CrisisLexT26 Tweets from 26 crises, labeled by informativeness, information type and source Nov 2014

This collection includes tweets collected during 26 large crisis events in 2012 and 2013, with about 1,000 tweets labeled per crisis for informativeness (i.e. “informative," or "not informative"), information type, and source.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Crisis** | **Country** | **Start / Duration** | **#Tweets** | **Category** | **Sub-Category** | **Type** | **Development** | **Spread** |
| 2012 Italy earthquakes | Italy | May / 32 days | 7,351 | Natural | Geophysical | Earthquake | Diffused | Instantaneous |
| 2012 Colorado wildfires | US | Jun / 31 days | 4,172 | Natural | Climatological | Wildfire | Diffused | Progressive |
| 2012 Philipinnes floods | Philipinnes | Aug / 13 days | 2,950 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2012 Venezuela refinery explosion | Venezuela | Aug / 12 days | 2,736 | Human-induced | Accidental | Explosion | Focalized | Instantaneous |
| 2012 Costa Rica earthquake | Costa Rica | Sep / 13 days | 2,193 | Natural | Geophysical | Earthquake | Diffused | Instantaneous |
| 2012 Guatemala earthquake | Guatemala | Nov / 20 days | 3,261 | Natural | Geophysical | Earthquake | Diffused | Instantaneous |
| 2012 Typhoon Pablo | Phillipines | Nov / 21 days | 1,944 | Natural | Meteorological | Typhoon | Diffused | Progressive |
| 2013 Brazil nightclub fire | Brazil | Jan / 16 days | 4,786 | Human-induced | Accidental | Fire | Focalized | Instantaneous |
| 2013 Queensland floods | Australia | Jan / 19 days | 1,223 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2013 Russian meteor | Russia | Feb / 19 days | 8,365 | Natural | Others | Meteorite | Focalized | Instantaneous |
| 2013 Boston bombings | US | Apr / 60 days | 157,454 | Human-induced | Intentional | Bombings | Focalized | Instantaneous |
| 2013 Savar building collapse | Bangladesh | Apr / 36 days | 4,070 | Human-induced | Accidental | Collapse | Focalized | Instantaneous |
| 2013 West Texas explosion | US | Apr / 29 days | 14,505 | Human-induced | Accidental | Explosion | Focalized | Instantaneous |
| 2013 Alberta floods | Canada | Jun / 25 days | 5,887 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2013 Singapore haze | Singapore | Jun / 19 days | 3,639 | Mixed | Others | Haze | Diffused | Progressive |
| 2013 Lac-Megantic train crash | Canada | Jul / 14 days | 2,342 | Human-induced | Accidental | Derailment | Focalized | Instantaneous |
| 2013 Spain train crash | Spain | Jul / 15 days | 3,681 | Human-induced | Accidental | Derailment | Focalized | Instantaneous |
| 2013 Manila floods | Phillipines | Aug / 11 days | 2,032 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2013 Colorado floods | US | Sep / 21 days | 1,778 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2013 Australia wildfires | Australia | Oct / 21 days | 1,982 | Natural | Climatological | Wildfire | Diffused | Progressive |
| 2013 Bohol earthquake | Phillipines | Oct / 12 days | 2,214 | Natural | Geophysical | Earthquake | Diffused | Instantaneous |
| 2013 Glasgow helicopter crash | UK | Nov / 30 days | 2,558 | Human-induced | Accidental | Crash | Focalized | Instantaneous |
| 2013 LA Airport shootings | US | Nov / 12 days | 2,730 | Human-induced | Intentional | Shootings | Focalized | Instantaneous |
| 2013 NYC train crash | US | Nov / 8 days | 1,066 | Human-induced | Accidental | Derailment | Focalized | Instantaneous |
| 2013 Sardinia floods | Italy | Nov / 13 days | 1,143 | Natural | Hydrological | Floods | Diffused | Progressive |
| 2013 Typhoon Yolanda | Phillipines | Nov / 58 days | 38,951 | Natural | Meteorological | Typhoon | Diffused | Progressive |

* **Contents**: ~250K tweets posted during 26 crisis events in 2012 and 2013, with most events having 2K-4K tweets.
* **Sampling method**: by keyword filtering from tweets included in the 1% sample at the Internet Archive.
* **Labels**: ~28,000 tweets (about 1,000 in each collection) were labeled by crowdsource workers according to informativeness (informative or not informative), information types (e.g. caution and advice, infrastructure damage), and information sources (e.g. governments, NGOs).
* **Data format**: comma-separated values (.csv) files containing tweet-ids for the unlabeled tweets, plus the text of the tweets and labels for the labeled ones. Also includes a JSON file with metadata about the collection, including the keywords used to select tweets.

If you use the CrisisLexT26 collection, please cite:

* A. Olteanu, S. Vieweg, C. Castillo. 2015. [What to Expect When the Unexpected Happens: Social Media Communications Across Crises](http://crisislex.org/papers/cscw2015_transversal_study.pdf). In Proceedings of the ACM 2015 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '15). ACM, Vancouver, BC, Canada.