Text Stream Trend Analysis using Multiscale Visual Analytics with Applications to Social Media Systems

Chad A. Steed
Oak Ridge National Lab
Oak Ridge, TN
csteed@acm.org

Justin Beaver
Oak Ridge National Lab
Oak Ridge, TN
beaverjm@ornl.gov

Paul L. Bogen II
Oak Ridge National Lab
Oak Ridge, TN
plbogen@gmail.com

Margaret Drouhard
University of Tennessee
Knoxville, TN
meg.drouhard@utk.edu

Joshua Pyle
Oak Ridge National Lab
Oak Ridge, TN
jpyle1@utk.edu

ABSTRACT
In this paper, we introduce a new visual analytics system, called Matisse, that allows exploration of global trends in textual information streams with specific application to social media platforms. Despite the potential for real-time situational awareness using these services, interactive analysis of such semi-structured textual information is a challenge due to the high-throughput and high-velocity properties. Matisse addresses these challenges through the following contributions: (1) robust stream data management, (2) automated sentiment/emotion analytics, (3) inferential temporal, geospatial, and term-frequency visualizations, and (4) a flexible drill-down interaction scheme that progresses from macroscale to microscale views. In addition to describing these contributions, our work-in-progress paper concludes with a practical case study focused on the analysis of Twitter 1% sample stream information captured during the week of the Boston Marathon bombings.

Author Keywords
Visual analytics; sentiment analysis; emotion classification; machine learning; multiscale; information visualization; intelligent user interface; text mining; data analytics; human-computer interaction.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Social media systems, such as Twitter, allow humans to broadcast their thoughts and experiences on a continuous basis. Due in large part to the widespread utilization of mobile technology, these systems are rapidly transforming public discourse and setting trends and agendas in nearly every facet of society. The rising value of these systems for global situational awareness is apparent as we see journalists increasingly turning to social media and similar online streams to detect and investigate breaking news events and identify sources of expertise [8]. In these situations, the ability to detect changes in sentiment, emotion, opinion, and behavior in social media streams for the health and safety of the public is both viable and of paramount importance.

However, the analysis of social media streams is a challenge because of the large and rapidly changing content delivered in the form of semi-structured textual information. Effectively understanding this information requires the ability to grasp key trends and interactively drill-down to increasingly detailed views of the data. More specifically, an intelligent visual analytics system is needed that carefully orchestrates the strengths of both human and computational machinery [9]. Computational power is utilized to efficiently consume, store, and process the streaming content. On the other hand, human intuition, creativity, high-bandwidth visual processing, and background knowledge are engaged via interactive and exploratory visual interfaces.

With these objectives in mind, we have developed a new visual analytics system, called Matisse (see Figure 1), that is designed to reveal key trends and associations in complex social media streams. Backed by a robust stream data management module and automated analytical algorithms, Matisse allows data-driven, human-directed analysis of social media streams via an intelligent user interface. More specifically, Matisse encourages a multi-scale analytical workflow that begins with high-level overviews and progresses to increasingly detailed views, which may include accessing the raw data records. Furthermore, the interface provides multiple views of the information using a coordinated interaction model—an efficient methodology involving linked brushings distributed across multiple visualizations [7].

The remainder of this paper provides a description of the following components of Matisse: (1) text stream data management; (2) automated analytics for positive/negative sentiment estimation and more detailed emotional classification; and (3) coordinated information visualizations that support multiscale drill-down exploration via human interaction. Then, we present a practical case study using tweets captured from the

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

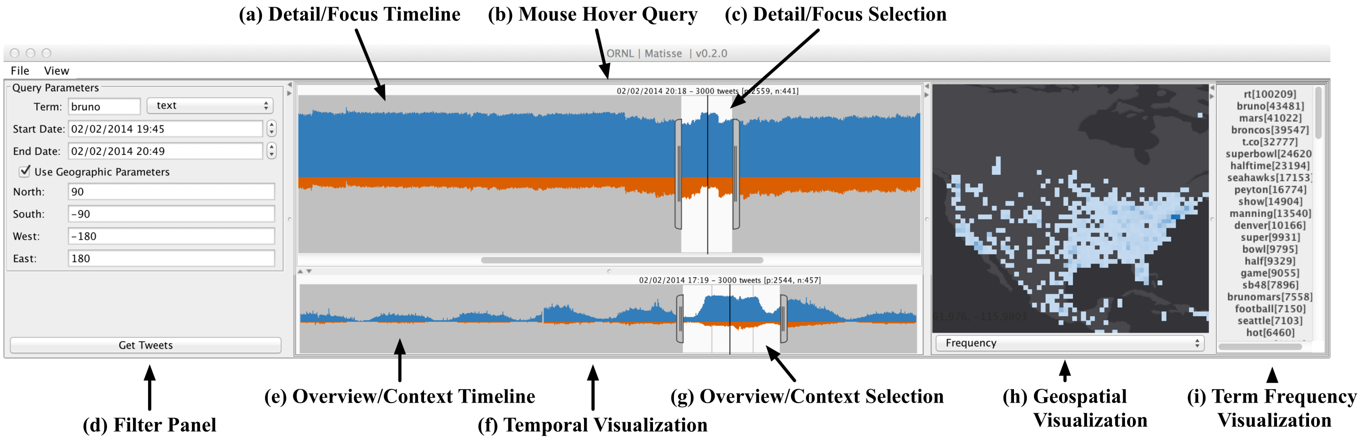


Figure 1. Matisse's main interface features a highly interactive visual canvas for exploring text stream information in temporal, geospatial, and term-frequency visualizations. The visualizations are linked using a coordinated model and a filter panel allows detailed record retrieval. In this Figure, tweets with terms related to Superbowl 48 are visualized using the bivariate timeline graphs (blue bars indicate positive tweet frequency and orange bars indicate negative tweet frequency). The detail/focus selection (c) reveals a noticeable disruption in positive/negative sentiment distribution during the Superbowl halftime show.

Twitter sample stream during the Boston Marathon bombings. This case study illustrates the system's effectiveness at revealing and flexibly exploring trends and associations at multiple levels of detail in a real-world scenario.

TEXT STREAM PROCESSING

Efficient management of streaming textual information is vital to any interactive analysis and visualization objectives. Matisse uses a customized stream management framework (see Figure 2) that is built upon a robust listener process and a mature open source search engine. The listener collects entries and periodically serializes the data into a compressed JavaScript Object Notation (JSON) formatted file archive. The serialization process is triggered at a customizable time interval (e.g. every second, minute, hour).

In a separate process, an indexer service continuously monitors the JSON archive. When new information is detected, the indexer extracts relevant information and adds it to a Lucene[1] index. After indexing, the data is available to the various analysis components through standard programming interfaces. Although the stream processing system is capable of consuming any textual information stream (e.g. RSS news feeds, weblog streams, other social media platforms), the current work focuses on trend analysis for the Twitter sample stream[2]. Twitter's sample stream API provides a continuous 1% random sample of all public tweets. The Twitter API also provides filtered stream queries over specific geographic areas and/or keywords of interest to increase the volume of relevant information.

The Matisse data management system can transform the streaming content into a variety of archival formats for subsequent analysis and visualization. In the current work, we focus on the utilization of Lucene indices, which provide efficient multi-faceted search capabilities over a variety of fields such as location, tweet text, and time. Our system also includes modules to analyze the data and store relationships based on implicit networks of users in a graph database such as Neo4J[3].

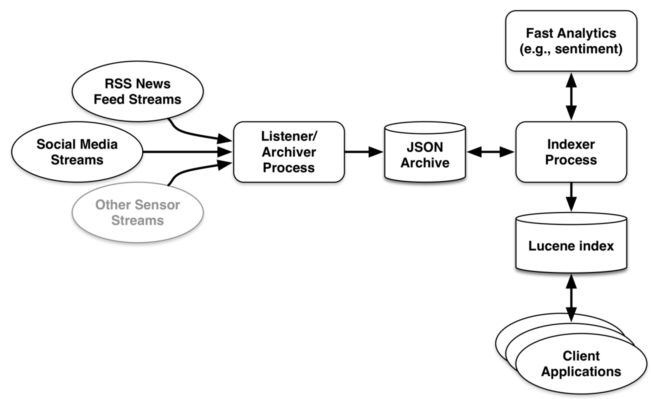


Figure 2. Matisse incorporates a robust data management system that consists of a listener process and an indexer process. Stream information is consumed and transformed into a Lucene index, which is available for client visualization and analytics applications. During indexing, analytical algorithms are applied to the data to classify the sentiment and emotion of the textual content.

To feed our multi-scale views, the system automatically computes supplemental summary information for the streams. These summaries are built based on temporal, geospatial, and textual context. For example, temporal summaries are created using statistical metrics calculated (e.g. frequency, averages, variance) for tweets collected in evenly distributed time interval objects (e.g. every minute, hour), or time bins. These time bin summaries are used in the linked visualizations to graphically represent trends in the information flowing through the system. The summary bins can also be calculated on-the-fly

[1] Lucene is an open source search engine available at `http://lucene.apache.org`.

[2] Twitter sample stream information is available at `http://dev.twitter.com`.

[3] Neo4J information is available at `http://www.neo4j.org`.

for use in our interactive multi-scale visualizations. Similar bin data structures are calculated and stored to describe the geospatial and term-frequency information in each time bin.

AUTOMATED ANALYTICS

The Matisse system leverages automated sentiment and emotion analytics via a supplemental pipeline that helps reduce the complexity of the raw information. In the following subsections, we describe two analytical capabilities: (1) Positive/negative sentiment estimation, and (2) fine-grained emotional classification.

Positive/Negative Sentiment Estimation

In the current system, we use an approach similar to Go et al. [4] for positive/negative sentiment estimation. The process for moving from raw text to a feature vector begins with converting all characters to lower case. Then, all tokens beginning with the '@' character are removed as these tokens represent to whom a tweet is directed and carry no sentimental information. The next step is to identify any instance of three or more of the same character in a row and reduce the segment to just two of the same character. For example, the following tweet, 'Iiii looooveeee yoooouuuu' becomes 'Ii loovee yoouu'. Character repetitions of three or more are replaced to prevent tokens like 'little' from becoming 'litle'. The process decreases to two characters to preserve the ability to distinguish the emphatic 'yoooouuuu' from the non-emphatic 'you'. Next, common contractions are identified (can't, won't, haven't, etc.) and replaced with their standard version (cannot, will not, have not). The next step removes various bracketing characters ('(', ')', '[', and ']') and compresses extra whitespace. The process then removes all URLs and replaces them with the token 'URL'. Finally, any remaining punctuation is removed.

The processed string is then segmented into tokens, stemmed using Porter's English stemmer, and filtered for stopwords using a custom stopword list that leaves in common sentimental words like 'want', 'not', 'should', 'could' that are often included in standard stopword lists. From these token lists a vector is generated consisting of the tokens in the filtered input. The process does not use counts as the short length of a tweet (at most 140 characters) makes re-occurrence of a token rare. Notable differences between our approach and Go et al. [4] include the discarding of '@' tokens, expansion of contractions, use of a full stemmer, the customized stopword list, and use of Boolean features as opposed to numeric features. Our process also does not utilize bigrams.

To train our classifier, we utilized the publicly available training set mentioned in Go et al. [4]. We discovered that while their work produced positive/neutral/negative sentiments, they do not provide training data for neutral tweets. For the current work, we opted to produce positive/negative sentiments only. The processed tweets from the Go et al. data set are used to train both a Python and Java naive Bayes classifier and a Java Maximum Entropy Classifier. For Python, we utilized *nltk* [1] along with *scikit-learn*[4]. For Java, we used *MALLET* [5] and *MinorThird* [3]. In our tests, the Python-based naive Bayes classifier achieved a 90% accuracy rate. Under Java, a naive Bayes classifier performed at a 79% accuracy rate and a maximum entropy classifier performed at an 82% accuracy rate. These results are comparable to the 83% classification accuracy reported in Go et al. [4] for unigram and bigram feature sets under a Maximum Entropy classifier.

[4]Information about *scikit-learn* is available at `http://scikit-learn.org`

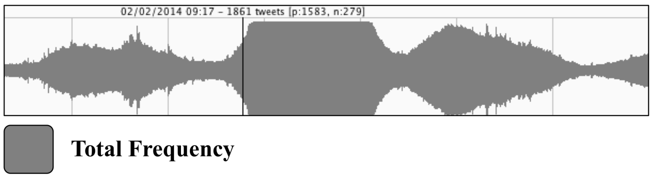
Emotional Classification using Machine Learning

To achieve fine-grained classifications beyond positive/negative sentiment, we also developed a machine-learning approach for estimating a text record's emotion in a manner that bypasses manual labeling of training data. This approach involves retrieving and then transforming the textual social media data into a collection of name-value pairs called a feature set. When labeled as specific emotions, these feature sets serve as examples that guide the learning algorithm in the automatic creation of a discriminative model.

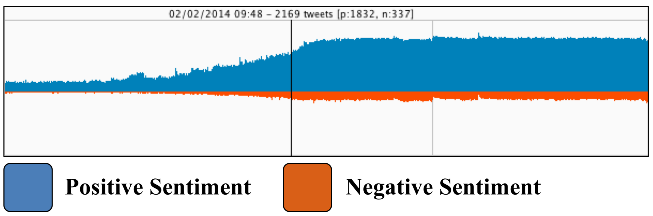
The challenges in training a discriminative model are selecting features that accurately represent the emotion associated with the underlying text, and retrieving a set of textual examples (and associated feature sets) that are unique to a specific emotion class. Once a model has been trained, it can be applied to new textual data to predict emotion. Our emotion classification model training process leverages features from both statistical and emotional text analysis, and it samples data based on the most pure examples of the various emotion classes.

The Affective Norms for English Words (ANEW) model [2] serves as the basis for our emotion quantification. ANEW is a dictionary of terms that have been quantified (interval scale) in terms of three dimensions: (1) Valence describes positivity/negativity, (2) Arousal indicates the excitability in the text, and (3) Dominance depicts the level of assertion by the author. For each text record, the Valence, Arousal, and Dominance (VAD) scores for all ANEW words in the text are averaged to obtain an overall ANEW emotion score for each dimension. For simplicity, the VAD scores are discretized over their score range. Valence is represented by the *unpleasant/neutral/pleasant* categories, Arousal is represented by the *subdued/neutral/active* categories, and Dominance is represented by the *dominated/in-control* categories. Thus, our approach to the VAD quantification of emotion for a document is one of 18 possible combinations of the VAD discretized categories. For example, a VAD categorization of *pleasant/neutral/dominated* indicates that the average raw VAD scores for the ANEW terms in the text mapped to the *pleasant* Valence category, the *neutral* Arousal category, and the *dominated* Dominance category.

The Term Frequency-Inverse Corpus Frequency (TF-ICF) term weighting method [6] is employed to determine the most significant terms in each textual record. TF-ICF calculates term frequency in a document by transforming the text into a vector space model [10], and approximating each term's weight based on weights determined from independent large text corpora. The most significant terms and ANEW emotion



(a) Total frequency (univariate display mode)



(b) Sentiment frequency (bivariate display model)

Figure 3. The overview+detail timeline visualization can display information in either a univariate mode (a) or a bivariate mode (b). In the univariate mode, bars are vertically centered. In the bivariate mode, the center line represents zero with top and bottom bars increasing in upward and downward directions, respectively.

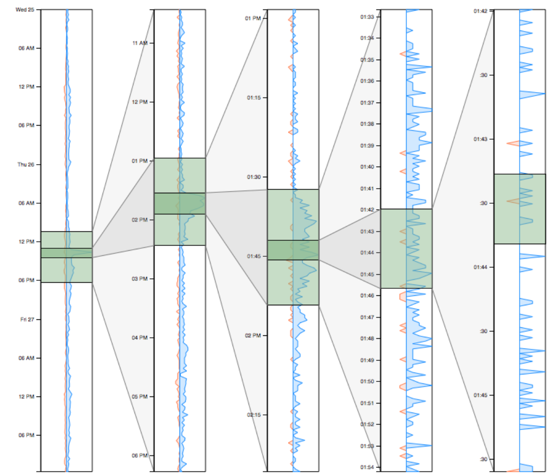


Figure 4. Matisse provides an alternative multi-scale temporal visualization that allows a gradual descent to increasingly detailed summaries of the time series information. Beginning the left overview, the user can drag time ranges of interest to display more detailed views of the data.

scores are merged as a feature set that is representative of the content and emotion associated with each textual record.

The process of transforming tweets into machine learning feature sets is the same for both training and prediction. First, the text is vetted for the presence of ANEW terms, and the VAD scores are averaged and discretized to obtain an overall categorical representation of emotion. Next, the text is converted into a vector space model and TF-ICF is used to identify the most significant terms. The VAD categories and most significant text terms are combined in a feature set to get an overall representation of the content and emotion associated with the text. We selected a maximum entropy learner from the *MinorThird* machine learning library [3] for multi-class emotion classification based on our prior experience in natural language processing. To train a machine learning model, we selectively retrieve tweets with a specific emotion class explicitly encoded as a hashtag (metadata encoded directly into the social media textual content that is intended to enable a topical search). This approach provides the purest representation of each sentiment class and enables automated labeling. Once all emotion classes achieve sufficient data representation from these retrievals, a classifier is built to predict the emotion in new unlabeled records.

INTERACTIVE MULTI-SCALE VISUALIZATION METHODS

The human-centered component of the Matisse system leverages an interactive canvas, shown in Figure 1, to allow visual exploration of the past and current state of activity in social media streams. The interactive canvas leverages the information produced by the previously mentioned analytical processing to render temporal, geospatial, term frequency, sentiment, and emotion visualizations. These separate data views are coordinated such that interactions in one display are systematically propagated to the other displays. Furthermore, the system supports drill-down investigations from high level overviews to detailed record listings using multi-scale representations. In the remainder of this section, we introduce key details about these visualization techniques.

Overview+Detail Temporal Visualizations

Matisse's data processing module aggregates the summary statistics for a user-defined unit of time (seconds, minutes, hours, etc.) to form a time series. As shown in Figure 3, this time series information is graphically encoded in a temporal visualization that represents the summary metric as a bar chart. The bar chart may encode the time series for a single variable or it can be split to show two variables. In Figure 3(a), the display shows the overall frequency of tweets per minute over a period of time where the height of the bar represents number of records with a time stamp that falls within the time bin. In Figure 3(b), the display shows the frequency of positive sentiment tweets as blue bars on the top and the frequency of negative sentiment tweets as orange bars on the bottom with a common baseline in the center representing zero. In the bivariate display variation, the bars can be added together to obtain the overall frequency for the time intervals. As the user moves the mouse cursor (see (b) in Figure 1), detailed summary information for the time bin is shown above the timeline as a mouse hover query.

In Figure 1, Matisse displays two timeline views in the main window: (1) an overview/context view and (2) a detail/focus view. The bottom view (see (e) in Figure 1) is the overview timeline, which provides an overall summary of the selected variable(s) for the entire time series. In this view, the time bin summaries are condensed and aggregated to fit the width of the window pane. Therefore, the time duration represented by each bar is determined by the width of the display and the bin dimensions are recalculated each time the window is resized. The user can select a time range in the overview (see (g) in Figure 1) by using the mouse to drag a time window box. When the drag operation is complete, the scrollable details timeline (see (a) in Figure 1) is regenerated with the most granular time bin information for the context selection time

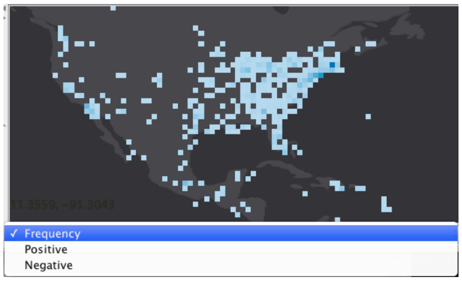


Figure 5. Matisse displays a geospatial heat map using a blue saturation color scale where darker more saturated colors indicate large values and lighter less saturated colors indicated small values. The heat map can be configured to show total frequency, positive sentiment frequency, or negative sentiment frequency.

range. In Figure 1, the detail timeline shows the minute level statistics. The user can also make time range selections in the details view (see (c) in Figure 1), which will propagate to the other coordinated visualizations.

Alternative Multi-scale Temporal Visualization

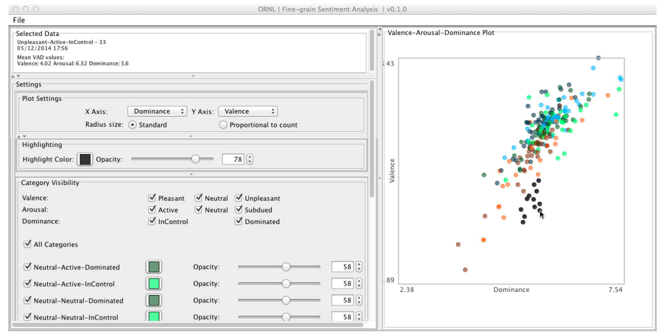
In addition to the overview+detail timeline, Matisse provides an alternative temporal visualization technique that allows gradual drill-down to increasingly detailed views via an extension to the standard brushing interaction. As with the overview+detail technique, tweet records are binned temporally and rendered as bars on a timeline. However, the multi-scale timelines are oriented vertically to maximize screen space.

The user starts by viewing one timeline that contains records over the entire temporal range. When the user brushes over a particular temporal range, a new timeline with finer granularity is displayed to the right of the brushed graph. As shown in Figure 4, the user can repeat this process to systematically increase the granularity of the display and explore trends in the data that are hidden in the more aggregated views.

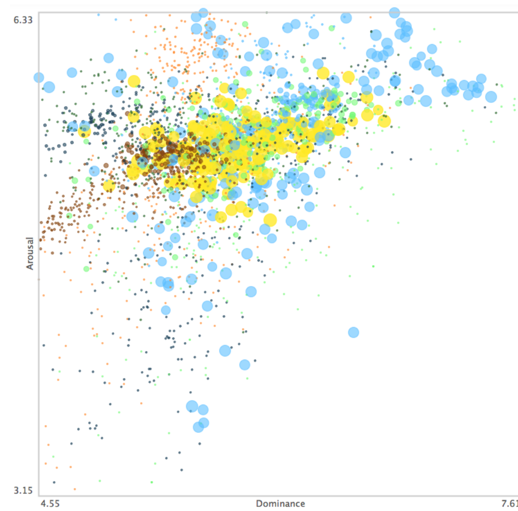
Figure 4 shows an example drill-down session. On the left timeline, a spike in positive sentiment is visible just after noon. Because the granularity is low, it is difficult to determine temporal qualities about the spike in sentiment. For instance, an analyst may want to know if the spike is due to tweets that are uniformly distributed in a time range, or if the spike is due to tweets that are sporadically spaced in the time range. By using the cascading array of timelines, the analyst discovers that the spike consists of tweets that are almost uniformly spaced around 1:45 p.m.

Geospatial Visualization

To the right of the temporal view (see (h) in Figure 1), Matisse displays a geospatial heat map for the selected time range. The map supports zooming in and out as well as panning operations. The color scale used in the map represents grid cells with higher tweet counts as darker and more saturated shades of blue and lower tweet counts as lighter and less saturated shades of blue. With this color scale, areas with more activity are displayed in a more visually salient manner to highlight geospatial trends.



(a) Full emotion visualization panel



(b) Emotion scatterplot with proportional point sizes

Figure 6. Matisse's emotion visualization panel lets the user interactively explore the results of the emotion classifier. In addition to zoom and pan interactions, the display can be configured to filter categories and encode frequency information in the point radii.

As shown in Figure 5, the map can be configured to visualize overall frequency, positive sentiment frequency, or negative sentiment frequency in the current version. Furthermore, additional derived statistical metrics can be created and visualized through extensions to the base map. Users can also select a geospatial region in the map view to set a spatial query for tweets that have latitude and longitude metadata falling within the query region. We observe that approximately 1.5% of tweets from the Twitter random sample stream include geographical location metadata fields. Consequently, a relatively small percentage of the tweet volume is reflected in the map display.

Term Frequency Visualization

At right of the geospatial view (see (i) Figure 1), the term frequency view shows the top ranked terms for the selected time range and spatial selection. To retain responsiveness, the top terms are calculated during the stream indexing process and stored within the time bins of the summary file. To quickly render the top terms for the selected time range, the top term summary information is read to populate the term view as the user interacts with the temporal visualization. When a user clicks on a term, the term is copied into the filter panel text field (see (d) in Figure 1) for subsequent retrieval of the most detailed tweet information.

Emotion Classification Visualization

As shown in Figure 6(a), Matisse offers a supplemental panel for visually investigating the emotional classification of selected text items. This visualization relies on the VAD emotion classifier described previously, and it uses the 18 VAD categories. Each point in the scatterplot represents a quantity of tweets assigned to one category for a particular time period. The default scatterplot view uses a standardized radius and has average Dominance values along the x-axis and average Arousal values along the y-axis. Every tweet in a tweet time bin from the main Matisse timeline view is assigned a category by the emotion classifier, and all of the tweets assigned to any given category will be aggregated and represented as a summary point in the scatterplot.

The emotion scatterplot supports zoom and pan interactions. When the user hovers the mouse cursor over points within the scatterplot, additional information about that point is displayed in the top left "Selected data" frame. Furthermore, the point and other points belonging to the same category are highlighted using a user-defined color.

The user may adjust both of the axes to represent the ranges of values for Valence, Arousal, and Dominance. As shown in Figure 6(b), the user may toggle the "Proportional to count" radio button to map the radius of each point to the total number of tweets assigned to the emotion category of a given tweet timeline bin. In addition, the user may adjust both color and opacity for each category as well as the highlight color. By default, point color is determined by color families mapped from Valence of the category. Each color family has two members: a lighter and less saturated color, representing an "InControl" Dominance category, and a darker, more saturated color that represents a category with the "Dominated" field. Finally, the user may select to interactively filter the categories displayed (*pleasant/neutral/unpleasant*, *active/neutral/subdued*, *dominated/in-control*) using the checkbox panel.

CASE STUDY: DRILL-DOWN TREND ANALYSIS OF THE TWITTER SAMPLE STREAM DURING THE BOSTON MARATHON BOMBING

In this section, we present a practical case study involving a real world scenario to illustrate the effectiveness of the Matisse visual analytics system. The scenario focuses on the discovery of significant trends from a collection of tweets that originated from the Twitter 1% sample stream during the week of the Boston Marathon bombings[5] (April 14–20, 2013). The objective of the case study is to illustrate how Matisse's interactive interface supports intuitive drill-down investigations that progress from clues in high-level overviews to dynamic queries of increasingly detailed evidence using a flexible, human-centered methodology.

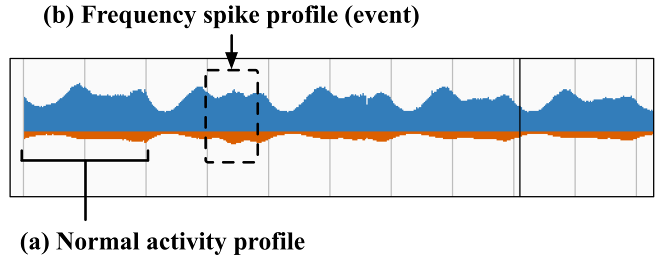


Figure 7. After initially loading the Twitter sample stream tweets into the overview timeline, we notice a frequency spike (b) on the second day by comparing it the profile of a typical time series (a). Typically, the afternoon bins are relatively flat as we see on the first day, but the spike marks the occurrence of a significant global event.

First, the processed Twitter stream index is opened and used to render the overview timeline visualization. As shown in Figure 7, the overview is configured to show the positive and negative tweet frequencies for the entire time series in a condensed bivariate timeline. In Figure 7(a), we note that the first day exhibits a normal timeline pattern for the Twitter 1% sample stream. That is, activity increases from a minimum in the early morning hours[6] to a peak at around noon. Then, activity drops slightly to a steady state in the afternoon and rises again in the evening until about midnight. Based on our experience, deviations from this typical pattern, such as the spike highlighted in Figure 7(b), indicate the occurrence of globally significant events. In this case, the deviation occurs during the afternoon on April 15, 2013 with a peak of 4,235 tweets (3,207 positive and 1,029 negative) at approximately 5:15 p.m. Furthermore, we note that the spike is visible in both the positive and negative tweet time series.

Having strong evidence of a significant event, we drill-down to gain more insight. In Figure 8, we select a time range in the overview timeline roughly centered on the day of the observed spike. Due to the linked view model, this selection causes the details timeline to render the minute-level time bin information for the selected time range. The resulting detail view confirms the increase in activity during the afternoon with much higher temporal fidelity. Using the mouse hover query, we observe a positive to negative tweet ratio of approximately 3:1. Visual inspection of the detail time series indicates that this ratio is generally consistent in the selected time range.

The geospatial view and the term frequency views (see Figure 8) reveal additional clues about the event prompting the activity spike. In the geospatial view, darker blue bins in the northeastern U.S. suggest the possible epicenter. In addition, top terms related to Boston refine our location estimation and terms such as "marathon" and "prayers" suggest possible ties to an event and feelings of compassion, respectively.

As shown in Figure 9, we drag an additional time range of interest in the detail view to restrict the information rendered in the geospatial and term frequency views to a smaller time

[5]Details regarding the Boston Marathon bombings are available at **`http://en.wikipedia.org/wiki/Boston_Marathon_bombings`**

[6]For the purposes of this case study time references are based on the Eastern Time Zone.

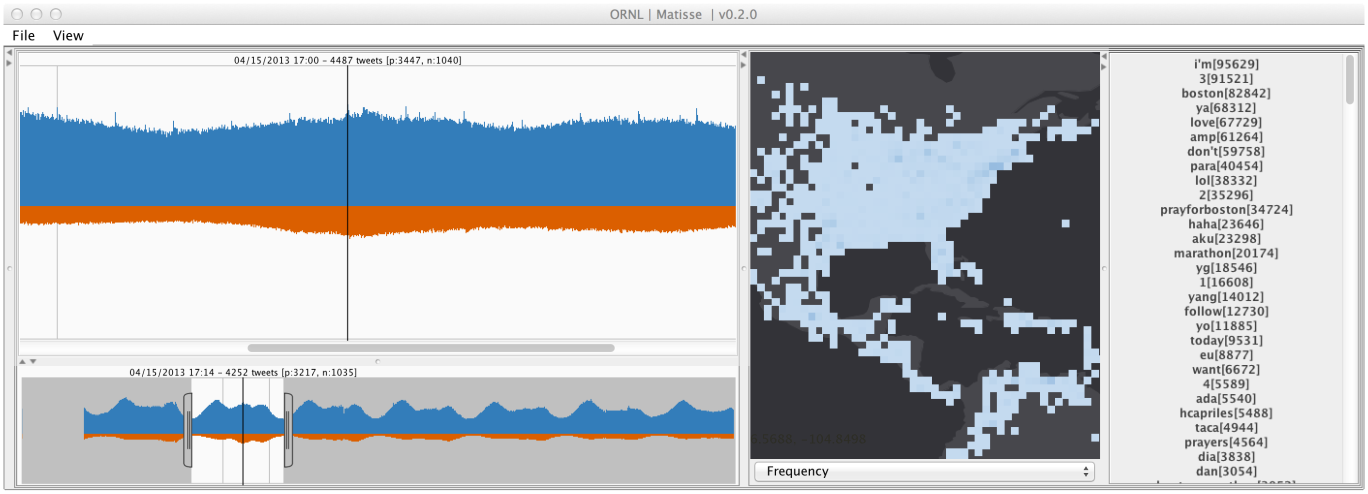


Figure 8. A frequency spike during the afternoon of April 15 prompts our selection of that day in the overview timeline. The detail timeline is rendered using the most detailed time bin summary information (minutes) providing more granularity. The spike in both positive and negative sentiment is still visible and the geospatial and term-frequency views indicate possible linkages to the Boston, MA area.

span (approximately 2 hours centered on 5 p.m.). The refreshed geospatial view reveals a high concentration of tweets around the Boston, MA region. In addition, the refreshed term frequency view provides additional clues. For instance, terms with "boston" and "marathon" suggest that something occurred at the Boston Marathon. In addition, terms such as "explosion" and "prayer" suggest a disastrous scenario.

At this point, we have progressed from an initial indication of atypical Twitter activity to the formulation of a rough hypothesis that people are discussing an event in the Boston area that may involve explosions and may be related to the annual Boston marathon. To test our hypothesis, we continue to drill-down to more detailed information by double-clicking on the term "explosion", which populates the query term text field in the left filter panel as shown in Figure 9. Then, we click the "Get Tweets" button to list all tweets containing this term during the selected time period. The resulting listing of tweets (not shown due to Twitter data usage restrictions) includes several descriptive responses to the Boston Marathon bombing and confirms our hypothesis about the details of the event that has caused the increase in Twitter activity.

Other terms in the view can be explored to reveal additional trends. The term "jfk" is listed because a fire at the JFK Presidential Library around the same time as the bombing sparked fears of a related explosion. Although subsequent investigations emphasized the library fire was due to mechanical issues and not an explosion, the rapid public's reaction illustrates the societal impact of social media systems such as Twitter.

In addition, we see that the term "maduro" appears in the term frequency view in response to the special election on April 14, 2013 to elect a new president of Venezuela after the death of Hugo Chàvez. Although unrelated to the bombings, the trend reveals the presence of multiple threads of globally significant communications that complicate event detection in social media streams. Despite multiple topics, it is straightforward with a visual exploration system like Matisse to formulate and confirm hypotheses from high-level overviews to detailed data investigations.

In this case study, we have illustrated the effectiveness of Matisse at delivering data-driven, human-guided analysis that starts with high-level overviews and systematically descends to detailed raw information through multiple levels of detail. Built upon a coordinated multiple view interaction model, the system fosters creative and flexible exploration of social media text streams in a manner that leverages both computational and human strengths. It is through the intelligent orchestration of computational power and human cognition that our approach alleviates the data complexities and exploratory nature of event detection in social media streams.

CONCLUSION

It would be impractical to manually analyze individual social media stream records to understand broad trends and associations. Likewise, overviews and aggregated views can be inconclusive without access to detailed records in their context. To be successful, both ends of the spectrum are needed as well as intermediate views. Matisse's visual analytics interface achieves this goal with coordinated visualizations that allow drill-down investigations of trending topics using multiscale visualizations. In addition to leveraging human interaction and cognition via interactive interfaces, Matisse capitalizes on computational power by augmenting its visualizations with information from automated sentiment/emotion analytics.

We are currently expanding Matisse to include additional visualizations and interaction schemes that support drill-down tasks. Encouraged from our emotion classification results, we plan to expand our approach and investigate novel techniques to incorporate human labeling in the model learning process. In addition, the system can be applied to other forms of streaming information and we are currently investigating its use in analyzing network traffic flow, online scientific simulations, and materials science experiments. This visual analyt-

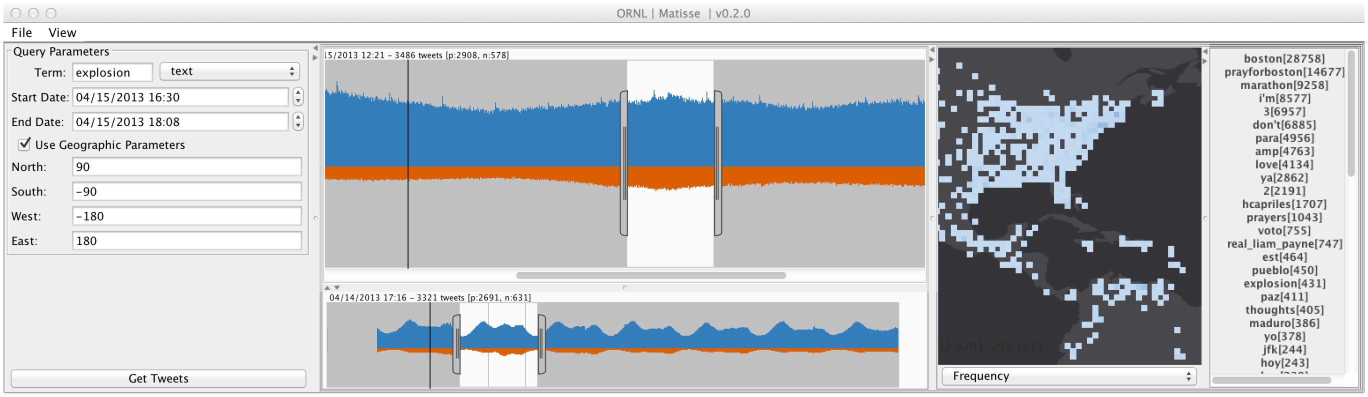


Figure 9. We restrict the geospatial and term-frequency views by selecting a smaller time range in the details timeline. The geospatial view and top terms provide strong evident that something disastrous has occurred at the Boston Marathon. To get more details, we execute a term query to list all tweets that include the term 'explosion'. The resulting list confirms our hypothesis and provides descriptive commentary on the event and public reaction, including discussion of a hoax involving related explosion at the JFK Presidential Library. We also note that the geospatial view and the term-frequency view indicate a parallel trend regarding the Presidential election in Venezuela.

ics system features multi-scale visualizations and coordinated views to allow significant trend discovery and exploratory investigations in challenging social media text streams.

ACKNOWLEDGMENTS

This work is sponsored by Oak Ridge National Laboratory LDRD project no. 6427. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

1. Bird, S., Loper, E., and Klein, E. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009.
2. Bradley, M. M., and Lang, P. J. Affective norms for english words (ANEW): Instruction manual and affective ratings. Tech. Rep. Technical Report C-1, University of Florida, 1999.
3. Cohen, W. W. Minorthird: Methods for identifying names and ontological relations in text using heuristics for inducing regularities from data, 2004. http://minorthird.sourceforge.net.
4. Go, A., Bhayani, R., and Huang, L. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* (2009), 1–12.
5. McCallum, A. K. MALLET: A machine learning for language toolkit, 2002. http://mallet.cs.umass.edu.
6. Reed, J. W., Yu, J., Potok, T. E., Klump, B. A., Elmore, M. T., and Hurson, A. R. TF-ICF: A new term weighting scheme for clustering dynamic data streams. In *Proceedings of the 5th International Conference on Machine Learning and Applications* (2006), 258–263.
7. Roberts, J. C. Exploratory visualization with multiple linked views. In *Exploring Geovisualization*. Elseviers, 2004, 159–180.
8. Steed, C. A., Potok, T. E., Patton, R. M., Goodall, J. R., Maness, C., and Senter, J. Interactive visual analysis of high throughput text streams. In *Interactive Visual Text Analytics Workshop* (Seattle, WA, Oct. 2012), 1–4.
9. Thomas, J. J., and Cook, K. A., Eds. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press, Los Alamitos, CA, 2005.
10. Wong, S. K. M., Ziarko, W., and Raghavan, V. V. On modeling of information retrieval concepts in vector spaces. *ACM Transactions on Database Systems 12*, 2 (1987), 299–321.