

# A Domain-Independent and Multilingual Approach for Crisis Event Detection and Understanding

Hernan Sarmiento

Department of Computer Science, University of Chile  
Santiago, Chile  
hsarmien@dcc.uchile.cl

## ABSTRACT

Most existing approaches that use social media for detecting and characterizing emerging crisis events are based on the analysis of messages obtained from social platforms using a predetermined set of keywords [2, 3]. In addition to keyword filters, messages must commonly be post-processed using supervised classification models to determine if messages are referring to a real-time crisis situation or not. However, keyword-based approaches have certain shortcomings; on the one hand they require specific domain knowledge of different crisis events to determine a set of keywords to filter relevant data about an emerging crisis situation; on the other hand, they require supervised methods to determine if the identified data actually corresponds to a new real-time crisis event. Hence, the creation of keyword-independent methods could also help generalize existing approaches so they can be used for cross-lingual events, since each language and culture can have its own particular terms to refer to a same event. The majority of these works also explain phenomena just for English messages. This limitation avoids replication of methodologies in other languages and countries where emergency events often occur. For this reason, researchers recently have focused on creating domain-independent and multi-lingual approaches for detecting and classifying social media messages during crisis events [1, 4]. These approaches have exploited low-level lexical features with the goal of reaching domain-transfer among different crisis events and languages. Nonetheless, most studies focused on crisis-related messages without testing non related crisis messages such as sporting events or music festivals.

The main objective of this work is to study and exploit cross-lingual domain-independent patterns for detecting and characterizing social media messages generated in collective activity related to unexpected high-impact real-world events in social media platforms, and specifically on emergency situations. The expected contribution is to develop novel techniques in order to provide multi-lingual and domain-independent detections, and characterizations of emergency situations. Such techniques should help us better understand the social media behavior during crises in affected locations around the world, independent of their language, domain and type of event. Some of these patterns arise independently of the particular type or domain of the crisis event, as well

as independent of the location, language and culture of the users that participate. Hence, our hypothesis is that there are patterns in the self-organized activity of the Web and social media users that emerge when a crisis situation starts to unfold in the physical world. Some of these patterns arise independently of the particular type or domain of the crisis event, as well as independent of the location, language and culture of the users that participate. We then propose the following research questions:

- *RQ1: Can we characterize collective patterns during crisis situations independently of their language and domain based on non-textual and low-level lexical features?*
- *RQ2: Are there differences among types of emergency situations (instantaneous, progressives, focalized and diffused) related to social media messages posted during these events?*
- *RQ3: Are non-textual and low-level lexical features sufficient at reducing the number of non-related emergency situations detected as crises in the Web and social media?*

## CCS CONCEPTS

• Information systems → Information extraction.

## KEYWORDS

Emergency Situations; Crisis Informatics; Social Media

### ACM Reference Format:

Hernan Sarmiento. 2019. A Domain-Independent and Multilingual Approach for Crisis Event Detection and Understanding. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*, July 21–25, 2019, Paris, France. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3331184.3331425>

## ACKNOWLEDGMENTS

This work has been partially funded by the Millennium Institute for Foundational Research on Data.

## REFERENCES

- [1] Prashant Khare, Grégoire Burel, Diana Maynard, and Harith Alani. 2018. Cross-Lingual Classification of Crisis Data. In *International Semantic Web Conference*. Springer, 617–633.
- [2] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief. In *ICWSM*.
- [3] Jazmine Maldonado, Jheser Guzman, and Barbara Poblete. 2017. A Lightweight and Real-Time Worldwide Earthquake Detection and Monitoring System Based on Citizen Sensors. In *Proceedings of the Fifth Conference of Human Computation and Crowdsourcing*. AAAI, 137–146.
- [4] Hernan Sarmiento, Barbara Poblete, and Jaime Campos. 2018. Domain-Independent Detection of Emergency Situations Based on Social Activity Related to Geolocations. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 245–254.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6172-9/19/07.

<https://doi.org/10.1145/3331184.3331425>