

Named Entity Resolution

Named Entity

- Named Entity is a Real world object which can be Denoted through a proper name.
- Named entity is a phrase that clearly identifies one items from a set of other items that have similar attributed.
- Named entities are often mined for marketing initiatives.
- Named entity can be first and last names, geographic locations, ages, addresses, phone numbers, companies and addresses persons, organizations, countries, places.

Named Entity

- Examples of Named Entity

1. Sachin played a spectacular match at the Eden Gardens today.
2. Github has a free and open source entity resolution software repository named Zingg.

Named Entity Recognition

- **Named entity recognition (NER)** is a subfield of Natural Language Processing that focuses on identifying and classifying specific data points from textual content.
- NER works with salient details of the text, known as *named entities* — single words, phrases, or sequences of words — by identifying and categorizing them into predefined groups.
- The categories encompass a diverse range of subjects present within the text, including individuals' names, geographic locations, organizational names, dates, events, and even specific quantitative values such as money and percentages.

Key Concepts of Named Entity Recognition

1. **POS tagging.** Standing for "part-of-speech tagging," this process assigns labels to words in a text corresponding to their specific part of speech, such as adjectives, verbs, or nouns.
2. **Corpus.** This is a collection of texts used for linguistic analysis and training NER models. A corpus can range from a set of news articles to academic journals or even social media posts.
3. **Chunking.** This is an NLP technique that groups individual words or phrases into "chunks" based on their syntactic roles, creating meaningful clusters like noun phrases or verb phrases.

Key Concepts of Named Entity Recognition

4. **Word embeddings.** These are dense vector representations of words, capturing their semantic meanings. Word embeddings translate words or phrases into numerical vectors of fixed size, making it easier for machine learning models to process.

NER Example

- Consider the sentence: "**Mary** from the HR department said that The **Ritz London** was a great hotel option to stay in **London**."
- "Mary" is labeled as PERSON, indicating that it is an entity representing a person's name.
- "The Ritz" is tagged as ORG, which stands for *Organization*. This means it is recognized as an entity that refers to companies, agencies, institutions, etc.
- "London" has been classified as GPE, which stands for *Geopolitical entity*. GPEs represent countries, cities, states, or any other regions with a defined boundary or governance.

Named Entity Recognition

- NER helps in understanding text, question answering, grouping together relevant information about entities for news, analysis etc.

Named Entity Resolution

- **Named Entity Resolutions** on the other hand is linking the same entity in different records where a common identifier is missing.
- **Named Entity Resolution** works on structured text in most cases, like customer or company records, though it may also be applied to long texts like product names and descriptions.
- Suppose there is a hospital registering a patient. Typical information collected would be first and last names, address, telephone number and date of birth.

Named Entity Resolution

- **Named Entity Resolution** is typically done through rule based systems, though recently a lot of work has happened on the ML/AI based approaches including deep learning.
- **Named Entity Resolution** is a fundamental technique used in medical research, census data analysis, CRM record deduplication, householding and other analytics and data quality functions.

Named Entity Resolution

- Suppose there is a hospital registering a patient. Typical information collected would be first and last names, address, telephone number and date of birth.
- First Name: Anne
- Last Name: Smith
- Address: 123, Milwauke Dr, Connecticut
- Phone: (123) 456 7890