Technical Review

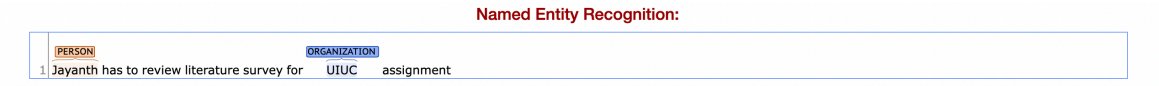
Usage of Named Entity Recognition in Financial Domain

Jayanth Chandra
NetID - jc101
5 November 2022

Introduction

Named entity recognition (NER) is an NLP based technique to identify mentions of rigid designators from text belonging to particular semantic types such as a person, location, organisation etc.

You can consider this example where I mention that I have to complete the assignment. The NER toolkit helps in distinguishing and characterising each word so that it belongs to particular entity.



Entities are the most important chunks of a particular sentence such as noun phrases, verb phrases, or both. Generally, Entity Detection algorithms are ensemble models of :

- Rule-based Parsing, python
- Dictionary lookups,
- POS Tagging,
- Dependency Parsing.

NER can automatically scan entire articles and pull out some fundamental entities in a text and classify them into predefined categories. Entities may be,

- Organizations,
- Quantities,
- Monetary values,
- Percentages, and more.
- People's names
- Company names
- Geographic locations (Both physical and political)
- Product names

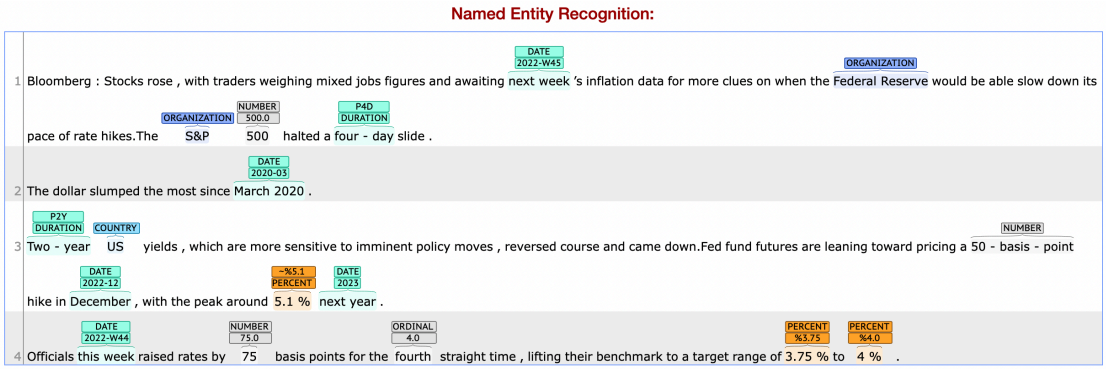
- Dates and times
- Amounts of money
- Names of events etc

Building a highly accurate NER algorithm requires a vast understanding of math, machine learning & image processing. Alternatively, using popular frameworks like PyTorch and Tensorflow and a few pre-trained models, we can build a Named Entity Recognition algorithm from custom data.

Financial Use Cases of NER

Unstructured text contains abundant information; the challenge is to find what's relevant. An estimated 80-90% of financial data is unstructured and the ability to analyze and act on this data presents a huge untapped opportunity. Unlocking the potential from unstructured data begins with recognizing and tagging entities within the data.

Consider the financial information from Bloomberg describing about the financial environment where we have applied NER on top of it.



- By extracting information from documents, such as numbers from financial reports, and connecting firm names to other content databases, NER enables you to parse data from documents and accelerate the pace and

scope of content collection. Businesses can use NER to build a structured database in order to organise and analyse unstructured data.

- Data extraction from massive amounts of PDFs and webpages is laborious, time-consuming, and prone to human mistake in the private markets and loans arena.Assessing profitability and credit risk can be facilitated and accelerated by using NER to tag and categorise pertinent data in order to extract information.
- You may track social media trends and extract entities from Twitter and Reddit discussions with the aid of NER. Reddit and Twitter have the potential to play a major factor in stock price changes, and NER can assist in locating the businesses, persons, and locations discussed on social media.
- NER can be a starting point for creating effective search algorithms. All documents can be subjected to NER in order to extract entities related to them and store them separately. The next time a user searches for a phrase, the search term would be matched with a reduced list of entities in each document, leading to a faster search execution. This is also one of the way to help in personalisation/recommendations of financial products.

References

a> https://economictimes.indiatimes.com/
b> https://corenlp.run/
c> https://insight.factset.com/financial-use-cases-for-named-entity-recognition-ner
d> https://nanonets.com/blog/named-entity-recognition-with-nltk-and-spacy/