

Apresentação anotada

ARQUIVAR A WEB

FAÇA-VOCÊ-MESMO!

WEBINÁRIO



25 de Novembro às 10h00

RICARDO BASÍLIO
FCCN PT



Versão
com notas

Passaporte Competências Digitais da Câmara Municipal de Lisboa

Arquivar a Web: faça-você-mesmo!

Ricardo Basílio, curador digital do Arquivo.pt

Lisboa, 25 de novembro de 2021



Breve apresentação do Arquivo.pt

- Serviço público (FCT-FCCN)
- Missão – preservar a Web portuguesa
- Acessível na Web em <https://arquivo.pt>
- Instalações físicas – Av. do Brasil – Lisboa
(Campus do LNEC)



Versão
com notas

Objetivo

- Utilizar ferramentas do Webrecorder.net para gravar páginas Web no próprio computador num formato normalizado

Versão
com notas

Agenda

- Introdução
- Tutorial
 - Instalar a Extensão ArchiveWeb.page
 - Gravar, guardar e reproduzir páginas
 - Guardar localmente ficheiros gravados em formato normalizado WARC
- Papel das comunidades locais na preservação de conteúdos da Web
- Conclusão

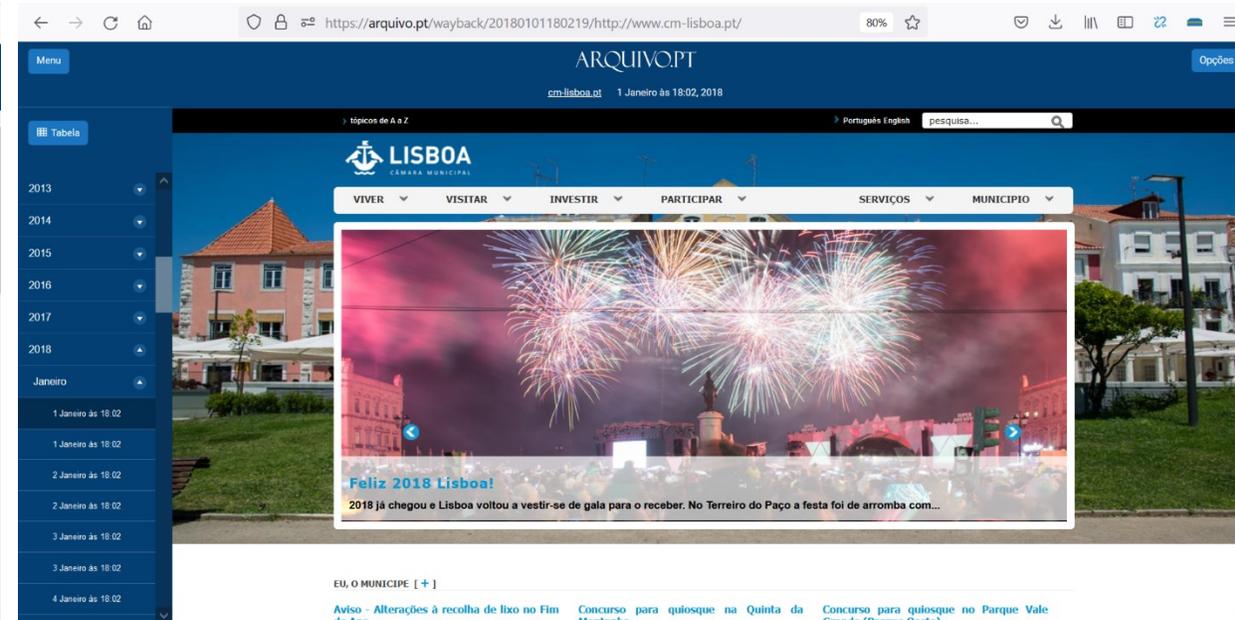
Introdução

Para que serve gravar a Web?

Antes... 1998



Depois... 2018



Os arquivos da Web servem para ver páginas do passado. Não são imagens planas. São páginas clicáveis e dinâmicas, “vivas”.

Vocabulário útil para começar a guardar páginas. Estes termos surgem frequentemente quando utiliza o ArchiveWeb.page ou outra ferramenta de gravação da Web.

A saber sobre os sites

O que é um URL

<https://www.lisboa.pt/>

O que é HTML

Uma das linguagens em que são escritos os sites; código

O que é um Browser

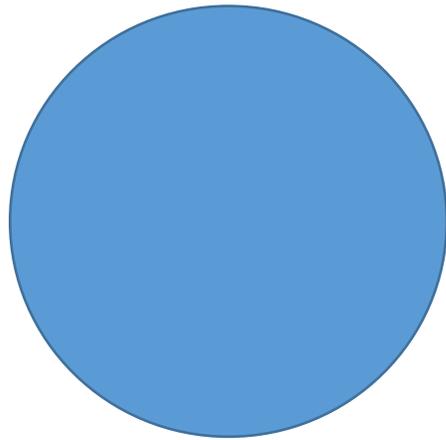
Software responsável por ler os sites nos nossos dispositivos. Exemplos: Chrome, Edge, Brave, Safari, Firefox

Formato WARC

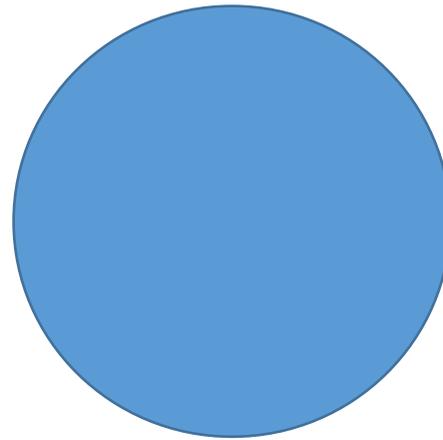
Formato normalizado para preservar sítios Web ISO 28500:2017

3 fases da Preservação da Web

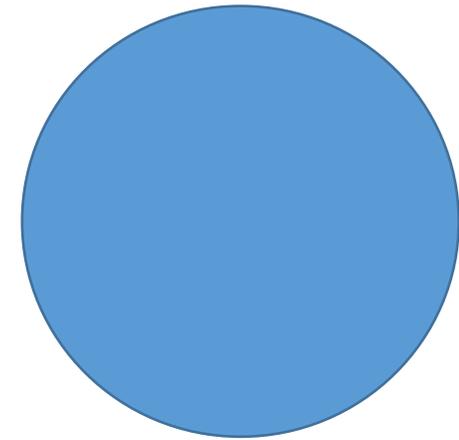
Neste slide procura-se mostrar que a extensão ArchiveWeb.page, assim como as ferramentas do Webrecorder.net, estão alinhadas com os arquivos da Web e por isso são ainda mais úteis. Tal como os grandes arquivos da web (Internet Archive, UK Web Archive, Arquivo.pt), também o ArchiveWeb.page grava, gera um ficheiro WARC e reproduz.



Gravar



Armazenar



Reproduzir

Parte 1

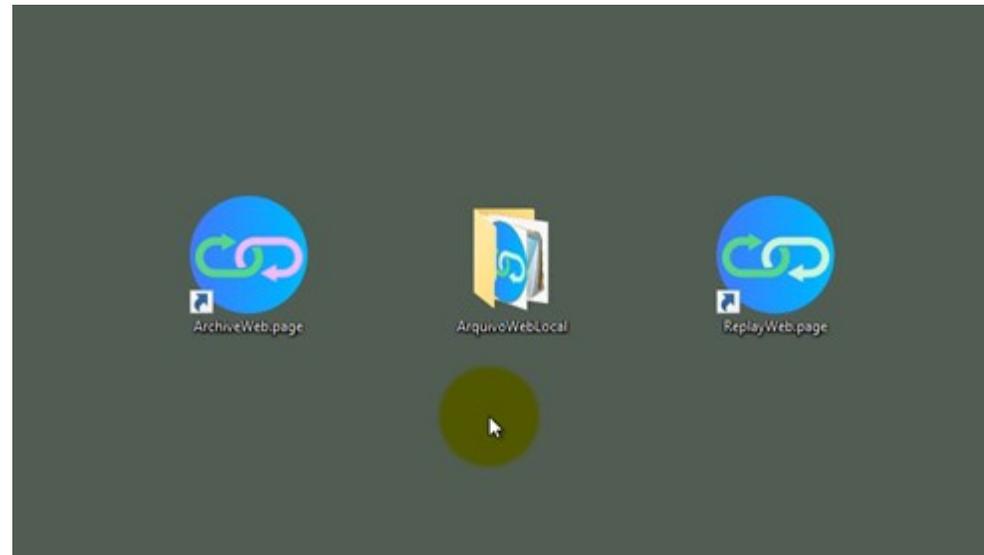
Tutorial

ArchiveWeb.page

1

Fluxo

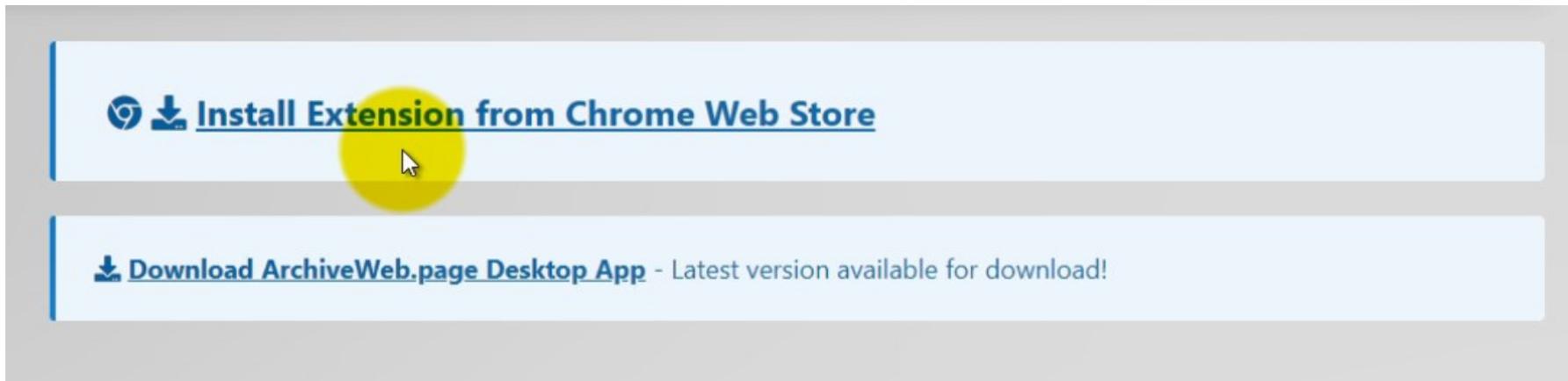
Gravar >> Armazenar >> Reproduzir



Ver gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=339>

Este vídeo mostra o fluxo de trabalho desejável num "Arquivo Web Local".

Instalar Extensão ArchiveWeb.page no browser



Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=390>

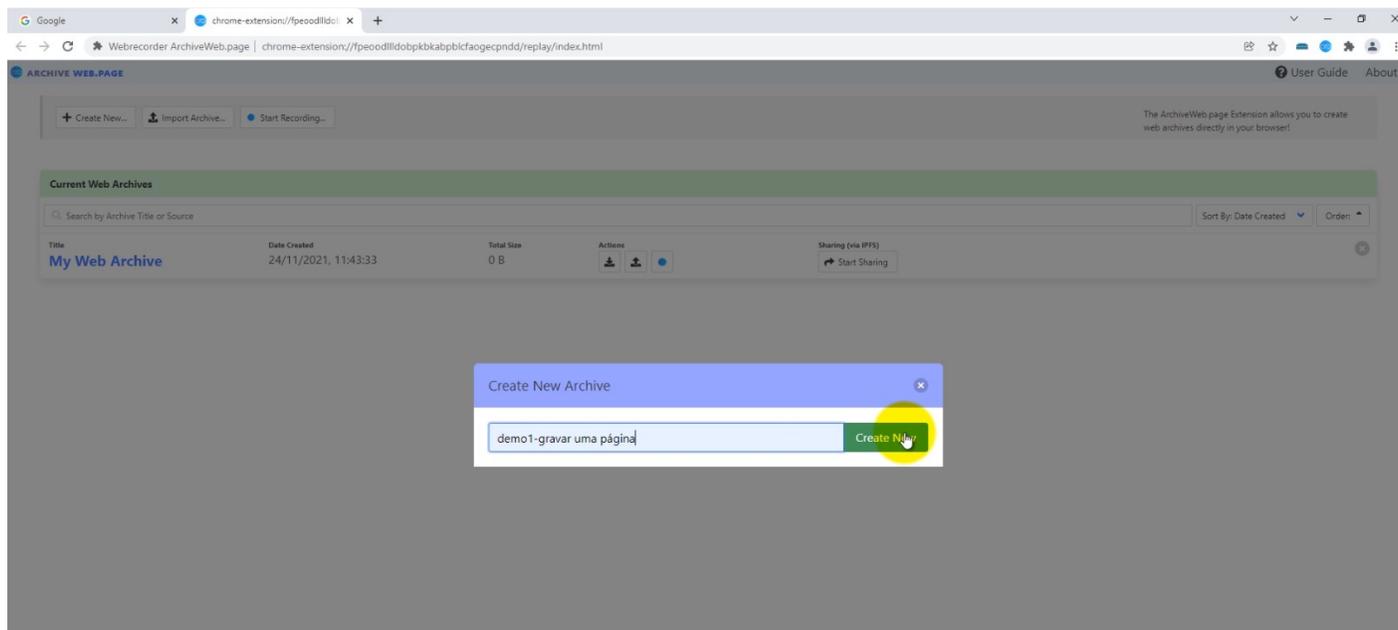
Este vídeo mostra a instalação da Extensão ArchiveWeb.page (Webrecorder.net) no browser Google Chrome ou em browsers baseados em Chrome. É sabido que no Safari e Firefox não funciona. Mas há solução.

Para instalar nos computadores Mac: instalar primeiro o Edge em Mac; em seguida, abrir o Edge e instalar a Extensão ArchiveWeb.page (Webrecorder.net).

Para instalar em Linux: utilizar o Chromium; instalar a extensão no Chromium.

3

Gravar uma página



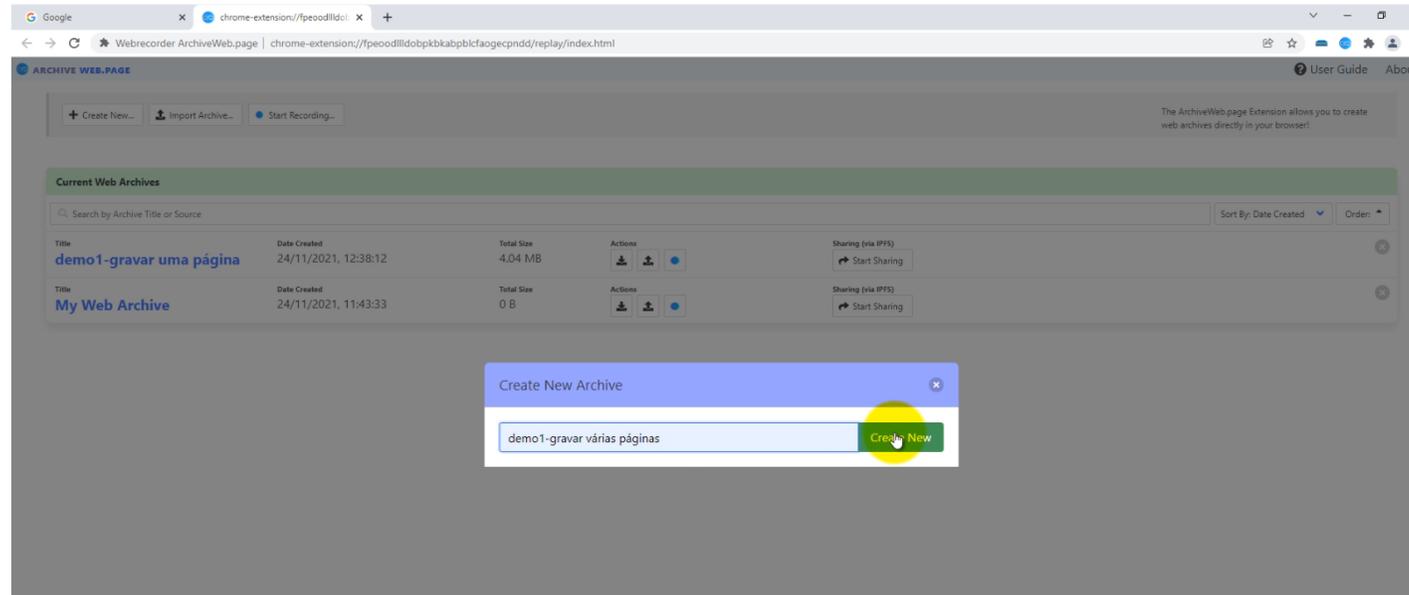
Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=598>

Este vídeo mostra a utilização da extensão de browser ArchiveWeb.page. Uma de cada vez. Clique a clique.

Propõe-se um percurso específico para usar a interface: abrir, criar pasta para conter gravações, gravar, ver gravação, voltar à pasta criada, exportar WACZ e WARC.

4

Gravar várias páginas



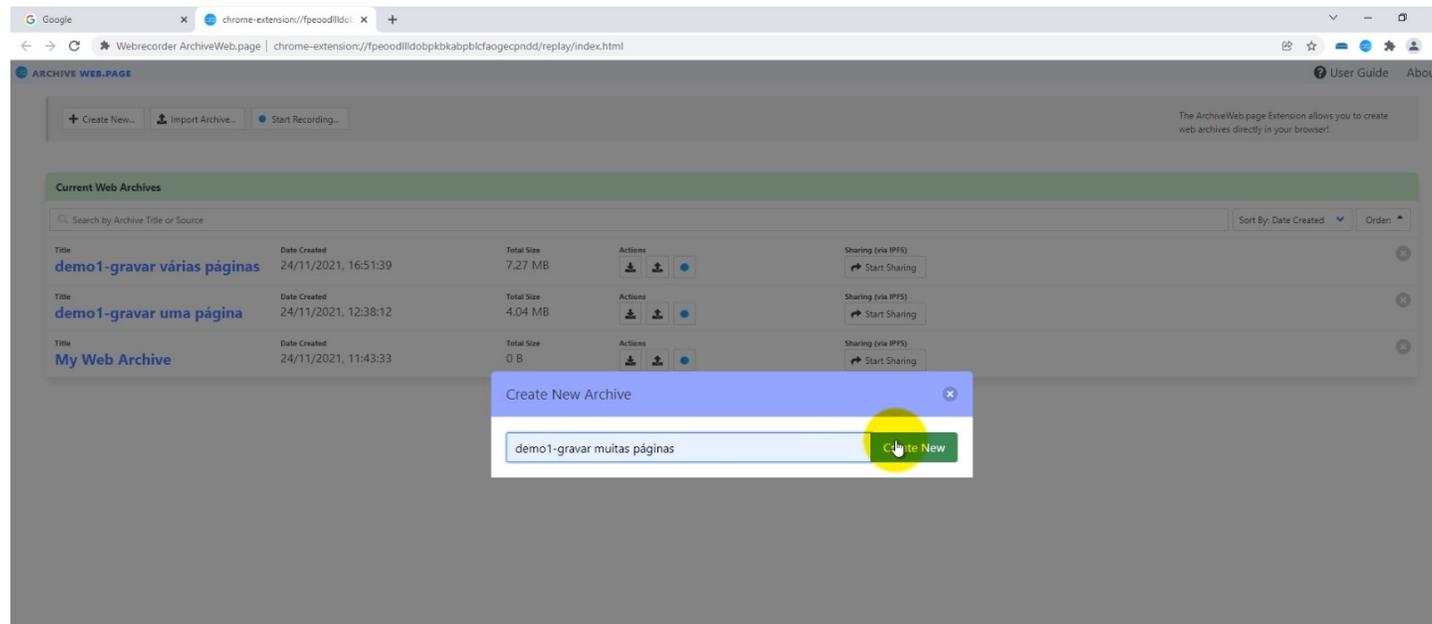
Ver gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=783>

Este vídeo repete o anterior mas com mais cliques, em várias páginas. Uma de cada vez. Clique a clique.

Propõe-se um percurso específico para usar a interface: abrir, criar pasta para conter gravações, gravar, ver e verificar gravação, voltar à pasta criada, exportar WACZ e WARC.

5

Gravar muitas páginas

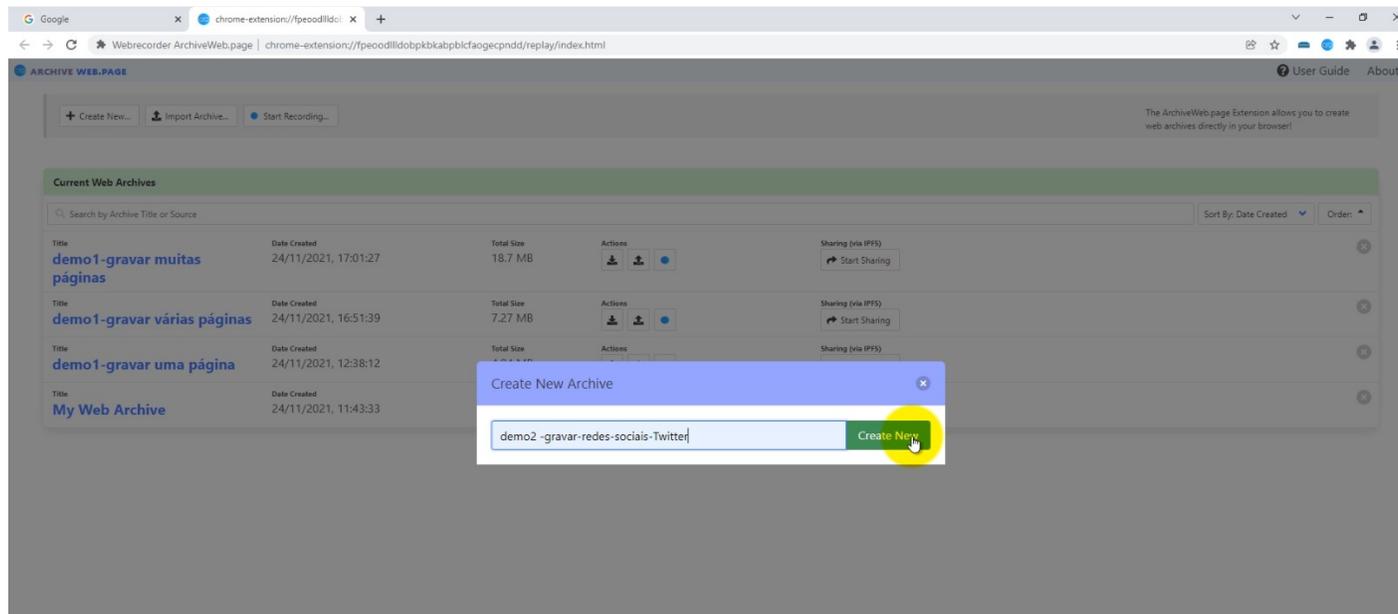


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=984>

Este vídeo amplia o procedimento com mais cliques, em páginas específicas, usando a extensão ArchiveWeb.page. Uma de cada vez. Clique a clique. E acrescenta a recomendação: abrir nova tab, abrir nova tab, etc. Permite controlar o processo de gravação sem perder a sequência desejada de páginas.

6

Gravar Twitter

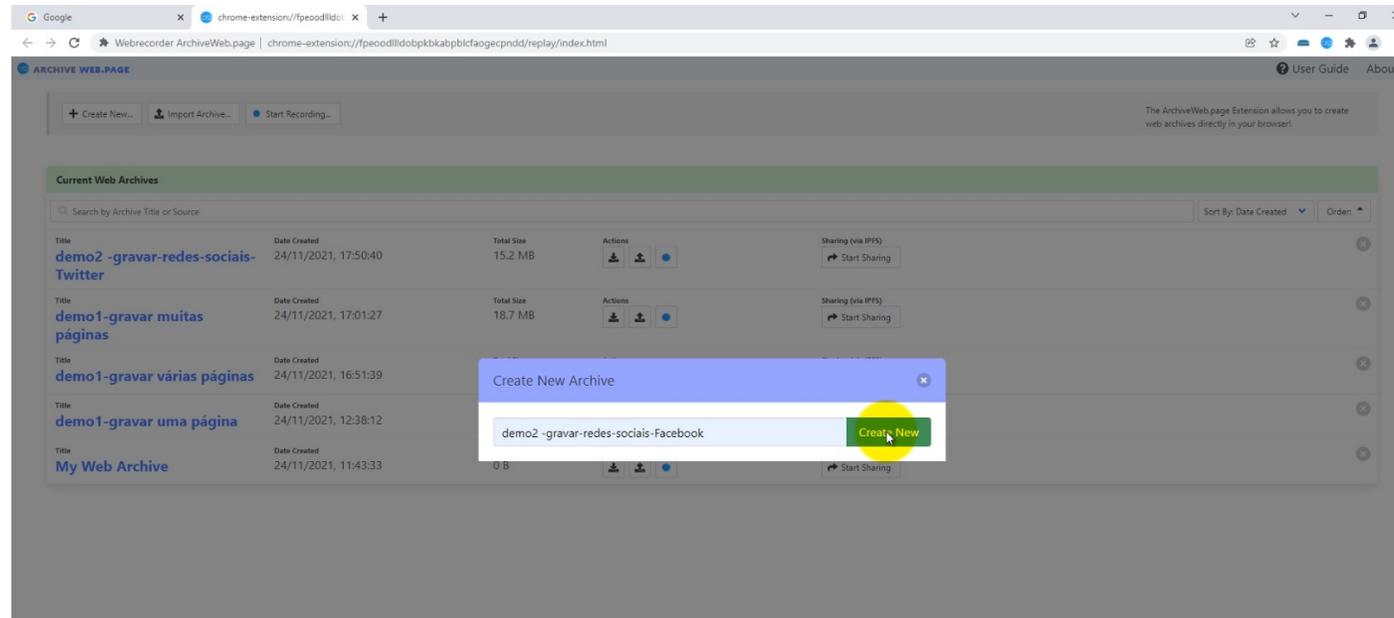


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=1351>

Este vídeo mostra a utilização da extensão para gravar Twitter. Correr verticalmente, alternativa com ou sem “piloto automático”. Grava o Twitter por for, sem login. Funciona bem. Os vídeos reproduzem. Posso personalizar a gravação e seguir links que interessem.

7

Gravar Facebook sem login

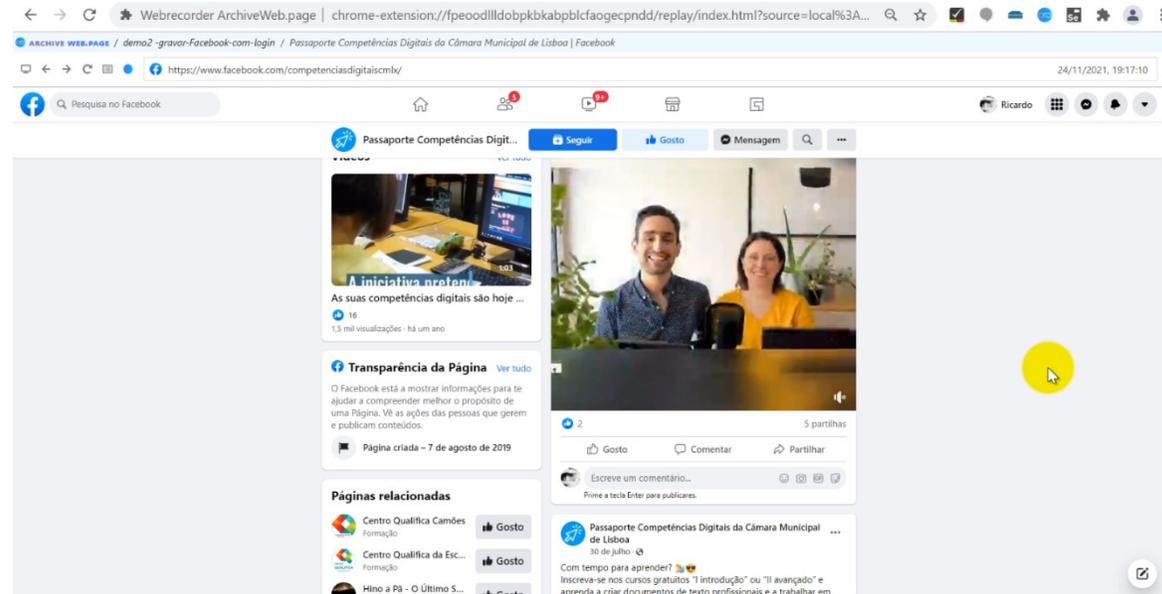


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=1531>

Este vídeo mostra a utilização da extensão para gravar Página do Facebook sem fazer login. Grava bem, tal como no Twitter. Os conteúdos “pesam” mais. Mas depois não reproduz bem. Depende do que as plataformas deixam ou não fazer. O Facebook não está a deixar.

8

Gravar Facebook com login

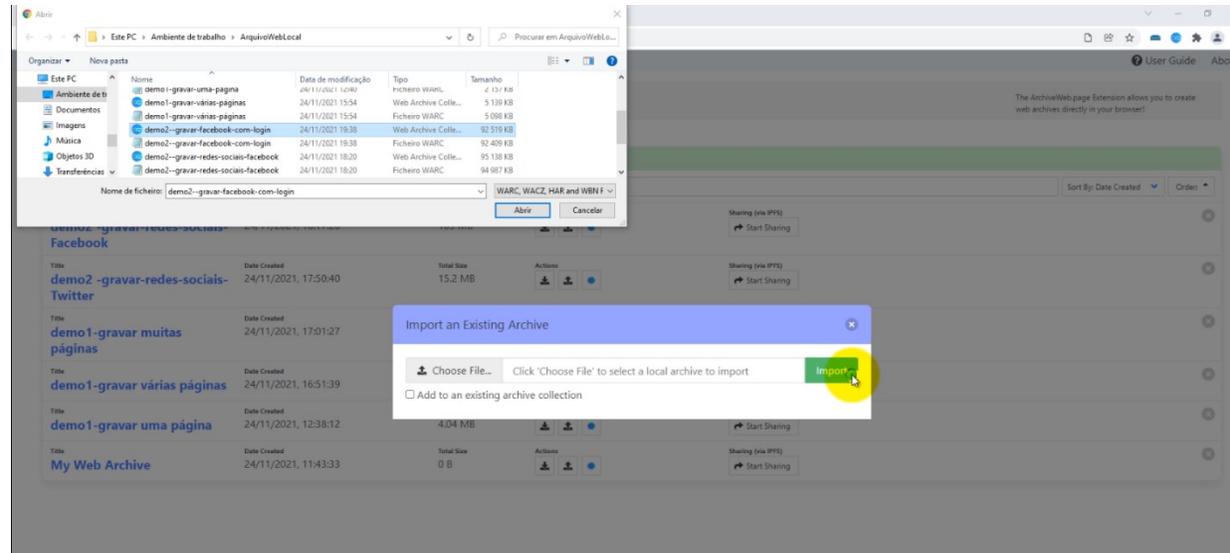


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=1759>

Este vídeo mostra a utilização da extensão para gravar Página do Facebook com login. Grava bem, tal como no Twitter. Os conteúdos “pesam” mais. Mas depois não reproduz bem. Na reprodução não passa à “pagina seguinte”.
Atenção: recomendado só para uso privado ou interno nas instituições. Deve fazer-se o login antes no browser e só depois de estar *logado* começar a gravar. Preencher o *user* e a *password* durante gravação pode resultar na escrita dessas credenciais que depois ficam guardadas no WARC.

8b

Importar gravados ficheiros para serem lidos

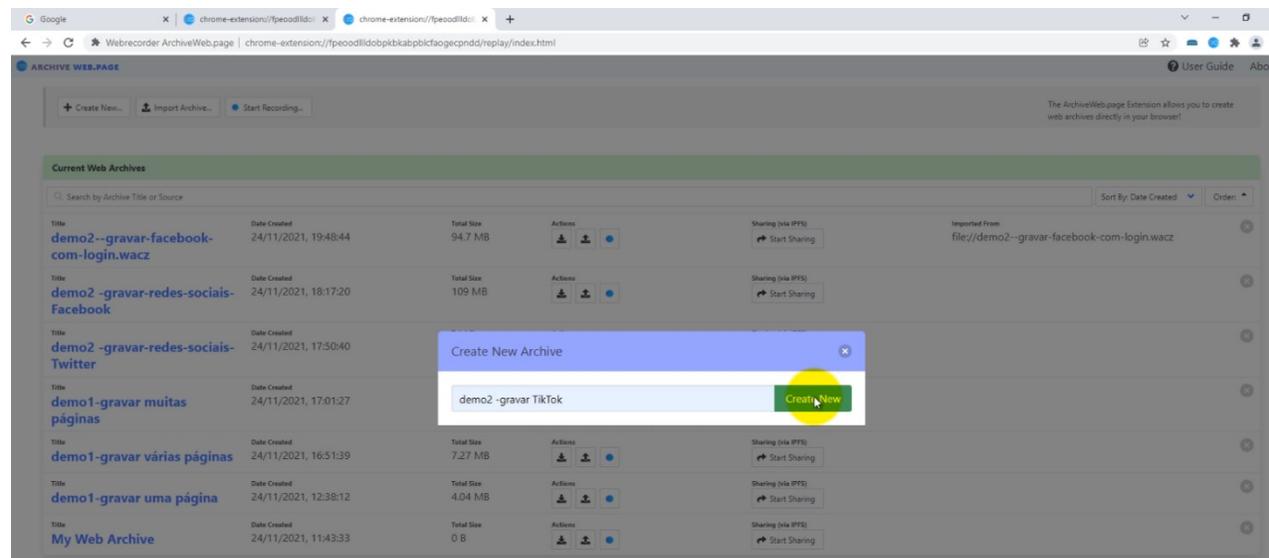


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=1879>

Este vídeo mostra como se pode importar para dentro do ArchiveWeb.page (Extensão ou Desktop App) ficheiros que foram gravados em outros lugares.

9

Gravar Tiktok

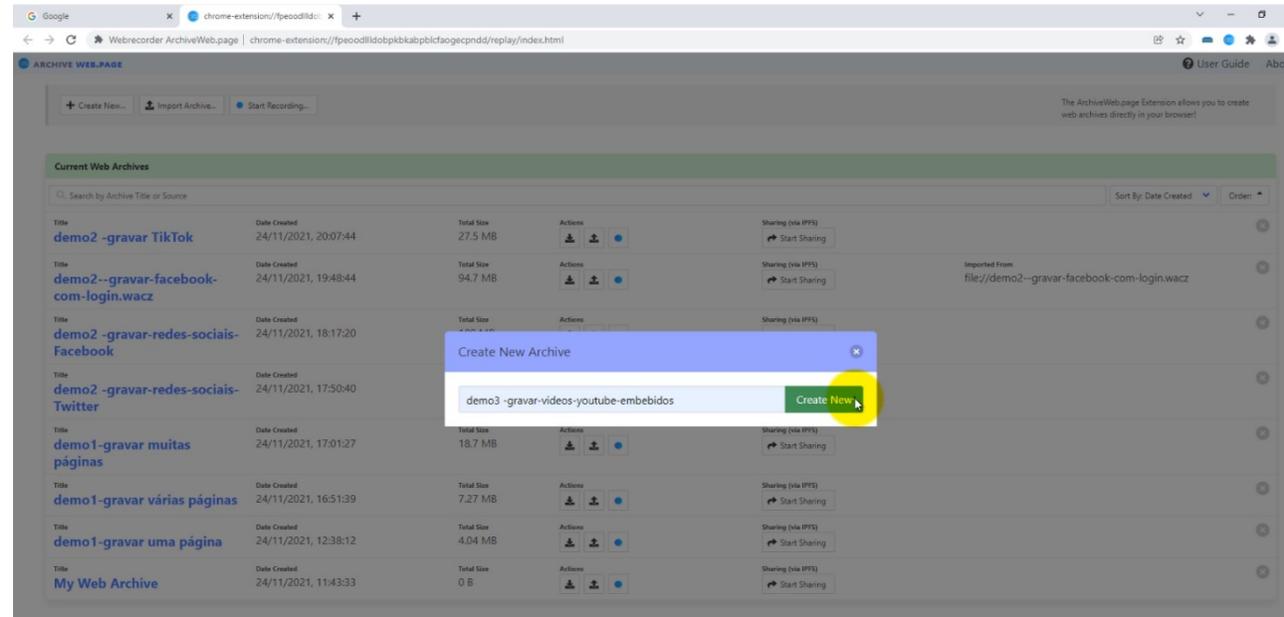


Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=1948>

Este vídeo mostra a utilização da extensão ArchiveWeb.page para gravar outras redes sociais, tais como o Tiktok. Experimente gravar ainda outras plataformas: Pinterest, Instagram, Vimeo, Wix, Blogspot.

10

Gravar vídeos do Youtube embebidos

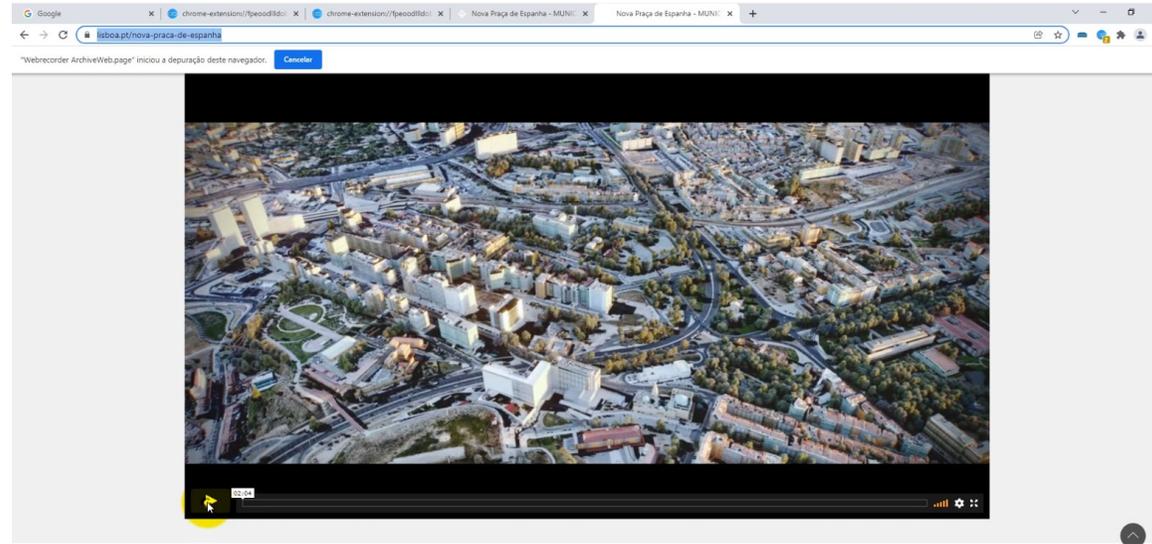


Ver gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=2007>

Este vídeo mostra a utilização da extensão para gravar vídeos embebidos nas páginas de Websites. Grava bem. Não é garantido que reproduza bem. Na gestão das Extensões “Manage Extensions”, no seu browser, onde pode ver quais as extensões que tem instaladas, verifique se tem instalada a app do Youtube. Se tiver, remova-a. Volte a gravar os seus vídeos do Youtube que agora já serão gravados e reprodutíveis no ArchiveWeb.page.

11

Gravar vídeos em outros formatos



Ver no gravação no Youtube: <https://youtu.be/KLJ6atqJh2Y?t=2227>

Este vídeo mostra a utilização da extensão para gravar vídeos embebidos nas páginas de Websites. É difícil gravar e reproduzir vídeo pois os formatos são muitos. Nem sempre software que lê é compatível com o formato em que o vídeo foi produzido. Ao longo do tempo os formatos mudam e o que funcionava antes deixa de funcionar depois. Neste caso correu bem. A utilidade de gravar vídeos no contexto em que foram publicados é precisamente manter a memória desse contexto.

Resumo

- O ArchiveWeb.page grava as páginas que escolhermos, clique a clique
- Conteúdos de websites, redes sociais, vídeos embebidos, são gravados com maior ou menor eficácia de acordo com cada plataforma
- Podemos exportar o conteúdo gravado num formato normalizado
- O ArchiveWeb.page reproduz os conteúdos no nosso próprio computador

Parte 2

Sobre o Webrecorder

O Webrecorder

“Web archiving for all”

- 2016 - Parte do Projeto Rhizome.org (Webrecorder.io)
- 2019 - Independente do Projeto Rhizome, dá suporte ao serviço Conifer
- 2020 - Webrecorder.net

Twitter: [@webrecorder_io](https://twitter.com/webrecorder_io)

Porque razão o Webrecorder.net é uma boa referência? A história desta iniciativa fala por si.
Já conta com alguns anos de experiência. Esta entre o melhor que se desenvolveu nos últimos anos.

O Webrecorder.net

Porque razão o Webrecorder.net é uma boa referência?
Tem o apoio de entidades internacionais importantes no domínio do Webarchiving.
É parceiro do Arquivo.pt que usa o software pyWB do Webrecorder para reproduzir as páginas Web.

Partners

Here are some of the partners that we work with or have collaborated with in the past:

Perma.cc 

 LOCKSS

RHIZOME

THE NATIONAL ARCHIVES

 hypothes.is

 IIPC
INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM

 CLOCKSS

 Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

 NATIONAL
LIBRARY
OF AUSTRALIA

 KIWIX

 MFB
FINO

 UKWA
UK WEB ARCHIVE

 DocNow

 ARQUIVO.PT

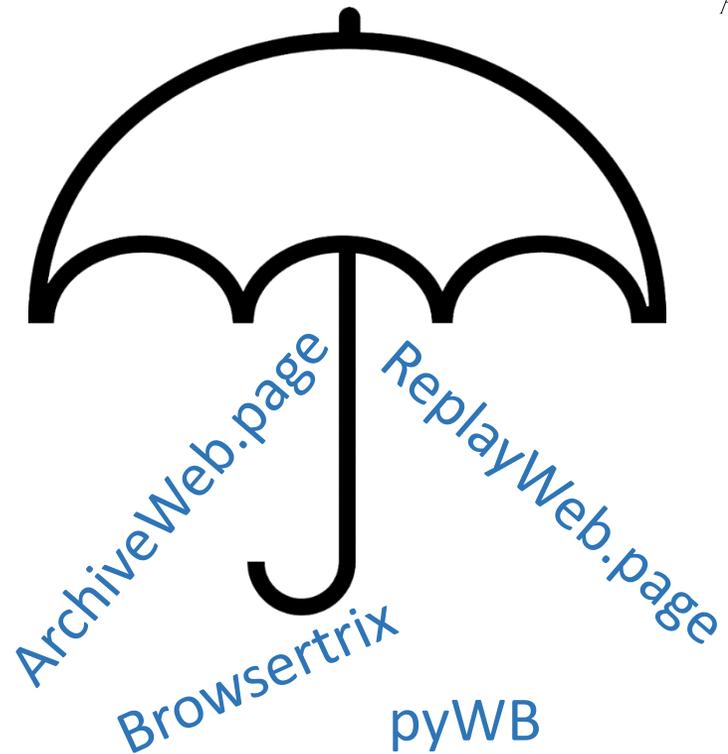
 STANFORD
UNIVERSITY PRESS

If you're an institution using our tools, let us know and we can add your logo here.

O Webrecorder.net

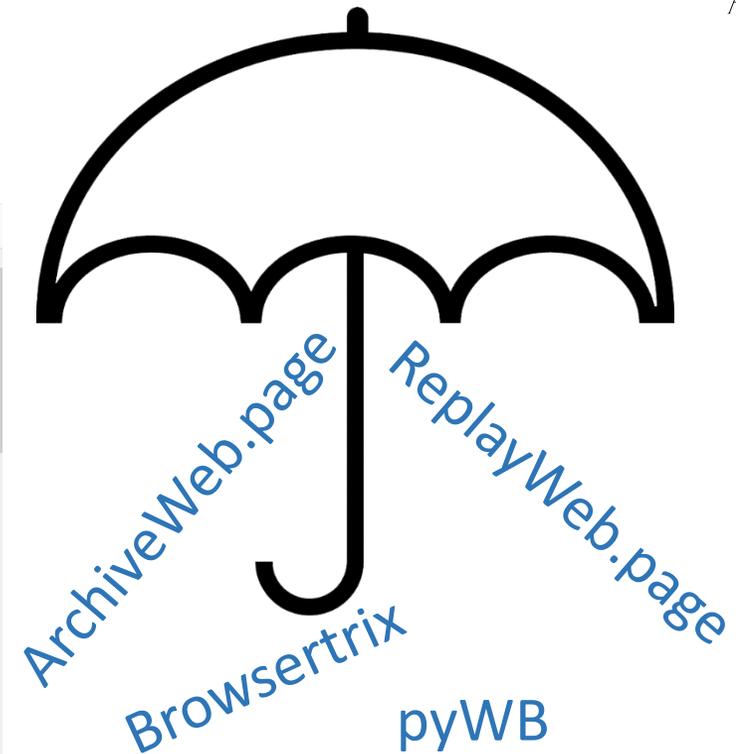
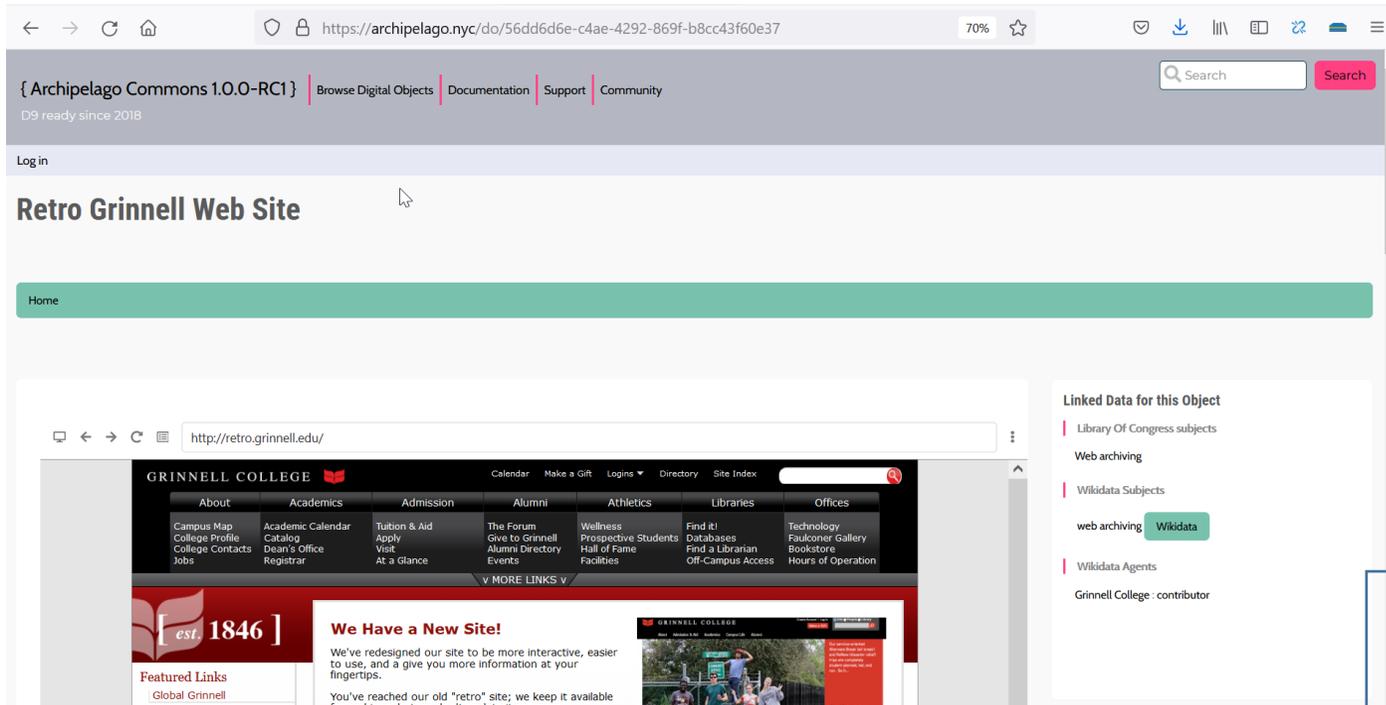
“Web archiving for all”

- Open source
- Centrado no utilizador não especializado
- Formatos normalizados
- Compatibilidade com arquivos da Web
- Apoio do International Internet Preservation Consortium (IIPC)



É aberto, gratuito, proativo e normalizado. Tudo o que se pode esperar.

O Webrecorder.net



Exemplo de integração do Webrecorder
 ArchiveWeb.page num repositório digital.
<https://archipelago.nyc/>

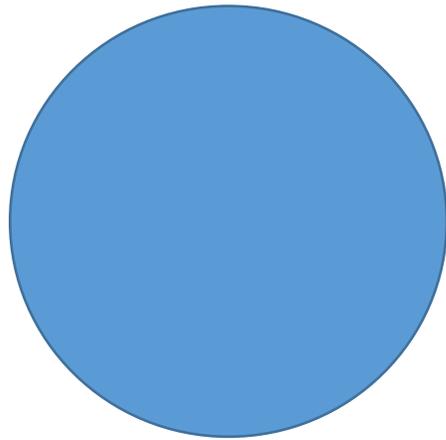
Este repositório digital inclui sites gravados com o ArchiveWeb.page, apresenta-os com metadados customizados e reproduz a gravação das páginas Web numa frame do próprio repositório.

A integração do ArchiveWeb.page em sites é possível, é fácil e está documentada no site Webrecorder.net. Veja a documentação e fale do assunto ao departamento de informática da sua organização:

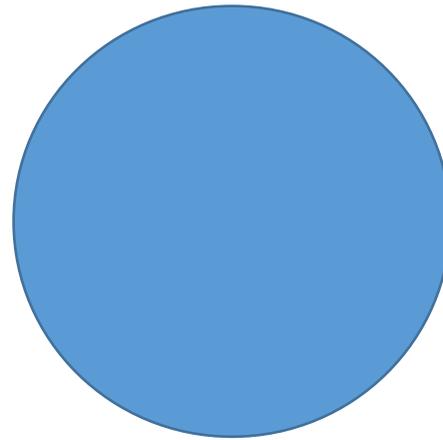
<https://replayweb.page/docs/embedding>

3 fases da Preservação da Web

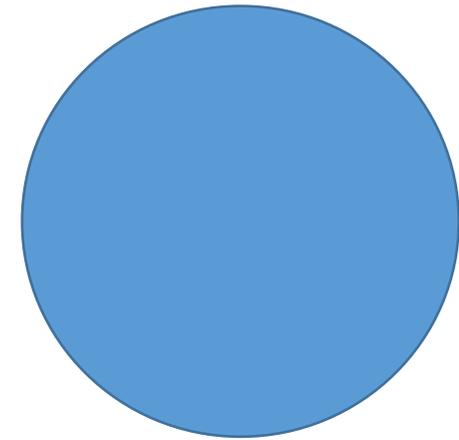
Neste slide e no seguinte mostramos que o fluxo geral que seguimos no Arquivo.pt pôde ser igualmente realizado com o ArchiveWeb.page. Coincidem também no formato que produzem, o WARC. O formato WACZ é uma especificação que contém mais elementos e que prepara tudo para a reprodução no ArchiveWeb.page. Mas no essencial o que está dentro de um WACZ é um WARC e os índices cdxj.



Gravar



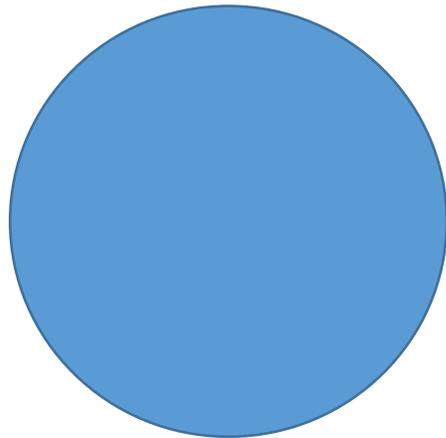
Armazenar



Reproduzir

ArchiveWeb.page

Neste slide e no anterior mostramos que o fluxo geral que seguimos no Arquivo.pt pôde ser igualmente realizado com o ArchiveWeb.page. Coincidem também no formato que produzem, o WARC. O formato WACZ é uma especificação que contém mais elementos e que prepara tudo para a reprodução no ArchiveWeb.page. Mas no essencial o que está dentro de um WACZ é um WARC e os índices cdxj.



Gravar



Armazenar

WACZ



Reproduzir

Formato WARC

- ISO 28500:2017
- Independente de plataforma
- Reproduzido por software específico (geralmente designado *wayback*)

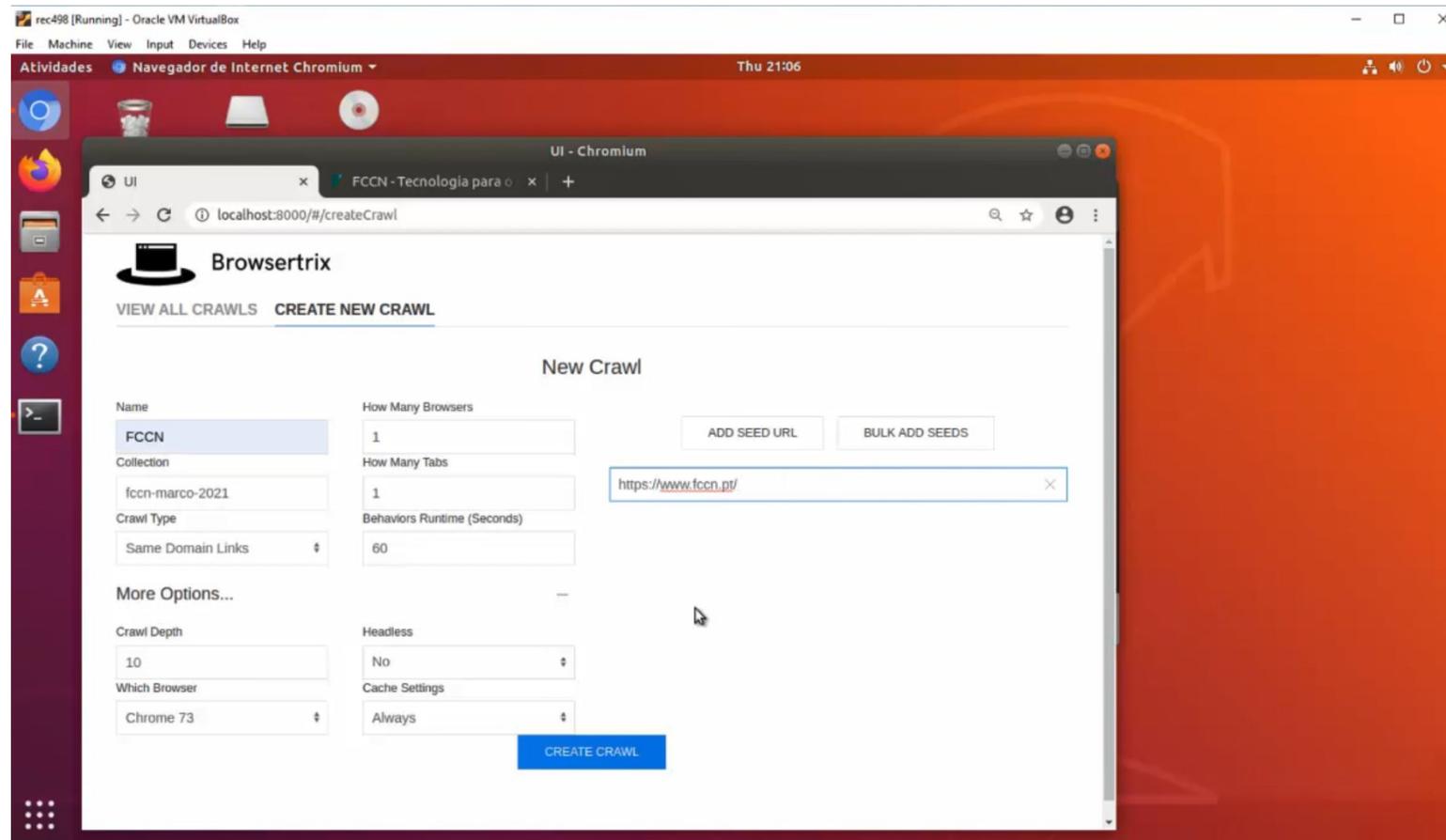
Insistimos no formato WARC que é, nesta altura, a “ferramenta” mais importante para os sistemas dos arquivos da Web.

Browsertrix

- Web crawler
- Baseado em Browser
- Vários perfis de recolha
 - Single page
 - Todos os links de um domínio
 - Profundidade personalizada
- Fácil de implementar para recolhas em pequena escala
- **[Guia para não especialistas: https://tinyurl.com/instalar-browsertrix](https://tinyurl.com/instalar-browsertrix)**
- [Documentação oficial: https://github.com/webrecorder/browsertrix](https://github.com/webrecorder/browsertrix)

Uma referência importante para responder à necessidade de gravar sites inteiros: de forma automática: o **Browsertrix**, também do <https://Webrecorder.net>

Demo – Gravar automaticamente um site inteiro com o Browsertrix-crawler



[Ver demo no Youtube](#)

Parte 3

Papel das comunidades locais
na preservação de conteúdos da Web

Caso de uso em que um arquivo municipal recolheu informação da Web sobre as vivências do Covid. A gravação de páginas Web foi um dos materiais recolhidos e depois enviados e partilhados com o Arquivo.pt que os disponibiliza nas sua interface de pesquisa.

O caso do Arquivo Municipal de Sines

XIII Encontro de História do Alentejo Litoral

DOCUMENTAR UMA PANDEMIA: O REGISTO DAS NOTÍCIAS DO CONCELHO DE SINES EM 2020-2021

DOCUMENTAR UMA PANDEMIA: O REGISTO DAS NOTÍCIAS DO CONCELHO DE SINES EM 2020-2021

Sines
Município

AM
ARQUIVO
MUNICIPAL

Colaboração para manter “vivo” o site do jornal O Corvo

Caso de uso de recolha de alta qualidade de páginas do jornal O Corvo, depois de ter chegado ao fim. Agora encontra-se no serviço Memorial do Arquivo.pt.

Jornal local O Corvo permanece vivo no Arquivo.pt

23 de Novembro, 2021

Samuel Alemão, fundador d'O Corvo, jornal digital e local sobre Lisboa que editou entre 2013 e 2019, quis preservar o projecto com dignidade e, por isso, trabalhou com a plataforma Arquivo.pt.

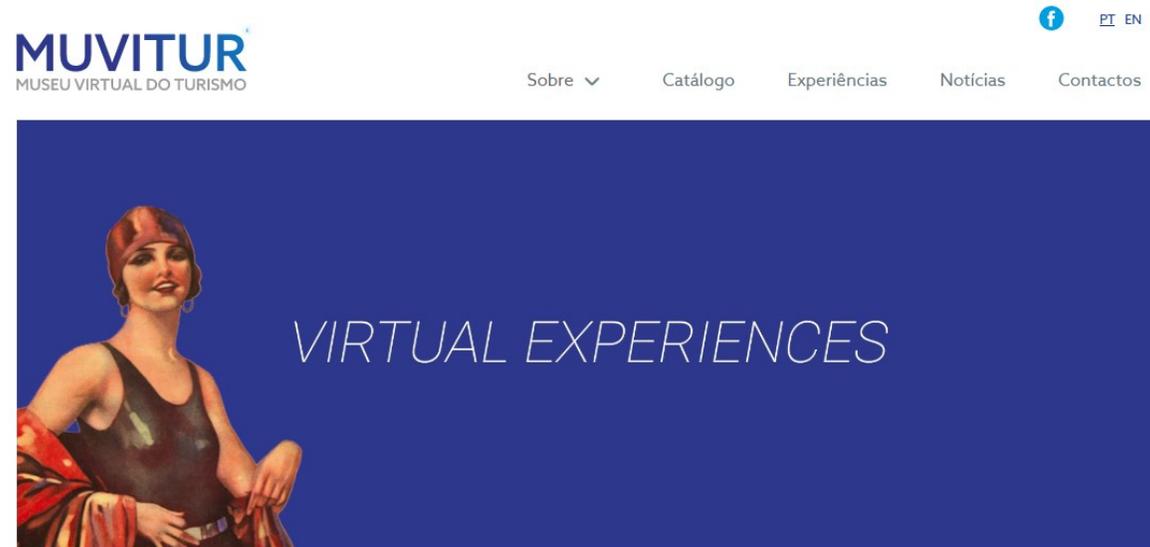


O jornal digital e local *O Corvo*, que marcou a cidade de Lisboa entre 2013 e 2019, transitou todo o seu arquivo para a plataforma portuguesa [Arquivo.pt](https://arquivo.pt). “É importante preservar o jornalismo que foi feito. A memória não pode ser rasurada, ainda para mais neste tempo em que vivemos e em que há uma poeira digital. Os artigos jornalísticos que foram feitos são documentos históricos”, disse Samuel Alemão, fundador do jornal, numa entrevista à revista *Lisboa Para Pessoas* em Julho.

<https://mediaalternativos.pt/2021/11/o-corvo-jornal-lisboa-arquivo-pt/>

Colaboração com o Museu Virtual do Turismo

Colaboração com o MUVITUR que resultou na melhoria da preservação de conteúdos acerca de turismo em Portugal. Procurou-se integrar de alguma forma esses conteúdos no Nyron, sistema que apresenta conteúdos numa interface Web.



<https://muvitur.eshte.pt/>

Livro sobre os Arquivos da Web para download

Estado da arte sobre arquivos da Web, cujo editor principal é Daniel Gomes, gestor do Arquivo.pt

Novidade!

Conheça o novo livro

The Past Web

Exploring web archives

arquivo.pt/livro

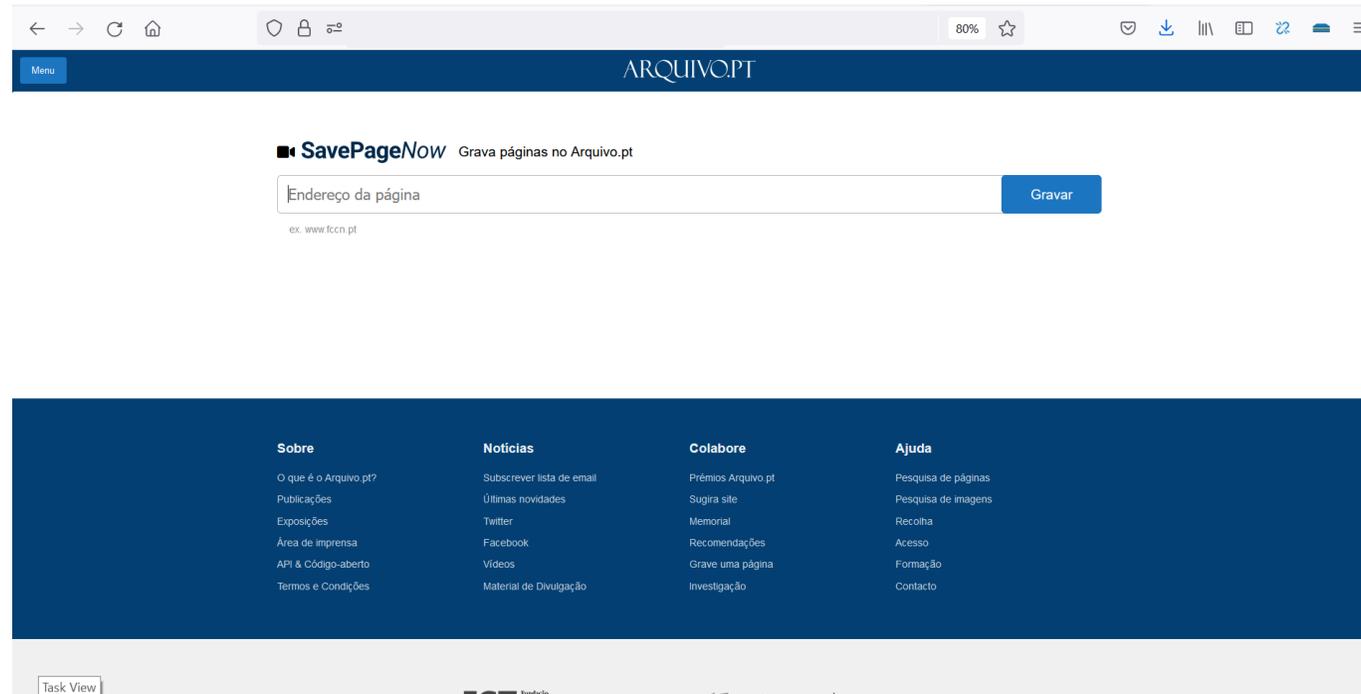
FCCN UNIDADE DA FCT
Tecnologia para o Conhecimento



<https://arquivo.pt/livro>

Gravar página na hora no Arquivo.pt – SavePageNow

Novidade principal de 2021 que responde à necessidade que os utilizadores têm manifestado: “Posso gravar uma página para ficar preservada no Arquivo.pt?” Sim, com este serviço pode gravar na hora um conteúdo Web que ache importante. Os conteúdos ficam automaticamente integrados no Arquivo.pt onde poderá vê-los depois.



Receber novidades: <https://arquivo.pt/inscrever>

Conclusão

- Grave conteúdos em formato normalizado
- Reutilize e analise em contexto local
- Contribua (juntamente com os arquivos da Web) para a preservação da memória das comunidades locais

Obrigado.

contacto@arquivo.pt



Extra: Instale no seu browser e experimente a “Extensão Arquivo.pt”

O que é e para que serve

- Foi 2º classificado do Prémio Arquivo.pt 2020 >> [apresentação dos autores](#)
- Permite aceder consultar o Arquivo.pt diretamente no seu browser
- Muito útil para incluir no seu fluxo de trabalho, sempre que quer aceder a páginas antigas da Web

Onde funciona e onde encontrar

- Google Chrome, Edge, Chromium em Linux, Opera e Brave >> [ver na Chrome Web Store](#)
- Firefox >> <https://addons.mozilla.org/en-GB/firefox/addon/arquivo-pt/>
- Safari: não disponível. Solução: para computadores Mac instale o Edge e no Edge instale a extensão.