

Public Opinion Aggregation by Annotation and Tagging of Online News Stories

Pranay Sethi

School of Information, University of Michigan

sethip@umich.edu

Keywords: sentiment analysis, public opinion mining, online commenting service, annotation, public opinion visualization

Abstract

Ubiquitous access to internet has resulted in more and more people going online to get their daily dose of news. In a 2010 survey conducted by the Pew Project for Excellence in Journalism, 41% of the respondents said they get most of their news online, 10% more than those who said they got most of their news from a newspaper. A lot of socio-technical factors have contributed to this phenomenal rise in adoption of online news in recent years. One of the biggest reasons why people are increasingly reading news online is because it facilitates discussion with peers (Nguyen 2010), offering different viewpoints which aid in forming a rounded personal opinion about the news story. The Pew survey found that 37% of online news users (and 51% of 18-29 year olds) think that commenting on news stories is an important feature to have. A lot of people tend to shape their opinion by reading discussion comments, reflective articles, blogs and even tweets about the news. Hence, an increasing number of people rely on online sources of news – be it news websites or news aggregator services like Digg, Reddit, Google Reader, Flipboard, Pulse etc.

The problem with these news websites and aggregators is that the only way people can gather public opinion is by actively searching through the endless stream of comments and feeds, filtering out spam (which is a growing problem) and then reading the relevant posts. A top trending story on Twitter will typically see multiple tweets per second, and keeping up with the rapid flow of incoming tweets is quite cumbersome and cognitively taxing. Hence it becomes increasingly difficult and time consuming for someone who wants to get the pulse of the people affected by a news story.

Furthermore, in certain scenarios people might want to look at more fine grained opinions. Currently, there is no elegant way to extract geographic and demographic impact of a news story. What is the public sentiment in Indonesia about the Arab Spring? How did the public opinion about the Wikileaks disclosures change as the story unfolded during the course of a year? It is very difficult and tedious to observe such patterns using the currently available news providers.

This work attempts to solve these problems by proposing a news aggregator platform which pulls news stories from various sources and also aggregates public responses, reflections, opinions and sentiments associated with those stories. This data is presented in ways that are easily understandable so readers can make better sense of the stories unfolding across the globe.

Such a news aggregator platform that gathers and display public opinion and sentiments about a story, must deal with various challenges –

1. Opinions are very subjective. Different people feel about a story in different ways. With such an enormous amount of diverse opinions and subjectivity, how can we possibly aggregate the responses into something that makes sense as a whole?

2. There isn't really a unified web standard for expressing opinion (in textual form). Some people tweet in 140 characters, while others write elaborate blog posts. Some websites employ tags which a reader can use to define and classify their public opinion, while others rely on threaded comments and comment ranking systems. How can a platform be flexible enough to adapt to all these varied standards so that it can extract valuable data from various sources? Perhaps the platform can create a new standard of expression on the web which is flexible and comprehensive enough to be used to express diverse views about every news story in the world.
3. How to filter out spam while extracting public opinion?
4. Once the platform has access to the data it needs, how should it be displayed to the reader in a way that makes sense? What forms of visualizations, illustrations and graphical representations can be employed to give the reader a holistic view of how people feel about a story?
5. How can the platform determine and convey effects of geographical, demographic and temporal variations as the story unfolds?

These are just a few out of possibly many issues which must be dealt with. Previous research on similar public opinion aggregation services has greatly focused on natural language processing, data mining and text categorization and clustering. Xiaojun (2010) proposed a framework for crawling the web for comments and applying various data mining algorithms on the data to extract relevant information. Diakopoulos and Shamma (2010) used tweets posted in conjunction with the live presidential debate between Barack Obama and John McCain to gauge public opinion. Brody and Diakopoulos (2011) studied the use of word lengthening to detect sentiment in microblogs.

This research proposes a solution – *The Opinionated Reader*, which relies on *sentiment tags* and annotations associated with a news story. The essential idea is to create a commenting, discussion and sharing plug-in which can be used by news websites and aggregators as a commenting solution for their news pages. Users wanting to share or comment on a news story through the plug-in are asked to tag the news story with *sentiment tags* and annotate the story with their reaction (happy/positive or sad/negative). These tags and annotations are stored, aggregated and linked to each news story. A mobile application provides the front-end interface for users to access the news stories and the aggregated sentiment associated with each story. The basic architecture is explained as follows:

The Opinionated Reader – Mobile/Tablet App

The app fetches news articles from various web sources, based on the interests and preferences configured by the user. In every news article, a portion of the screen real estate is reserved for *Opinions* which shows graphical visualizations and illustrations of the public opinion surrounding the news story. These visualizations include:

- A *Sentiment Graph* indicating the popular tags associated with the story (E.g.: “Shocking”, “Inspiring”, “Amusing” etc.). See *Figure 1* for example visualization.
- A *Positivity Graph* which plots the level of positivity associated with the story on a time scale from when the news broke. See *Figure 2* for example visualization.

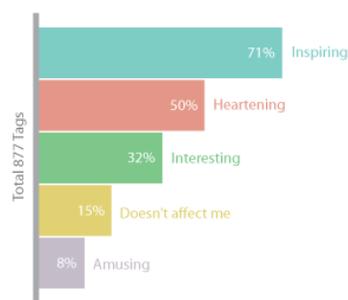


Figure 1. Example Sentiment Graph

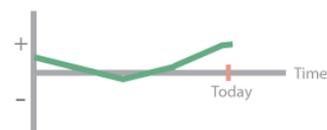


Figure 2. Example Positivity Graph

The user can choose to see these visualizations for a particular time period in the evolution history of the news story, or for a specific country. The app also facilitates people to tag and annotate news articles from within its interface.

The Opinionated Reader – Commenting and Sharing Web Plug-in

These days, a common way of adding discussion and commenting functionality to news websites is by using 3rd party services (like DISQUS). The Opinionated Reader is a similar service which can be embedded into the news articles of various news websites to enable commenting and sharing. When someone wishes to comment on an article, the comment is directed through this plug-in, which allows the users to annotate the article with the sentiment tags and reactions along with their comments. The Opinionated Reader saves this information along with the commenter's location and date of comment (See Figure 3).

Comment on this story

"Bay Area bids adieu to Discovery"

Name/Alias:

Location:

Type your comment here

Reaction to story:

Sentiment Tags: separate multiple tags by comma

Popular tags for this story: heartening, inspiring, sad, exciting, doesn't interest me

powered by The Opinionated Reader

Figure 3. The Opinionated Reader Web Plug-in - 'Add Comment' dialog mockup

The Opinionated Reader – Back-end

The Back-end maintains a database of news items extracted from RSS feeds of various news websites. Each news article is linked with the sentiment tags and reaction/positivity annotations extracted from the comments and annotations gathered by the commenting plug-in. This data is used by the mobile/tablet app to generate visualizations (Sentiment Graph and Positivity Graph). The back-end also performs data mining on the tags and annotations for geographies and tracks the opinions across time.

The back end system responds to queries received from the mobile app with the news story and associated tags and annotations, which are then rendered by the mobile app for the user.

