



Named Entities in NLP

Introduction & Experiments

Siyao Logan Peng



Named Entities

The essence of a document can often be captured by the **entities** and **relations** that are mentioned

- **Entity**: object, person, organization, date, etc.
- **Relation**: property that links one or several entities

Among all entities, **named entities** are often the most important for:

- Text summarization
- Question answering
- Information retrieval
- Sentiment analysis

Named Entities (NEs)

- NE involves identification of **proper names** in texts, and classification into a set of predefined categories.
- Three universally accepted categories: **person, location and organisation**
- **Persons**: entities including humans, individuals or groups.
- **Location**: entities including geographical entities such as geographical areas like names of countries, cities, continents and landmasses, bodies of water, and geological formations.
- **Organization**: entities including to corporations, agencies, and other groups of people defined by an established organizational structure

Example: [U.N.]ORG official [Ekeus]PER heads for [Baghdad]LOC .

NER output

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**].

More on PER

- first, middle and last names of people,
- animals and fictional characters
- Aliases

These fine-grained definitions are from the CoNLL 2003 corpus:

<https://www.cnts.ua.ac.be/conll2003/ner/annotation.txt>

More on ORG

- Organizations: companies (press agencies, studios, banks, stock markets, manufacturers, cooperatives)
- subdivisions of companies (newsrooms)
- brands
- political movements (political parties, terrorist organisations)
- government bodies (ministries, councils, courts, political unions of countries)
- publications (magazines, newspapers, journals)
- musical companies (bands, choirs, opera companies, orchestras)
- public organisations (schools, universities, charities)
- other collections of people (sports clubs, sports teams, associations, theaters companies, religious orders, youth organisations)

More on LOC

- roads (streets, motorways)
- trajectories
- regions (villages, towns, cities, provinces, countries, continents, dioceses, parishes)
- structures (bridges, ports, dams)
- natural locations (mountains, mountain ranges, woods, rivers, wells, fields, valleys, gardens, nature reserves, allotments, beaches, national parks)
- public places (squares, opera houses, museums, schools, markets, airports, stations, swimming pools, hospitals, sports facilities, youth centers, parks, town halls, theaters, cinemas, galleries, camping grounds, NASA launch pads, club houses, universities, libraries, churches, medical centers, parking lots, playgrounds, cemeteries)
- commercial places (chemists, pubs, restaurants, depots, hostels, hotels, industrial parks, nightclubs, music venues)
- assorted buildings (houses, monasteries, creches, mills, army barracks, castles, retirement, homes, towers, halls, rooms, vicarages, courtyards)

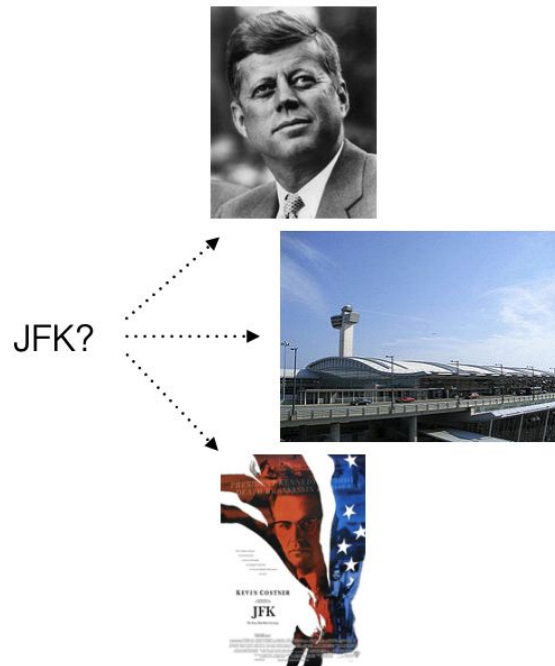
More on MISC

- words of which one part is a location, organisation, miscellaneous, or person
- adjectives and other words derived from a word which is location, organisation, miscellaneous, or person
- Religions
- political ideologies
- Nationalities
- Languages
- programs
- events (conferences, festivals, sports competitions, forums, parties, concerts)
- wars
- sports related names (league tables, leagues, cups)
- titles (books, songs, films, stories, albums, musicals, TV programs)
- Slogans
- eras in time
- types (not brands) of objects (car types, planes, motorbikes)

Ambiguity in NER

NER systems often have to deal with several important types of ambiguity:

- **Reference resolution:** the same name can refer to different entities of the same type. For instance, JFK can refer to a former US president or his son.
- **Cross-type Confusion:** the identical entity mentions can refer to entities of different types. For instance, JFK also names an airport, several schools, bridges, etc.



Entity Type Ambiguity

[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.

Now, we will start with
the fun Mentimeter experiments!
Hooray!

Disclaimer: by participating in the mentimeter
experiments, you agree that the data could be used
for research purposes.

[England] won the World Cup

England won the [World Cup]

The years I spent as soccer team manager of [the Republic of Ireland] were the best years of my life.

[White House] spokesman Mike McCurry said Clinton plans to have regular news conferences during his second term.

[ALPINE] SKIING - WOMEN 'S WORLD CUP
SUPER G WINNER PROFILE .

I bear witness that there is no [God] .

Snow said the plane last reported to air traffic control at about 3 A.M. local time / 1:30 A.M.

[EST]

UK bookmakers [William Hill] said on Friday
they have lengthened the odds of a
Conservative victory

"William Hill" a British gambling company

The granddaughter of Italy's [Fascist] dictator
Benito Mussolini

“Fascism is a far-right, authoritarian,
ultranationalist political ideology and
movement” – from Wikipedia

Barbarians captain Rob Andrew , the man who kicked [Australia] to defeat with a last-ditch drop-goal in the World Cup quarter-final in Cape Town .

LA CLIPPERS AT [NEW YORK]

[LA CLIPPERS] AT NEW YORK

He will not be considered for a test match against [Australia] starting on October 10

Now German/Bavarian

Nacha kimmt im #BR no' amoi de
#Starkbierprobe am [#Nockherberg].

Wissts es no , beim [BER] drobm , was hamma
mia da glacht Aber ob's Minga mit da zwoatn
Stammstreckn etzad besser macht ?

Is wie bei „ [Lucifer] “ : schert si ned um Andere
 , is Egozentrisch , is Narzisstisch , glaubt , Nix
 geht ohne Eam , leugnet Probleme Anderer

De Entscheidung fia [Atlanta] hot aba aa dem IOC Kritik eibrocht , do's infoigedessn ois kaifle erschiena is , de gwinnorientiertn Mitglieda vom IOC ham se gega de de traditionella Wert vo de olympischn Bewegunga vapflichtatin duachgsetzt .

@pachlerroman Ein wuida Hund , unsa
[Roman] . Owa gfoit ma , de Leifbälänz da
Jugend testn. 👍 😊
An Guadn Moang auf Wean. ☕

Zua Bekanntgob ham des
Organisationskomitee und de Regierung vo
Griechnland de Feia zum 90.
Jubiläum vo de 1896a Spui am 15. Aprui 1986
im oidn **[Athena Parlamentsgebäude]** gnutzt .

Kurze Abkühlung aufm Rückweg von [da Post]

.

[Erdinger Weißbier] (mit " ß " 👍) geng
Erdinger Urweisse (mit " ss " 👎) Heit oomdz
geht's auf ! 🙌 <https://t.co/XqabAY7J3Y>

Comments?
Questions?
Thank you!

References

Slides are borrowed from:

- <https://course.khoury.northeastern.edu/cs6200sp15/slides/m07.s08%20-%20named%20entity%20recognition.pdf>
- <https://cs.uwaterloo.ca/~ppoupart/teaching/cs886-fall13/slides/cs886-lecture12.pdf>