Food for thought

Train a NER model

Serena Sensini @ Strive School

Abstract

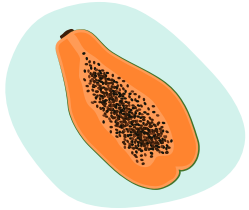
Issue

How to teach a system to recognise new entities such as food?

Topics

- What is spaCy
- What is NER
- How does it work
- Live coding

Table of Contents



INTRO

Let’s define

STEP-BY-STEP

Let’s analyze

LIVE CODING

Let’s work

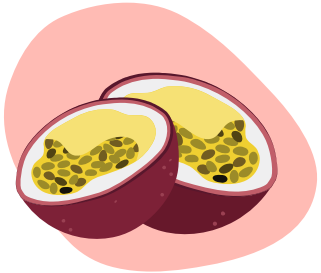
Who Am I

SERENA SENSINI

Enterprise Architect @ Dedalus spa

Founder of TheRedCode.it

Author of 4 tech books



Introduction

What hell is NER

NLP

aka *natural language processing*

- Subfield of AI that aims to process and analyze natural language data
- Examples
  - Speech recognition, natural language understanding and generation

Just an example

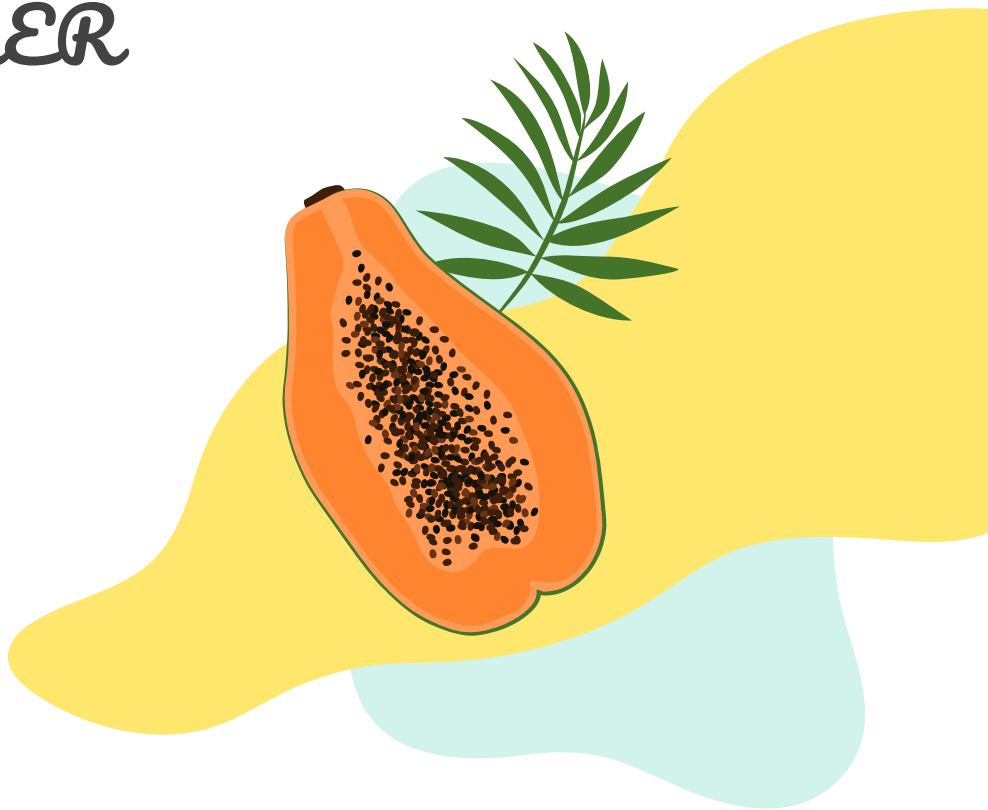
Human-machine interaction

- A human talks to the machine;
- The machine acquires audio;
- The audio is converted to text;
- The text is processed;
- A response is processed;
- The machine responds to the human by playing the audio file resulting from his response.

NER

aka: *Named Entity Recognition*

- process of automatic identification of the entities present in a text and consequent classification



spaCy

- Library designed for production use that helps you build applications that process and “understand” text
- Counting on 64+ languages support
- Different trained models to achieve different tasks

How it should work

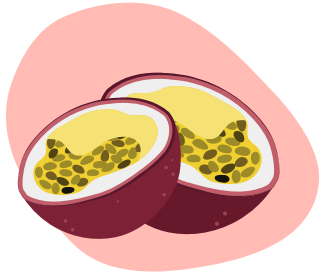
FOOD

Don't try to make pizza and pineapple!

How it works

FOOD

Don't try to make pizza and pineapple!



Step-by-step

How do I do it

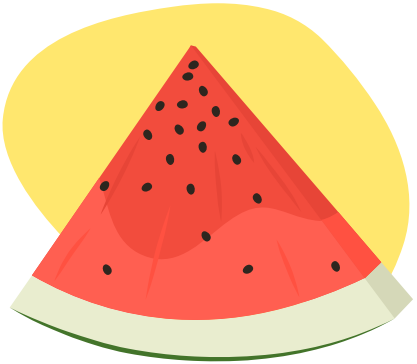
How it works

LOAD MODEL

```
import spacy

nlp = spacy.load('en_core_web_lg')

with open('text.txt') as file:
    article = file.read()
```



APPLY IT

```
doc = nlp(article)
for ent in doc.ents:
        print(ent.text, ent.label_
```

Example

PERSON DATE ORG

When John started working driving cars at Google in 2007...

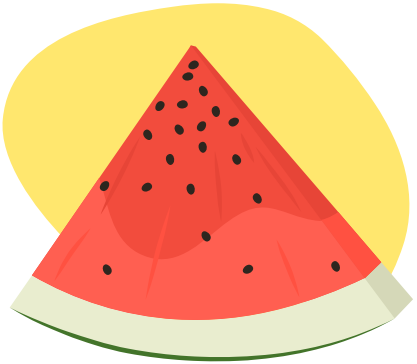
How to update it

LOAD MODEL

```
import spacy

nlp = spacy.load('en_core_web_sm')

with open('text.txt') as file:
    article = file.read()
```



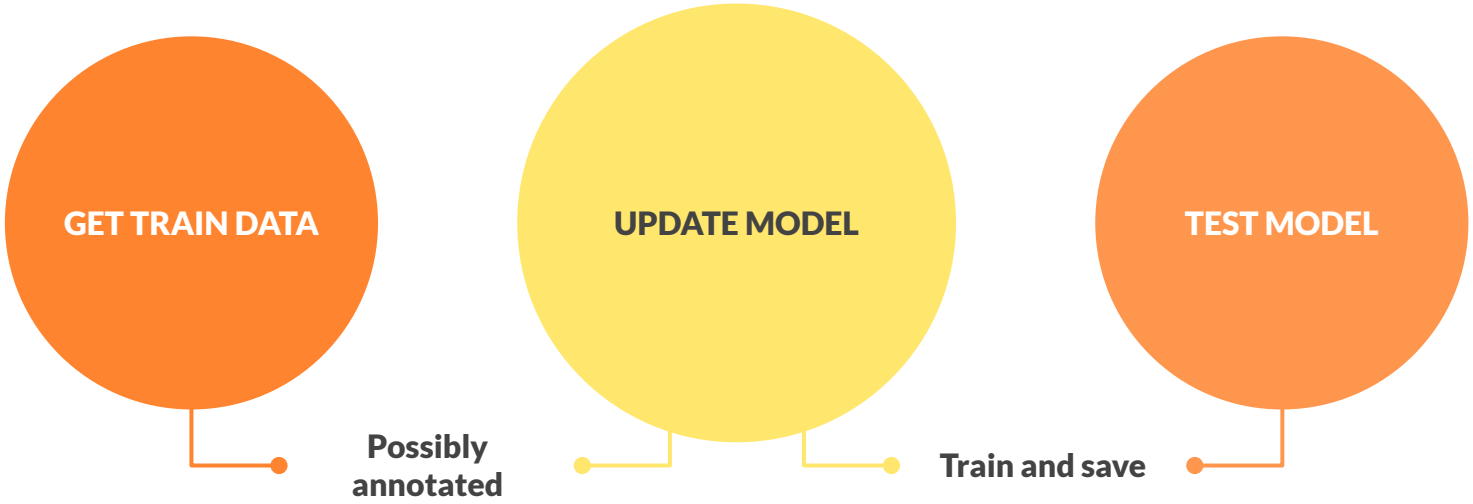
UPDATE IT

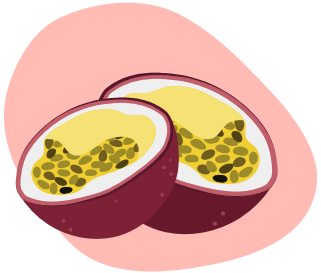
```
?
```

Ingredients

- Python 3.x installed in your system
- Trained language model installed (here the instruction)
- Your favorite IDE
- A dataset

Recipe


GET TRAIN DATA
UPDATE MODEL
TEST MODEL
Possibly annotated
Train and save



Live coding

Hands off!

Whoa!

That could be hard!

Thanks!

ANY DOUBT, QUESTION OR...?

Reach me on LinkedIn!
info@theredcode.com
https://theredcode.it