Named Entity Recognition Annotation Guideline

Named Entity Recognition (NER) is a classification task that identifies phrases in a text that refer to entities or predefined categories (such as dates, person, organization and location names).

Example of NER: John [PERSON] is going to Kenya [LOCATION] on Monday [DATE]

For annotation, we will be following the Message Understanding Conference (MUC-6) named entity definition guideline for identifying Persons, organizations, locations, dates and times, see https://cs.nyu.edu/faculty/grishman/NEtask20.book_1.html

You may also use
https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-entities-guidelines-v6.6.pdf

Person Annotation [PER]

1. **Personal names** (including first names, middle names and last names) should be tagged **excluding titles** e.g Mr, Mrs, President, Professor, e.t.c. Examples:

President Donald Trump [PER]

Donald Trump Jr. [PER]

2. Family names should be tagged. Example:

The Kennedy [PER] family

Abiy Ahmed's [PER] wife.

Some tokenization may separate "Ahmed" and " 's", in that case, we have, For example in CoNLL format, where 'O' is the no-entity label.

Abiy, PER

Ahmed, PER

's, O

wife, O

3. **Personal names that refer** to an organization, location, events, law, disease, prizes **should not be tagged with PER tag**. Example,

St. Michael

Nobel Prize

George Washington University was established in 1821.

The White House is located in Washington.

Organization Annotation [ORG]

1. **Corporate designations** such as "Co." are organizations. Example:

   The Bridgestone Sports Co.

2. **Proper names that include** sports teams, stock exchanges, multinational organizations, political parties, unions, government parastatals. **Many of them comes in abbreviations** e.g WHO, NCDC, NASDAQ, EU, AU

   The World Health Organization [ORG] (WHO [ORG])

   AU [ORG]

3. **Proper names referring to facilities** e.g factories, hotels, universities, airports, hospitals, churches except (a) the name is clearly the name of the structure/place and not of an organization, or (b) the name is known to be the name of an organization but is used only in reference to the facility as a structure/place. Example:

The Frankfurt Airport Management has new rules for Aircraft.

I am traveling from the Frankfurt Airport …*will not be tagged since it appeared as a LOC*

I will stay at the IBIS Hotel [ORG]

The Supreme Court of Nigeria [ORG]

The convict will be taken to court.

The patient will be transferred to the General Hospital Ikoyi.

The patient is in the hospital.

Location Annotation [LOC] - - GPE and non-GPE

1. All country names,region names, state names and city names.

   United States of America [LOC]

   USA [LOC]

   Cape Town [LOC]

   Niger Delta Region [LOC]

2. **Multiword expressions** in which place names are separated by a comma are to be tagged as separate instances of LOCATION. Example:

   Cairo [LOC], Egypt [LOC]

   Washington [LOC] , D.C. [LOC]

3. **Do not label Nationalities and street addresses as locations.**

   An American is traveling to Abuja [LOC]

   A Nigerian hospital

   53140 Gatchell Road

4. Place names can sometimes be part of an organization. The annotation will follow these guidelines: (a) If there is a corporate designator, it marks the end of the organization name; (b) if there is no corporate designator, the "of " is part of the organization name. Example:

Hyundai of Korea, Inc. [ORG]

Hyundai, Inc.[ORG] of Korea [LOC]

John [PER] is working in BOSCH [ORG], Germany [LOC].

University of California [ORG] in Los Angeles [LOC]

5. Non-GPE locations like mountain name, river name, body of water should be tagged e.g

River Niger

DATE / TIME Annotation [DATE]

Tag all absolute and relative dates or periods, including days, months, years. We could combine DATE and TIME annotation as [DATE]

1. **Absolute date expressions are to be tagged**. Absolute date expression must indicate a specific segment of date i.e the particular day, season, financial quarters, years, decade or a particular century. Example:

Monday [DATE]

10th of October [DATE]

April 6, 2020 [DATE]

Summer [DATE]

the Autumn [DATE] report

Winter 2020 [DATE]

fourth quarter [DATE]

Third quarter of 1991 [DATE]

first half of the year [DATE]

1995 [DATE]

1980s [DATE]

19th century [DATE]

2. **Relative time expressions should also be tagged**. Example:

July last year [DATE]

this June [DATE]

next summer [DATE]

thirty days before the end of the year [DATE]

3. **Special days, such as holidays**, that are referenced by name should be tagged. Example:

because of the observance of All Saints' Day [DATE]

The Christmas day [DATE]

4. **Expression indicating periods between two dates should be tagged**. Example:

We are on vacation between July 1 and July 8 [DATE]

Her visit is from July 1 to July 8 [DATE]

Time: this refers to times smaller than a day.

5. **Absolute time expressions are to be tagged**. Absolute time expression must indicate a specific segment of time i.e a particular minute and hour. Example:

20 minutes after 10 [DATE]

midnight [DATE]

twelve o'clock noon [DATE]

noon [DATE]

5 p.m. EST [DATE]

5:40 [DATE]

6. Time expressions including the city time zone should be tagged. Example:

1:30 p.m. Chicago time [DATE]

7. Expression indicating periods between two times should be tagged. Example:

The election is from 2 p.m. to 4 p.m. [DATE]