

Style Guide

We define guidelines to assist in the decision whether to tag a token as a entity trigger. The main purpose of entity triggers is to explain why their related named entity belongs to the class it does. In the case of this paper, entity triggers explain why a named entity is a location specifically. We aim to utilize entity triggers as explanation on their own without further justification. Thus, when entity triggers are tagged for dataset creation, a person examining the dataset should be able to understand why a entity trigger was chosen without confusion or need for elaboration. This should allow us to take a step towards being able to use entity triggers to improve the explainability of the system trained on them.

No guidelines for annotating entity triggers exist yet, so we develop our own. These guidelines were chosen based on the intuition of the author and common patterns noticed by the author in existing trigger-annotated datasets. More research should be done in determining which patterns often make for good trigger explanations.

Entity triggers must be associated with a single named entity. These guidelines assume that the sentence being annotated only has one named entity in it. If a sentence has multiple named entities, the sentence should be duplicated for each named entity. Each duplication should have one unique named entity each.

It should be noted that it is possible there are no tokens suitable for explanation in a sentence. In this case, we simply do not tag any entity triggers. Additionally, should we come across an entity in our dataset which was tagged as a location entity despite not being a location, we do not tag any entity triggers as well. We write down the index of these falsely-tagged entities so that we may choose to cull these entries at a later time.

1. Tag phrases which define the named entity. If such phrases are known to be a location, it is easily inferred that the named entity is a location. Examples:

- Local currency depreciated in countries such as Brazil, Chile and South Africa.
LOC
- New York is a bustling city. It is nicknamed “The City That Never Sleeps.”
LOC

2. Tag phrases which describe features of the location. If a named entity is said to contain municipalities or mountains, it is likely a location. Examples:

- Office locations were reduced in the US and Mexico.
LOC
- We eventually arrived in the indigenous community of Mechahuasca.
LOC

3. Tag phrases which the location describes as an attributive noun. In other words, if the location acts like an adjective. Examples:

- A Kansas woman bought the entire stock to donate to Nebraska flood victims.
LOC
- A Kansas woman bought the entire stock to donate to Nebraska flood victims.
LOC

4. Tag verb phrases which describe events that imply a location. Verbs tend to refer to specific classes of entities: the object following a phrase such as “travelled to” is likely describing a location. Examples:

- Celebrities quickly evacuate California as wildfires rage.
LOC

- We eventually arrived in the indigenous community of Mechahuasca.
LOC
5. Tag prepositions referring to the location. Prepositions often denote a direction or spatial relationship. This helps to infer locations, particularly together with verbs from guideline 4. Examples:
- A mountain biking race from Luxembourg to the Netherlands collects donations.
LOC
 - Begging 4 help in Canoa, pacific coast. There is no way to get there by road.
LOC
6. Tag other locations in the same clause as the named entity. For instance, if a named entity is among a list of locations, it is likely a location itself as well. Examples:
- Local currency depreciated in countries such as Brazil, Chile and South Africa.
LOC
 - Begging 4 help in Canoa, pacific coast. There is no way to get there by road.
LOC
7. Tag pronouns referring to the named entity. All types of pronouns may aid in inferring the class of an entity, including but not limited to personal pronouns like “it,” demonstrative pronouns like “that” and relative pronouns like “which.” These pronouns may help distinguish inanimate locations from animate entities with similar names, such as a person named “Virginia” being referred to by “she” or “who” in the sentence. Examples:
- New York is a bustling city. It is nicknamed “The City That Never Sleeps.”
LOC
 - Begging 4 help in Canoa, pacific coast. There is no way to get there by road.
LOC
8. Tag adjectives describing details of the location. Certain adjectives are often used with locations specifically. Examples:
- We went on a holiday to picturesque Florence for our honeymoon.
LOC
 - New York is a bustling city. It is nicknamed “The City That Never Sleeps.”
LOC
9. Tag punctuation tokens only if they are part of words which help to infer the class of the named entity, such as periods or hyphens in words. Enclosing symbols such as parentheses or quotation marks are not considered part of the words they enclose. Examples:
- 70,000 tarps to the U.S. Virgin Islands and Puerto Rico since Hurricane Maria.
LOC
 - Storm Irma moving from Saint-Martin to Florida.
LOC
10. Do not tag adjectives describing things other than the location as in guideline 8. Example:
- Celebrities ~~quickly~~ evacuate California as wildfires rage.
LOC

11. Do not tag punctuation unless it fits guideline 9. Ignore other punctuation such as parentheses, hashtags, commas, periods.
12. Do not tag parts of hashtag names. As hashtag names include no spaces, only tag these if the entire hashtag name is the entity trigger. Do not tag the hashtag symbol itself. Example:

- Call if you wanna donate something #Amatrice #Accumoli ~~#PrayForItaly~~
LOC