

Named Entity Annotation Guide

Group Number: FILL IN YOUR NUMBER FROM THE SPREADSHEET

Group Members: FILL IN YOUR GROUP MEMBER STUDENT UNIKEYS

The initial annotation guide was based on "An Annotated Dataset of Literary Entities," David Bamman, Sejal Popat and Sheng Shen, NAACL 2019.

We consider six types of entities: Person (PER), Facility (FAC), Geo-political entities (GPE), Location (LOC), Vehicle (VEH), Organisation (ORG). Note, we will not annotate 'Other' or 'Misc' entities, a practise that is used in some Named Entity annotations.

Each annotation is a labelled span, i.e., a start position, an end position, and a label. The format of annotation is:

One line per annotation, with the following style:

((line start, token start), (line end, token end)) - label

For example, if a Person entity spanned the first two tokens in the third line, you would have:

((2, 0), (2, 1)) - PER

For items that span a single token you may save it in one of two ways (either is acceptable):

((0, 5), (0, 5)) - PER

(0, 5) - PER

Note:

- The numbering starts from 0
- Tokens are specified by splitting the raw text file on whitespace
- Blank lines count when determining the line number

Annotations can nest, for example:

Text: "My sister's friend ..."

Annotations:

((0, 0), (0, 1)) - PER

((0, 0), (0, 2)) - PER

There are two annotations here because there are two people being mentioned: (1) my sister, and (2) my sister's friend.

Person (PER)

A single person indicated by a proper name (“Tom Sawyer”) or common entity (“the boy”); or set of people, such as “her daughters” and “the Ashburnhams”.

General examples:

my mother

Jarndyce

the doctor

a fool

his companion

Facility (FAC)

Functional, primarily man-made structure designed for human habitation (buildings, museums), storage (barns, parking garages), transportation infrastructure (streets, highways), and maintained outdoor spaces (gardens). Rooms and closets within a house are the smallest possible facility.

General examples:

the house

the room

the garden

the drawing-room

the library

Geo-political entities (GPE)

Single units that contain a population, government, physical location, and political boundaries. This includes not only cities that have known geographical locations within the real world (“London”, “New York”), or nations (“England”, “the United States”), but also both named and common imagined entities as well (“the town”, “the village”).

General examples:

London

England

the town

New York

the village

Location (LOC)

Entities with physicality but without political entities. This includes named regions without political organisation (“the balkans”, “the South”) and planets (“Mars”). It also includes geologically designated areas describing natural settings, such as “the sea”, “the river”, “the country”, “the valley”, “the woods”, and “the forest”.

General examples:

the sea

the river

the country

the woods

the forest

Vehicle (VEH)

A physical device primarily designed to move an object from one location to another.

General examples:

the ship

the car

the train

the boat

the carriage

Organisation (ORG)

Defined by the criterion of formal association, e.g. “the army”, “the University of Sydney” (note, as an administrative entity, distinct from the church as a facility with a physical location).

General examples:

the army

the Order of Elks

the Church

Blodgett College