

# Annotation instruction

## 1. Introduction

In this task, we want to know about Hurricane Sandy by reading the tweets in 2012, the year Hurricane Sandy happened. We want to know how it happened, what was its influence on people, what people were thinking, what they exactly experienced during the disaster, what was the damage, etc.

So in this task, we will annotate tweets about Hurricane Sandy. All tweets contain keyword “Sandy”, but some of them may not give us information about the hurricane. The goal of this task is to judge if the tweets are RELEVANT to Hurricane Sandy, and label the different aspects (in this task we call them TOPICS) that tweets give useful information to.

There is another important concept in this task, called ENTITY. The entity can be a city, a small area, an organization, a river, a beach, etc. Tweets about the same entity are put together, and are shown to you in time order. In the tweets group of the same entity, for any tweet in it, other tweets (especially the tweets posted before this one) can be considered as its CONTEXT. The context may help you understand a tweet.

## 2. Procedure

The annotation procedure includes two steps for every tweet:

**First step:** Judge the tweet RELEVANT or not based on whether it provides USEFUL information about Hurricane Sandy. If it is irrelevant, do NOT give it any tag, JUST PASS and ignore it.

**Second step:** Based on the topic(s) that the tweet talks about, give it one or more TOPIC tags.

NOTICE: Pay attention to the **hashtag**(#XXX) and **mention**(@XXX) besides the **plain text** when you read the tweets.

## 2.1 How to judge if a tweet is relevant or not?

Basically, a tweet can be considered relevant if its content has ALL the following features:

- **Related to Hurricane Sandy.** The content of some tweets may have nothing to do with Hurricane Sandy, even though sometimes they contain hashtags like “#sandy”.

Examples of irrelevant tweets:

- *Does anyone know of any good drink specials in the West Village ? #sandy #ZoneA @ Casa Tilbury <http://t.coKbi5Euni>*
- *My #Mom loves weather + can't get enough of my updates . #Love her recommendation #Sandy #NYC @ West Village <http://t.coOtM2GVBO>*

- **Useful.** The tweet can tell you something that you care about. For example, if tweets only talk about things so personal that other people/strangers usually would not care about at all, it would be irrelevant.

Example of irrelevant tweet:

- *one last trip outside before sandy strikes . @ Upper East Side <http://t.co6ZBapEj9>*

- **Understandable.** If you cannot understand what a tweet is talking about, even after you look at its context, just consider it irrelevant. If you can figure out what it probably means, and you feel that most people, even children, could also, then it is still understandable. But if you use specialized knowledge that most people would not have, then it is not relevant.

Example of irrelevant tweet:

- *Take that , #sandy @ East Village <http://t.coPyzrRaMG>*

And here we provide some examples of relevant tweets:

- *Stocking up at an East Village bodega because the supermarket lines are too long #Sandy2k12*
- *ConEd called to alert that they are probably shutting power preemptively in the east village ... not stoked #sandy*

## 2.2 How to give a tweet topic tag(s)?

In this task, TOPIC is a subject that Twitter users often tweet about during the Hurricane Sandy event. Based on the topic(s) that the tweet talks about, give it **one or more** TOPIC tags.

### 2.2.1 Topics

We have built a typology of topics. This typology has two levels, and 9 topics that can be tagged (those have acronyms) in total.

Now we will describe every topic and give you some examples to help you understand each of them. Here is the format of our examples. The [...] part before every tweet is the topic tags for this tweet. And topics are separated by stars \*.

**[TE\*WD]** Temp is cooler outside , wind is up in the East Village . Lucky we stocked up on torches and more booze than food . #sandy

t.a. = tweets about

1) Description of Hurricane: Direct description about the hurricane.

- **WinD** (WD): t.a. wind
- **RaiN** (RN): t.a. rain  
Example:
  - a) **[RN\*F]** *Haven't had much rain but the Hudson River don't care #HurricaneSandy #Frankenstorm*
- **Temp** (TE): t.a. temperature during hurricane

2) The Influence of Hurricane on Human society: about the influence of hurricane on different aspects of human society.

- **Flood/river/ocean** (F): t.a. water in/from **natural environment**, e.g. water in **river**, **wave in the ocean**, **storm surge**, or **flood on streets**.  
Example:
  - a) **[F]** *The waves are angry ! Time to evacuate ! We are out ! #sandy #hurricane #crazy #waves #evacuate @ Long Beach*
  - b) **[F]** *Staten Island streets flooded by Hurricane Sandy*  
*http:t.co0VHZArcz*

c) [F] *Bring it on Sandy #sickbday #sandy #hurricane #flood @ Long Beach - East End*

- **Tree (T):** t.a. trees.

Example:

a) [T] *Sanitation Dept . clearing downed trees in West Village . #sandy #HurricaneSandy <http://t.copXQ7UtTc>*

- **Explosion/Fire (EF)**

Example:

a) [EF] *#broadway #fireandrescue #sandy #hurricane #longisland @ LONG BEACH , NY <http://t.codVRMpEXa>*

- **Chemical Spill (CS):** E.g. oil spill

- **Building (B):** t.a. building, including **window, door, facade, etc.** E.g. people **tape up the window, lock the doors** or a building damaged in hurricane.

Example:

a) [B] *Many businesses in the West Village have taped up their windows in preparation for #Sandy <http://t.coHewfAo6g>*

b) [B] *@twc\_hurricane BREAKING : FDNY responding to report of a 4-story apartment building collapse in Chelsea , Manhattan . #Sandy #NYC*

### 3) General

- **Situation (S):** 1) **General** description of the **situation** of the hurricane, of people, of a city, or other things. 2) The tweet is relevant, but none of other topics can be tagged.

Example:

a) [S] *#TheHurricane is really bad in some parts of Staten Island and Jersey . I hope everyone is safe and warm . #Sandy watching @NY1headlines .*

b) [S] *Hurricane Sandy is doing a number on Staten Island right now .*

- **Subjective (SU):** If the tweet contains people's **explicitly expressed** subjective information, e.g. **emotion (worried, fear, etc.)**, **prayer, hope, wish, desire, etc**, tag this topic. **DO NOT infer here**, because we **do not tag implicitly expressed** subjective information.

Example:

- a) [SU] *Prayers go out to everyone effected by hurricane sandy especially to my family in Staten Island .*
- b) [F\*SU] *Help Staten Island #hurricanesandy #majorflood #help #desperate*
- c) [SU] *Stay safe New York ! #hurricanesandy #newyork #batterypark @ Battery*
- d) [S] *@RadioMoBounce why isn't Staten Island getting recognized ? It was the hardest hit borough #StatenIsland #Sandy #aftermath*

We can infer that the user is angry here, but since there are **no words that explicitly express** his anger, we do **NOT give it SU tag**.

Here is the typology of all topics:

❖ Description of Hurricane

- WinD(WD)
- RaiN(RN)
- Temp(TE) : temperature

❖ The Influence of Hurricane on Human Society

- Flood/river/ocean (F): e.g. waves, flood on street, storm surge...
- Tree(T)
- Explosion/Fire(EF)
- Chemical Spill(CS): e.g. oil spill
- Building(B): window, door, building collapse...

❖ General

- Situation(S): 1) general description 2) tweet is relevant, but no other topics can be tagged here
- SUBjective(SU): Explicit subjective information. e.g. emotion(worried, fear, desperate, etc.), prayer, hope, wish, desire...

### 2.2.2 Some things to notice

- Usually when you see a **relevant** tweet, but **none of the topics** can be tagged, use **Situation (S)**.
- **Multiple topic tags can be tagged to one tweet**, because one tweet can talk about different topics, but **usually no more than THREE topics** for one tweet. Here is some statistics in gold-standard corpus, 16% of tweets are irrelevant, 75% one topic, 8% two topics, 1% three topics.
- The content of the tweet may be limited, so when what the tweet is talking about is not so clear, you can do some **reasonable inference** based on the context and your background knowledge, but **NO excessive association and imagination**. Also notice that: the background knowledge you use **MUST NOT be specialized knowledge**; use the knowledge that you think most people have.
- You may just see PART of all the topics during your annotation.
- Remember that **negation does NOT affect the topic tagging**. For example, for tweets “Wind is up here.” and ”There is no wind”, topic WinD should be tagged for both of them.