



CS221.N11.KHTN – NLP

8.3 Named Entities & Named Entity Tagging

Presenter: Duy Dat Ng

Authors: Duy Dat Ng, The Viet Le, Huu Khoa Tran, Hoang Vu Huynh



Named Entity



A named entity is anything that can be referred to with a proper name.



Janet



Stanford University



Colorado



Named Entity Recognition (NER)



The NER is to find spans of text that constitute proper names and tag the type of the entity.



PER (person)



ORG (organization)



LOC (location)



Named Entity Recognition



GPE (geo-political entity)



However, the term named entity also includes dates, times, kinds of temporal expressions and numerical expressions like prices.

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**].



Many applications will also need to use specific entity types like proteins, genes, commercial products, or work of art.

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	Turing is a giant of computer science.
Organization	ORG	companies, sports teams	The IPCC warned about the cyclone.
Location	LOC	regions, mountains, seas	Mt. Sanitas is in Sunshine Canyon .
Geo-Political Entity	GPE	countries, states	Palo Alto is raising the fees for parking.

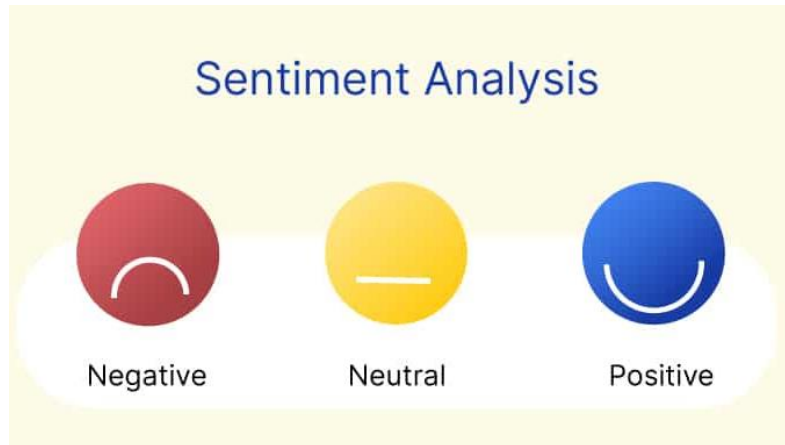
Figure 8.5 A list of generic named entity types with the kinds of entities they refer to.



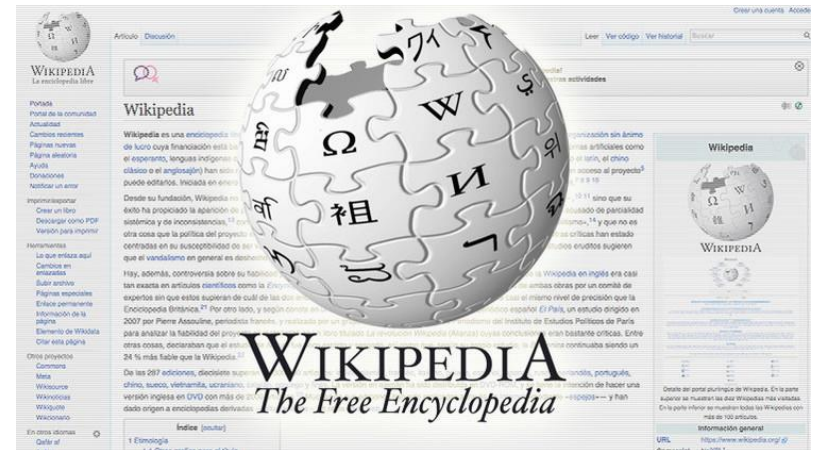
Named Entity Recognition



NER is a useful first step in lots of NLP tasks.



Sentiment Analysis



Wikipedia



Limitations



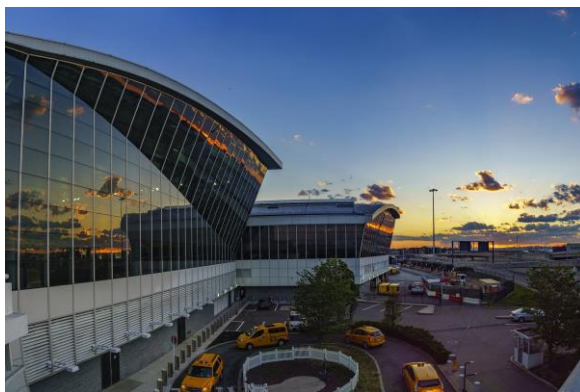
Examples of type ambiguities

[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.

Figure 8.6 Examples of type ambiguities in the use of the name *Washington*.



person



airport

JFK



bridge



school



IO, BIO, BIOES taggings



Examples of tagging formats

Jane Villanueva of United Airlines Holding discussed the Chicago route.

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

Figure 8.7 NER as a sequence model, showing IO, BIO, and BIOES taggings.



Other example for BIO format

Học sinh Trường THPT Hà Nội – Amsterdam học tại Đại học Công nghệ Thông tin.

