SHARI- An Integration of Tools to Visualize the Story of the Day

Shawn M. Jones

Alexander C. Nwala  
*Old Dominion University*

Martin Klein

Michele C. Weigle  
*Old Dominion University*

Michael L. Nelson  
*Old Dominion University*

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs

Part of the Computer Sciences Commons

**Original Publication Citation**

SHARI – An Integration of Tools to Visualize the Story of the Day

Shawn M. Jones  
smjones@lanl.gov  
Los Alamos National Laboratory  
Los Alamos, New Mexico

Alexander C. Nwala  
anwala@cs.odu.edu  
Old Dominion University  
Norfolk, Virginia

Martin Klein  
mklein@lanl.gov  
Los Alamos National Laboratory  
Los Alamos, New Mexico

Michele C. Weigle  
mweigle@cs.odu.edu  
Old Dominion University  
Norfolk, Virginia

Michael L. Nelson  
mln@cs.odu.edu  
Old Dominion University  
Norfolk, Virginia

ABSTRACT

Tools such as Google News and Flipboard exist to convey daily news, but what about the news of the past? In this paper, we describe how to combine several existing tools and web archive holdings to convey the “biggest story” for a given date in the past. StoryGraph clusters news articles together to identify a common news story. Hypercane leverages ArchiveNow to store URLs produced by StoryGraph in web archives. Hypercane analyzes these URLs to identify the most common terms, entities, and highest quality images for social media storytelling. Raintale then takes the output of these tools to produce a visualization of the news story for a given day. We name this process SHARI (StoryGraph Hypercane ArchiveNow Raintale Integration). With SHARI, a user can visualize the articles belonging to a past date’s news story.

KEYWORDS

news, web archives, memento, storytelling, visualization, summarization

1 INTRODUCTION

AlNoamany et al. [1] introduced how to use social media storytelling to summarize web archive collections. Collections on specific topics exist at various web archives [6]. Klein et al. [7] have built collections from web archives by conducting focused crawls. Jones developed Hypercane [4] to intelligently sample mementos from larger collections. Jones also developed Raintale [3] for generating social media stories to summarize groups of mementos, providing visualizations that employ familiar techniques, like cards, that require no training for most users to understand. What if we want to tell stories from web archives with semi-current news articles?

Nwala et al. [9, 10] have focused on finding seeds within search engine result pages (SERPs), social media stories, and news feeds. As part of this research, Nwala et al. also developed StoryGraph [11], a tool that analyzes multiple news sources every ten minutes and automatically determines the news story or stories that dominate the media landscape at that time. Aturban et al. developed ArchiveNow [2], a tool that accepts live web URI-Rs and submits them to web archives to produce memento URI-Ms. We have tied StoryGraph together with tools from the Dark and Stormy Archives Toolkit1 to produce visualizations summarizing the biggest StoryGraph story of a given day.

The StoryGraph Hypercane ArchiveNow Raintale Integration (SHARI) [5] process automatically creates stories summarizing news for a day. Figure 1 details what each tool contributes to the story. Figure 2 shows the steps of the SHARI process. In step 1, with the StoryGraph Toolkit, we query the StoryGraph service for the URI-Rs belonging to the biggest story of the day. In step 2, Hypercane converts these URI-Rs to URI-Ms by first querying the LANL Memento Aggregator via the Memento Protocol [12]. For each URI-M that does not have a memento, Hypercane creates a memento by calling ArchiveNow [2]. In step 3, Hypercane runs the mementos through spaCy3 to generate a list of named entities, sorted by frequency. In step 4, Hypercane runs the mementos through sumgram [8] and generates a list of sumgrams, sorted by frequency. In step 5, Hypercane scores all of the mementos’ embedded images. In step 6, Hypercane runs the mementos through newspaper3k4 to extract each article’s publication date and orders the URI-Ms by that date. In step 7, Hypercane consolidates the entities, terms, image scores, and ordered URI-Ms into a JSON file containing the structured data for the summary. During this step, Hypercane uses the highest scoring image as the striking image for the summary. In Figure 1,

1https://oduwsdl.github.io/dsa/software.html
3https://spacy.io/
4https://newspaper.readthedocs.io/en/latest/
the highest-ranking image is the UK Prime Minister addressing his country about the COVID-19 pandemic. In step 8, Raintale renders the output as Jekyll HTML based on the contents of this JSON file, a template file, and information on each memento provided by MementoEmbed. In step 9, the SHARI script publishes the summary story to GitHub Pages for distribution. Figure 3 shows the output of our dsa_tweeter bot which announces the story after publication.

3 SUMMARY AND FUTURE WORK

SHARI produces a familiar yet novel method of viewing news for a day in the past. SHARI can create stories for today, yesterday, and back to StoryGraph’s creation on August 8, 2017. It is different from other storytelling services like Wakelet6 because SHARI is entirely automated. The stories produced by SHARI are different from services like Google News6 or Flipboard7 because those tools focus on current events and personalized topics. Because StoryGraph samples content from multiple sides of the political spectrum, the SHARI process can provide a visualization of articles not tied to one interest area or even a single side’s terminology. This process works because each component is loosely coupled, has high cohesion, has explicit interfaces, and engages in information hiding. Each command passes data in the expected format to the next.

We are exploring how to produce and render other news stories for a given day and any given period of time. We are researching how to best visualize significant events that span substantial periods of time, like the entire COVID-19 news story. Though StoryGraph is an existing service that gathers current news, we also want to apply its algorithm directly to mementos and tell the news stories of past events like the Hurricane Katrina disaster. One day, through SHARI, historians, journalists, and other researchers may glance at the news for any date.

REFERENCES


