

Ciclo de Webinars

Arquivar a Web: faça você mesmo

Como preservar as referências na Wikipédia?



Breve apresentação do Arquivo.pt

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

Objetivo

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

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?

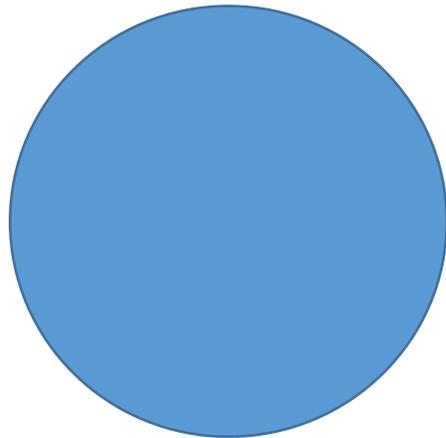
The screenshot shows the Portuguese Wikipedia page for "Seleção Portuguesa de Futebol" as it appeared on November 8, 2013. The page title is "Seleção Portuguesa de Futebol". The main text describes the national team's history, mentioning its debut in 1921 and its participation in the 2006 FIFA World Cup. A sidebar on the right contains a list of external links and categories, including "Alcunhas", "Associação", "Confederação", "Patrocinadores", "Material desportivo", "Treinador", "Capitão", and "Mais participações". The page is part of the ARQUIVOPT project.

Página da Wikipédia sobre a Federação Portuguesa de Futebol em 2013

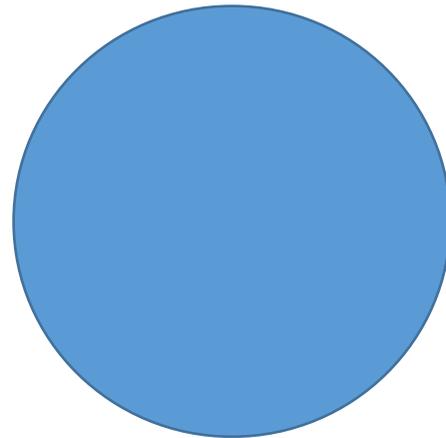
The screenshot shows the Portuguese Wikipedia page for "Seleção Portuguesa de Futebol" as it appeared on December 25, 2020. The page title is "Seleção Portuguesa de Futebol". The main text is more detailed, mentioning the team's participation in the 2016 UEFA Euro and the 2018 FIFA World Cup. A sidebar on the right contains a list of external links and categories, including "Alcunhas", "Confederação", "Material desportivo", "Treinador", "Capitão", "Mais participações", and "Melhor artilheiro". The page is part of the ARQUIVOPT project.

Página da Wikipédia sobre a Federação Portuguesa de Futebol em 2020

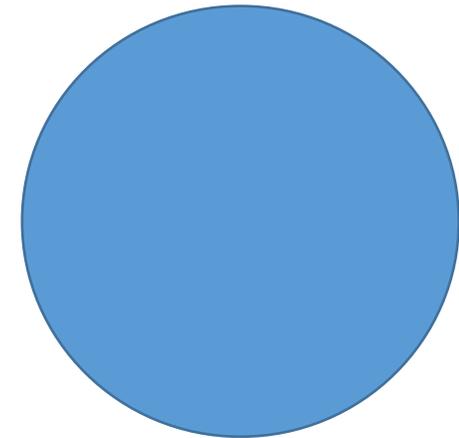
3 fases da Preservação da Web



Gravar



Armazenar



Reproduzir

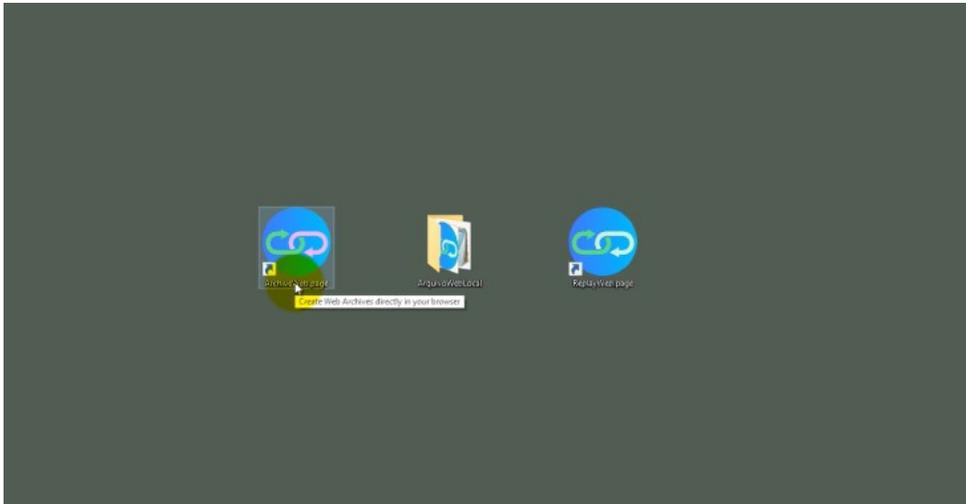
Parte 1

Tutorial

ArchiveWeb.page

Fluxo de trabalho

Gravar >> Armazenar >> Reproduzir



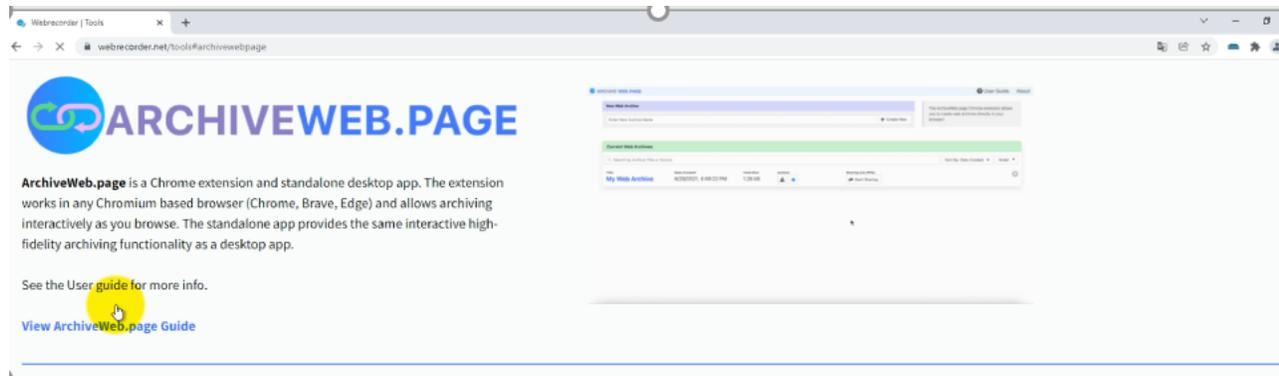
Este vídeo mostra o fluxo de trabalho desejável num “Arquivo Web Local”:

- 1- Grava-se páginas no Webrecorder
- 2- Faz-se o download para o meu “ArquivoWebLocal” (trata-se de uma pasta onde simplesmente armazeno os meus ficheiros WARC)
- 3- Lê-se os ficheiros gravados quando e onde se quiser utilizando o software adequado (neste caso pode ser o ReplayWeb.page). Pode ser também o ArchiveWeb.page).

https://youtu.be/bHu_8Av1ESA?t=269

O Arquivo.pt também lê os ficheiros WARC, por isso havendo interesse podem ser integradas gravações feitas com o ArchiveWeb.page.

Instalar Extensão ArchiveWeb.page no browser



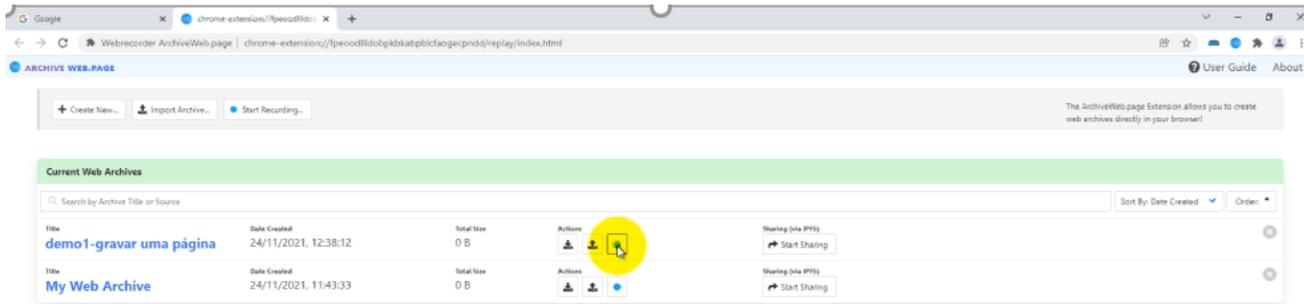
Este vídeo mostra a instalação da extensão de browser (Chrome ou baseados em Chrome, tal como o Brave e o Edge) para páginas na hora. Uma de cada vez. Clique a clique.

Para instalar em iOS: instalar o Edge; em seguida, instalar a extensão no Edge.

https://youtu.be/bHu_8Av1ESA?t=429

3

Gravar uma página



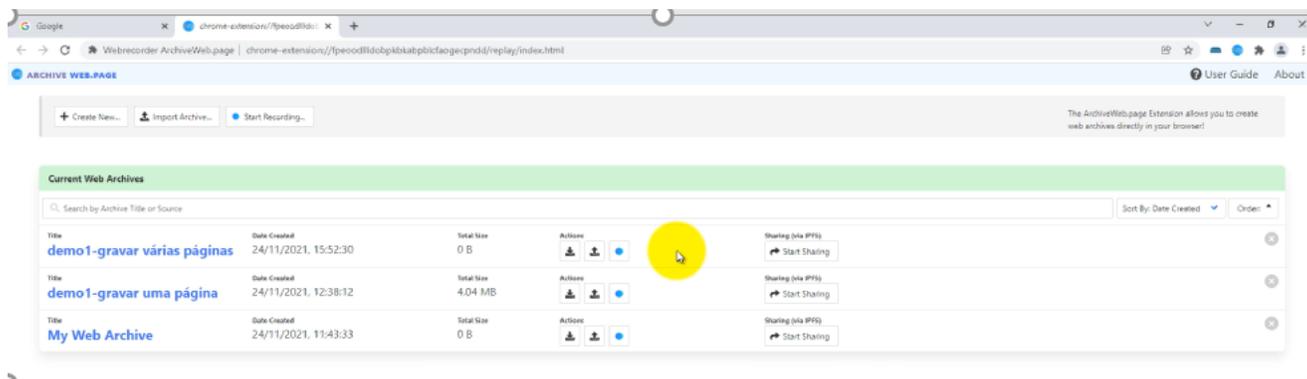
https://youtu.be/bHu_8Av1ESA?t=554

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.

O exemplo utilizado é o da Página da Câmara Municipal de Lisboa.

4

Gravar várias páginas

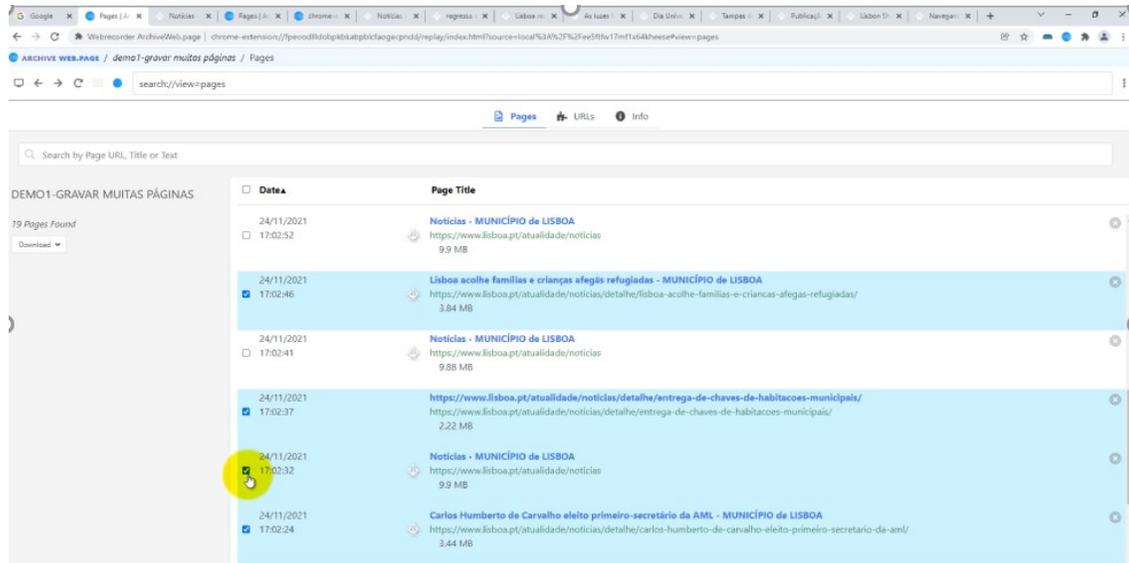


https://youtu.be/bHu_8Av1ESA?t=769

Este vídeo repete o anterior mas com mais cliques, em várias páginas, usando a extensão ArchiveWeb.page. O processo é o seguinte: uma ligação 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.

5

Gravar muitas páginas



https://youtu.be/bHu_8Av1ESA?t=909

Este vídeo amplia o procedimento com mais cliques, em páginas específicas, usando a extensão ArchiveWeb.page. Uma ligação de cada vez. Clique a clique. E acrescenta a recomendação: abrir nova tab, abrir nova tab.

Sugestão para o caso da Wikipédia: escolher página sobre a Covid-19 que tem mais de 1000 referências externas. Como fazer para gravar todas estas referências? Se for com o ArchiveWeb.page terá de ser **ABRIR NOVA TAB UMA A UMA** e nunca fechar qualquer dessas tabs. Fechar uma tab que se abriu faz parar o processo de gravação. Se parar podemos sempre continuar a fazer gravações na mesma sessão de gravação e acrescentar novas gravações. Por isso é que este processo de gravação "à unha" tem de ser muito controlado por quem grava para não perder o norte. Vantagem: qualidade garantida. Temos a certeza de que tudo ficou gravado. Alternativa: outras ferramentas de gravação automática que vamos referir mais adiante, como por exemplo, o Browsertrix.

6

Gravar Twitter

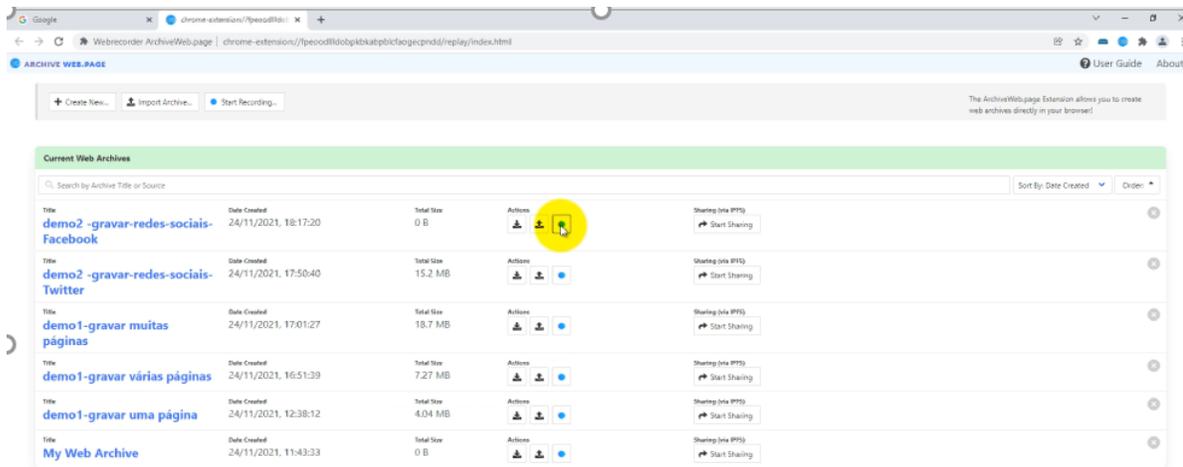


Este vídeo mostra a utilização da extensão para gravar Twitter. Correr verticalmente, alternativa com ou sem “piloto automático”. Grava o Twitter por for, sem login. Funciona bem. Os vídeos reproduzem. Posso personalizar a gravação e seguir links que interessem.

Recomendação: gravar por fora (sem fazer login) estes canais, seguindo a regra, conteúdos públicos gravam-se de forma pública.

https://youtu.be/bHu_8Av1ESA?t=1099

Gravar Facebook sem login



https://youtu.be/bHu_8Av1ESA?t=1284

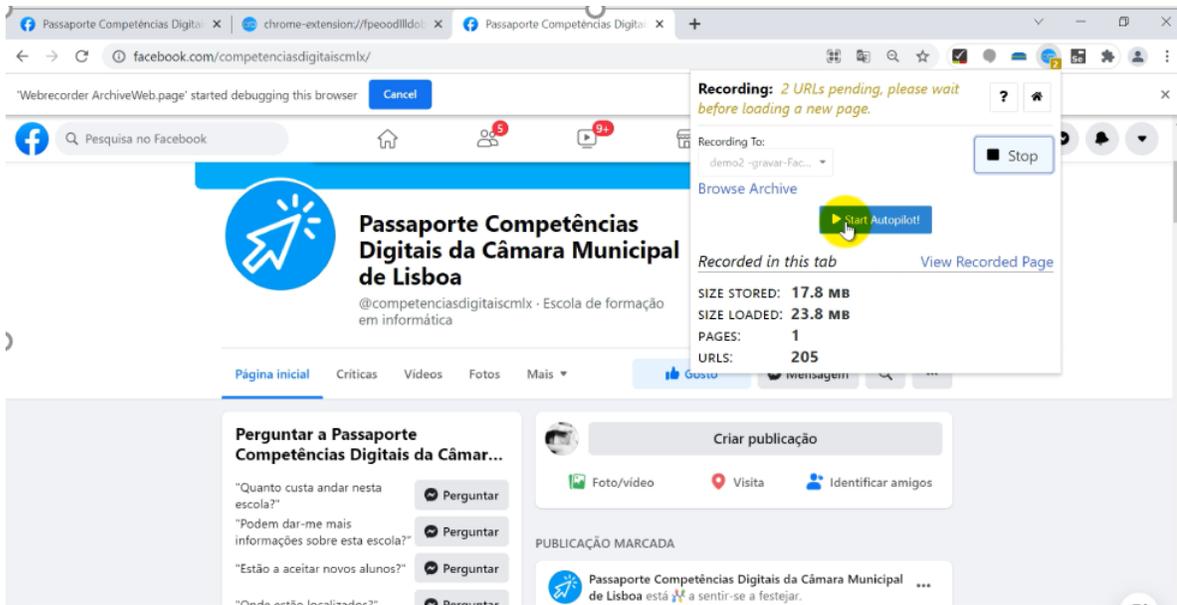
Este vídeo mostra a utilização da extensão para gravar Página do Facebook sem fazer login. Grava bem, tal como no Twitter. Os conteúdos “pesam” mais.

Com o Facebook há dificuldade em reproduzir os conteúdos, apesar terem ficado gravados.

A preservação depende do que as plataformas deixam ou não fazer.

As instituições devem ter isso em conta quando publicam conteúdos. Devem garantir formas alternativas de preservarem os conteúdos produzidos.

Gravar Facebook com login



Este vídeo mostra a utilização da extensão para gravar Página do Facebook com login.

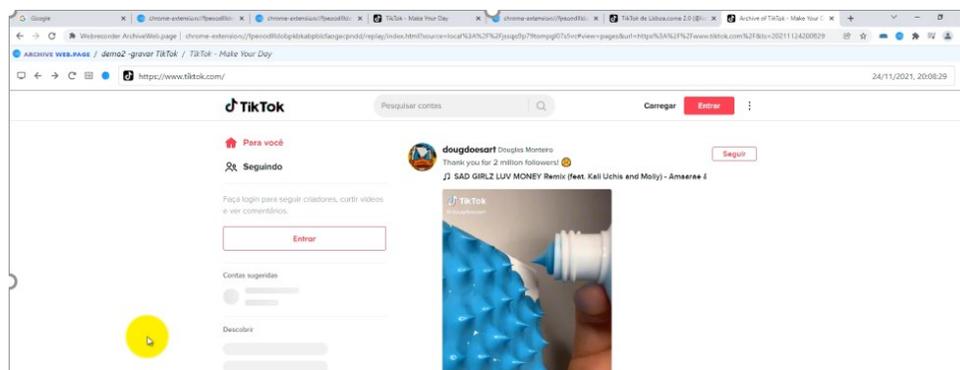
Atenção: recomendada-se só para uso privado ou interno nas instituições.

Grava bem, tal como no Twitter. Os conteúdos “pesam” mais. Mas depois não reproduz bem. Na reprodução não faz o scroll down para conteúdos gravados mais abaixo na sequência de publicações.

https://youtu.be/bHu_8Av1ESA?t=1468

9

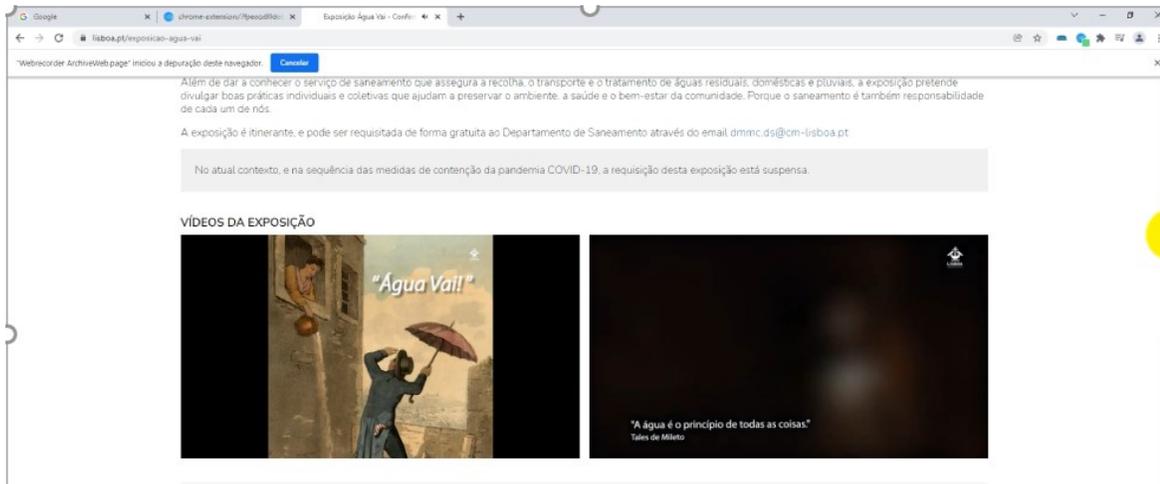
Gravar Tiktok



Este vídeo mostra a utilização da extensão ArchiveWeb.page para gravar outras redes sociais, tais como o Tiktok.

https://youtu.be/bHu_8Av1ESA?t=1672

Gravar vídeos do Youtube embebidos

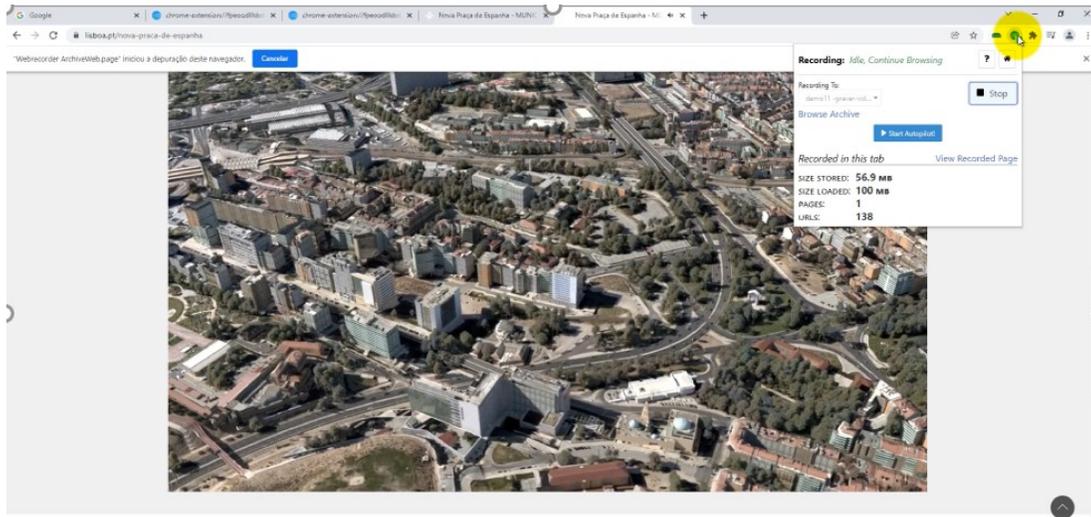


https://youtu.be/bHu_8Av1ESA?t=1752

Este vídeo mostra a utilização da extensão para gravar vídeos embebidos nas páginas de Websites. Neste caso funcionou bem, pois o software que lê nesta versão, à data de 2021, é compatível com o Youtube. Pode mudar no futuro. Que fazer quando as compatibilidades mudam ao longo do tempo? A gravação de vídeos do Youtube deve ser complementada com outras formas de preservação destes conteúdos.

Nos sites institucionais é útil publicar o link para os vídeos institucionais do Youtube.

Gravar vídeos em outros formatos



https://youtu.be/bHu_8Av1ESA?t=1872

Este vídeo mostra a utilização da extensão para gravar vídeos embebidos nas páginas de Websites. Neste caso funcionou bem.

É difícil gravar e reproduzir vídeo pois há muitos formatos. Nem sempre o software que lê é compatível. Ao longo do tempo os formatos mudam e o que funcionava antes deixa de funcionar depois.

A preservação digital deve fazer-se complementarmente entre várias instituições e utilizando várias técnicas. O arquivo da página Web é útil pois fornece o contexto em que o vídeo foi publicado.

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)

0 Webrecorder.net

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



 UKWA
UK WEB ARCHIVE

 DocNow

 ARQUIVO.PT

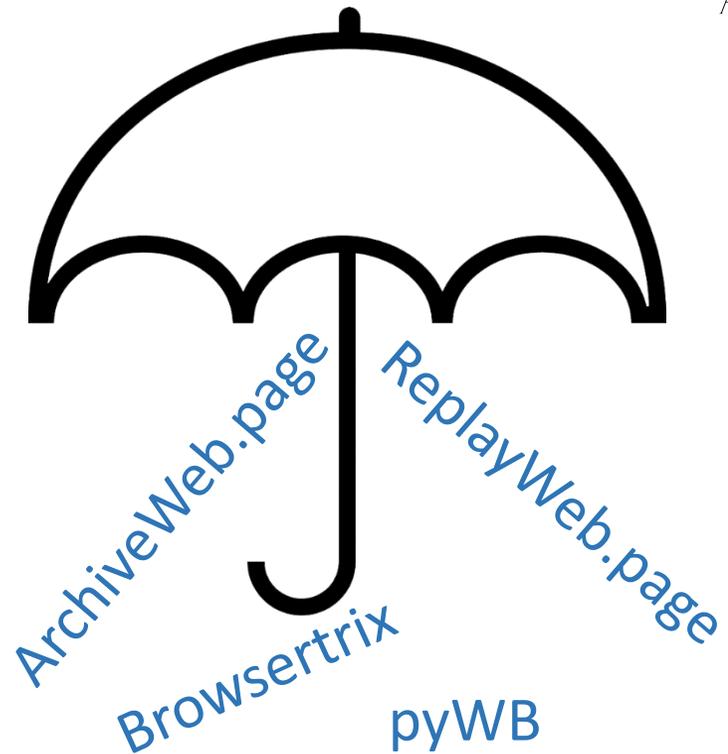
 STANFORD
UNIVERSITY PRESS

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

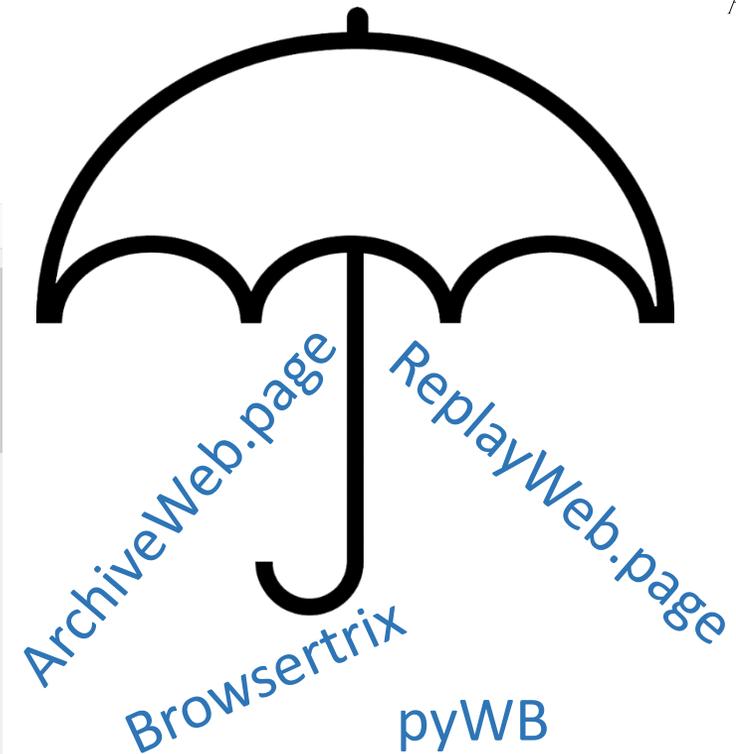
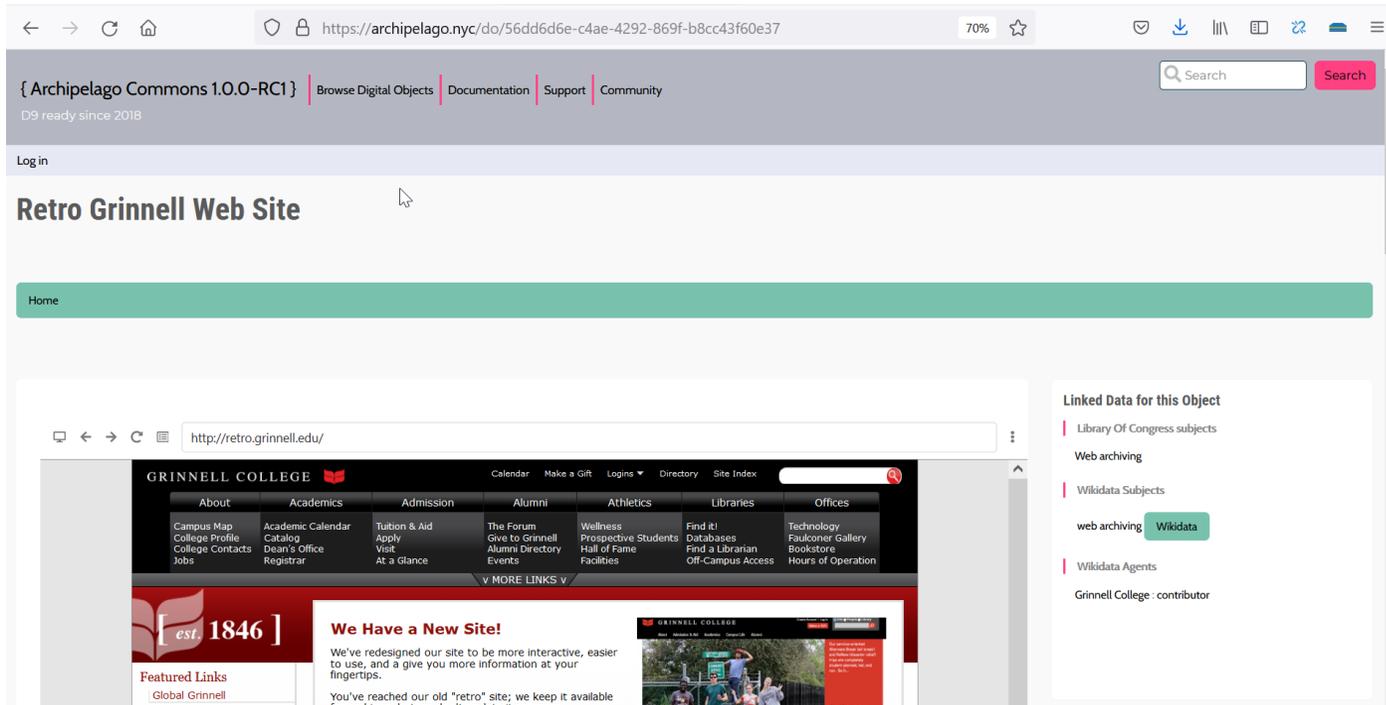
O Webrecorder.net

“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)

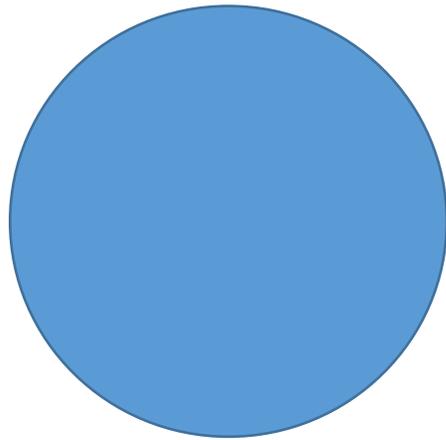


O Webrecorder.net



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

ArchiveWeb.page



Gravar



Armazenar

WACZ



Reproduzir

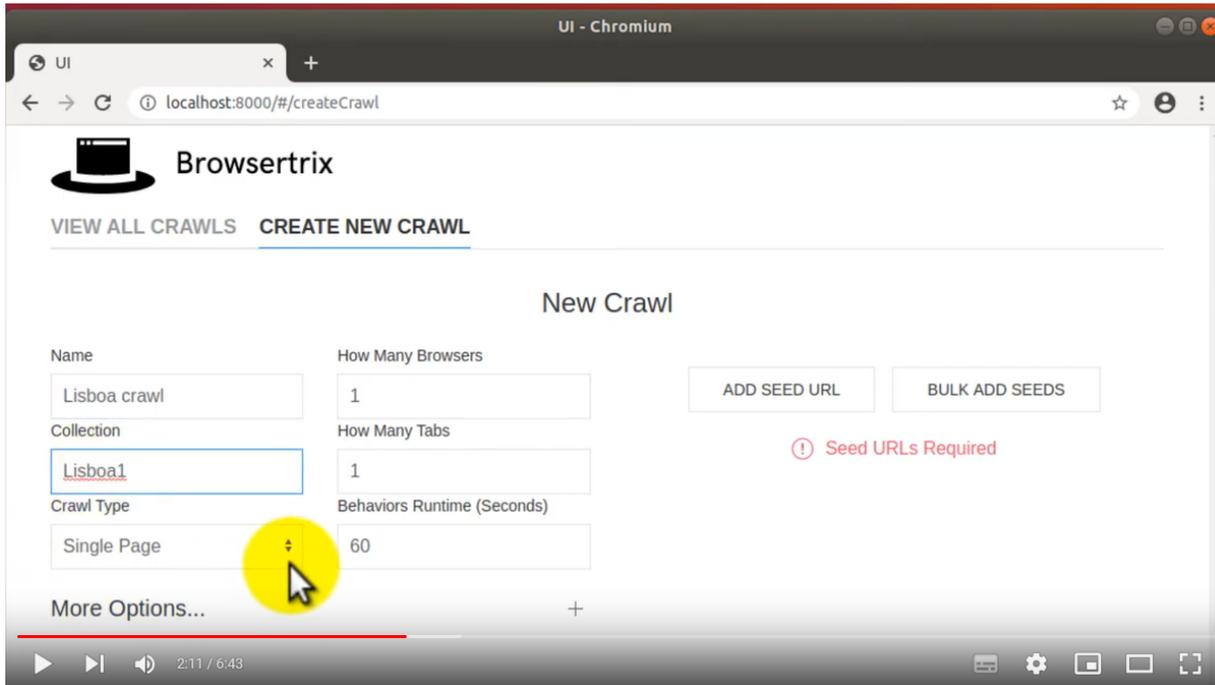
Formato WARC

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

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>**
- Documentação oficial: <https://github.com/webrecorder/browsertrix>

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



O Browsertrix, como o próprio nome sugere, é baseado em browser. Ao gravar os conteúdos, fá-los passar por um browser. O curador pode ver em tempo real a gravação nesse “browser remoto” e até interagir, clicando em botões e provocando comportamentos numa página, a fim de que tudo fique gravado

<https://youtu.be/7JNoMo302NA>

Parte 3

Papel das comunidades locais na preservação de conteúdos da Web

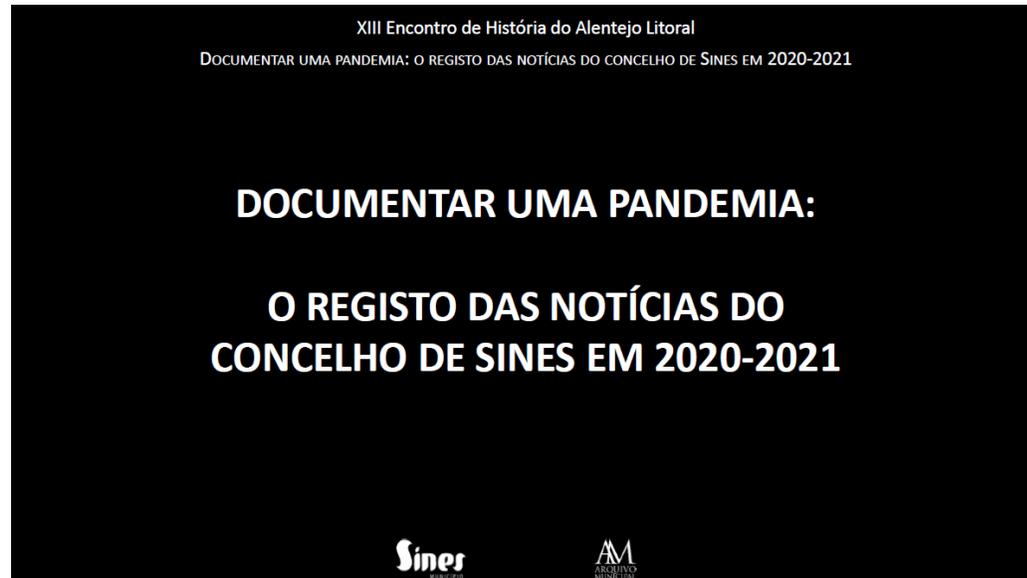
O papel das comunidade é importantíssimo para preservar bem a Web.

Fortalece a comunidade de prática.

Melhora efetivamente a qualidade da preservação dos sites institucionais ou de âmbito local.

Há milhares de sites portugueses que precisam do cuidado da comunidade para ficarem gravados – muitos são esquecidos – e para ficarem gravados com qualidade.

O caso do Arquivo Municipal de Sines



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.

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](#).

Para ver como funciona procure num motor de busca ou aceda ao site do jornal <https://ocorvo.pt/>

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

Colaboração com o Museu Virtual do Turismo

MUVITUR
MUSEU VIRTUAL DO TURISMO

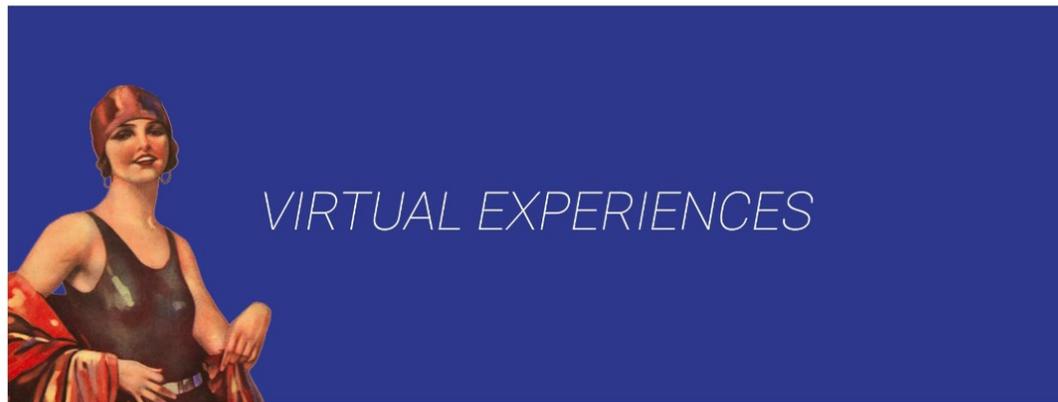
Sobre ▾

Catálogo

Experiências

Notícias

Contactos



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

Novidade!

Conheça o novo livro

The Past Web

Exploring web archives

arquivo.pt/livro

FCCN UNIDADE DA FCT
Tecnologia para o Conhecimento



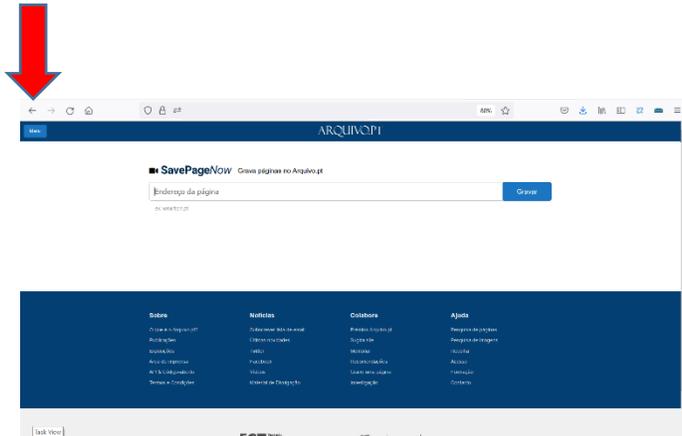
<https://arquivo.pt/livro>

Gravar página na hora no Arquivo.pt – SavePageNow (1)

The image shows a browser window displaying the Arquivo.pt website. A red arrow points to the 'Menu' button in the top navigation bar. The main content area features the 'SavePageNow' interface, which includes a search bar labeled 'Endereço da página' and a 'Gravar' button. Below the search bar, there are sections for 'Sobre', 'Notícias', 'Colabore', and 'Ajuda'. A sidebar menu is open on the left, showing options like 'English', 'Copiar link', 'Páginas', 'Imagens', 'Grave páginas', and 'Sobre'. A red arrow points to the 'Grave páginas' option in the sidebar. The sidebar also shows a calendar for August 6th and buttons for 'Páginas', 'Imagens', and 'Narrativa'. The main content area also displays a 'Prémio Arquivo.pt 2022' section.

<https://arquivo.pt/savepagenow>

Gravar página na hora no Arquivo.pt – SavePageNow (2)



O SavePageNow é serviço que o Arquivo.pt criou para a que a comunidade possa participar ativamente na gravação de páginas Web com melhor qualidade.

O SavePageNow é como o “ArchiveWeb.page” mas funciona do lado do Arquivo.pt, o que ainda é mais fácil.

Grave clique a clique, uma ligação de cada vez.
Em menos de 48 horas os conteúdos estarão acessíveis no Arquivo.pt.

<https://arquivo.pt/savepagenow>

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

