

# Named Entity Recognition

*Linguistic Resources for Natural Language Processing  
LM Language Technologies and Digital Humanities  
2024-25*

**Cristina Bosco**

# Named Entity Recognition

NER is the task of associating Named Entities, anything that can be referred to with a proper name, with a tag. The most commonly used tags are PERSON, LOCATION, ORGANIZATION and GEO-POLITICAL-ENTITY.

For example, it applies to:

Isaac Newton > PER

San Francisco > LOC

University of Turin > ORG

Italian Government > GPE

But often is used to dates, times and prices.

# Named Entity Recognition

NER consists in both finding the spans of text that correspond to the proper names and associating them the proper entity label.

Citing high fuel prices, [ORG **United Airlines**] said [TIME **Friday**] it has increased fares by [MONEY **\$6**] per round trip on flights to some cities also served by lower-cost carriers. [ORG **American Airlines**], a unit of [ORG **AMR Corp.**], immediately matched the move, spokesman [PER **Tim Wagner**] said. [ORG **United**], a unit of [ORG **UAL Corp.**], said the increase took effect [TIME **Thursday**] and applies to most routes where it competes against discount carriers, such as [LOC **Chicago**] to [LOC **Dallas**] and [LOC **Denver**] to [LOC **San Francisco**].

# Named Entity Recognition

NER is very useful as preliminary step for other more complex tasks.

To assign the **sentiment** to a message it can be relevant to know the nature of the entity towards which the sentiment is expressed, that is the target.

In **question answering**, it is crucial to identify the entities cited in the question and their nature to generate a proper answer.

# Named Entity Recognition

NER is very useful as preliminary step for other more complex tasks.

To assign the **sentiment** to a message it can be relevant to know the nature of the entity towards which the sentiment is expressed, that is the target.

In **question answering**, it is crucial to identify the entities cited in the question and their nature to generate a proper answer.



# Named Entity Recognition

NER can be difficult to perform when

- The name of the named entity is a compound, that is a multi-word-expression
- The name of the named entity is ambiguous

[PER Washington] was born into slavery on the farm of James Burroughs.  
[ORG Washington] went up 2 games to 1 in the four-game series.  
Blair arrived in [LOC Washington] for what may well be his last state visit.  
In June, [GPE Washington] passed a primary seatbelt law.

# Named Entity Recognition

The most used approach to address the NER challenges is called BIO tagging.

It consists in marking in the exact span of text in which the named entity is cited.

This can be done marking the beginning (B), the internal part(s) (I) and the end of each named entity (O) (the first token which is out of the named entity span).

In these days, [Elton]**B** [John]**I** [was]**O** in [New]**B** [York]**I** [for]**O** a concert.

# Named Entity Recognition

[PER Jane Villanueva] of [ORG United] , a unit of [ORG United Airlines Holding] , said the fare applies to the [LOC Chicago ] route.

Words	BIO Label
Jane	B-PER
Villanueva	I-PER
of	O
United	B-ORG
Airlines	I-ORG
Holding	I-ORG
discussed	O
the	O
Chicago	B-LOC
route	O
.	O



# Named Entity Recognition

The advantage of the BIO tagging is that it associates un label to each element of the named entity, also when the named entity is composed by more than one token.

A similar approach is applied on multi-word expressions in PoS tagging.