Article

Towards Automated Situational Awareness Reporting for Disaster Management—A Case Study

Klaus Schwarz [1,2,]*[ ], **Daniel Arias Aranda** [2][ ] **and Michael Hartmann** [1]

1   Berlin School of Technology, SRH Berlin University of Applied Sciences, 10587 Berlin, Germany
2   Department of Business and Economics, University of Granada, 18071 Granada, Spain
*   Correspondence: klaus.schwarz@srh.de

Abstract: Disasters do not follow a predictable timetable. Rapid situational awareness is essential for disaster management. People witnessing a disaster in the same area and beyond often use social media to report, inform, summarize, update, or warn each other. These warnings and recommendations are faster than traditional news and mainstream media. However, due to the massive amount of raw and unfiltered information, the data cannot be managed by humans in time. Automated situational awareness reporting could significantly and sustainably improve disaster management and save lives by quickly filtering, detecting, and summarizing important information. In this work, we aim to provide a novel approach towards automated situational awareness reporting using microblogging data through event detection and summarization. Therefore, we combine an event detection algorithm with different summarization libraries. We test the proposed approach against data from the Russo-Ukrainian war to evaluate its real-time capabilities and determine how many of the events that occurred could be highlighted. The results reveal that the proposed approach can outline significant events. Further research can be carried out to improve short-text summarization and filtering.

Keywords: disaster management; automated reporting; sustainability in disaster management; situational awareness reporting; microblogging

1. Introduction

A disaster is a sudden accident or natural catastrophe that causes significant damage or loss of life [1]. Disasters can occur suddenly anywhere on Earth. In addition to natural disasters, there are also human-made disasters such as war [2]. Some might happen once, while others could be recurrent disasters such as earthquakes or tsunamis [3]. Even though it is impossible to avoid them, we must contain their consequences and act quickly once they occur. To achieve this, societies need to be as well-prepared as possible. Hence, disaster management refers to the totality of all coordinated measures in the areas of disaster prevention, disaster preparedness, disaster management, and post-disaster recovery, including the ongoing evaluation of actions taken in these areas. The phases of disaster management are structured in a continuous cycle—the Disaster Management Cycle—as depicted in Figure 1. The Disaster Management Cycle illustrates the ongoing process of reducing the impact of disasters. The overarching goal of disaster management is to reduce or avoid potential losses, provide rapid and appropriate assistance to victims, and ensure the fast rebuilding of infrastructure. The entire cycle includes the design of public policies and plans to mitigate the impacts on people, property, and infrastructure [4]. According to the definition of Al-Madhari et al. [1], disaster management actors are involved in the immediate response and recovery phases when a disaster occurs. For the most part, these are humanitarian organizations. The four phases of disaster management outlined here (in addition to the disaster itself) do not regularly occur in isolation. Often, the phases of the cycle overlap, and their lengths depend heavily on the severity of the disaster.
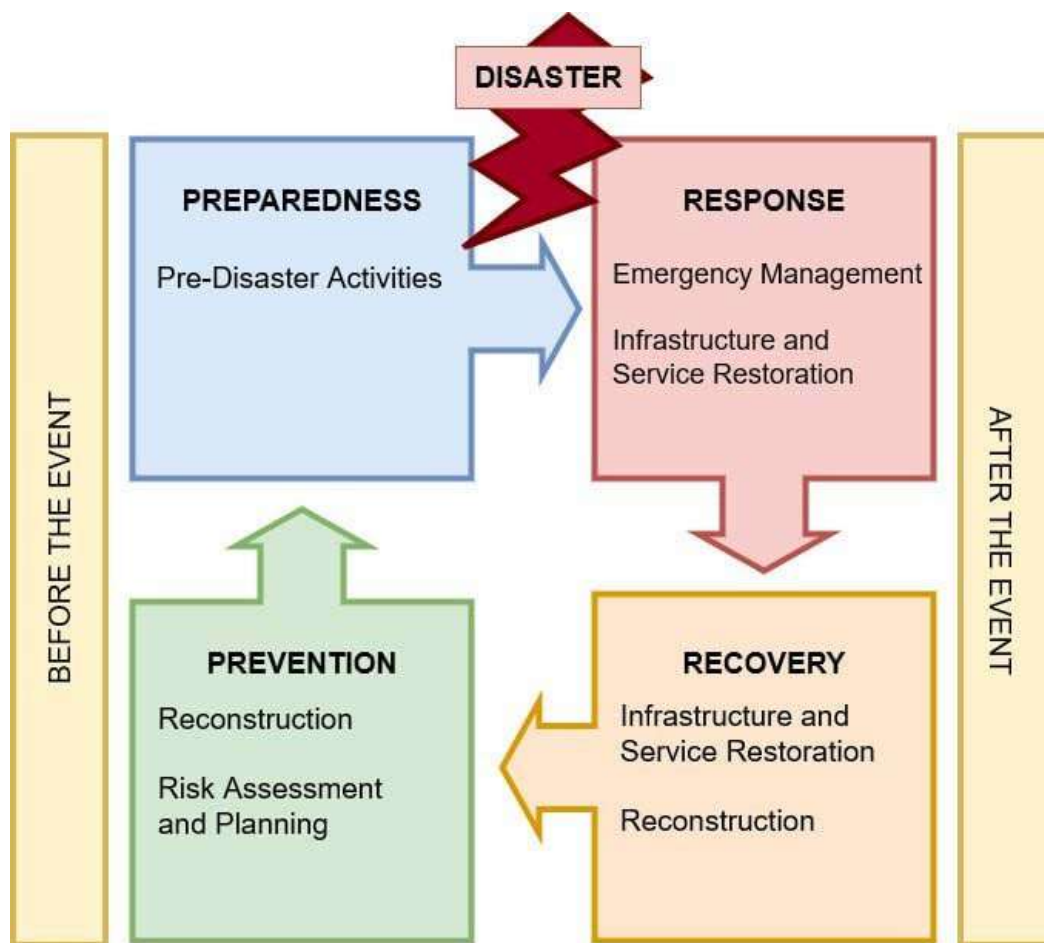
Figure 1. The Disaster Management Cycle.

The Disaster Management Cycle includes four phases. These are prevention, preparedness, response, and recovery. The prevention phase contains measures to reduce vulnerability to the consequences of a disaster. These include, for example, efforts to protect against injury, loss of life, and property. Strengthening public infrastructure, medical care, and other actions that increase a community's resilience to disasters are also included.

The preparedness phase is about understanding how a disaster might affect the community and how to build the capacity to respond to and recover from a disaster through education, outreach, and training. This may include business community involvement, pre-disaster strategic planning, and other logistical preparedness activities.

The response phase is the direct response to a disaster. It addresses the immediate threats the disaster poses through actions such as saving lives, meeting humanitarian needs such as food, shelter, clothing, public health, and safety, as well as cleaning up, assessing damage, and resource distribution. As the mission progresses, the focus shifts from addressing the immediate emergency to conducting repairs, restoring utilities, establishing public services, and completing cleanup. The most pressing emergency issues are assessed and addressed. There is often a degree of chaos during this phase, which can last for a while depending on the disaster's nature and the damage's extent.

The recovery phase is the fourth and final phase in the Disaster Management Cycle. It is typically the most long-term phase in the cycle and usually lasts for decades. It requires thoughtful strategic planning and actions to address a disaster's more severe or permanent impacts. In this work, we focus on the preparedness phase. The preparedness phase has been chosen, as we take the first step towards automated situational awareness reporting using microblogging data through event detection and summarization. The approach we

propose was developed using real data from past war events. The system proposed with this work is settled within the training phase of the preparedness cycle. Meanwhile, the long-term goal of this research is to develop a system that delivers sustainable improvement to the recovery phase of the Disaster Management Cycle. The National Incident Management System (NIMS) [5] defines preparedness as "a continuous cycle of planning, organizing, training, equipping, exercising, evaluating, and taking corrective action in an effort to ensure effective coordination during incident response".

This cycle depicts a broader system to prevent, respond to, recover from, and mitigate natural disasters, acts of terrorism, and other human-made disasters. There are five phases of the preparedness cycle [6]. These are:

1. Plan
   - The planning phase basically covers the entire cycle. The strategic and operational planning carried out in this section defines requirements and provides a benchmark for evaluating implementations. The planning elements show which capacities must be available during a major disaster [7].

2. Organize/Equip
   - Organization and equipment provide the human and technical resources needed to build and meet modernization and sustainability requirements. Organization and equipment include identifying the competencies and skills that staff should have and ensuring that an organization has the proper personnel. It also includes identifying and procuring standard equipment that an organization might need in emergencies [7].

3. Train
   - Training provides first responders, homeland security officials, emergency management officials, private and non-governmental partners, and other personnel with the knowledge, skills, and abilities needed to perform key tasks required during a specific emergency situation [6].

4. Exercise
   - Exercises enable entities to identify strengths and incorporate them within best practices to sustain and enhance existing capabilities. They also provide an objective assessment of gaps and shortfalls within plans, policies, and procedures to address areas for improvement prior to a real-world incident [7].

5. Evaluate
   - The final phase of the preparedness cycle is evaluation. In this phase, organizations gather lessons learned, develop improvement plans, and pursue corrective actions to address gaps and deficiencies identified during exercises or real-world events [6].

Situational awareness is an essential point in most phases of the cycle. Accordingly, it becomes essential in the response phase. Targeted actions can only be taken with a precise overview of the situation. As pointed out by Arias-Aranda et al., flaws in the constant monitoring of all actions cause critical failures in achieving crucial milestones of humanitarian actions [8]. Therefore, fast and undistorted situational awareness reports can contribute to the improvement of humanitarian responses in a sustainable way.

Evaluating microblogging sites such as Twitter has become essential to obtaining situational awareness in disaster situations [9–15]. As Phengsuwan et al. pointed out in their Survey about the use of social media data in disaster management, "Several publications have proposed the exploitation of social media data for disaster management, with Twitter being one of the most significant social media data sources used for disaster management. The temporal and spatial information extracted from Twitter is critical information for supporting decision-making in disaster management" [16].

The nature of these microblogging services enables their users to publish messages with a limited number of characters. However, this limited number of characters allows

people to distribute status updates quickly and easily. These are not exclusively important or relevant messages. Most status updates are highly irrelevant and concern social or private matters. However, recent research shows that this type of social media yields critical information in disaster situations because people publish, share, and report critical information [17–19].

Nonetheless, the many news stories during a disaster event, the accompanying social statements worldwide, and the many social messages of those who have something to say that is unrelated to the topic are a problem in the evaluation of this goldmine that is yet to be exploited [20]. However, the number of status updates that occur and are relevant during disasters holds great potential for helping in disaster management in the long term. Occurring events can be found and tracked during a disaster.

According to the definition of Hasan et al. [21], an event in the context of social media is "[. . .] an occurrence of interest in the real world which instigates a discussion on the event-associated topic by various users of social media, either soon after the occurrence or, sometimes, in anticipation of it".

By bundling and following these events, a news stream in the form of a story evolves, which is essential for decision-makers and first responders. This particularly timely stream of information can also help related work in a variety of other disaster-related fields and phases, such as the use of unmanned aerial vehicles (UAVs) for reconnaissance [22], vehicle routing and relief supply distribution [23], as well as volunteer assignment [24], to determine mission locations, mission requirements, and success rates. Furthermore, recordings of on-site information can help to understand complex situations in retrospect. Finally, the resulting records can help ensure adequate preparation and training in the future after a disaster. However, this valuable opportunity comes with its problems. Even after filtering irrelevant messages, personal communications, spam, and background noise, much relevant information remains. Although unmanageable to organize manually, these pieces of information need to be logically summarized. Of course, evaluating and summarizing this information by hand would require a large workforce and a considerable amount of time and resources. This would be unsustainable since we are dealing with vast amounts of data in this context.

To assess the following research questions, this work aims to provide a novel approach towards automated situational awareness reporting using microblogging data through event detection and summarization, by combining an event detection algorithm with different summarization libraries.

- Is the system able to process the data of the previous time window in the current time window?
- What percentage of the events in a dataset that is as realistic as possible is the system able to detect?
- Which of the three libraries for automated text summarization provides the best results?
- How could the results be improved?

This paper addresses the first step towards automated situational awareness reporting for disaster management. As previously described, Twitter provides one of the most significant sources for this purpose. Although we think our approach can be applied to other text-focused social media, they are outside the scope of this paper. The rest of this paper is ordered as follows. In Section 2, we outline related work, and in Section 3, we introduce the architecture of our proposed system. Sections 4 and 5 explain the design of our experiments and discuss the results. Finally, we summarize the work in Section 6 and give an outlook.

2. Related Works

To discuss related works, the following subdivision is made. The first part discusses works with related goals in the same disaster management/preparedness cycle phases. The second part discusses works that describe related components. Therefore, Section 2.1

presents different event detection approaches, while Section 2.2 presents different approaches for automated text summarization. To conclude, in Section 2.3 a summary is given.

Works with related goals can be divided into two major categories. The first category represents ontology-based approaches [25–27], while the second category concerns AI-based approaches [12,28–30].

Nguyen et al. [31] propose an interpretable classification-summarization framework that first classifies tweets into different disaster-related categories, and then, summarizes those tweets. Instead of focusing on performance measures, their proposed approach focuses on the decision-making process to give decision-makers an interpretable context. To achieve this, they employ a BERT-based multi-task learning approach and labeling. Accordingly, there are several limitations. In contrast to our approach, real-time capabilities are sacrificed to achieve higher accuracy. While we focus on delivering undistorted information fast to provide decision-makers with rapid situational awareness reports, their work mainly focuses on providing explanations or rationales. However, fast, undistorted reports can be used to better simulate the conditions of a disaster and to better train individuals in the appropriate behavior. It allows emergency personnel to practice reacting quickly until more detailed reports are available.

Mukherjee et al. [32] introduce an approach called MTLTS, a multi-task framework to obtain trustworthy summaries from crisis-related microblogs. Unlike other researchers in the field, they concentrate on a supervised approach. While they claim that their approach generalizes well across domains, outperforms the strongest baselines for the auxiliary verification/rumor detection task, and achieves high scores overall in the verified ratio of summary tweets, this approach requires time-consuming training and the labeling of datasets. The latter, itself, requires a time-consuming process, and thus, a high level of human resources.

Rudra et al. [33] propose a framework that classifies tweets to extract situational information, and then, summarizes this information. Their proposed approach relies on their observation that tweets in disaster situations often contain disaster-related and non-disaster-related information alongside numerical information such as casualties. While this approach is developed to meet rapid performance, it does not perform event detection, and thus, relies on a keyword search provided by the Twitter API.

2.1. Event Detection

To ensure targeted and sustainable automated awareness reporting in disaster management, the initial detection and tracking of events is an important starting point. It is necessary to find out what happened, and when and where it happened. The microblogging service Twitter has emerged as the most popular and widely used service to obtain information about real-world events [34].

The works of Hasan et al. [35] and Li et al. [36] provide a comprehensive overview of the field of event detection methods applied to streaming data from Twitter. For this purpose, representative social media event detection techniques are extracted and presented in a categorized overview according to common features. Subsequently, various aspects of the subtasks and challenges related to event detection are discussed. Our proposed approach uses these results as a basis and overview of the commonly used event detection algorithms. For this purpose, we use a complex, large, real-world dataset for evaluation that is publicly available. In contrast to the presented work, we focus on fast, automated situational awareness reporting. The approaches presented in the included works often have entirely different goals and are only tested on small artificial or non-public datasets.

Unankard et al. [34] propose the approach of Location-Sensitive Emerging Event Detection (LSED). This approach is used to detect emerging hotspot events from microblog messages. The goal of the message detection in this work is specifically to help governments or organizations prepare for and respond to unexpected events and disasters. The approach is also based on correlations between the user's location and the event's location obtained

from microblog messages. The researchers report that this approach provides better real-world event leverage results than traditional TF-IDF and hashtag approaches. Still, they also disclose that it is "[. . .] difficult to effectively and efficiently process a large number of noisy messages." [34]. This leads to significant variations and errors in the results, which needs further optimization in future work. Knowing the location and time associated with the user and event can be vital to increase the event detection accuracy and tracking performance. The distinction from this work is that it refers to the early detection of possible disaster scenarios. It is not intended to be used for news reporting or even report writing, but solely for government agencies to issue early warnings.

Osborne et al. [37] propose a system based on local sensitive hashing (LSH). As in our proposed approach, the main focus is on disaster detection. The LSH approach is not new and is derived from previous work [38]. Since the accuracy of LSH is not high enough by itself, a content classifier based on a passive-aggressive algorithm is developed and trained. Similar to our proposed system, a dictionary is also used, which is used to help detect relevant tweets in the disaster domain. However, in contrast to our work, a weakly supervised approach is used here to learn new words successively.

2.2. Text Summarization

Summarizing Twitter messages is complicated. While the brevity of microblogging-based messages can have advantages in event detection, it poses many problems in summarization [39]. Short messages with a high repetition factor pose significant problems for standard algorithms. Rudrapal et al. [39] state that standard summarization algorithms can be roughly divided into two camps: those based on summary content or those based on events. Content-based approaches can be divided into two categories. So-called abstract summarization tries to create entirely new text from given text. This new text may contain components that were not present in the original text. Furthermore, there is the extractive summary. Here, the features are weighted again, and essential parts of the text are extracted. Text parts with lower weights are omitted in this type of algorithm. Again, event category-based approaches can also be divided into two categories: generic and domain-specific approaches.

Li et al. [40] present two summarization approaches for events previously detected by the system. The first approach uses semantic types of event-related terms and ranks the messages based on the terms computed from the analysis of these semantic terms. The second approach uses graph convolutional networks and tweet relation graphs to detect hidden tweet features. Furthermore, Li et al. state that "There are many studies on text summarization, but only a few of them focus on social media" [40].

In this context, we propose, for our approach, the benchmarking of three standard libraries, namely SpayCi [41], Gensim [42], and NLTK [43], to evaluate the absolute baseline.

2.3. Summary

This section first discussed work with related objectives in the same disaster management/preparedness cycle phases. This was followed by a discussion of work describing related components. Each of the previous works shows a related component or approach. However, the works dealing with related components mainly show solutions at different disaster management/preparedness cycle phases or do not deal with disaster-based objectives. The approaches are either unsuitable in real-time, represent only an experiment, or are tested on small or non-public datasets.

On the other hand, works with related goals always have a deep neural network-based component in at least one part. As Li et al. state in their work, the disadvantage of deep neural networks is that they have velocity issues when used to discover new events from data streams in real time. This work focuses on delivering undistorted information fast to provide decision-makers with rapid situational awareness reports. Fast, undistorted reports can be used to better simulate the conditions of a disaster and to better train individuals in

the appropriate behavior. Our approach allows emergency personnel to practice reacting quickly until more detailed reports are available.

3. Proposed Approach

The system we propose in this work consists of three main components that work together as follows. The first component involves loading and sorting data from the Twitter API. The data loaded in this way are first cleaned of usernames, hashtags, and other so-called stop words via a cleaning process. This is followed by the operation of the event detection and clustering module. At this stage, the previously cleaned data are assessed and rated using a term frequency–inverse document frequency algorithm and checked against a Dictionary containing unique disaster-related words. Simultaneously every tweet is hashed using a sophisticated hash function. Afterward, highly ranked data are clustered. Finally, the data are summarized. The system uses the Twitter API to preprocess simple filtering tasks. To achieve this, only relevant keywords, hashtags, and locations are tracked and retrieved from the Twitter API. The data are always fetched within a time window for processing. In the course of this work, the data from such a time window are called a chunk. A chunk contains all tweets from a given time window. One of the goals of the modules of our system is to process the messages of the last time window in the given time window. The following section describes the components of the system, as depicted in Figure 2.
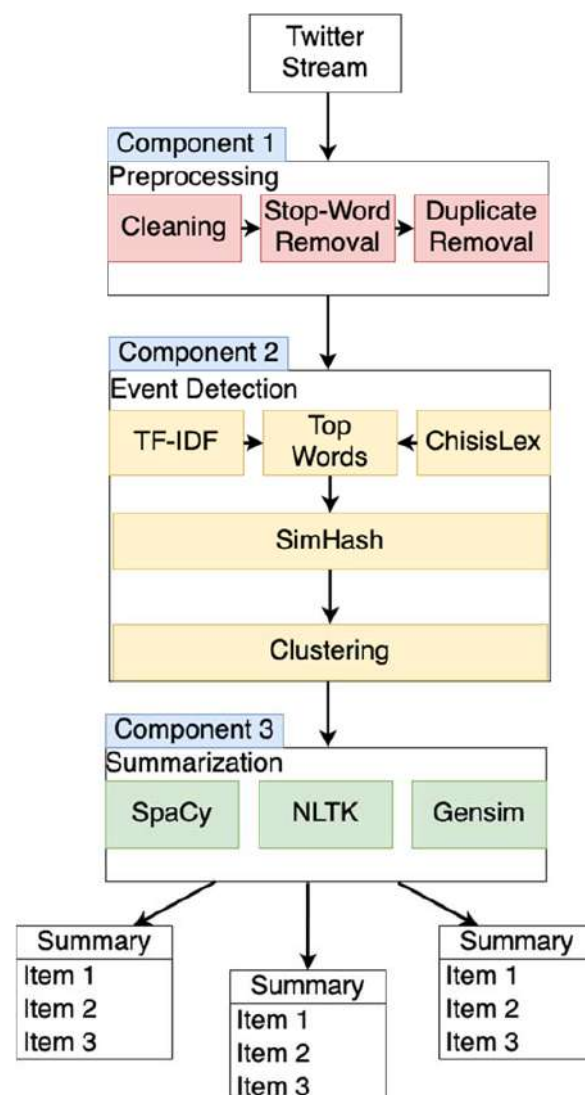

Figure 2. The proposed architecture.

3.1. Preprocessing

In this first processing step, data previously retrieved from the Twitter API are normalized, presorted, and cleaned. Since the text data to be processed are unfiltered raw data, it is to expected that they will contain a high proportion of noise and symbols that cannot be processed. Weblinks, hashtags, and so-called retweets, as well as usernames, which usually start with an @ symbol in Twitter data, are removed as a first step. Additionally, so-called stop words are filtered. Stop words are words whose processing is problematic because they occur particularly frequently but have no deeper meaning. Especially in emotional crisis situations, profanity can occur, so these words are also filtered.

Furthermore, only English-language messages are processed at this research stage, which is why all messages in other languages are filtered out. The filtering of the data is performed in three steps. In the first step, all messages are cleaned using the regular expressions described above, and capitalization is removed. After this, a specially created stop word dictionary is used to pre-filter the messages. The second step of filtering is performed using a stop word list from the SpaCy library [41]. Finally, duplicates are filtered by employing the Levenshtein distance [44]. The filtered data are then passed to the event detection stage.

3.2. Event Detection

The data are systematically evaluated for latent events in this second step. In the first step, the term frequency–inverse document frequency (tf-idf) [45] is calculated over the entire chunk. This is a statistical measure for assessing the relevance of terms in documents. To achieve this, first, the relative term frequency is calculated to prevent particularly frequently appearing words from being rated exceptionally highly. Secondly, the inverse document frequency is calculated. Here, the specificity of a term is measured. The matching occurrence of a rare term leads to a higher evaluation.

In our system, this first step creates a list of top words. This list of top words is now matched with CrisisLex [46]. CrisisLex contains a list of words that occur particularly frequently in disaster situations. Simultaneously, a hash is formed over each tweet in the current chunk using the SimHash algorithm [47]. Simhash is a so-called locality-sensitive hashing method (LSH). Unlike cryptographic hashes, where the slightest difference leads to significantly different hash values, with LSH, similar inputs lead to close hash values. This closeness can later be calculated easily and quickly as it only represents a one-dimensional distance. The advantages of the SimHash algorithm are a very efficient computation time, and that it takes so-called features in hashed elements into account so that unique semantic properties are not lost during the hashing process.

Relevant tweets are now filtered from the entire chunk using matching and scoring based on the previously selected top words. The previously created hash value is used to refine this set of relevant messages further and evaluate their significance.

The system now possesses an overview of the properties of the current chunk. At this stage, relevant tweets can be clustered according to their hash values. In our proposed system, we assume that important events generate many similar messages. Close hash values indicate similar messages. These similar messages are bundled into clusters. A larger cluster potentially represents more critical events.

These clusters are now passed on to the summary module. At this point, the previously determined term frequency–inverse document frequency (tf-idf) is reset. Only the list of hashes of tweets that made it into the final clusters is kept, in order to merge similar clusters into a time-divided story later.

3.3. Summarization

According to the problems described in Section 2. We use three well-known standard libraries at this point to assess a baseline. For this reason, our system relies on so-called Extractive Summarization. Here, features are weighted again, and essential text parts are extracted. An advantage of this approach is that the previously created clusters do not

have to be filtered. Due to the fuzziness of the extraction, duplicates are not significantly noticeable. Furthermore, former top words can be reused at this stage. For the summarization processes, the previously mentioned libraries SpayCi [41], Gensim [42], and NLTK [43] are used again. The verbal set of a cluster is now processed further. First, essential words are again determined and contextualized with the whole text. Afterward, in this way, weighted sentences are summarized. Since our approach aims to assess a baseline of fast summarization, no further filtering is conducted. Finally, clusters with similar hash values are merged in future time windows to obtain a chronological story.

4. Experiment

Testing automation for situational awareness reporting under realistic conditions is difficult. First, a reliable source of disaster events is needed. As McCreadie et al. [48] have pointed out, Wikipedia is not only well-known to contain newsworthy events, but is also a good source for disaster event information and a commonly used benchmark in related research [49–51].

The timeline of current events provided by Wikipedia was used as an evaluation basis to establish a measurement scale [52]. Wikipedia maintains a broad list of events, including those found in the dataset we used. As a disclaimer, we want to mention that this list of events and their real-world backgrounds are not verifiable by us. It remains unclear whether the list is complete. Due to the nature of Wikipedia, which allows anyone to edit the records stored there, Wikipedia could only be used as a rough basis.

As a next step, a baseline of actual events that occurred in September 2022 was needed to evaluate the system. In order to achieve this, the timeline mentioned above was assessed by hand to extract every event that occurred in that timeframe.

Many available datasets contain either irrelevant data or are limited to data that include disaster events exclusively and, therefore, are already filtered to a high percentage. To test the proposed system adequately, we used the Ukraine Conflict Twitter Dataset [53]. This dataset has been updated daily since 23 February 2022, and is expected to continue receiving daily updates until 30 January 2023 [53]. It was published under the CC0 Public Domain license and currently includes about 16 gigabytes of records on various hashtags. The publishers of the dataset specify the "Twitter-Verse" as the geospatial coverage. The dataset is highly relevant to the nature of the proposed system.

Nevertheless, it contains misinformation, spam, duplicates, advertisements, noise, and private conversations. To test the proposed system, we used all the data of September 2022 from the abovementioned dataset. As a final step, the proposed system was executed three times consecutively, and each time, another summarization library was tested.

The partial dataset we used contains about two million tweets. The proposed system approached the whole month of September in timeframes of 24 h. For each timeframe, the proposed system started detecting events at exactly 00:00 and stopped at 24:00. To detect multiple co-occurring events. The dataset was continuously evaluated in chunks. The chunk being evaluated always represented a period of precisely 15 min.

5. Discussion

In this work, we aimed to provide a novel approach towards automated situational awareness reporting using microblogging data through event detection and summarization. Therefore, we combined an event detection algorithm with different summarization libraries. Furthermore, we tried to answer the following questions:

- Is the system able to process the data of the previous time window in the current time window?
- What percentage of the events in a dataset that is as realistic as possible is the system able to detect?
- Which of the three libraries for automated text summarization provides the best results?
- How could the results be improved?

Regarding the first question, "Is the system able to process the data of the previous time window in the current time window?", this research shows that the proposed system can perform the task of processing the given data in the given time window very well. It even works faster than in the assigned time window. This result proves the practical feasibility of the proposed system and that it allows for the reliable processing of data in a given time frame.

Furthermore, we asked questions regarding what percentage of actual occurring events the proposed system can find in the given dataset, and which of the three libraries provided the best results. In order to be able to assess these questions, it was first necessary to determine the percentage of matches between events that actually happened and events detected by the summarization system on a given day. For this purpose, it was tested using a close-to-reality dataset. Table 1 shows the result of this comparison. The left column of the table shows the respective date of each day in the dataset analyzed by the system. Each day, we started detecting events at exactly 00:00 and stopped at 24:00 that day. In the second column is displayed the total number of tweets for that day. To the right of this column is the number of events that actually occurred on that day.

Table 1. Evaluation results for September 2022.

Date	Num. of Tweets	Num. of Events	SpaCy	Gensim	NLTK
Sep 01	48,262	2	2	0	0
Sep 02	53,668	4	4	2	3
Sep 03	40,732	2	1	0	2
Sep 04	40,151	2	3	1	2
Sep 05	44,695	0	2	0	1
Sep 06	61,619	2	1	0	1
Sep 07	46,496	1	3	1	1
Sep 08	47,830	0	0	0	0
Sep 09	70,401	1	2	0	1
Sep 10	64,953	6	9	3	4
Sep 11	67,526	8	7	6	8
Sep 12	62,944	2	5	3	2
Sep 13	63,866	2	3	2	2
Sep 14	75,392	2	2	1	2
Sep 15	67,230	1	1	0	2
Sep 16	63,125	1	2	0	1
Sep 17	49,592	1	2	0	2
Sep 18	42,426	1	1	0	1
Sep 19	49,332	2	3	0	2
Sep 20	57,409	3	5	1	3
Sep 21	104,313	4	6	3	4
Sep 22	78,507	0	2	0	0
Sep 23	77,368	3	3	2	3
Sep 24	63,866	0	0	0	0
Sep 25	52,998	0	1	0	0
Sep 26	60,625	1	1	1	1
Sep 27	68,866	4	5	3	3
Sep 28	80,527	3	3	2	2
Sep 29	66,984	3	4	3	2
Sep 30	69,930	3	3	1	1
	184,1633	64	134.375%	54.6875%	87.5%

To evaluate the events that actually occurred, we used the data from the "Timeline of the Russian Invasion of Ukraine 2022: Phase 3" of the Wikipedia current events website [52]. This timeline of the third phase of the Russian invasion of Ukraine in 2022 covers the period from 29 August 2022, when Ukrainian forces retook significant territory in counterattacks in southern and eastern Ukraine, to the present day.

On the positive side, Wikipedia briefly summarizes each event, which allows for benchmarking of the events covered by the proposed system. On the negative side, Wikipedia does not list temporal information about the events. Thus, we were not able to determine the delay with which events could be detected for the first time or the delay with which they appeared in the report.

To evaluate the events that actually occurred, we manually extracted the data from the above timeline for the month of September. Then, we ran the system proposed in this work with one of each of the three libraries to detect and summarize events from the dataset. We were then able to compare how many hits each of the libraries achieved to determine what percentage of the actual number of events matched the number of events detected in the summaries.

The number of actual events varied from 0 to 8. A total of 1,841,633 tweets could be counted on the 30 evaluated days of September. The number of actual events that were manually extracted is 64 in the whole month of September. The SpaCy library detected 134.375% events, the Gensim library detected 54.6875%, and the NLTK library detected 87.5%. This means the SpaCy library achieved performance far above the expected average compared the actual number of events that were manually extracted.

However, the results obtained do not always correspond with those that can be read in the Wikipedia timeline, and include lots of sidenotes and unimportant information. The Gensim library performed the worst. With only 54.6875%, it detected just over half of the minimum detectable war events. The NLTK library achieved a good average with a performance of 87.5%. The most interesting result was achieved with the SpaCy library. Here, we were able to outperform Wikipedia by 34.375%. We see that the total number of tweets in a day ranged from 40,151 (on 4 September 2022) to 104,313 (on 21 September 2022). This variance of 64,162 tweets comes from the fact that war events do not happen every day.

To answer the last question regarding how the results could be improved, we assessed every report that the system produced. Every report consisted of multiple textual summarizations consisting of several sentences, where each summarization visualized a time window of 15 min. These time windows included multiple events that occurred simultaneously. These co-occurring events then formed stories over multiple time windows.

We manually read through every report to assess ways to improve the results. We can easily see the progression and importance of stories on days such as 16 September, with 63,125 tweets. Stories, as we find them in the media, can be well tracked in the summarizations and number of messages on Twitter. Take the example of September 16 mentioned above, which, according to Wikipedia, holds one significant war event. The SpaCy and NLTK summaries report a powerful explosion in Luhansk in the late morning of September 16 in the time window between 11:45 and 12:00 UTC. The events are clearly traceable; the location, the number of victims, and their names and affiliations are traceable in the text. In addition, SpaCy's early afternoon summary in the time window between 14:15 and 16:30 UTC reports that there were deaths and injuries from shelling in Velyka Kostromka Oblast. The text gives the exact number of dead and injured victims, as well as the exact cause and origin of the event.

The detection rates perform weakly when there is a lot of propaganda, spam, and hate speech among the messages. This significantly distorts the results. The dataset used has a special significance regarding its connection to the conflict in Ukraine. Unlike natural disasters, here, there are several parties with different views. The view of the attacker and that of the defense. At this point, the best way to improve the summarization results is to improve the filtering for hate speech, spam, and sentiment.

6. Conclusions and Outlook

In this work, we provided a novel approach towards automated situational awareness reporting using microblogging data through event detection and summarization. We combined an event detection algorithm with different summarization libraries. We then

tested the proposed approach against data from the Russo-Ukrainian war to evaluate its real-time capabilities, and tried to answer multiple questions regarding its performance. In the process of this work, we evaluated our system using a dataset containing 16GB of data and several million tweets. The system was shown to be able to detect significant events in almost every case in the process. However, minor events and side notes caused problems and could not always be appropriately detected, or were largely over-detected at times. A large amount of spam, noise, and hate speech also proved problematic. The automated awareness reports allowed conclusions to be drawn about the real-world events the system detected. However, propaganda and hate speech could also be clearly recognized for several days or hours with low message rates. Future work on this topic could focus intensively on spam filtering, sentiment analysis, rumors, and fake news detection. A possible future proposal should also address social data such as likes, retweets, sentiment, and social scoring to prevent hate speech from getting out of hand. We believe our solution is quite suitable for the creation of quick initial reports and can contribute to sustainable and improved social awareness in disaster management.

References

1. Al-Madhari, A.F.; Keller, A.Z. Review Of Disaster Definitions. *Prehosp. Disaster Med.* **1997**, *12*, 17–21. [CrossRef] [PubMed]
2. Zibulewsky, J. Defining Disaster: The Emergency Department Perspective. *Bayl. Univ. Med. Cent. Proc.* **2001**, *14*, 144–149. [CrossRef]
3. Hanna Sawalha, I. Behavioural Response Patterns: An Investigation of the Early Stages of Major Incidents. *Foresight* **2018**, *20*, 337–352. [CrossRef]
4. Malilay, J.; Heumann, M.; Perrotta, D.; Wolkin, A.F.; Schnall, A.H.; Podgornik, M.N.; Cruz, M.A.; Horney, J.A.; Zane, D.; Roisman, R.; et al. The Role of Applied Epidemiology Methods in the Disaster Management Cycle. *Am. J. Public Health* **2014**, *104*, 2092–2102. [CrossRef] [PubMed]
5. Anderson, A.I.; Compton, D.; Mason, T. Managing in a Dangerous World—The National Incident Management System. *Eng. Manag. J.* **2004**, *16*, 3–9. [CrossRef]
6. Gujral, U.P.; Johnson, L.; Nielsen, J.; Vellanki, P.; Haw, J.S.; Davis, G.M.; Weber, M.B.; Pasquel, F.J. Preparedness Cycle to Address Transitions in Diabetes Care during the COVID-19 Pandemic and Future Outbreaks. *BMJ Open Diabetes Res. Care* **2020**, *8*, e001520. [CrossRef]
7. Ramsbottom, A.; O'Brien, E.; Ciotti, L.; Takacs, J. Enablers and Barriers to Community Engagement in Public Health Emergency Preparedness: A Literature Review. *J. Community Health* **2018**, *43*, 412–420. [CrossRef]
8. Arias-Aranda, D.; Molina, L.-M.; Stantchev, V. Integration of Internet of Things and Blockchain to Increase Humanitarian Aid Supply Chains Performance. *Dyna* **2021**, *96*, 653–658. [CrossRef]
9. Rudra, K.; Ganguly, N.; Goyal, P.; Ghosh, S. Extracting and Summarizing Situational Information from the Twitter Social Media during Disasters. *ACM Trans. Web.* **2018**, *12*, 1–35. [CrossRef]
10. Lamsal, R.; Kumar, T.V.V. Twitter-Based Disaster Response Using Recurrent Nets. *Int. J. Sociotechnol. Knowl. Dev.* **2021**, *13*, 133–150. [CrossRef]
11. Mukhtiar, W.; Rizwan, W.; Habib, A.; Afridi, Y.S.; Hasan, L.; Ahmad, K. Relevance Classification of Flood-Related Twitter Posts via Multiple Transformers. *arXiv* **2022**, arXiv:2301.00320.
12. Karimiziarani, M.; Jafarzadegan, K.; Abbaszadeh, P.; Shao, W.; Moradkhani, H. Hazard Risk Awareness and Disaster Management: Extracting the Information Content of Twitter Data. *Sustain. Cities Soc.* **2022**, *77*, 103577. [CrossRef]

13. Karimiziarani, M.; Moradkhani, H. Social Response and Disaster Management: Insights from Twitter Data Assimilation on Hurricane Ian. *SSRN Electron. J.* **2022**. [CrossRef]

14. Lamsal, R.; Harwood, A.; Read, M.R. Socially Enhanced Situation Awareness from Microblogs Using Artificial Intelligence: A Survey. *ACM Comput. Surv.* **2022**, *55*, 1–38. [CrossRef]

15. Takahashi, B.; Tandoc, E.C.; Carmichael, C. Communicating on Twitter during a Disaster: An Analysis of Tweets during Typhoon Haiyan in the Philippines. *Comput. Hum. Behav.* **2015**, *50*, 392–398. [CrossRef]

16. Phengsuwan, J.; Shah, T.; Thekkummal, N.B.; Wen, Z.; Sun, R.; Pullarkatt, D.; Thirugnanam, H.; Ramesh, M.V.; Morgan, G.; James, P.; et al. Use of Social Media Data in Disaster Management: A Survey. *Future Internet* **2021**, *13*, 46. [CrossRef]

17. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26 April 2010; ACM: New York, NY, USA, 2010; pp. 851–860.

18. Shan, S.; Zhao, F.; Wei, Y.; Liu, M. Disaster Management 2.0: A Real-Time Disaster Damage Assessment Model Based on Mobile Social Media Data—A Case Study of Weibo (Chinese Twitter). *Saf. Sci.* **2019**, *115*, 393–413. [CrossRef]

19. Muzamil, S.A.H.B.S.; Zainun, N.Y.; Ajman, N.N.; Sulaiman, N.; Khahro, S.H.; Rohani, M.M.; Mohd, S.M.B.; Ahmad, H. Proposed Framework for the Flood Disaster Management Cycle in Malaysia. *Sustainability* **2022**, *14*, 4088. [CrossRef]

20. Pastor-Galindo, J.; Nespoli, P.; Gomez Marmol, F.; Martinez Perez, G. The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. *IEEE Access* **2020**, *8*, 10282–10304. [CrossRef]

21. Hasan, M.; Orgun, M.A.; Schwitter, R. A Survey on Real-Time Event Detection from the Twitter Data Stream. *J. Inf. Sci.* **2018**, *44*, 443–463. [CrossRef]

22. Yu, X.; Li, C.; Yen, G.G. A Knee-Guided Differential Evolution Algorithm for Unmanned Aerial Vehicle Path Planning in Disaster Management. *Appl. Soft. Comput.* **2021**, *98*, 106857. [CrossRef]

23. Rabiei, P.; Arias-Aranda, D. Introducing a Novel Multi-Objective Optimization Model for Vehicle Routing and Relief Supply Distribution in Post-Disaster Phase: Combining Fuzzy Inference Systems with NSGA-II and NRGA. In Proceedings of the 6th International Conference on Transportation Information and Safety: New Infrastructure Construction for Better Transportation, ICTIS 2021, Wuhan, China, 22–24 October 2021; pp. 1226–1243. [CrossRef]

24. Rabiei, P.; Arias-Aranda, D.; Stantchev, V. Introducing a Novel Multi-Objective Optimization Model for Volunteer Assignment in the Post-Disaster Phase: Combining Fuzzy Inference Systems with NSGA-II and NRGA. *Expert Syst. Appl.* **2023**, *226*, 120142. [CrossRef]

25. Garg, P.K.; Chakraborty, R.; Dandapat, S.K. OntoDSumm: Ontology Based Tweet Summarization for Disaster Events. *arXiv* **2022**, arXiv:cs.SI/2201.06545.

26. Garg, P.K.; Chakraborty, R.; Dandapat, S.K. EnDSUM: Entropy and Diversity Based Disaster Tweet Summarization. *arXiv* **2022**, arXiv:2203.01188.

27. Wu, K.; Li, L.; Li, J.; Li, T. Ontology-Enriched Multi-Document Summarization in Disaster Management Using Submodular Function. *Inf. Sci.* **2013**, *224*, 118–129. [CrossRef]

28. Banerjee, S.; Mukherjee, S.; Bandyopadhyay, S.; Pakray, P. An Extract-Then-Abstract Based Method to Generate Disaster-News Headlines Using a DNN Extractor Followed by a Transformer Abstractor. *Inf. Process. Manag.* **2023**, *60*, 103291. [CrossRef]

29. Saini, N.; Saha, S.; Bhattacharyya, P. Microblog Summarization Using Self-Adaptive Multi-Objective Binary Differential Evolution. *Appl. Intell.* **2022**, *52*, 1686–1702. [CrossRef]

30. Vitiugin, F.; Castillo, C. Cross-Lingual Query-Based Summarization of Crisis-Related Social Media: An Abstractive Approach Using Transformers. In Proceedings of the HT 2022: 33rd ACM Conference on Hypertext and Social Media—Co-located with ACM WebSci 2022 and ACM UMAP 2022, Barcelona, Spain, 1 July–28 June 2022; pp. 21–31. [CrossRef]

31. Nguyen, T.H.; Rudra, K. Towards an Interpretable Approach to Classify and Summarize Crisis Events from Microblogs. In Proceedings of the ACM Web Conference 2022, Virtual, 29 April 2022. [CrossRef]

32. Mukherjee, R.; Vishnu, U.; Peruri, H.C.; Bhattacharya, S.; Rudra, K.; Goyal, P.; Ganguly, N. MTLTS: A Multi-Task Framework to Obtain Trustworthy Summaries from Crisis-Related Microblogs. In Proceedings of the 15th ACM International Conference on Web Search and Data Mining (WSDM), Tempe, AZ, USA, 21–25 February 2022; pp. 755–763. [CrossRef]

33. Rudra, K.; Ghosh, S.; Ganguly, N.; Goyal, P.; Ghosh, S. Extracting Situational Information from Microblogs during Disaster Events: A Classification-Summarization Approach. In Proceedings of the International Conference on Information and Knowledge Management, Proceedings, Melbourne, Australia, 19–23 October 2015; pp. 583–592. [CrossRef]

34. Unankard, S.; Nadee, W. Sub-Events Tracking from Social Network Based on the Relationships between Topics. In Proceedings of the 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Pattaya, Thailand, 11–14 March 2020; pp. 1–6.

35. Hasan, M.; Orgun, M.A.; Schwitter, R. Real-Time Event Detection from the Twitter Data Stream Using the TwitterNews+ Framework. *Inf. Process. Manag.* **2019**, *56*, 1146–1165. [CrossRef]

36. Li, Q.; Chao, Y.; Li, D.; Lu, Y.; Zhang, C. Event Detection from Social Media Stream: Methods, Datasets and Opportunities. In Proceedings of the 2022 IEEE International Conference on Big Data, Osaka, Japan, 17–20 December 2022; pp. 3509–3516. [CrossRef]

37. Osborne, M.; Moran, S.; McCreadie, R.; von Lunen, A.; Sykora, M.; Cano, E.; Ireson, N.; Macdonald, C.; Ounis, I.; He, Y.; et al. Real-Time Detection, Tracking, and Monitoring of Automatically Discovered Events in Social Media. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; Association for Computational Linguistics, Stroudsburg, PA, USA, 23–24 June 2014; pp. 37–42.

38. Petrovic, S.; Osborne, M.; Lavrenko, V. Streaming First Story Detection with Application to Twitter. In Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10), Los Angeles, CA, USA, 2–4 June 2010; Association for Computational Linguistics: Seattle, WA, USA, 2010; pp. 181–189.

39. Rudrapal, D.; Das, A.; Bhattacharya, B. A Survey on Automatic Twitter Event Summarization. *J. Inf. Process. Syst.* **2018**, *14*, 79–100. [CrossRef]

40. Li, Q.; Zhang, Q. Twitter Event Summarization by Exploiting Semantic Terms and Graph Network. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15347–15354. [CrossRef]

41. Montani, I.; Honnibal, M.; Honnibal, M.; van Landeghem, S.; Boyd, A.; Peters, H.; McCann, P.O.; Samsonov, M.; Geovedi, J.; O'Regan, J.; et al. Explosion/SpaCy: V3.1.6: Workaround for Click/Typer Issues 2022. Available online: https://zenodo.org/record/1212303 (accessed on 3 April 2023).

42. Rygl, J.; Pomikálek, J.; Řehůřek, R.; Růžička, M.; Novotný, V.; Sojka, P. Semantic Vector Encoding and Similarity Search Using Fulltext Search Engines. In Proceedings of the 2nd Workshop on Representation Learning for NLP, Vancouver, BC, Canada, 3 August 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; Volume 3, pp. 81–90.

43. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2009; Volume 6.

44. Navarro, G. A Guided Tour to Approximate String Matching. *ACM Comput. Surv.* **2001**, *33*, 31–88. [CrossRef]

45. Rajaraman, A.; Ullman, J.D. Data Mining. In *Mining of Massive Datasets*; Cambridge University Press: Cambridge, UK, 2011; pp. 1–17.

46. Olteanu, A.; Castillo, C.; Diaz, F.; Vieweg, S. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; Volume 8, pp. 376–385. [CrossRef]

47. Manku, G.S.; Jain, A.; das Sarma, A. Detecting Near-Duplicates for Web Crawling. In Proceedings of the 16th international conference on World Wide Web, New York, NY, USA, 12 May 2007; ACM: New York, NY, USA, 2007; pp. 141–150.

48. McCreadie, R.; Macdonald, C.; Ounis, I. Insights on the Horizons of News Search. Available online: http://terrierteam.dcs.gla.ac.uk/publications/richardmSSM2010.pdf (accessed on 25 April 2023).

49. Li, C.; Sun, A.; Datta, A. Twevent: Segment-Based Event Detection from Tweets. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, HI, USA, 29 October–2 November 2012; pp. 155–164. [CrossRef]

50. Stilo, G.; Velardi, P. Efficient Temporal Mining of Micro-Blog Texts and Its Application to Event Discovery. *Data Min. Knowl. Discov.* **2016**, *30*, 372–402. [CrossRef]

51. Osborne, M.; Petrovi'cpetrovi'c, S.; Mccreadie, R.; Macdonald, C.; Ounis, I. Bieber No More: First Story Detection Using Twitter and Wikipedia. In Proceedings of the TAIA'12, Portland, OR, USA, 12–16 August 2012; pp. 16–76.

52. Timeline of the 2022 Russian Invasion of Ukraine: Phase 3—Wikipedia. Available online: https://en.wikipedia.org/wiki/Timeline_of_the_2022_Russian_invasion_of_Ukraine:_phase_3 (accessed on 1 February 2023).

53. BwandoWando Ukraine Conflict Twitter Dataset 2023. Available online: https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows/versions/363 (accessed on 13 February 2023).