



Search Medium



Write



# NLP-Entity Coreference Resolution



Sarang Mete · Following

Published in Analytics Vidhya · 2 min read · Nov 1, 2022



28



...

Explore multiple libraries





Photo by [Tim Mossholder](#) on [Unsplash](#)

## Definition:

*Coreference Resolution* resolves situation where two or more words within a text refer to the same entity. The entities resolved may be a person, place, organization, or event. [Read more..](#)

## Applications:

Machine Translation, information extraction, question answering, summarization

## Training:

Coreference is pairwise classification task. Coreference resolution typically requires a pre-processing pipeline comprising a variety of NLP tasks (e.g., tokenization, lemmatization, named entity recognition, part-of-speech tagging). We will not discuss how to do training. You can read about how hugging face has developed neuralcoref [here](#) and [here](#).

## Libraries for application:

1. [huggingface/neuralcoref](#)

2. [allennlp](#):

It has some drawbacks and solution to these is [here](#)

3. [Stanford Core NLP](#)

4. [coreferee](#): mixture of neural networks and programmed rules

I've used coreferee.

Usage:

```
text = "Tom was in Texas with his wife and he said that he liked it"
doc = nlp(text)

doc._.coref_chains.print()

0: Tom(0), his(5), he(8), he(11)
1: Texas(3), it(13)
```

In above example 'it' resolved to 'Texas' but 'Texas' resolved to None because Texas is original entity.

# Basic logic to get resolved text:

1. Get all token words

2. Check resolution of each token

3. If it is None, it means it's original entity or there is nothing to resolve, so ignore None

4. If it is not None, then replace token with resolved/original entity

5. Combine all tokens back to get resolved text

```
doc = nlp(text)
# Get token list
tok_list = list(token.text_with_ws for token in doc)
print(tok_list)
for index, _ in enumerate(tok_list):
```

```
#Check resolution of each token
if None != doc._.coref_chains.resolve(doc[index]):
    new_token = ""
    #If it is not None, then replace token with resolved/original entity
    for resolved_token in doc._.coref_chains.resolve(doc[index]):
        new_token = new_token + resolved_token.text + " "
    tok_list[index] = new_token
resolved_text = "".join(tok_list)
```

Tom was in Texas with Tom wife and Tom said that wife liked Texas

I've created a complete end to end project . You can refer it [here](#).

## Entity Coreference Resolution

Enter text for Entity Coreference Resolution

Although he was very busy with his work, Peter had had enough of it. He and his wife decided they needed a holiday. They travelled to Spain because they loved the country very much."

**Submit**

model\_output:

Although Peter was very busy with Peter work, Peter had had enough of work . Peter and Peter wife decided Peter wife needed a holiday. Peter wife travelled to Spain because Peter wife loved the Spain very much."

Image by Author

The main challenges I've solved in this project:

1. Create processing logic for coreferee output

## 2. Create production ready code.

If you liked the article or have any suggestions/comments, please share them below!

Let's connect and discuss on [LinkedIn](#)

Entity Resolution

NLP

Streamlit

Spacy

Knowledge Graph

### More from the list: "NLP"

Curated by Himanshu Birla



Jon Gi... in Towards Data ...

#### Characteristics of Word Embeddings



. 11 min read . Sep 4, 2021



Jon Gi... in Towards Data ...

#### The Word2vec Hyperparameters



. 6 min read . Sep 3, 2021



Jon Gi... in

#### The Word2ve



. 15 min rea

[View list](#)



## Written by Sarang Mete

162 Followers · Writer for Analytics Vidhya

Following

Senior NLP Engineer

### More from Sarang Mete and Analytics Vidhya



 Sarang Mete

## NLP-Semantic search using elasticsearch and embeddings

Semantic Search

3 min read · Nov 14, 2022

18  1 

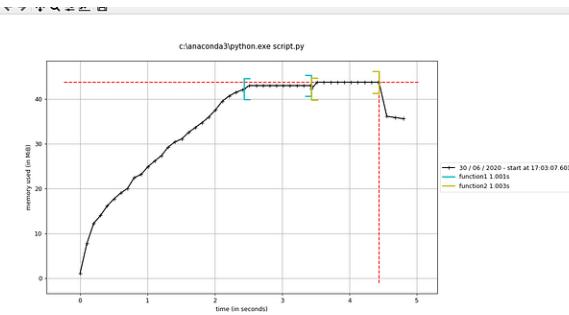


 Sajal Rastogi in Analytics Vidhya

## Wifi -Hacking using PyWifi

4 min read · Feb 6, 2021



Kia Eisinga in Analytics Vidhya

## How to create a Python library

Ever wanted to create a Python library, albeit for your team at work or for some open...

7 min read · Jan 27, 2020



2K

24



...



Sarang Mete in Analytics Vidhya

## MLOps-Calculate Memory Consumption of Python Code

Python developers should be aware of the memory consumption of the code they are...

3 min read · Dec 16, 2021



10

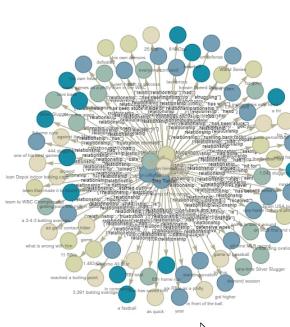
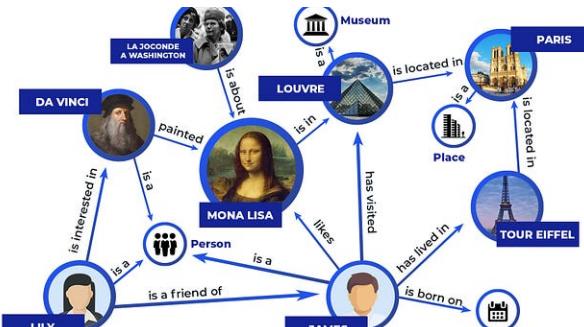


...

[See all from Sarang Mete](#)

[See all from Analytics Vidhya](#)

## Recommended from Medium



 Alla Chepurova in DeepPavlov

## Improving Knowledge Graph Completion with Generative LM...

Combining both internal LM knowledge and external data from KG

13 min read · Sep 5



36



...



Wenqi Glantz in Better Programming

## 7 Query Strategies for Navigating Knowledge Graphs With...

Exploring NebulaGraph RAG Pipeline with the Philadelphia Phillies

★ · 17 min read · 4 days ago



501



4



...

## Lists



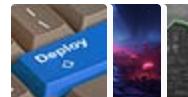
### Natural Language Processing

669 stories · 283 saves



### The New Chatbots: ChatGPT, Bard, and Beyond

13 stories · 133 saves



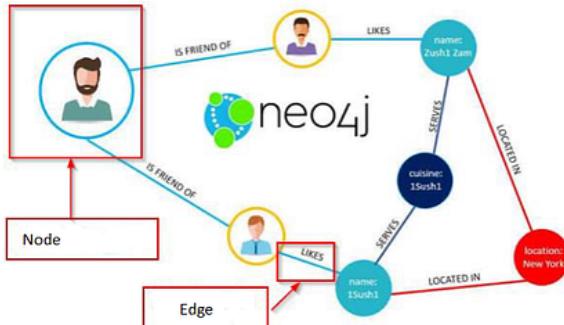
### Predictive Modeling w/ Python

20 stories · 452 saves



### New\_Reading\_List

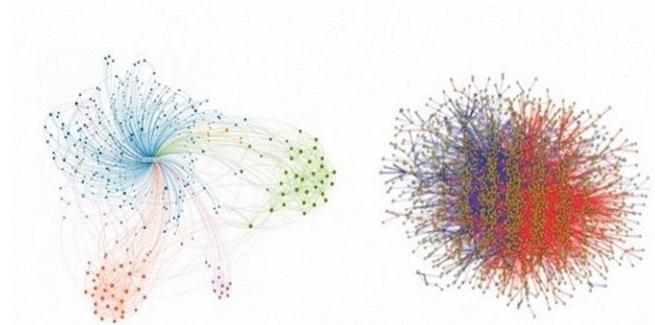
174 stories · 133 saves


 Serbu Claudiaa

## How to use Neo4j

A graph-based database management system represents an alternative to the...

9 min read · Sep 15


 Lei Wang in graphscope

## Graphs and Graph Applications

In this post, we will introduce basic concepts of graphs, and some typical applications of...

7 min read · May 22



Artificial Intelligence (AI ORG), an ever-evolving field, has witnessed remarkable growth since its inception. Dating back to the Dartmouth Conference ORG in 1956 DATE, AI ORG has emerged as a multidisciplinary domain encompassing machine learning, natural language processing (NLP ORG), computer vision, and robotics. Recent breakthroughs, like the introduction of deep learning techniques in the early 2010s DATE, have accelerated AI ORG advancements. Tech giants like Google ORG, IBM ORG, and Microsoft ORG have invested heavily in AI ORG research and development. Significant milestones include the landmark victory of IBM ORG's Deep Blue ORG over Garry Kasparov PERSON in 1997 DATE and the emergence of voice assistants like Apple ORG's Siri in 2011 DATE. AI ORG continues to shape industries across healthcare, finance, and transportation, fueling innovation and transforming the way we live and work.



 Sahithi Reddy

## Named Entity Recognition using Python and Spacy.

NER, short for Named Entity Recognition, is an NLP method that detects and categorizes...

4 min read · Jun 11



 Oxdevshah in AI Skunks

## Knowledge Graphs: A Comprehensive Analysis

Knowledge graphs are becoming increasingly important in NLP due to their ability to mode...

8 min read · Apr 9

See more recommendations