeglądać (Rys. 7)

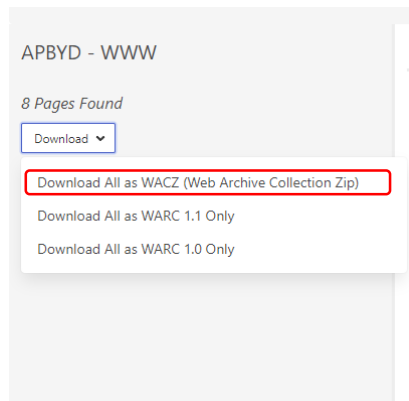


Rys. 6



Rys. 7

Zarchiwizowane zasoby należy wyeksportować. W tym celu należy kliknąć w przycisk „Download” i wybrać opcję „Download All as WACZ (Web Archive Collection Zip)” (Rys. 8) i wskazać miejsce, gdzie plik ma zostać zapisany. W przypadku dużej ilości danych może to zająć dłuższą chwilę.



ReplayWeb.page - instrukcja

1. Informacje wstępne

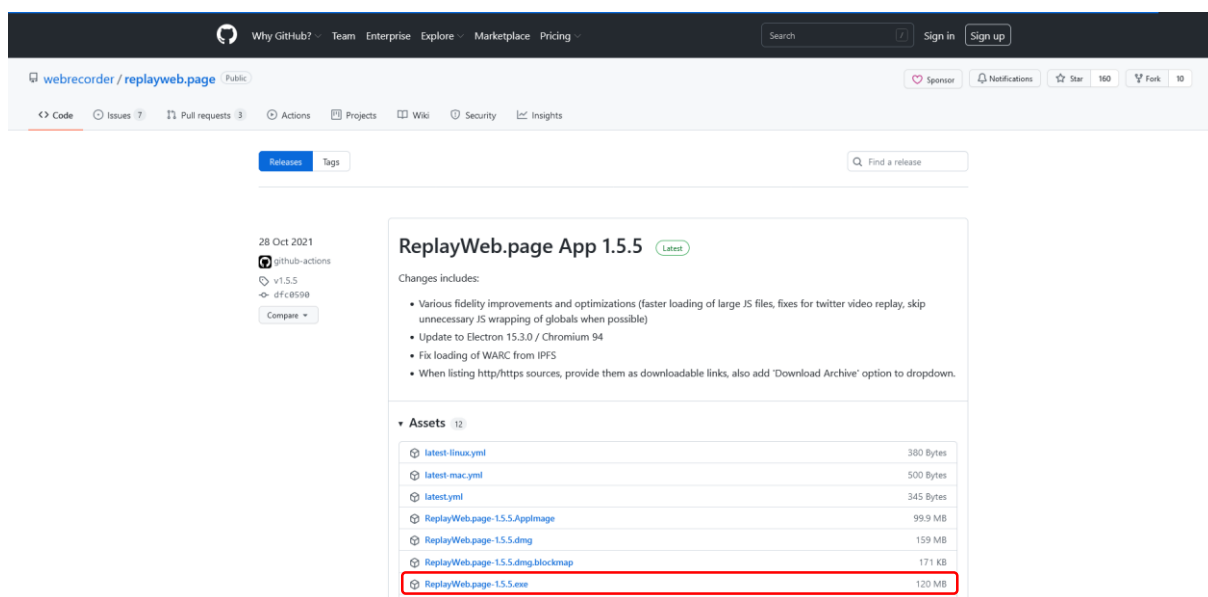
Strony WWW archiwizowane za pomocą ArchiveWeb.page lub innego narzędzia, zapisywane są do plików w formatach WARC lub WACZ. Do otwierania tego rodzaju plików służy program ReplayWeb.page, stworzony w ramach projektu Webrecorder. Pozwala on na odtwarzanie stron WWW tak, aby jak wyglądały jak w wersji online wraz zachowaniem ich funkcjonalności (np. przemieszczanie się za pomocą linków, korzystanie z multimediów).

ReplayWeb.page jest dostępne jako samodzielna aplikacja desktopowa oraz przeglądarkową dostępną pod adresem: <https://replayweb.page/>. W instrukcji zostanie przedstawiona wersja desktopowa, jednakże funkcjonowanie w przeglądarce internetowej wygląda identycznie. Podobnym narzędziem, bazującym na tych samych rozwiązaniach jest Conifer (dawny Webrecorder.io). Można z niego korzystać za pomocą przeglądarki, wymaga jednak założenie darmowego konta w serwisie. Dostępny jest pod adresem <https://conifer.rhizome.org/>.

Instrukcja w języku angielskim, przygotowana przez twórców oprogramowania jest dostępna pod adresem: <https://replayweb.page/docs/>.

2. Pobieranie i instalacja oprogramowania

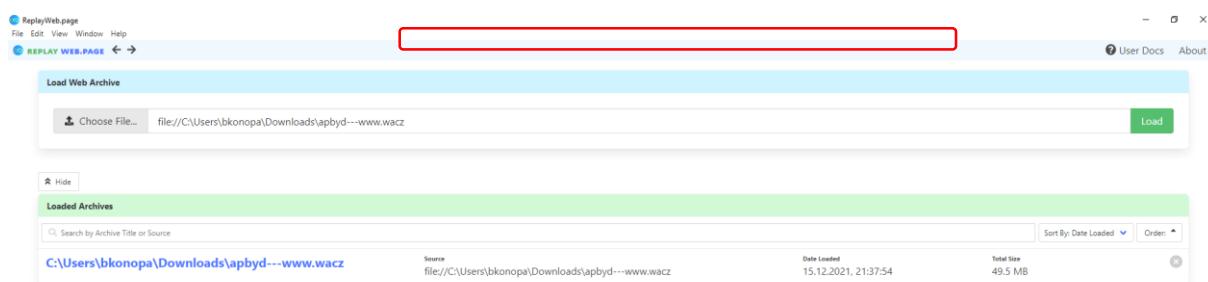
Oprogramowanie w wersji desktopowej możliwe jest z portalu GitHub pod adresem: <https://github.com/webrecorder/replayweb.page/releases>. Należy pobrać najnowszą wersję z rozszerzeniem .exe. (Rys. 1) oraz zainstalować (system Windows może ostrzegać przed niesprawdzonym źródłem oprogramowania)



Rys. 8

3. Uruchomienie i okno główne programu

Po instalacji na pulpicie komputera pokaże się ikona ReplayWeb.page, którą należy uruchomić, po czym pokaże się okno główne programu (Rys. 2). Zobaczyć możemy w nim zobaczyć dwa pole. W pierwszym z nich możemy wgrać do programu plik w formacie WARC lub WACZ. Aby to zrobić należy kliknąć „Choose File...”, wyszukać odpowiednik plik i go wybrać. Gdy w polu obok pojawiły się lokalizacja pliku należy kliknąć w przycisk „Load”. Drugie pole jest listą wgranych kolekcji stron wgranych do ReplayWeb.page. Oprócz niego nazwy mamy podaną lokalizację pliku, datę dodania pliku oraz jego rozmiar. Obok jest przycisk pozwalający usunąć pozycję z listy.

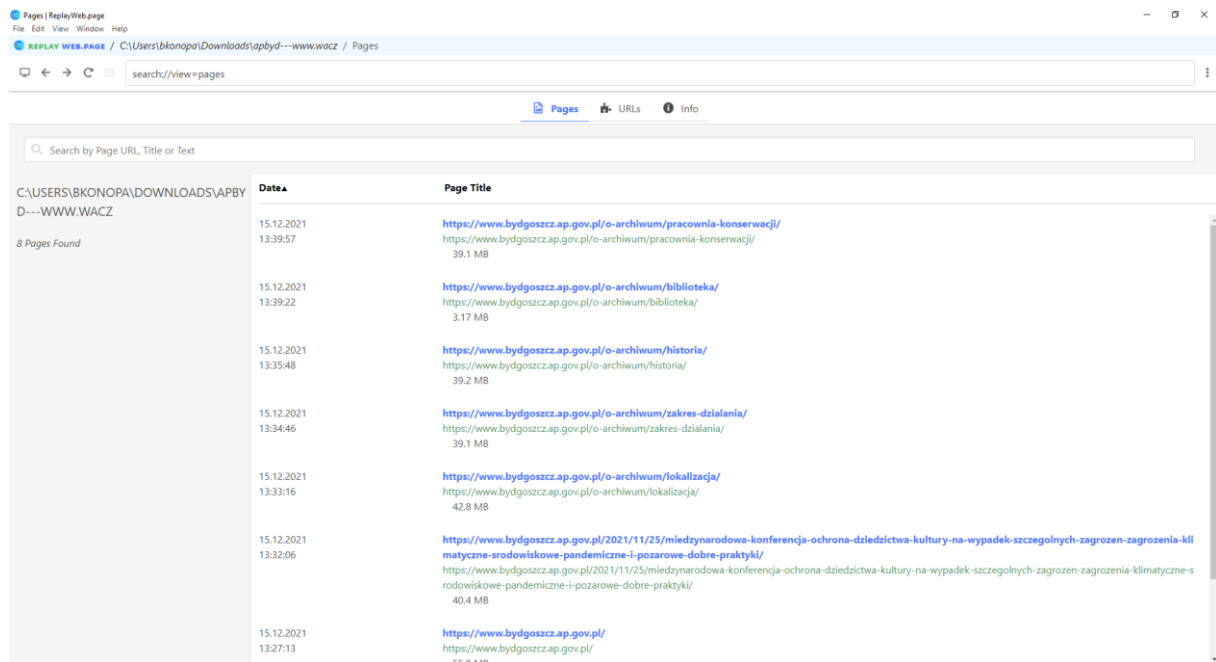


Rys. 9

4. Przeglądanie zarchiwizowanych stron

Po wgraniu pliku lub wybraniu kolekcji z listy pojawią się okno wyświetli się lista stron zarchiwizowanych (Rys. 3). Po lewej stronie widoczny jest nazwa kolekcji oraz liczba stron. Po prawej stronie widoczna jest lista stron z informacją o dacie i godzinie wykonania jej kopii, tytułem i adresem URL oraz rozmiarem. Możliwe jest przeglądanie też poszczególnych elementów storn w zakładce „URLs” oraz podstawowych informacji o kolekcji w zakładce „Info”. Po kliknięciu w wybraną możemy ją przeglądać w widoku przypominającym prostą

przeglądarką internetową. Obok adresu strony podana jest data i czas wykonania kopii (Rys. 4).



Rys. 10

