Machine Learning Tutorial   Data Analysis Tutorial   Python - Data visualization tutorial   NumPy   Pandas   Op

Python | Named Entity Recognition (NER) using spaCy

Named Entity Recognition (NER) is a standard NLP problem which involves spotting named entities (people, places, organizations etc.) from a chunk of text, and classifying them into a predefined set of categories. Some of the practical applications of NER include:

- Scanning news articles for the people, organizations and locations reported.
- Providing concise features for search optimization: instead of searching the entire content, one may simply search for the major entities involved.
- Quickly retrieving geographical locations talked about in Twitter posts.

NER with spaCy

spaCy is regarded as the fastest NLP framework in Python, with single optimized functions for each of the NLP tasks it implements. Being easy to learn and use, one can easily perform simple tasks using a few lines of code.

Installation :

```
pip install spacy
python -m spacy download en_core_web_sm
```

Code for NER using spaCy.

```python
import spacy

nlp = spacy.load('en_core_web_sm')

sentence = "Apple is looking at buying U.K. startup for $1 billion"

doc = nlp(sentence)

for ent in doc.ents:
    print(ent.text, ent.start_char, ent.end_char, ent.label_)
```

Output

```
Apple 0 5 ORG
U.K. 27 31 GPE
$1 billion 44 54 MONEY
```

In the output, the first column specifies the entity, the next two columns the start and end characters within the sentence/document, and the final column specifies the category.

Further, it is interesting to note that spaCy's NER model uses capitalization as one of the cues to identify named entities. The same example, when tested with a slight modification, produces a different result.

```python
import spacy

nlp = spacy.load('en_core_web_sm')

sentence = "apple is looking at buying U.K. startup for $1 billion"

doc = nlp(sentence)

for ent in doc.ents:
    print(ent.text, ent.start_char, ent.end_char, ent.label_)
```

Output

```
U.K. 27 31 GPE
$1 billion 44 54 MONEY
```

The word "apple" no longer shows as a named entity. Therefore, it is important to use NER before the usual normalization or stemming preprocessing steps.

One can also use their own examples to train and modify spaCy's in-built NER model. There are several ways to do this. The following code shows a simple way to feed in new instances and update the model.

```python
import spacy
from spacy.gold import GoldParse
from spacy.language import EntityRecognizer

nlp = spacy.load('en', entity = False, parser = False)
```

```python
doc_list = []
doc = nlp('Llamas make great pets.')
doc_list.append(doc)
gold_list = []
gold_list.append(GoldParse(doc, [u'ANIMAL', u'O', u'O', u'O']))

ner = EntityRecognizer(nlp.vocab, entity_types = ['ANIMAL'])
ner.update(doc_list, gold_list)
```

By adding a sufficient number of examples in the doc_list, one can produce a customized NER using spaCy.

spaCy supports the following entity types:
PERSON, NORP (nationalities, religious and political groups), FAC (buildings, airports etc.), ORG (organizations), GPE (countries, cities etc.), LOC (mountain ranges, water bodies etc.), PRODUCT (products), EVENT (event names), WORK_OF_ART (books, song titles), LAW (legal document titles), LANGUAGE (named languages), DATE, TIME, PERCENT, MONEY, QUANTITY, ORDINAL and CARDINAL.

References

- https://spacy.io/

Don't miss your chance to ride the wave of the data revolution! Every industry is scaling new heights by tapping into the power of data. Sharpen your skills and become a part of the hottest trend in the 21st century.

Dive into the future of technology - explore the Complete Machine Learning and Data Science Program by GeeksforGeeks and stay ahead of the curve.

Last Updated : 18 Jun, 2019

4

Previous

Next

Share your thoughts in the comments

Add Your Comment

Similar Reads

Named Entity Recognition in NLP	Named Entity Recognition

NLP	Named Entity Chunker Training	Python	Word Similarity using spaCy
Python	PoS Tagging and Lemmatization using spaCy	Python	Perform Sentence Segmentation Using Spacy
Install Specific Version of Spacy using Python PIP	HTML Cleaning and Entity Conversion	Python
Python program to create dynamically named variables from user input	Python - Move Files To Creation and Modification Date Named Directories

A　　Anannya …

Article Tags :　　Machine Learning ,　Python ,　Python Programs

Practice Tags :　Machine Learning,　python

Company

About Us

Legal

Careers

In Media

Contact Us

Advertise with us

GFG Corporate Solution

Placement Training Program

Explore

Hack-A-Thons

GfG Weekly Contest

DSA in JAVA/C++

Master System Design

Master CP

GeeksforGeeks Videos

Geeks Community

Languages

Python

Java

C++

PHP

GoLang

SQL

R Language

Android Tutorial

Tutorials Archive

DSA

Data Structures

Algorithms

DSA for Beginners

Basic DSA Problems

DSA Roadmap

Top 100 DSA Interview Problems

DSA Roadmap by Sandeep Jain

All Cheat Sheets

Data Science & ML

Data Science With Python

Data Science For Beginner

Machine Learning Tutorial

ML Maths

Data Visualisation Tutorial

Pandas Tutorial

NumPy Tutorial

NLP Tutorial

Deep Learning Tutorial

HTML & CSS

HTML

CSS

Web Templates

CSS Frameworks

Bootstrap

Tailwind CSS

SASS

LESS

Web Design

Django Tutorial

Python Tutorial

Python Programming Examples

Python Projects

Python Tkinter

Computer Science

Operating Systems

Computer Network

Database Management System

Web Scraping

OpenCV Tutorial

Python Interview Question

Software Engineering

Digital Logic Design

Engineering Maths

DevOps

Git

AWS

Docker

Kubernetes

Azure

GCP

DevOps Roadmap

Competitive Programming

Top DS or Algo for CP

Top 50 Tree

Top 50 Graph

Top 50 Array

Top 50 String

Top 50 DP

Top 15 Websites for CP

System Design

High Level Design

Low Level Design

UML Diagrams

Interview Guide

Design Patterns

OOAD

System Design Bootcamp

Interview Questions

JavaScript

JavaScript Examples

TypeScript

ReactJS

NextJS

AngularJS

NodeJS

Lodash

Web Browser

Preparation Corner

Company-Wise Recruitment Process

Resume Templates

Aptitude Preparation

Puzzles

Company-Wise Preparation

School Subjects

Mathematics

Physics

Chemistry

Biology

Social Science

English Grammar

World GK

Management & Finance

Management

HR Management

Finance

Income Tax

Organisational Behaviour

Free Online Tools

Typing Test

Image Editor

Code Formatters

Code Converters

Currency Converter

Marketing	Random Number Generator
	Random Password Generator

More Tutorials

Software Development

Software Testing

Product Management

SAP

SEO - Search Engine Optimization

Linux

Excel

GeeksforGeeks Videos

DSA

Python

Java

C++

Data Science

CS Subjects