

Parts of Speech

Natural Language Processing Bootcamp

- **Section Goals**
 - Understand how to retrieve Parts-of-Speech using Spacy
 - Understand how to use Named Entity Recognition with Spacy
 - Visualize POS and NER
 - Perform Sentence Segmentation

Parts of Speech Basics

Natural Language Processing Bootcamp

- Most words are rare, and it's common for words that look completely different to mean almost the same thing.
- The same words in a different order can mean something completely different.

Natural Language Processing Bootcamp

- Even splitting text into useful word-like units can be difficult in many languages.
- While it's possible to solve some problems starting from only the raw characters, it's usually better to use linguistic knowledge to add useful information.

Natural Language Processing Bootcamp

- That's exactly what spaCy is designed to do: you put in raw text, and get back a **Doc** object, that comes with a variety of annotations.

Natural Language Processing Bootcamp

- In this lecture we'll take a closer look at coarse POS tags (noun, verb, adjective) and fine-grained tags (plural noun, past-tense verb, superlative adjective).

Natural Language Processing Bootcamp

- Keep in mind, English is an incredibly complex language with many rules (and even more exceptions to rules).
- Teaching English is outside the scope of this course, but check out the resource references to learn more about various POS tags and what they mean in English.

Natural Language Processing Bootcamp

- Let's explore Part of Speech (POS) tags in more detail with Spacy!

Visualizing POS

Natural Language Processing Bootcamp

- We just saw how to extract POS tags from tokens and obtain information about those POS.
- Let's review how to visualize Parts of Speech structure with **displacy**.
- Pay attention to “dep”

Named Entity Recognition

Natural Language Processing Bootcamp

- Named-entity recognition (NER) seeks to locate and classify named entity mentions in unstructured text into pre-defined categories such as the person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

Natural Language Processing Bootcamp

- Our goal is to read in raw text such as:
 - Jim bought 300 shares of Acme Corp. in 2006.
- And add additional NER information:
 - [Jim]_{Person} bought 300 shares of [Acme Corp.]_{Organization} in [2006]_{Time}.

Natural Language Processing Bootcamp

- Let's explore NER with Spacy, as well how to add our own custom entities.

Named Entity Recognition

Part Two

Natural Language Processing Bootcamp

- We've seen how to add a single term as our own NER (e.g. adding "Tesla").
- But what if we have several terms to add as possible NERs?
- In this continued lecture, we will go over how to add in multiple phrases as NERs.

Natural Language Processing Bootcamp

- For example, if we are working with a vacuum company, we might want to add both **vacuum cleaner** and **vacuum-cleaner** as PROD (product) NERs.
- Let's see how we can do this.

Visualizing NER

Natural Language Processing Bootcamp

- Let's review how to visualize NER with Spacy and Displacy.

Sentence Segmentation

Sentence Segmentation

- In Spacy Basics we saw briefly how Doc objects are divided into sentences.
- In this lecture we'll learn how sentence segmentation works, and how to set our own segmentation rules to break up docs into sentences based on our own rules.