

Storia: Summarizing Social Media Content based on Narrative Theory using Crowdsourcing

Joy Kim
Stanford University
jojo0808@stanford.edu

Andrés Monroy-Hernández
Microsoft Research
amh@microsoft.com

ABSTRACT

People from all over the world use social media to share thoughts and opinions about events, and understanding what people say through these channels has been of increasing interest to researchers, journalists, and marketers alike. However, while automatically generated summaries enable people to consume large amounts of data efficiently, they do not provide the context needed for a viewer to fully understand an event. Narrative structure can provide templates for the order and manner in which this data is presented to create stories that are oriented around narrative elements rather than summaries made up of facts. In this paper, we use narrative theory as a framework for identifying the links between social media content. To do this, we designed crowdsourcing tasks to generate summaries of events based on commonly used narrative templates. In a controlled study, for certain types of events, people were more emotionally engaged with stories created with narrative structure and were also more likely to recommend them to others compared to summaries created without narrative structure.

Author Keywords

Social computing; crowdsourcing; creative collaboration; storytelling.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Collaborative computing

INTRODUCTION

Social media today allows millions of people from all over the world to share and discuss their thoughts about commonly experienced events. There has been increasing interest among researchers, journalists, and marketers alike in using social media to understand what people say about these events; a large body of research explores summarizing emotions and reactions as seen on Twitter [23, 17, 5], and news articles and blog posts often integrate social media content and visualizations into their text.

However, while existing automated approaches excel at identifying moments of public attention, it is often up to the viewer to create their own interpretation of the data (or seek out interpretations provided by journalists and bloggers). This can be difficult because tweets, Facebook posts, and other social media messages are generally created in the moment [20], and so sometimes lack the context needed to make sense to future viewers. In addition, some viewers may not be familiar enough with the event to understand the jargon, idioms, or other specialized language used by social media authors.

On the other hand, manually authored stories punctuated with a curated set of social media posts, such as those created through Storify [2], can provide unifying commentary that bridge gaps in information: for example, a news article about Brazil's dramatic loss to Germany in the FIFA 2014 World Cup semi-finals describes not just the final score but also compares it against past matches to highlight the intensity of the loss; it also points out that Brazil's top scorer and top defender were unable to play during the match due to circumstances external to the match itself. However, the process for creating these stories is limited to those who have the time, skills, and resources to learn about an event's context by observing the event, doing research, and conducting interviews.

To achieve both the scalability of automated approaches and the coherency of manually authored stories, we propose *summarizing social media content based on narrative structure*: rather than depending on a trained storytelling expert, interpreting social media in terms of narrative elements — such as beginnings, middles, ends, characters, goals, and climaxes — may reveal what information needs in order to make sense to viewers. This approach makes use of crowdsourcing to interpret data at scale and automatically generate summaries with these narrative elements in mind.

We hypothesize that designing crowdsourcing tasks around narrative elements can help non-expert crowd workers collaborate in addressing this challenge. In this paper, we first *identify and fill narrative gaps* in a social media record. For example, if a name appears in a social media feed, we may want to know more about who they are and their significance with respect to the overall story (i.e. a “character” introduction). Then, we *link content to narrative categories*—for example, we may be able to recognize certain tweets as descriptions of conflict between the story's characters. Specifically, we refer to *narrative categories* [6] as a simple template for structuring social media content with respect to storytelling roles.

To explore this approach, we created a prototype crowdsourcing system called *Storia* (Figure 1). *Storia* takes, as input, data from an automated system that detects moments that occur during a public event (such as a sports game) and uses crowdsourcing to output a written story about the event. *Storia* consists of two crowdsourcing modules that ask crowd workers to 1) gather missing narrative context and 2) write paragraphs for each important moment in the event based on social media content, using narrative categories as a template. For four social media events, we compared stories generated by *Storia* with stories that were crowdsourced without using narrative structure through a controlled study. We asked 30 participants to evaluate each story with respect to how well it conveys an event to someone who had missed it, and found that, for certain events, *Storia* stories were recommended three times as often by participants due to its emotional content.

To summarize our contributions, in this paper we:

- contribute a technique for recovering missing information from social media feeds by identifying and filling narrative gaps,
- demonstrate the application of narrative theory in designing crowdsourcing workflows for generating stories, and
- explore the limitations of narrative summarization by studying its output given different types of social media events.

Our results set the stage for constructing concise emotional experiences out of multiple viewpoints and deriving lessons for applying narrative theories to approaches for crowdsourcing creative work.

RELATED WORK

Storia focuses on using an *underutilized source of content* (that is, social media) to craft narratives, rather than surfacing trends and themes for analysis or (re)constructing a logical description of events. It turns to past work in narrative theory, social media summarization and curation, and crowdsourcing creativity to inform its design.

Narrative Theory

Narrative theory [9] stems from the idea that people use narrative as a basic cognitive strategy for making sense of various aspects of the human experience (such as time and change). In fact, the presence of narrative can significantly alter how an experience is perceived [21]. We may be able to frame social media content as the product of people attempting to understand experiences with others; when a user creates and posts content, they act as a narrator conveying some experience to an audience.

Storia attempts to incorporate specific theories about how stories are understood in the design of its crowdsourcing tasks. In his theory of narrative categories, Cohn approaches storytelling in comics with respect to cognition, examining visual elements in terms of narrative syntax [6]. Cohn argues that individual comic panels can be mapped to four basic narrative functions or roles that control narrative dramatization and pacing:

- the *peak* depicts the culmination of an action set in motion during the narration; it can stand alone as a (blunt) summary of the narrative.
- the *establisher* sets up the relationships of all characters involved in the story,
- the *initial* starts the action that eventually culminates in the peak,
- and the *release* depicts the aftermath or reaction to the peak, providing a sense of closure or creating room for anticipation for the next part of the story.

Furthermore, these roles can act hierarchically: a group of panels can together fulfill a narrative role for the larger story. In other words, these roles form “sentences” that make up the narrative arcs of a comic. *Storia* uses this theory as the basis for a structured form used by crowd workers to write text (rather than visual) summaries based on social media data.

Social Media Summarization and Curation

Projects such as Narrative Science [1] point to the value of transforming large amounts of quantitative data into natural language summaries to facilitate an accessible understanding of an event. A large body of work focuses specifically on text summarization of events in social media, particularly on Twitter [17, 5]. This work focuses mostly on using text analysis, sentiment analysis, and machine learning [15, 7, 10] to surface important moments out of social media chatter and on generating understandable summary text automatically. Other work aggregates Twitter content into visual summaries [18, 22]. However, while visual summaries may *suggest* a narrative, they do not offer a narrative on their own. Both types of summaries often result in a simple list of highly-tweeted moments (such as the goals in a soccer game) that is detached from the emotional ups and downs of the overall event. We complement this past work; rather than try to detect important events, we start with a data set that has already grouped social media content into important moments using existing techniques and attempt to form a *narrative* that provides the context behind various pieces of information and conveys a sense of dramatic structure.

In social media curation, people *manually* organize social media content in order to engage in sensemaking—for example, by collecting posts from a specific conversation or event. A major aspect of research studying social media curation involves developing systems that can assist human curators in sorting through large amounts of content, often by providing automatic techniques for recommending new and relevant content [8, 25]. We extend this approach; in this paper, machines assist by providing *structure* rather than direct suggestions, and crowd workers then interpret the provided structure to generate new content.

Crowdsourcing Stories and Reports

Crowdsourcing is often used as a tool to generate ideas and break down creative tasks into smaller pieces [12]. In NewsPad [16] and Eventful [3], a story author asks crowd workers to create or retrieve content for news stories—for example, by asking workers to go to a particular event and take a specific set of photos. By delegating work to multiple people and

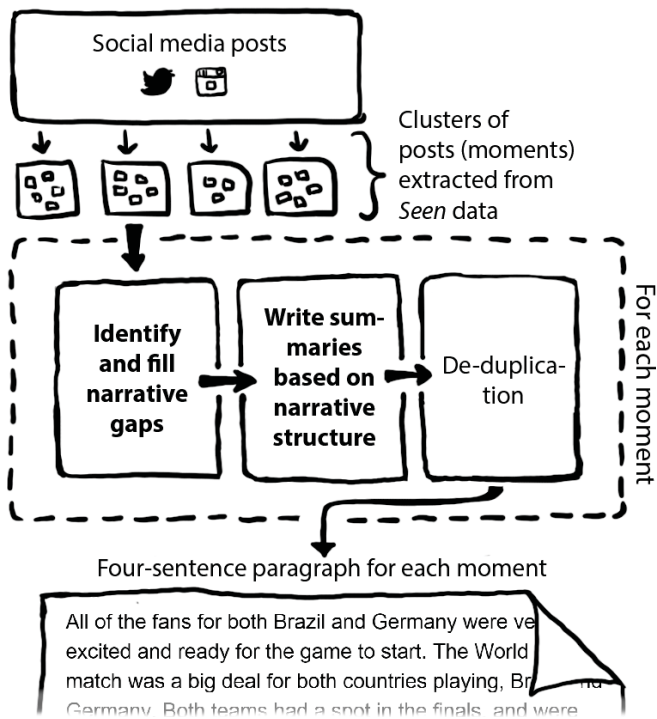


Figure 1. Overview of the Storia system.

drawing from content that has been already created, an author can more quickly collect a diverse set of content than if they were to work alone.

The application of crowdsourcing to more artistic work is also an emerging area of research. For example, in Ensemble [11], authors maintained an outline of a short fiction story to guide crowd workers in generating ideas and contributing content. While machines have showed promise in their ability to logically reorder or generate content according to narrative templates and structures, Storia hypothesizes we can augment this past work by making use of people’s unique capacity to understand and create emotion.

STORIA

Storia is a system that generates summaries of social media events through crowdsourcing tasks designed based on narrative theory.

Data

Storia is based on content from *Seen*¹, an online service that creates clusters of social media posts from Instagram, Twitter, and Vine based on time and keyword given a social media hashtag. Each cluster represents trending sub-issues or points in time related to the hashtag (e.g. tweets about the first goal of the #GERvsARG soccer game). In this paper, we call each cluster of posts a *moment*; several moments make up an *event*.

Design Challenge: Narrative Gaps

To understand the role social media play in how people construct a story about an event, we conducted preliminary interviews with 10 participants (six male, four female) recruited

¹<http://seen.co/about>

from research volunteer mailing lists. We showed participants a variety of views of the 2014 FIFA World Cup final match between Germany and Argentina (a raw social media feed, a timeline of clustered social media posts, and a news article) and observed them as they used the views to learn about factual and emotional aspects of the event. Afterwards, we asked about the strategies they use for finding information about events, the role social media play in these strategies, perceptions of the social media view they were given, and motivations behind sharing information and social media posts with others. Each session lasted about 45 minutes.

Through these interviews, we found three common themes in the type of information participants looked for while attempting to construct a picture of the event:

First, there was a lack of understanding of the *relationship between moments* during the event; even when social media posts were divided into clusters, participants had difficulty identifying the discrete parts that made up an event and how those parts related to one another:

I was trying to figure out the order of events like when that damn free kick happened. I have no idea.

Participant 5, timeline condition

Some participants had an easier time making predictions about where information might be located based on the fact that the story views shown to participants presented information in rough chronological order:

So I’m kind of like, from when everybody was like, “Oh, Germany won.” And just going right back from there and looking for “free kick” and a name in the post.

Participant 3, timeline condition

However, this strategy required *domain knowledge* to be useful; the participant above knew that free kicks usually happen near the end of a soccer match.

Second, participants were also confused about the *relationship between actors*. When asked whether Brazilians generally rooted for Argentina or Germany during the final match, eight participants stated they didn’t know. (Brazilians cheered for Germany, as Argentina is traditionally their rival with respect to soccer.) Thus, the potential conflict and outcomes at stake for each of the event’s “characters” were not clear to participants.

Lastly, most participants saw no *relationship between the event and their own lives*. Participants expressed little motivation to share or act on information found through social media with friends and family if they were not already interested or invested in the event in some way. We note that this is not necessarily an intrinsic weakness of social media but was also affected by participants’ individual interests; nevertheless, participants indicated that emotional investment was necessary for them to take further action regarding the event.

It is unsurprising that participants had difficulty finding this information (even when the information they were looking for was available in the data). Social media posts are generally made in the moment [20], and so may make little sense

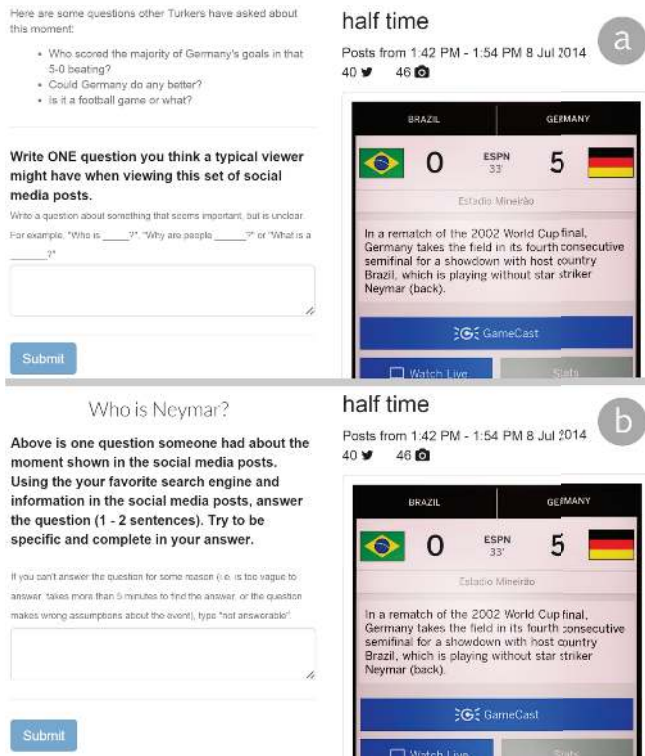


Figure 2. Crowd workers can (a) view the social media posts for the moment they are assigned to and ask a question that will be (b) answered by another crowd worker.

to future readers; these posts might refer to people using pronouns or nicknames, or simply comment on a moment without describing what they are referring to (“Did you see that!?!?”). Furthermore, finding information is a learned skill, and the difficulties people encountered may not be solely inherent to social media; supporting information retrieval and search is a large research area on its own [14]. Given these difficulties, automating the creation of narratives may allow people to understand large amounts of data more easily.

Extracting Narrative from Social Media

With these challenges in mind, we developed a prototype story creation system called Storia (Figure 1) comprised of three crowdsourcing modules. Storia takes as input a set of raw social media content clustered by moment, and outputs a written story consisting of several four-sentence paragraphs (one paragraph per moment).

Storia uses Amazon Mechanical Turk, an online crowdsourcing platform where workers can perform short microtasks for pay. In all tasks, crowd workers were from the U.S. who had a Mechanical Turk approval rating of over 90%.

Identifying and Filling Narrative Gaps

We first eliminate the narrative gaps we observed in our preliminary interviews. To alleviate confusion workers might have about unfamiliar names or terms encountered while processing social media content, we wanted to add context to the characters and actions that might appear in the final story.

Complete the Story

Help write a story about 2014 FIFA World Cup - Brazil vs Germany! You've been assigned to complete one part of the form below will guide you in transforming a set of social media posts into the next paragraph of the story.

Step 1: Skim story so far Step 2: Skim social media posts for the next part Step 3: Write the next paragraph

The fans prepared to watch Germany vs Brazil. Some were excited to hear the announcers attempt to pronounce the name of the German player Schweinsteiger. To the surprise of some, the Korean announcers were able to pronounce the name reasonably well.

Complete the Story

Help write a story about 2014 FIFA World Cup - Brazil vs Germany! You've been assigned to complete one part of the form below will guide you in transforming a set of social media posts into the next paragraph of the story.

Step 1: Skim story so far Step 2: Skim social media posts for the next part Step 3: Write the next paragraph

esta masacre, esto, seriously stop

Posts from 2:36 PM - 2:46 PM 8 Jul 2014 10 4



Complete the Story

Help write a story about 2014 FIFA World Cup - Brazil vs Germany! You've been assigned to complete one part of the form below will guide you in transforming a set of social media posts into the next paragraph of the story.

Step 1: Skim story so far Step 2: Skim social media posts for the next part Step 3: Write the next paragraph

How are answers by Turkors to questions from Turkors about these social media posts to help you write great sentences.

What are the Germans doing?

Why are people so anxious?

How is Brazil winning Germany score so many goals?

Which best describes when the social media posts were created?

During the event

Set the scene. Write ONE sentence about who is involved in this paragraph and what they were doing before the action started.

For example, "The runners were approaching the finish line."

Write in past tense, please!

Start the action. Write ONE sentence about how something began.

For example, "One of the runners saw his chance to slip through the other racers and started running hard."

Write in past tense, please!

Describe the main event. Write ONE sentence about what happened at the moment people posted these photos and tweets.

Figure 3. Storia guides summary writing for crowd workers. Workers are provided with (a) the story written so far, (b) the social media posts for a moment of the event, and (c) a structured form with prompts based on narrative categories.

Two Mechanical Turk tasks (Figure 2) collected this information for each moment:

Ask questions. For this task, we showed crowd workers a chronologically ordered stream of social media posts for a randomly chosen moment and asked workers what questions they thought a typical viewer might ask when viewing this content. Rather than suggest types of questions to ask based on narrative structure, we gave workers free rein to ask any question relevant to the goal of understanding the moment represented by the social media feed. We wanted to be open to the possibility that workers might ask unexpected types of questions, but workers did tend to ask questions related to narrative elements such as character and progression of plot, such as questions about names (“Who is Oscar?”), jargon specific to the event (“What is a free kick?”), and why some action was occurring (“Why was everyone booing Fred?”),

reflecting the types of narrative gaps we observed in preliminary interviews.

Answer questions. Here, we showed workers a question created by a worker from the previous task and asked them to briefly answer the question using a search engine or the information present in the social media stream. These workers saw the same stream of social media shown to the worker who had asked the question.

We collected two or three questions for each moment, and collected at least one answer for each question.

Writing Summaries using Narrative Categories

Second, we link content to narrative roles to generate a story structure for the event. In this module, we asked crowd workers to write one four-sentence paragraph for each moment of the event. Storia utilizes Cohn's four basic narrative categories (establisher, initial, peak, release) as a simple narrative template that crowd workers use as a base for the paragraph they write. Workers were provided with a view of paragraphs written by other workers for the overall story so far (Figure 3a), the set of social media posts about the moment for which they were writing their paragraph (Figure 3b), and the set of questions and answers generated by workers in the previous step. A structured form (Figure 3c) prompted workers in mapping the information available in the social media feed to each of the Cohn's basic narrative categories. The wording of the prompts used in the form differed slightly (Table 1) depending on when the worker thought the moment occurred (i.e., near the beginning, middle, or end of the event). This process resulted in a story consisting of several paragraphs (one paragraph per moment), where each paragraph was composed of four sentences that map to each of Cohn's four basic narrative categories.

We collected at least three paragraphs per moment then launched another task to ask other workers to vote for the paragraph they thought best represented the set of social media posts belonging to the moment. The highest voted paragraph became the representative paragraph for that moment. Votes were weighted slightly more if the voting worker indicated that he or she had watched the event.

De-duplication

Once we had one paragraph per moment, we ran a redundancy elimination task on Mechanical Turk. This is because the structured data we received from Seen.co did not produce mutually exclusive clusters—sometimes content appeared in multiple clusters, making it possible for the same content to inform multiple summaries by workers.

We showed workers a random paragraph from the story and asked them to select other paragraphs in the story they thought could replace their assigned paragraph without drastically affecting the story's overall flow or the information available to a reader. Workers were then asked to vote for the paragraph that best represented the entire group of paragraphs that they had selected.

We then grouped paragraphs by similarity. If at least two workers indicated that they thought Paragraph A was similar

to Paragraph B, we considered Paragraph A and Paragraph B as true duplicates. If one of the two duplicate paragraphs was already in a group, we simply added the other paragraph to the same group; similarly, if both paragraphs were already in groups, their respective groups were combined. We then tallied the votes for all paragraphs and used the highest-voted paragraph from each group for the final story.

EVALUATION

We hypothesized that recognizing and organizing social media data according to narrative roles could help workers overcome narrative gaps present in social media in order to produce evocative and automatic summaries of social media events.

Method

We tested our hypothesis through a controlled study comparing the output of two crowdsourcing workflows: the workflow used by Storia, described above (see Appendix A), and a control version of the workflow where we asked crowd workers to write four-sentence paragraphs for each moment of the event, but without prompts to guide the summary writing phase or the questions and answers generated about the event (see Appendix B). All other aspects of the control workflow remained the same as in the Storia workflow: paragraphs in the control condition went through a de-duplication process similar to that of the Storia condition in order to create a story with one paragraph for each moment.

We ran both the Storia and control workflows over social media posts about four different events (Table 2), and randomly sampled up to 12,000 posts from the entire corpus of content for each event. Stories ranged from sports events to television specials; in all cases, we chose widely viewed events based on topics that most crowd workers would be familiar with. Furthermore, we chose events that were well-structured (making it easy to compare crowd-created interpretations of the event with the actual sequence of events) and had some element of emotional arousal (and thus suitable for narration rather than just description). We may be able to apply our findings to events that are interpreted and processed over an indeterminate amount of time (such as natural disasters and other breaking news), but leave this as future work.

To measure how well a non-expert reader might understand the generated text stories for each condition, we asked 30 Mechanical Turk workers to complete a short evaluation task. As we were interested in emotional reactions of a general population rather than an objective sense of the quality of stories written, we chose not to evaluate stories with experts. Participants were shown the Storia and control stories for a random event in random order, then asked to choose the story they would be more likely to recommend to someone who had wanted to see the event but missed it. We also asked workers to briefly explain their choice. Participants were paid \$0.30 for this two-minute task. Participants' free-form responses to the task were analyzed to look for themes in how participants justified their choice.

A second within-subjects evaluation task asked 30 additional Mechanical Turk workers to evaluate stories according to

	Establisher	Initial	Peak	Release
Beginning	Make an introduction. Write ONE sentence about anyone or anything that hasn't appeared in the story yet, with a brief description about who/what they are.	Provide context. Write ONE sentence briefly summarizing the event as a whole.	Describe goals. Write ONE sentence about what the people or things you wrote above want to do during this event.	Describe the stakes. Write ONE sentence about what it would mean if they achieve their goal (or if they don't).
Middle	Set the scene. Write ONE sentence about who is involved in this paragraph and what they were doing before the action started.	Start the action. Write ONE sentence about how something began.	Describe the main event. Write ONE sentence about what happened at the moment people posted these photos and tweets.	Resolve the action. Write ONE sentence describing the aftermath or the reaction to the moment you see in these photos and tweets.
End	Summarize the result. Write ONE sentence about the important ending result of the event as a whole.	Summarize the reaction. Write ONE sentence about how people reacted to the result you wrote about in the sentence above.	Describe the consequences. Write ONE sentence about what the end result of the event means for the future.	Look to the future. Write ONE sentence about the next event or the next goal for the characters in this story.

Table 1. Prompts guided summary writing by workers. Storia prompts changed depending on whether the worker thought a moment occurred at the beginning, middle, or end of the event.

Event	# posts	# moments	Event Date
Sochi Winter Olympics Opening Ceremony	4691	27	7 Feb. 2014
2014 FIFA World Cup Semi-finals	1483	45	8 Jul. 2014
State of the Union (SOTU) 2015	11921	48	20 Jan. 2015
Glee Series Finale	5574	29	20 Mar. 2015

Table 2. The events used to generate stories through Storia and the control system. Content was randomly sampled from the entire corpus of posts for each event.

Task	# of HITS	\$ per HIT
Ask questions	2 per moment	\$0.10
Answer questions	2 per question	\$0.20
Write summaries	3 per moment	\$0.50
Voting for summaries	5 per moment	\$0.15
De-duplication	≥ 2 per moment	\$0.30

Table 3. Summary of the tasks and costs for both the Storia workflow and the control workflow.

more fine-grained dimensions (such as informativeness) using 7-point Likert scales. The stories from both conditions for a randomly chosen event were shown in random order. Participants were paid \$0.40 for this task. For all evaluation tasks, participants were restricted to Mechanical Turk workers who had not participated in any of the story creation tasks.

Results

Event type seemed to have a strong effect on the differences we observed between the Storia and control stories for each event. For this reason, we divide this section into two parts — the first section describes the results for the FIFA and Winter Olympics events, and the second section describes the results for the SOTU and Glee events.

“Placing me back in the game”

For the FIFA and Winter Olympics events, crowd workers followed the establisher-initial-peak-release pattern in the para-

graphs written for Storia stories, as seen in this example paragraph from the FIFA semi-finals Storia story:

The fans are sitting in front of their TVs and smartphones getting excited about the match as it starts.

Germany scores their first goal against Brazil, and the fans are going wild rooting for Germany.

Germany then goes on to scored their second, then their third and finally their fourth goal against Brazil, who has zero goals.

Fans cannot believe what they're seeing and they're wondering if this is a match or a bloodbath because Germany has completely demolished Brazil.

In contrast, paragraphs written for control stories conveyed less of a dramatic arc, and instead tended to dwell on the same idea for most of its sentences. For example, each sentence in the following paragraph from the Winter Olympics control story mentions that viewers are ready for the event:

People watching the ceremony announced they were ready for it to begin.

The people watching were ready to support their individual nations.

People tweeted out picture of themselves wearing gear showing their commitment to their country.

Some people even tweeted out pictures of babies getting ready for their first opening ceremony.

For these two events, participants significantly preferred Storia stories over control stories (FIFA: $\chi^2(1) = 6.5333, p < 0.05$; Winter Olympics: $\chi^2(1) = 10.8, p < 0.01$); participants preferred the Storia story over the control story 73% of the time for the FIFA event and 80% of the time for the Winter Olympics event.

Participants who picked Storia stories appreciated the large amount of detail included and felt that they were a more complete view of the event. Notably, most of the participants also justified their choice with some variant of “I felt like I was getting a vivid recap” or “the story captured the emotion”:

While it would be simple to just say that [the Storia story] is longer etc... It actually really expresses more emotion, more detail, and the ability to get a real feel for how the game went, the sentiments involved, everything to make it a better read!

Participant, FIFA event

Participants who voted for the control story stated they chose it because it was more concise, conveying major points about the event without including extraneous information:

[The Storia story] seems like a lot of non-quality information designed to entertain... while [the control story] is more informative.

Participant, Winter Olympics event

To these participants, the control story seemed more professional. However, the participants that chose Storia stories stated they *did not* pick control stories for very similar reasons: the control story felt like a bland generalization or a brief report. The second set of participants corroborated this sentiment; Friedman tests indicated that participants thought Storia stories had more interesting introductions (FIFA: $\chi^2(1) = 8.067, p < 0.01$; Winter Olympics: $\chi^2(1) = 5.762, p < 0.05$), giving Storia stories mean scores of 4.679 ($SD = 1.307$) for the FIFA event and 5.111 ($SD = 1.22$) for the Winter Olympics event, and control stories mean scores of 3.571 ($SD = 1.501$) for the FIFA event and 4.37 ($SD = 1.363$) for the Winter Olympics event.

Participants also thought Storia stories for these events were more informative (FIFA: $\chi^2(1) = 16.2, p < 0.01$; Winter Olympics: $\chi^2(1) = 7.1176, p < 0.01$), giving Storia stories mean scores of 5.679 ($SD = 0.612$) for the FIFA event and 5.519 ($SD = 1.087$) for the Winter Olympics event, and control stories mean scores of 4.393 ($SD = 1.343$) for the FIFA event and 4.778 ($SD = 1.086$) for the Winter Olympics event.

Lastly, participants felt that the Storia stories for these events were more likely to make readers feel as if they were there (FIFA: $\chi^2(1) = 15.385, p < 0.01$; Winter Olympics: $\chi^2(1) = 4.262, p < 0.05$), giving Storia stories mean scores of 5.143 ($SD = 1.079$) for the FIFA event and 5 ($SD = 1.144$) for the Winter Olympics event, and control stories mean scores of 3.25 ($SD = 1.404$) for the FIFA event and 4.111 ($SD = 1.528$) for the Winter Olympics event.

"I disliked both stories, but..."

The stories for the Glee and SOTU events were characteristically different from the stories for the FIFA and Winter Olympics events. For example, in both SOTU stories, workers added their own opinions regarding certain politicians using words not present in the provided set of social media content:

President Obama walked into the SOTU with the normal pomp and circumstance.

President Obama came out firm, expressing his overwhelming victories to the obstinate and denialist Congress.

John Boehner had the same, half asleep, half drunk look on his face.

With that, the gauntlet was thrown.

The control condition went so far to parody political relationships to the point of absurdity, which extended through several paragraphs:

Vladimir Putin was not happy, the endless enlargement of NATO could not stand!

John McCain on the other hand couldn't help but peer over at his sore buddy Putin and laugh to himself.

But the murmuring in the crowd quieted as Barack Obama approached the podium...

"Suck it Putin!" he exclaimed as he ripped off his shirt and exposed the "Superman S" on his undershirt! "I'm Barack Obama, you got that" and he flew away.

For these stories, there was no significant effect of study condition on participants' preferences (Glee: $\chi^2(1) = 0.133, n.s.$; SOTU: $\chi^2(1) = 1.2, n.s.$). For the Glee story, most participants stated they based their choice on writing quality rather than on emotional aspects. The second set of participants reflected this, as there was no significant difference in which story they thought would be more likely to make readers feel like they were at the event. In hindsight, this makes sense, as these events were meant to be televised rather than attended.

Opinions on which story participants preferred for the SOTU event was divided on the control story's use of parody—some workers thought it was amusing, but others disapproved:

[The control story] was juvenile and unintelligent in too many places... [The Storia story] provides a better summary that is more sophisticated and intelligent (even though it's not great either).

Participant, SOTU event

This is reflected in the results from the second evaluation task; neither Glee story was seen as more informative than the other ($\chi^2(1) = 7.118, n.s.$), but the Storia story for the SOTU event was seen as more informative ($\chi^2(1) = 10.889, p < 0.01$), receiving a mean score of 4.96 ($SD = 1.695$) while the control story received a mean score of 3.28 ($SD = 1.969$). Neither story, in both events, was seen as having a more interesting introduction (Glee: $\chi^2(1) = 3.556, n.s.$; SOTU: $\chi^2(1) = 2, n.s.$).

Overall, the approach of framing social media summarization around narrative seemed to be more effective for the FIFA and Winter Olympics events. For these events, participants did perceive the Storia story as having higher emotional value, choosing to recommend Storia stories to someone who wanted to learn more about the event.

DISCUSSION

Through an evaluation of narratives generated by crowdsourcing tasks based on narrative theory, we found that certain types of stories written with narrative guidance were more emotionally engaging and better suited for conveying an event to someone who had missed it. Other events did

not benefit as clearly from narrative guidance. Here, we discuss the strengths and limitations of this approach as seen in our results, as well as potential broader impact on future approaches for crowdsourcing creativity.

Interpreting Social Media

Notably, study participants did not observe emotional differences between the Storia and control stories for the Glee and SOTU events. We speculate this is because the narrative gaps in these events were too large—in the Glee event’s case, workers had little information about the overall arc of the finale episode, making it difficult to construct an emotional summary even with information retrieved by other crowd workers through questions and answers. Similarly, the SOTU event had no inherent narrative arc at all, with no clear winner, loser, or conflict. Because of this, the only meaningful thing workers could write about were their own opinions (or fictitious storylines). This is in contrast to sports events, which can be understandable and interesting even when one is not acquainted with the players or teams due to its familiar narrative and emotional pattern—a struggle to win, the joy of victory, and the pain of loss.

A strategy often employed by journalists for writing stories about events that do not have their own narrative arcs is to take a specific viewpoint. Storia may be able to accommodate this strategy, as it does not provide deterministic output; depending on the workers who participate in the process, one could view a State of the Union address from the perspective of a Democrat *or* from the perspective a Republican. One could also view events from the perspective of a person watching the event at home, or from the perspective of the event organizers, or from the perspective of players or performers, each with their own goals and hopes for the event.

Extending this further, this might allow us to apply Storia’s approach to other types of social media, such as question-answering sites (like Quora) or discussion sites (like reddit). Rather than generating a story about an event (which implies a narrative structure due to having a beginning and end), the crowd could, for example, identify multiple stances for opinion pieces based on an online discussion. These perspectives could then be available to anyone wanting to learn about different interpretations of the event. Stories are not reproductions of reality but representations of it (whether consciously designed or not). We have other means of learning what happened in reality (e.g. through recordings); this paper tries to look at how we can expose and examine an audience’s interpretation of what happened.

Converging Creative Goals

Crowdsourcing creative work typically requires splitting the project into smaller tasks *a priori* and stitching together the results of each sub-task to create the final product [12]. As a result, the success of a creative task depends heavily on the design of its sub-tasks. While Storia follows this approach in the sense that an event is described through a set of individual paragraphs, Storia also allows the crowd to write sub-tasks for itself: the crowd is able to ask questions that other workers can answer, for example.

The idea of asking crowd workers to generate crowdsourcing tasks for themselves is not new [13]; however, Storia points toward a strategy of allowing the crowd to *identify gaps in understanding*, which, in turn, becomes work for other crowd workers. That is, it dynamically structures creative work around testing and iteration rather on predetermined sub-tasks that eventually merge together. Enabling this flexibility may be even more important for creative work (such as story writing), which may have no objective solution and require workers to converge on common creative goals.

Limitations

As explained previously, we chose to apply Storia’s approach to events that had a clear beginning and an end, in order to ensure we would be able to compare crowd-created interpretations of the event with what actually happened. For this reason, we were able to use an approach where social media is automatically clustered into parts and the crowd linearly transforms these clusters into a story. This approach might be less suitable for events concerning natural disasters, breaking news, and other ongoing, developing stories—Storia currently relies on a story structure that is static, and interpreting the importance or emotional valence of certain moments may be difficult as new information and public opinion develops.

Despite the fact that participants appreciated Storia’s level of detail, both Storia and control stories were prone to factual errors. However, in this paper, our goal was to enable scalable creation of evocative and experientially-oriented summaries of social media events rather than aim for factual accuracy; we suspect additional strategies (such as asking a human editor to proofread stories generated by our system) can effectively address these issues.

Storia also limited itself to producing text output, despite the large number of visual content that appeared in social media feeds shown to workers. Images can be effective tools for expressing difficult-to-describe or intangible things such as emotion or atmosphere; in the New York Times’ article about Brazil’s loss to Germany in the 2014 FIFA World Cup semi-finals [4], photos and videos of stunned fans and grieving players accompany the text. Exploring how crowd workers can make use of multiple types of media while generating their stories is left as future work.

CONCLUSION

In this work, we explored the strategies for creating stories about events based on social media data that convey a public interpretation of an event. Our prototype, Storia, drew together narrative theories and crowdsourcing to create a system that generates collaborative creative work by *finding and filling narrative gaps* and *linking content to narrative roles*. Through a controlled study, we compared stories generated by Storia to stories crowdsourced with no narrative guidance, and found that, for certain events, participants found Storia stories more emotionally engaging and more appropriate for communicating what it felt like to view an event.

Are emotional stories better than objective ones? Journalists often struggle to maintain a balance between objectivity

and emotion when writing their stories. In journalism studies, the “emotionalization” of stories is often associated with sensationalism and the decline in quality of journalistic stories [19]; on the other hand, a case study of Pulitzer Prize-winning articles revealed that winning stories rely heavily on emotional storytelling, using strategies such as anecdotal introductions and expressions of affect in order to draw attention to complex and important social and political topics [24].

We did not attempt to compare stories generated by Storia with professionally written news articles and stories. Storia does not attempt to automatically generate Pulitzer-prize winning articles, nor does it try to solve the problem of distinguishing emotionality from tabloidism; it is obvious that our generated stories are nowhere near professional quality. However, motivated by evidence that emotional perspectives of events are valued, it takes a first step towards considering emotional and narrative arcs in the automatic generation of event summaries. People join social media networks to ask questions, read comments, and make connections—all to seek out what others think and how they feel. Social media is rich with declarations of emotions; by distilling the chatter of the social web, we may be able to bring out its voice.

ACKNOWLEDGMENTS

We would like to thank our colleagues at Microsoft Research, study participants, and Mechanical Turk workers for their valuable feedback and participation, as well as Neil Cohn for his insights on narrative theories. We also thank Mor Naaman and Seen.co for their help with their dataset. This material is based upon work supported by the NSF GRFP under Grant No. DGE-114747.

APPENDIX

A. Example Storia Story

The fans prepared to watch Germany vs Brazil. Some were excited to hear the announcers attempt to pronounce the name of the German player Schweinsteiger. To the surprise of some, the Korean announcers were able to pronounce the name reasonably well. The fans were impressed at the ability of the announcers to pronounce Schweinsteiger.

All of the fans for both Brazil and Germany were very excited and ready for the game to start. The World Cup match was a big deal for both countries playing, Brazil and Germany. Both teams had a spot in the finals, and were very anxious to win the game to proceed. Losing the game would mean they would have no chance at the World Cup.

Brazil and Germany had both worked very hard to get to this point, and only one would be able to continue. The fans were going wild as the opening ceremonies began. Both teams played very well. Only one team actually won.

David Luiz held the jersey of his injured teammate Neymar. The team began to sing the national anthem. Their enthusiasm shone brightly as they sang the anthem like a war cry. Fans speculated whether David Luiz could lead the team to victory as they geared up for the match.

The fans are sitting in front of their TVs and smartphones getting excited about the match as it starts. Germany scores their first goal against Brazil, and the fans are going wild rooting for Germany. Germany then goes on to score their second, then their third and finally their fourth goal against Brazil, who has zero goals. Fans cannot believe what they're seeing and they're wondering if this is a

match or a bloodbath because Germany has completely demolished Brazil.

Germany was on fire and scored goal after goal. The Brazilian fans started to become angered while the Germans were more and more elated. Anger turned to despair for Brazilian fans as Germany extended their lead by an incredible margin. As loss seemed inevitable, the fans lamented the absence of Brazilian forward Neymar, who had been injured in a previous match.

The game was already heavily in Germany's favor. German player Lahm made a good tackle on Marcelo, winning the ball fairly. The fans were very impressed with Lahm's performance. It was just one more show of dominance by Germany.

Germany's leading scorer Thomas Muller drove towards the goal. Muller scored the first goal of the game, and the fans went wild. The fans remarked that this was Mueller's tenth World Cup goal, an impressive feat. Unfortunately for Brazil, things were only going to get worse from here.

Brazil and Germany were locked in a tense World Cup game. Fans are in disbelief about what has happened so far. Fans are stunned that Germany is dominating Brazil. Fans were in utter disbelief at the way the game was playing out.

The German players looked to still have a full tank of gas half way into the game. The Brazilians tried without luck to stop the German attack, but German technique was to hog the ball. Again, German took a shot at the goal with Schurrle moving into the goal box! One tweeter posted "They look like they've got concrete boots on

Brazil and Germany played a fierce game of soccer. Germany's goalkeeper blocked every one of Brazil's attempts to score a goal. The game ended in embarrassment for Brazil as they ended the game with 0 points, due to Germany's defensive goalkeeper. Germany has now advanced to the finals, leaving Brazilian fans fuming across the world.

With Brazil's weakness in depth, they had to play Fred as a striker. He showed promise in qualifying leading up to the event but has really struggled in Brazil. As he looked invisible on the pitch and Germany ran up the score, people wondered why he was even on the field. After the 7-0 scoreline flashed, fans wondered how a nation such as Brazil can have such little depth.

Germany managed to defeat Brazil 8-0. The Germany fans were ecstatic, while the Brazilians were shocked. Brazil, previously thought to be one of the best, will have this game rubbed in their faces until the end of time. Germany, boosted by their win, will be helped in the next stage.

Germany destroyed Brazil in the World Cup semifinal. Fans were astounded by the massacre. Brazil will suffer a setback on the international soccer stage. Germany looks to win in the final.

With Germany up 7 to 0, it appeared Brazil would go scoreless in this World Cup semifinal. With only minutes left, Brazilian player Oscar managed to score one goal. Now the score was 7 to 1, so at least Brazil had put some points on the board. Many fans thanked Oscar for helping give a small shred of respect to the Brazilian rout.

B. Example Control Story

Neuer, Lahm, Boateng, Kroos, Schweinsteiger and Müller comprised Germany's starting lineup in the semi-finals. Dante made it to Brazil's semi-finals, playing his first game for the World Cup. It had been speculated that Willian would replace Neymar, but it turned out to be Bernard instead. Dante replaced Thiago Silva, making Brazil's semi-final lineup Fred, Oscar & Bernard.

While casual fans thought Brazil would win, die-hard soccer fans thought Germany had the better team. It wasn't surprising how Germany played, but it was how Brazil played. For some reason it just seemed that Brazil didn't show up to start the game and things

steamrolled. Germany was the best team at this World Cup and deserved the title.

Thomas Müller scored the first point in the semi-finals! Left unguarded, he snuck the ball in through the corner. The fans went wild, as Germany put heavy pressure on Brazil with the early point. With an impressive record of scores and assists in World Cup games, Müller has truly shown himself as a strong asset to Germany's team.

After thirty minutes of the Germany-Brazil game, the game was over. With a score of 0-5 Germany fans were already celebrating. While some felt bad for Brazil fans, most were happy with the results. German fans were ecstatic that the chance to play in the World Cup final was given so easily to them.

Fans around the world had much to say about Brazil's disappointing performance in the 2014 FIFA World Cup. Some took to social media sites like Twitter to bash the Brazilian team. German fans were excited about their team and took every opportunity to make fun of Brazil. Brazilians were upset about their team.

As Germany crushed, you had to feel sorry for Brazil. Goal keeper Manuel Neuer was in great form, shutting down Brazil. Neymar was sulking, mighty Brazil was in trouble. Finally, deep in stoppage time, Brazil got some relief as they had scored.

REFERENCES

1. Narrative Science.
<http://www.narrativescience.com/>.
2. Storify. <http://storify.com/>.
3. Elena Agapie and Andrés Monroy-Hernandez. 2014. Eventful: Crowdsourcing Local News Reporting. Collective Intelligence Conference. <http://research.microsoft.com/apps/pubs/default.aspx?id=211606>
4. Sam Borden. 2014. World Cup 2014: Host Brazil Stunned by Germany in Semifinal. *The New York Times* (July 2014).
5. Deepayan Chakrabarti and Kunal Punera. 2011. Event Summarization Using Tweets. In *Proc. ICWSM (ICWSM)*.
6. Neil Cohn. 2013. Visual narrative structure. *Cognitive science* 37, 3 (2013), 413–452.
7. N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *IEEE Symposium on VAST*. 115–122.
8. Kevin Duh, Tsutomu Hirao, Akisato Kimura, Katsuhiko Ishiguro, Tomoharu Iwata, and Ching-Man Au Yeung. 2012. Creating Stories: Social Curation of Twitter Messages. In *Proc. ICWSM (ICWSM)*.
9. Andrew Gibson. 1996. *Towards a postmodern theory of narrative*. Edinburgh University Press, Edinburgh.
10. Renato Kempter, Valentina Sintsova, Claudiu Musat, and Pearl Pu. 2014. EmotionWatch: Visualizing Fine-Grained Emotions in Event-Related Tweets. In *Proc. ICWSM*.
11. Joy Kim, Justin Cheng, and Michael S. Bernstein. 2014. Ensemble: Exploring Complementary Strengths of Leaders and Crowds in Creative Collaboration. In *Proc. CSCW*. 745–755.
12. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The Future of Crowd Work. In *Proc. CSCW*. ACM, New York, NY, USA, 1301–1318.
13. Anand Kulkarni, Matthew Can, and Björn Hartmann. 2012. Collaboratively Crowdsourcing Workflows with Turkomatic. In *Proc. CSCW*. ACM, New York, NY, USA, 1003–1012.
14. Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46.
15. Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. 2011. Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration. In *Proc. CHI*. ACM, New York, NY, USA, 227–236.
16. J. Nathan Matias and Andrés Monroy-Hernandez. 2014. NewsPad: Designing for Collaborative Storytelling in Neighborhoods. In *Proc. CHI*. 1987–1992.
17. Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. 2012. Summarizing Sporting Events Using Twitter. In *Proc. IUI*. ACM, New York, NY, USA, 189–198.
18. Brendan O'Connor, Michel Krieger, and David Ahn. 2010. TweetMotif: Exploratory Search and Topic Summarization for Twitter. In *Proc. ICWSM*.
19. Mervi Pantti. 2010. The value of emotion: An examination of television journalists notions on emotionality. *European Journal of Communication* 25, 2 (2010), 168–181.
20. Steven Schirra, Huan Sun, and Frank Bentley. 2014. Together Alone: Motivations for Live-tweeting a Television Series. In *Proc. CHI*. 2441–2450.
21. Edward F Schneider. 2004. Death with a Story. *Human communication research* 30, 3 (2004), 361–375.
22. David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill. 2011. Peaks and Persistence: Modeling the Shape of Microblog Conversations. In *Proc. CSCW*. ACM, New York, NY, USA, 355–358.
23. Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita. 2010. Summarizing Microblogs Automatically. In *Proc. HLT*. Association for Computational Linguistics, Stroudsburg, PA, USA, 685–688.
24. Karin Wahl-Jorgensen. 2013. The strategic ritual of emotionality: A case study of Pulitzer Prize-winning articles. *Journalism* 14, 1 (2013), 129–145.
25. Changtao Zhong, Sunil Shah, Karthik Sundaravadivelan, and Nishanth Sastry. 2013. Sharing the Loves: Understanding the How and Why of Online Content Curation. In *Proc. ICWSM (ICWSM)*.