

Current: A News Project

<http://www.aNewsProject.com>

Zoe Fraade-Blanar
NYU Interactive Telecommunications Department
721 Broadway, 4th floor
New York, NY 10003
410.419.8183
Fraade@gmail.com

ABSTRACT

Current is a real-time data visualization of internet memes as they carry out lifecycle activities of birth, evolution and decline in reaction to the daily news cycle. By visually anthropomorphizing the capricious nature of public attention we are able to spotlight missed opportunities in news coverage, and, potentially, recover news readership that has been lost to more sensational sources.

Keywords

Algorithmic art, memes, memescape, stacked graph, news, newspapers, web, media, *Processing*, php, API's, *Current*

INTRODUCTION

Effective and hard-hitting news coverage is one of the basic foundations of democracy. It's so important that it is easy to forget that the news media dedicated to creating it is also a business. Truth and trustworthiness in news is, at heart, an advertising ploy with unexpectedly positive consequences. The existence of good, hard news coverage cannot be taken for granted.

Right now news is famously suffering from budgetary troubles, mostly because of advertising-based loss. Each visitor to a news website is a precious resource, yet at the same time readers are turning to other sources for their media input. A standard newspaper can contain less than 10% hard news coverage[9]. It relies on the rest of the paper to support the few important articles: horoscopes, celebrity scandals, wedding announcements, all of these are intended to subsidize reporters to investigate more significant but less interesting topics. However, since the advent of the Web there are faster, more targeted, bloggier sources for readers to find this all-important soft news [6].

This means that there exists a basic disconnect between traditional news sources and their readers. During the summer of 2009 I worked with the New York Times Analytics Group looking at ways to understand incoming traffic behavior. One common effect they had documented was that the newsroom sometimes celebrated a spike in traffic, for example on the day of an important political event, when the true cause for the upsurge was much more commonplace - in one case an article about Neti Pots.[10]

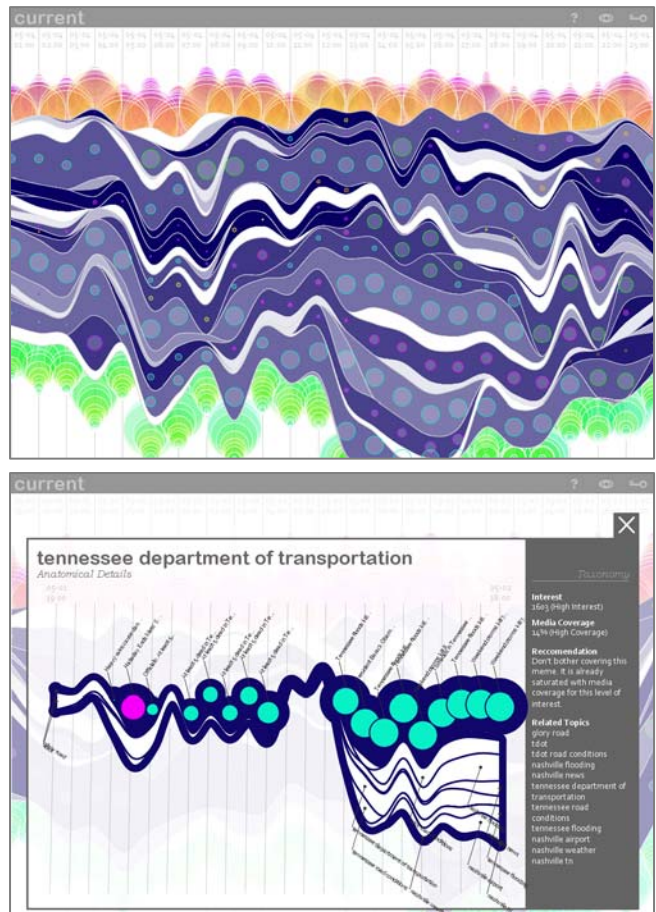


Figure 1: Primary and secondary interface screens within *Current*

BACKGROUND

Current is my third attempt to use data visualization to tackle the question of what topics news media should be covering. My first try, the *Overlooked Newsreader*, was a project that attempted to guilt editors into expanding their coverage by pointing out which international news was being ignored.[8] With *Scales*, a whimsical data-inspired environment featuring exchange rates as swimming fish, I tried to engage users in a comparison of various economies without giving enough direction for effective action. Both of these approaches failed: the correct solution cannot

simply wag a finger at moral oversight, but it also must offer some guidance on how to correct the issue.

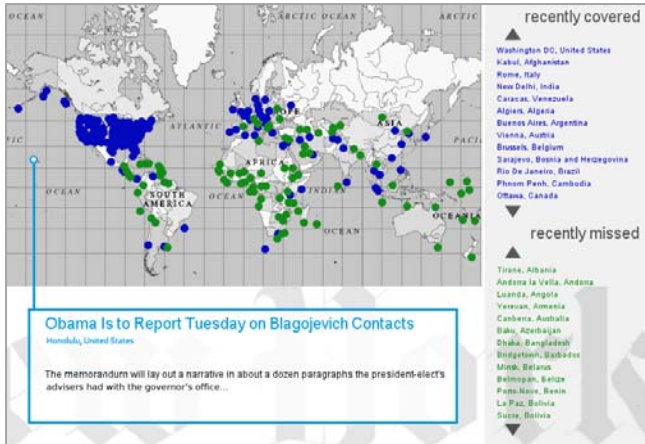


Figure 3: The Overlooked Newsreader

At base, any solution to News’ problems must offer a path to financial success in addition to advice on maintaining journalistic integrity. *Current* seeks to fulfill this purpose by supplying the ability to differentiate which news items are most likely to draw web traffic to a news source. In this way it is not the percentage of soft news that increases at the expense of hard news coverage, but the nature of the soft news itself that becomes more enticing. That way hard news may continue to stay safely alive and funded.

CONCEPT DESCRIPTION

For the purposes of *Current*, the meme is a useful metaphor for the rise and fall of public interest. In cognitive philosophy a meme is any unit of cultural information that is passed via imitation.[3] For example, religions, the words to the national anthem, or a group preference for Macs over PC’s.

Memes can be thought of as living ‘thought organisms’ with agency, control, and a selfish motivation. This means that, like the genes that inspired their naming, memes are non-judgmental about their content. A successful meme is the one that lasts longest; it is not the prettiest, most truthful, or even most helpful to a host organism. A harmful meme can wipe out a sizable percent of the population before it’s carriers might reconsider its usefulness, for example, a meme that advocate behaviors that transmit lethal diseases.[9]

Up until now a poll has been the preferred method for finding out what the public is thinking. This means sampling a small subsection even if the subjects are truthful. *Current’s* power is in showing a snapshot of what the entire internet-using population in America has been thinking about for the last 24 hours, in the form of their collective Google search history. This information is then condensed for readability into memes.

To make the purpose more clear, *Current’s* interface is a single, consolidated diagram of pre-rated content set up to show existing opportunities in under-covered memes: Each

highlighted meme represents untapped traffic that could be channeled to a specific news site if only an article existed on the topic.

SOFTWARE

User Scenario

Prior to *Current* an editor in an office at a newspaper deciding what to cover might have a number of sources for news tips – favorite blogs, searches and RSS feeds. *Current* mediates between this editor’s need to drive traffic to the website and the need to cover important, albeit less sensational topics.

For them, using *Current* should not be an intimidating experience. One of the lessons from *Scales* was that users are willing to try a beautiful application when they might be unwilling to use a boring one. The memescape is shown in consolidated, visual form because the human eye is so much better than the brain for picking up trends, an important consideration for an editor on a tight deadline.

Development Process

There are a number of different data sources that might have allowed *Current* to arrive at its final form. *Yahoo*, *Technorati*, *Twitter*, and *Bing* all display ‘Trending Topics’ lists for time periods of varying modularity. These are lists of fastest rising topics; that is to say, topics that have recently made large gains in the public imagination. This is an important differentiation from ‘most popular topics’ which for the most part are sexual or technical in nature and do not change often.

However the *Google Hot Trends*[4] list is the most attractive of these tools because it provides an hourly readout of this type of information and provides enough of it to chart the individual rise and fall of search terms. The effect was originally noticed through a simple data test during the fall 2009 ITP class *Crafting with Data*. Information from the *Hot Trends* tool was logged for 24 hours and then consolidated into memes using a basic spreadsheet macro. The result was an unusual chart of

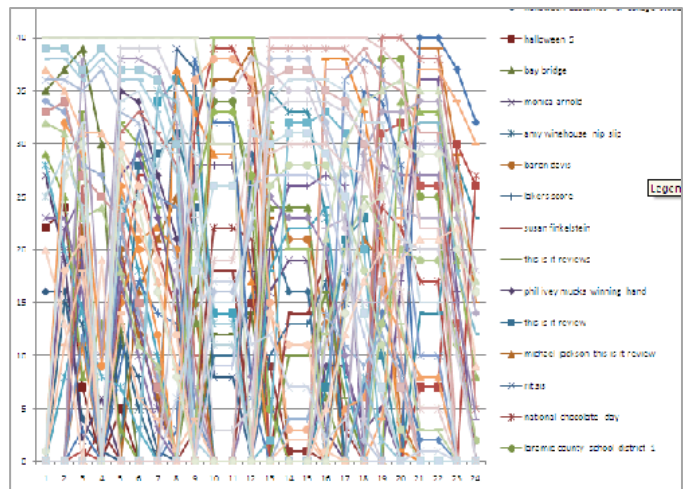


Figure 2: Initial diagram of meme behavior

upside-down loops and curves that hinted at complex and informative behavior.

An early proof of concept incorporated a simple 24-hour rolling index of meme items, cross referenced with information from the *New York Times API* [7]. This led to a form of the software that could directly display which memes had and had not received coverage during the prior day. A more advanced version was developed for the *ITP Winter Show* [5] using *Google News* for the first time to cross-reference memes with the top five largest newspapers in the United States. The interface was simplistic but effective: clicking the name of a news source displayed which memes had recently been associated. This method allowed patterns to emerge; for example, *USA Today* had a much higher percentage of coverage for these often sensational topics than a more traditional news source such as the *Wall Street Journal*.

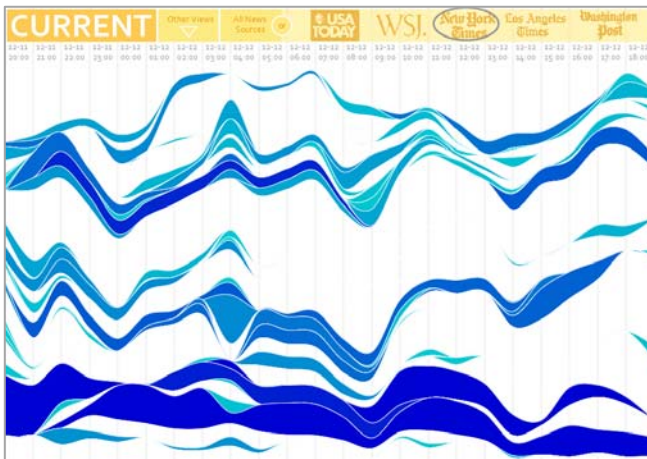


Figure 5: Early prototype of *Current* software showing only single-source coverage

Product Description

The present version of *Current* takes this interaction a step further. Instead of checking for specific source coverage, the software analyzes and stores all major hourly news coverage. This allows users to see which news stories have been successful in fertilizing meme interest, and which have been left out of the Interest Flow. The visualization is in the form of a stacked graph in which it is not the placement but the thickness and volume of each line that represents the fluctuating level of interest

To allow ease of visual scanning, each meme is colored according to the percentage of coverage it has attracted: darker memes are highly fertilized with news while lighter memes are less so. Pure virgin white memes have 0% coverage, and would be a good candidate for any editor's attention.

More complete information about each meme can be found through the Meme Inspector screen, including Taxonomy and Anatomical Details such as Related Memes, Percent Coverage, and Final Recommendations about the meme's journalistic viability

Additional functionality is added in the form of the "Eye" tool, a set of 'best bets' for an editor in a rush. This list, accessible at the top of the application along with the Question and Key tools, is comprised of the Most Successful (longest lasting) Meme, Most Fertilized Meme (highest news coverage), the Most Carnivorous Meme (the most sub-memes eaten), and a list of current memes with the lowest coverage.

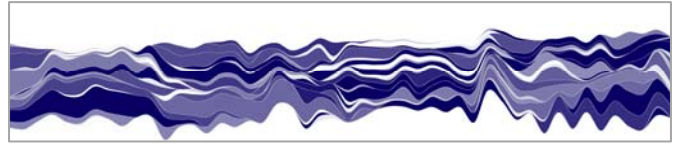


Figure 4: A 48-hour stream of meme data

This final version of *Current* can also show timeframes longer than 24 hours. Meme classification can depend on the context in which it is presented; for example, the further the chronological distance from the launch of the iPhone the more it becomes part of an overall "Smart Phones" meme and less about that particular product. The ability to show longer time intervals counteracts this effect.

Technical Specifications

The application that makes up *Current* is comprised of three distinct parts: A number of PHP scripts that do the work of collecting, processing, and serving meme and news data on an hourly basis, a large database for storage and organization, and a Processing (java) front-end application to make the data visually explorable.

Please see the Appendix for full technical details.

Testing

In designing the interface for *Current* the primary concern was usability. The amount of information displayed on any specific screen can be astronomical, and at times the very colors and shapes that have been added to make the data more lucid can become overwhelming to a user who has no prior knowledge of the application. This was the origin of the Key area at the top of the screen, which contains a complete description of sizes, colors, and keyboard actions.

An early test also uncovered that users were having difficulty understanding that the circles along the outside and within the information stream both referred to the same type of data; too much news in one place blurred the outlines of a news event and created an overall glow effect. This was solved through giving each news event an outline to modularize it from the surrounding competition.

OBSERVATIONS

Over time, an inspection of the *Current* interface displays a number of interesting phenomenon.

Despite the ongoing battle for resources between memes, the news volume or visual height of the Interest Flow remains about stable. Partly this is an artifact of the tools used, but it may also point to the limited range of human interest as a resource. In this way it is possible that *Current* demonstrates the finite nature of human attention.

Memes take a number of tactics in pursuing the longevity that marks their success, and a small number particularly achieve it through becoming highly carnivorous. That is to say, they attain bulk through the eating of smaller related memes. Because of the nature of the language processing used, in *Current*, the more carnivorous a meme, the 'hungrier' it becomes. This effect leads to at least one or two extremely bloated super-memes during any given time period.

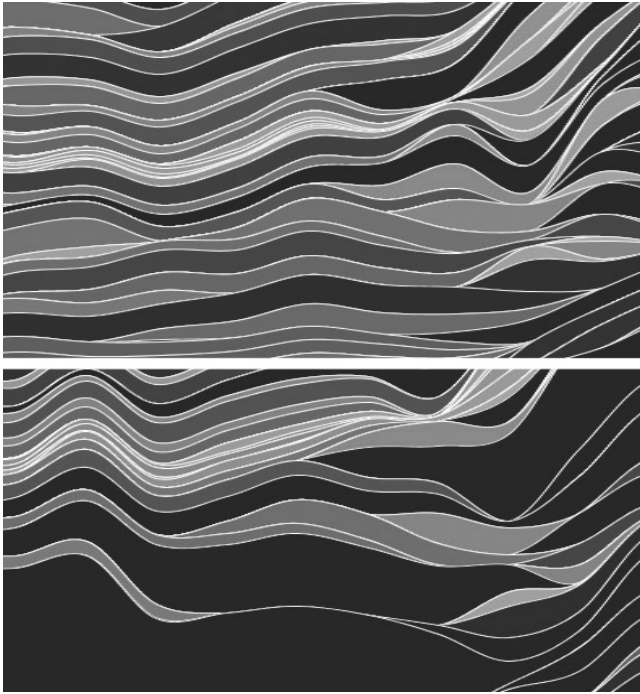


Figure 7: Carnivorous memes eating smaller, related memes, shown before and after consolidation

Because of the nature of public interest, sensational topics such as celebrity scandals and music releases are what most easily capture latent public interest. More serious but slowly-developing stories, such as health care, are less likely to fire the imagination of a large group of people all at the same time. Therefore it is the soft news such as Entertainment, Sports, and Health, depicted in the warmer spectrum of colors and entering through the top of the interface, that tend to make it into the Interest Flow most often. There is often significantly more National, International, or SciTech news in 'cooler' colors visibly pushed to the outside of the Interest Flow at the bottom of the interface. It seems that users are less inspired to repeatedly search for hard news in synchronous numbers, perhaps particularly because it is more important.

One unexpected finding from *Current* was that the heavy fertilization of a meme by news did not necessarily preclude that meme from being a good bet for coverage. A small but significant number of memes attracting heavy coverage are still found to retain a relatively light color. The implication is that there is such a massive amount of interest in these memes that, even though they have already been covered by a large number of news sources, they can

still absorb yet more. Of course, it is interesting to note that if *Current* was properly used by enough news sources, all light-colored areas would immediately become saturated by coverage.

Ongoing issues

An ongoing issue throughout the development of *Current* has been the beta status of the *Google Hot Trends* tool. What started as a top-100 rising searches tool had become a top-40 rising searches tool by the beginning of the project. Halfway through development it again was updated to the top-20 tool. While the tool now remains stable at twenty results an hour, it occasionally misses hourly updates. On at least one occasion as many as twelve hours have passed without an update, invalidating any cross referencing during that time.

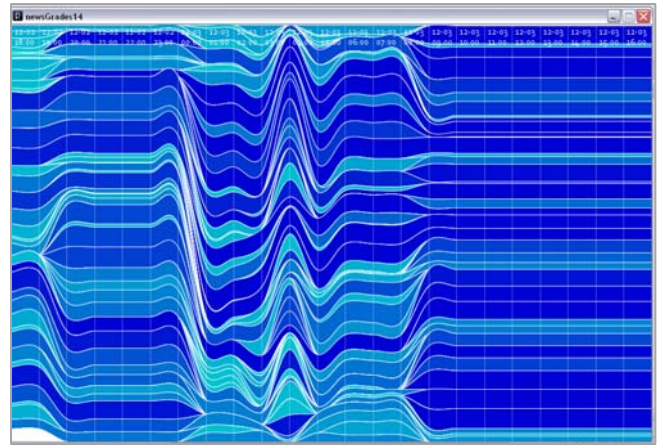


Figure 6: Pauses in Google Hot Trends Service

Another Google issue is found in the proprietary nature of Google's analytic statistics. While it is true that the information about memes derived from the *Hot Trends* tool is correct relative to other memes, the overall search volume doubtless changes throughout the span of the day. These are numbers that Google is not comfortable releasing, so early prototypes artificially accounted for this variation through sine wave motion. Future versions of *Current* should derive traffic estimates based on overall internet traffic at the time, assuming that, as the most visited website on the web, its traffic will be inversely proportional to overall web speed. It is important not to assume an evenness of search input that is, in reality, an artifact of the tool algorithm.

At the moment an additional, unexplained issue can be found in the Google RSS news feeds which have twice unexpectedly ceased feeding information during the span of the project. During both failures, data resumed flowing a number of hours later without outside input. Switching to a unique IP for the involved server did not affect the situation. In the future this effect is something that should be more thoroughly investigated.

On the Processing side of the application, because of the Bezier curves used to chart the upper and lower bounds of

each meme in the interface, the 'meme selection' algorithm is not as exact as it might be. At times, particularly for thin and highly curved areas, the mathematical line functions used to triangulate mouse events and their real-life meme shape do not perfectly overlap. This can lead to incorrect or multiple meme selection.

The final Processing application is exceptionally graphics-intensive and requires a powerful computer to display. As such it is too heavy to run as a web applet, which would otherwise be the most attractive final form.

CONCLUSIONS

Current is part of a small but growing collection of news-related technology that has been created, not to prop up a failing news infrastructure, but to help support whatever direction news takes as it continues its ongoing evolution. Success in news media in the coming years will be the domain of those institutions that are willing to act within the prevailing paradigm, and that means they will need to work with their readers and not against them. My hope is that *Current* will help them in this mission.

Acknowledgements

The *Current* project could not have been completed without the assistance of the following advisors: Rob *Faludi*, James *Robinson*, Clay *Shirky*, Daniel *Shiffman*.

APPENDIX: TECHNICAL SPECIFICATIONS

Data Sources

RSS Feeds

News data is collected from separate Google News RSS feeds

Google Hot Trends Tool

A trending topics tool from Google Zeitgist. All information is scraped from the html.

PHP Scripts

Grabber.php

A cron job-triggered script to scrape the *Google Hot Trends* tool hourly, process the information and store it in the database

Newser.php

A script to process the various Google News RSS feeds, extract keywords through the *Term Extraction Web Service API*[11] and store the information in the database

Consolidator.php

A cron job-triggered language processor to consolidate similar term information into overarching memes

Matcher.php

A script to match database news items with database term information

Harvester.php

A user-triggered script to export database information to an XML feed

Database Layout

Hourly

(date, term, score)

Hourly_News

(identity,_date, headline, keywords, category, score)

Consolidated

(date, term, score, all_terms, term_history, articles)

Processing(java)

Classes

Current.pde

DatabaseHandler.pde

Navigator.pde

TermSystem.pde

Term.pde

NewsItem.pde

SecondaryNavigator.pde

SecondaryTerm.pde

Libraries

SimpleML, an XML parsing library by Daniel Shiffman [2]

Other resources

Yahoo Term Extraction Web Service API

A web service to find keywords in news headlines

REFERENCES

1. "Current:A News Project." Binaryspark. Zoe Fraade-Blonar, 01/12/2009. Web. December 12, 2009. <http://www.anewsproject.com>.
2. Daniel, Shiffman. "Tutorials: SimpleML." Learning Processing. N.p., n.d. Web. <http://www.learningprocessing.com/tutorials/simpleml/>.
3. Dawkins, Richard. *The Selfish Gene*. New York: Oxford University Press, 1976. Print.
4. "Google Trends Hot Searches." Google. Google, n.d. Web. <http://www.google.com/trends/hottrends>.
5. "ITP Winter Show 2009." ITP.NYU.edu. N.p., November 9th, 2009. Web. 5 May 2010. <http://itp.nyu.edu/sigs/news/itp-winter-show-2009/>.
6. Milken Institute. "2008 Global Conference: The Future of Print Media: How to Adapt to the Digital Age." *Shaping the Future*. (April 28-30, 2008): DVD.
7. "New York Times Article Search API." The New York Times. The New York Times, n.d. Web. http://developer.nytimes.com/docs/article_search_api/.
8. "Overlooked Newsreader." Binaryspark. Zoe Fraade-Blonar, December 2008. Web. December 2009. <http://www.binaryspark.com/reader/>.
9. Susan, Blackmore. *The meme Machine*. Oxford: Oxford University Press, 1999. Print.
10. Sweeney, Camille. "Short, Stout, Has a Handle on Colds ." *New York Times* 1 January 2008: Web.

December 2009.
<http://www.nytimes.com/2008/01/03/fashion/03skin.html>.

11. "Yahoo Term Extraction Web Service." Yahoo Developer Network. Yahoo, n.d. Web.
<http://developer.yahoo.com/search/content/V1/termExtraction.html>.