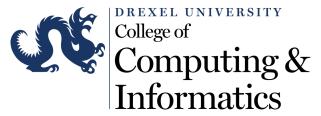
SAVING ADS

Assessing and Improving Web Archives' Holdings of Online Advertisements

Christopher Rauch (cr625@drexel.edu), Travis Reid (treid003@odu.edu)

Advised by Profs. Mat Kelly, Alex H. Poole, Michael Nelson, Michele C. Weigle







mkelly@drexel.edu ahp56@drexel.edu@drexel.edu mln@cs.odu.edu mweigle@cs.odu.edu



Introduction

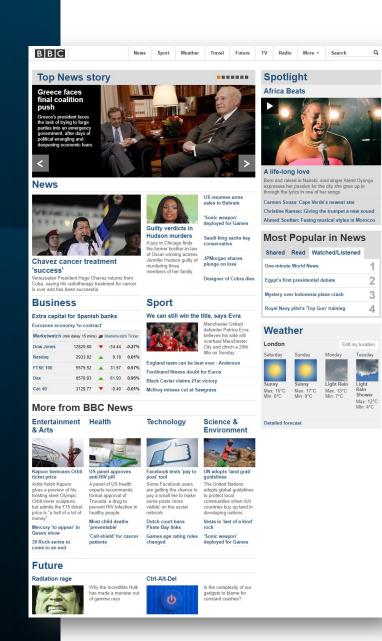
- The World Wide Web is a staging ground for transformations in marketing, advertising, and cultural representation.
- Online advertisements capture the essence of their times by reflecting prevailing economic trends, societal norms, and cultural shifts.
- The shift from traditional to digital marketing introduced challenges in preserving inherently ephemeral digital artifacts.

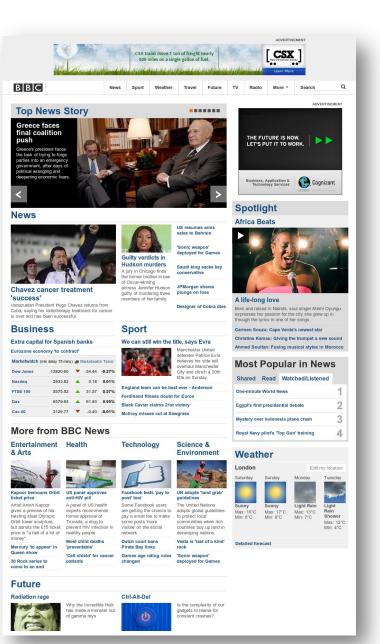


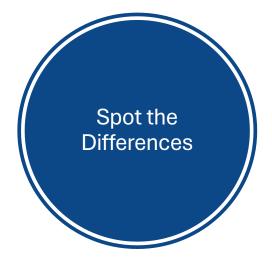
The Saving Ads Project

- Collaborative initiative between Drexel University and Old Dominion University
- Supported by a grant from the Institute of Museum and Library Services (IMLS)
- Focuses on archiving digital marketing materials in their broader, networked context
- Examines the evolution of online advertising, the complexities of archiving such fluid content, and the strategies developed to address these challenges
- Highlights the significance of archiving digital marketing to document this era of digital 'small' history for future generations
- Underscores the broader implications of these archival efforts for subsequent scholarship

Does it Matter?

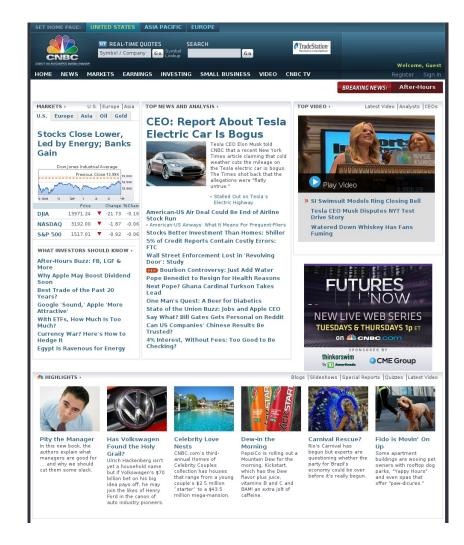


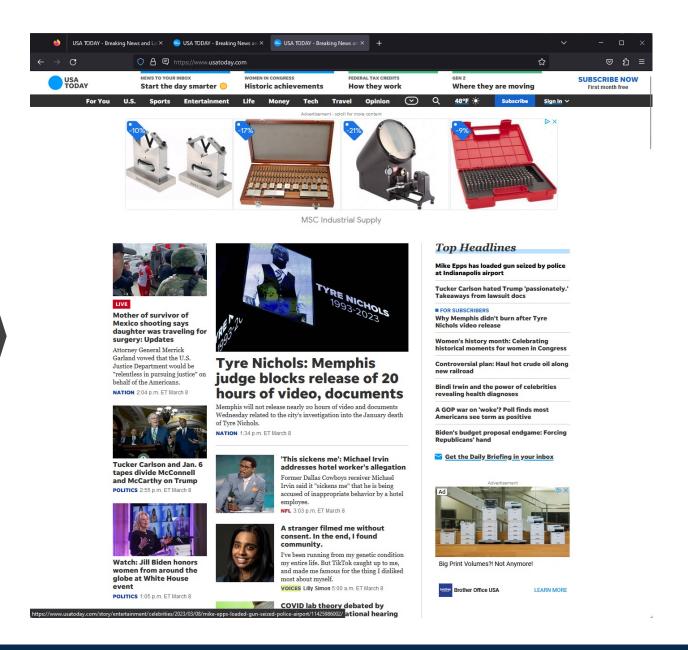




April 25, 2024









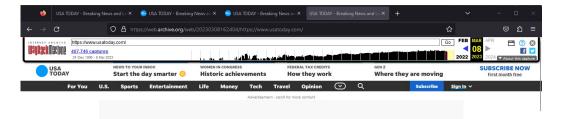
USA TODAY - Breaking News and Low X USA TODAY - Breaking News an X USA TODAY - Breaking News

MSC Industrial Supply

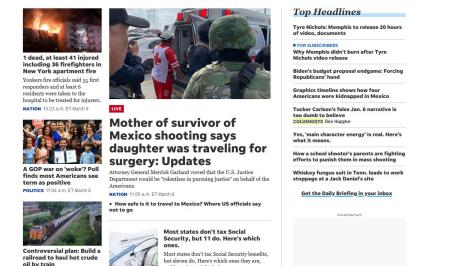


container + data-item for ad collection

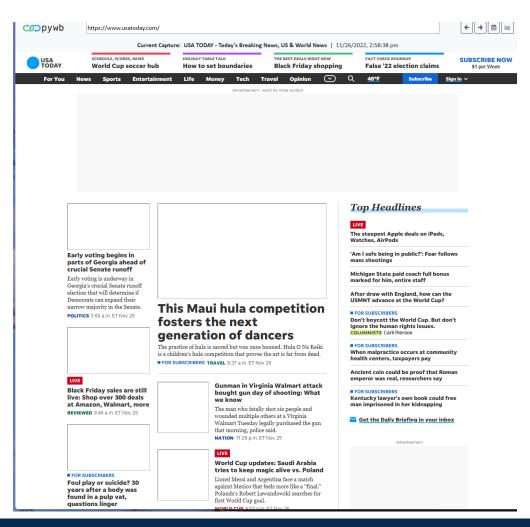
</aside>



Internet Archive



Sized container but no data-item



Retrieved from common crawl

Out of Context

• not possible to retrieve the containing page (small chance at reproduction)



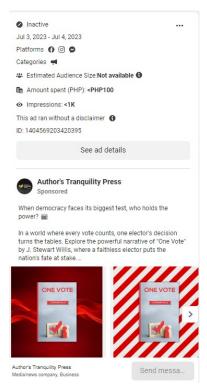




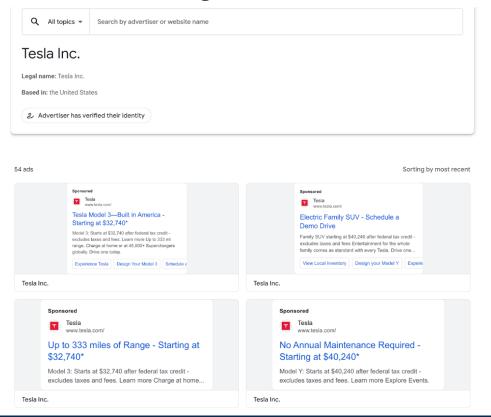
Out of Context

Vendor archives

Facebook



Google





Out of Context

• Industry Archives: AdBeat.com

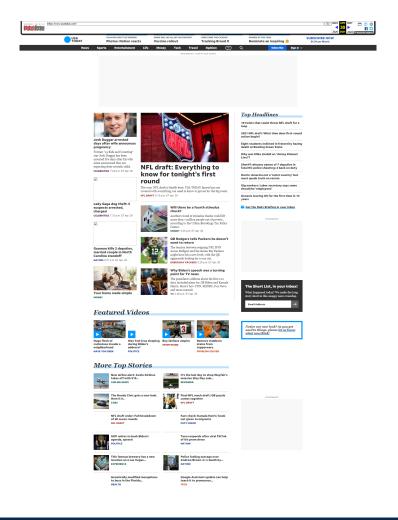




Cost Benefit

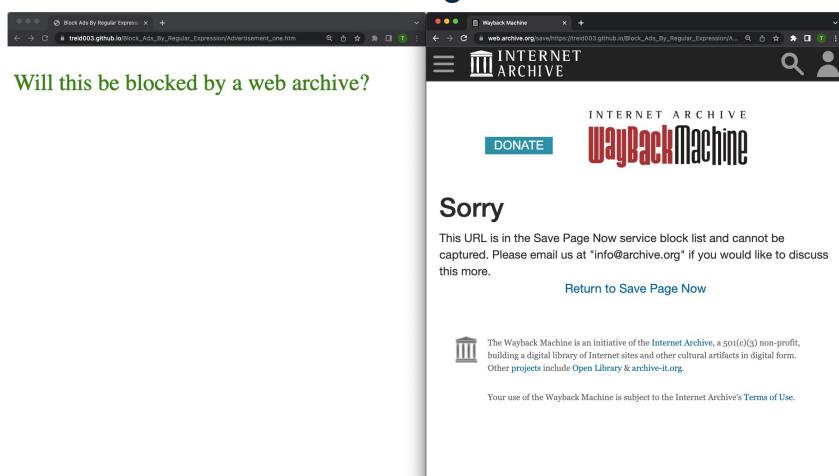






Archiving Issues

Web Archives Blocking Ads



URLs that include an ad related name were blocked by Internet Archive's Save Page Now last year before August 2023

File name: Advertisement_one.htm

URI-R: https://treid003.github.io/Block_Ads_By_Regular_Expression/Advertisement_one.htm

Ad Related Filename Blocked

An ad related file name previously caused a URL to be blocked by Internet Archive's Save Page Now

- displayAds.js
- imgAd.jpg
- videoAd.mp4
- webAd.png

File name: imgAd.jpg

URI-R: https://savingads.github.io/files/imgAd.jpg



Sorry

This URL is in the Save Page Now service block list and cannot be captured. Please email us at "info@archive.org" if you would like to discuss this more.

Return to Save Page Now

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form.

Other projects include Open Library & archive-it.org.

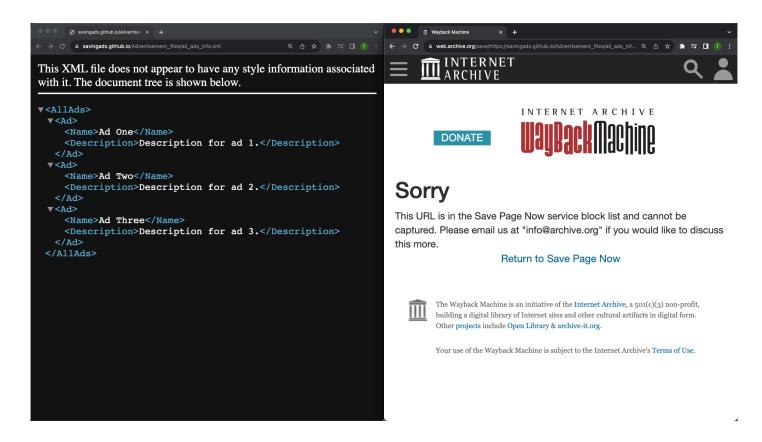
Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

Ad Related Directory Name Blocked

Example directory names that can cause a URL to be blocked:

- displayAds
- videoAd
- webAd
- ads
- Advertisement files

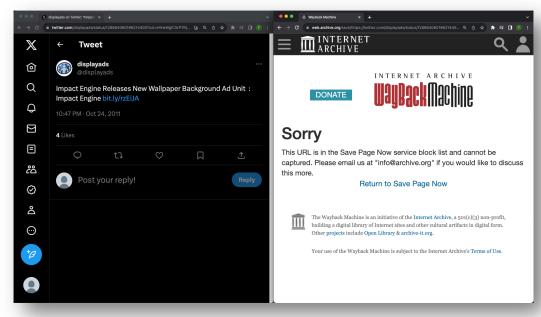
Directory name: Advertisement_files URI-R: https://savingads.github.io/Advertisement_files/all_ads_info.xml



Social Media Accounts Blocked

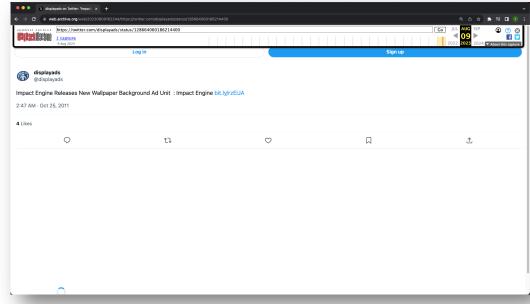
Social media accounts with an ad related username were also getting blocked, but after talking to Wayback Machine staff these accounts can now be archived

Account name: displayads
Tweet blocked during July 2023



URI-R https://twitter.com/displayads/status/128664060186214400?cxt=HHwWgICSrP7RjckDAAAA

Account name: displayads
Tweet archived during August 2023

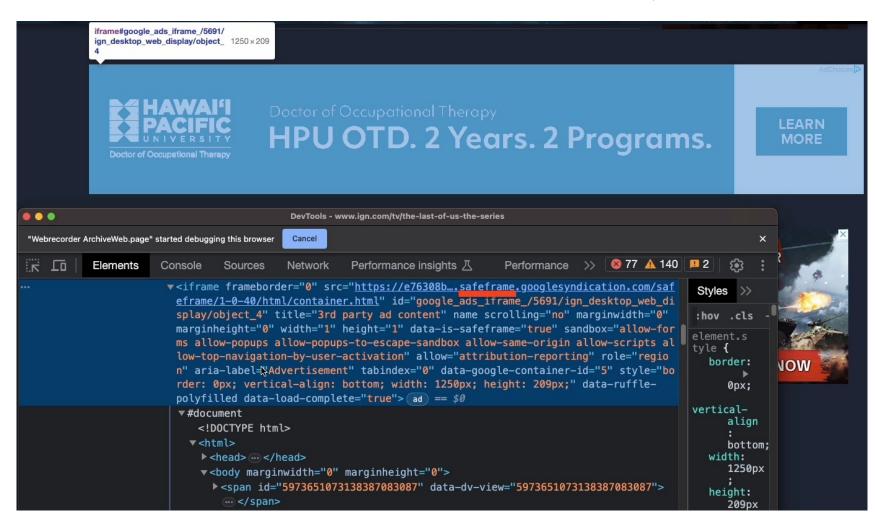


https://web.archive.org/web/20230809162244/https://twitter.com/displayads/status/12866406018621440

URI-M:

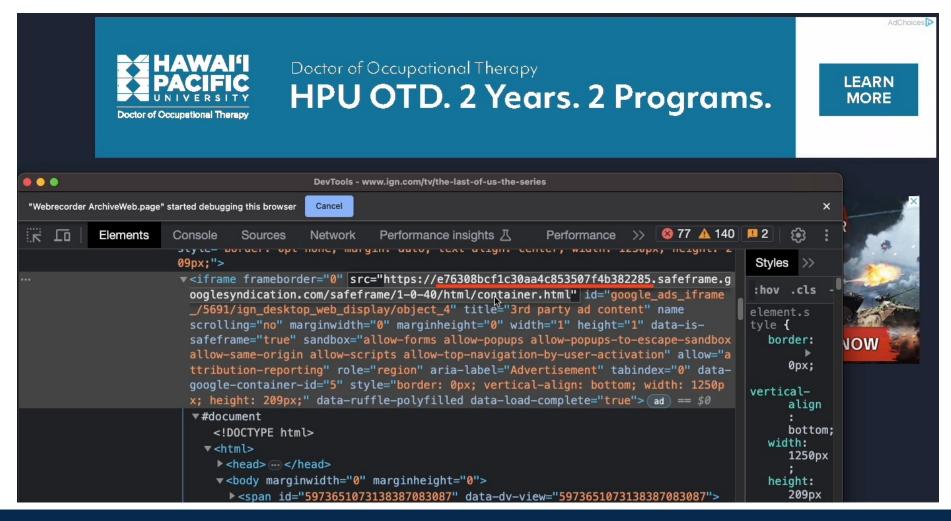
Replay Issues

Ad Delivery Networks and Replay (Google SafeFrame)



A Google SafeFrame is an <iframe> that is used to load web advertisements

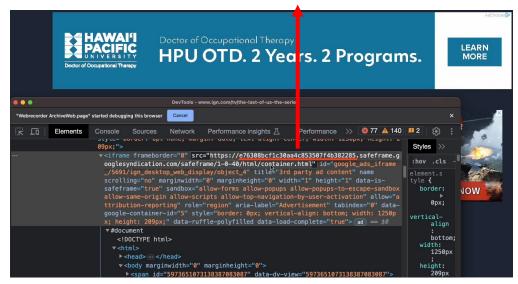
Most Google SafeFrames have a URL with a random value in the subdomain



Google SafeFrame Random Values

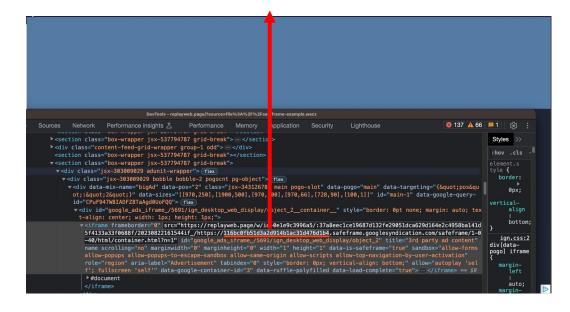
Live Web Page

Random value in subdomain: e76308bcf1c30aa4c853507f4b382285



Archived Web Page

Random value in subdomain: 116bc0fb51d3a2d914b1ac31d476d1b4

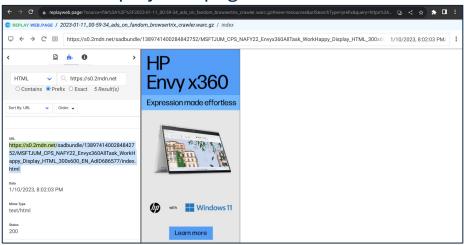


Processes and Tools: Finding Archived

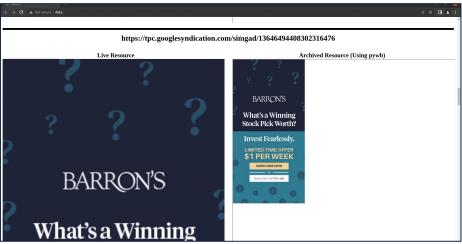
Ads

April 25, 2024

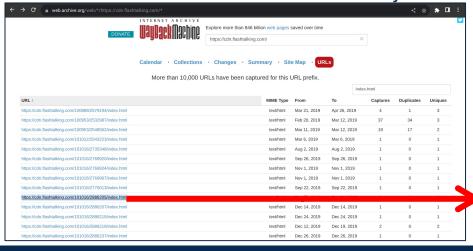
ReplayWeb.page's URL Search

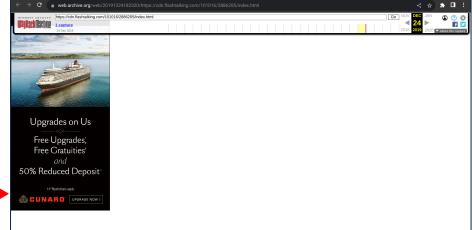


Our Tool For Displaying Ads From A WARC file



Wayback Machine's URL Search





Conclusion

https://sites.google.com/view/treid803/home

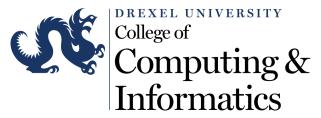
Discussion/Questions

SAVING ADS
Assessing and Improving Web Archives' Holdings of Online Advertisements

treid003.github.io

Christopher Rauch (cr625@drexel.edu), Travis Reid (treid003@odu.edu)

Advised by Profs. Mat Kelly, Alex H. Poole, Michael Nelson, Michele C. Weigle







mkelly@drexel.edu ahp56@drexel.edu@drexel.edu mln@cs.odu.edu mweigle@cs.odu.edu