Data Collection and Annotation Pipeline for Social Good Projects

Christoph Scheunemann[*]**, Julian Naumann, Max Eichler, Kevin Stowe, Iryna Gurevych**

Ubiquitous Knowledge Processing Lab (UKP Lab), Department of Computer Science
Technical University of Darmstadt
`https://www.ukp.tu-darmstadt.de/`

Abstract

Vast amounts of data are generated during crisis events through both formal and informal sources, and this data can be used to make a positive impact in all phases of crisis events. However, collecting and annotating data quickly and effectively in the face of crises is a challenging task. Crises require quick, robust, and efficient annotation to best respond to unfolding events. Data must be accessed and aggregated across different platforms and sources, and annotation tools must be able to utilize this data effectively.

This work describes an architecture built for rapid collection and annotation of data from multiple sources which can then be built into machine learning and data analysis models. We extract data from social media via multiple systems for Twitter data collection, as well as building architecture for the collection of news articles from diverse sources. These can then be input into the INCEpTION annotation framework, which has been adapted to allow for easy management of multiple annotators, aiming to improve functionality to facilitate the application of citizen science. This allows us to rapidly prototype new annotation schema across a diverse array of data sources, which can then be deployed for machine learning. As a use case, we explore annotation of COVID-19 related Tweets and news articles for case prediction.
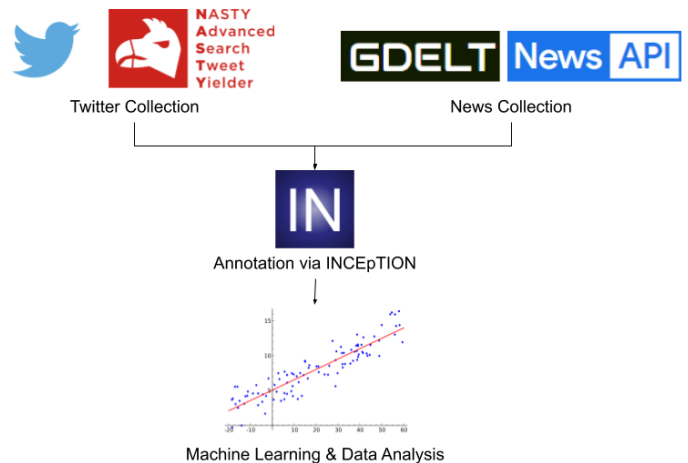
Figure 1: Overview of data collection and annotation pipeline. We collect data from Twitter using the official API and Nasty scraper, and from news using GDELT and News API. These are then fed to INCEpTION for annotation, and then used to build machine learning models.

Introduction

Data collection and annotation is a difficult process; this difficulty is compounded in crisis situations. In order to be effective, data collection and annotation needs to be quick and comprehensive. This requires analyzing "real-time" data sources (like social media) as well as data sources that are more broad (such as published news). This information is critical to build an understanding of the event in all phases of crisis, which is in turn necessary to provide adequate relief to impacted areas, build resilience, provide information to affected populations, and generally mitigate harm.

In order to facilitate the rapid collection and analysis of relevant information, we build an architecture for collection and annotation of data from multiple sources (Figure 1). First, we explore Twitter, building a system to extract not only live-streaming Tweets, but also older tweets. This allows users to quickly deploy a system that collects Tweets as they arrive (for instance, during the time immediately after a disaster event occurs), and then later collecting relevant data from before that point to complement the live stream. We also implement functionality to collect users' streams as well as reply graphs, both of which can increase our understanding of crisis events.

We also build architecture for collection of news articles based on keywords. This collection spans thousands of sources, allowing for rapid extraction of relevant news articles, giving a more comprehensive view of crisis events. While social media can be an effective lens to view crises, utilizing only one data source necessarily introduces the biases inherent to certain platforms. Adding additional data sources, particularly more formal sources, allows for not only a more thorough analysis of known problems but also new, interesting contributions combining informal social data with more formal sources.

We use both of these systems for data collection, and pro-

---

[*]First three authors with equal contribution.

vide infrastructure for then inputting the collected data to the INCEpTION platform for annotation. INCEpTION (Eckart de Castilho et al. 2018) is a powerful, flexible, web-based annotation platform. We supplement its architecture with functionality for improved management of multiple annotators over multiple tasks. This facilitates the use of citizen science (Gura 2013) by allowing project managers to easily manage and assess a large quantity of non-expert annotators that can contribute to relevant projects.

To highlight the utility of this architecture, we present a case study concerning the COVID-19 pandemic. We collect relevant data via our Twitter and news crawling architectures, supply it to the INCEpTION, and can then quickly deliver it to annotators for detailed coding. We use this output to then make predictions about COVID-19 cases based not only on previous case numbers, but also on social media reactions and news responses. This combination of social media and news events provides a novel viewpoint on this humanitarian crisis, highlighting the necessity of rapid data collection and annotation for multiple sources, and providing a blueprint for applying this architecture for social good projects. A practical step by step walkthrough of our work, including data collection from Twitter and news sources, import to INCEpTION, and making use of workflow improvements, is available at https://github.com/UKPLab/social-good-data-pipeline.

Related Work

There are numerous perspectives on collecting social media and other data during crisis situations (Reuter et al. 2017; Imran et al. 2015; Spence, Lachlan, and Rainear 2016; Castillo 2016). There are many issues that need to be accounted for (for a relevant bibliography, see Palen et al. (2020)). In order to be maximally impactful, data collection needs to be quick and efficient. In addition, flexibility is necessary to handle the changing nature of crisis events. Capturing data from multiple sources is essential to alleviate biases associated with using only a single type of data. Of course, there are also numerous ethical considerations that need to be made, particularly with regard to collection and distribution of data from populations that may be at risk. Even in public data such as Twitter, there are issues regarding users' understanding of their data that researchers need to be aware of (Fiesler and Proferes 2018).

The primary aspect we attempt to address is the massive amount of data created during these events. Keyword collection is often insufficient: keywords can both overgenerate (yielding excessive irrelevant data) as well as undergenerate (missing important data that lacks a given keyword). Consider the following example tweets (Stowe et al. 2018b):

1. Rock you like a HURRICANE!!

2. Just bought a 100 tea-light candles so yes, if we lose power, my apartment will look like The Bachelor finale

In collecting data for a possible hurricane event, using the *hurricane* keyword will yield (1), despite the fact that it is likely irrelevant. Conversely, given the context of a hurricane, (2) very likely gives relevant information pertaining to

the pre-crisis phase, exhibiting fear about incoming power losses and taking preparatory actions. However, it is unlikely that a basic list of keywords will capture this tweet, excepting perhaps *power*, which will again greatly increase the number of false positives.

For this reason, data is often first collected with keywords, then machine learning algorithms are applied to better filter find more relevant information. Our methodology supports this pattern: we build tools for quickly extracting relevant data based on keywords. This data is then sent to an annotation pipeline where fine-grained tags can be applied by trained or volunteer workers. Then, using this labelled data, we can build supervised machine learning models for better data processing and analysis.

There are a variety of related approaches for data collection. Perhaps most relevant is the Artificial Intelligence for Digital Response (AIDR), a platform for collecting and classifying social media messages during disasters (Imran et al. 2014). AIDR allows users to define search queries and collect Twitter data, which can then be further filtered using user-defined machine learning systems.

Another relevant systems is the Social Data Collection (SDC) of Reuter et al. (2017). This architecture allows for search and analysis of social media data through a user-friendly graphical interface, allowing non-experts to quickly build and analyze relevant data sets.

The architecture of Anderson et al. (2015) also provides a solution for Twitter collection in emergency situations. Their system is designed to scale, incorporating a variety of different technologies to process, store, and analyze massive incoming data streams via the Twitter firehouse. Their architecture is ideally suited to processing the extreme volume of data available on Twitter, but requires substantial setup and expertise, as well as relying on the official data stream from Twitter. Our methodology is orthogonal in that it is built for rapid, lightweight data collection, requiring little technical expertise.

Our framework differs others dealing with social media in a number of critical areas. First, we collect both informal social media data (via Twitter) and more formal news sources. This allows for a broader understanding of events. Second, our system retrieves both real-time and historic data, making it practical for all phases of crises. Finally, our architecture is built to seamlessly pass data to an annotation framework built to support citizen-science based annotation. This allows for the rapid collection and annotation of diverse data, which can then leveraged for machine learning purposes.

Twitter Collection

Our Twitter collection focuses on four different ways of scraping Twitter data, while relying on two different platforms for obtaining the data. The (1) method of obtaining data works by accessing the official Twitter API, which we access through the Tweepy library.[1] The (2) method relies on the NASTY scraper, which accesses the Twitter website by simulating http requests and processing of the resulting
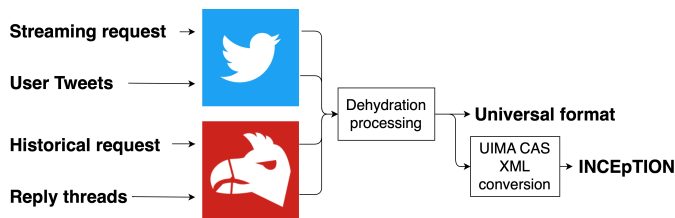
---

[1]https://www.tweepy.org

Figure 2: Twitter processing pipeline based on the official Twitter API and the NASTY project.

HTML.[2] See Figure 2 for an overview of the different ways of obtaining data.

Live Streaming API

The first way of obtaining Tweets is by using the streaming feature of the Twitter API. First, the researcher supplies keywords which they are interested in. The scraper then accesses the live stream of globally tweeted messages and filters it so that at least one of the keywords is contained in every Tweet which is then in turn passed of for further processing, i.e. meta data stripping. Through this method we are able to collect roughly 4,100,000 Tweets a day for our case study (described below). It utilizes the unpaid Twitter API, allowing easy access for researchers. It represents the most efficient way to gather data quickly. The three other use cases our scraper provides are supplementary to the streaming of live data.

Historical Scraping

The historical feature provides a way to gather data from the past. This is useful when certain dates from the past become interesting retrospectively, e.g. how populations reacted at the very beginning of a pandemic before researchers began their analysis. Similarly to the streaming feature, the researcher provides keywords which are then used in the request to the Twitter website. As this feature relies on the NASTY scraper instead of the official Twitter API, it is slower than the streaming feature and provides less data overall. In our tests, the historical feature was able to find up to roughly 710,000 Tweets for a specific search while the streaming feature was able to find 4,100,000 Tweets for the same time frame. 73% of the 710,000 historical Tweets were duplicates of the live stream data while the other 27% were distinct from the streaming data, indicating the potential usefulness of combining both methods. Combination is done in a duplicate avoidance fashion as explained further below.

User Collection

The user-collection feature provides researchers with an easy way of obtaining Tweets from specific users. This can be useful in cases where specific Twitter accounts provide useful information on an continual basis or where users of interest can be observed over a continual basis to observe

change in behavior. Previous research has shown the difficulty of analyzing tweets in isolation (Stowe et al. 2018a), and user streams allow us to better understand tweets within context.

Reply Threads

Reply thread collection is the last supplemental method the scraper provides. Given a Tweet ID it collects all the replies in a tree-like fashion. This supplements the user collection, and is also useful to researchers for examining Twitter discourses which are more in-context than just one-off messages.

Post-Processing

In case of topics which are not bound to one region but are rather trans-national, we provide a naive filtering implementation which provides the possibility to filter by language after the Tweets are collected. This implementation relies on the language meta-tag provided by Twitter, acknowledging that it is not always correct.

As all four methods potentially gather overlapping data, efficient and scalable duplicate detection was implemented in order to build data sets of good quality.

In the case of disaster studies which unfold over variable time spans, efficient data management is important to ensure scalable processing of the collected data. As crisis events unfold, the needs of researchers may change, and thus we need flexibility in terms of data collection. Therefore, we included a filtering feature for Tweet dehydration, whereby researchers can decide which metadata of a Tweet is interesting for their research and which can be filtered out after obtaining the Tweets.

Finally, we provide an easy way to export the data set from the default, universally-accessible format to the UIMA XML CAS format which is required for use in INCEpTION, in order to facilitate the pipeline from data collection to annotation. Another publishing feature also enables the publishing of the data set by providing a way to strip all meta data from the Tweet as is required by the Twitter terms of service.

Legality of Collecting Data From Twitter

As we are using the NASTY scraper to scrape Twitter data users of our application are potentially in violation of Twitter's terms of service as this is expressly prohibited and only allowed via certain bots (ie. Googlebot). Morally, it is questionable if citizens can be prohibited from freely accessing information in any way they deem appropriate from platforms which are used by politicians around the world.

Legally, in many jurisdictions it is still perfectly legal to use our tool, and by extension the NASTY scraper, as per the following laws:

1. In Germany, up to 75% of any publicly accessible data may be copied for academic purposes.[3]
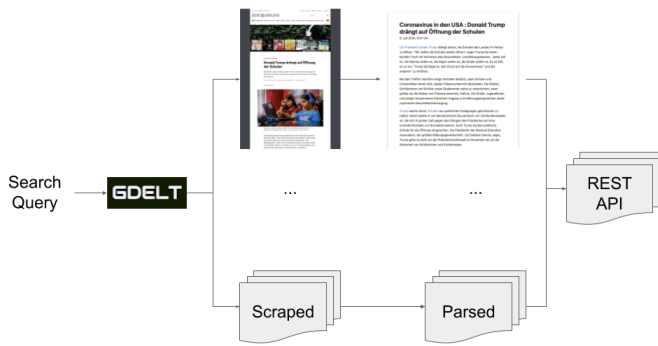
Figure 3: News data collection pipeline based on The GDELT Project.

2. In the United States of America, some courts have affirmed the right to scrape publicly available information.[4]

3. Depending on the jurisdiction, it is unclear if the Twitter's terms of service extend to users of our scraper, as users never agreed to the terms and could therefore not be bound to their conditions.

For further discussion on this topic, see documentation via the NASTY scraper repository.[5]

News Collection

Our news collection uses two data sources (1) The GDELT Project[6] and (2) News API[7] both of which provide access to a diverse set of world wide news outlets. GDELT provides access to a set of near real-time news articles and blogs for the last 3 month while News API focuses on news outlets. By providing data from these two sources researchers can easily compare data quality and adjust to different use cases, i.e. through the inclusion of blogs or a sole focus on news. While these two APIs have significant built-in functionality regarding search, they only provide researchers with URLs and some metadata. This leaves researchers with the task of scraping and parsing websites manually for data collection. We implemented an end-to-end data pipeline to crawl, scrape, and parse news sites, so researchers can put their focus towards experiments instead of data collection.

GDELT and News API both provide URLs plus additional metadata matching a search query. A variety of filters including language, country,[8] domain, and date range can be applied to each search query. Results can be further enhanced by using "and" / "or" syntax to find a subset of results, e.g. *COVID-19 AND (lockdown OR masks)*.

---

[4]http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf

[5]https://github.com/lschmelzeisen/nasty#legal-and-moral-considerations

[6]https://www.gdeltproject.org

[7]https://newsapi.org

[8]Country selection is only available on GDELT data.

We developed a data collection pipeline visualized in Figure 3 which passes search query and filters to GDELT or News API and scrapes each returned URL for its raw HTML data. To make the data more usable we apply boilerplate detection (Kohlschütter, Fankhauser, and Nejdl 2010) to extract the article title and main content. Boilerplate detection looks at the number of words and the link density to distinguish text from navigation, advertisement, and other sections of a website. While long(-ish) text usually features a high number of words with a low link density, the opposite is true for boilerplate which can then be removed. To further simplify the usage of our data we tokenize each article on a sentence level, and extract metadata like title, description, timestamp, domain, and language.

We provide JSON data via multiple REST API endpoints which vastly reduces the setup time for data collection and gives easier access to researchers from non-technical backgrounds. It is as simple as determining a search query plus optional filters and sending a http request. In Python a full data collection process can be set up in less then 5 lines of code.

Currently we support English, German, and Italian news articles for both data sources, though this could eventually be extended to up to a maximum of 65 languages supported by GDELT and 14 supported by News API.

Due to a focus on concurrent data processing our pipeline returns results in seconds which allows researchers to quickly try multiple search queries to gather a fitting data set for their needs. This is especially critical during a crisis where time is of the essence and circumstances might change quickly. If one searches for a hundred news articles, we create a unique thread to scrape and parse each of those. This is important since we are often faced with slow servers, timeouts, and sometimes sites that block our requests because of their noncompliance with GDPR laws. Since we frequently encounter websites that are inaccessible, we rarely retrieve all requested results. We mitigate this by adding a small buffer of additional sites which fill the gap if a process fails. This allows us to return as many news articles as possible without sacrificing performance. With our data pipeline researchers can rapidly try new search queries and evaluate how these changes affect the rest of their work.

In addition to raw data collection, the pipeline performance allows for the development of end user applications. Our data pipeline has been employed for a near real-time argument mining web application, showing its practicality for collection of news data.[9]

INCEpTION

We extended INCEpTION (Eckart de Castilho et al. 2018), a fully functional tool to annotate data from different possible data sources. INCEpTION is an online annotation platform that incorporates many related tasks into a joint web-based platform.[10] In this section we explain what INCEpTION is and how an annotator can annotate data. Furthermore, we

---

[9]http://asg.ukp.informatik.tu-darmstadt.de

[10]https://www.informatik.tu-darmstadt.de/ukp/research_6/current_projects/inception/index.en.jsp
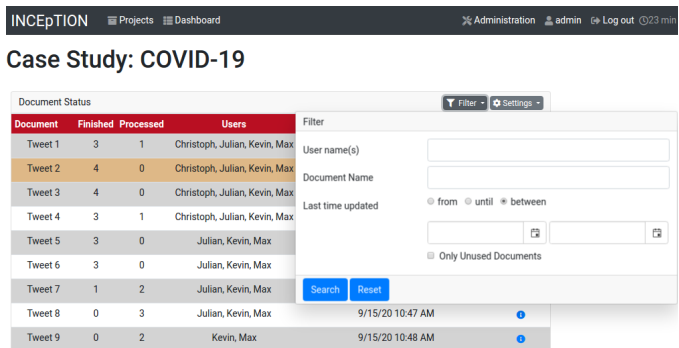
Figure 4: INCEpTION workload manager with a Twitter annotation project featuring an overview of finished and processed Tweets in addition to the filter UI.

will give an introduction on how we improved usability and functionality to support citizen science. However, following our walkthrough, which is linked in the introduction, is recommended as it will explain in detail how INCEpTION can be set up and how our features can be used as they are currently experimental only and therefore must be enabled manually.

Project structure

A project is the baseline for any annotation workflow in INCEpTION. All of the documents must be available in that specific project in order to be annotated. A project may consist of many annotators who can work simultaneously on any of the documents.

In order to add annotators to a project, every annotator requires an account on the corresponding server. Annotators only have the ability to annotate the documents they are provided with. While these accounts currently need to be manually created, part of our ongoing work is the incorporation of volunteer account creation, allowing non-experts to participate in annotation projects.

Data can be imported to INCEpTION directly using files in the UIMA XML CAS format. This format is output by both our Twitter and news crawling tools, simplifying the processing pipeline. Annotation is done document by document. When an annotator is finished with a specific document, they must mark it as "finished" to continue with the next one until all documents are processed.

Workload

The new workload manager seen in Figure 4 is our primary addition to INCEpTION. Our goal is to make INCEpTION more flexible regarding the monitoring of the documents within a project and thereby becoming more suitable for citizen science. We want to give project managers not only more control for the documents and the annotation workflow, but also a better and faster overview of their projects, allowing them to better manage a greater number of volunteer annotators.

Our new workload can be accessed only by a project manager and mainly consists of a substantial but easy to under-

stand table containing all the data a project manager needs from their documents. We added a comprehensive amount of filters and made the table sortable to increase its readability. We implemented filtering for the documents a specific annotator is working on, as well as for documents that have not been annotated at all. These documents can then be either cut out of the project or assigned manually to annotators to increase their progress.

Furthermore, the project manager is also able to get detailed information on how each of their annotators is complying with their work, e.g. by getting exact feedback on how many of their assigned documents are already finished in their annotation status and how many are still left to be done.

Our new workload system also enables a new workflow in the annotation process. Now, annotators can be automatically assigned to documents rather than making their own decisions. This prevents the case of having one specific document annotated too often, whereas others are not annotated at all. In addition, we decided, that it is significantly better for the annotation process if documents are finished linearly. Therefore, we disabled the ability for annotators to switch between their assigned documents. Instead, they must first finish the document they are currently working on before getting to the next one.

Another area we have focused on is performance. The larger the project, the bigger the difference in load times. Even for very small projects with only 20 documents, our workload page is three times faster than the old monitoring page (40ms compared to 120ms). This was mainly achieved by refreshing only necessary parts of the page and not using and recreating many smaller web resources.

To maintain consistency with older projects, we give researchers the possibility to decide which workload manager to use for their project: Either the previous monitoring type or our new workload manager. Switching between both of them is available at any time. The previous monitoring type also contains an overview table, but lacks important features for a project manager, including filtering and sorting.

In summary, we build a data pipeline for collecting both social media data and news from formal and informal sources. We updated the INCEpTION platform to improve management of annotation projects, improving speed and performance in order to facilitate larger, citizen-science based methodology for annotation. This architecture will facilitate quicker development of projects combining NLP practitioners and on-the-ground workers to develop high-quality understanding of crisis events.

Case Study: COVID-19

To showcase the usability of our data collection and annotation pipeline, we designed a small case study to predict COVID-19 case numbers. We decided to include Twitter sentiment (positive, neutral, or negative) as a potential indicator of peoples' behavior and news articles which are referencing an increase or decrease in measures by government or private entities (increased quarantine/mask regulations vs reopening schools and businesses).

We set up our Twitter scraper which collects publicly available COVID-19 related Tweets by searching for the keywords "corona", "covid", "epidemic", "pandemic", and "lockdown". A smaller, random subset of these Tweets is then exported into the UIMA XML CAS format which is used by INCEpTION. We try to distinguish between a positive sentiment, e.g. people who are optimistic about the current situation, negative sentiment, e.g. people who are unhappy with their government actions, and neutral, i.e. unrelated, unspecific Tweets or factual reporting of facts without expressing opinion. Our underlying assumption is that people with a positive sentiment are more likely to follow government rules which (theoretically) leads to a decrease case numbers.

For news data we collect articles mentioning *COVID-19 AND (lockdown OR masks OR measure OR opening)*. We annotate these articles into those mentioning an increase in restricting measures, e.g. lockdown or required masks in super markets, a decrease, e.g. opening of schools or allowing international travel, and those that do not mention or are merely discussing current measures. To develop these annotation guidelines INCEpTION provides automatic calculation of inter annotator agreement for two annotators.

After receiving both annotated Twitter sentiment data as well as news article about lockdown measures we train a neural network which takes those two features as well as officially reported COVID-19 case numbers in order to predict case number development for a short time frame, i.e. a time frame of five days into the future. This work is in progress; for documentation on the results, as well as a comprehensive walkthrough of our procedure, see https://github.com/UKPLab/social-good-data-pipeline.

Conclusions and Future Work

Many projects in NLP start with the collection and annotation of data. This holds especially true during a novel crisis where there might be no existing data set available. This is a time consuming endeavour and costs valuable resources when time is of the essence. We developed a general purpose data collection and annotation pipeline for researchers to rapidly gather Twitter and news data from various sources in near real-time, which then can be annotated using the INCEpTION annotation platform. We extended INCEpTION to accommodate citizen science by allowing researchers to manage a group of annotators and their workload.

As future work we would like to improve the robustness of our Twitter scraper, as the unofficial NASTY library relies on website parsing and is not supported by Twitter and therefore sometimes experiences performance issues and needs patching. With regard to news collection, extraction of user comments on news articles would be useful to better understand the article in context, public reactions, and its likely effects on the population.

INCEpTION will be further optimized towards the needs of citizen science, therefore, another important planned feature is the invitation of annotators and researchers via their *.edu* addresses, which will allow a more flexible login mechanism. As we want to increase INCEpTIONs ability to cope up with citizen science, granting many people from academia all around the globe the ability to create accounts and login by following an email link will make it much easier to find annotators for projects. Still, in order to maintain a good quality of the annotation work, we want to add the possibility for project manager to "mark" annotators who are performing poorly. These annotators will then not be offered any new documents after dropping below a certain threshold. Therefore, one of our key features will be a standardized test document each project automatically contains and which must be annotated by all annotators.

Having immediate access to social media data from Twitter and news articles across the world while utilizing an annotation platform for citizen science allows researchers to focus on their analysis of data and development of machine learning models to better understand how society reacts during an ongoing crisis. This data processing pipeline facilitates the better use of data for humanitarian efforts, particularly those that occur during crisis events.

Acknowledgements

References

Anderson, K. M.; Aydin, A. A.; Barrenechea, M.; Cardenas, A.; Hakeem, M.; and Jambi, S. 2015. Design Challenges/Solutions for Environments Supporting the Analysis of Social Media Data in Crisis Informatics Research. In *2015 48th Hawaii International Conference on System Sciences*, 163–172.

Castillo, C. 2016. *Big Crisis Data: Social Media in Disasters and Time-Critical Situations*. ISBN 9781107135765. doi:10.1017/9781316476840.

Eckart de Castilho, R.; Klie, J.; Kumar, N.; Boullosa, B.; and Gurevych, I. 2018. Linking Text and Knowledge Using the INCEpTION Annotation Platform. In *14th IEEE International Conference on e-Science, e-Science 2018, Amsterdam, The Netherlands, October 29 - November 1, 2018*, 327–328. IEEE Computer Society. doi:10.1109/eScience.2018.00077. URL https://doi.org/10.1109/eScience.2018.00077.

Fiesler, C.; and Proferes, N. 2018. âœParticipantâ Perceptions of Twitter Research Ethics. *Social Media + Society* 4(1): 2056305118763366. doi:10.1177/2056305118763366. URL https://doi.org/10.1177/2056305118763366.

Gura, T. 2013. Citizen science: Amateur experts. *Nature* 496: 259–261. doi:"10.1038/nj7444-259a".

Imran, M.; Castillo, C.; Diaz, F.; and Vieweg, S. 2015. Processing Social Media Messages in Mass Emergency: A Survey. *ACM Comput. Surv.* 47(4). ISSN 0360-0300. doi:10.1145/2771588. URL https://doi.org/10.1145/2771588.

Imran, M.; Castillo, C.; Lucas, J.; Meier, P.; and Vieweg, S. 2014. AIDR: Artificial Intelligence for Disaster Response. doi:10.1145/2567948.2577034.

Kohlschütter, C.; Fankhauser, P.; and Nejdl, W. 2010. Boilerplate detection using shallow text features. In Davison, B. D.; Suel, T.; Craswell, N.; and Liu, B., eds., *Proceedings of the Third*

International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010, 441–450. ACM. doi:10.1145/1718487.1718542. URL https://doi.org/10.1145/1718487.1718542.

Palen, L.; Anderson, J.; Bica, M.; Castillos, C.; and Crowley, J. 2020. Crisis Informatics: Human-Centered Research on Tech Crises: A Guided Bibliography Developed by Crisis Informatics Researchers. URL ffhal-02781763f.

Reuter, C.; Ludwig, T.; Kotthaus, C.; Kaufhold, M.-A.; Radziewski, E. v.; and Pipek, V. 2017. Big Data in a Crisis? Creating Social Media Datasets for Crisis Management Research. In Ziegler, J., ed., *i-com: Vol. 15, No. 3*, 249. Berlin: De Gruyter. doi:10.1515/icom-2016-0036.

Spence, P. R.; Lachlan, K. A.; and Rainear, A. M. 2016. Social media and crisis research: Data collection and directions. *Computers in Human Behavior* 54: 667 – 672. ISSN 0747-5632. doi:https://doi.org/10.1016/j.chb.2015.08.045. URL http://www.sciencedirect.com/science/article/pii/S0747563215005270.

Stowe, K.; Anderson, J.; Palmer, M.; Palen, L.; and Anderson, K. 2018a. Improving Classification of Twitter Behavior During Hurricane Events. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, 67–75. Melbourne, Australia: Association for Computational Linguistics. doi:10.18653/v1/W18-3512. URL https://www.aclweb.org/anthology/W18-3512.

Stowe, K.; Palmer, M.; Anderson, J.; Kogan, M.; Palen, L.; Anderson, K. M.; Morss, R.; Demuth, J.; and Lazrus, H. 2018b. Developing and Evaluating Annotation Procedures for Twitter Data during Hazard Events. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, 133–143. Santa Fe, New Mexico, USA: Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W18-4915.