NAMED ENTITY RECOGNITION (NER)

Spoorthi Podduturi
What is NER?

Locate and classify elements in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced using the new discovery of nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell–Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.
Why Named Entity Recognition?

No longer feasible for human beings to process enormous data to identify useful information.

Named entity itself may be the answer to a particular question (Question-Answering)
Categories:

- XML style format described in the Message Understanding Conference (MUC).
  - "ENAMEX" tags are used for names,
  - "NUMEX" tags are used for numerical entities, and
  - "TIMEX" tags are used for temporal entities.

Unlike Robert, John Briggs Jr contacted Wonderful Stockbrockers Inc in New York and instructed them to sell all his shares in Acme.

NER involves two tasks: find and classify names in text, for example:

The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.
The measure behaves a bit funnily for IE/NER when there are **boundary errors** (which are common):

First **Bank of Chicago** announced earnings...

- This counts as both a fp and a fn
- Selecting nothing would have been better
Three Standard Approaches of NER

- Hand-made Rule-based NER
- Machine Learning-based NER and
- Hybrid NER.
Advantages:

- Named entities can be indexed, linked off, etc
- Sentiment can be attributed to companies or products
- A lot of IE relations are associations between named entities
- For question answering, answers are often named entities.

Disadvantages:

- Entities may firstly be difficult to find, and once found, difficult to classify.
- Making distinctions based on world knowledge is difficult. (e.g., location or Person)
Future Scope:

- NER for different languages can be improved
- Entity identification in the molecular biology, bioinformatics can be improved
- Reducing the annotation labor by employing semi-supervised learning.
Sources


Questions??