Media-N

Top Two News Words (By Hour): Ten Years of Reduction and Reflection

RICK VALENTIN
Assistant Professor of Arts Technology, School of Music, College of Fine Arts, Illinois State University

ABSTRACT

The *Top Two News Words* project began in 2007 as a gallery piece featuring a computer and dot matrix printer linked to an online parsing routine which gathered headlines from fifteen major news sources hourly, and analyzed and reduced these headlines to the two most frequently occurring words. The resulting pairs were printed each hour on a continuous sheet of computer paper, creating a linear document of the 24/7/365 news cycle. Since 2008, the online component of the piece has been running automatically, without its physical half, publishing hourly word pairs via RSS and on Twitter and building an online archive of nearly 90,000 hours of news.

Top Two News Words has consistently evoked questions of bias from its audience: "Why only these sources? Why only sources in English? Who are you to decide what is a major news source?" This is, of course, one of the desired outcomes of the project. A deeper question, which is reflected in the recent controversy and surprise over Facebook's use of human curators for trending topics, is why don't we investigate for bias in supposedly neutral online news aggregators such as Google? And, is it even possible to filter news programmatically without bias? I seek to use this project to illustrate the simple concept that curation, bias and reduction are not the antithesis of awareness in a world of continuous, direct news but are an essential part of navigating and understanding this world.

The *Top Two News Words* project began in 2007 as a gallery piece featuring a computer and dot matrix printer linked to an online parsing routine. It gathered headlines from fifteen major news sources hourly, and analyzed and reduced these headlines to the two most frequently occurring words. The resulting pairs were printed each hour on a continuous sheet of computer paper, creating a linear document of the 24/7/365 news cycle.

The piece grew from a personal frustration with the never-ending stream of news available online via RSS (Really Simple Syndication). I found myself spending large amounts of time consuming this information but having little time to digest what I had read. I had the superficial sense of being connected with the world while at the same time having a deeper feeling of isolation. I was so overwhelmed by data that I had lost the ability to place the information into a personal and historical context, and suspected that the information appearing important was actually distracting filler without substance (a complaint that now feels common and quaint ten years later).

<u>Figure 1.</u> Top Two News Words (By Hour), 2007, Computer/Printer/Paper/Software

The system was built to focus the broad expanse of the 24-hour news cycle into an easily digestible form that would allow an observer to look at it from a distance, as a whole, rather than from the inside where one could easily be overwhelmed. The resulting installation, while seemingly cold and mechanical with its monochrome computer display, impact printer and continuous paper sheet with monospaced text, fostered introspection in the viewer and initiated discussion among groups of observers: *"Why the word MEGA?"- "Oh, someone won the lottery"; "What happened in Japan?"- "A reactor leaked."* Dialogue that had disappeared because we had begun to consume our media in digital isolation was rekindled by an inert collection of the very technologies that created this isolation.

Since 2008, the online component of *Top Two News Words* has been running automatically, without its physical half, publishing hourly word pairs via RSS and on Twitter and building an online archive of nearly 90,000 hours of news. While this transition to an exclusively digital format has eliminated the spontaneous real world conversations that occurred around the physical installation, the project has maintained its ability to invoke a level of reflection in the viewer that cannot be achieved through the continuous digital stream of current events crossing our social media feeds. A viewer can quickly scroll back through the past, remembering events that were so important at the time but have quickly disappeared from our minds. For example, these select pairs from 2015: PARIS HEBDO, ALPS GERMANWINGS, SHOOTING OREGON, mark high profile moments from recent history that have disappeared from our collective digital consciousness almost as quickly as they appeared.

```
TOP TWO NEWS WORDS (BY HOUR) 2016            TRUMP CLINTON          TRUMP CLINTON
                                              TRUMP PHILLY          TRUMP OBAMA
                                              TRUMP OBAMA           TRUMP SYRIA
                  TRUMP OBAMA                 TRUMP OBAMA           TRUMP CLINTON
                  TRUMP OBAMA                 TRUMP OBAMA           TRUMP CLINTON
                  TRUMP OBAMA                 TRUMP OBAMA           TRUMP PHILLY
                  TRUMP CLINTON               TRUMP OBAMA           TRUMP OBAMA
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP OBAMA
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP OBAMA
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP OBAMA
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP OBAMA
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               TRUMP CLINTON         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         TRUMP CLINTON
                  TRUMP CLINTON               CLINTON TRUMP         CLINTON TRUMP
                  TRUMP CLINTON               CLINTON TRUMP         CLINTON TRUMP
                  TRUMP CLINTON               CLINTON TRUMP         CLINTON TRUMP
```

Figure 2. Top Two News Words (By Hour), 2016, Web Site.

From a less granular perspective, yearly top twos compiled from the data can elicit complex narratives in the reader's mind:

2016: TRUMP CLINTON (as of September 2016)
2015: OBAMA SHOOTING
2014: OBAMA UKRAINE
2013: OBAMA SYRIA
2012: OBAMA SYRIA
2011: OBAMA LIBYA
2010: OBAMA CHINA
2009: OBAMA HEALTH
2008: OBAMA IRAQ
2007: IRAQ BUSH

"Obama dominates from 2008-15 but his opponents in 2008 and 2012 don't make an appearance- what does "TRUMP CLINTON" augur for the election of 2016?"

"The appearance of HEALTH in 2009 and SHOOTING in 2015 reflect an interest in domestic US issues– or does it?" (Public health and shootings are not exclusively domestic concerns).

Even the greatest reduction, comparing the first hourly pair of the project (in 2007), IRAN IRAQ, to the most recent hourly pair (as of this writing in September of 2016), CLINTON TRUMP, creates a moment of deep reflection. *How are the top issues of 2007 connected to the top issues of today?*

Reduction, filtering and abstraction are frequently viewed as negatives when speaking of recorded history– every moment lost or simplified is an opportunity to introduce bias, to change

the true narrative. Our contemporary online, crowdsourced media environment has often been upheld as a counter to the elite, traditional media gatekeepers of the 20th century (newspapers, radio, television) but this pendulum swing towards unfettered access to real time events, without the addition of informed commentary or verification from a (perceived) authority or expert has also become a point of contention. The complaint of a society blinkered by media overlords has been replaced with the concept of a society overwhelmed by too much, often frivolous information. It is obvious that information and filtering need to coexist, in balance, in order for even the most dedicated individual to navigate the online world.

As we moved towards this digital culture built on the network model of many, diverse, interconnected nodes (individual citizens, individual news events, individual eyewitnesses, individual observers linked through nearly infinite pathways) rather than a society with top-down dissemination of news, information and authority, the question became: how does one create reflection, insight and narrative from this overwhelming web of data? It is not enough to provide access to all information, we must have a method of efficiently traversing (and eliminating some of) the information we are receiving. The solution most frequently offered was that of the algorithm, an automated, unbiased filtering process to replace the human filters of the past.

Top Two News Words has consistently evoked questions of bias from its audience: *"Why only these sources? Why is it so US-centric? Who are you to decide what is a major news source?"* One of the original desired outcomes of this project in 2007 was to bring these questions to light and connect them to supposedly neutral online aggregators such as Google News. The recent controversy and surprise over Facebook's use of human curators for trending topics illustrates that this question of bias in online aggregators of news is still relevant.

```php
parsenews.php

32    );
33    foreach( $feeds as $feed) {
34      $xmlparse = new XMLParser($feed, 'link', 1);
35    if($xmlparse->data != '') {
36      $xml = $xmlparse->getTree();
37    //  echo $xml['RSS']['CHANNEL']['TITLE']['VALUE'] . "\n";
38      foreach($xml['RSS']['CHANNEL']['ITEM'] as $item) {
39        $content .= strtolower($item['TITLE']['VALUE']) . " ";
40      }
41    }
42    }
43
44    $content = ereg_replace('[^a-zA-Z ]', ' ', $content);
45
46    $freq = array();
47    $tok = strtok($content, " ");
48    while ($tok !== false) {
49    //   echo "Word=$tok<br />";
50      if((strlen($tok) > 3) || $tok == "flu" || $tok == "sex") {
51        if(array_key_exists($tok, $freq)) {
52          $freq[$tok]++;
53        } else {
54          $freq[$tok] = 1;
55        }
56      }
57      $tok = strtok(" ");
58    }
```

<u>Figure 3.</u> Top Two News Words (By Hour), 2007, Code.

Top Two News Words is, at its core, an algorithm with the same basic characteristics of larger scale news aggregation:

1. Collect current events curated by online news sources and/or curated by individuals on social media.
2. Rank each event based on its popularity.
3. Display only the highest ranking events in an easily digestible list.
4. Rinse and repeat.

There are differences of scale and complexity but there remains a common thread of reduction and simplification (which one could argue is also a core concept of journalism). So why is an artwork which uses a similar algorithm seen as biased, when algorithmic results from an online service are not? The simple answer is that art, even when technologically-based, is perceived as an individual, human statement, while algorithmic results from an faceless digital corporation are perceived as "clean" and free of human interference, a product of computing rather than a product of humankind. The strict stereotypes of art=human and technology=machine are still in play.

This explains why there was such an uproar when it was discovered that humans were involved in the curation Facebook's Trending Topics. We collectively still believe in the ideal of an impartial machine delivering unadulterated truth, much in the same way traditional news outlets of the past were believed to be unbiased conduits for world events (Walter Cronkite and CBS News come to mind).

To solve this perceived problem of bias, Facebook removed the human authored synopses of trending topics and replaced them with the raw sharing data (e.g. "Mars: NASA Rover Captures Incredible 360 View of Mars" replaced by "Mars: 4.2K people talking about this") reinforcing the impression of impersonal, automated curation, with numbers replacing narrative. [1]
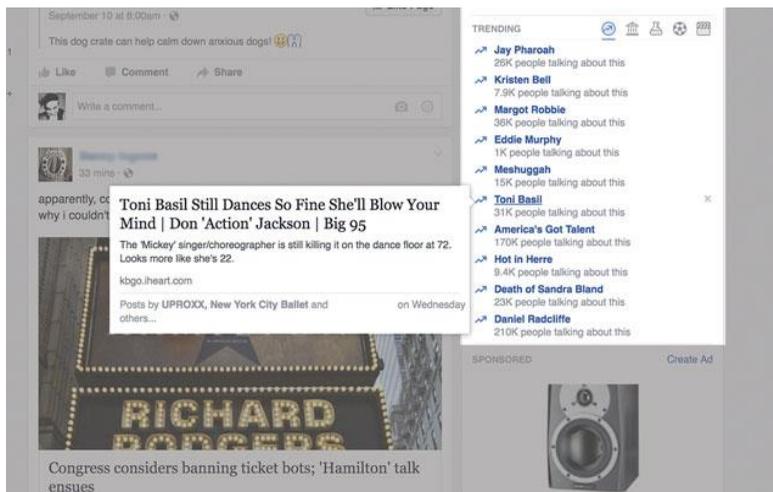
Yet, in its announcement that Trending Topics would be reworked, Facebook also admitted that humans were still part of the equation:

> *"There are still people involved in this process to ensure that the topics that appear in Trending remain high-quality — for example, confirming that a topic is tied to a current news event in the real world. The topic #lunch is talked about during lunchtime every day around the world, but will not be a trending topic." [2]*

A human, probably an engineer, made the decision that lunch was not a trending topic- it was not decided by a computer, not by an artificial intelligence, and especially not by the "crowd", who are apparently obsessed with lunch. This parallels the programming decision I made to ignore most words three letters or less when parsing the news – a news aggregator that consistently says the top two news words are "THE" and "AND" would have little resonance. At some point in the chain of curation, a human must intervene in order to make the information accessible to the outside world. Whether or not that moment occurs in a software design meeting or editorial board meeting does not reduce individual or group bias.

When challenged on the veracity of its Trending Topics, Facebook chose to obfuscate human involvement while simultaneously confirming that humans are an essential part of the process and as a result reinforced the misconception that algorithms can provide a pure truth, rather than acknowledging that algorithmic curation is an extension of the human programmer and has all the inherent pitfalls of traditional curation. This also has the danger of reinforcing the idea that programmers are more adept at curating news than those (hopefully) trained in journalism or the arts.

I do not believe we are currently in a cultural battle between engineers and artists (which may have been inferred from the above example), but we are actually in a battle of outdated perceptions and labels. One can say that the role of the contemporary artist is to interpret and reconfigure the complex terrain of modern life and foster understanding in an overwhelmed populace. The role of the programmer becomes a reducer of real-world complexity into a simplified series of commands and approximations in the form of algorithms. And because networked technology is not a simple mechanical device, but rather the conduit through which we now experience and define our world, the role of artist and role of programmer are merging in the realm of curation. This hybrid persona of programmer/artist must not be a simplistic mashup of programmatic skills used for the creation of traditional art. Instead, the artist should be versed in the programmatic methodologies that form the foundation of our digital culture and must build work from digital culture itself.

So am I saying that all the artists should learn to program and go to work at Facebook? Infiltrate the beast and reinvent it from the inside? Not at all. In fact, the example of Trending Topics reinforces the view that there is an institutional momentum within social networking empires to downplay bias and humanity at every step, a drive to reinforce the fiction of social technologies as a free networks regulated by an invisible, impartial, inhuman force. Facebook's solution for complaints of bias in their news feeds was to feed more information (view counts) to an audience already gorged with information and in desperate need of more filters and context. The advantage

of the artist's role in contemporary culture is that artists are allowed to exhibit bias, while media outlets and news are not.

In addition, simply using established online aggregators as an artistic platform does not appear to be an effective way of revealing these systems as inherently human rather mechanical. The transition of *Top Two News Words* from a physical piece to the virtual environment of Twitter resulted in a reduction of discourse, gallery conversations were replaced with superficial likes and retweets. A possible reason for this is the nature of Twitter itself; while it is a "social" network, it is not a direct replacement for all social interaction, especially when it is used as a method of consuming news. According to the Pew Research Center for Internet, Science & Tech, the audience on Twitter for news (dubbed the "Broadcast Network") *"… are often connected only to the hub news source, without connecting to one another."*[3] News on Twitter is consumed in much the same way as traditional news sources are consumed, with little interaction, making the network an ineffective method for sparking analytical discussions about the curation of news itself.
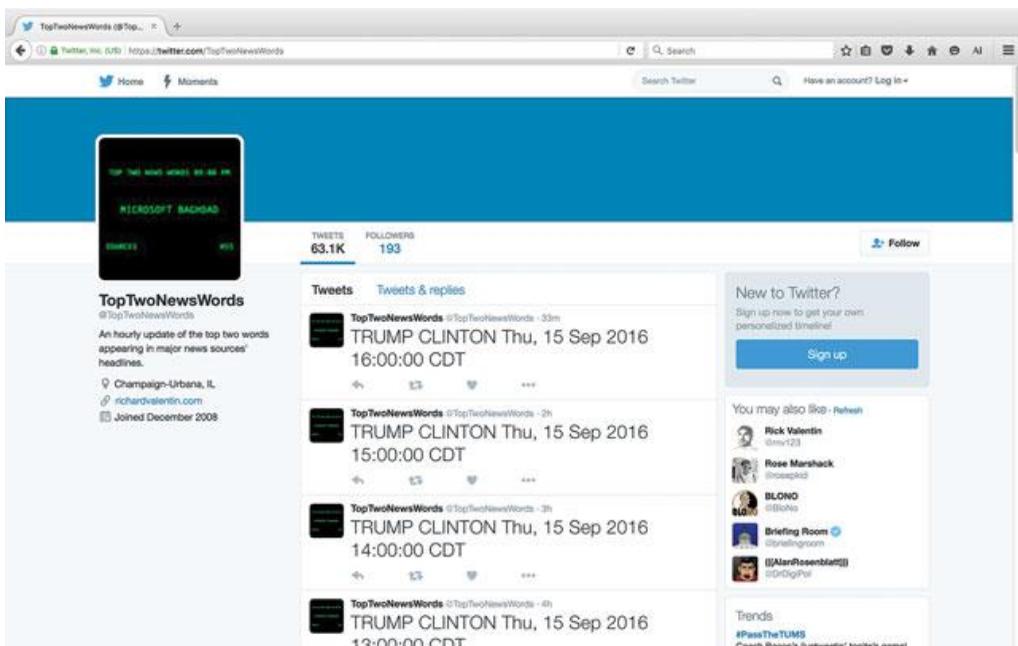

Figure 5. Top Two News Words (By Hour), Twitter Feed.

So what is the solution? I can't say that standalone projects like *Top Two News Words* are the ideal approach to explore and highlight concerns over contemporary news consumption and the consumption of information about our world as a whole, but there are some key characteristics that are worth noting. *Top Two News Words* is a parasitic work of art by its very nature. Not only does it require established online news sources to feed its simple algorithm, it requires the viewer to have an ambient awareness of current events. The word pairs are meaningless without context. The project has evolved and will evolve along with the life cycle of news curation: from its beginnings as a location based artwork, engendering discussion among viewers (like the newspaper kiosks of old); to its transition from physical to digital (like the print newspapers of

old); to its current position as a representative of algorithmic bias; to its possible future role as a document of the decay of centralized media outlets in the online world. In theory the project can run indefinitely with little maintenance. As its sources fall by the wayside, the algorithm's pairs will become less relevant and finally disappear.

Artwork meant to critique or expose the inner workings of the online world must operate within that world and adapt and evolve within this digital realm– it cannot be a precious algorithmic music box, turning in isolation, creating its own reality. If engineers treat our digital environment as merely a technical problem to be solved and if artists view this digital realm as inhuman, we will be lost in an environment defined by antiquated perceptions of technology as non-human. It is the programmer/artist's role in the 21st Century to highlight the human side of technology, bridge the gap and reveal that we are all becoming hybrids who exist simultaneously within the digital and the physical.

REFERENCES AND NOTES

1. Facebook Newsroom website, "Search FYI: An Update to Trending," August 26, 2016, accessed September 15, 2016, http://newsroom.fb.com/news/2016/08/search-fyi-an-update-to-trending/.
2. Ibid.
3. Pew Research Center website, "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters," February 20, 2014, accessed September 15, 2016, http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

AUTHOR BIO

Rick Valentin has been an Assistant Professor of Arts Technology at Illinois State University since 2010. His research and teaching focus on the collection and visualization of personal and historical data and the impact of digital media on these processes. Rick has been a professional musician for over 25 years, having released 14 albums of original music with the bands Thoughts Detecting Machines, Poster Children and Salaryman and has performed extensively throughout the United States and Europe. Technology has always played an essential role in his creative output, including pioneering work in enhanced CD design, podcasting and web presence. He holds a Bachelor of Science degree in Computer Science from the University of Illinois at Urbana-Champaign and a Master of Fine Arts in New Media from UIUC. http://www.rickvalentin.com