

## Summarizing web archives through storytelling with the Dark and Stormy Archives Project

Shawn M. Jones

Los Alamos National Laboratory Research Library Prototyping Team

Web Science and Digital Libraries Research Group Old Dominion University

CNI Fall 2020 Meeting

Thanks to:

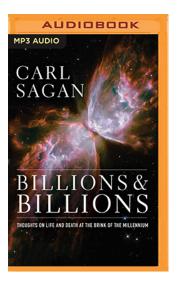


SIGIR Special Interest Group on Information Retrieval



# The abundance of data in web archives





### Web archives contain numerous documents



000



**Brewster Kahle** @brewster kahle

Wayback Machine now has 898,570,440,000 URL's -serving millions of users as a foundation for what has been said on a constantly shifting WWW. (up 17 billion URLs in less than 1 month ago twitter.com/brewster\_kahle...) Go @internetarchive !

Rewster Kahle @brewster kahle · Jan 10

Wayback Machine just grew to 881,352,519,000 web URL's. That is 881 Billion. For every one that becomes important in the news or in someones personal world, we crawl and store millions of others just-in-case. go @internetarchive

2:20 PM · Feb 5, 2020 · Twitter Web App

- https://twitter.com/brewster kahle/status/1225167435399036939
- https://www.webarchive.org.uk/en/ukwa/info/fag
- https://archive-it.org/collections/3572

#### How big is the archive?

The UK Web Archive collects millions of websites each year and billions of individual assets (pages, images, videos, pdfs etc.). As of 2017 we have collected approximately 500TB of data and increasing this by over roughly 60 – 70 TB a year.



# Mementos are the documents in web archive collections

Los Alamos National Laboratory	1
Research Library	

	Today: The Wall ( X +		F (	2 ☆	lii\ 🖸	8 0
Conifer		About	Report Bug	Help	Sign Up	Log
USA Today: T	he Wall					
Overview Browse A	1					
Bookmarks						
	USA Today: The Wall					
	USA Today's <i>The Wall</i> is a complex journalism project repor Mexico, using classic articles, video features, podcasts and navigation and an annotated map, accessible on multiple pla	rirtual reality,				
	This Webrecorder archive contains the website with all its e the exception of helicopter aerial footage of the complete b certain point are derived map coordinates; these coordinate the map's zoom level, to comprehensively archive at the mo	order: URLs t s are depend	hat would start the	video at a		
	The VR part has not been archived as it is not accessible on	the web.				
	Any links leading outside of the project, for example to the	JSA Today m	ain site, have not b	een included.		
	Bookmarks					
	List automatically created from visible pages on 2018-0	-02 00-48-0	6		_	
	List automatically created from visible pages on 2010 of	02.00.40.0				
	<ol> <li>The Wall – Interactive map explorin https://www.usatoday.com/border-wall/us-mexico-in/</li> </ol>	·		r		
	11/12/2017, 1:03:27 PM	eractive-bord				
	<ol> <li>The Wall – 13 documentaries explo https://www.usatoday.com/border-wall/documentarie 11/14/2017, 2:31:05 AM</li> </ol>		order and Tru	ımp's wal	u	
	3. The Wall – An in-depth examination	of Dona	ıld Trump's I	oorder		
	wall					
	https://www.usatoday.com/border-wall/ 11/14/2017, 2:37:45 AM					
	<ol> <li>A human smuggler, and the wall that https://www.usatoday.com/border-wall/story/human /559784001/ 11/14/2017, 2:38:07 AM</li> </ol>			gally-methods		
	5. Border vigilantes, and the wall they https://www.usatoday.com/border-wall/story/vigilant			er		

#### Collections with Conifer/Webrecorder

Web archive collections consist of **seeds** and crawls of those seeds.

The *mementos* from these crawls are versions of pages from the time of the crawl.

The mementos are the documents in our collections.

			Ε … ♥ ☆	in m
' ŵ	U 🎽 https://arch	ive-it.org/collections/3649		lin 🖸
_			F 🗉 🕥	Login
ARCHIVE-IT	IOME EXPLORE	for cultu	ding web archiving service collecting and accessing and heritage on the web <i>ilt at the Internet Archive</i>	Î
xplore >> Interne	et Archive Global Events >>	2013 Boston Marathon Bombing		
	ARCHIVE	2013 Boston Marathon Bom Collected by: Internet Archive Global Events	bing	
		Archived since: Anr. 2013		
E Z	<u> </u>	Description: This collection is comprised of contributions fr includes news articles, blogs, social media sites, and organizz Monethon Bombing and its of enrolfeste	rom curators and subject matter exp	erts; and
		Marathon Bombing and its aftereffects Subject: Spontaneous Events	anoma websites related to the 2013 l	
		Subject: Spontaneous Events		
Narrow You	Denk			
Narrow iou	Ir Results	Sites for this collection are listed below. Narrow your results a specific URL or to search the text of archived webpages.	t left, or enter a search query below t	o find a site,
Subject	Sort By: Count (A-Z)	Enter search terms here	Search	Clear
Manhunt (35)			Scarch	_
Dur Marathon (21 Reddit (19)		Sites Search Page Text		
Dzhokhar Tsarna Tamerlan Tsarna				
More V		Page 1 of 4 (318 Tot	tal Results) Next F	age 🕨
Date		Sort By: Title (A-Z)   Title (Z-A)   URL (A-Z)   URL (Z-Z)	Z-A)	
	Sort By: Count (A-Z)	URL: http://abclocal.go.com/ktrk/feature?	?section=news/local&	
4/19/2013 (36) 4/15/2013 (28)		id=9065470		
4/18/2013 (26) 4/16/2013 (25)		Captured once on Apr 19, 2013		
4/17/2013 (25)		Title: Brothers Who Lost Legs in Boston I	-	
More 🔻		URL: http://abcnews.go.com/blogs/headli lost-legs-in-boston-bombing-reunited/	ines/2013/04/brothers-wh	0-
Туре	Sort By: Count (A-Z)	Captured 2 times between May 2, 2013 and May 10, 2013		
News Article (131	)	Subject: Runners, Victim Type: News Article		
Blog (35) Social Media (25)		Date: 4/30/2013		
	site (12)	Title: The feminist librarian: a few more		
		URL: http://annajcook.blogspot.com/2013 and-flowers.html	3/04/a-few-more-thoughts	-cats-
Institutional Web Photos (12) More V				
Photos (12) More V		Captured once on Apr 22, 2013		
Photos (12)	Sort By: Count (A-Z)	Captured once on Apr 22, 2013 Subject: Dzhokhar Tsarnaev, Blogs Type: Blog		
Photos (12) More V	Sort By: Count   (A-Z)	Captured once on Apr 22, 2013 Subject: Dzhokhar Tsarnaev, Blogs		
Photos (12) More V Language Dutch (1)	Sort By: Count   (A-Z)	Captured once on Apr 22, 2013 Subject: Dzhokhar Tsarnaev, Blogs Type: Blog Date: 4/20/2013 Title: The feminist librarian: Some Thoug		
Photos (12) More V Language Dutch (1) English (1)	Sort By: Count   (A-2)	Captured once on Apr 22, 2013 subject: Datokhar Tarnare, Blogs Type: Blog Date: 4202013 Title: The feminist librarian: Some Thoug URL: http://annajcook.blogspot.com/2011		
Photos (12) More V Language Dutch (1) English (1)	Sert By: Count   (A-2)	Captured ance an Apr 22, 2013 Subject: Doublar Tsraneev, Blogs Type: Blog Date: 4/20/2013 Title: The feminist librarian: Some Thoug URL: http://annajcook.blogspot.com/2011 Captured ance an Apr 22, 2013 Subject: Blog		
Photos (12) More V Language Dutch (1) English (1)	Sort By: Count   (A-Z)	Captured ance an Apr 22, 2013 subject: Dhokkar Tamarev, Blogs Type: Blog hat: 4202013 Title: The feminist librarian: Some Thoug URL: http://annajcook.blogspot.com/2011 Captured once on Apr 22, 2013		

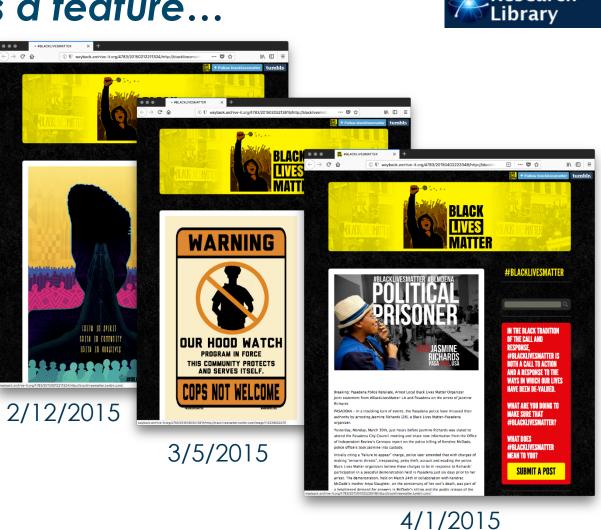
Collections with Archive-It



# Web archives have many versions (mementos) of the same page and this is a feature...





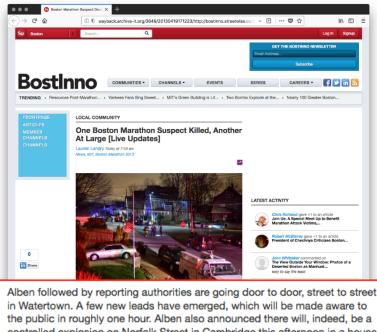


University of Utah Office of Admissions from the University of Utah Web Archive Collection

Tumblr Black Lives Matter Blog from the #blacklivesmatter Collection

### Different versions (mementos) allow us to see an unfolding news story





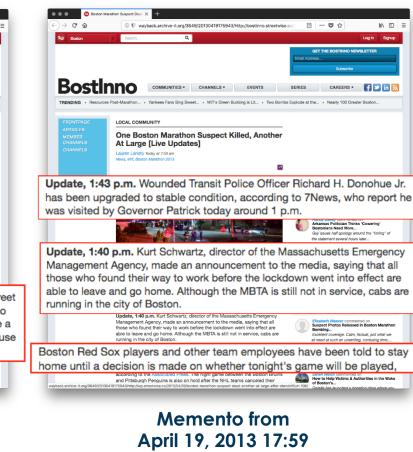
the public in roughly one hour. Alben also announced there will, indeed, be a controlled explosion on Norfolk Street in Cambridge this afternoon in a house they secured earlier today.

Robert McElaney gave +1 to an article Suspect Photos Released in Boston Maratho

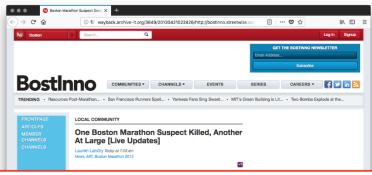
everything we've asked, but we need some more time.

Deveau assured the town authorities will not be leaving until the neighborhood is safe, and Menino said they will keep the public informed as news comes in.

Memento from April 19, 2013 17:12 Searching for suspects, City on lockdown



Officer Donahue in hospital, Lockdown loosened, Will the Red Sox game be cancelled?



**Updated, 10:20 p.m.** "We have closed an important chapter in this tragedy," said President Barack Obama in a press conference at the White House, minutes after a briefing came to a close in Watertown.

Obama acknowledged the coordination among state, federal and local officials, proclaiming, "They all worked as they should—as a team. We are extremely grateful for that."

He also honored Sean Collier, the MIT police officer who lost his life at the hand of the Marathon bombing suspects, and acknowledged the questions that remain.

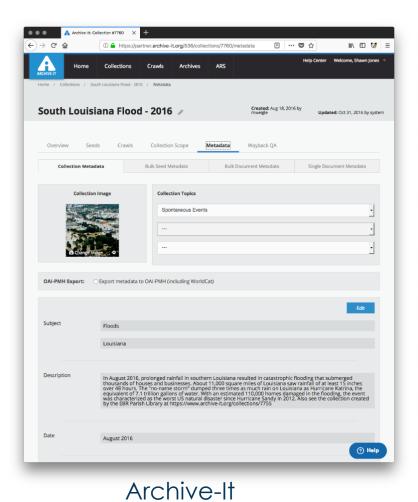
Why did young men who grew up and study here...resort to such violence?

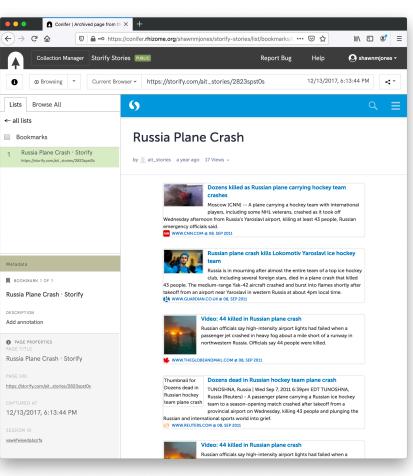
How did they plan and carry out these attacks, and did they receive any help?

"The wounded deserve answers," Obama said, announcing all necessary resources will continue to be deployed to determine what happened.

#### Memento from April 24, 2013 2:24 Suspect Found, Officer Collier lost life, Obama speaks

# Tools exist that easily allow users to add content to web archives...





#### Conifer/Webrecorder



Save	Page Now
http://	
SAVE PAGE	
for use as a tr	b page as it appears now usted citation in the future. sites that allow crawlers.
_	et Archive's Page Now



Archive.Today's On-demand Web Form





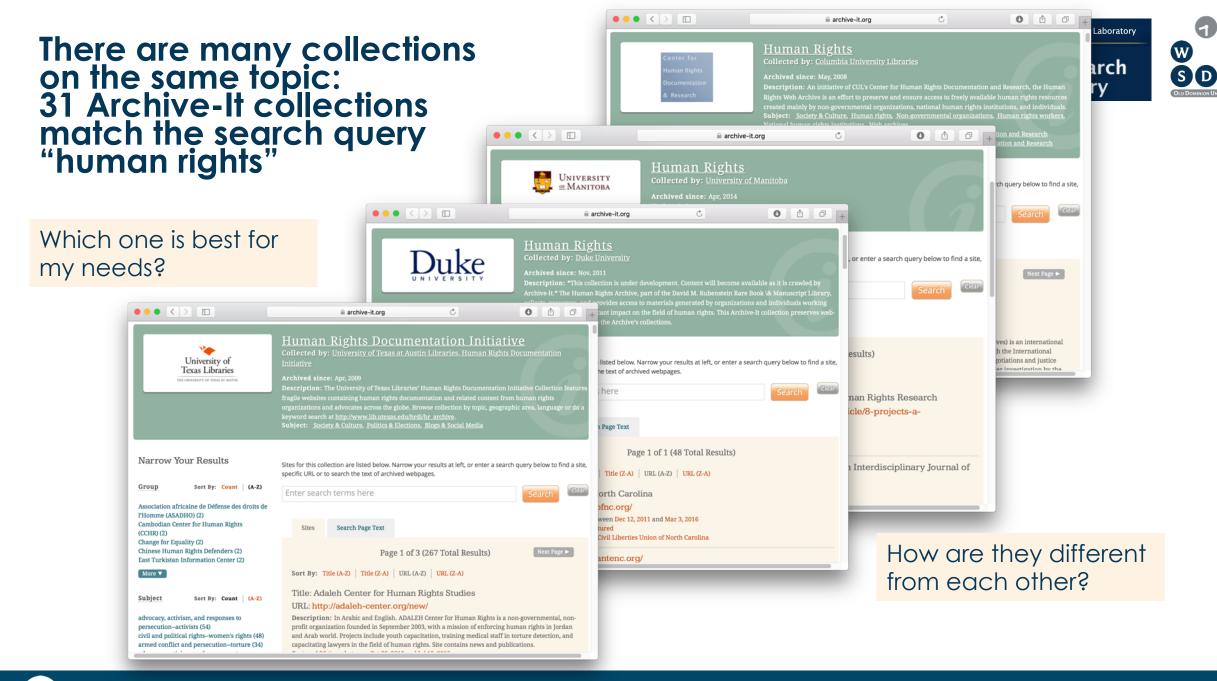
These collections live a life after their curator has stopped adding to them.

The collection curator is not the only user of the collection!



### Consider Archive-It as an example...





#### 🤊 @shawnmjones @StormyArchives



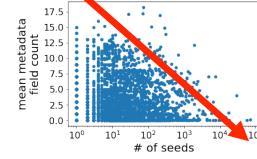
### Metadata is there, but does not always help

Because metadata is optional it is not always present.

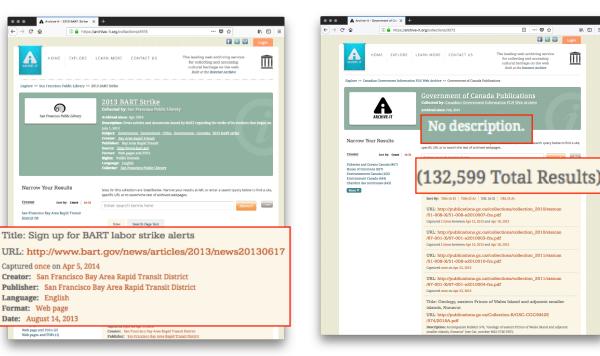
Metadata on Archive-It collections:

- many different curators
- different organizations
- different content standards
- different rules of interpretation
- it is inconsistently applied

This means that a user cannot reliably compare metadata fields to understand the differences between collections.



🥑 @shawnmjones @StormyArchives



#### 9 seeds with metadata

132,599 seeds no metadata

11

Paradox: More seeds = more effort More seeds = greater user need for metadata

Jones, S.M., Weigle, M.C., & Nelson, M.L. (2019). Social Cards Probably Provide For Better Understanding Of Web Archive Collections. In ACM CIKM '19. https://doi.org/10.1145/3357384.3358039

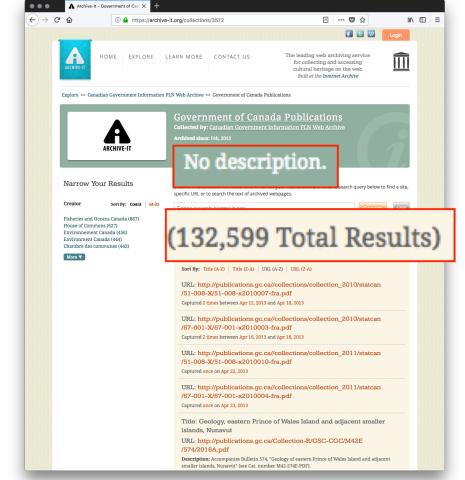


### Reviewing mementos manually is costly

Some collections have 1000s of seeds

Each seed can have many mementos

In some cases, this can require reviewing 100,000+ documents to understand the collection



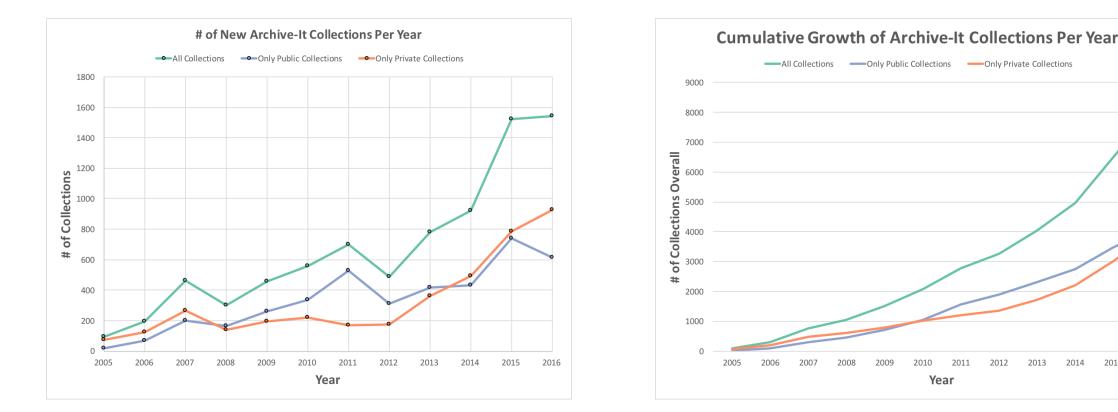
This collection has 132,599 seeds, many with multiple mementos

### In addition to more mementos in web archives, more Archive-It collections are added every year



2015

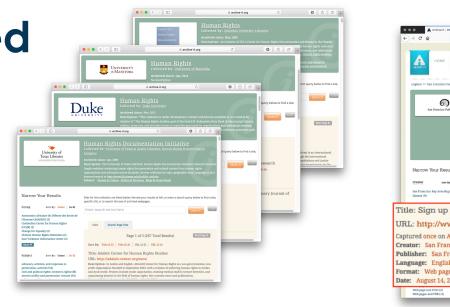
2016

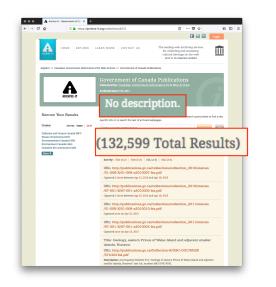


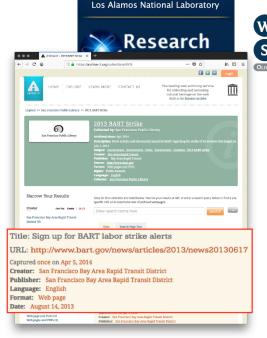
#### More than 15,000 collections exist

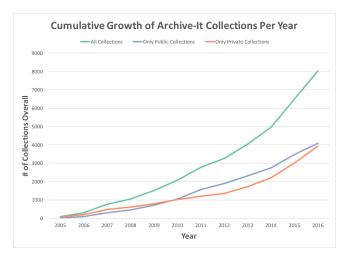
### The problem, summarized

- S There are multiple collections about the "same concept."
- § The metadata for each collection is non-existent, or inconsistently applied.
- § Many collections have 1000s of seeds with multiple mementos.
- S There are more than 15,000 collections.
- § Human review of these mementos for collection understanding is prohibitively expensive.





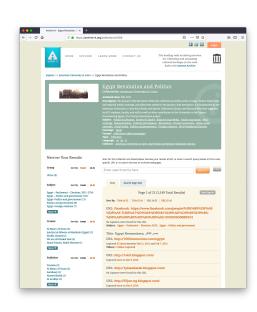






# **Existing Solutions**

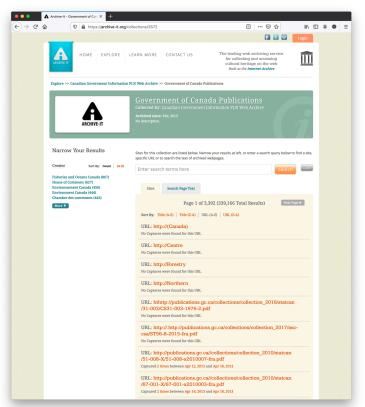


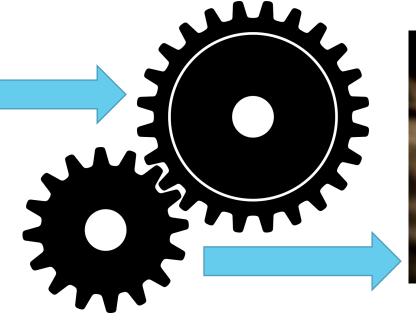




# Ideally, a user would be able to glance at a visualization and gain understanding of the collection









#### Web Archive Collection

Automated Solution Visualization that conveys understanding at a glance





٠								1 m	P
						1 A A A A A A A A A A A A A A A A A A A			a
					1	1 1		h .	18
								1	
						1 . · · · · · · · · · · · · · · · · · ·			
11	2008-07-02 00:00:00		1						1
	2. Non-governmental organizations (NGOs) => A	Abuelas de Plaza de Mayo				1.1			-
:	<ol> <li>Non-governmental organizations (NGOs) =&gt; 8</li> </ol>								-
÷	2. Non-governmental organizations (NGOs) => 8 2. Non-governmental organizations (NGOs) => 8	Jurma Campaign UK							
	2. Non-governmental organizations (NGOs) => F	Front Line				1			
1	2. Non-governmental organizations (NGOs) => H	Human Rights First		1.1		1.			
	<ol> <li>Non-governmental organizations (NGOs) =&gt; J</li> <li>Non-governmental organizations (NGOs) =&gt; P</li> </ol>	Kurdish Women's Rights Watch				1.			1
	2. Non-governmental organizations (NGOs) => 5	SANGONET					1		1 C
	<ol> <li>Non-governmental organizations (NGOs) =&gt; 1</li> <li>Non-governmental organizations (NGOs) =&gt; 1</li> </ol>	Tibetan Centre for Human Rights and Democracy			1 1	L .			<u> </u>
-	<ol> <li>Non-governmental organizations (NGOs) =&gt; 1</li> <li>Non-governmental organizations (NGOs) =&gt; 1</li> </ol>	libet information technology web		-	· ·	1.			18 - C
-	2. Non-governmental organizations (NGOs) => U	United Nations Assistance Mission for Iraq		1.1					· · · · ·
	4. Blogs by individuals => 看不见的西藏~唯色博著	5	]						
						1.		• :	· · · · · ·
									1
					_				<b>.</b>
						· · · · ·			<u> </u>
									4
									<b></b>
1									
		1				£			
					-				-
						1.1			
		1		1.1					14 C
								1	- ·
		· · · · ·				1			a -
-+-					- · · :=			1	i i
								1.00	3
						·		1 m 2 m	<b>1</b>
						1			·
								1.1.1	
								· ,	11
008-0	7 2008-10 2009-01	2009-04 2009-07 20	09-10 2010-01	2010-04	301	0-07 201	0-10 201	1-01 201	11-04
~18-0	2008-10 2009-01	2009-07 20	2010-01	2010-04	2010	201	201	201	11-04
	Sep '08 Jan '09	May '09 Se	p '09 jan '10		May *10		Sep '10	Jan '11	

Timelines provide an overview, but do not provide understanding at a glance.

🥑 @shawnmjones @StormyArchives

## Thumbnails work well for small collections, but do not scale.

K. Padia, Y. AlNoamany, and M. C. Weigle, (2012) "Visualizing digital collections at Archive-It," In JCDL '12. https://doi.org/10.1145/2232817.2232821

17

What if we could apply a visualization technique that users already know how to process?



# Users would require no training and could instead get information immediately.

Users could quickly decide which collection meets their needs!



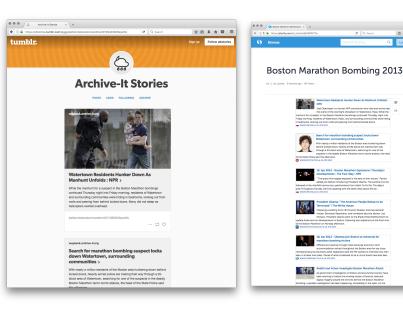


☆ é \$ \* ♥ :

Log In Sign Up

# Social media storytelling

🥑 @shawnmjones @StormyArchives



19

#### Web surrogates provide a visual summary of a web resource drawn from the content of the resource Instances

#### Home | Michele C. Weigle

#### www.cs.odu.edu/~mweigle -

Michele C. Weigle, Michael L. Nelson, Deborah Kempe (Frick Art Reference Library and New York Art Resources Consortium), Pamela Graham (Columbia University Libraries), and Alex Thurman (Columbia University Libraries), "Visualizing Webpage Changes Over Time", NEH/IMLS Digital Humanities Advancement Grant HAA-256368-17, Oct 2017 - ...

Research Methods · Publications · Schedule/Travel · Teaching · Working With Me

#### Text snippet (example from Bing)

Instances archived after 01-jul-2008 and before 31-jan-2010



#### Browser Thumbnail (example from UK Web Archive)



2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections? A social card generated by Facebook for my previous blog post. Rich links, snippet, social snippet, social media card, Twitter car.. WS-DL.BLOGSPOT.COM

#### Social Card (example from Facebook)

S. M. Jones. "Let's Get Visual and Examine Web Page Surrogates." https://ws-dl.blogspot.com/2018/04/2018-04-24-lets-get-visual-andexamine.html, 2018.



#### www.loc.gov the library of congress 1278,906 3,868 **R** 81 **2,717,124** 25,124,927 web captures from 1996 to 2017

#### Text + Thumbnail (example from Internet Archive)





# Social cards are summaries of web resources



Long URL:

https://www.google.com/maps/dir/Old+Dominion+University,+Norfolk,+VA/Los+Alamos+National+Laboratory,+New+Mexico/ @35.3644614,-109.356967,4z/data=!3m1!4b1!4m13!4m12!1m5!1m1!1s0x89ba99ad24ba3945:0xcd2bdc432c4e4bac!2m2!1d-76.3067676!2d36. 8855515!1m5!1m1!1s0x87181246af22e765:0x7f5a90170c5df1b4!2m2!1d-106.287162!2d35.8440582

The same URL represented as a social card:



Old Dominion University to Los Alamos National Laboratory Find local businesses, view maps and get driving directions in Google Maps.

§ URLs can be difficult for people to comprehend

§ Social cards provide a title, small text snippet, and striking image from the web page behind a URL

VS.



### And we use these cards all the time!



WS-DL.BLOGSPOT.COM 2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections?



2017-08-11: Where Can We Post Stories Summarizing Web ... The Web Science and Digital Libraries Research Group at Old Dominion University. ws-dl.blogspot.com

#### Twitter



#### 2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections? >

The Web Science and Digital Libraries Research Group at Old Dominion University.

#### Tumblr

ws-dl.blogspot.com > 2017/08 > 2017-08-11-where-ca... \*

Where Can We Post Stories Summarizing Web Archive ...

 $\label{eq:linear} \mbox{Aug 11, 2017} - \mbox{In effect, social cards are visualizations of web objects, piercing the veil} \hdots \mbox{Earlier} attempts at visualizing \mbox{Archive-It collections by Padia and} \hdots \hdo$ 



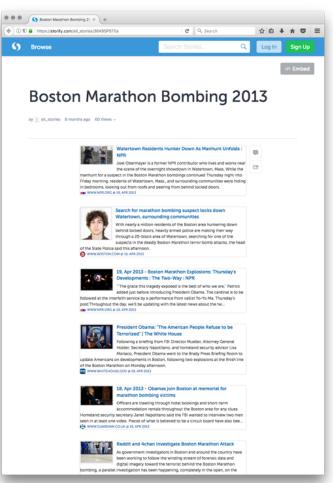
Facebook

2017-08-11: Where Can We Post Stories Summarizing Web ... https://ws-dl.blogspot.com/2017/08/2017-08-11-where-can-we-post-stories.html -Aug 11, 2017 · In Yasmin AlNoamany's Dark and Stormy Archives (DSA), she uses social cards in Storify stories to summarize mementos from Archive-It collections. These stories take the form of 28 high ... Author: Shawn M. Jones

Bing

# Social media storytelling uses groups of social cards to provide a "summary of summaries"



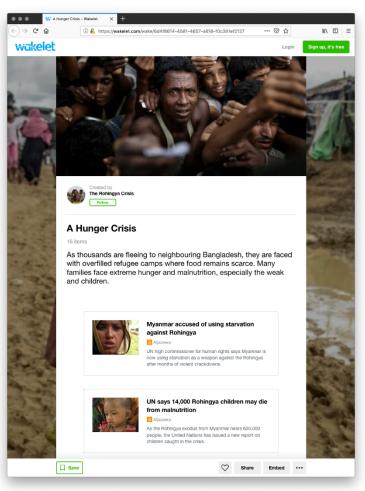


Each social card summarizes a web resource.

Each story groups the social cards, summarizing the topic.

Social cards contain the same information in the same place on each card, allowing for easy comparison.

We want to use this technique to summarize web archive collections **because users are already familiar with this visualization paradigm**.



#### 2 resources are shown in this Wakelet story

6 resources are shown in this Storify story

Social media storytelling uses an interface that users already understand how to process!



# Users would require no training and could instead get information immediately.

Users could quickly decide which collection meets their needs!

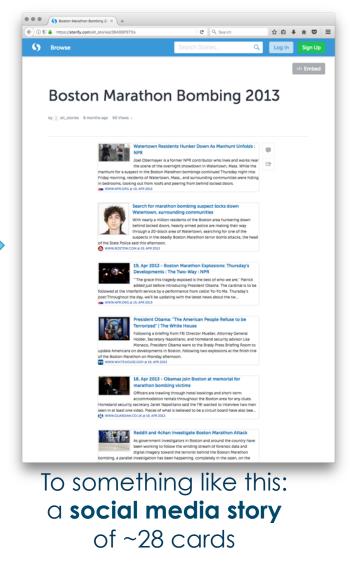




# Our solution: applying social media storytelling to web archives

tps://archive-it.org/collections/3649	マロピ 9, Search 会自 🖡 会 🛛 🖇
	🖬 🖬 💹 🛛 Login
HOME EXPLORE L	EARN MORE CONTACT US The leading web archiving service for collecting and accessing cultural heritage on the web Built at the <i>lowernet Archive</i>
lore >> Internet Archive Global Events >> 20	13 Boston Marathon Bombing
ARCHIVE	2013 Boston Marathon Bombing Collected by: Internet Archive Giobal Events Archived alone: Apr. 2013 Beespitale: This collection is comprised of contributions from curators and subject matter experts and includes news articles. Naps, seed intelliation, and organizational websites related to the 2013 Boston Marathon Bombing and Its athereffects Subject: <u>Sportaneous Events</u>
arrow Your Results	Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.
ubject Sort By: Count (A-2)	Enter search terms here
Janhunt (35) Jur Marahan (21) Johokhar Tsarnaev (15) Janerian Tsarnaev (13)	Sites Search Page Text Page 1 of 4 (318 Total Results) Net Page 1
More V	Seet By: Title (A-2)   Title (Z-A)   UEL (A-2)   UEL (Z-A)
ate Sort By: Count (A-2)	
/19/2013 (36) /15/2013 (26) /18/2013 (26) /19/2013 (25) /17/2013 (25)	URL: http://doi.ocal.go.com/ktrk/festure?section=news/local& id=906470 Capterd mor on Apr 18, 2013 Title: Brothers Who Lost Legs in Boston Bombing Reunited
More V	URL: http://abcnews.go.com/blogs/headlines/2013/04/brothers-who- lost-legs-in-boston-bombing-reunited/
ype Sort By: Count (A-2)	Captured 2 times between May 2, 2013 and May 10, 2013 Subject: Runners, Victim
iews Article (131) Nog (35) Iocial Media (25)	Subject: Runners, Victim Type: News Article Date: 4/30/2013
News Article (131) Nog (15) cocial Media (25) notitutional Website (12)	tablec: Runser, Victo Type: Irwy Ardic Date: 4302013 Title: The feminist librarian: a few more thoughts + cats and flowers
Sews Article (131) Bog (35)	Subject: Runners, Victim Type: News Article Date: 4502003
Vews Article (131) Blog (15) Octal Media (25) ratizutional Website (12) Totos (12)	salject Rusers, Vitin Type: Kreve Andre Date: 4902033 Title: The feminist librarian: a few more thoughts + cats and flowers URL: http://annajcook.blogspot.com/2013/04/a-few-more-thoughts-cats-

From this: 318 seeds with 2421 mementos





### Storytelling With Web Archives Has Multiple Use Cases...

### **1. Promotion of the collection**

Storytelling allows a curator to promote a collection, making others aware of it

# 2. Exploring aspects of the collection

S Users can explore a collection and expose specific sides of a news story or focus on specific people or places

### 3. Summarization

§ Web archive collections are too large for manual review – we need a summary to understand what they contain

	HIt - 2013 Boston Marati × +			
ë û	🛈 🔒 https://arch	ive-it.org/collections/3649	E \cdots 🛡 🏠	lii V
			F 🗉 🛛	Login
ARCHIVE-IT	HOME EXPLORE	LEARN MORE CONTACT US	The leading web archiving service for collecting and accessing cultural heritage on the web <i>Built at the Internet Archive</i>	
Explore >> Inte	rnet Archive Global Events >> :	2013 Boston Marathon Bombing		
		2013 Boston Marathon Collected by: Internet Archive Global Eve Archived since: Apr, 2013 Description: This collection is comprised of con includes new articles, blogs, acclain mella sites, Marathon Bombing and its aftereffects Subject: Spontaneous Events		
Narrow Y	our Results	Sites for this collection are listed below. Narrow y specific URL or to search the text of archived web		low to find a site,
Subject	Sort By: Count (A-Z)	Enter search terms here	Sea	Clear
Manhunt (35) Our Marathon Reddit (19) Dzhokhar Tsar Tamerlan Tsar More V	naev (15)	Sites Search Page Text Page 1 of 4	4 (318 Total Results)	iext Page ►
Date	Sort By: Count (A-Z)	Sort By: Title (A-Z)   Title (Z-A)   URL (A-	Z) URL (Z-A)	
4/19/2013 (36) 4/15/2013 (28) 4/18/2013 (26) 4/16/2013 (25) 4/17/2013 (25) More ▼		URL: http://abclocal.go.com/ktrk id=9065470 Captured once on Apr 19, 2013 Title: Brothers Who Lost Legs in URL: http://abcnews.go.com/blo lost-legs-in-boston-bombing-re	Boston Bombing Reunited gs/headlines/2013/04/brothers-	-who-
<b>Type</b> News Article (1 Blog (35)		Captured 2 times between May 2, 2013 and M Subject: Runners, Victim Type: News Article Date: 4/30/2013		
Social Media (2 Institutional W Photos (12) More ▼		Title: The feminist librarian: a fe URL: http://annajcook.blogspot.o and-flowers.html Captured once on Apr 22, 2013	•	
Language Dutch (1)	Sort By: Count   (A-Z)	Subject: Dzhokhar Tsarnaev, Blogs Type: Blog Date: 4/20/2013		
English (1) Russian (1)		Title: The feminist librarian: Sor URL: http://annajcook.blogspot.e Captured once on Apr 22, 2013 Subject: Blogs Type: Blog Date: 4/19/2013	-	ml

# Some services equate "stories" with ephemera, which is not what we want



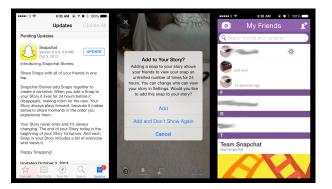


Facebook stories

Image ref: https://techcrunch.com/2018/04/05/facebook-stories-default



Instagram stories Image ref: https://buffer.com/library/instagram-stories



#### Snapchat stories

Image ref: https://techcrunch.com/2013/10/03/snapcha gets-its-own-timeline-with-snapchat-stories-24-

These platforms delete the user's stories 24 hours after they are posted.

This form of social media storytelling is the opposite of what we are looking for. We want the stories to be artifacts themselves.



Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

# Existing social platforms do not work well for archived resources (mementos)



	https://twitter.com/(/momenta/8950500927925			立日
y Home	Moments	Search Twitter	Q, Hos	an account? Log i
	All Stories Gat_stories - 10m wayback.archive-it.org/	3649/201304191	n v	~
	AlT Stories Stat_stories - 10m wayback.archive-it.org/	3649/201304191	u 🛛	~
	All Stories Gal_stories - 10m	2010/201001101	n v	~
	wayback.archive-it.org/	3049/201304191		
	AIT Stories Cost_stories - 10m		12 V	~
	wayback.archive-it.org/	3049/201304191		
	All Stories Sail_stories - 10m		u 🛡	~
	wayback.archive-lt.org/	3649/201304191		
	AT Stories		12 🗢	~

@shawnmjones @StormyArchives





Archive-It Stories http://wayback.archive-it.org/.../8uoVeyShelDu.../story.html



Patriot's Day heroes: Marathon spectators become first- responders - The Boston...

BOSTONGLOBE.COM

August 3 at 5:27pm



Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

### Existing card services create a confusing experience for pages from web archives (mementos)



#### ARCHIVE-IT



#### Egypt - This Just In - CNN.com Blogs

To alleviate U.S. pressure on the Egyptian army, Saudi Arabia would consider matching the estimated \$1.5 billion the United States provides Egypt in military aid each year if Washington decides to cut its aid to the troubled North African nation, an Arab diplomat says.

Read the article on wayback.archive-it.org >

powered by embedly

#### embed.ly social card

#### This is confusing.

Who published these resources? Archive-It? CNN?

It is a matter of trust, because different sources are attributed.



Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

29

# Current social media storytelling solutions don't work with archival content



## Our stories should not be temporary.

# Our summaries should not present users with confusing attribution.



Web archive collections are vast and hard to understand



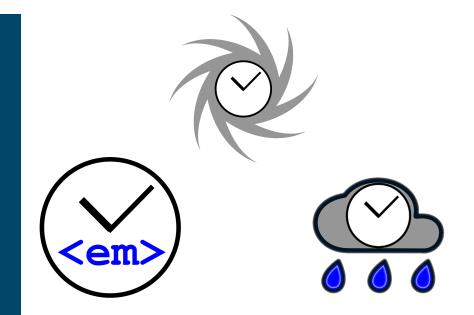
## Understanding them is difficult with existing visualizations

Existing social media storytelling does a poor job with archived pages (mementos)



## Solving the problem with the Dark and Stormy Archives Project

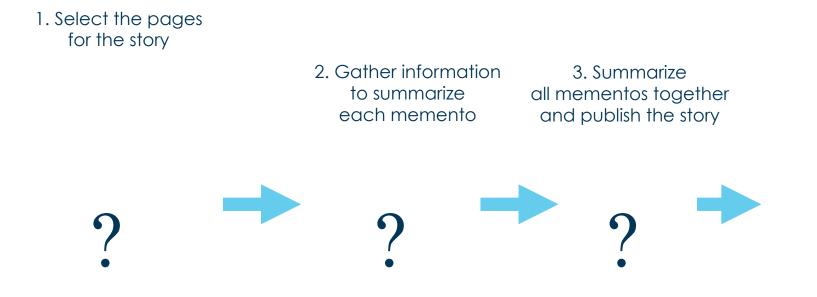




32

### Telling a story with web archives has 3 basic steps

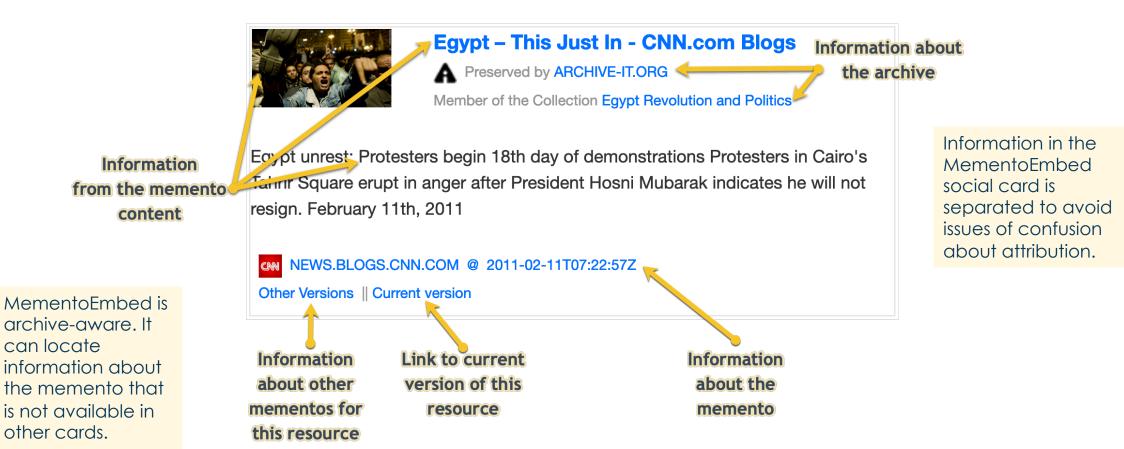




Øshawnmjones @StormyArchives

### Neither social media services nor card services were reliable for web archive storytelling, so we created MementoEmbed...



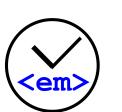


🕥 @shawnmjones @StormyArchives

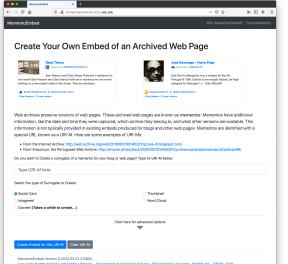
Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

34

# We can summarize *individual* pages with MementoEmbed in different ways







All of these summaries represent the same page about the Egyptian Revolution.

@shawnmjones @StormyArchives



Egypt – This Just In - CNN.com Blogs
Preserved by ARCHIVE-IT.ORG
Member of the Collection Egypt Revolution and Politics

Egypt unrest: Protesters begin 18th day of demonstrations Protesters in Cairo's Tahrir Square erupt in anger after President Hosni Mubarak indicates he will not resign. February 11th, 2011

NEWS.BLOGS.CNN.COM @ 2011-02-11T07:22:57Z Other Versions || Current version

Cards



Imagereels



#### Browser Thumbnails



Word Clouds



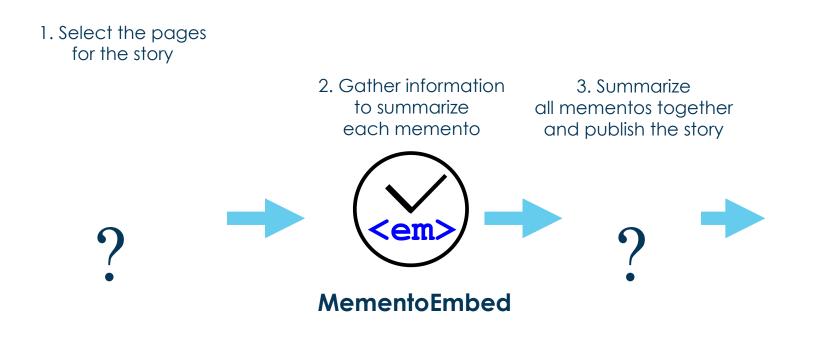
Docreels

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

35

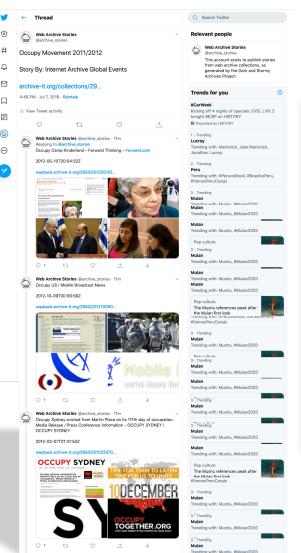
# Now we can produce cards, but how do we create stories?

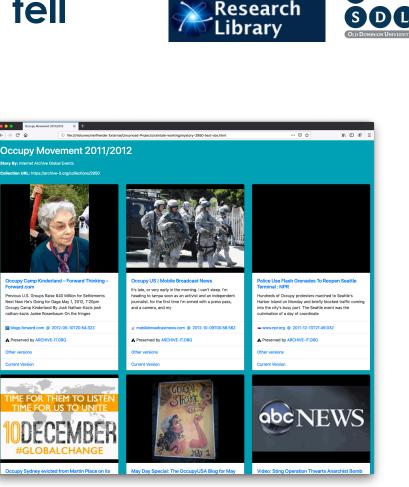




# MementoEmbed just produces information for individual pages, so we created Raintale to tell stories...

- § Raintale generates social media stories for mementos
- § Raintale publishes to file formats, like HTML, or social media services, like Twitter
- § Raintale also supports templates for custom output





Los Alamos National Laboratory

1

W

37

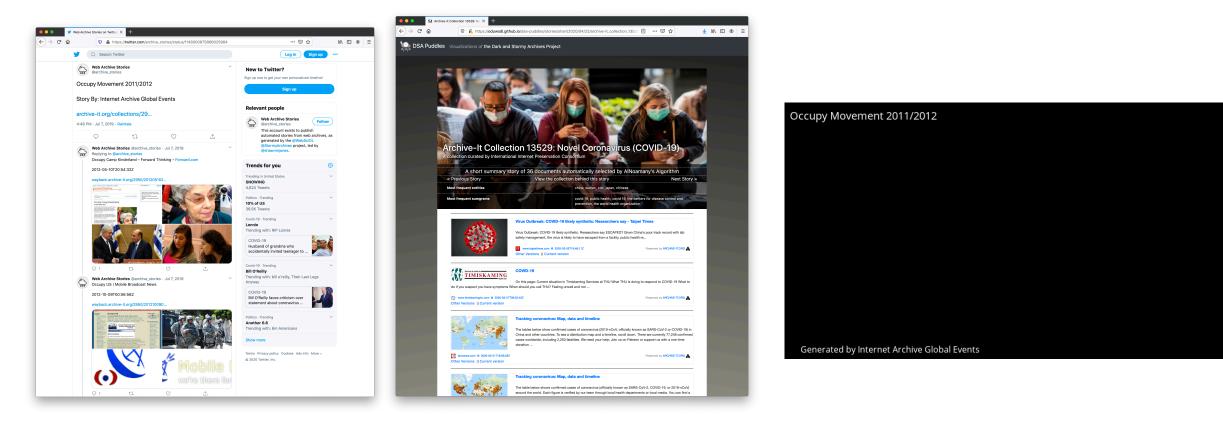
#### https://oduwsdl.github.io/raintale/

StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

## We can publish *many* Mementos with **Raintale** in a variety of formats





Twitter Threads with screenshots, images, titles, dates, links GitHub Pages with cards and much more Videos with top ranked images and sentences to be uploaded to YouTube, Twitter, etc. (experimental)

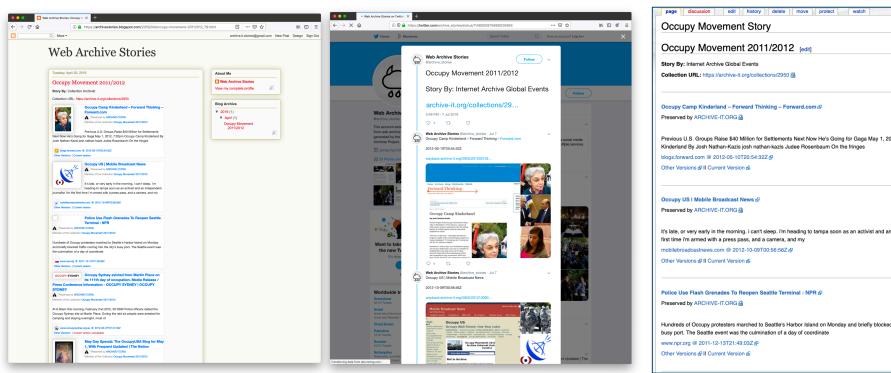


Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

#### Raintale uses templates to let you customize the look and destination media of your story

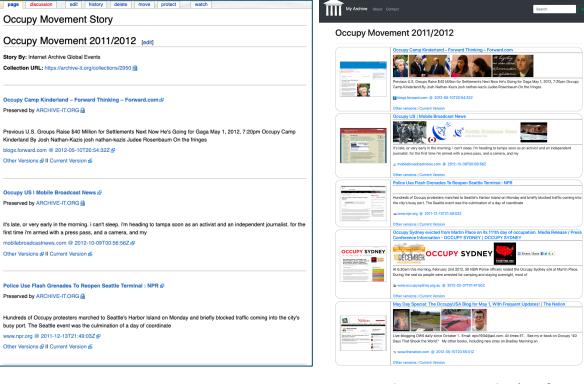


1



MementoEmbed cards in Blogger

#### Customized Twitter Threads



MediaWiki Pages

Brand your own stories for your institution

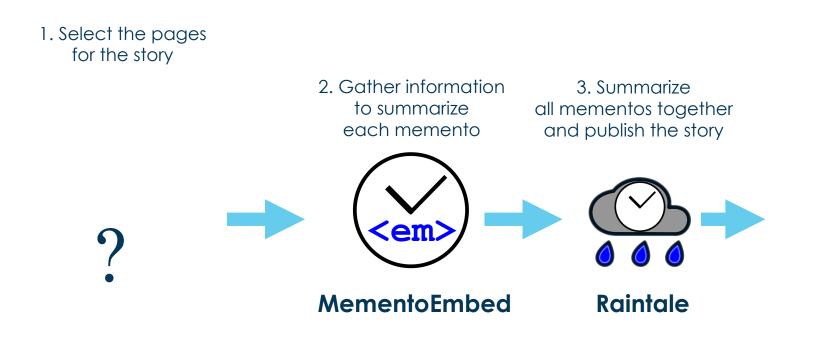
39



Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

## Now we can create stories, but how do we select mementos?





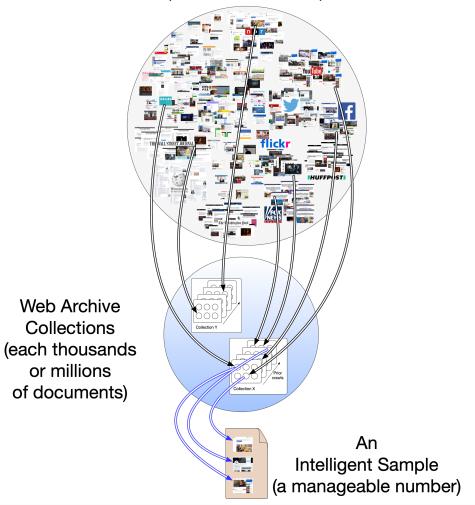


## Hypercane provides intelligent sampling of web archive collections

- **§** Works with Memento-compliant archives
- § Discovers mementos
- § Can create new mementos
- § Reports on:
  - § collection metadata
  - § named entities
  - § terms
  - § image analysis
- § Provides summarization steps:
  - § Clustering
  - § Filtering
  - § Scoring
  - § Ordering



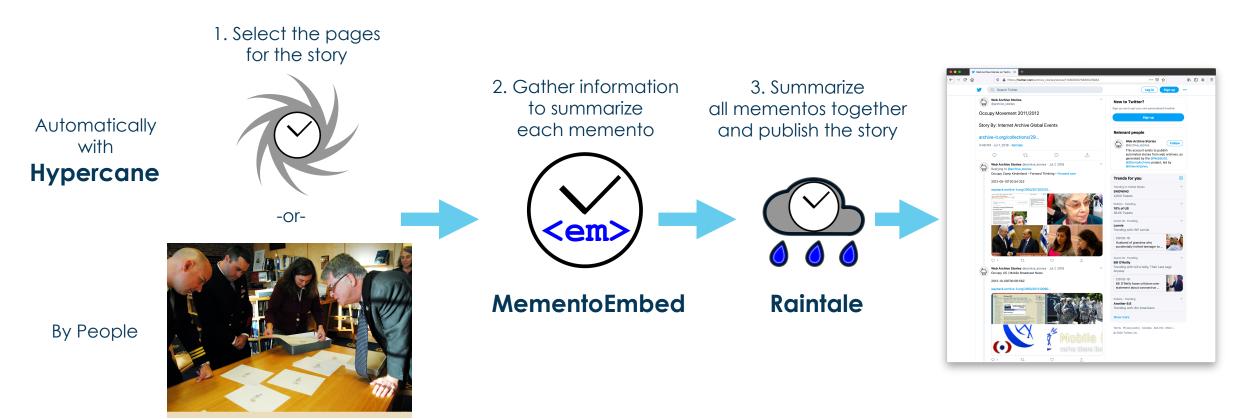
The Web in the "perpetual now" (near ∞ documents)



S. M. Jones. "Hypercane Part 1: Intelligent Sampling of Web Archive Collections

# The DSA Toolkit provides a solution for each stage of the storytelling lifecycle

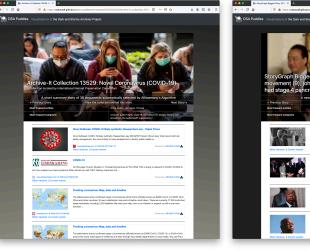


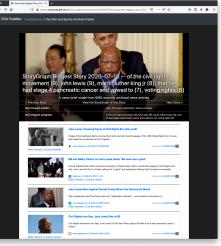




# Examples of the DSA Toolkit at work

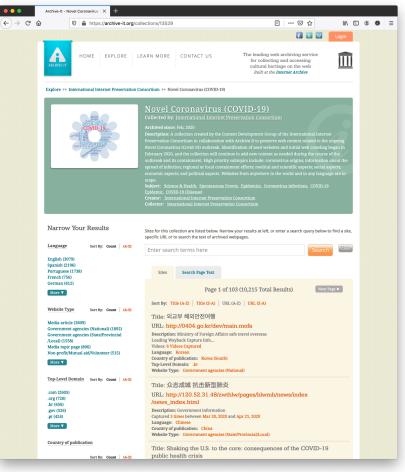




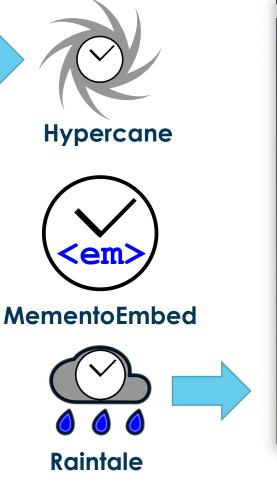


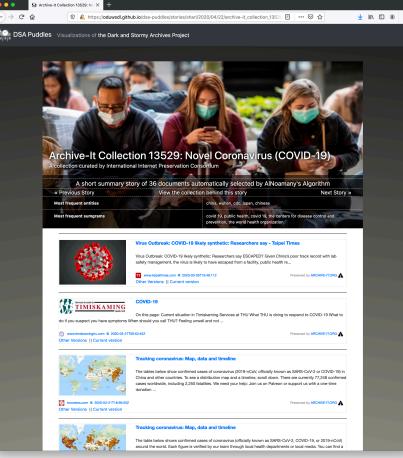
## We automatically summarize public Archive-It collections





From this: 23,376 mementos



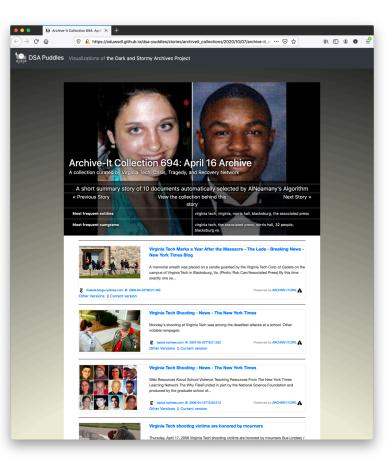


To this: a sample of 36 mementos visualized as social cards, phrases, and images

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <u>https://arxiv.org/abs/2008.00137</u>

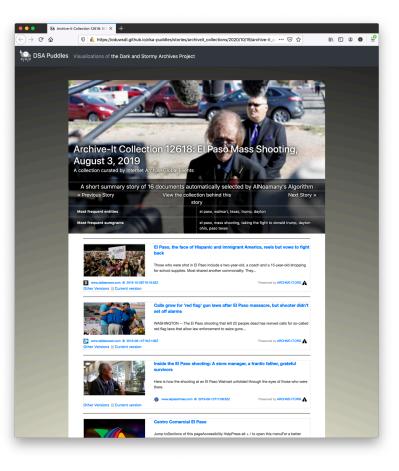
### With the DSA Toolkit, we can compare Archive-It collections about mass shootings







Norway



El Paso

Virginia Tech

🥑 @shawnmjones @StormyArchives

## But the DSA Toolkit is <u>NOT</u> just for Archive-It collections...



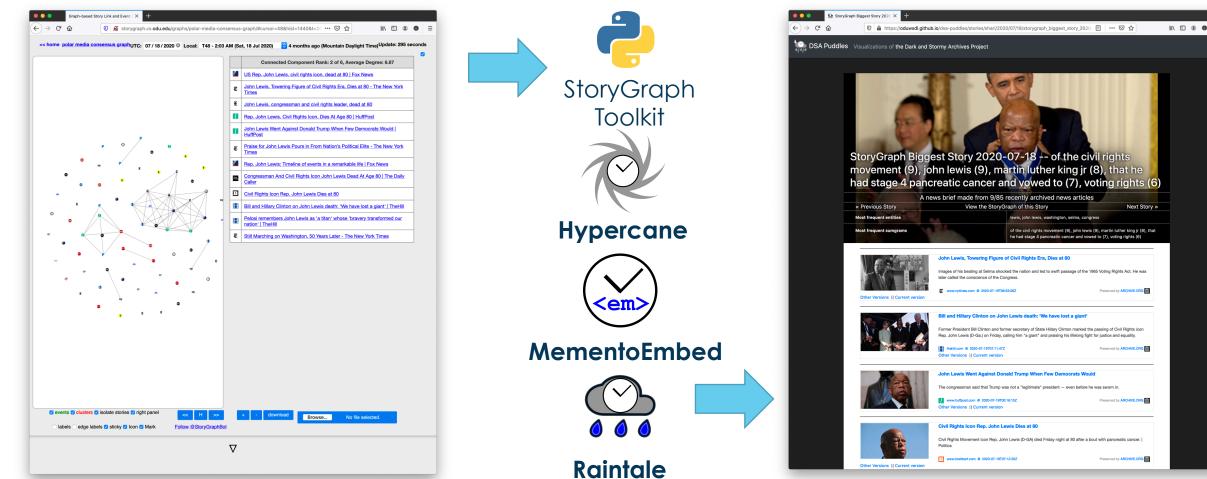
## The DSA Toolkit works with Memento-compliant Archives





## We automatically produce daily news summaries





To this: news articles of the biggest story of the day

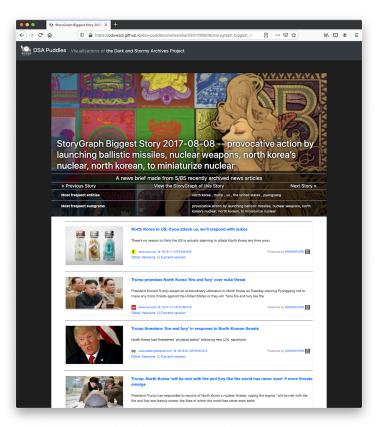
Jones, S.M., Nwala, A.C., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). SHARI --An Integration of Tools to Visualize the Story of the Day. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. https://arxiv.org/abs/2008.00137

From this: news articles extracted by StoryGraph

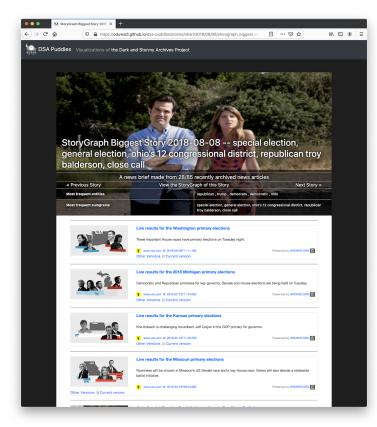
🕥 @shawnmjones @StormyArchives

# This allows us to review news from the same day in different years

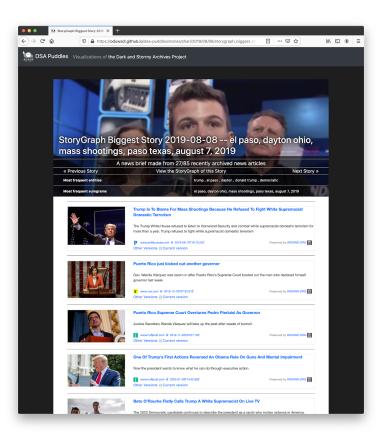




August 8, 2017 Nuclear Provocation by North Korea



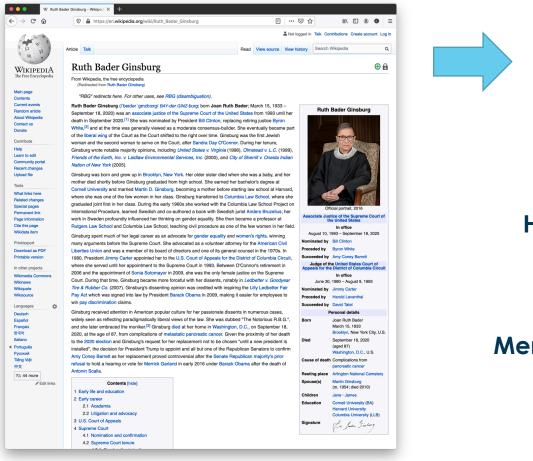
#### August 8, 2018 US Primaries and Special Elections



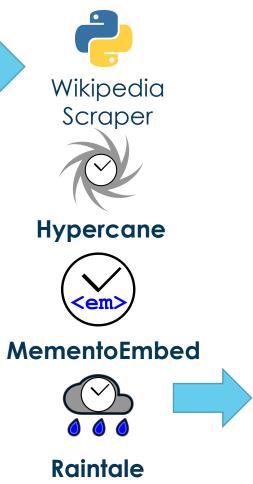
August 8, 2019 Aftermath of Shootings at El Paso and Dayton

## We can automatically summarize Wikipedia references





From this: Wikipedia references as mementos

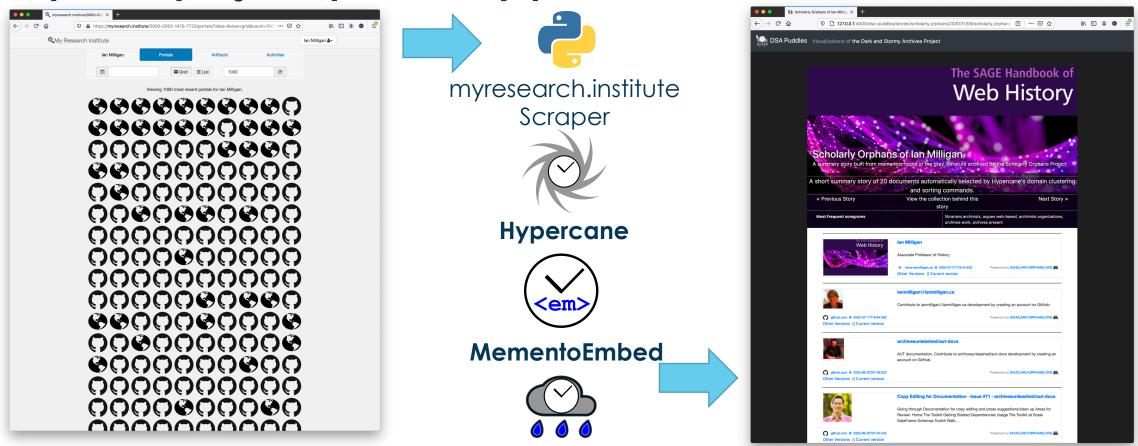


	ub.io/dsa-puddles/stories/wikipedia_reference_summaries/2020/1 🗉 🚥 🔂 🕅 🕚
ddles Visualizations of the Dark and	d Stormy Archives Project
	The second se
Summary of W	ikipedia References for Ruth Bader
Ginsberg	
A summary story built from n Ginsberg	mementos found in the reference section of the Wikipedia Article Ruth Bader
A short summary stor	y of 19 documents automatically selected by Hypercane's LDA topic modeling and DSA scoring commands
« Previous Story	View the Wikipedia article Next Story » behind this story
Most frequent entities	ginsburg, ruth bader ginsburg, supreme court, washington, the supreme court
Most frequent sumgrams	supreme court justice ruth bader ginsburg, the aclu's women's rights project, supreme court associate justice ruth bader,
	pancreatic cancer, new york
	The Rehnquist Court
The Rehnquist Court	Pages displayed by permission of ABC-CLIO. Copyright.
=	C booksprovek.com 9 (2006)-1911/328312 Preserved by APCHRE.OND C
	Ginsburg, Ruth Bader   Federal Judicial Center
	Born 1933 in Brooklyn, NYFederal Judicial Service: Judge, U.S. Court of Appeals for the
Federal Judicial Cen	Iter District of Columbia CircuitNominated by Jimmy Carter on April 14, 1980, to a seat vacated by Harold Leventha
JE www.tjc.gov @ 2018-04-29T15:55 Other Versions    Current version	
	JUDGE GINSBURG'S GIFT
The Washington Post	By Linda K. Keber 90, Linda K. Kerbar August 1, 1993 On the first day of her confirmation hearings, Ruh Baber Ginsburg told the story of Gwendolyn Hoyt. Hoyt was convicted in 1957 of the munder of
bip www.washingtorpost.com @ 201 Other Versions    Current versio	

To this: a sample providing new insight into the references not seen in the article

#### We can summarize a scholar's grey literature from the web archive collections of the Scholarly Orphans project, previously presented at CNI





From this: 1000 mementos of grey literature from the Scholarly Orphans project

Raintale

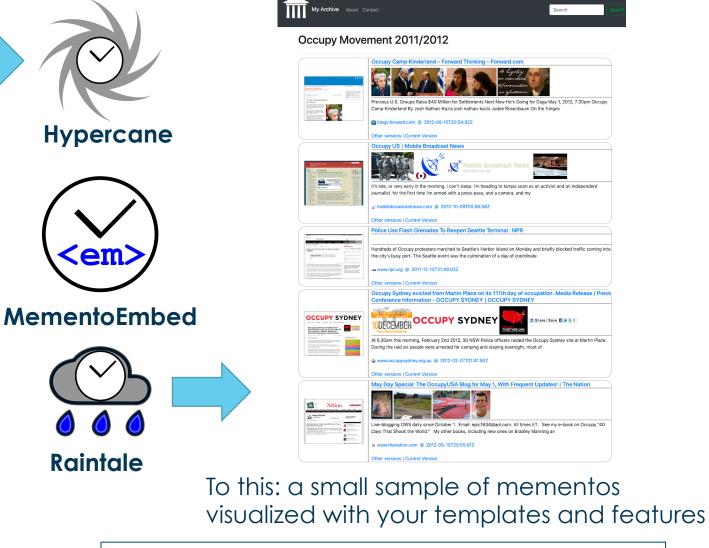
To this: 20 mementos representing a scholar's work

## You can automatically summarize your web archive collection here



From this: your web archive collection

🕥 @shawnmjones @StormyArchives



Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <u>https://arxiv.org/abs/2008.00137</u>



## For More Information on Dark and Stormy Archives



It was a dark

and stormy night.



- Dark and Stormy Archives Project
  - § Website: https://oduwsdl.github.io/dsa/
  - § Example visualizations: <u>https://oduwsdl.github.io/dsa-puddles/</u>



δ

ξ

- Hypercane Intelligently Sample Mementos
  - § Website: <a href="https://oduwsdl.github.io/hypercane/">https://oduwsdl.github.io/hypercane/</a>
  - § Documentation: <a href="https://hypercane.readthedocs.io/en/latest/">https://hypercane.readthedocs.io/en/latest/</a>
  - § Source code: <u>https://github.com/oduwsdl/hypercane</u>



- Raintale Tell Stories with Many Mementos
  - § Website: <a href="https://oduwsdl.github.io/raintale/">https://oduwsdl.github.io/raintale/</a>
  - § Documentation: https://raintale.readthedocs.io/en/latest/
  - § Report Issues: <u>https://github.com/oduwsdl/raintale/issues</u>
  - § Technical Report: https://arxiv.org/abs/2008.00137



- MementoEmbed Summarize Individual Mementos
  - § Documentation: <a href="https://mementoembed.readthedocs.io/en/latest/">https://mementoembed.readthedocs.io/en/latest/</a>
  - § Report Issues: <a href="https://github.com/oduwsdl/MementoEmbed/issues">https://github.com/oduwsdl/MementoEmbed/issues</a>
  - § Technical Report: https://arxiv.org/abs/2008.00137