Statler

Summarizing Media through Short-Messaging Services

David A. Shamma
Yahoo! Research
Santa Clara, CA USA
aymans@acm.org

Lyndon Kennedy
Yahoo! Research
Santa Clara, CA USA
lyndonk@yahoo-inc.com

Elizabeth F. Churchill
Yahoo! Research
Santa Clara, CA USA
churchill@acm.org

ABSTRACT

Statler is a tool to examine short messages from Twitter that relate to a broadcast media event. Using twitter messages sent at the time of widely televised events, Statler shows segmentation, trending topics, level-of-interest, and Tweet geo-locations. Statler examines the first U.S. presidential debate of 2008 and the Inauguration of Barack Obama by applying current methodologies for examining implicit media annotation as collected through Twitter. Future versions will explore other genres as well as be optimized for live event tracking.

Author Keywords

Experience mining, TV, Twitter, inauguration, conversation, commentary, broadcast, event, participation, community, follower, social media

ACM Classification Keywords

H.5.1 Information interfaces and presentation: Multimedia Information Systems—*Video*; H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces—*Synchronous interaction; Collaborative computing*

General Terms

Human Factors

INTRODUCTION

While watching TV, people chat with their friends on the sofa, over the phone, or using the Internet. Recently, with the rise of short-message services, websites like Twitter have become a destination for people wanting to post or read comments about first run TV shows and events. During the 2008 U.S. Presidential Debates, Current TV ran a program called *Hack the Debate*[1]. They called for people to live-post their comments about the debate to Twitter. Through a semi-automated editoral process, Current TV filtered the tweets for content and displayed them on live TV, overlaid on the debate. This program continued through the U.S. Election and Inauguration. From Current TV's work, many U.S. TV news shows now call for people to submit comments via short message services (such as Twitter or Facebook). Other TV shows, like Bravo's Top Chef, have celebrities tweet comments during show's first air times.

CSCW 2010, February 610, 2010, Savannah, Georgia, USA.

Figure 1. A screen shot of Statler watching the first U.S. Presidential Debate. The video is played (left) while an interactive Table of Contents (ToC) is displayed (right). The ToC contains automatically determined segments and shows terms from the tweets as well as from the debate's closed captioning. Seeking on the video marks the current segment in the ToC.

In this article, we introduce Statler[1], a tool of identifying video content and insitu commentary from community annotation. While many users tend to inspect, *or follow*, a stream of tweets that relate to a TV show, Statler examines the semantics and structure of the Twitter messages, in addition to the messages' content. Statler automatically finds chapters segments, important moments, and topics of momentary and sustained conversation.

RELATED WORK

Unlike traditional work that examines news content for segmentation [2], Statler explores the relationship between the news media and community annotation. Recently Shamma et al. [4] demonstrated several methods that use Twitter messages about a media event to segment and topically identify trends. Statler presents these findings in a live application. While a few of these systems have been built, most notably the MTV Video Music Awards Twitter visualization [3], Statler's goal is not to show overall volume, but rather identify interesting moments within the tweet stream itself.

STATLER

Statler currently explores the first U.S. presidential debate of 2008 as well as the inauguration of Barack Obama. The

[1]The full demo can be found at `http://bit.ly/statler`

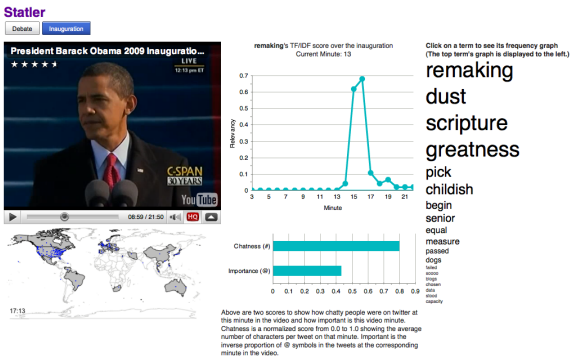

Figure 2. A screen shot of Statler watching the inauguration of Barack Obama. The video is played (left) while a meter of importance and chattness is displayed (bottom middle) along with a ranked list of tweeted terms (far right). The trend for the top ranking term of the current second is displayed next to the term list (center top). Below the video is a world map highlighting the countries tweeting at that minute.

debate tweets, $\approx$ 3000, came from Twitter's search interface and the inauguration tweets, $\approx$ 53000, were pulled from Twitter's constant rate data mining feed. Each sample represented 90 minutes of event footage in total. The two feeds varied in structure, we explore different features visible to each particular data set.

Presidential Debate

The debate interface consists of two main components, a table of contents (ToC) and topic trends, see Figure 1. The table of contents is generated automatically by examining tweet volume [4] and is wired to the video player. Since people often want at ToC but use the video player to find segments [5], Statler connects the two—scrubbing on the video moves the ToC's highlighted region.

For each segment, the ToC shows key terms, from the tweets and from the debate's closed captioning. This gives insight into the content of the segment. Terms are scored by comparing the frequency of a given term in tweets sent during the segment against its frequency over all tweets. The highest-scoring terms are then included in a short set of terms shown for each segment.

The Obama Inauguration

For the inauguration, Statler displays two aggregate metrics, trending terms and a world map of tweet locations, see Figure 2. The aggregate metrics are Importance and Chatness; each score is normalized between 0.0 and 1.0 and displayed as a bar chart. The Importance metric is a function of how many `@`mentions were used on any given minute. More `@` symbols means more conversation and, hence, people are paying less attention to the event. Similarly, how many characters typed per tweet per minute is used as a metric of how much time people spent typing their tweet that minute. More chatty means more to discuss or announce. At Obama's swearing in, we find few `@` symbols in the stream where as later, people were more chatty, see Figure 3. Above the bar graphs, Statler displays a world map that highlights the origin countries of the tweets for that minute.

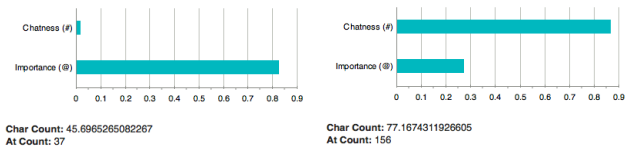

Figure 3. Left: Minute 12:05 of the inauguration was the swearing in of Barack Obama; it features a significant drop in `@`mentions ($> 2\sigma$). Right: Towards the end of Obama's inaugural address, people were typing more and chatting amongst each other more.

The system also displays the salient terms being used on Twitter for any given minute. At a given point in time, it calculates the frequency of each term during the previous two minutes and weighs that by the typical term frequency. These terms, which are being used more frequently in this recent window than usual are then shown to the user to capture the topics of discussion emerging on Twitter around each point in time. A small graph showing frequency over time can be displayed for any current trending term.

FUTURE WORK

Statler demonstrates how to understand content semantics derived from implicit community annotation. We have begun to explore metrics, based on current methodologies [4], and application designs which can reveal these insights. Future editions of Statler will work *live*, which will allow for easier exploration of new genres outside of the news domain. Different users may wish to see different trends; we wish to tailor Statler for various applications from journalists to casual TV watchers. Additionally, we wish to add some filtration such as following a set of terms or geo-restricting the display to a set of countries.

REFERENCES

1. Current TV: Hack the Debate. `http://current.com/topics/88834922_hack-the-debate/`.

2. C. Y. Low, Q. Tian, and H. Zhang. An automatic news video parsing, indexing and browsing system. In *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia*, pages 425–426, New York, NY, USA, 1996. ACM Press.

3. E. Rodenbeck. Kanye west is an a**hole and other twitter moments from the mtv video music awards. `http://content.stamen.com/kanye_west_is_an_a__hole_and_other_twitter_moments`, September 2009. Accessed October 12th, 2009.

4. D. A. Shamma, L. Kennedy, and E. F. Churchill. Tweet the debates: Understanding community annotation of uncollected sources. In *WSM '09: Proceedings of the international workshop on Workshop on Social Media*, Beijing, China, 2009. ACM. *In press*.

5. E. G. Toms, C. Dufour, J. Lewis, and R. Baecker. Assessing tools for use with webcasts. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital loibraries*, pages 79–88, New York, NY, USA, 2005. ACM Press.