

Article

DisKnow: A Social-Driven Disaster Support Knowledge Extraction System

João Boné ^{1,2,*} , Mariana Dias ^{1,2} , João C. Ferreira ^{1,2,3}  and Ricardo Ribeiro ^{1,4} ¹ Instituto Universitário de Lisboa (ISCTE-IUL), 1649-026 Lisbon, Portugal;

Mariana_Rebello@iscte-iul.pt (M.D.); jcafa@iscte-iul.pt (J.C.F.); ricardo.ribeiro@iscte-iul.pt (R.R.)

² INOV INESC Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal³ Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL), 1649-026 Lisbon, Portugal⁴ INESC-ID Lisboa, 1000-029 Lisbon, Portugal

* Correspondence: Joao_Bone@iscte-iul.pt

Received: 9 July 2020; Accepted: 31 August 2020; Published: 2 September 2020



Abstract: This research is aimed at creating and presenting DisKnow, a data extraction system with the capability of filtering and abstracting tweets, to improve community resilience and decision-making in disaster scenarios. Nowadays most people act as human sensors, exposing detailed information regarding occurring disasters, in social media. Through a pipeline of natural language processing (NLP) tools for text processing, convolutional neural networks (CNNs) for classifying and extracting disasters, and knowledge graphs (KG) for presenting connected insights, it is possible to generate real-time visual information about such disasters and affected stakeholders, to better the crisis management process, by disseminating such information to both relevant authorities and population alike. DisKnow has proved to be on par with the state-of-the-art Disaster Extraction systems, and it contributes with a way to easily manage and present such happenings.

Keywords: disaster management; natural language processing; information extraction; crowdsourcing; automatic knowledge base construction; knowledge graphs

1. Introduction

Throughout the world, earthquakes, floods, fires, and other natural hazards cause tens of thousands of deaths and billions of euros in economic losses each year [1]. These numbers only rise if we take into consideration other risk-inducing disasters, like terrorism or even pandemics, as we have seen in 2020 with the spread of COVID-19. Advances in information technology and communications, combined with the introduction and upsurge of social media apps, have created a new world of emergency and disaster management services by allowing impacted people to produce real-time georeferenced information on critical incidents [2]. In social networks, such as Twitter, Facebook, Instagram, and others, humans volunteer plentiful and free information about their surrounding environments. This information can be leveraged as a category of sensor networks. Social networks have been playing a huge role in disaster management. However, although they are used to communicate emergency information and urgent requests between authorities and those affected by disasters [3], this flow of information is still not without its flaws. A considerable slice of relevant information is still being posted on these media, with no connection to the real stakeholders of said disasters, and thus, very limited impact. Handling this information has the potential to extract crowd-approved knowledge and use that knowledge to increase mass situational awareness in disaster scenarios, but it does not come without its challenges. One of the obstacles of being able to use this information is filtering it, considering social media posts tend to vary widely in both their subjects and utility, ranging from off-topic to relevant disaster-related information.

Gathering actionable information can improve disaster response and significantly reduce both property and human casualties [4]. Today, due to the mobile devices people constantly have in their possession, and sensing information they are continually posting, we can see social networks not only as a means to socialize but also as vast networks of boundless information, waiting to be used in a vast range of fields, including society-threatening scenarios. These apps allow people to register surrounding events and essential related information, such as where they started, who is being affected, and what kind of threats should the population and first responders expect to face. This activity could provide useful information, in the form of alerts, towards competent authorities and help them act in these crises. An example of these works is [5], where the authors applied social data extraction to flood inundation, to gain in-depth insight into the situation. Out of these social networks, Twitter has been widely used as a source of information. Around 140 million people use Twitter, with some redundancy, as users reuse or modify posted tweets. Duplication drives the information to spread through the service, and previous research has indicated that 29% to 47% of crisis-related tweets are re-tweets, and that 60% to 75% of all tweets are near duplicates of at least one other tweet [6].

Twitter has been widely used as a source of information in several works. From 2010 on, it has been seeing a growing usage as a source of useful information for disaster support management systems [7], and in 2013, Williams, Terras, and Warwick pointed it out as having the most potential for data extraction, when compared to other social networks [8]. Studies show that in disaster scenarios, victims, responders, and volunteers increasingly use Twitter to share and access situation-awareness reports, and to provide help during these crises. Relevant Examples of its application are: In 2010, Li and Rao have shown that information about an earthquake in China had initially appeared on Twitter, three minutes before the official authorities announced it [9]. In Japan 2011 the same happened with information about an earthquake, reported in the work of Tatsubori [10]. Another case happened in Paris 2015 terrorist attacks, where eyewitnesses sent tweets to notify others about what was happening [11]. These examples are not, however, unique. Twitter has been widely used in several types of disasters, such as wildfires and floods [12–15], terrorist attacks [16], humanitarian aid [17], and other accidents [18]. Some studies even show that Twitter can provide more accurate information than the one obtained through traditional data-collection methods [19,20]. Despite its perks, Twitter's usage as an information source provides interesting challenges, due to very high message volumes, lack of essential metadata, missing generalized georeferenced information, and being in constant change.

Information extraction techniques are, nowadays, one of the most relevant topics in disaster management, and many researchers have examined these approaches in order to establish pre and post-disaster response programs [21]. Its potential, however, does not come without challenges. There is a massive scarcity of formalized data, such as domain-specific annotated data sets, as well as severe problems regarding the use of social media data, such as the lack of methods to filter trustful information and to exploit noisy social media data streams of high volume, in near real-time [22]. Although geotagging is supported on Twitter, only approximately 1% of published tweets are geotagged [23]. The short nature of tweets, having a maximum size of 280 characters, makes applying the current state-of-the-art machine learning algorithms—that build latent topic models—quite difficult.

Our major approach is to handle this useful information in the context of disaster scenarios to provide knowledge to local operation authorities. This, however, requires a way to efficiently store and connect data, which can be addressed by database technologies, such as graphs. Graphs are extremely capable when extracting insight from data is a top priority, as they are mainly made of nodes and connections, and translating that to the real world, it means we can have access to a logical map of entities and their relationships. In addition to their structure, graphs are also very efficient performance-wise, as in contrast to traditional databases, their performance stays nearly constant even as their data grows. Their structure also allows for high flexibility when expanding to incorporate new data.

Taking into account the work performed on disaster monitoring through social media, and these mentioned techniques, we identified a gap in the way knowledge is being delivered to stakeholders, as most systems use social media for extracting data that only later is reported as a whole. DisKnow aims to fill this gap. It improves on such systems by not only treating each piece of data as possible new information, but also automatically connecting the knowledge it extracts. This is made possible through the usage of state-of-the-art classifiers for filtering tweets that are related to disasters, and graphs that allow for natural and automated representations of both pieces of individual knowledge, and the connections between them. With these technologies combined, our system is continuously learning from tweets, and therefore expanding its knowledge. This allows DisKnow to present its contributions to disaster management stakeholders in a way that they can employ real-time updated knowledge to make the best possible decisions, in scenarios where every second matters.

In this article, we will be presenting the current state-of-the-art for both disaster monitoring through social media and Disaster Extraction techniques. We will then present our system, whose pipeline will be divided and detailed. Lastly, we will provide both a comparison with other Disaster Extraction techniques, and a simulated scenario, to set forth DisKnow's contributions in what would be a real emergency.

2. Related Work

2.1. Social Media for Disaster Management

A disaster can be defined as a source of danger, and its consequences can adversely affect the world in different ways. Digital social networks have both emerged and evolved greatly in the 2000s, and the large amount of information shared can be of great help in disaster management and in other critical situations [24]. Major systematization work on Twitter messages was carried out by Steiger et al. in 2015 [7], where they highlight that 46% of papers were about event detection. Furthermore, 71% of the computer science reviewed studies have no specific application context, and are mainly focused on creating system architectures. This led the authors to the conclusion that disaster management is an opportunity to apply such knowledge to, and could significantly benefit from it, in order to strengthen situation awareness and improve emergency response. It shows that event detection on Twitter is a popular domain and plays an essential role in Twitter research. This is done based on a keyword search directly over Twitter's Application Programming Interface (API). It is usual to use control vocabularies to identify specific predefined events to identify what tweets will be captured and retrieved through the API. Examples of these approaches are the keyword-based filtering to identify predefined events like forest fires [25], earthquakes [26], and floods [27]. Although keyword-based data extraction might be useful, its standalone usage has several associated problems, like noise and bias. There is also a high rate of false results associated with these methods, such as detecting that there is going to be a fire in New York from the tweet "Arcade fire are going to be playing in NYC". An example of these problems is the work of Bosley et al. [28], where they identify that only 25% of the tweets generated by their initial keyword search were related to their topic. In 2013, through a more detailed analysis of the #TFF hashtag for the Tobacco Free Florida media campaign, the authors found that out of 3104 selected tweets, only 1% were actually related to it [29]. Therefore, a need for methods that can go beyond keyword filtering has been, and still is, one of the main focuses of Twitter research efforts to date.

The research literature on social networking and social media applied to disasters and crisis situations, is still quite limited. Some researchers, however, have examined these approaches in order to establish pre and post-disaster response programs [21]. One example would be Sakaki et al. [30] where Twitter was used to construct an earthquake reporting system in Japan, with an extremely positive result of 96% of said earthquakes over Richter' magnitude 3 being detected, directly through monitoring tweets. In another example use-case, demonstrating that collecting trustworthy information is crucial for disaster management, Uddin et al. [31] used Apollo [32] to collect and evaluate tens of thousands of tweets about New York City's gas availability during and shortly after 2012's Hurricane

Sandy to determine the accuracy of individual tweets, while considering unknown source reliability and uncertain provenance. Their results have shown that 90% of the tweets believed by Apollo were true, which is very relevant given that less than half of the tweets were actually true. CrisisTracker [33] has also proven to be a great example of using social media in disaster management, and probably one of the most successful ones. Although it has been used to monitor events like Fukushima's nuclear disaster, its largest field trial dealt with the 2012 civil war in Syria. During the 8-day study, CrisisTracker processed an average of 446,000 tweets a day and managed to successfully reduce the information to consumable stories, being successful in raising the situational awareness of such disaster areas. In practice, it took about 30 min after an isolated incident occurred before CrisisTracker could reduce the overload of social media information to a consumable story.

2.2. Disaster Classification

Disaster classification can be seen as a very specific application of event detection. The latter is a task whose purpose is to identify related stories from a continuous stream of textual data, and considers an event as being something that occurs at a certain time in a certain location. Farzindar and Wael [34] have surveyed the methods used in event detection from social media, and have reported the usage of various machine learning techniques, such as classifiers, clustering, and language models. Recently, however, deep learning has emerged on most natural language processing (NLP) tasks, and event detection was not an exception. Several works have introduced the application of deep learning techniques to address event detection at a sentence level. This has been done by first identifying event triggers—which, in a sentence can be for example verbs or derivational nominalizations—and then classifying them into specific types.

Deep learning models have proven to be more tolerant to domain and context variations, when compared with keyword-based models, one of the reasons being that the usage of word embeddings allows for a more generic and richer representation of words and their relationship [35]. There is not, however, much work in regard to disaster classification in tweets using machine learning, as this is generally done using keyword-based approaches, and labeled tweets for this task are scarce. Nevertheless, a few works have been published in recent years, showing promising results.

In 2017, Burel et al. [36] contributed to the field with the use of a dual convolutional neural network (Dual-CNN), to both see if tweets were, or not, related to a disaster, and which type of disaster they were related to. This Dual-CNN has an additional semantic representation layer which vectorizes information regarding the tweet's detected named entities. The authors have reported very little difference between applying Dual-CNN and a single CNN, for this specific task. Despite the semantic knowledge used by the first, its results have not shifted, and in some tasks it has even performed worse. More recently, in 2019, Sit et al. [37] applied long short-term memory (LSTM) neural networks to the disaster relatedness task, while comparing its results with linear machine learning algorithms, like support vector machines (SVM) and logistic regressions. Although the results have not improved over Burel et al. [36], the deep learning approaches have clearly improved on the results achieved by both SVM and logistic regression models. Additionally in 2019, Alam et al. [38] used several models for the disaster type classification task, ranging from traditional machine learning models, such as naïve Bayes classifier, random forests, and SVM, to a deep learning CNN. Although the CNN has outperformed most models, the SVM has shown results on par with it, coming up to an impressive 93% accuracy.

3. Proposed Approach

DisKnow was jointly developed at Iscte—Instituto Universitário de Lisboa and INOV INESC Inovação. It is a system that retrieves tweets in real-time, filtering noisy data to better its performance and reliability. It then processes these tweets, using a language independent approach, and translates them to English, subsequently applying deep learning techniques to classify whether a tweet refers to real disasters or not. For relevant tweets, the extracted information is connected and related as different

types of nodes, that are represented in the knowledge base, for a noise-free concise representation available to both population and decision makers. This knowledge is constantly being expanded and updated, thus achieving a desired level of adaptability to real threat scenarios, which are also continuously evolving, and require swift and informed responses.

The knowledge base consists of a dynamic knowledge graph [39]. This allows us to structure entities as a knowledge graph in which new nodes and relationships are constantly being added as new context is detected, to expand and upgrade its knowledge in real-time. This knowledge is disaster-centered, as disaster occurrences are represented by specific nodes which then connect to locations, dates, people, organizations, etc. In addition to having several node types, such as the disasters and locations, each node also has several attributes. This allows for flexibility when dealing with various types of data associated with entities. As for its management, our knowledge base uses the Neo4j (<https://neo4j.com/>) graph database management system. The interactions with it are done through in-house developed custom methods, using a mixture of the Cypher query language, and the Py2neo (<https://py2neo.org/v4/>) toolkit. An overview on DisKnow's architecture is presented in Figure 1.

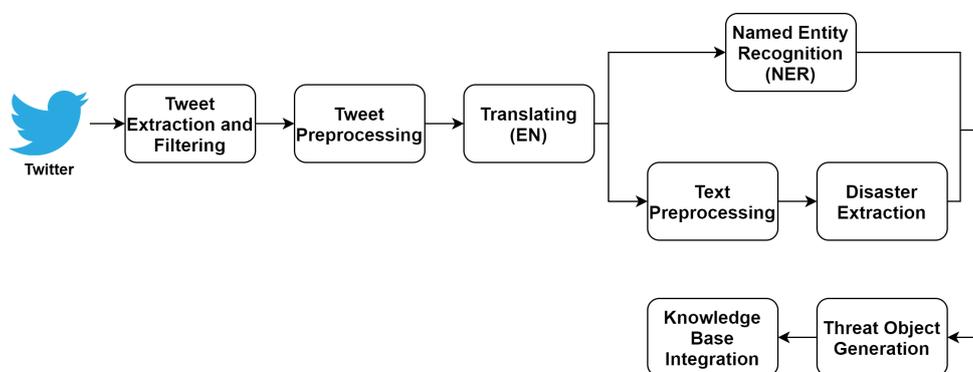


Figure 1. DisKnow's process pipeline.

1. **Tweet Extraction and Filtering:** The first step of our pipeline is the extraction of tweets and their related information, from the social media Twitter. To avoid overloading the system, these tweets are gathered through periodic API calls with several filters on them. These filters include the tweets' language and geolocation, which is done by defining a center and a radius, and calculating, for each tweet, the geodesic distance between the tweet's location and the defined center. Both the tweet's location and defined center are geocoded using data from OpenStreetMap [40], allowing for the computation of the mentioned geodesic distance. Relevant tweet information, such as each tweet's posting datetime and full text is also saved for further use.
2. **Tweet Preprocessing:** This step focuses on cleaning the textual content that is specific to tweets. As we have previously mentioned, one of the challenges of using social media information, is the huge amount of noisy data it pertains. This processing pipeline uses hand-made techniques, such as regex rules, that aim to take care of specific tweet noise, such as retweeting information, hyperlinks, user mentions, and hashtags. Firstly, any information regarding retweeting, hyperlinks, and user mentions is removed, and secondly hashtag symbols are removed and their content is split by upper-case starting words.
3. **Translating:** Translation is a natural step here, as DisKnow is being implemented as a component of a European project (grant agreement n833088) which aims to be put into effect in several countries. With this in mind, and to be able to use state-of-the-art models, which mostly exist for English, the input language is firstly detected, and then the cleaned text is translated to English. This is done using Google's Neural Machine Translation (GNMT) [41], which consists in a deep LSTM using both residual and attention connections. This model has achieved a BLEU [42] score of 38.95 in the commonly evaluated en-fr translation task, with a vocabulary size of around

- 32,000 words. It is also reported to reduce translation errors by more than 60%, when compared to its competition, the Phrase-based Machine Translation (PBMT) models [41].
4. **Text Preprocessing:** Preprocessing is an essential step for natural language algorithms. Although activities like Named Entity Recognition (NER) benefit from inputting entire texts with punctuation and the original formatting, most algorithms benefit from further processed text. This step focuses on additionally preparing tweets for Disaster Extraction. It starts by tokenizing the tweets using Natural Language Toolkit (<https://www.nltk.org/>) (NLTK) algorithms, and proceeds to using regex rules to lowercase all the tokenized words. Lastly, English stop words are removed, as to lessen the noise inputted to the following steps. Although lemmatization or stemming are also frequently used, in our case we found that our Disaster Extraction algorithm slightly benefited from not using them.
 5. **Named Entity Recognition:** NER is used as a tool to extract entities from the aforementioned tweets. These entities can be person names, companies, geographical locations, times, dates, etc. Their role is to complement a detected threat, in order to give users more insight. For this purpose, our system makes use of Spacy (<https://spacy.io/>) NER pipeline. As this pipeline is open, we are able to add new layers to it, to better embrace specific disaster detection characteristics, such as detailed locations and local businesses.
 6. **Disaster Extraction:** The next step is, without any doubt, one of the most critical steps of our system. It consists of two sequential tasks: Firstly, classifying tweet relatedness to disasters, that is, classifying if a tweet's content is referring to a disaster event. Secondly, classifying the disaster type a tweet is referring to. To accomplish both of these tasks we have implemented two distinct convolutional neural networks (CNNs). The first CNN takes care of the first task, and if it classifies the tweet as being related to disasters, the second CNN then classifies which disaster it is, out of a predefined set of disasters it has been trained with, which are fire, flood, earthquake, explosion, and none. This last category exists as a double-check, seeing that the first CNN can detect disaster events that are not modeled by the second CNN. The second neural network outputs a list of confidence levels regarding all disaster classes.

In the event of the second CNN classifying, a tweet with close confidence levels, $\text{Conf1} - \text{Conf2} < \text{ConfMargin}$, where Conf1 refers to the highest confidence the CNN outputs, Conf2 refers to the second highest confidence, and ConfMargin refers to the minimum margin set in the model, a weighted check is triggered. This check consists of adding to each of these confidence levels the confidence of a lexicon-based model. This model involves the preprocessing of the message with the use of tokenization and stemming techniques, and then comparing its result with predefined lexicons for each type of disaster. The results, LexCount , then go through a min-max normalization:

$$\text{LexCount}' = \frac{\text{LexCount} - \min(\text{LexCount})}{\max(\text{LexCount}) - \min(\text{LexCount})} \quad (1)$$

Lastly, a weighted ponderation of these values is added to the output confidence of the neural network, allowing for a better classification of the disaster a tweet refers to. This component's pipeline is also represented in Figure 2.

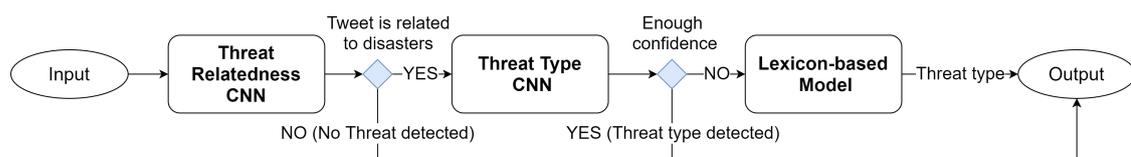


Figure 2. Disaster extraction component pipeline.

The data used to train both networks were obtained as a subset of two CrisisLex [43] disaster datasets:

1. CrisisLexT6 (<https://crisislex.org/data-collections.html#CrisisLexT6>);
2. CrisisLexT26 (<https://crisislex.org/data-collections.html#CrisisLexT26>).

Although they do not have the same data, as a set, they have English tweets across more than 30 large disasters which took place between 2012 and 2013. These include, for the sake of example, the 2012 Colorado Wildfires and the famous 2013 Boston Marathon Bombings. On top of this, to better comprehend the domain where they are being applied, our classification models use the 200d Glove [44] Twitter embeddings (<https://nlp.stanford.edu/projects/glove/>). These pretrained embeddings have explicitly been trained from more than 2 billion tweets, with about 30 billion tokens, and more than 1.2 million unique vocabulary, where each word is represented as a 200-dimensional vector. The usage of these embeddings is a huge advantage in tweet information extraction, as most pretrained embeddings are either trained with Wikipedia data, web crawling data, or news.

As shown in Figure 3, our CNNs are fed integer encoded tweets, which have formerly been padded to 26 words to guarantee homogeneity. The embedding layer is responsible for converting these integer vectors into 26 by 200 matrices, where 200 represents the vectorial depiction of each tweet's word, according to the embeddings we have used. Following the embedding layer, the CNNs have a convolutional layer with a Rectified Linear Unit (ReLU) activation where 128 feature detectors slide through the data in one dimension—hence being a 1D convolutional layer—for 5 steps. The global max pooling layer follows, which is responsible for the dimensionality reduction of the data. The network then has a set of three dense layers—that also use ReLU activation—with decreasing number of nodes and dropout rates, which help us prevent overfitting our data. This is extremely important because we are dealing with data relating to particular scenarios/disasters, so we want to keep our models from overfitting specific data, which can be done, to some extent, by ignoring random neurons during the training phase. The network's output is then yielded through a dense layer, which varies from having one node and a sigmoid activation in the relatedness classification CNN and having five nodes and a softmax activation in the type classification CNN.

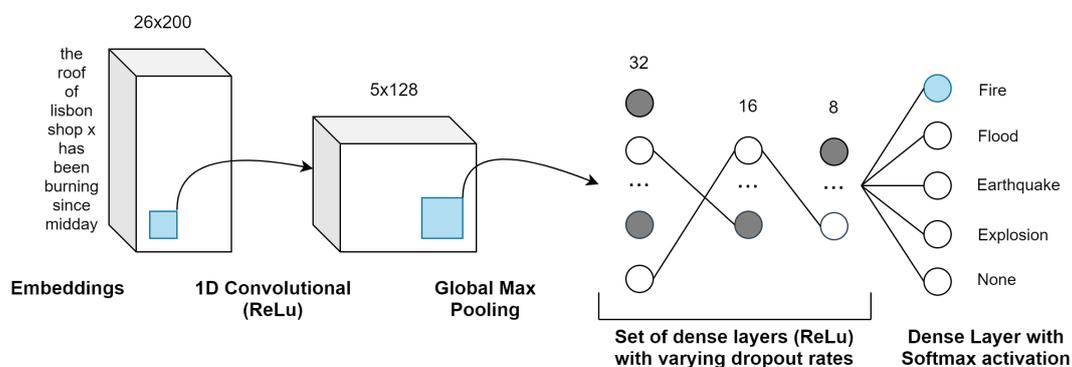


Figure 3. Threat type classification convolutional neural network (CNN).

1. **Threat Object Generation:** The second to last step of our pipeline consists of joining all of the information extracted from the previous steps. That is, the tweet information extracted by the Tweet Extraction and Filtering module, all of the entities extracted by the Named Entity Recognition module, and the threat type classified by the Disaster Extraction, are all combined to form a disaster object. This step is essential, as it creates a sense of identity for the detected disasters, joining their occurrences to all of the information that define them. We consider the location, date, and type of threat to be the minimum necessary information for creating and comparing a threat. Therefore, this module also serves as an extra filtering phase, as it discards any threats the system has not been able to extract enough information from.
2. **Knowledge Base Integration:** Lastly, the detected disasters must be included in our knowledge graph. This step involves all validations regarding the previous existence of our gathered data, as well as the creation of new nodes and relationships between them, to express the identity of

the extracted disasters in a way that facilitate human consumption and future system integrations. To further deal with social media unreliability, all disaster nodes have an attribute which is incremented when the same disaster is detected from different tweets. This attribute can then be used as a detection threshold, to allow for better filtering of relevant threats. If new information regarding an already existing disaster is extracted from the previous steps, it is also validated and then connected with that disaster. This mechanic allows for our system to be continually learning, expanding its knowledge, and bettering the knowledge it already has.

4. Evaluation

Our system is evaluated in two specific tasks: Disaster Extraction, and top-down testing. Disaster Extraction aims to compare our disaster classification algorithms with the current state-of-the-art systems, in terms of model evaluation metrics. The top-down testing, however, consists of a presentation of our system's behavior when confronted with a set of tweets having both relevant and noisy information.

4.1. Disaster Extraction Algorithms

To compare our models with the current state-of-the-art literature on tweet disaster classification, we have divided this information in the two aforementioned classes: Event relatedness classification and Event type classification. We have then selected the best results of each relevant publication, being essential to highlight that all of these have been published between 2018 and 2019. The compared models use different algorithms and implementation techniques, but all of them use subsets of the same datasets—*crisisLex*—to train.

Although not all the compared studies have shown their accuracy, precision, or recall, we have decided to also include these measures, as they may prove to be interesting in comparison with future works. To compare with the state-of-the-art however, as it is the most common measure for text classification, we use the F1-Score. This measure considers both precision and recall, and is often preferred in the field of information retrieval, as it offers a better estimate of wrongly classified cases, when compared to accuracy.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Our models have been trained using a subset of close to 20,000 labeled tweets. These tweets were balanced for the relatedness classification task, having around 54% related tweets and the remaining 46% unrelated. For the second task, however, there was a need to use under-sampling techniques to balance major classes—as our models were greatly overfitting those same classes—ending up with close to 5000 tweets.

DisKnow's CNN related classifier has fared exceptionally well in this task. We have achieved an increase of almost 10 p.p of F1-Score when compared to the current state-of-the-art [36], with a test F1-Score of 92. This improvement can be explained by the usage of Twitter-specific embeddings, and variations in the model's architecture, such as pre-processing methods—which change the input and reduce the noise the network is exposed to—and the neural network architecture as well. Detailed results are shown in Table 1.

Table 1. Comparative results of the event relatedness classification task.

Model	Test Accuracy	Test Precision	Test Recall	Test F1-Score
CNN [36]	-	-	-	0.838
LSTM [37]	0.748	-	-	0.751
CNN—DisKnow	0.903	0.921	91.9	0.920

As for the second task, disaster type classification, DisKnow has also achieved good results, although not as good as the state-of-the-art. A deep investigation on these results led us to believe

that our type-classification CNN has a slight overfit to the flood class, sometimes misclassifying unrelated tweets. This problem is, however, fixed with the lexicon-based model that was presented in the proposed approach, although not reflected in these results, as to correctly compare the models. Despite this, further improvements may be needed in the implementation of this model. Detailed results are shown in Table 2.

Table 2. Comparative results of the event type classification task.

Model	Test Accuracy	Test Precision	Test Recall	Test F1-Score
SVM [36]	-	-	-	0.997
CNN [38]	0.930	-	-	0.928
CNN—DisKnow	0.884	0.888	0.876	0.882

4.2. Top-Down Demonstration

Our top-down demonstration consists in a simulation that was led in the municipality of Barreiro, a Portuguese city nearby Lisbon. This scenario aims to show what DisKnow’s normal operation would be like, in a real scenario, from the moment it retrieves tweets, to the representation of the extracted information, in the Knowledge graph. As a result of the sheer impossibility of creating a large scale disaster for testing purposes, we had to resort to a simulated demonstration. Despite its smaller scale, this demonstration aims to show the contribution of DisKnow in disaster scenarios, automatically providing filtered and summed-up information to both local relevant authorities and population. In this demonstration, we have posted six relevant tweets at different moments. One hundred and thirty-six tweets were gathered by DisKnow. These tweets included the six tweets that we created and 130 other unknown tweets. Out of all these tweets, only the original six have successfully had impact in DisKnow’s knowledge base, the remaining having been filtered through the system’s pipeline, by either the Disaster Extraction or the Threat Object Generation component. Table 3 shows the amount of tweets outputted by each filtering component of the system.

Table 3. Information filtering throughout the system.

Component	Number of Tweets
Tweet Extraction and Filtering	136
Disaster Extraction (First CNN)	9
Disaster Extraction (Second CNN)	7
Threat Object Generation	6

The simulation starts with a fire in the middle of a local factory called FISIFE. The flames are seen by three citizens that were passing by, and each posts a different tweet reporting the occurrence, using different levels of expressiveness. A few minutes later, someone reports that the fire has also spread to the nearby Vodafone store. Eventually, local news report tweets about the factory director having got a third-degree burn from the strong flames. About ten minutes after the fire started, a citizen also reports that the local Remax shop has been flooded at 12:30, and two of its workers have been injured.

Figures 4–6 lay out the knowledge base on different moments of the simulation. Out of the initial three detections, two were written in Portuguese, and one in French. The follow-up tweets were all sent in Portuguese, and were written with different styles, to encompass what a real scenario would be like. Excluding all unrelated tweets, and keeping the order described in the scenario, the tweets detected by DisKnow were:

1. **Estão chamas no telhado da FISIFE Barreiro!**
2. **Au secours ! Fisipe brûle à Barreiro**
3. **#Chamas #FISIFE fogo no Barreiro!**

4. O incêndio da FISIFE espalhou-se para a loja da Vodafone do Barreiro
5. #Catástrofe #Fogo O director da fábrica FISIFE Barreiro João Alberto ficou com queimaduras de 3º grau devido às fortes labaredas
6. #Ajuda Está a transbordar água na Remax do Barreiro desde as 12h30. A Marta Almeida e a Fernanda Torre estão feridas

After the extraction and filtering, the next steps in DisKnow's pipeline are related to processing and translating the tweets, as well as extracting both named entities and referred disasters. These steps are represented in DisKnow's second to fifth modules, and the output is shown in Table 4.

Table 4. Processed tweets and extracted information.

Translated Tweet	Entities	Disaster
1. Flames are on the roof of FISIFE Barreiro	FISIFE (Org) Barreiro (GPE)	Fire
2. Help ! Fisipe burns in Barreiro	Fisipe (ORG) Barreiro (GPE)	Fire
3. Flames FISIFE fire in Barreiro	FISIFE (ORG) Barreiro (GPE)	Fire
4. The FISIFE fire was spread to the Vodafone do Barreiro store	FISIFE (ORG) Vodafone (ORG) Barreiro (GPE)	Fire
5. Fire catastrophe FISIFE factory director Barreiro João Alberto suffered 3rd degree burns due to strong flames	FISIFE (ORG) Barreiro (GPE) João Alberto (PER)	Fire
6. Help Water is overflowing at Remax do Barreiro since 12:30 pm Marta Almeida and Fernanda Torre are injured	Remax (Org) Barreiro (GPE) 12:30 (TIME) Marta Almeida (PER) Fernanda Torre (PER)	Flood

After the first three tweets were successfully processed, DisKnow has deemed them as referring to the same threat, which is done in the knowledge base integration module. The system's translating module allows for DisKnow to process information in any of the 109 languages GNMT supports [41], as demonstrated with the language variation in the second tweet. In Figure 4, we can see the representation of the threat, as nodes and connections. There is a central node that represents the threat occurrence, and this node is connected through a specific relationship to each of the five adjacent nodes, which have different types and attributes. In Figure 4, we can see these adjacent nodes representing dates, times, locations, threat types, and organizations, each of these nodes matching information extracted from the NER and Disaster Extraction modules of DisKnow. The representation of adjacent knowledge helps decision makers to not only know that a threat—in this case a fire—is ongoing, but to also know when and where it was first reported, and who is being affected by it. After the first detection, each subsequent tweet reporting that same threat incremented the "detections" attribute of the node, thus showing that it has had been confirmed by several citizens.

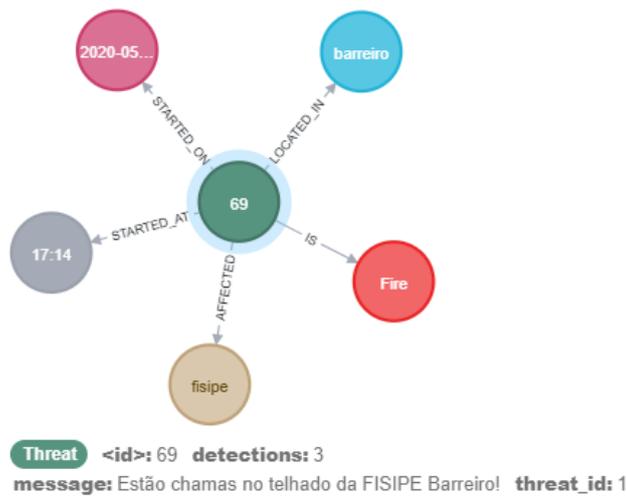


Figure 4. Knowledge graph after the first 3 tweets.

The impact of the fourth and fifth tweet is demonstrated in Figure 5. These tweets also referred to the same threat, but added new information, which was the fire spreading to a nearby store, and the director of the factory being injured. DisKnow correctly detected that this was an already reported occurrence, and proceeded to extract new information, which was then represented in the knowledge graph, by creating the new person and organization nodes and connecting them to the ongoing fire threat, with the AFFECTED relationship.

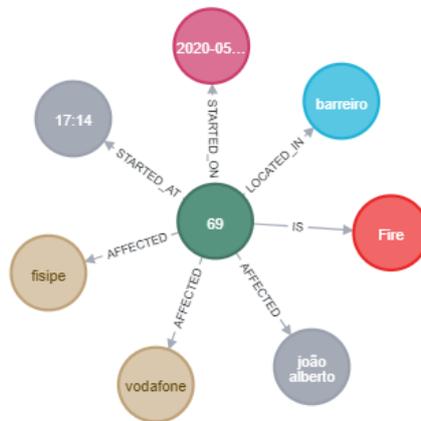


Figure 5. Knowledge graph after the first 5 tweets.

Lastly, in Figure 6 the result of the sixth tweet is included. This tweet refers to a completely different threat that was reported. Seeing that the citizen specified a start time for the flood, DisKnow used that value, instead of the time the tweet was published at. It is also interesting to observe that both threats are connected by their location and date, which is extremely valuable, as these connections can later be used for inference.

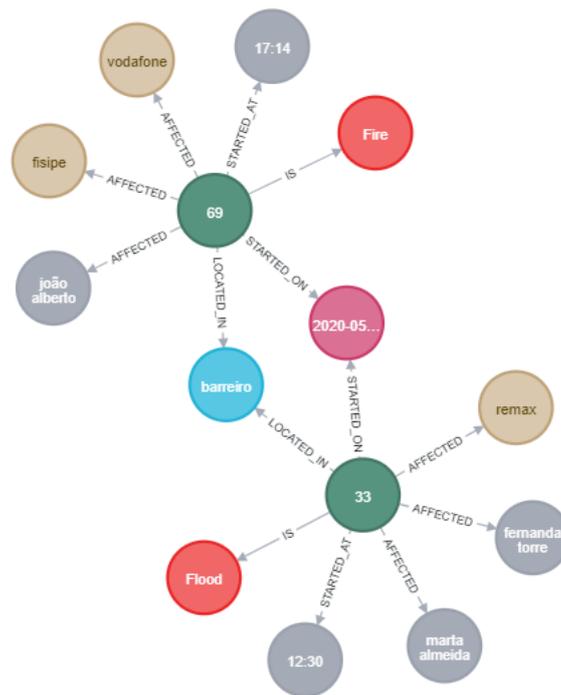


Figure 6. Knowledge graph after all 6 tweets.

5. Conclusions

Disaster related casualties have an extensive impact in our society. Disaster management focuses on preventing and minimizing the risks these scenarios face humanity with. Social media, although raw, presents a way of using humans as sensors to detect such hazards with the utmost brevity.

DisKnow builds upon existing tools, contributing with a dynamic way of extracting and representing spatial and temporal relationships, as well as providing this knowledge to decision-makers, in real-time. Its capability to detect disasters and related information, through social media, and then create intelligible representations, can be a powerful tool when dealing with scenarios requiring real-time decision-making, as every single bit of extra information can make the difference. The extraction and connection of related information allow our system to present a strong contribution, that is further expanded by the capacity of our knowledge graph to grow and encompass new information and update its own, presenting a reliable and up-to-date source of knowledge, and allowing for quicker and more informed decisions in scenarios where scarcity of time is often an issue. DisKnow's relevance has been evaluated by the administrative division of the municipality of Barreiro, representing an end-user, according to its role as the local authority of civil protection in the city of Barreiro and surrounding areas. The feedback has had a positive focus on the ability of using citizens as a network of volunteer sensors to correctly extracting and generating real-time knowledge about occurring threats, as well as the potential for learning during disaster responses.

Despite the good results and feedback, there is still much work to be done. Future improvements to DisKnow should include a way to filter trustworthy information, besides the sheer number of detections a disaster has. Other existing components such as the classification models and the NER component are also a continuous effort and should see some improvements which would better the system's extracted information. These improvements should also be complemented with the introduction of new node types to the graph architecture, as to expand the information DisKnow can offer its end users. Finally, we have seen that with its growth, a knowledge graph becomes incomprehensible for direct human usage. With this in mind, DisKnow also needs an automatic way to

communicate its knowledge to decision makers, whether it be general population, or first-responders. Digital assistants are a suitable contender for this task, in a nearby future.

DisKnow's contribution can prove to be very useful in providing overview knowledge in disaster scenarios. As a tool, with few scenario-related tweaks, it is ready for implementation in real disasters that might happen in the nearby future, or even current ones, like the 2020 COVID-19 pandemic.

Author Contributions: J.B. is a Master student and has performed all of the development work. M.D. is a Master student that has supported the development and validated the NER process. R.R. and J.C.F. are thesis supervisors and have organized all work and performed work revision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Infrastress Project—The Infrastress project is a H2020 financed project, aiming to address cyber-physical security of Sensitive Industrial Plants and Sites (SIPS), improve resilience and protection capabilities of SIPS exposed to large scale, combined, cyber-physical threats and hazards, and guarantee continuity of operations, while mitigating effects in the infrastructure itself, the environment, and citizens in vicinity, at reasonable cost. This work was also supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020.

Acknowledgments: We would like to thank the European Commission for supporting this research work under the Horizon 2020 EU Framework Programme for Research (project Infrastress, grant agreement n833088). We also thank all 27 European partners for supporting our project with their specific acknowledgement and contributions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dilley, M.; Chen, R.S.; Deichmann, U.; Lerner-Lam, A.L.; Arnold, M. *Natural Disaster Hotspots: A Global Risk Analysis*; World Bank: Washington, DC, USA, 2015.
2. Poblet, M.; García-Cuesta, E.; Casanovas, P. Crowdsourcing tools for disaster management: A review of platforms and methods. In *International Workshop on AI Approaches to the Complexity of Legal Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 261–274.
3. Feldman, D.; Contreras, S.; Karlin, B.; Basolo, V.; Matthew, R.; Sanders, B.; Serrano, K. Communicating flood risk: Looking back and forward at traditional and social media outlets. *Int. J. Disaster Risk Reduct.* **2016**, *15*, 43–51. [[CrossRef](#)]
4. Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; Meier, P. Extracting information nuggets from disaster-Related messages in social media. In *Proceedings of the 10th International ISCRAM Conference*, Baden, Germany, 12–15 May 2013.
5. Fohringer, J.; Dransch, D.; Kreibich, H.; Schröter, K. Social media as an information source for rapid flood inundation mapping. *Nat. Hazards Earth Syst. Sci. (NHESS)* **2015**, *15*, 2725–2738. [[CrossRef](#)]
6. Rogstadius, J.; Kostakos, V.; Laredo, J.; Vukovic, M. A real-time social media aggregation tool: Reflections from five large-scale events. In *Proceedings of the CSCWSmart Workshop at ECSCW 11*, Siegen, Germany, 24–28 September 2011.
7. Steiger, E.; De Albuquerque, J.P.; Zipf, A. An advanced systematic literature review on spatiotemporal analyses of twitter data. *Trans. GIS* **2015**, *19*, 809–834. [[CrossRef](#)]
8. Williams, S.A.; Terras, M.M.; Warwick, C. What do people study when they study Twitter? Classifying Twitter related academic papers. *J. Doc.* **2013**, *69*, 384–410. [[CrossRef](#)]
9. Li, J.; Rao, H.R. Twitter as a rapid response news service: An exploration in the context of the 2008 China earthquake. *Electron. J. Inf. Syst. Dev. Ctries.* **2010**, *42*, 1–22. [[CrossRef](#)]
10. Tatsubori, M.; Watanabe, H.; Shibayama, A.; Sato, S.; Imamura, F. Social web in disaster archives. In *Proceedings of the 21st International Conference on World Wide Web*, Lyon, France, 16–20 April 2012; pp. 715–716.
11. Trending, B.B.C. How the Paris Attacks Unfolded on Social Media. Available online: <https://www.bbc.com/news/blogs-trending-34836214> (accessed on 20 May 2020).
12. Vieweg, S.; Hughes, A.L.; Starbird, K.; Palen, L. Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, 10–15 April 2010; pp. 1079–1088.

13. Starbird, K.; Palen, L.; Hughes, A.L.; Vieweg, S. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, GA, USA, 6–10 February 2010; pp. 241–250.
14. Starbird, K.; Palen, L. *Pass it on? Retweeting in Mass Emergency*; International Community on Information Systems for Crisis Response and Management: Seattle, WA, USA, 2010; pp. 1–10.
15. Palmer, J. Emergency 2.0 is coming to a website near you. *New Sci.* **2008**, *198*, 24–25. [CrossRef]
16. Busari, S. Tweeting the Terror: How Social Media Reacted to Mumbai. CNN International. Available online: <https://edition.cnn.com/2008/WORLD/asiapcf/11/27/mumbai.twitter/> (accessed on 22 May 2020).
17. Meier, P. Behind the Scenes: The Digital Operations Center of the American Red Cross. Available online: <http://irevolution.net/2012/04/17/red-cross-digital-ops> (accessed on 22 May 2020).
18. Lamont, I. A Plane Lands on the Hudson River, and Twitter Documents It All. Available online: <http://www.computerworld.com/s/article/9126181> (accessed on 22 May 2020).
19. Graham, M. What Can Twitter Tell Us about Hurricane Sandy Flooding? Visualised. Available online: <http://www.guardian.co.uk/news/datablog/2012/oct/31/twitter-sandy-flooding> (accessed on 22 May 2020).
20. Graham, M.; Poorthuis, A.; Zook, M. Digital Trails of the UK Floods-How Well Do Tweets Match Observations. Available online: <https://www.theguardian.com/news/datablog/2012/nov/28/data-shadows-twitter-uk-floods-mapped> (accessed on 22 May 2020).
21. Kim, J.; Hastak, M. Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manag.* **2018**, *38*, 86–96. [CrossRef]
22. Lanfranchi, V.; Mazumdar, S.; Blomqvist, E.; Brewster, C. Roadmapping Discussion Summary: Social Media and Linked Data for Emergency Response. In Proceedings of the 10th Extended Semantic Web Conference, Montpellier, France, 26–30 May 2013.
23. Rogstadius, J.; Kostakos, V.; Laredo, J.; Vukovic, M. Towards real-time emergency response using crowd supported analysis of social media. In Proceedings of the International Conference on Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada, 7–12 May 2011.
24. Ahmed, A. Use of Social Media in Disaster Management. In Proceedings of the 32nd International Conference on Information Systems, Shanghai, China, 4–7 December 2011.
25. De Longueville, B.; Smith, R.S.; Luraschi, G. “OMG, from here, I can see the flames!” A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In Proceedings of the International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 73–80.
26. MacEachren, A.M.; Robinson, A.C.; Jaiswal, A.; Pezanowski, S.; Savelyev, A.; Blanford, J.; Mitra, P. Geo-twitter analytics: Applications in crisis management. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011; pp. 3–8.
27. Murthy, D.; Longwell, S.A. Twitter and disasters: The uses of Twitter during the 2010 Pakistan floods. *Inf Commun. Soc.* **2013**, *16*, 837–855. [CrossRef]
28. Bosley, J.C.; Zhao, N.W.; Hill, S.; Shofer, F.S.; Asch, D.A.; Becker, L.B.; Merchant, R.M. Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication. *Resuscitation* **2013**, *84*, 206–212. [CrossRef] [PubMed]
29. Kim, A.E.; Hansen, H.M.; Murphy, J.; Richards, A.K.; Duke, J.; Allen, J.A. Methodological considerations in analyzing Twitter data. *J. Nat. Cancer Inst. Monogr.* **2013**, *2013*, 140–146. [CrossRef] [PubMed]
30. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake shakes Twitter users: Real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 851–860.
31. Uddin, M.Y.S.; Al Amin, M.T.; Le, H.; Abdelzaher, T.; Szymanski, B.; Nguyen, T. On diversifying source selection in social sensing. In Proceedings of the Ninth International Conference on Networked Sensing (INSS), Antwerp, Belgium, 11–14 June 2012; pp. 1–8.
32. Le, H.; Wang, D.; Ahmadi, H.; Uddin, M.Y.S.; Ko, Y.H.; Abdelzaher, T.; Wang, H. *Apollo: A Data Distillation Service for Social Sensing*; Tech. Rep.; University of Illinois Urbana-Champaign: Champaign, IL, USA, 2012.
33. Rogstadius, J.; Vukovic, M.; Teixeira, C.A.; Kostakos, V.; Karapanos, E.; Laredo, J.A. CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM J. Res. Dev.* **2013**, *57*, 1–4. [CrossRef]
34. Farzindar, A.; Khreich, W. A survey of techniques for event detection in twitter. *Comput. Intell.* **2015**, *31*, 132–164.

35. Nguyen, T. H.; Grishman, R. Event detection and domain adaptation with convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 365–371.
36. Burel, G.; Saif, H.; Fernandez, M.; Alani, H. On Semantics and Deep Learning for Event Detection in Crisis Situations. In Proceedings of the 14th International Conference (ESWC 2017), Portorož, Slovenia, 28 May–1 June 2017.
37. Sit, M. A.; Koylu, C.; Demir, I. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: A case study of Hurricane Irma. *Int. J. Digit. Earth* **2019**, *12*, 1205–1229. [[CrossRef](#)]
38. Alam, F.; Ofli, F.; Imran, M. CrisisDPS: Crisis Data Processing Services. In Proceedings of the International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, 19–22 May 2019.
39. He, H.; Balakrishnan, A.; Eric, M.; Liang, P. Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017.
40. Haklay, M.; Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18. [[CrossRef](#)]
41. Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Klingner, J. Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv* **2016**, arXiv:1609.08144.
42. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. BLEU: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.
43. Olteanu, A.; Castillo, C.; Diaz, F.; Vieweg, S. Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In Proceedings of the Eighth International AAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014.
44. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing EMNLP, Doha, Qatar, 25–29 October 2014; pp. 1532–1543.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).