What is NLP?

Definitions and Explanation

Definitions

- **Natural language**: language that evolved naturally by humans to use to communicate with each other. Mostly used in contrast with constructed languages like Dothraki (a language that was made up for a fictional group of people to speak in a television show) or programming languages like Python. English and Spanish are examples of natural languages.
- **Natural Language Processing**: a branch of artificial intelligence that deals with computers understanding, speaking, and writing natural language.

DISCUSSION

- How do the sites we played with fit the definition of NLP?
- What are some natural languages you know of? Non-natural languages?

Definitions

- **Stop words**: words that do not add to the meaning of a text, like “the”
- Older NLP algorithms deleted stop words from their models. Now that we have more computing power, and more data, people have started putting stop words back in.
- There is no one complete list of stop words; different Python libraries use different lists.

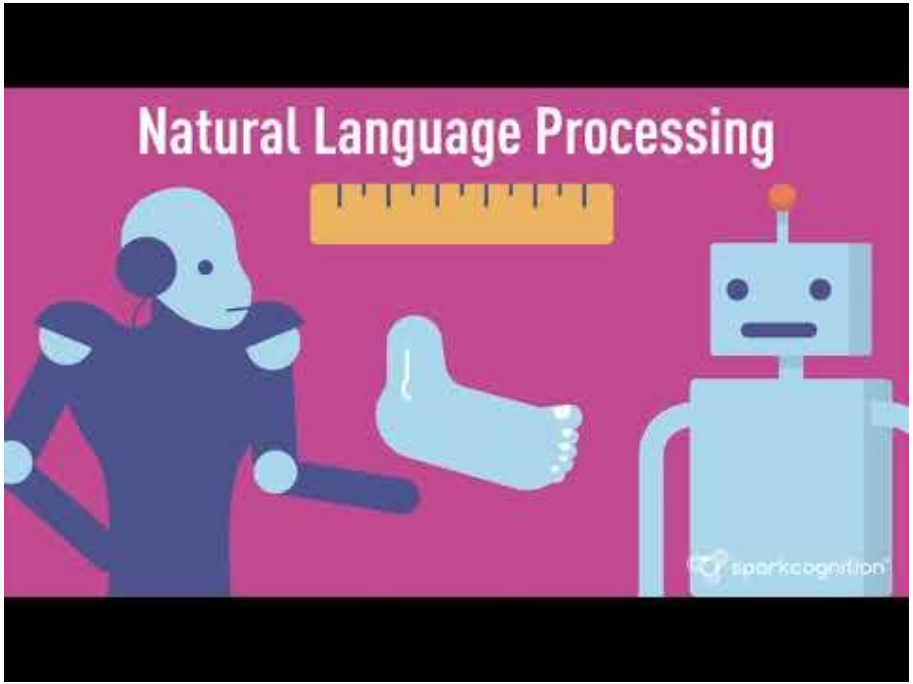
DISCUSSION

- What are some words you think should be stop words?
  - Why might it be better to either take them out or put them in?

Definitions

- **Matrix** (plural matrices): a grouping of numbers laid out in a table
    - Computers turn words into data they understand and store that data in matrices. (You'll probably learn more about matrices in a later Math class; this is all you need to know about them for now.)
- **Vector**: a set of numbers used to describe a position. In NLP, vectors are used to map the position of words in a space of hundreds or thousands of dimensions.
- **Count Vector**: a way of describing words as numbers based on where the word appears in sentences.

Basics of Natural Language Processing



Basics of NLP - Video Overview

- Computers like to deal with structured data. Most data is in text form rather than structured formats that are easy for computers to understand.
- Computers can't "read" texts, but they can find patterns in text. Computers find patterns by turning huge amounts of text into groups of numbers called matrices.
- NLP works by having computers count how many times groups of words are together.
- Some groups of words happen together often in a few documents. This means there's something going on with those documents and they might have a similar topic.
    - Ex: if there are a lot of documents about oil rigs dropping items centering around the Gulf of Mexico, we know something must be happening in the Gulf of Mexico.

Comprehension Questions

Summarize what you learned from the video

- How do humans and computers read differently?
- Without NLP, what can a computer “read”?
- How does a computer “understand” text?
- Why is NLP useful?

Exit ticket: Unit 1.02 - What is NLP?

Describe what Natural Language Processing (NLP) is in your own terms.