

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



DREXEL UNIVERSITY
College of
**Computing &
Informatics**



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

BBC

Top 1

Greece's final push
Greece's push for the final task of the government's political agenda is deepening

News

Chavez's succession
Venezuelan President Hugo Chavez, saying he is over and hand over the

Business

Extra capital
Eurozone economic
Marketwatch
Dow Jones
Nasdaq
FTSE 100
Dax
CAC 40

More Entertainment & Arts

Kanye West's new album
Artist Kanye West's new album 'The Life of Pablo' gives a preview of his new sound, but admits the price is 'a little bit of money'

Future

Radiation

[illegible]



SET HOME PAGE: UNITED STATES ASIA PACIFIC EUROPE

CNBC
PART OF BUSINESS WORLDWIDE

REAL-TIME QUOTES
Symbol / Company Go

SEARCH Go

WELCOME, Guest
Register Sign In

HOME NEWS MARKETS EARNINGS INVESTING SMALL BUSINESS VIDEO CNBC TV CNBC PRO

MARKETS U.S. Europe Asia
U.S. Europe Asia Oil Gold

Stocks Close Lower, Led by Energy; Banks Gain

Dow Jones Industrial Average
Previous Close 15,526
15,448
15,204

Price Change %Change
DJIA 16436.41 76.23 0.47%
NASDAQ ---
S&P 500 1879.55 4.74 0.25%

WHAT INVESTORS SHOULD KNOW

After-Hours Buzz: FII, LGF & More
Why Apple May Boost Dividend Soon
Best Trade of the Past 20 Years?
Google 'Sound,' Apple 'More Attractive'
With ETFs, How Much Is Too Much?
Currency War? Here's How to Hedge It
Egypt Is Ravenous for Energy
ECB's Weidmann: No Euro Intervention
Which ETFs Are Investors Buying?
Your Salary Information for Sale
Frugal Year of the Snake Unshored In
Weird ETFs, Revisited
Dow 30 After-Hours Quotes

4% Interest, Without Fees: Too Good to Be Checking?

Recent rates are not a thing of the past. Thanks to a group dedicated to giving local banks a even chance against the big boys.
Banks Still Raising, Hiding Fees

CEO: Report About Tesla Electric Car Is 'Unreasonable'

Stalled Out on Tesla's Electric Highway
American-US Air Deal Could Be End of Airline Stock Run
Stocks Better Investment Than Homes: Shiller
Wall Street Enforcement Lost in 'Revolving Door': Study
Why SI Swimsuit Has Buzz Despite Leak—Kate Upton
Bourbon Controversy: Just Add Water
Pope Benedict to Resign for Health Reasons
Next Pope? Ghana Cardinal Turkson Takes Lead
One Man's Quest: A Beer for Diabetics
State of the Union Buzz: Jobs and Apple CEO
Say What? Bill Gates Gets Personal on Reddit
Can US Companies' Chinese Results Be Trusted?
What Immigration Reform Means for the Economy
Highlights From NY Fashion Week
Models on Twitter: Tech Hits the Runway
5% of Credit Reports Contain Costly Errors: FTC
Fed Taking 'Forceful Action' on Economy: Yellen
Pentagon Cuts Only Possible Under Sequestration: Dean
Chasing Big Payoff, Savers Lose on Complex Bets
Stars of 'Nerd Philanthropy' Dominate Top 50 Givers

TOP VIDEO Latest Video | Analysts | CEOs

Play Video

Maria Presses White House on Spending Cuts
Gartman: Gold Breaks Down... 'Serious Trouble'
SI Swimsuit Models Ring Closing Bell

HIGHLIGHTS

Pity the Manager
In this new book, the authors explain what managers are good for, and why we should cut them some slack.

Has Volkswagen Found the Holy Grail?
Ulrich Hackenberg isn't yet a household name but Volkswagen's \$70 billion bet on his big idea pays off: he may join the likes of Henry Ford in the canon of auto industry pioneers.

Celebrity Love Nests
CNBC.com's third-annual Homes of Celebrity Couples collection has houses that range from a young couple's \$2.5 million "starter" to a \$43.5 million mega-mansion.

Dew-in the Morning
PepsiCo is rolling out a Mountain Dew for the morning. Kickstart, which has the Dew flavor plus juice, vitamins B and C and BANI an extra jolt of caffeine.

Carnival Rescue?
Rio's Carnival has begun but experts are questioning whether the party for Brazil's economy could be over before it's really begun.

Fido is Movin' On Up
Some apartment buildings are wooing pet owners with rooftop dog parks, "Toggy Hours" and even spas that offer "paw-dicures."

Blogs | Slideshows | Special Reports | Quizzes | Latest Video

SET HOME PAGE: UNITED STATES ASIA PACIFIC EUROPE

CNBC
PART OF BUSINESS WORLDWIDE

REAL-TIME QUOTES
Symbol / Company Go

SEARCH Go

WELCOME, Guest
Register Sign In

HOME NEWS MARKETS EARNINGS INVESTING SMALL BUSINESS VIDEO CNBC TV

MARKETS U.S. Europe Asia
U.S. Europe Asia Oil Gold

Stocks Close Lower, Led by Energy; Banks Gain

Dow Jones Industrial Average
Previous Close 15,526
15,448
15,204

Price Change %Change
DJIA 15971.24 21.73 -0.16
NASDAQ 3192.00 -1.87 -0.06
S&P 500 1517.01 -0.92 -0.06

WHAT INVESTORS SHOULD KNOW

After-Hours Buzz: FII, LGF & More
Why Apple May Boost Dividend Soon
Best Trade of the Past 20 Years?
Google 'Sound,' Apple 'More Attractive'
With ETFs, How Much Is Too Much?
Currency War? Here's How to Hedge It
Egypt Is Ravenous for Energy

CEO: Report About Tesla Electric Car Is Bogus

Tesla CEO Elon Musk told CNBC that a recent New York Times article claiming that cold weather cuts the mileage on the Tesla electric car is bogus. The Times shot back that the allegations were "flatly untrue."
Stalled Out on Tesla's Electric Highway

American-US Air Deal Could Be End of Airline Stock Run

American-US Airways: What It Means For Frequent-Fliers
Stocks Better Investment Than Homes: Shiller
5% of Credit Reports Contain Costly Errors: FTC
Wall Street Enforcement Lost in 'Revolving Door': Study
Bourbon Controversy: Just Add Water
Pope Benedict to Resign for Health Reasons
Next Pope? Ghana Cardinal Turkson Takes Lead
One Man's Quest: A Beer for Diabetics
State of the Union Buzz: Jobs and Apple CEO
Say What? Bill Gates Gets Personal on Reddit
Can US Companies' Chinese Results Be Trusted?
4% Interest, Without Fees: Too Good to Be Checking?

TOP VIDEO Latest Video | Analysts | CEOs

Play Video

SI Swimsuit Models Ring Closing Bell
Tesla CEO Musk Disputes NYT Test Drive Story
Watered Down Whiskey Has Fans Fuming

HIGHLIGHTS

Pity the Manager
In this new book, the authors explain what managers are good for, and why we should cut them some slack.

Has Volkswagen Found the Holy Grail?
Ulrich Hackenberg isn't yet a household name but Volkswagen's \$70 billion bet on his big idea pays off: he may join the likes of Henry Ford in the canon of auto industry pioneers.

Celebrity Love Nests
CNBC.com's third-annual Homes of Celebrity Couples collection has houses that range from a young couple's \$2.5 million "starter" to a \$43.5 million mega-mansion.

Dew-in the Morning
PepsiCo is rolling out a Mountain Dew for the morning. Kickstart, which has the Dew flavor plus juice, vitamins B and C and BANI an extra jolt of caffeine.

Carnival Rescue?
Rio's Carnival has begun but experts are questioning whether the party for Brazil's economy could be over before it's really begun.

Fido is Movin' On Up
Some apartment buildings are wooing pet owners with rooftop dog parks, "Toggy Hours" and even spas that offer "paw-dicures."

Blogs | Slideshows | Special Reports | Quizzes | Latest Video

Case Study: USA Today

USA TODAY - Breaking News and Local News | <https://www.usatoday.com>

NEWS TO YOUR INBOX Start the day smarter | WOMEN IN CONGRESS Historic achievements | FEDERAL TAX CREDITS How they work | GEN Z Where they are moving | [SUBSCRIBE NOW](#) First month free

For You U.S. Sports Entertainment Life Money Tech Travel Opinion 48°F | [Subscribe](#) | [Sign In](#)

Advertisement - scroll for more content.

MSC Industrial Supply

Top Headlines

Mike Epps has loaded gun seized by police at Indianapolis airport

Tucker Carlson hated Trump 'passionately.' Takeaways from lawsuit docs

FOR SUBSCRIBERS

Why Memphis didn't burn after Tyre Nichols video release

Women's history month: Celebrating historical moments for women in Congress

Controversial plan: Haul hot crude oil along new railroad

Bindi Irwin and the power of celebrities revealing health diagnoses

A GOP war on 'woke'? Poll finds most Americans see term as positive

Biden's budget proposal endgame: Forcing Republicans' hand

[Get the Daily Briefing in your inbox](#)

Ad

Big Print Volumes? Not Anymore!

[Brother Office USA](#) | [LEARN MORE](#)

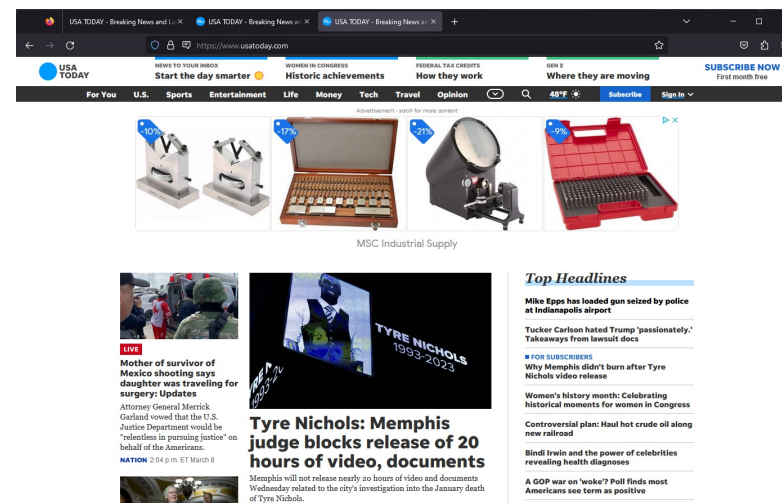
COVID lab theory debated by national hearing

<https://www.usatoday.com/story/entertainment/celebrities/2023/03/08/mike-epps-loaded-gun-seized-police-airport/1142988802/>

Case Study: USA Today

container + data-item for ad collection

Live

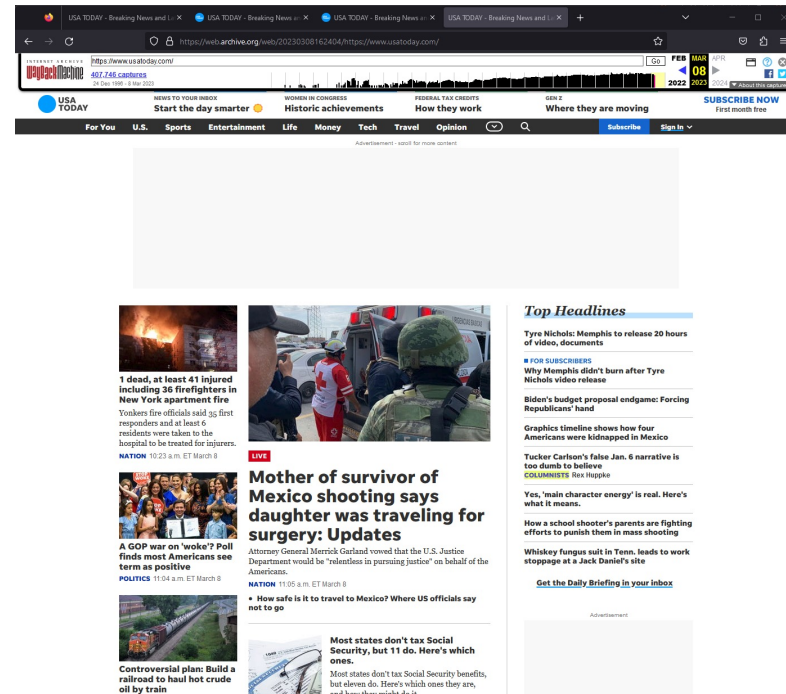


```
<aside aria-label="advertisement" class="gnt_m gnt_x gnt_x__lbl gnt_x_hi">
  <div id="ad-slot-7103-usatoday-high_impact-homepage-2" class="gnt_x_sl gnt_x_hi"
    data-google-query-id="CPKqy4GLzf0CFUzKswodfc8B0g">
    <div id="google_ads_iframe_7103/usatoday/high_impact/homepage_0__container__"
      style="border: 0pt none; min-width: 728px; min-height: 90px; width: 970px; height: 250px; display: inline-block;"
      class="gnt_x_sl_gpt"><iframe style="border: 0px none; vertical-align: bottom;"
        src="USA%20TODAY%20-%20Breaking%20News%20and%20Latest%20News%20Today_files/container.htm"
        id="google_ads_iframe_7103/usatoday/high_impact/homepage_0" title="3rd party ad content" name=""
        scrolling="no" marginwidth="0" marginheight="0" data-is-safeframe="true"
        sandbox="allow-forms allow-popups allow-popups-to-escape-sandbox allow-same-origin allow-scripts allow-top-navigation-by-user-activation"
        role="region" aria-label="Advertisement" tabindex="0" data-google-container-id="2"
        data-load-complete="true" width="970" height="250" frameborder="0"></iframe></div>
    </div>
  </aside>
```

Case Study: USA Today

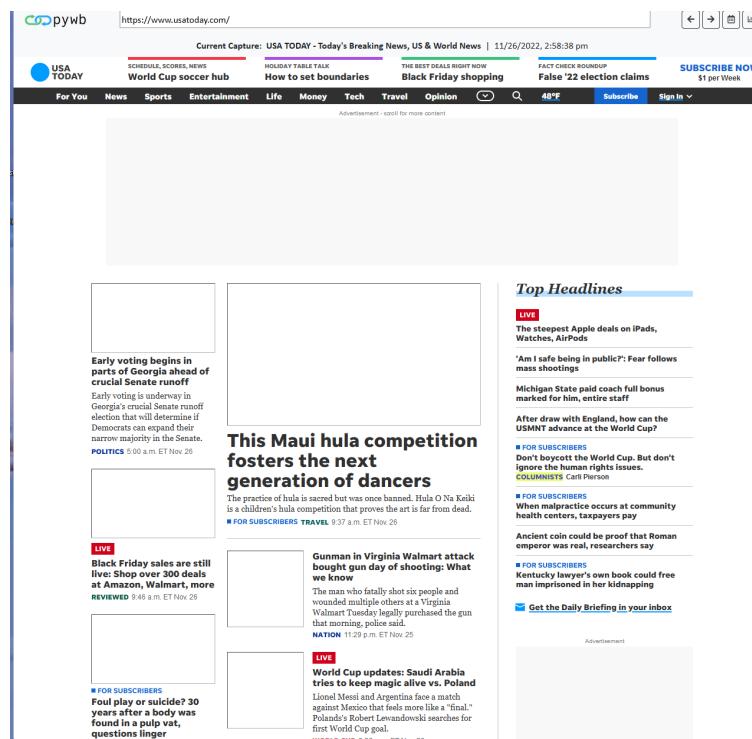
Internet
Archive

Sized container but no data-item



```
<aside aria-label="advertisement" class="gnt_m gnt_x gnt_x_lbl gnt_x_hi">
  <div id="ad-slot-7103-usatoday-high_impact-homepage-2" class="gnt_x_sl gnt_x_hi">
    <div id="google_ads_iframe_7103/usatoday/high_impact/homepage_0_container_"
      style="border: 0pt none; width: 1000px; height: 0px; min-width: 728px; min-height: 90px;" class="gnt_x_sl_gpt">
    </div>
  </div>
</aside>
```

Case Study: USA Today



Retrieved from common crawl

Out of Context

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






- **Vendor archives**


Google

📌 Active

Started running on Jul 3, 2023

Platforms  


Categories 

This ad has multiple versions 

ID: 283738683752008

3 ads use this creative and text

See summary details




Get U.S Citizenship

Sponsored • Paid for by Get U.S Citizenship

While USCIS's free resources are a great starting point, customized study materials can elevate your learning journey!

Our 100 Civics Questions Bundle, personalized with video, audio, and PDF materials, not only includes specifics about YOUR representative, senators, ...



CIVICQUESTIONS.GETUSCITIZENSHIP.COM

Pass Your Interview On The First Try! Risk free

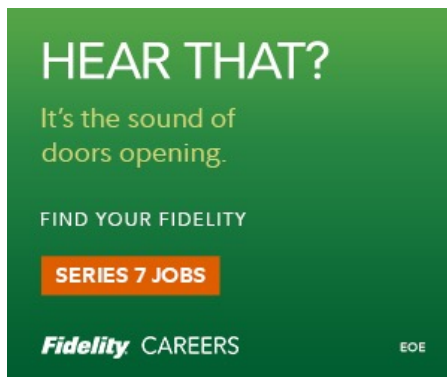
🛑 STOP WORRYING about the U.S. Citizenship Test

We've got you covered with our 100 Civics Questions...

Learn More

Out of Context

- Industry Archives: AdBeat.com



Cost Benefit

USA TODAY

THANKS TO THE POWER OF VACCINES: Photos: Nathan reacts

WHEN WILL WE ALL BE VACCINATED? Vaccine rollout

HOW DOES THE CORONA TRACKING BOARD X?

WOMEN OF THE YEAR: Nominate an inspiring

SUBSCRIBE NOW \$4.99 per month

Home Sports Entertainment Life Money Tech Travel Science

SET TO KNOW

AIRKNIT^x BOXER BRIEFS

Like you're floating on air.

SHOP NOW MACK WELDON

Top Headlines

Josh Dugger arrested days after wife announces pregnancy
 "Famous" TV host and "Celebrity" star Josh Dugger has been arrested five days after his wife Katie announced they are expecting their seventh child.
CELEBRITIES 7:24 p.m. ET Apr. 28

NFL draft: Everything to know for tonight's first round
 The 2021 NFL draft is finally here. USA TODAY Sports has you covered with everything you need to know to get set for the big event.
en.usatoday 1:13 p.m. ET Apr. 28

Lady Gaga says she's charged
CELEBRITIES 7:23 p.m. ET Apr. 28

Overman kills 2 drag queens, married couple in North Carolina standoff
NATION 6:27 p.m. ET Apr. 28

Phillips helps health partners enable better care
Money from Phillips

Will there be a fourth stimulus check?
 Another round of stimulus checks could hit more than 7 million people out of poverty, according to the Urban-Brookings Tax Policy Center.
MONEY 3:03 p.m. ET Apr. 28

QB Rodgers tells Packers he doesn't want to return
 The tension between reigning NFL MVP Aaron Rodgers and the Green Bay Packers might have hit a new level, with the QB apparently looking for a way out.
GREEN BAY PACKERS 6:00 p.m. ET Apr. 28

Why Biden's speech was a turning point for TV news
 The president's address about his five 100-day outlook plans for 2021 Biden and Kamala Harris, seen here on CNN, MSNBC, Fox News and more.
TV 2:49 p.m. ET Apr. 28

THE FIRST SURVIVOR OF ALZHEIMER'S DISEASE IS

Featured Videos

Huge Rock at construction inside a neighborhood
NEWS YOU WERE ASKED

Was Ted Cruz sleeping during Biden's address?
POLITICS

Boy identifies umpire who threw out home run
SPORTS

Remove stubborn stains from topware
PROVIDED BY THE USER

USA TODAY

THANKS TO THE POWER OF VACCINES: Photos: Nathan reacts

WHEN WILL WE ALL BE VACCINATED? Vaccine rollout

HOW DOES THE CORONA TRACKING BOARD X?

WOMEN OF THE YEAR: Nominate an inspiring

SUBSCRIBE NOW \$4.99 per month

Home Sports Entertainment Life Money Tech Travel Science

SET TO KNOW

AIRKNIT^x BOXER BRIEFS

Like you're floating on air.

SHOP NOW MACK WELDON

Top Headlines

Josh Dugger arrested days after wife announces pregnancy
 "Famous" TV host and "Celebrity" star Josh Dugger has been arrested five days after his wife Katie announced they are expecting their seventh child.
CELEBRITIES 7:24 p.m. ET Apr. 28

NFL draft: Everything to know for tonight's first round
 The 2021 NFL draft is finally here. USA TODAY Sports has you covered with everything you need to know to get set for the big event.
en.usatoday 1:13 p.m. ET Apr. 28

Lady Gaga says she's charged
CELEBRITIES 7:23 p.m. ET Apr. 28

Overman kills 2 drag queens, married couple in North Carolina standoff
NATION 6:27 p.m. ET Apr. 28

Phillips helps health partners enable better care
Money from Phillips

Will there be a fourth stimulus check?
 Another round of stimulus checks could hit more than 7 million people out of poverty, according to the Urban-Brookings Tax Policy Center.
MONEY 3:03 p.m. ET Apr. 28

QB Rodgers tells Packers he doesn't want to return
 The tension between reigning NFL MVP Aaron Rodgers and the Green Bay Packers might have hit a new level, with the QB apparently looking for a way out.
GREEN BAY PACKERS 6:00 p.m. ET Apr. 28

Why Biden's speech was a turning point for TV news
 The president's address about his five 100-day outlook plans for 2021 Biden and Kamala Harris, seen here on CNN, MSNBC, Fox News and more.
TV 2:49 p.m. ET Apr. 28

THE FIRST SURVIVOR OF ALZHEIMER'S DISEASE IS

Featured Videos

Huge Rock at construction inside a neighborhood
NEWS YOU WERE ASKED

Was Ted Cruz sleeping during Biden's address?
POLITICS

Boy identifies umpire who threw out home run
SPORTS

Remove stubborn stains from topware
PROVIDED BY THE USER

USA TODAY

THANKS TO THE POWER OF VACCINES: Photos: Nathan reacts

WHEN WILL WE ALL BE VACCINATED? Vaccine rollout

HOW DOES THE CORONA TRACKING BOARD X?

WOMEN OF THE YEAR: Nominate an inspiring

SUBSCRIBE NOW \$4.99 per month

Home Sports Entertainment Life Money Tech Travel Science

SET TO KNOW

AIRKNIT^x BOXER BRIEFS

Like you're floating on air.

SHOP NOW MACK WELDON

Top Headlines

Josh Dugger arrested days after wife announces pregnancy
 "Famous" TV host and "Celebrity" star Josh Dugger has been arrested five days after his wife Katie announced they are expecting their seventh child.
CELEBRITIES 7:24 p.m. ET Apr. 28

NFL draft: Everything to know for tonight's first round
 The 2021 NFL draft is finally here. USA TODAY Sports has you covered with everything you need to know to get set for the big event.
en.usatoday 1:13 p.m. ET Apr. 28

Lady Gaga says she's charged
CELEBRITIES 7:23 p.m. ET Apr. 28

Overman kills 2 drag queens, married couple in North Carolina standoff
NATION 6:27 p.m. ET Apr. 28

Phillips helps health partners enable better care
Money from Phillips

Will there be a fourth stimulus check?
 Another round of stimulus checks could hit more than 7 million people out of poverty, according to the Urban-Brookings Tax Policy Center.
MONEY 3:03 p.m. ET Apr. 28

QB Rodgers tells Packers he doesn't want to return
 The tension between reigning NFL MVP Aaron Rodgers and the Green Bay Packers might have hit a new level, with the QB apparently looking for a way out.
GREEN BAY PACKERS 6:00 p.m. ET Apr. 28

Why Biden's speech was a turning point for TV news
 The president's address about his five 100-day outlook plans for 2021 Biden and Kamala Harris, seen here on CNN, MSNBC, Fox News and more.
TV 2:49 p.m. ET Apr. 28

THE FIRST SURVIVOR OF ALZHEIMER'S DISEASE IS

Featured Videos

Huge Rock at construction inside a neighborhood
NEWS YOU WERE ASKED

Was Ted Cruz sleeping during Biden's address?
POLITICS

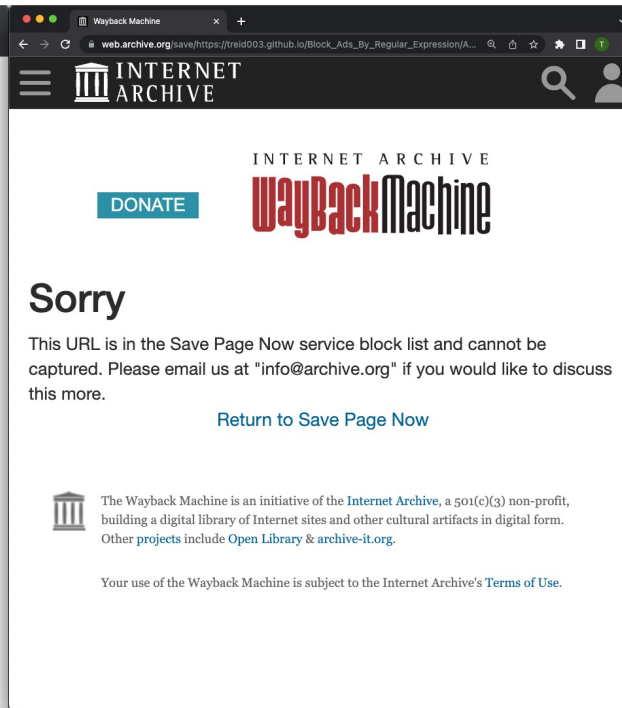
Boy identifies umpire who threw out home run
SPORTS

Remove stubborn stains from topware
PROVIDED BY THE USER

Archiving Issues

Web Archives Blocking Ads

Will this be blocked by a web archive?



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

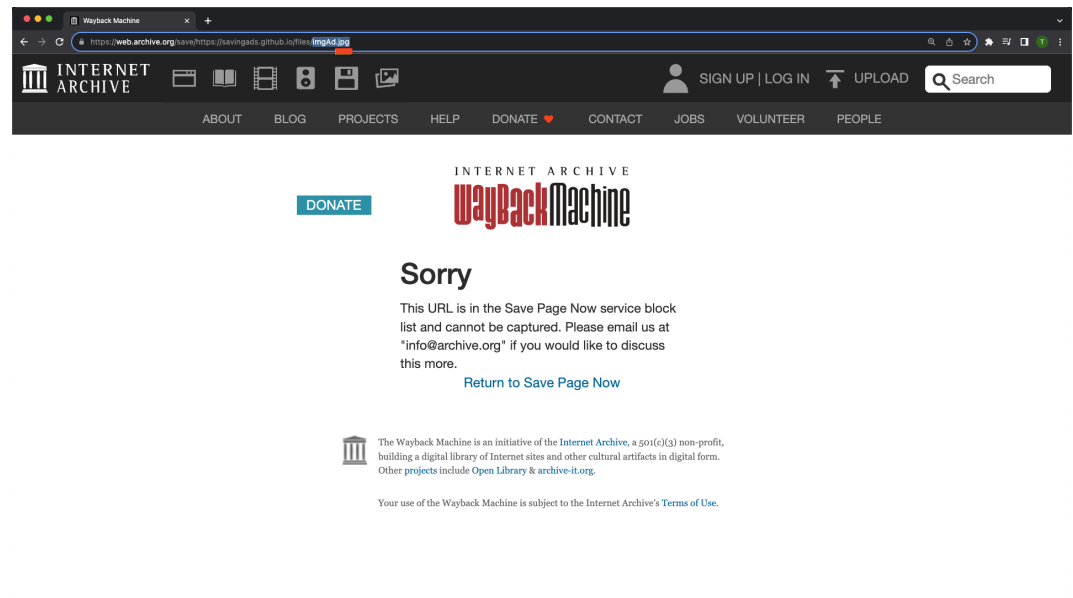
Ad Related Filename Blocked

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>



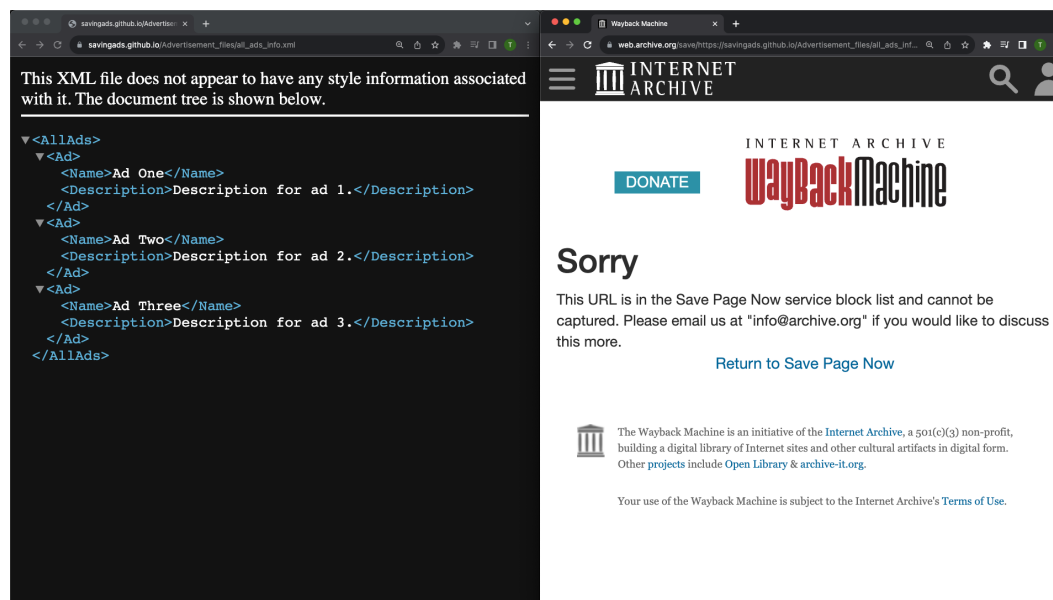
Ad Related Directory Name Blocked

Directory name: Advertisement_files

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

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

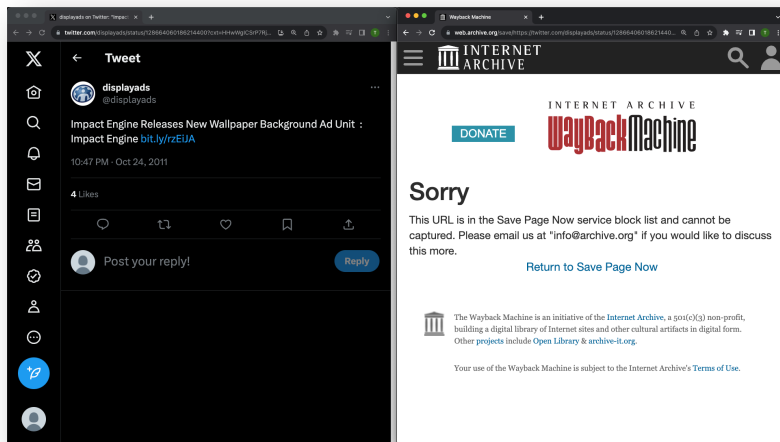
- displayAds
- videoAd
- webAd
- ads
- Advertisement_files



Social Media Accounts Blocked

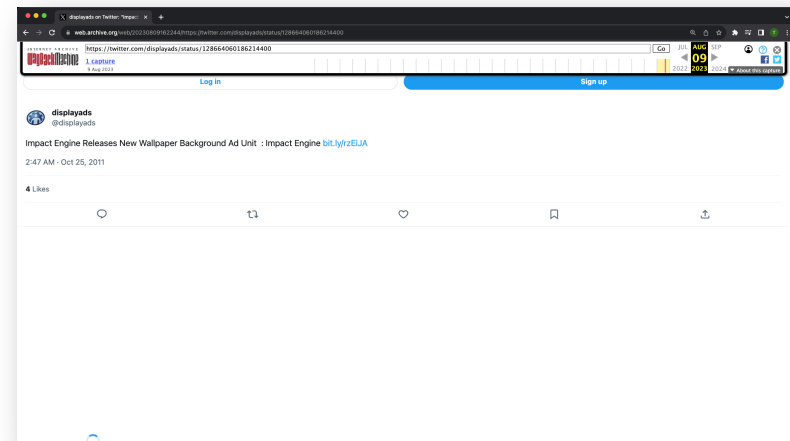
Social media accounts with an ad related username were also getting blocked, but after talking to Wayback Machine staff these accounts can now be archived

Account name: displayads
Tweet blocked during July 2023



URI-R
<https://twitter.com/displayads/status/128664060186214400?cxt=HHwWglCSrP7RjckDAAAA>

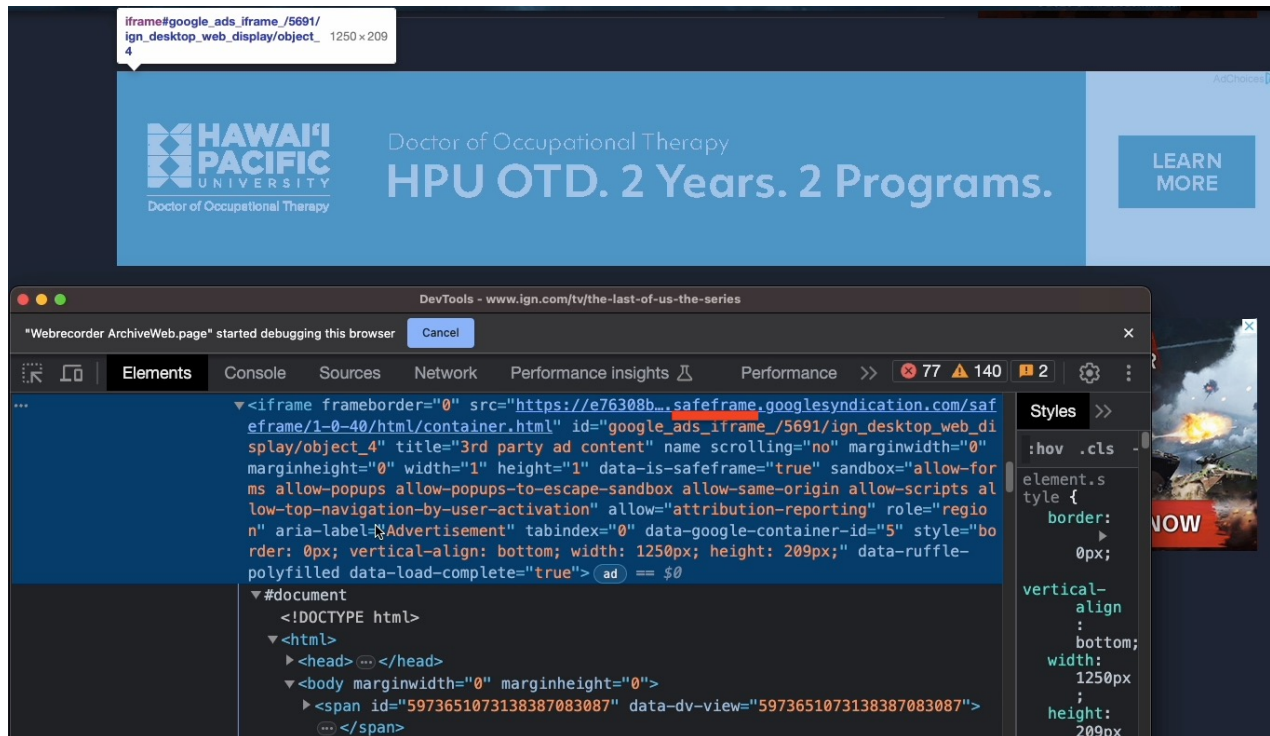
Account name: displayads
Tweet archived during August 2023



URI-M:
<https://web.archive.org/web/20230809162244/https://twitter.com/displayads/status/128664060186214400>

Replay Issues

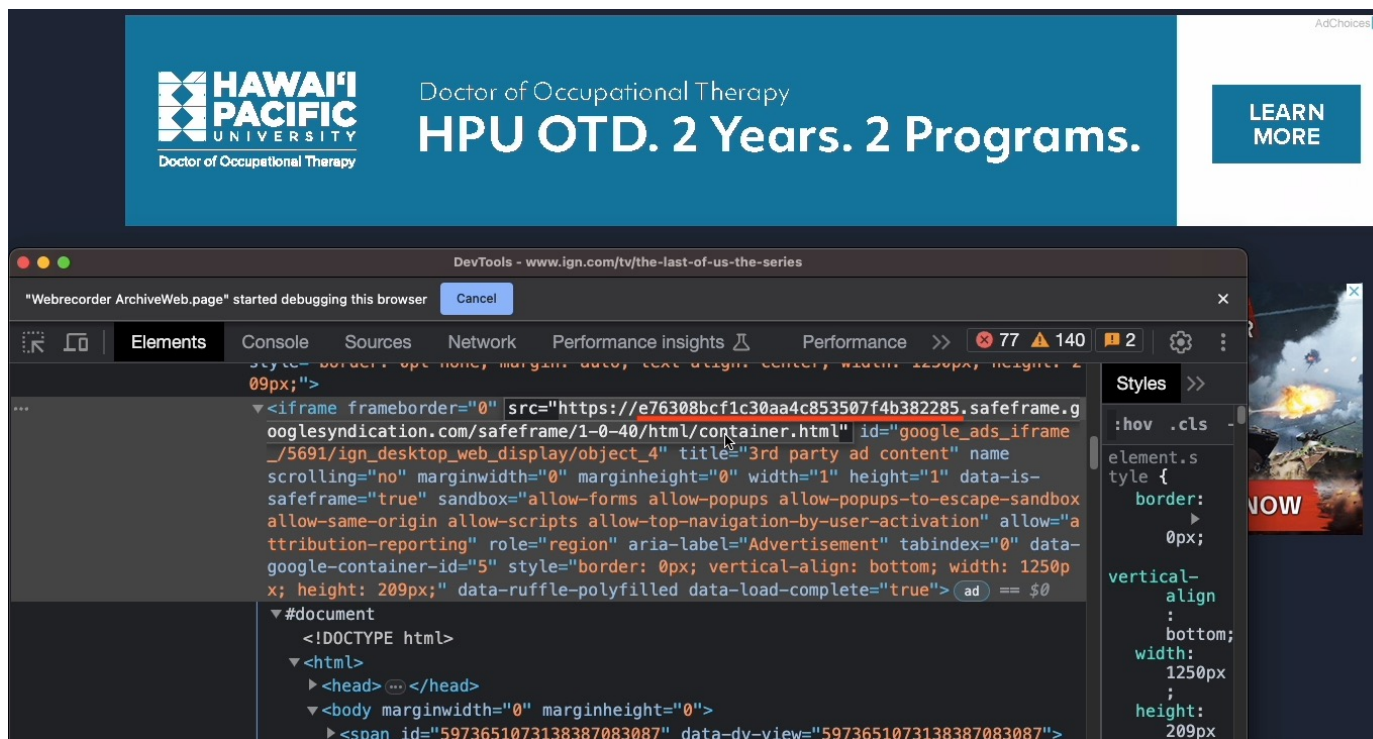
Ad Delivery Networks and Replay (Google SafeFrame)



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

Google SafeFrame

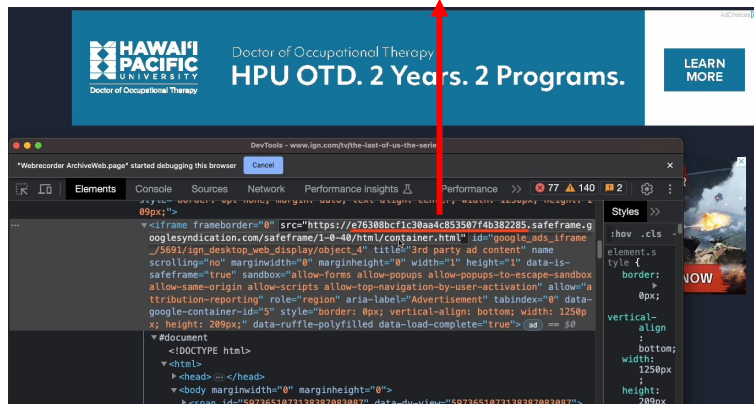
Most Google SafeFrames have a URL with a random value in the subdomain



Google SafeFrame Random Values

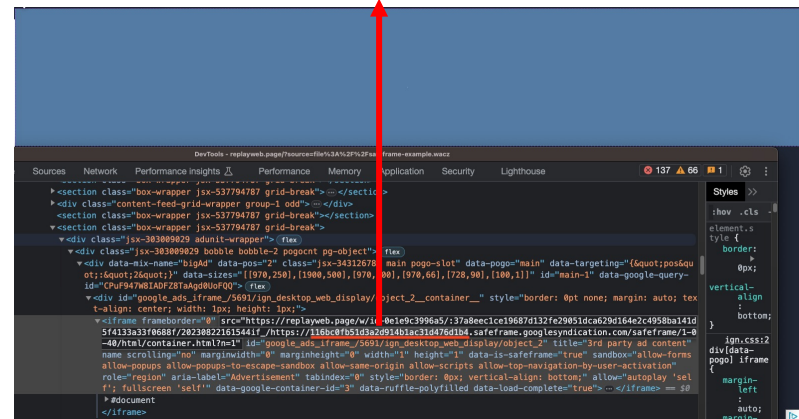
Live Web Page

Random value in subdomain:
e76308bcf1c30aa4c853507f4b382285



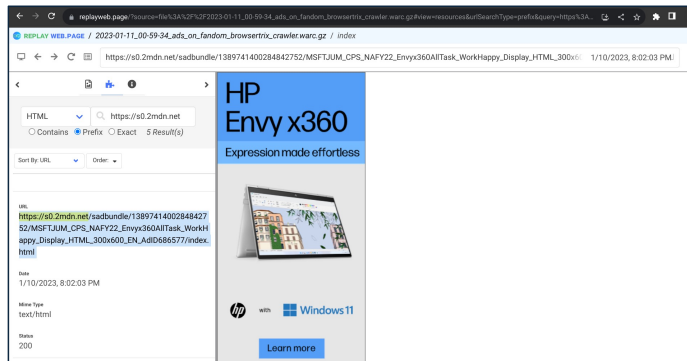
Archived Web Page

Random value in subdomain:
116bc0fb51d3a2d914b1ac31d476d1b4

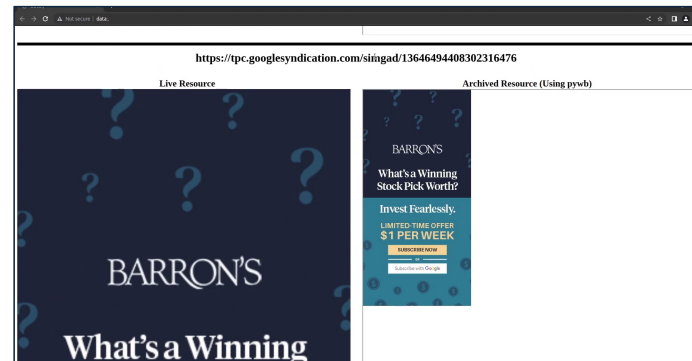


Processes and Tools: Finding Archived Ads

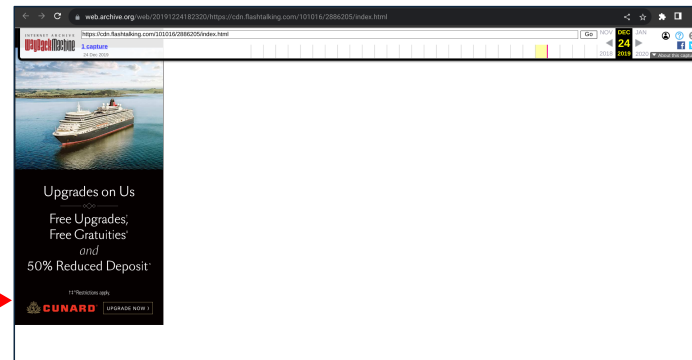
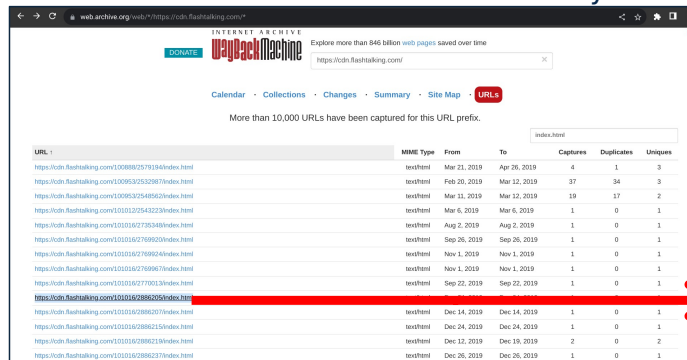
ReplayWeb.page's URL Search



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



DREXEL UNIVERSITY
College of
Computing &
Informatics



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