

ILMT NER Tagset

1. Person (**NEP**)

- a. Salutations such as Ms., Mr., Shri., Smt., Dr., Sir, Padmashri etc., are not included in the Person entity
- b. Honorifics like Ji, Sir are also not part of the Person named entity
- c. For languages such as Telugu where honorifics like Garu is attached in the name token, they will be tagged PERSON
- d. Ex:- Shri NEP(Atal Vihari Bajpayee)
- e. Ex:- NEMI (Major) NEP(Dhyan Chand)
- f. Ex:- (Padma Vibhushan) NEMI Ustad NEP (Zakir Hussain) Ji is the chief guest for the event.
 - i. Here, the award names will be tagged as NEMI

2. Location (**NEL**)

- a. Location Names - Can also include Airport names
- b. Ex:- A cyclone is formed in the NEL(Arabian Sea).
- c. Ex:- If there are location specifiers such as chowk, nagar, lane, street, nadi (river), sagar (sea) in an entity, then include that specifier in the named entity
 - i. My friend lives in NEL(Indira Nagar, 2nd Lane).
 - ii. Buy sweets from NEL(Laal Chowk).
- d. Ex:- if there are direction specifiers to a location, then the whole phrase will be tagged as NEL
Elections are going to be held in NEL (Eastern and Southern Gujarat).

3. Organization (**NEO**)

- a. Names of Organizations - can include names of institutes, schools
- b. Companies, agencies, institutions, etc.
- c. Do not mark “high court” as NEO, if there is a

particular high court such as Hyderabad High Court, then mark it as NEO

- d. Ex:- NEO (International Institute of Information Technology, Hyderabad)
- e. k, Kanpur) - Even though it has an Abbreviation, we will tag it NEO as it is the name of an organization
- f. Ex:- NEO (National Thermal Power Corporation) (NEO (NTPC))
- g. Ex:- NEO (BJP), NEO (Bhajapaa)
- h. Ex:- He is a member of NEO (Indian National Congress) (NEO (INC)).
- i. My cousin studies in NEO (Shri Teg Bahadur University).

~~4. Date (**NEDA**)~~

- ~~a. Includes month names, years~~
- ~~b. In some case, need to be marked taking the whole sentence in context~~
~~Ex:- NEDA (21 march) ko shadi he, mein NEDA (20) ko pahunchunga~~
- ~~c. All words in a duration need to be marked separately as NEDA~~
~~Ex:- NEDA (January) - NEDA (February) NEDA (2020)~~
- ~~d. If there is a single date, we will tag the whole data as a single entity~~
~~(1st January, 2023) NEDA~~

5. Time, Day, & Date (**NETI**)

- a. Include the complete span of the time expression
- b. Include weekday names, time units, month names, years.
 - i. Ex:- I will reach at NETI (7 on Monday)
- c. Duration
 - i. I will be available NETI (from 5 to 7) for the meeting [Time duration]

- ii. We are going on vacation NETI (from 23rd January to 27th January). [Date Duration]
- iii. The cyclone will be active NETI (from 5 AM on 13th July to 1 PM on 14th July). [Date and Time duration]
- d. Relative time spans
 - i. NETI (Last Sunday), some friends visited my home.
 - ii. India witnessed a boom in the telecom sector NETI (last year).
 - iii. My friend interned at Google for NETI (3 months starting from this June).

6. Terms (**NETE**)

- ~~a. Domain terms in a domain will be marked NETE~~
- ~~b. Ex: In computer science domain~~
- ~~NETE (Databases) are designed to store information.~~
- ~~c. Ex: Do NETE (T3), NETE (T4), NETE (TSH) tests.~~

7. Artifacts (**NEAR**)

- a. An artifact is usually a simple object (such as a tool or ornament) showing human workmanship or modification as distinguished from a natural object. The artifacts can be both physical or virtual (software, online platforms)
- b. Can not be an abstract entity such as thought, ideal, philosophy
 - i. Marxism, Gandhivaad is not NEAR.
- c. Names of Books, journals, movies, TV shows, etc.
 - i. Ex:- Dr. NEP (APJ Abdul Kalam) penned the book "NEAR (Wings of Fire)".
- d. Name of softwares, online portals
 - i. I have NEAR (Windows XP) on my computer.
 - ii. NEAR (Mygov portal) gives the citizens direct access to raise their complaints.

- e. Poem titles, song titles, paper titles, slogans are not tagged as NEAR.
- f. Gadgets, sim card names
 - i. Palash uses an NEAR (Apple iphone).
 - ii. NEAR (Airtel SIM) is the cheapest in the market.
- g. Any monument
 - i. NEAR (Taj Mahal) is a wonderful monument.
- h. Any vehicle brand
 - Ex:- Ram drives a NEAR (Maruti Alto).

~~8. Number (**NEN**)~~

- ~~a. Any number mention will be marked NEN~~
- ~~b. Ex:- NEN(5 crore) rupees~~

9. Facility (NEF)

- a. This is a new tag which is introduced
- b. Previously, airports, ports, hospitals, financial institutions, educational institutions, railway stations, hotels, restaurants, press are tagged NEO, now they will be tagged NEF
- c. NEF (Rajiv Gandhi International Airport) got its new terminal this year.
- d. R&D Showcase is an annual event that showcases the projects and research at NEF (International Institute of Information Technology, Hyderabad).

~~10. URL (**NEU**)~~

- ~~a. Any url will be tagged as URL~~
- ~~b. Ex:- Please visit the link NEU (<https://www.iiit.ac.in>).~~
- ~~c. Ex: The padma awardees list can be accessed from NEU (<https://www.padmaawards.gov.in/padmaawardees.aspx>).~~

11. Miscellaneous (**NEMI**)

- a. All named entities which do not belong to the above 9 categories can be tagged miscellaneous

- b. Designations/Posts such as President, Prime Minister, Cabinet Minister, Major, Lieutenant, Justice
- c. Language names - Hindi, Tamil, English, Urdu, Kannada
- d. Script names - Devanagari, Tamil, Pali
- e. Award and competition names - Padma Bhushan, Bharat Ratna, Olympics, Asian Games, 14th Filmfare Award

Decisions Related to Named Entity Annotation

- 1. We will only tag a single level of Named Entities without any nesting.
- 2. Numbers will automatically be identified and tagged. Annotators need not mark the number (NEN) category.
- 3. Domain term marking will be done as a separate task. Therefore, the tag NETE has been dropped. In Governance domain (Yojana), all names of schemes/yojanas, policies, abhiyaans/movements will be tagged as domain terms. Therefore, these are not included in the named entity annotation.
- 4. Article names will not be tagged as NEMI.
- 5. Fruit and flower names are not included as named entities.
- 6. Cult groups, religion names, caste names are not named entities.
- 7. Forum names are not named entities.
- 8. Measurement units such as kilometer, liter, meter, c.m. are not named entities.