Variation in Situational Awareness Information due to Selection of Data Source, Summarization Method, and Method Implementation

M. Janina Sarol*, **Ly Dinh***, **Jana Diesner**
University of Illinois at Urbana-Champaign
mjsarol, dinh4, jdiesner@illinois.edu

Abstract

The increasing amount of information available about crisis events calls for tools that help people to efficiently and reliably sift through and summarize this information. When using such tools, practitioners have to make choices that can affect not only the relevant information retrieved, but also their understanding or situational awareness of a crisis. We present a study that assesses the impact of some of these choices on the resulting situational awareness information. We focus on commonly used sources (news, tweets, and blogs), text mining methods (topic modeling and text summarization), and a disaster-related text mining task (needs detection). Our main application context is the 2010 Haiti earthquake, supplemented by data about 11 other events. Our results show that situational awareness information retrieved about the same crisis event can be different or even conflicting based on these choices. We also found that analyzing news data can be helpful in aiding response efforts as they contain unique situational awareness information not found in other considered sources, including first-responder accounts.

Introduction

Situational awareness (SI) is defined as "the ability to identify, process, and comprehend the critical information about an incident" (U.S. Department of Homeland Security (DHS) 2008). Tasks to gain SI include gathering and analyzing data, and disseminating relevant information to other stakeholders and the public (FEMA 2018). During crisis events, people within and outside of crisis zones use multiple types of sources (e.g., social media and mainstream news) to learn, communicate, and exchange information, such as updates on where to seek shelter (Imran et al. 2013), locations of damaged infrastructure (Caragea et al. 2011), and how to donate (Olteanu, Vieweg, and Castillo 2015).

There is abundant information available to crisis responders, affected communities, and the general public that allows each of these groups to gain SI (Sarter and Woods 1991). However, large and diverse amounts of information can result in information overload (Hiltz and Plotnick 2013). Being able to distill this information before responders arrive at disaster sites may increase the efficiency of crisis response. In a post hoc analysis of the U.S.' response to the Haiti earthquake, the deployed responders did not have full SI before they arrived in Haiti, which impacted planning and coordination (Cecchine et al. 2013). To mitigate information overload and quickly obtain SI, a number of studies has focused on automatically (e.g., via machine learning (Imran, Mitra, and Castillo 2016)) extracting relevant information (Weil et al. 2008), such as alerts about dangerous situations, missing people, and building damages, from sources such as Twitter (Verma et al. 2011), blogs and online forums (Li and Chen 2008; Kitamoto 2005), and news articles (Tanev, Zavarella, and Steinberger 2017). In this paper, we refer to information related to situational awareness about an event and extracted from text data as situational awareness information (SII), which is different from raw data or information about crisis events.

While prior studies have applied text mining and machine learning methods to analyze sizable disaster-related corpora, the effects of data and method selection choices on the obtained SII have not been examined in detail. A number of studies used a single data source for studying SI (e.g., Twitter). Also, studies that did use multiple methods for SI analysis did not explicitly discuss the implications of their choices on their results. This paper is based on the premise that researchers and practitioners need to be aware of meaningful variations in their results due to their research design choices. In the context of this paper, we define "variation" as the difference in SII that one obtains about the same event based on human choices for how to extract information. We base this definition on statistics and observational research literature, which specifies that variation is an effect that can be caused by using different techniques for gathering information (Grimes and Schulz 2002). We test for variations in text mining results due to the selection of specific data sources, analysis methods, and implementation of a method or algorithm (e.g., preprocessing or formula tweaks).

We seek to answer the following research questions:

- RQ1: What, if any, are the differences in situational awareness information (SII) depending on source of information?
- RQ2: What, if any, are the differences in situational awareness information (SII) depending on text summarization method?
- RQ3: What, if any, are the differences in situational

*These authors contributed equally to this work.

awareness information (SII) depending on implementation of a text summarization method?

- RQ4: What are practical implications of source, method, and implementation choices on resembling first responders' accounts?

Research questions 1 to 3 focus on identifying variations in analysis results due to use of different data sources, namely, tweets, news, and blogs (RQ1); text summarization methods, namely, COWTS-TFIDF (Rudra et al. 2015) and SumBasic (Nenkova and Vanderwende 2005) (RQ2); and implementations of summarization methods (RQ3). We chose text summarization as a specific text mining method as it helps with handling information overload and is a common method for condensing large amounts of information. Results obtained from these research questions are then used to answer RQ4. To answer these questions, we examine the 2010 Haiti earthquake event in detail. We also use a sample of text documents from 11 other crisis events to supplement our results for RQ2 and RQ3.

We expect to see differences in results when different sources, methods, and implementations are used. After all, one of the primary goals of machine learning research is to improve the results from one algorithm to another. However, usually, these results are looked at through the lens of accuracy (or in general, their ability to resemble annotations of ground truth data). In real-time applications (e.g., during a disaster), there is often no ground truth data available, hence information is taken as is. Thus, it is important to understand the differences in information obtained when using different data sources, methods, and implementations. This paper makes the following contributions:

- We qualitatively and quantitatively assess how choices about data sources, analysis methods, and method implementations lead to differences in SII about the same event.
- We show how arbitrary as well as default choices or settings can impact and skew SII.
- We show how we can leverage readily available and previously validated text mining methods to extract SII from various types of text data about crisis events.

Related Work

Data Sources for Gaining Situational Awareness

Increasing amounts of information relevant to crisis events have been generated and shared online (Reuter, Hughes, and Kaufhold 2018; Palen and Anderson 2016). This represents a shift in main sources for gaining SI from television, radio, and print media to real-time or near-time updates, e.g., via Twitter, blogs, and online news. Researchers in crisis informatics (Palen and Anderson 2016; Reuter, Hughes, and Kaufhold 2018) have sought computational means to detect and examine crisis-related information from different text-based sources (Verma et al. 2011; Abel et al. 2011; Tkachenko, Jarvis, and Procter 2017) in a timely and unbiased manner. Using Twitter as a primary source, Verma and colleagues (2011) extracted linguistic features (unigrams, bigrams, parts-of-speech, subjective cues, register, tone) to detect tweets that contain content related to SI. They achieved over 80% accuracy. Olteanu and colleagues (2015) classified tweets based on informativeness, finding that 32% of the tweets contained useful information, 20% were about emotional support, 10% about donations and volunteers, 10% about caution and advice, and 7% about infrastructure damages.

While Twitter has become a useful source of information for emergency response, tweets are often short and noisy, which pose difficulties for text mining tasks (Schulz, Ristoski, and Paulheim 2013). Several studies found that linking news articles to tweets increased accuracy of text mining tasks such as semantic linking (Abel et al. 2011; Meij, Weerkamp, and De Rijke 2012). Abel and colleagues (2011) found that in comparison to tweets, news articles provide more relevant information on over 20 types of entities such as persons, organizations, and locations. Schmierbach and Oeldorf-Hirsch (2012) compared perceived credibility of sources and messages from tweets and mainstream news, and also found tweets to be less credible due to limited context and shortness.

In addition to Twitter and mainstream news, blogs and discussion forums are also important sources for SII. Several scholars have used Flickr blog geo-tags to improve the detection of floods and their anticipated locations (Tkachenko, Jarvis, and Procter 2017; Shklovski et al. 2010). Shklovski and colleagues (2010) found that Myspace blogs were used by victims of hurricane Katrina to communicate recovery efforts and community re-building.

Text Mining Methods for Gaining Situational Awareness

Researchers have proposed and used various automated text mining methods to process, classify, filter, and analyze sizable amounts of text data to aid responders in gaining SI (Hiltz and Plotnick 2013; Olteanu, Vieweg, and Castillo 2015). Reliable and validated text mining methods are necessary to produce trustworthy results (Reuter, Hughes, and Kaufhold 2018; Diesner 2015a,b). Rudra and colleagues (2015) proposed a summarization framework where content words (verbs, nouns, numerals) were extracted and used as a dominant features. Herbane (2010) used keywords-in-context to extract terms associated with "crisis" and "disaster" in the context of business crises. They found common themes in the keywords, such as lack of control, uncertainty and pressure, and financial impacts. Ramage, Dumais, and Liebling (2010) applied topic modeling, along with TFIDF, as features to classify tweets in terms of four dimensions, namely substance (events, ideas, people), social (communication and social activities), status (personal updates), and style (language use and tones).

Variations in Text Mining Results due to Choices about Sources, Methods, and Implementations

Researchers have to make a number of choices throughout the research process that relate to the selection of sources (Olteanu et al. 2019), experimental design (González-Bailón et al. 2014), and methods and parameters (Diesner 2013, 2015a). With multiple options of sources, methods and algorithms available, researchers need to be aware of the assumptions associated with their design choices to avoid undesirable variations or biases. Olteanu and colleagues' (2019) examination of online social data found that each source contained different design affordances, and behavioral norms among users and content contributors. Hargittai (2015) suggested that data triangulation (collecting data from multiple sources on the same topic) can help mitigate biases.

Prior studies have also suggested that different implementations of the same text mining method may yield different results (Diesner and Carley 2008; Uysal and Gunal 2014). Diesner and Carley (2008) tested four design decisions needed to implement a part-of-speech (POS) tagger (e.g., removing noise from training data, post-processing of unknown words) and found that performance of POS taggers depend on researchers' choices for some design decisions. Diesner (2013) examined the impact of methodological choices for extracting relations from text data on the resulting network structure, finding that using reference resolution changed the weight of 76% of all nodes and 23% of all edges, which substantially altered the network structure. Levy, Goldberg, and Dagan (2015) found that changing one hyperparameter setting in word2vec yielded better performance than using another algorithm or increasing the training data. We build upon this prior body of literature by bringing it to the context of SI and studying variations in SII due to various choices about sources, text summarization methods, and implementation details.

Methods

Experimental Design

We conducted our experiments with a "one-factor-at-a-time" (Czitrom 1999) design, in which only one factor (e.g., source) is varied while the other two factors (e.g., method and implementation) are kept constant in each experiment (as shown in Table 1). This experimental design is suitable to answer our first three research questions as it treats factors as being independent of one another. This allows us to test for the differences due to data, method, and implementation separately, and avoids confounding of our results.

RQ	Source	Method	Implementation
1	**Variable**	Constant	Constant
2	Constant	**Variable**	Constant
3	Constant	Constant	**Variable**

Table 1: One-factor-at-a-time Experimental Design

Data Collection and Preprocessing

Data Collection Our main application scenario is the 2010 Haiti Earthquake. Consistent with official documentation from the U.S. Military's Operation Unified Response (Cecchine et al. 2013), we gathered data according to the timeframe associated with the initial *Response* phase of the disaster, i.e., from January 12 to February 4, 2010.

We collected text data from Twitter and blogs using the Crimson Hexagon[1] platform and mainstream news articles using ProQuest[2]. Each of these sources were gathered post hoc through a number of databases that were already cleaned and organized. These data represents something of a "best case" scenario for analysis as documents were mostly (or all) about the Haiti earthquake. Starting with a relatively clean dataset lowered the chance of our results being affected by noisy and unrelated data.

To further validate our results for RQ2 and RQ3, we reused a set of tweets from 11 disaster events from CrisisNLP (Imran, Mitra, and Castillo 2016). We took these datasets as is, i.e., we did not remove any additional tweets. Lastly, for RQ4, we collected first responders' accounts through interviews and from official situational reports.

Table 2 shows a summary of the datasets we used in this study along with their source and number of documents per dataset. Overall, we used 12 disaster events and 15 different datasets for analysis.

Event	Data Source	Number of Documents
2010 Haiti earthquake	Blogs	14,424
	News	7,512
	Twitter	26,849
	FR Accts.	402
2013 Pakistan earthquake	Twitter	1,881
2014 California earthquake		1,701
2014 Chile earthquake		1,932
2014 Ebola crisis		1,774
2014 Hurricane Odile		1,262
2014 India floods		1,820
2014 MERS crisis		1,358
2014 Pakistan floods		1,769
2014 Typhoon Hagupit		2,010
2015 Cyclone Pam		2,004
2015 Nepal earthquake		3,003

Table 2: Total Documents per Data Source

Data Preprocessing We converted all words to lowercase characters. We identified meaningful n-grams (phrases) using Autophrase (Shang et al. 2018), and we processsed these n-grams as single tokens, e.g., Dominican Republic to dominican-republic. We performed whitespace tokenization, and removed numbers, words containing numbers (e.g., 1st), non-Latin characters, and punctuations (except for dashes, which signify phrases). We also removed URLs (tokens starting with 'http://', 'https://', and 'www.') and stopwords from the scikit-learn stopword list. For tweets, we also removed mentions (i.e., @user), which is a common preprocessing method for analyzing Twitter data.

Impact of Source Selection on SII Variation

We applied three text mining methods, i.e., topic modeling, text summarization, and needs detection, to three data

[1]https://forsight.crimsonhexagon.com/ch/home

[2]https://www.proquest.com

sources about the Haiti earthquake, i.e., blogs, news, and Twitter data. We chose topic modeling and text summarization because they are general-purpose text mining methods that aim to capture the gist of a set of documents, and needs detection because it is a domain (i.e., disaster) related information extraction task. Topic modeling extracts the salient themes contained in a set of documents as unlabeled word vectors (Blei, Ng, and Jordan 2003). (Extractive) text summarization identifies the most salient sentences in a corpus (Eisenstein 2019). Needs detection identifies the resources needed during a disaster, e.g., by extracting tweets that mention resources needed (Basu et al. 2017) or creating a ranked list of resources (Sarol et al. 2020).

For topic modeling, we used Latent Dirichlet Allocation (LDA) as implemented in Mallet (McCallum 2002). We combined documents from all sources to created a single topic model with 15 topics. LDA generates a document-topic distribution, which represents how frequent a topic is discussed in a particular document. To identify the main topic discussed in a document, we decided on the following threshold: if a topic has a distribution greater than 50% in a single document, then we consider that topic as the main topic for that document. There can only be one main topic per document, and some documents may not have a main topic. We compared the distributions of the documents per topic across the data sources.

To compare summaries, we used COWTS, a disaster-specific text summarization method proposed by Rudra and colleagues (2015) for extracting a set of most "important and informative" (p.140) tweets that capture the main points of a corpus. The goal of COWTS is to maximize the presence of important content words in the generated summary, where a word's importance is based on its TFIDF score. To increase the method's efficiency, we only selected the 100 words with the highest TFIDF scores as content words and kept their scores as weights. As the method was developed for short documents (i.e., tweets), we split the documents into sentences using the NLTK sentence tokenizer, and then generated sentence-based summaries using COWTS. We set the following thresholds: minimum sentence length = 10 words and maximum summary length = 100 words. We qualitatively compared the generated summaries in terms of major topics/themes observed and their coverage. We refer to the output of the summarization method as summary sentences.

For needs detection, we used a method introduced by Sarol and colleagues (2020), which produces a set of keywords that represent resources needed (e.g., food, water, shelter). The ranked list is produced by finding the top terms whose word embeddings are closest (in terms of cosine similarity) to the words 'needs' and 'supplies'. We herein considered the top 20 retrieved terms. We generated word embeddings by training a word2vec (Mikolov et al. 2013) model per data source. Finally, we qualitatively compared the list of needs results across the data sources.

Impact of Method Selection on SII Variation

We compared COWTS to a general-purpose summarization method, namely SumBasic (Nenkova and Vanderwende 2005), which repeatedly selects the highest scoring sentence containing the content word with the highest probability of occurring at a given iteration. The sentence score is computed by averaging the probabilities of the content words occurring in the sentence. To limit the effects of preprocessing steps and parameter settings on the methods' differences, we kept the following steps the same across both summarization methods: tokenization, stopword list, minimum length per sentence in the summary (10 words), and maximum length of all sentences in the summary (100 words). To ensure that the results are not just reflective of a specific dataset or crisis event, we qualitatively compared the SII outputs generated by the methods on all non-First Responder (FR) datasets (details of the datasets in Table 2).

Impact of Method Implementation on SII Variation

We compared two different implementations of COWTS, for which we differed the selection of content words. For this purpose, we first used words with high TFIDF scores (we call this COWTS-TFIDF), and then used words with high keyword-in-context (KWIC) scores (we call this COWTS-KWIC). The KWIC score of each word represents the frequency of a term appearing in the same sentence as and within 7 non-stop words of the keyword. The keyword we selected for each event is the type of the disaster (e.g., for the 2013 Pakistan earthquake, the keyword is 'earthquake'). For the Haiti earthquake, as we had a larger number of documents, we used the keyword 'Haiti earthquake'. For both TFIDF and KWIC implementation, we selected the top 100 words as content words. We qualitatively compared the results of the implementations for all 14 non-FR datasets (details of the datasets in Table 2).

Results

Impact of Source Selection on SII Variation

We ran topic modeling, text summarization, and needs detection on three corpora about the Haiti earthquake: blogs, news articles, and Twitter. Table 3 shows the results of topic modeling as the distribution of the 15 topics over each data source. The topics were labeled by one author (L Dinh) and verified by another author (MJ Sarol). We initially labeled the first 5 topics as being about the *Haiti situation*. After examining a random sample of documents associated with each of these 5 topics, we determined that each topic discussed a different facet of the Haiti earthquake: 1) the Haiti earthquake in the news; 2) further information and facts about the earthquake (e.g., 7.0 magnitude); 3) the plight of the Haitian people; 4) relief efforts, including search and rescue missions; and 5) aftershocks and the possibility of earthquakes in other regions.

For news data, 43.7% of the articles contained main topics, with the most frequently discussed topics being *Haiti situation (relief efforts)* (10.8% of news articles) and *international aid* (6.6% of news articles). For blogs data, 39.6% of the articles contained main topics, with the most frequently discussed topics being *affect* (8.2% of blogs) and *donations* (4.4% of blogs). Twitter content had the lowest proportion of articles containing the main topics, at 15.1%. In tweets,

Topics (top 8 words)	Label	Blogs	News	Twitter
haiti earthquake user news victims video haiti-earthquake relief	Haiti news	1.5%	0.1%	6.1%
haiti earthquake people port-au-prince quake country magnitude buildings	Haiti earthquake	2.5%	3.2%	1.4%
haiti people country haitians world government haitian haiti's	Haitian people	2.8%	2.4%	0.0%
people port-au-prince rubble food days water earthquake survivors	Relief efforts	1.4%	10.8%	0.3%
earthquake haiti earthquakes oil january earth caused years	Aftershocks	1.7%	0.9%	0.1%
haiti relief earthquake donate donations money efforts people	Donations	4.4%	1.8%	5.9%
haiti earthquake hope show music telethon raise benefit	Fundraising	2.8%	4.7%	0.5%
haiti aid u.s relief military government haitian port-au-prince	International aid	2.5%	6.6%	0.3%
haiti medical people water team food supplies work	Medical aid	1.9%	1.3%	0.1%
people time good back day things life make	Affect	8.2%	0.3%	0.1%
school pounds february january students day saturday sunday	Schools	0.8%	6.4%	0.0%
god people haiti earthquake world israel church pray	Religion	2.7%	0.4%	0.1%
children haiti haitian earthquake family parents families orphans	Children	1.8%	3.8%	0.2%
obama jan u.s year state president news posted	U.S government	2.8%	1.0%	0.0%
climate posted jan data denny global-warming science people	Climate change	1.8%	0.0%	0.0%
	Total	39.6%	43.7%	15.1%

Table 3: Distribution of Main Topics for Each Data Source (Blogs, News, Twitter)

other main topics were *Haiti situation (news)* (6.1% of all articles) and *donations* (5.9% of all articles).

Tables 4, 5, and 6 show the results for summarizing texts per data source via COWTS-TFIDF. Bolded words are the content words identified using TFIDF. Of the three sources, only the news summary contained an entire list of sentences that pertained to SII updates about the Haiti situation, including sentences about the magnitude of the earthquake (sentence 1 in Table 5), ongoing relief efforts (3, 4, 5, 7, and 9), needs (6), and casualties (2, 8). The Twitter summary contained sentences about the magnitude (sentence 1 in Table 4), requests for help, donation, or prayers (2, 3, 5, 6, and 7), number of affected people (4), and a link to updates (8). The blogs summary contained sentences about the act of donating (sentence 1 in Table 6), requests for donations (2), religion (3, 5, and 9), school (4), ongoing relief efforts (6), the magnitude (7), and needs (8).

There was minimal overlap in SII between Twitter, news, and blogs summaries using COWTS-TFIDF. We found only one sentence that discussed the same information in all three summaries (sentences 1, 1, and 7, from the Twitter, news, and blogs summaries, respectively), which was about the magnitude of the earthquake. However, in the extracted sentences from Twitter and the news, the magnitude of the earthquake was 7.0, while in the sentence from the blogs summary, the magnitude was 7.3. We found one case where two sentences mentioned the same information, but one of the sentences contained more detail. Both blogs and news summaries contained a single sentence about the need for food, water, and medical items (sentence 8 in Table 6), but the news summary also mentioned tents (sentence 6 in Table 5). Overall, 7 summary sentences from each of the three sources contained specific information that was not present in the summary sentences from the other data sources.

Table 10 shows the needs we extracted per source. Both blogs and news data featured a notable proportion of resource-oriented words (55% for blogs, 50% for news), with blogs containing 11 such terms (food, equipment, resources, medications, medicines, shelter, supply, clean-water, medical-equipment, goods, medicine), and news containing 10 (resources, medical-care, medicines, clean-water, equipment, stocks, medicine, goods, food, shelter). Twitter data featured 2 resource terms (pounds, shoes), which were different from the terms found in blogs and news in that they did not address the most common essential needs. Overall, we found that Twitter data contained more action-oriented terms, mainly about donations (e.g., "collecting").

The Autophrase algorithm identified "urgently needed" as a salient phrase in both blogs and news data. For tweets, "need"-related terms were "need", "needs", "needed", and "needing". We extended the needs detection algorithm to extract urgent needs by having it return the terms closest to the embeddings of "urgently-needed" and "supplies". The list of the resulting urgent needs is shown in Table 11. Both blogs and news mentioned basic needs for immediate relief, such as medicine and water. In blogs, we found 5 terms that were specifically related to medical needs (medical-equipment, medicines, medical-assistance, medical-personnel, and medications), and the same number was found in news (i.e., medicines, medical-care, medicine, medical-assistance, and drugs).

Impact of Method Selection on SII Variation

Table 7 shows the results for summarizing the collected Haiti earthquake tweets using SumBasic. The words in bold have the highest probability of occurring in the Twitter dataset (in order of probability, i.e., "relief" has the highest probability in the Twitter dataset). Compared to COWTS-TFIDF, a larger proportion of summary sentences was about *donations* (75%; sentences 1, 2, 3, 6, 7, and 8 in Table 7), while the two remaining sentences were about the *Haiti situation*.

Even though the sentences came from the same dataset, no single sentence appeared in both outputs. However, a couple of sentences contained the same information. For exam-

1	catastrophic **quake**'**hits** haiti: fears of **major disaster** as it is **hit** by 7.0 **magnitude** earthquake.
2	__user__ **pls help** __user__ bring **aid** to haiti after **today**'s **devastating** earthquake!
3	we **need** the **prayers** and **help watch** – haiti earthquake **aftermath** __link__
4	red-cross says 3 **million people affected** by the earthquake in haiti
5	**new post**: **hope** for haiti, **hope** for the **world** __link__ ****donation info****please rt** #haiti #**donations** #earthquake
6	after __user__ **cnn** plea, i **donated** 2 haiti earthquake **fund.just text** "**yele**" to 501501 to **donate** $5, **visit** __link__
7	**support victims** of the earthquake in haiti by **texting** '**yele**' (**wyclef**'s foundation) to 501501 ($5) & or 'haiti' (**red**...
8	**list** of **relief efforts**, **news updates** on haiti earthquake **response**: __link__

Table 4: COWTS-TFIDF Results for Haiti Twitter Data

1	the 7.0 **magnitude quake struck just** before 5pm **local time** on **tuesday**.
2	many **thousands** of **people** in the **caribbean nation** are **dead**.
3	**children**'s **charity** plan-international **said** it had raised-pounds 610,000 in **emergency aid donations**.
4	are trying to **raise money** to **help** with the **relief effort** for their **home country**.
5	in **port-au-prince**, the **capital**, **foreign** rescue-teams scoured **buildings** for **survivors** under the **rubble**.
6	they **need** basic **supplies like food**, **water**, **medical** equipment and tents.
7	as a result a uk **international** search & **rescue team** has been deployed.
8	a **haitian government minister said yesterday** that 150,000 **bodies** had been counted.
9	the **world**'s **nations** have pledged $1 billion (**pounds** 616 **million**) in **aid**.

Table 5: COWTS-TFIDF Results for Haiti News Data

ple, the fourth sentence in both COWTS-TFIDF and SumBasic summaries stated that 3 million people were affected. Sentences about giving $5 donations by texting 'yele' to 501501 were also present in both summaries (sentences 6 and 7 for COWTS-TFIDF, and sentence 3 for COWTS-KWIC). Also, none of the other donation-related sentences from the original tweets linked to the same organization.

There were no sentences with conflicting information. While sentence 1 from COWTS-TFIDF and 5 from SumBasic mentioned different earthquake magnitudes, the first sentence referred to the original earthquake that occurred on January 12, 2010, while the second one referred to another earthquake that occurred 8 days later, on January 20, 2010. In general, we found that even though both Haiti earthquake summaries contained more differences than similarities, of the three sources, SumBasic and COWTS-TFIDF results overlapped the most for Twitter data.

Table 8 shows that for news, the SumBasic summary sentences greatly differed from the COWTS-TFIDF results. For instance, there were no sentences about ongoing relief efforts in the SumBasic results, whereas the COWTS-TFIDF summary contained 5 of these sentences. There was only one sentence in the SumBasic summary discussing the need for a relief system: "the relief system might not be working yet in haiti". The SumBasic results contained an estimated death count (sentence 8 in Table 8: "the government says around 170,000 were killed, at least 200,000 injured and a million left homeless."), while the COWTS-TFIDF results mentioned that there were thousands of dead people (sentence 2 in Table 5). For blogs, SumBasic results also differed greatly from COWTS-TFIDF results.

For the other 11 disaster events, results were similar to the results for the Twitter analysis of the Haiti earthquake in that there was moderate overlap between the information in the COWTS-TFIDF and SumBasic results. COWTS-TFIDF produced a total of 85 summary sentences (7.7 per event), while SumBasic produced 77 sentences (7 per event). Table 12 shows example sentences with different types of overlapping information. In total, we found 43 pairs of sentences with overlapping information, 40 sentences with information only present in the COWTS-TFIDF summaries, and 31 sentences with information only present in the SumBasic summaries. Of these 43 pairs of sentences, 18.6% (n=8) appeared in both COWTS-TFIDF and SumBasic summaries; 55.8% (n=24 non-identical sentence pairs) contained the same information; 7% (n=3) contained conflicting information; 16.3% (n=7) had one sentence with more information; and 2.3% of the sentence pairs (n=1) contained information not in the other sentence.

Impact of Method Implementation on SII Variation

Table 6 shows the summary sentences extracted for blogs using COWTS-TFIDF, and Table 9 shows the summary sentences for blogs using COWTS-KWIC. Only one sentence appeared in both summaries: "-jose irazuzta, md support project-hope's relief efforts inhaiti donate today". An additional sentence in the COWTS-KWIC summary was about donations (sentence 9). Two sentences from the COWTS-KWIC summary link to another website and article (sentences 1 and 6). Unlike the COWTS-TFIDF results, where 3 sentences were religion-related, none of the sentences in the COWTS-KWIC summary addresssed religion. For Haiti, the COWTS-TFIDF summary contained a sentence about the magnitude of the earthquake, a sentence about the NYC search and rescue taskforce on its way to Haiti, and a sentence with current needs. The COWTS-KWIC summary, on the other hand, contained a sentence about searching for victims a day after the earthquake and a sentence about members of the U.S. fire departments ready for deployment.

COWTS-KWIC produced 89 summary sentences for the 11 other disaster events. 46 pairs of sentences from both summaries contained overlapping information, 36 sentences

1	does donating **aid** to a foreign **country make** you less **american**?
2	-jose irazuzta, md **support** project-hope's **relief efforts** in haiti **donate today**
3	despite **great** difficulties, **god**'s **people** have found a **way** to **help** others.
4	i was **like** i **don**'t **need** full **time day care just** a preschool.
5	deng ming-dao do you **think haitian government** has **read** this meditation?
6	source: 29. **january** 14, yeshiva **world news** - (**international**; **new** york) nyc search & **rescue** taskforce heads to haiti.
7	the **magnitude** 7.3 **quake hit** close to the **capital port-au-prince**, officials **said**.
8	**right** now they **need money** for **food**, **water** and **medical supplies**.
9	how else will these **children know** the **love** of christ?

Table 6: COWTS-TFIDF Results for Haiti Blogs Data

1	if you can, here's how you can help w/ haiti earthquake **relief** __user__ __link__
2	go to __link__ or __link__ to donate to the earthquake **victims** in haiti.
3	please text **yele** to 501 501 now and $5 will go toward earthquake
4	over 3 million **people** have been affected by the earthquake in haiti.
5	i just heard on the **news** that another earthquake 6.1 just hit haiti!
6	you have donated $3million to red-cross earthquake relief **efforts** in #haiti by texting "haiti" to 90999. than ...
7	devastating earthquake in haiti to **aid** the children go to __link__
8	rt __user__ help now by donating to our haiti earthquake response **fund**: __link__

Table 7: SumBasic Results for Haiti Twitter Data

contained information only present in the COWTS-TFIDF summaries, and 40 sentences contained information only present in the COWTS-KWIC summaries. 36 sentences in the COWTS-TFIDF and COWTS-KWIC summaries conveyed the same information (24 from the same sentence, 12 from different sentences). We found 3 pairs of sentences with conflicting information, 6 pairs of sentences where one sentence had more information, and 1 one pair where both sentences contained information not in the other sentence.

Implications of Data Source, Method, and Implementation Choices

Our first research question asked about the differences in SII depending on source. To answer this question, we compared the outputs of three different methods: topic modeling, text summarization, and needs detection on three data sources: blogs, news, and Twitter documents.

1	but we have to move them out of haiti first, he **said**.
2	we have so much and there are all these **people** with nothing.
3	it should be, for there is much they can do to **help**.
4	it's been a week since **port-au-prince** was destroyed by an earthquake.
5	even before the quake, haiti was getting almost that much in **aid**.
6	the best canada can do for **haitian** children is to give them support in their own country.
7	the **relief** system might not be working yet in haiti.
8	the **government** says around 170,000 were killed, at least 200,000 injured and a million left homeless.

Table 8: SumBasic Results for Haiti News Data

1	for more **information** you can go to the **website** __link__
2	-jose irazuzta, md **support** project-hope's **relief efforts** in haiti **donate today**
3	imagine a **world** rocked by a **tragedy**, where **thousands** of **people** have died.
4	haiti-earthquake 2010: **update** with **live news** from **twitter** and blogs haiti-earthquake: breakingews, **updates** (**video**).
5	specially-trained members of **american** fire departments **prepared** to deploy the **day** after the **quake**.
6	there's a very **interesting article** on the **haitian disaster** at global view.
7	20.residents search for **victims** after an earthquake in **port-au-prince january** 13, 2010.
8	"human-beings are in some ways **like** bees," **professor** haidt **said**.
9	**help** me **raise money** for mercy-corps' **response** to the haiti-earthquake.

Table 9: COWTS-KWIC Results for Haiti Blogs Data

The topic modeling results showed that each source features different main topics: blogs focused on *donations* and *affect*-oriented content, news on relief efforts by different agencies (i.e., *Haiti situation (relief efforts)* and sources of *international aid*), and Twitter on updates about the Haiti situation (i.e., *Haiti situation (news)* and *donations*). Topic modeling results suggest that news contained more SII than blogs and tweets, and news contained more SII about the *Haiti situation* (17.4% of news articles are about one of the 5 *Haiti situation* topics).

Looking at the distribution of main topics, blog data was topically diverse or scattered as each topic was the primary topic for at least one document. Blogs also contained more types of narratives (topics/themes) than Twitter and news that were less relevant to SII, such as *religion* and *climate change*. In contrast to that, tweets showed the smallest percentage of main topics, and a notable number of topics was only discussed in 0.1% of all tweets. The Twitter result can be due to the fact that tweets only consist of a few words such that the topic model may not have enough words to

Blogs	News	Twitter
assistance	resources	collecting
medical-assistance	medical-care	provide
food	assistance	accepting
needed	medicines	continue
need	clean-water	save
equipment	urgently-needed	support
distribution	equipment	auction
resources	medical-assistance	credible
medications	essential	pounds
medicines	stocks	pledge
shelter	needed	emergency
supply	medicine	free
clean-water	life-saving	others
medical-equipment	necessary	providing
goods	capacity	shoes
medicine	providing	qualify
capacity	goods	invited
teams	food	reputable
provisions	shelter	americares
priorities	facilities	immediate

Table 10: Needs Detected from Blog, News, and Twitter

Blogs	News
medical-equipment	stocks
medicines	medicines
medical-assistance	life-saving
equipment	distributing
potable-water	equipment
supply	clean-water
shipments	goods
medical-personnel	supply
life-saving	medical-care
rations	vital
distribution	kits
goods	water-purification
hygiene-kits	medicine
kits	providing
packages	lifesaving
medications	medical-assistance
transport	shipments
heavy-equipment	deliveries
fuel	drugs
delivery	facilities

Table 11: Urgent Needs Detected from Blogs and News. Twitter omitted as 'urgently-needed' was not salient phrase.

properly categorize tweets.

Our text summarization results suggest that news data contained the most SII. Each of the sentences in the COWTS-TFIDF summary for news data contained a topic relevant to the Haiti situation, whereas blogs and Twitter also discussed other themes such as religion. All COWTS-TFIDF summaries contained a sentence about the magnitude of the earthquake. Aside from this similarity, the summaries for each source contained different situational narratives: Twitter mentioned the number of people affected, blogs mentioned that the NYC search and rescue taskforce was on their way to Haiti, and news mentioned that the UK international search and rescue team was on their way. A small but notable difference between the blogs and news data is the summary sentence about needs. The news summary also mentioned tents as a need, while the blogs summary only mentioned food, water, and medical items. While this could be seen as a minor difference, this detail is important during a disaster as shelter is a basic human need.

Another important dimension of SI is the ability to identify the needs of different stakeholders to appropriately respond to those needs. We compared the needs detected in blogs, news, and Twitter content. In blogs and news, the main resources retrieved from needs detection were food, water, and medical supplies, while for Twitter, resources identified were (British) pounds and shoes, which relate to raising funds and providing donations to victims in need.

Overall, our results from applying topic modeling, text summarization, and needs detection suggest that one gains different SII from each data source. News consistently contained the most amount of SII, and blogs featured relevant results for needs detection. Twitter contained information about donations and fundraising activities.

Our second research question asked about differences in SII depending on choice of method. We implemented two text summarization methods, one specifically developed for extracting summaries from disaster data (COWTS, specifically COWTS-TFIDF) and the other as a general purpose method (SumBasic). Results from using them to synthesize each of the 12 Twitter datasets showed medium overlap in the information conveyed by the COWTS-TFIDF and SumBasic summary sentences. However, on the news and blogs datasets, there is only a small overlap. A small portion of sentences contained conflicting information.

Time was a factor that contributed to the presence of conflicting information; the 3 cases of conflicting pairs of sentences arose because one tweet was older than the other. In one of these cases, the sentence from the older tweet mentioned that the airport was closed, while the other sentence mentioned that the airport was open for flights. In another case, the two sentences contained different death toll information (11 versus 327 people killed). Upon examining the tweets, the two tweets discussed different earthquakes of similar magnitudes that happened within 4 days of each other in the same general area.

Inspecting the information that only one summarization method retrieved, we found cases of missing sentences that could be crucial. For instance, "man did not have symptoms; tested pos after contact w/1st patient." only appeared in the SumBasic results of the 2014 MERS crisis. This was relevant information as it suggested that asymptomatic persons may still carry the virus.

Results from summarizing news and blogs data further suggest that the narrative about an event can differ depending on the chosen summarization method. Summaries of the news data generated by using COWTS-TFIDF mentioned relief efforts by several agencies, while the summary as per

Same sentence appearing in two summaries:	
#mers is a relatively new respiratory illness, spread b/w people in close contact.	
Same information appearing in non-identical sentences:	
according to officials a new island formed from a deadly massive 7.7m earthquake in #pakistan.	new island formed by a m7.7 #earthquake yesterday off the coast of pakistan.
Conflicting information:	
officials estimate 6.0 earthquake that hit california's wine-country caused $1 billion in damage.	napa earthquake damage, insurance losses could hit $4b __link__ #wine #winery
One sentence has more information:	
floods wash away homes in nepal, india, 180 dead __link__ #world	now- many dead in nepal and india floods__link__
Both sentences contain information not in the other	
rt __user__: cracked wine casks, damaged historical buildings and coffee shops.	sad about all lost wine but at least no fatalities.

Table 12: Sample Sentences with Overlapping Information

SumBasic mentioned that "the relief system might not be working yet in haiti". The SI gained about a disaster response can be different based on these sentences. A person reading the COWTS-TFIDF summary may be impressed with the rapid response by multiple nations and agencies, and might be dismayed by the lack of relief efforts when relying on the SumBasic results.

To answer our second research question, we found that variations in SII can occur depending on the choice of summarization method. Our results show that conflicting information may arise from using the two text summarization methods we tested and can lead to different SI of the event.

Our third research question asked about differences in SII depending on method implementation. We compared two implementations of the COWTS method that differed in content word selection. We found differences in narratives learned about a disaster due to this choice. The set of summarizing sentences for blogs and news with COWTS-TFIDF contained a sentence about resources needed, while the COWTS-KWIC results lacked that information. The COWTS-TFIDF summary for blogs contained a sentence about the NYC search and rescue heading to Haiti, whereas the COWTS-KWIC summary contained information about the Haiti residents searching for the victims.

For topic modeling, we compared outputs based on 10, 15, and 20 topics, and found 15 topics to comprehensively cover the data without being repetitive. Our results based on 10 topics missed out on salient topics such as *international aid* and *medical aid*, which were present in the 15 and 20 topic results. Additionally, results from 10 and 20 topics both contained the category of *miscellaneous*, which is not relevant SII. The results with 15 topics, on the other hand, do not contain the miscellaneous or broad categories, and capture the same set of topics found in 20-topics result.

In the original needs detection approach from Sarol and colleagues (2020), only the term "needs" was used for term extraction. We also considered other variations of the word "needs" (e.g., "need", "needed"). However, we also ended up only using "needs" as the resulting list of terms contained more resource-oriented words than the list returned when using other conjugations and variations. We also experimented with using the average of the word embeddings from different word forms, but the number of resource-oriented words was still highest when using "needs". These results suggest that even small choices in the implementation of methods can lead to different results and interpretations. Interestingly, using both "needs" and "need" produced "shoes" as one of the top terms for Twitter.

When using text mining methods, a non-trivial task is to select preprocessing steps. To do that, two of the authors qualitatively evaluated the outputs of each method from using different preprocessing steps (e.g., tokenization, phrase detection). For producing our final results, we used the preprocessing steps that led to the results containing most coherent information, as evaluated by the authors.

Another non-trivial task is to remove irrelevant documents (e.g., documents not about the disaster or not containing SII). We did not perform this step for the non-Haiti earthquake datasets. This resulted in some summaries having irrelevant information. For instance, COWTS-TFIDF, COWTS-KWIC, and SumBasic all selected the following sentence from the 2013 Pakistan earthquake: "rt __user__: another attack #peshawar - more than 30 dead & 80+ injured."; which is not related to that earthquake.

While data gathered from online sources have been shown to aid emergency responders in constructing SI (Verma et al. 2011; Olteanu, Vieweg, and Castillo 2015), they may entail misinformation (Castillo, Mendoza, and Poblete 2011), false rumors (Starbird et al. 2014), polarization of opinions, and echo chambers (Barberá et al. 2015), all of which may confound response efforts. We also found that misinformation can seep into the summary sentences. Summary results from COWTS-TFIDF, COWTS-KWIC, and SumBasic all contained a sentence posing a question whether salt water can cure Ebola. The World Health organization debunked this theory (2014). With the proliferation of fake news, misinformation, and disinformation, it is even more imperative that methodological choices are examined in more detail.

Overall, our results have shown that choices about method implementation can also lead to variations in SII. While the proportion of similar sentences when using two different implementations was higher than when using two different methods, the same problems arose: we also found conflicting information and majority of information not being present in both summaries. This meant that choices, from data collection to preprocessing to method application, need to be made in an informed manner and empirically tested.

Im summary, our results suggest that human choices about data sources, methods, and method implementation can cause variations in SII. These differences can lead to first responders and the public gaining different impressions of the same event. For first responders, the list of needs extracted from one source or with one method may be different or less comprehensive than the ones obtained by using another source or method. For the public, their impression of the response by government and non-governmental agencies may be different depending on the source, method, or implementation, and might for example lead to either satisfaction or dissatisfaction with the response.

Case Study: Comparison of Results to First Responders' Accounts

Our last research question asked for practical implications of choices about data source, method, and method implementation, specifically when comparing text mining results to first responder (FR) accounts. We compared the results obtained for research questions 1-3 to actual accounts of FRs on-scene at Haiti to determine how SII extracted via different data sources, methods, and implementations aligned with FR accounts. Our FR data comprised of transcribed interviews with first responders (n=5) and situational reports released by governmental and non-governmental agencies (n=397) during the initial Haiti earthquake response. We ran the text summarization (COWTS-TFIDF, COWTS-KWIC, and SumBasic) and needs detection on the FR accounts and compared the results to those for RQ 1-3. Finally, we also compared all results based on our narratives learned from FR accounts, which were primarily objective and fact-based.

The COWTS-TFIDF-based summary of FR accounts contained 9 sentences. Five of them discussed aid given by different agencies, 2 the *Haiti situation*, and 2 did not discuss any topic, but instead provided links to further information. The summary from news data was closest to the FR data summary, containing sentences about *international aid* and the *Haiti situation*. In the summary of FR data, the relief organizations mentioned were the United Nations Stabilization Mission in Haiti, United Nations, Pan American Health Organization, World Health Organization, World Food Programme, UNICEF, and International Organization for Migration. None of these organizations were mentioned in any of the summary sentences from news, blogs, and Twitter.

Our needs detection results for FR data returned a majority of action-oriented terms (top 3 terms were "ensure", "help", and "work"), and no specific resource-oriented terms (only term returned was "resource"). These results are similar to the distribution of action- and resource-oriented terms in Twitter. However, a closer look showed that Twitter and FR needs detection results did not have a single word in common. Blog and news-based needs detection results only had one term in common with FR results: "resources". Thus, we found no similarities between Twitter, blogs, and news data to the FR accounts when only comparing the actual needs detection results of the 4 data sources.

The World Food Programme report from January 15, 2010 (3 days after the earthquake), identified 'search and rescue, medical services and supplies, clean water and sanitation, emergency shelter, food, logistics and telecommunications" as urgent priorities (2010). The urgent needs detected from blogs and news data (shown in Table 11) lists several terms related to these priorities (e.g., medical-assistance, potable-water). These results suggest that blogs and news are reasonable sources for detecting needs similar to those expressed in FR data.

We found that COWTS-TFIDF and SumBasic produced comprehensive summaries of FR accounts: the COWTS-TIFDF summary contained two sentences about the limited status of the airport, and the SumBasic summary contained sentences about damaged hospitals and problems with the distribution of water due to fuel shortages. Both summaries included narratives on which agencies are providing aid: the United Nations Stabilization Mission in Haiti, United Nations, Pan American Health Organization, World Health Organization, and World Food Programme were mentioned in both summaries. The United Nations Stabilization Mission in Haiti, in particular, was one of organizations that led the Haiti earthquake response and relief efforts (Cecchine et al. 2013). This information was captured by both the COWTS-TFIDF and SumBasic summaries, as both contained the following sentence: "minustah is providing search and rescue operations, security, and assistance.".

We found that FR data summarization using COWTS-KWIC produced less informative sentences than COWTS-TFIDF. The only sentences that contained situational updates were "the paho/who emergency-operations-center-situation-report will issue situation reports as the situation requires." and "distribution will increase over the coming days to target 60,000 people.". None of the other summaries from news, blogs, and Twitter contained this information.

One sentence that did not contain any situational update but important information nonetheless was: "guide to humanitarian giving for the haiti-earthquake". The document that contained this sentence included a link to a website with consolidated information about helping the Haitian people. This information was also not present in any of the news, blogs, and Twitter summaries.

Overall, our text summarization results of FR data showed that SII extracted from FR accounts is vastly different to SII from news, blogs, and Twitter data. This finding does not mean that the information from FR accounts is not being reported in these data sources; more likely, it means that news, blogs, and Twitter prefer to focus on other SII (e.g., relief efforts by other agencies or countries), or that the methods we choose did not find matching information. Our in-depth comparison of narratives from the FR accounts support our above-mentioned finding that news contain the most SII as well as SII that is not present in the FR accounts (e.g., other relief efforts).. Therefore, news can be a reliable supplemental source of information for FRs. We also conclude that news and blogs can be sources of salient information about urgent needs and resources that should be prioritized in relief efforts. Our results further support findings from prior work that Twitter data contain less SII than the other sources (Schmierbach and Oeldorf-Hirsch 2012; Abel et al. 2011).

Conclusions

In this study, we showed how human choices about selecting data sources, text summarization methods, and tweaks to these methods (implementations) can affect what we learn about a disaster. Different choices can lead to different narratives of the same event; these differences can be about resources needed or the response by the government and other agencies. When creating a disaster management tool that uses text mining, practitioners are faced with a myriad of seemingly arbitrary choices. We show that these choices must be carefully considered as they affect more than just the accuracy of text mining methods; they affect the narratives in which the disaster responses are evaluated. These can have notable effects, from first responders not having the accurate information when providing relief efforts, to the public misjudging the relief efforts due to incomplete information. While developing new methods is important, in our study, we show that it is equally important to assess choices made during the development these methods.

Acknowledgments

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2015-ST-061-CIRC01. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. We appreciate helpful feedback and continuous support from the Critical Infrastructure Resilience Institute (CIRI) and U.S. Coast Guard. We also would like to thank our collaborators from Florida International University and University of Southern California for valuable feedback, and Dr. Nicolas LaLone from the University of Nebraska at Omaha for providing guidance and feedback throughout the project.

References

Abel, F.; Gao, Q.; Houben, G.-J.; and Tao, K. 2011. Semantic enrichment of twitter posts for user profile construction on the social web. In *Extended Semantic Web Conference*, 375–389.

Barberá, P.; Jost, J. T.; Nagler, J.; Tucker, J. A.; and Bonneau, R. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 26(10): 1531–1542.

Basu, M.; Ghosh, K.; Das, S.; Dey, R.; Bandyopadhyay, S.; and Ghosh, S. 2017. Identifying post-disaster resource needs and availabilities from microblogs. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 427–430.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022. 10.1162/jmlr.2003.3.4-5.993.

Caragea, C.; McNeese, N.; Jaiswal, A.; Traylor, G.; Kim, H.-W.; Mitra, P.; Wu, D.; Tapia, A. H.; Giles, L.; Jansen, B. J.; et al. 2011. Classifying text messages for the Haiti earthquake. In *Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 1–10. 10.1.1.370.6804.

Castillo, C.; Mendoza, M.; and Poblete, B. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web (WWW)*, 675–684.

Cecchine, G.; Morgan, F. E.; Wermuth, M. A.; Jackson, T.; and Schaefer, A. G. 2013. *The US military response to the 2010 Haiti earthquake: Considerations for Army leaders*. Rand Corporation.

Czitrom, V. 1999. One-factor-at-a-time versus designed experiments. *The American Statistician* 53(2): 126–131. 10.1080/00031305.1999.10474445.

Diesner, J. 2013. From texts to networks: detecting and managing the impact of methodological choices for extracting network data from text data. *KI-Künstliche Intelligenz* 27(1): 75–78. 10.1007/s13218-012-0225-0.

Diesner, J. 2015a. Small decisions with big impact on data analytics. *Big Data & Society* 2(2): 1–6.

Diesner, J. 2015b. Words and networks: How reliable are network data constructed from text data? In *Roles, Trust, and Reputation in Social Media Knowledge Markets*, 81–89. Springer. 10.1007/978-3-319-05467-4_5.

Diesner, J.; and Carley, K. M. 2008. Looking Under the Hood of Stochastic Machine Learning Algorithms for Parts of Speech Tagging. Technical report, Carnegie Mellon University Institute of Software Research.

Eisenstein, J. 2019. *Introduction to natural language processing*. MIT Press.

FEMA. 2018. Emergency Operations Center (EOC) Skillset: Situational Awareness. https://www.fema.gov/media-library-data/1537442017190-8f3bac7c67dfd13c2ac954bdf9b379c3/NQSEOCSkillset_SituationalAwareness.pdf. Accessed: 2019-09-10.

González-Bailón, S.; Wang, N.; Rivero, A.; Borge-Holthoefer, J.; and Moreno, Y. 2014. Assessing the bias in samples of large online networks. *Social Networks* 38: 16–27. 10.1016/j.socnet.2014.01.004.

Grimes, D. A.; and Schulz, K. F. 2002. Bias and causal associations in observational research. *The Lancet* 359(9302): 248–252. 10.1016/S0140-6736(02)07451-2.

Hargittai, E. 2015. Is bigger always better? Potential biases of big data derived from social network sites. *The Annals of the American Academy of Political and Social Science* 659(1): 63–76. 10.1177/0002716215570866.

Herbane, B. 2010. Small business research: Time for a crisis-based view. *International Small Business Journal* 28(1): 43–64. 10.1177/0266242609350804.

Hiltz, S. R.; and Plotnick, L. 2013. Dealing with information overload when using social media for emergency management: Emerging solutions. In *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 823–827.

Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; and Meier, P. 2013. Extracting information nuggets from disaster-related messages in social media. In *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 1–10.

Imran, M.; Mitra, P.; and Castillo, C. 2016. Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, 1638–1643.

Kitamoto, A. 2005. Digital typhoon: Near real-time aggregation, recombination and delivery of typhoon-related information. In *Proceeding of the 4th International Symposium on Digital Earth (ISDE)*, 1–16. 10.1.1.146.2372.

Levy, O.; Goldberg, Y.; and Dagan, I. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics (TACL)* 211–226. 10.1162/tacl_a_00134.

Li, M.; and Chen, A. 2008. A Web mining based measurement and monitoring model of urban mass panic in emergency management. In *Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 366–370. 10.1109/FSKD.2008.495.

McCallum, A. K. 2002. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu. Accessed: 2018-06-15.

Meij, E.; Weerkamp, W.; and De Rijke, M. 2012. Adding semantics to microblog posts. In *Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM)*, 563–572. 10.1145/2124295.2124364.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* 1–12.

Nenkova, A.; and Vanderwende, L. 2005. The impact of frequency on summarization. Technical report, Microsoft Research. MSR-TR-2005.

Olteanu, A.; Castillo, C.; Diaz, F.; and Kiciman, E. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2: 1–33.

Olteanu, A.; Vieweg, S.; and Castillo, C. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 994–1009. 10.1145/2675133.2675242.

Palen, L.; and Anderson, K. M. 2016. Crisis informatics—New data for extraordinary times. *Science* 353(6296): 224–225. 10.1126/science.aag2579.

Ramage, D.; Dumais, S.; and Liebling, D. 2010. Characterizing microblogs with topic models. In *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)*, 130–137.

Reuter, C.; Hughes, A. L.; and Kaufhold, M.-A. 2018. Social media in crisis management: An evaluation and analysis of crisis informatics research. *International Journal of Human–Computer Interaction* 34(4): 280–294.

Rudra, K.; Ghosh, S.; Ganguly, N.; Goyal, P.; and Ghosh, S. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *Proceedings of the 24th International on Conference on Information and Knowledge Management*, 583–592. 10.1145/2806416.2806485.

Sarol, M. J.; Dinh, L.; Rezapour, R.; Chin, C.-L.; Yang, P.; and Diesner, J. 2020. An Empirical Methodology for Detecting and Prioritizing Needs during Crisis Events. In *Findings of the Association for Computational Linguistics: EMNLP*, 4102–4107. 10.18653/v1/2020.findings-emnlp.366.

Sarter, N. B.; and Woods, D. D. 1991. Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology* 1(1): 45–57.

Schmierbach, M.; and Oeldorf-Hirsch, A. 2012. A little bird told me, so I didn't believe it: Twitter, credibility, and issue perceptions. *Communication Quarterly* 60(3): 317–337.

Schulz, A.; Ristoski, P.; and Paulheim, H. 2013. I see a car crash: Real-time detection of small scale incidents in microblogs. In *Extended Semantic Web Conference*, 22–33.

Shang, J.; Liu, J.; Jiang, M.; Ren, X.; Voss, C. R.; and Han, J. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 30(10): 1825–1837. 10.1109/TKDE.2018.2812203.

Shklovski, I.; Burke, M.; Kiesler, S.; and Kraut, R. 2010. Technology adoption and use in the aftermath of Hurricane Katrina in New Orleans. *American Behavioral Scientist* 53(8): 1228–1246. 10.1177/0002764209356252.

Starbird, K.; Maddock, J.; Orand, M.; Achterman, P.; and Mason, R. M. 2014. Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston marathon bombing. In *Proceedings of the International Conference on Information (iConference)*, 1–9.

Tanev, H.; Zavarella, V.; and Steinberger, J. 2017. Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media. In *Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 592–602.

Tkachenko, N.; Jarvis, S.; and Procter, R. 2017. Predicting floods with Flickr tags. *PloS One* 12(2): 1–13.

U.S. Department of Homeland Security (DHS). 2008. National Response Framework. Technical report, United States. Https://www.fema.gov/pdf/emergency/nrf/nrf-core.pdf.

Uysal, A. K.; and Gunal, S. 2014. The impact of preprocessing on text classification. *Information Processing & Management* 50(1): 104–112. 10.1016/j.ipm.2013.08.006.

Verma, S.; Vieweg, S.; Corvey, W. J.; Palen, L.; Martin, J. H.; Palmer, M.; Schram, A.; and Anderson, K. M. 2011. Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the 5th International Conference on Weblogs and Social Media (ICWSM)*, 385–392. 10.1.1.371.9621.

Weil, S. A.; Foster, P.; Freeman, J.; Carley, K.; Diesner, J.; Franz, T.; and Gorman, J. 2008. Converging approaches to automated communications-based assessment of team situation awareness. *Macrocognition in Teams. Theories and Methodologies* 277–300. 10.1201/9781315593166-16.

World Food Programme. 2010. Emergency Operation 200110 - Haiti, Food Assistance to earthquake affected populations in Haiti. Https://reliefweb.int/report/haiti/wfp-emergency-operation-200110-haiti-food-assistance-earthquake-affected-populations.

World Health Organization. 2014. Ebola: Experimental therapies and rumoured remedies. https://www.who.int/mediacentre/news/ebola/15-august-2014/en/. Accessed: 2020-05-10.