

Annotation Guide

This document was created to standardize the rules for annotating spatial and thematic entities in texts from automated transcripts and tweets.

The rules are detailed in paragraphs 1, 2 and 3. Two examples of annotated texts are presented in paragraph 4.

A spatial entity is defined as a language element that refers to a single, concrete entity, belonging to a specific domain : the location mentioned on the text.

Example : ‘France’, ‘Dolomites’, etc, are spatial entities.

A thematic entity is defined as a language element that refers to a single, concrete entity, belonging to a specific domain : the topic, subject of the text.

Example : ‘bird flu’, etc, is a thematic entity.

1. General Annotation

- An entity is considered with the name of the entity and its label (spatial, thematic, etc).

For any n-gram representing a spatial or thematic entity, we annotate it with its label between bracket and parenthesis. [*name_entity*, *label_entity*]

Example : [‘France’, ‘spatial’]

- We do not take into account the determiners when annotating.

Example : the sentence “*I went to the United States*” will be annotated “*I went to the [‘United States’, ‘GPE’]*”

2. Annotation of spatial entities

- When annotating spatial entities, a distinction must be made: we do not annotate a place name as a place name if it semantically represents an organization.

Example : “*Barcelona will continue the construction*” Barcelona will not be annotated as a spatial entity because Barcelona is considered here as an organization.

- When a spatial entity have a name of the city, etc on it that represented a building, name of river, etc..., we do not annotate only the name of the city but the entire spatial entity

Example : the sentence “*I went to the Delhi Zoo*” will be annotated “*I went to the Delhi Zoo*” as a spatial entity, and not only “Delhi”.

- We do not annotate relative spatial entities; relation types as orientation, distance are not considered.

Example : the sentence “The event was in the South of France” will be annotated : “The event was in the South of [‘France’, ‘GPE’].

There are three semantic types representing spatial entities. With these different types, we can better represent entities e.g. between countries, mountains, roads, etc. The three types (GPE, LOC, and FAC) are described below:

a. Type GPE : Geopolitical entity

The type ‘GPE’ is specific for entities representing countries, cities, states, etc.

Example : [‘France’, ‘GPE’] ; [‘Europe’, ‘GPE’] ; [‘Hérault’, ‘GPE’]

b. Type LOC : Non-GPE locations

The type ‘LOC’ is specific for entities representing mountain ranges, bodies of water, etc.

Example : [‘Mount Everest’, ‘LOC’]

c. Type FAC

The type ‘FAC’ is specific for entities representing buildings, airports, highways, bridges, etc.

Example : [‘Delhi Zoo’, ‘FAC’]

3. Annotation of thematic entities

There are several semantic types for annotating corpora. For the MOOD project, we take into account these three semantic types that are based on UMLS : ‘Virus’, ‘Sign or Symptom’ and ‘Disease or Syndrome’.

a. Type Virus

The type ‘Virus’ is specific for entities that are defined like this : “An organism consisting of a core of a single nucleic acid enclosed in a protective coat of protein. A virus may replicate only inside a host living cell. A virus exhibits some but not all of the usual characteristics of living things.”

Example : ['coronavirus', 'Virus'] ; ['H5N1', 'Virus']

b. Type Sign or Symptom

The type 'Sign or Symptom' is specific for entities that are defined like this :
“An observable manifestation of a disease or condition based on clinical judgment, or a manifestation of a disease or condition which is experienced by the patient and reported as a subjective observation.”

Example : ['sore throat', 'Sign or Symptom'] ; ['fever', 'Sign or Symptom']

c. Type Disease or Syndrome

The type 'Disease or Syndrome', is specific for entities that are defined like this : “A condition which alters or interferes with a normal process, state, or activity of an organism. It is usually characterized by the abnormal functioning of one or more of the host's systems, parts, or organs. Included here is a complex of symptoms descriptive of a disorder.||Any specific disease or syndrome that is modified by such modifiers as "acute", "prolonged", etc. will also be assigned to this type. If an anatomic abnormality has a pathologic manifestation, then it will be given this type as well as a type from the 'Anatomical Abnormality' hierarchy, e.g., "Diabetic Cataract" will be double-typed for this reason.”

Example : ['bird flu', 'Disease or Syndrome'] ; ['flu', 'Disease or Syndrome']

4. Examples

Automatic Transcript Sample:

"Breaking news coming in ['bird flu', 'Disease or Syndrome'] hits ['delhi zoo', 'FAC'] seven positive cases of ['bird flu', 'Disease or Syndrome'] have been reported"

Automatic Transcript Sample:

"['bird flu', 'Disease or Syndrome'] outbreak continues to spread in more states across the country the ['avian influenza', 'Disease or Syndrome'] has been confirmed in the national capital and ['mumbai', 'GPE'] in ['delhi', 'GPE'] eight samples taken from the dead birds have tested positive remember crows and ducks were found dead at various parks of ['delhi', 'GPE'] over the past few days

Automatic Transcript Sample:

"symptoms begins with two to eight days and can seem like ['common flu', 'Disease or Syndrome'] like ['cough', 'Sign or Symptom'] ['fever', 'Sign or Symptom'] ['sore throat', 'Sign or Symptom'] ['muzzle legs', 'Sign or Symptom'] ['headache', 'Sign or Symptom'] and ['shortness of breath', 'Sign or Symptom'] and ['sneezes', 'Sign or Symptom']

Symptom'] etc and for prevention wash your hands with warm water and soap keep yourself sanitized”