Los Alamos
NATIONAL LABORATORY
EST. 1943

LA-UR-20-26143
Approved for public release; distribution is unlimited.

Title: MementoEmbed and Raintale for Web Archive Storytelling

Author(s): Jones, Shawn Morgan
Klein, Martin
Weigle, Michele C.
Nelson, Michael L.

Intended for: Web Archiving and Digital Libraries 2020, 2020-08-05 (Wuhan, China)

Issued: 2020-08-11

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

MementoEmbed and Raintale for Web Archive Storytelling

Shawn M. Jones · Martin Klein
Los Alamos National Laboratory
Research Library Prototyping Team

Michele C. Weigle · Michael L. Nelson
Old Dominion University
Web Science and Digital Libraries Research Group

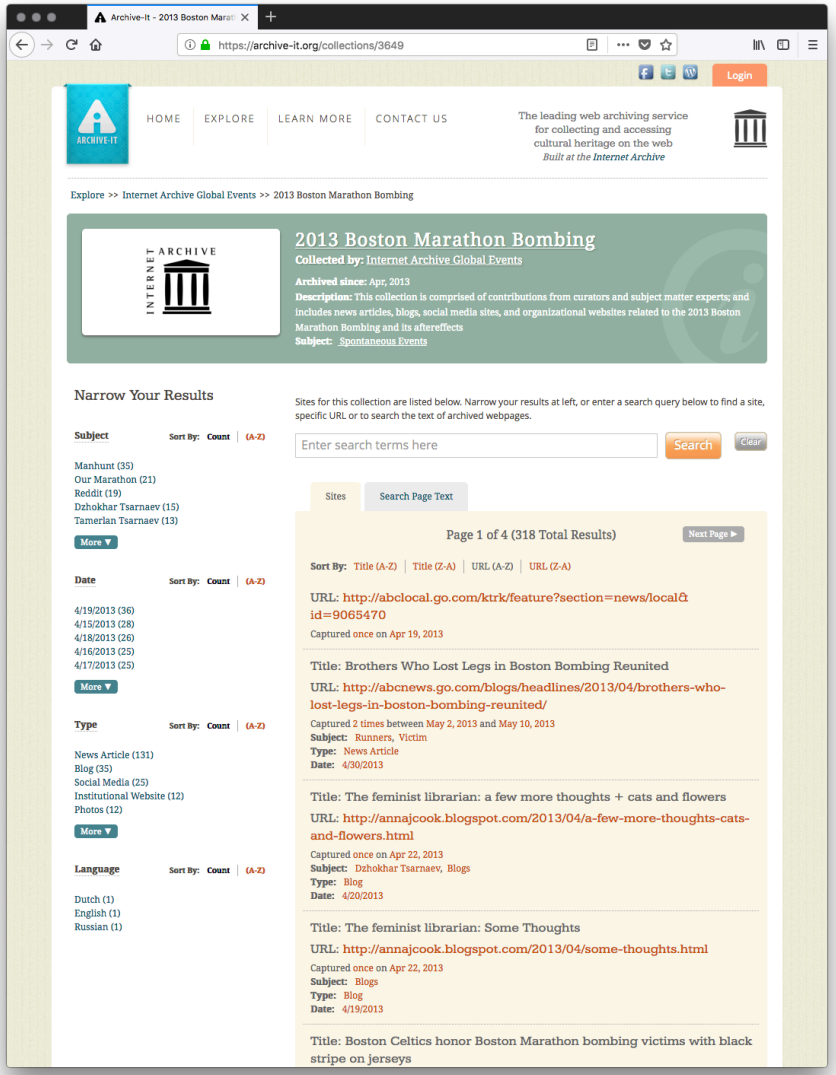
Thanks to:

INSTITUTE of Museum and Library SERVICES

@shawnmjones @WebSciDL @StormyArchives

Storytelling With Web Archives Has Multiple Use Cases…

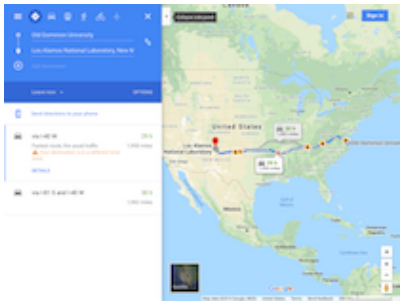
1. **Promotion of the collection**
   - Storytelling allows a curator to promote a collection, making others aware of it
2. **Exploring aspects of the collection**
   - Users can explore a collection and expose specific sides of a news story or focus on specific people or places
3. **Summarization**
   - Web archive collections are too large for manual review – we need a summary to understand what they contain



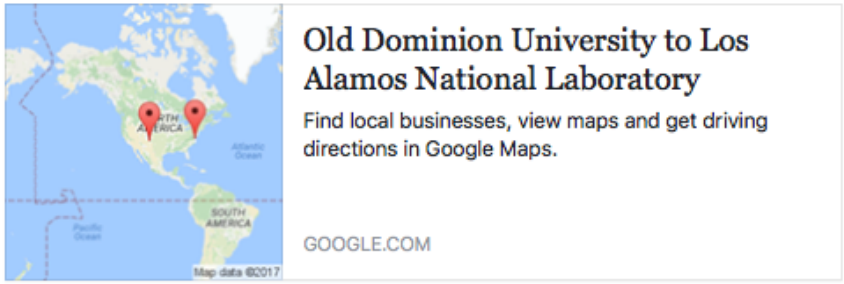
Surrogates provide a visual summary of the content behind a URI…

Long URI: 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!1 s0x89ba99ad24ba3945:0xcd2bdc432c4e4bac!2m2!1d-76.3067676!2d36 .8855515!1m5!1m1!1s0x87181246af22e765:0x7f5a90170c5df1b4!2m2!1 d-106.287162!2d35.8440582

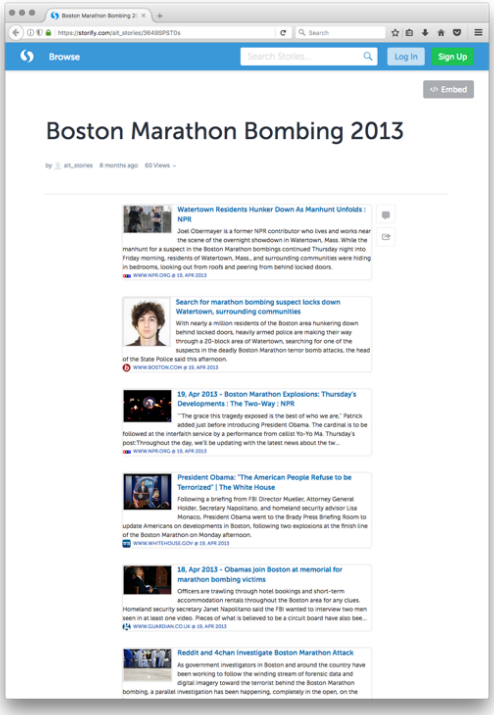
The same URI represented by a browser thumbnail surrogate:



The same URI represented by a social card surrogate:



Social media storytelling uses groups of surrogates to provide a "summary of summaries"



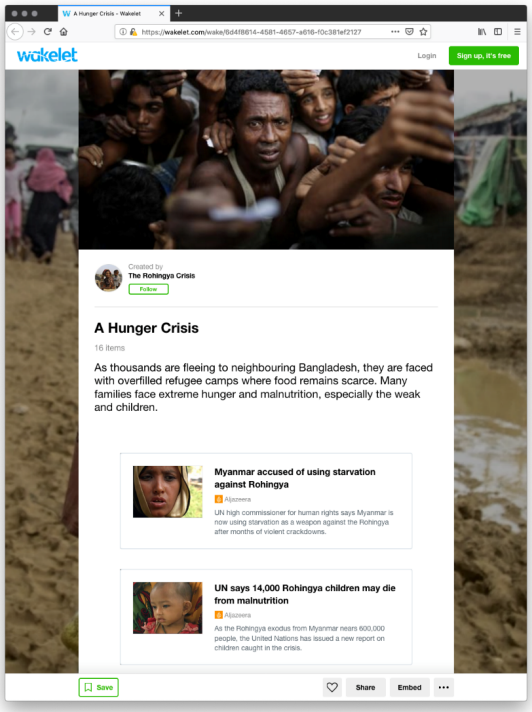
6 resources are shown in this Storify story

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

Telling a story with mementos has 3 basic steps

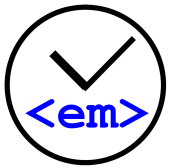
1. Select the pages for the story

Automatically with **Hypercane**

-or-

By People

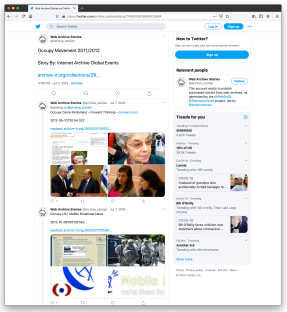
2. Gather information to summarize each memento



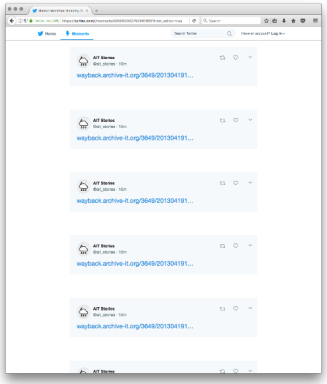
MementoEmbed

3. Summarize all mementos together and publish the story

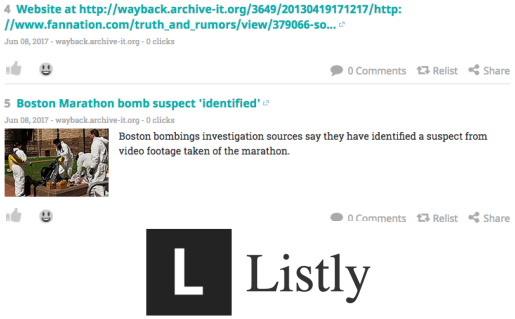
Raintale



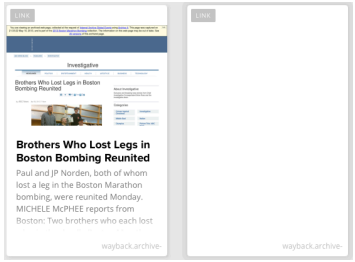
Existing social platforms do not reliably produce surrogates for mementos



twitter



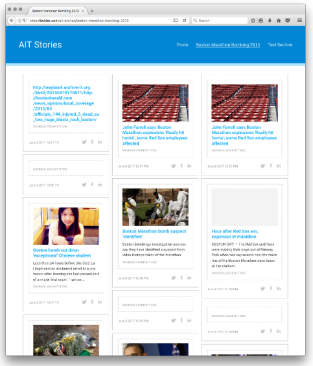
Listly



Channelkit

Instagram

pearltrees



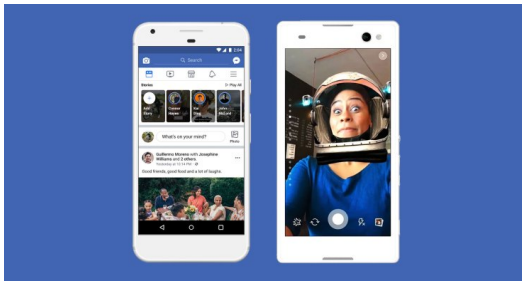
flockler

If we cannot rely upon the service to generate a social card for a memento, our system must then do the work to create our own.

@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "Where Can We Post Stories Summarizing Web Archive Collections?" https://ws-dl.blogspot.com/2017/08/2017-08-11-where-can-we-post-stories.html, 2017.

Some services have stories, but not long term storytelling?

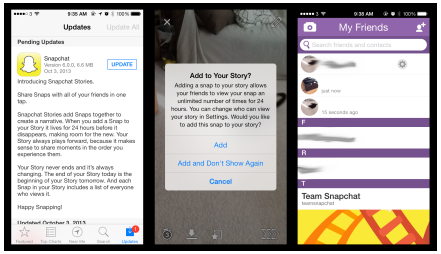


Facebook stories

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

Instagram stories

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



Snapchat stories

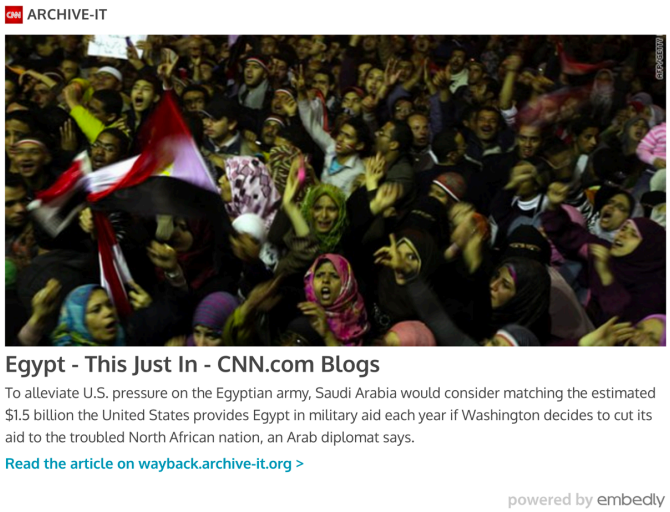
Image ref:
https://techcrunch.com/2013/10/03/snapchat-gets-its-own-timeline-with-snapchat-stories-24-hour-photo-video-tales/

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.

S. M. Jones. "Raintale – A Storytelling Tool for Web Archives." https://ws-dl.blogspot.com/2019/07/2019-07-11-raintale-storytelling-tool.html, 2019.

Existing card services create a confusing experience for pages from web archives



embed.ly social card

embed.rocks social card

Who published these resources?
Archive-It?
CNN?
Is the story author sharing fake news?

@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "A Preview of MementoEmbed: Embeddable Surrogates for Archived Web Pages." https://ws-dl.blogspot.com/2018/08/2018-08-01-preview-of-mementoembed.html, 2018.

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

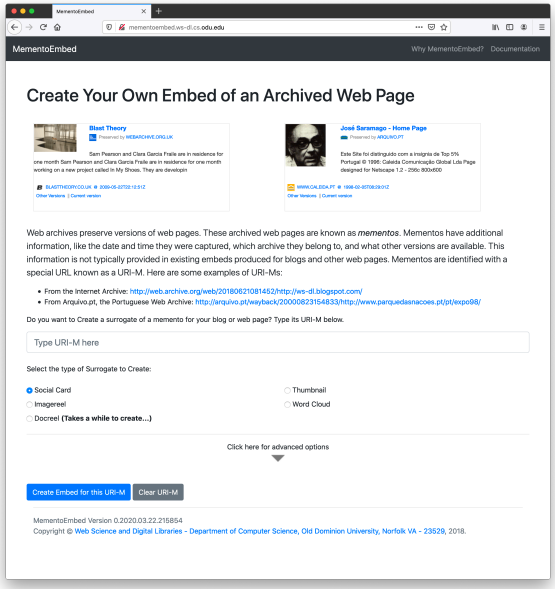
MementoEmbed is archive-aware. It can locate information about the memento that is not available in other cards.

Information in the MementoEmbed social card is separated to avoid issues of confusion about attribution.

@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "A Preview of MementoEmbed: Embeddable Surrogates for Archived Web Pages." https://ws-dl.blogspot.com/2018/08/2018-08-01-preview-of-mementoembed.html, 2018.

Summarizing *Individual* Pages With MementoEmbed



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

Cards

Browser Thumbnails

Imagereels

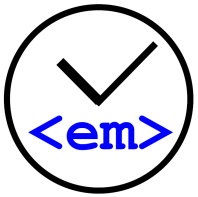
Word Clouds

Docreels

@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "A Preview of MementoEmbed: Embeddable Surrogates for Archived Web Pages." https://ws-dl.blogspot.com/2018/08/2018-08-01-preview-of-mementoembed.html, 2018.

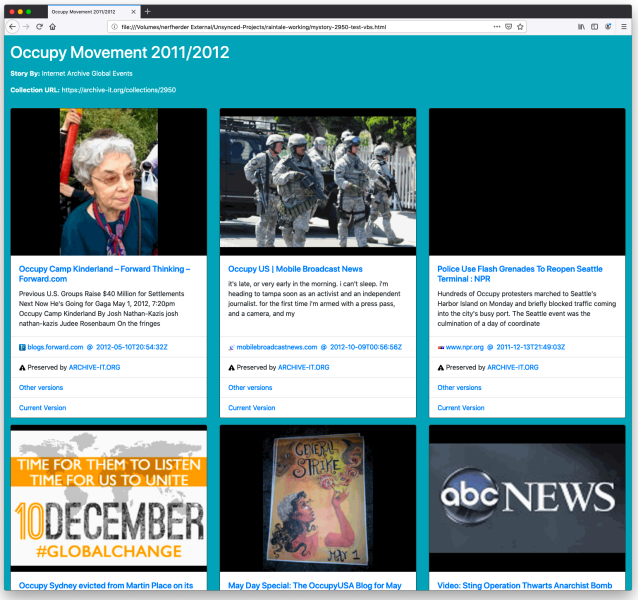
MementoEmbed's API



- Titles and text snippets
- Paragraph and sentence ranking
- Image analysis and scoring
- Thumbnails, imagereels, wordclouds, docreels
- Seed metadata, if present
- Original resource favicons, domains, URIs, current link status
- META tag content

MementoEmbed just produces cards and information for individual pages, so we created Raintale to tell stories…

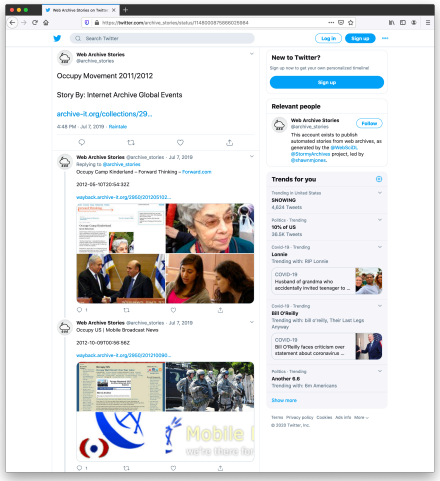
- Raintale is a client of MementoEmbed's API
- Raintale generates file output or publishes content to services like Twitter
- Raintale takes a text file consisting of the URLs for the story:
  - This file could be generated by Hypercane
  - This file could also be generated by an end user! You can manually select memento URLs to insert into the story.



@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "Raintale – A Storytelling Tool for Web Archives." https://ws-dl.blogspot.com/2019/07/2019-07-11-raintale-storytelling-tool.html, 2019.

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



Twitter Threads
with screenshots,
images, titles, dates, links

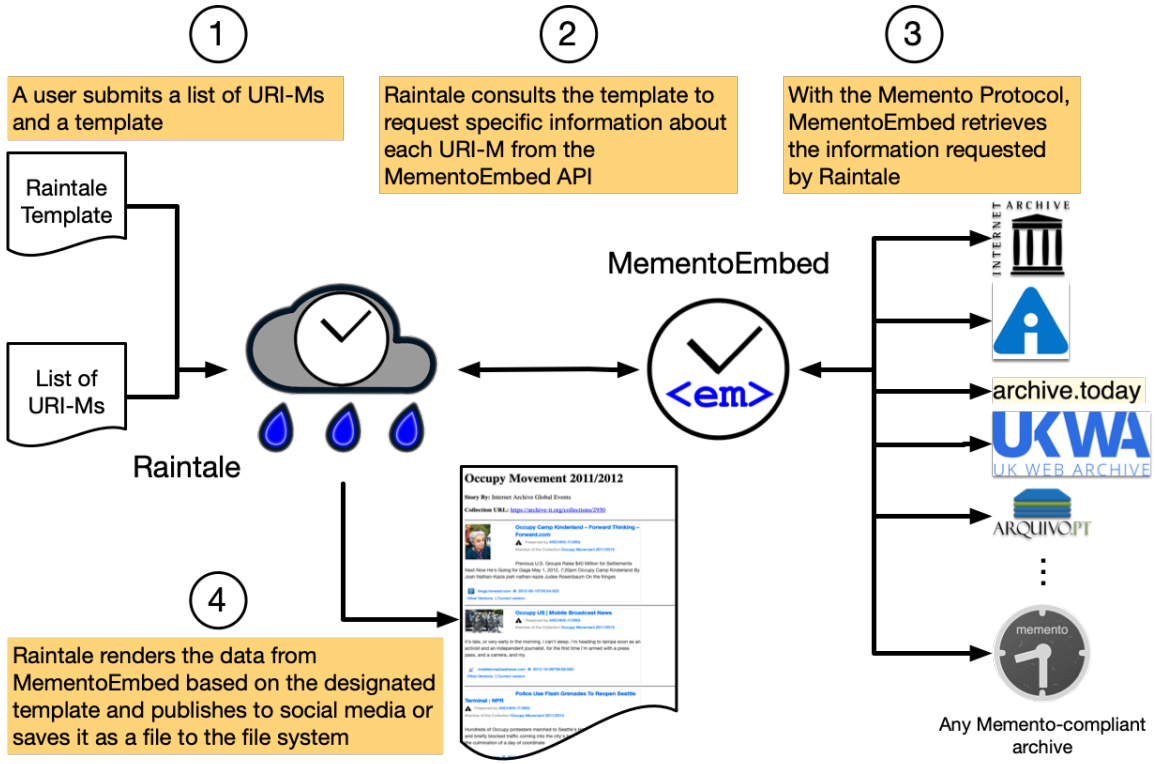
GitHub Pages
with cards and much more

Videos
with top ranked images and sentences
to be uploaded to YouTube, Twitter, etc.
(experimental)

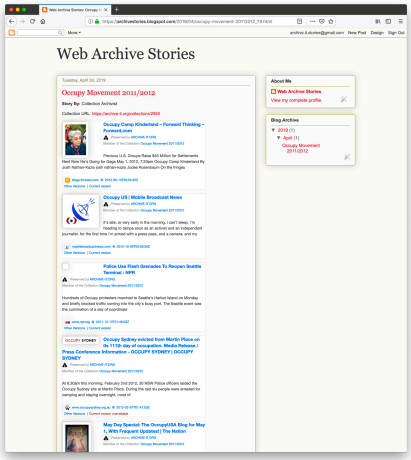
@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "Raintale – A Storytelling Tool for Web Archives." https://ws-dl.blogspot.com/2019/07/2019-07-11-raintale-storytelling-tool.html, 2019.

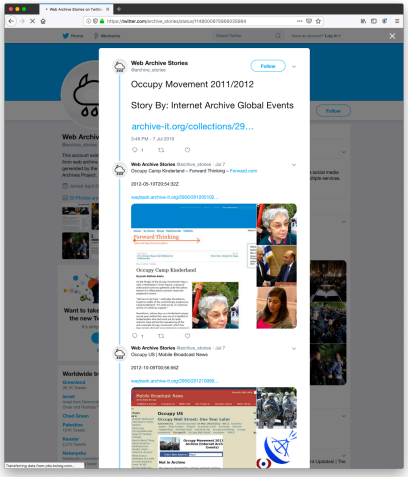
Raintale can do this because it is a client of MementoEmbed's API



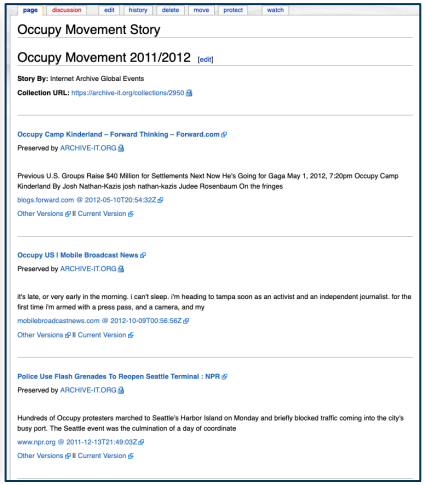
Raintale uses templates to let you customize the look and destination media of your story



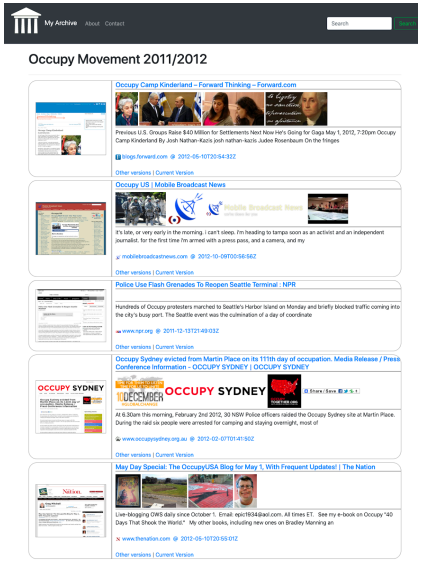
MementoEmbed cards in Blogger



Extracted components via Twitter Thread



Extracted components via MediaWiki Pages



Extracted components as Bootstrap cards used by the fictional "My Archive"

@shawnmjones @WebSciDL @StormyArchives

S. M. Jones. "Raintale – A Storytelling Tool for Web Archives." https://ws-dl.blogspot.com/2019/07/2019-07-11-raintale-storytelling-tool.html, 2019.

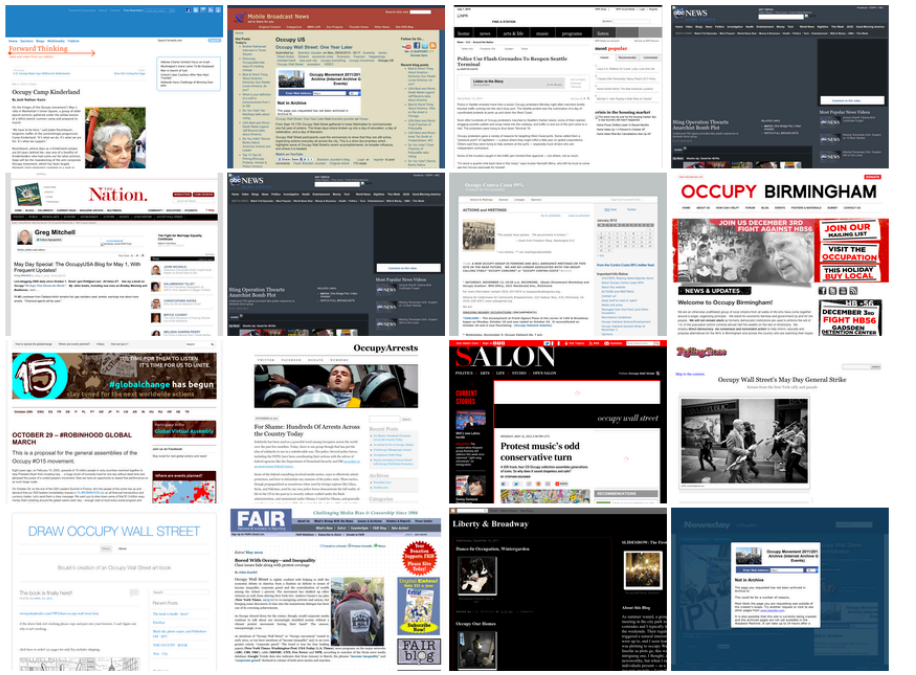
With Raintale templates the story can contain more than just cards

```
<p><h1>{{ title }}</h1></p>

{% if generated_by is not none %}
<p><strong>Story By:</strong> {{ generated_by }}</p>
{% endif %}

{% if collection_url is not none %}
<p><strong>Collection URI:</strong> <a href="{{
collection_url }}">{{ collection_url }}</a></p>
{% endif %}

<hr> <table border="0"> <tr>

{% for element in elements %}
{% if element.type == 'link' %}

<td><a href="{{ element.surrogate.urim }}"><img src="{{
element.surrogate.thumbnail|prefer remove_banner=yes
}}"></a></td>

{% if loop.index is divisibleby 4 %}

</tr><tr>

{% endif %}

{% endif %}

{% endfor %} </table>
```

Occupy Movement 2011/2012

Story By: Internet Archive Global Events

Collection URL: https://archive-it.org/collections/2950

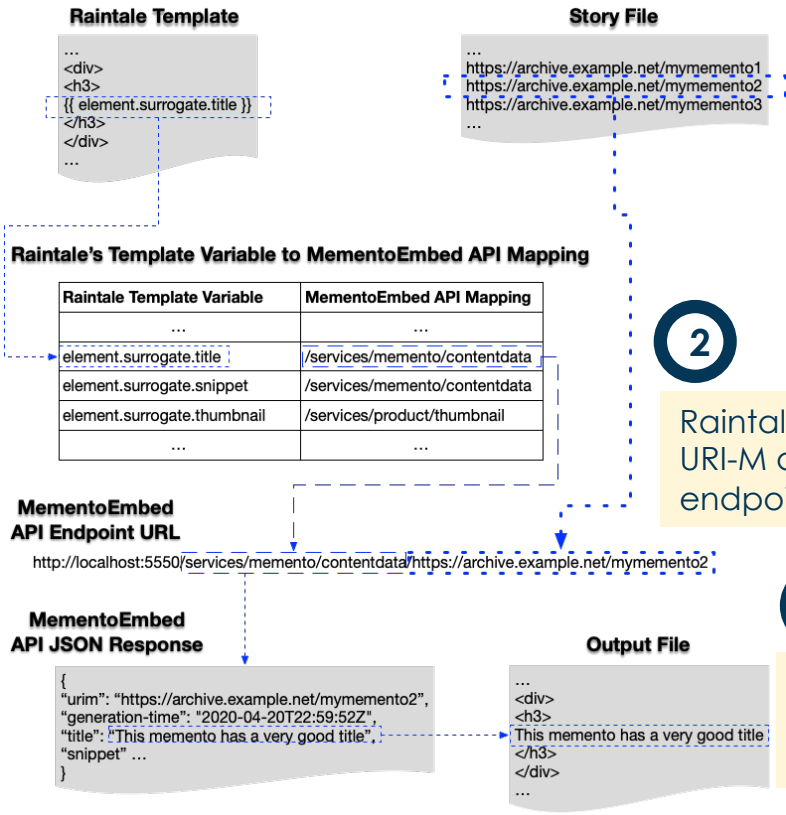


@shawnmjones @WebSciDL @StormyArchives

Raintale maps variables from the template to MementoEmbed API endpoints to generate data

1

Raintale extracts all variables from the template and maps them to MementoEmbed API endpoints



2

Raintale iterates through each URI-M and appends it to the endpoint

3

Raintale queries the endpoint for the variable and replaces it in the template

@shawnmjones @WebSciDL @StormyArchives

Future Work

MementoEmbed

- More summarization algorithms for generating page descriptions:
  - LexRank
  - Luhn's Algorithm
  - Edmundson's Algorithm
  - Latent Semantic Analysis
  - SumBasic
  - Kullback-Lieber Sum
- Allow clients to specify boilerplate removal method
  - Boilerpipe
  - Beautiful Soup
- Allow users to upload their own surrogate template

Raintale

- In addition to MementoEmbed, support for other web archiving services that provide memento information:
  - Carbon Date
  - Memento Damage
  - Archival Fixity Information
- Raintale GUI
- Direct publishing to
  - Blogger
  - GitHub
  - MediaWiki
  - Instagram
  - Tumblr
  - Pinterest
- Other output formats:
  - PowerPoint
  - PDF

For More Information on Dark and Stormy Archives

- Dark and Stormy Archives Project: https://oduwsdl.github.io/dsa/

- Raintale:
  - Website: https://oduwsdl.github.io/raintale/
  - Documentation: https://raintale.readthedocs.io/en/latest/
  - Report Issues: https://github.com/oduwsdl/raintale/issues
  - Technical Report: https://arxiv.org/submit/3303795

- MementoEmbed:
  - Documentation: https://mementoembed.readthedocs.io/en/latest/
  - Report Issues: https://github.com/oduwsdl/MementoEmbed/issues
  - Technical Report: https://arxiv.org/submit/3303795