Coreference Resolution

Co-reference Resolution (CR) is the task of deciding whether two entity mentions refer to the same instance or not.

For example in:

> Joe Biden appeared at the event at 8pm. The president was wearing a Louis Vuitton Tuxedo.

The objective is to identify that Joe Biden and The president are the same entity.

Coreference Detection is related to Relationship Extraction (RE) - in fact you could even say that CR is a special case of RE in the sense that we are interested in the special relationship between entity mentions when they both refer to the same entity.

In-Document Coreference Resolution

This is the "normal" CR case in which you're trying to resolve mentions of entities within the same document e.g. a single news article.

Approaches

- **2022-10-23** A recent blog post from explosion / spaCy shows how they have implemented end-to-end CR in their excellent NLP pipeline but as of writing they do not provide a trained model and they require you to have a copy of the Ontonotes dataset.

Cross-Document Coreference Resolution

Cross-Document Coreference Resolution (CDCR) is when you try to link named entity references across multiple input documents. A use case might be identifying that a number of news articles do actually refer to the same person (e.g. "Joe Biden", "The President").

CDCR is challenging because there are so many possible entities and thus $O(n^2)$ comparisons to make between candidates.

Approaches

- In 2021 we proposed CD^2CR - a CDCR approach across documents and domains that allows us to match mentions of people, places, technologies etc across scientific papers and news articles that discuss them.

Revision #5
Created 2022-10-23 06:59:01 UTC by James
Updated 2024-02-08 21:05:14 UTC by James