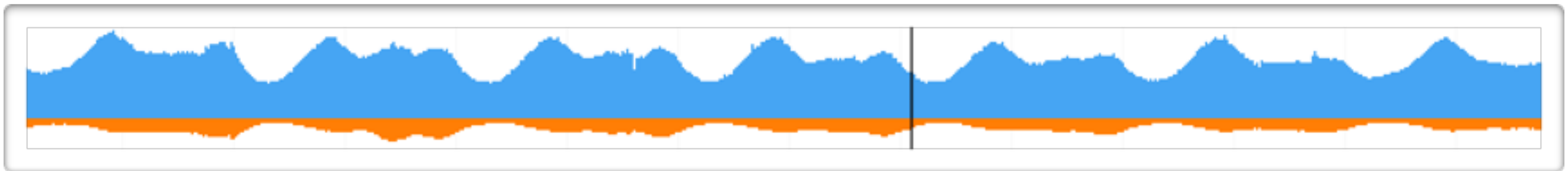
Matisse: A Visual Analytics System for Exploring Emotion Trends in Social Media Text Streams



Chad A. Steed (csteed@acm.org), Margaret Drouhard,
Justin Beaver, Joshua Pyle , and Paul L. Bogen II

IEEE International Conference on Big Data
November 1st, 2015

This research is sponsored by Oak Ridge National Laboratory (ORNL)
Laboratory Directed Research and Development (LDRD) no. 6427.

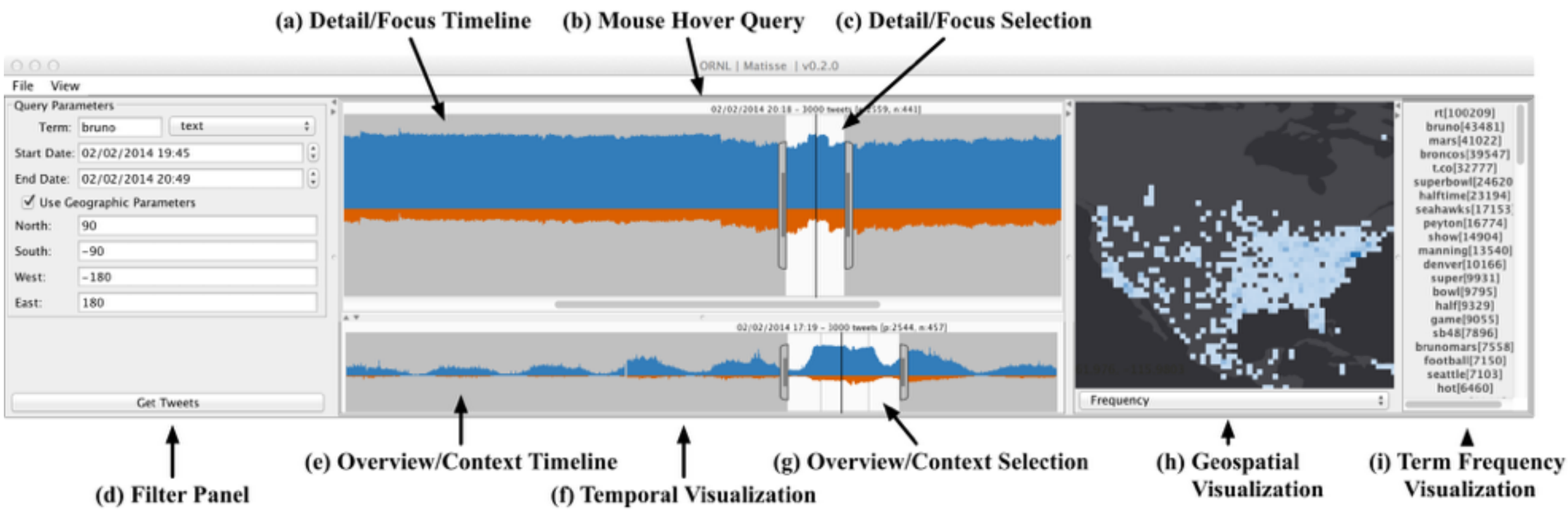
OAK RIDGE National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

Motivation

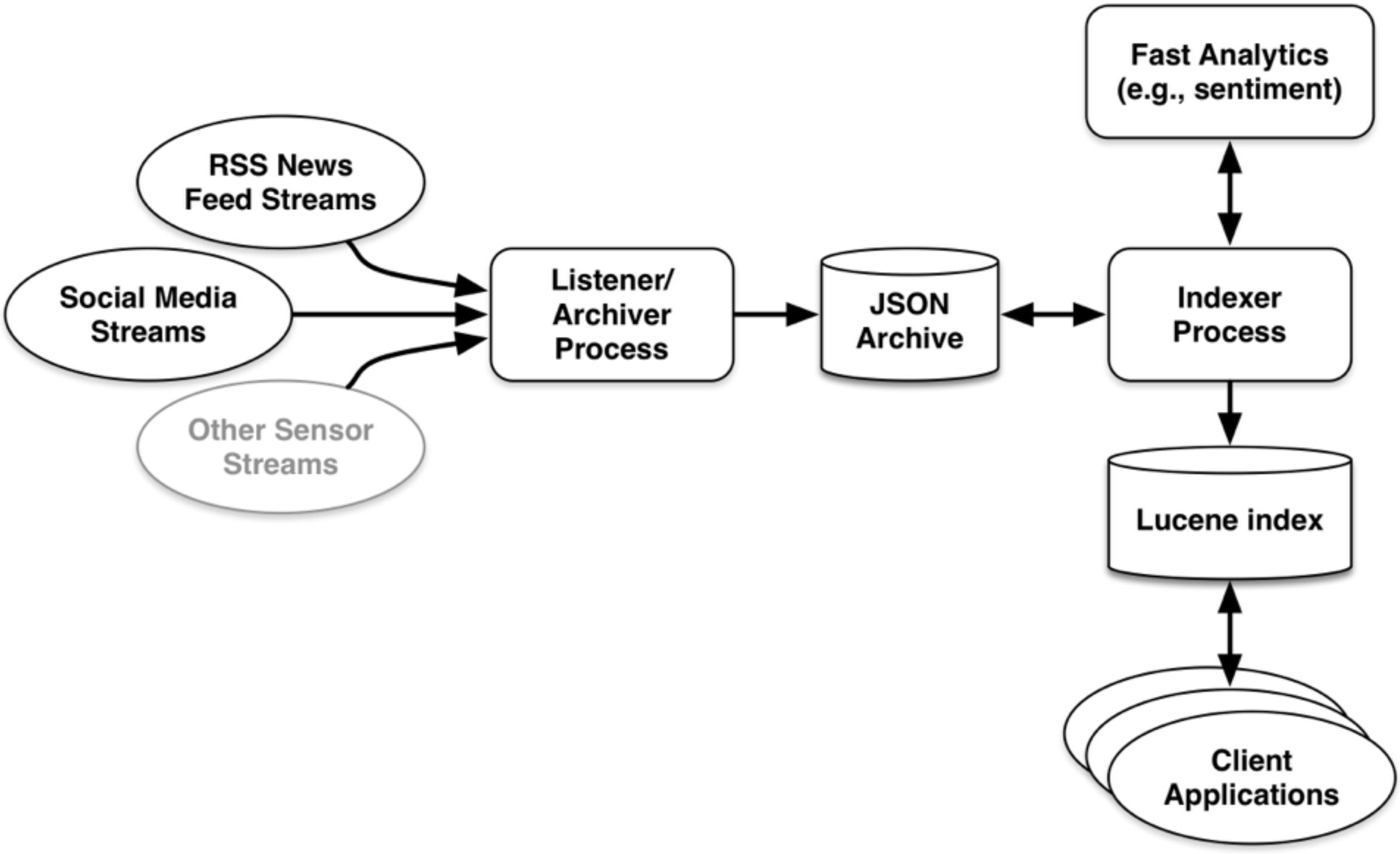
- Social media text streams represent a tremendous resource for understanding global events in real-time.
- But these streams are a challenge to analyze due to large and rapidly changing semi-structured textual information.
  - Too large for a pure visualization approach
  - Too exploratory for a pure analytics approach
- Automated analytics are needed to guide users to key trends and interactive techniques to drill-down to more detailed views.
  - Ideal scenario for a visual analytics approach

Matisse: Exploratory Analysis of Text Streams



Design Goal: Facilitate multi-faceted, exploratory analysis of social media text streams from overviews to detailed investigation

Stream Data Management



Positive / Negative Sentiment Analytics

- Estimate positive / negative sentiment based on textual content
- Initial processing step modifies raw text to accommodate some nuances of Twitter content
- Porter's English stemmer applied and custom stop words are removed
- Train classifier (naive Bayes and Java Maximum Entropy) using pre-coded tweets from Go et al. [4] to create a feature vector.
- Comparable classification accuracy to Go et al. [4] results (80% - 90%)

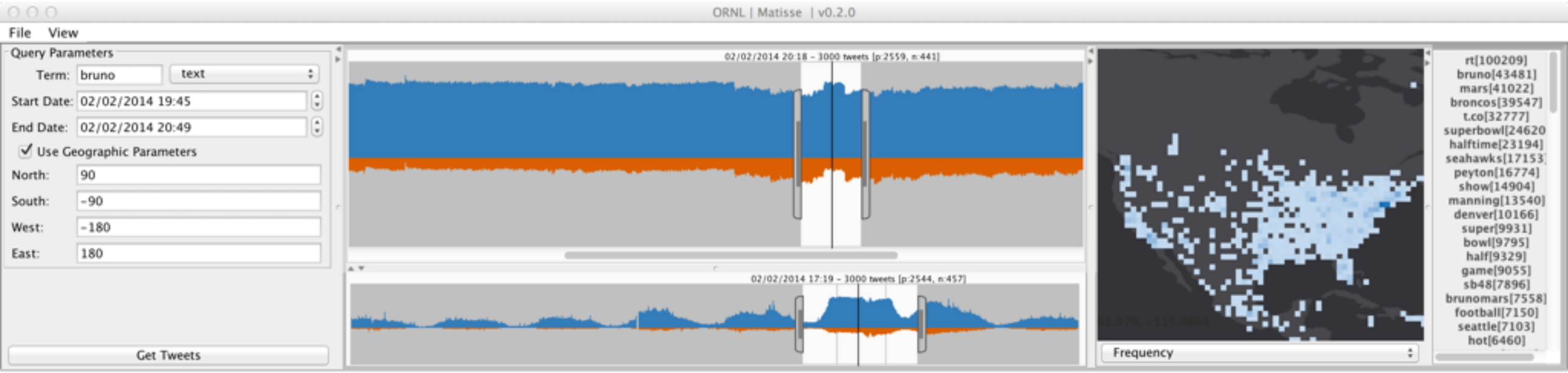
Emotion Classification using Machine Learning

- Estimate textual content's emotion using a ML approach that avoids manual labeling of training data
- Leverages statistical and emotional text analytics trained on pure examples of various emotion classes
- Leverages ANEW [2] model
  - Valence - positivity/negativity
  - Arousal - excitability
  - Dominance - assertion level of the author

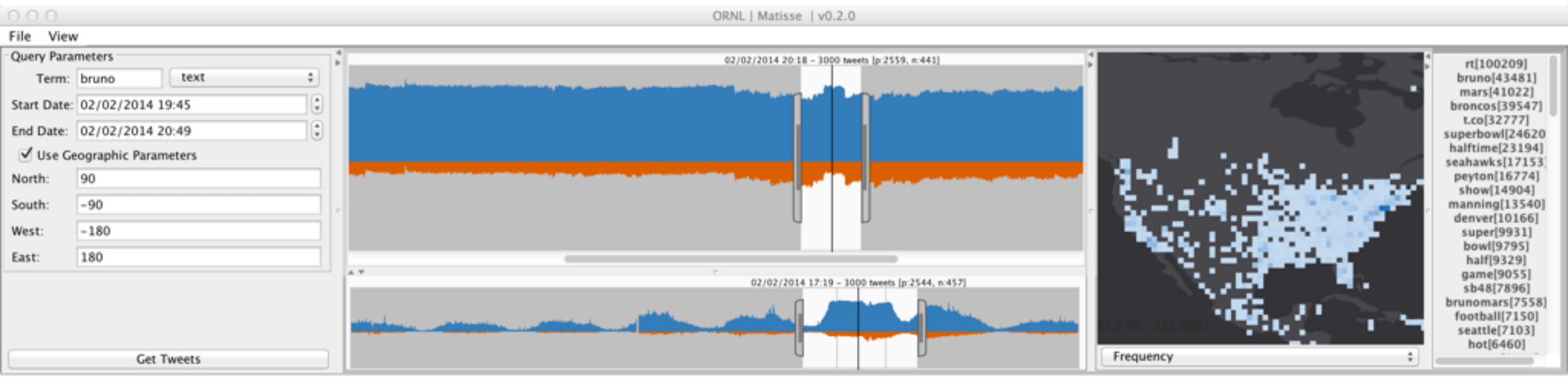
Emotion Classification using Machine Learning

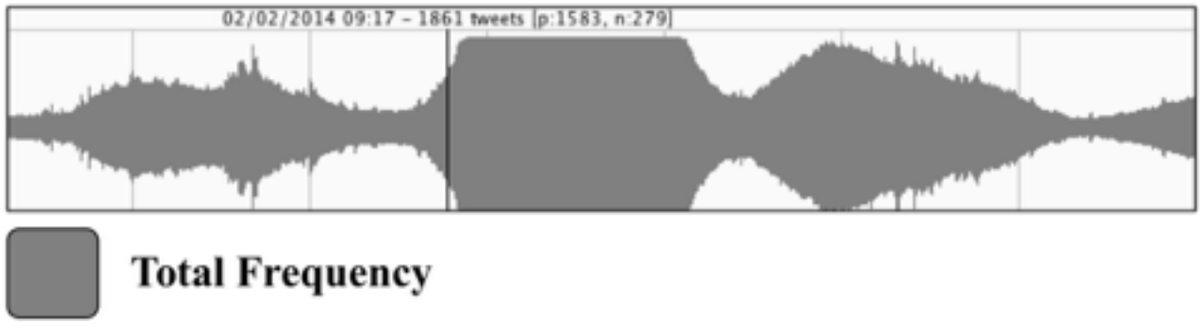
- TF-ICF term weighting method predicts significant terms using a vector space model
- Significant terms and ANEW scores are merged as a feature set representative of the content and emotion of the textual record
- A maximum entropy learner from the MinorThird ML library is used to train the model
- Training process selects tweets with emotion class explicitly encoded as a hashtag
    - Provides pure representations of each emotion class for automated labeling.
- Classifier is built to predict emotion in new unlabeled records

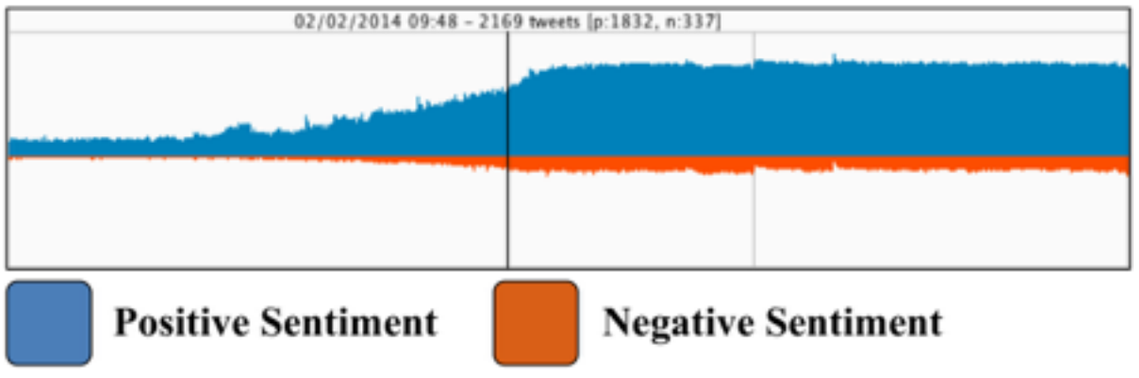
Geospatial + Term Frequency Visualizations



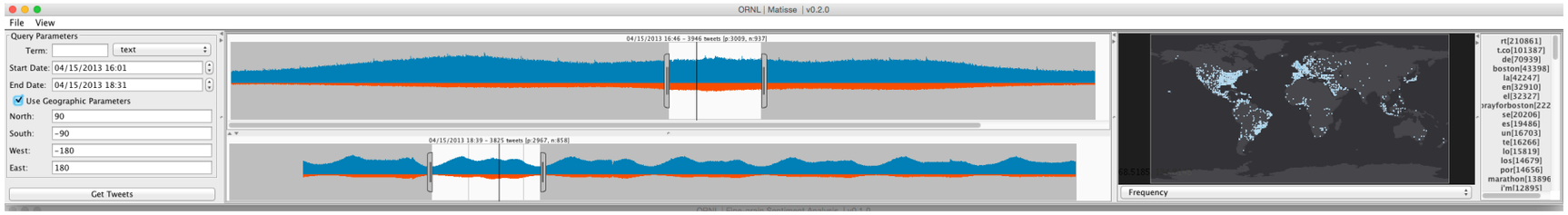
Multi-scale Temporal Visualizations

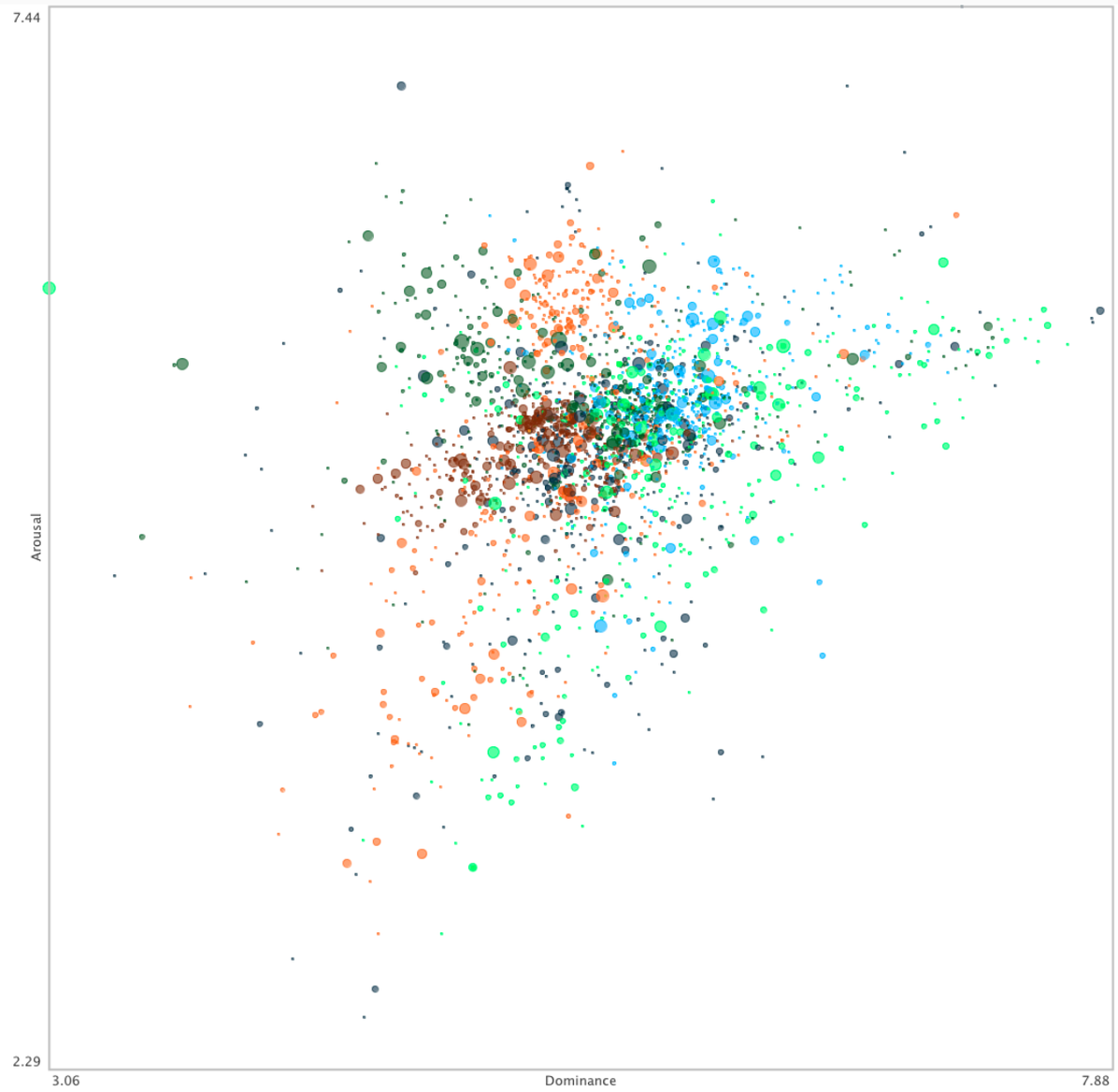




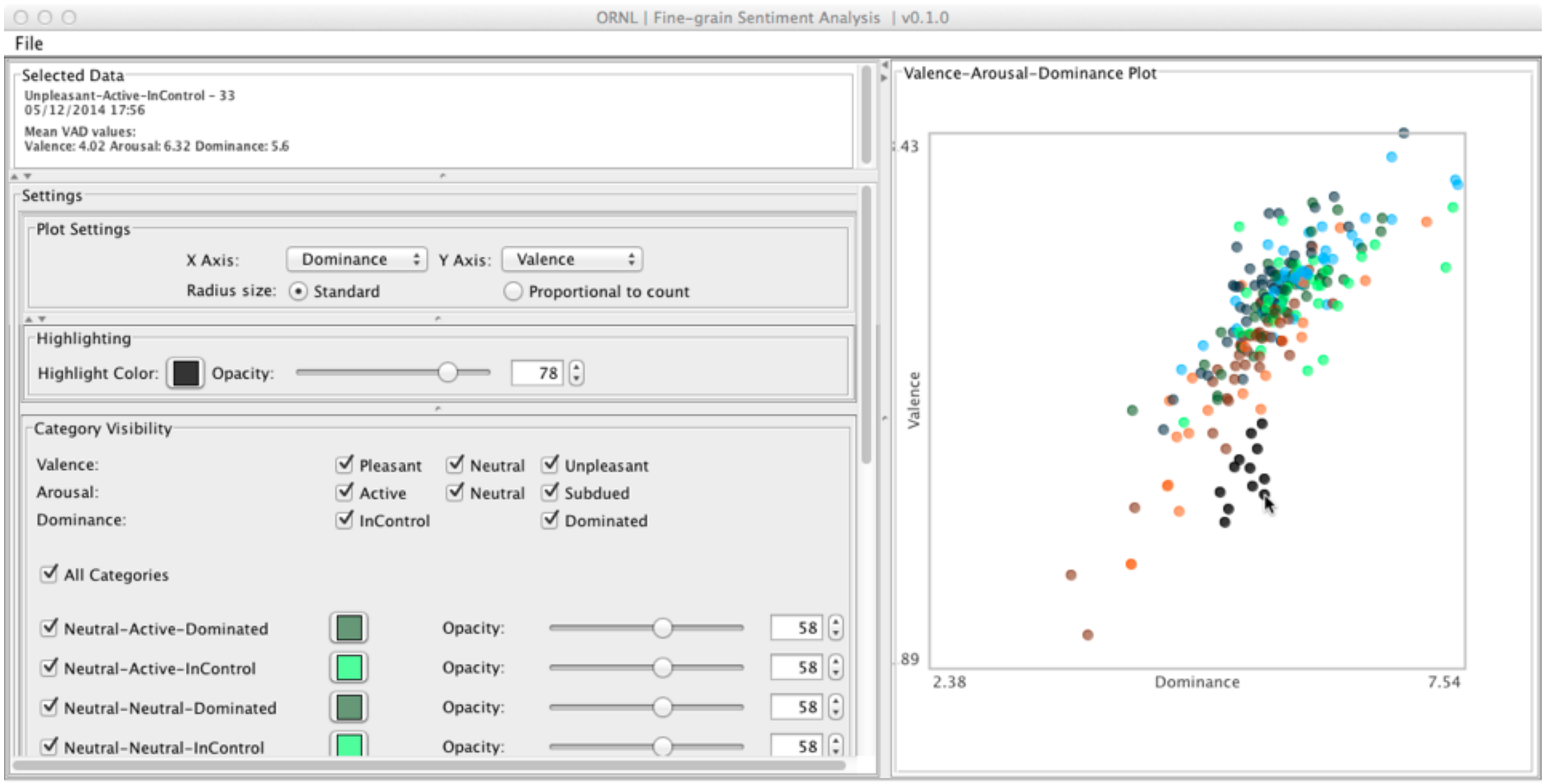


Visualizing Emotion Classifications

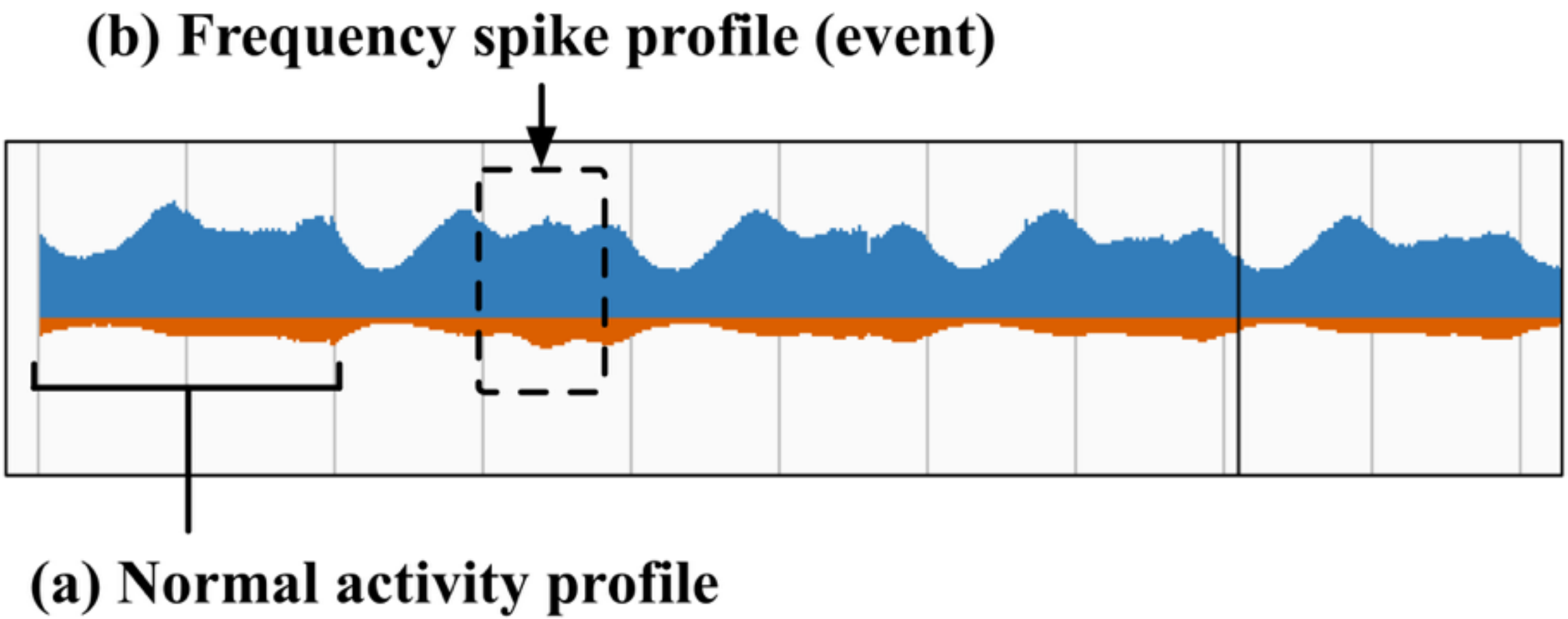




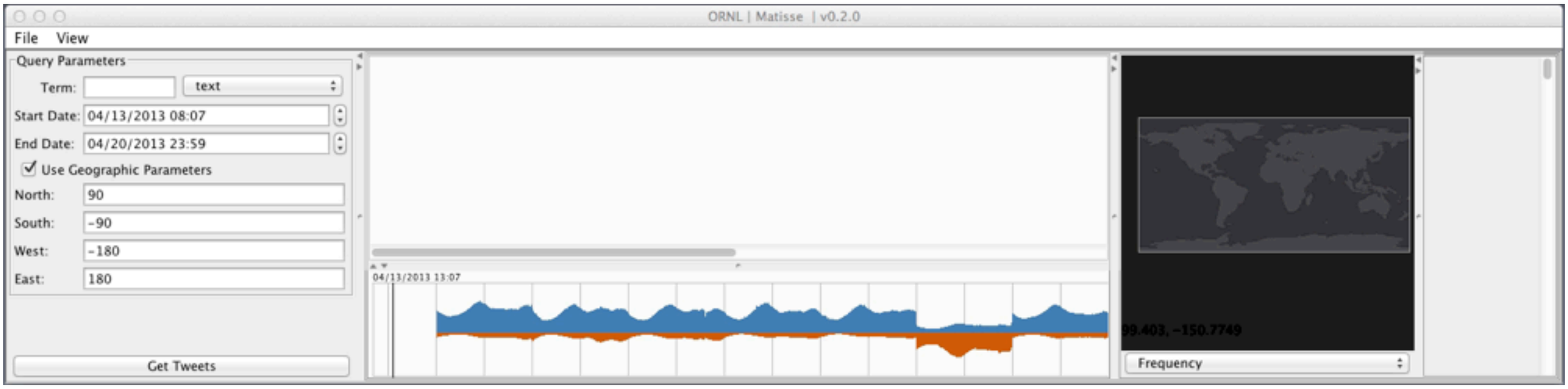
Visualizing Emotion Classifications



Temporal Event Detection



Case Study: Boston Marathon Bombing



Twitter 1% Sample Stream
Week of Boston Marathon Bombing
14-20 April, 2013

Conclusions

- Multi-scale visualizations enable scalable exploratory analysis
  - More intermediate views and scaled analytics will improve
- Automated analytics (sentiment / emotion) help guide analysis
  - ML emotion classification needs validation
- Linked views and interactions foster more creative analysis.
  - Additional views and interactions can help (e.g., graphs)