

Procesamiento de Lenguajes Naturales - NLP

El procesamiento de lenguajes naturales (NLP) es una área de las ciencias de la computación y la inteligencia artificial que se encarga de las interacciones entre los lenguajes de los ordenadores y los humanos (natural). NLP se utiliza para aplicar modelos de ML a texto y lenguaje.

Ejemplo:

Enseñar a una máquina a entender palabras del idioma escrito y hablado es el objetivo fundamental del NLP. Cuando dictamos algo a nuestro iPhone o Android, nuestra voz se convierte a texto y ahí es cuando los algoritmos de NLP entran en acción.

La historia del NLP empieza desde 1950, aunque existe trabajo encontrado desde periodos anteriores. En 1950, Alan Turing publicó el artículo [Computing machinery and intelligence](#) el cual proponía lo que hoy llamamos [test de Turing](#) como criterio de inteligencia.

Hasta los años 80, la mayoría de sistemas de procesamiento de lenguajes naturales se basaban en conjuntos de reglas manuscritas. Desde finales de la década de los 80, hubo una revolución en el mundo del NLP con la introducción de algoritmos de Machine Learning para procesar el lenguaje natural.

https://es.wikipedia.org/wiki/Procesamiento_de_lenguajes_naturales

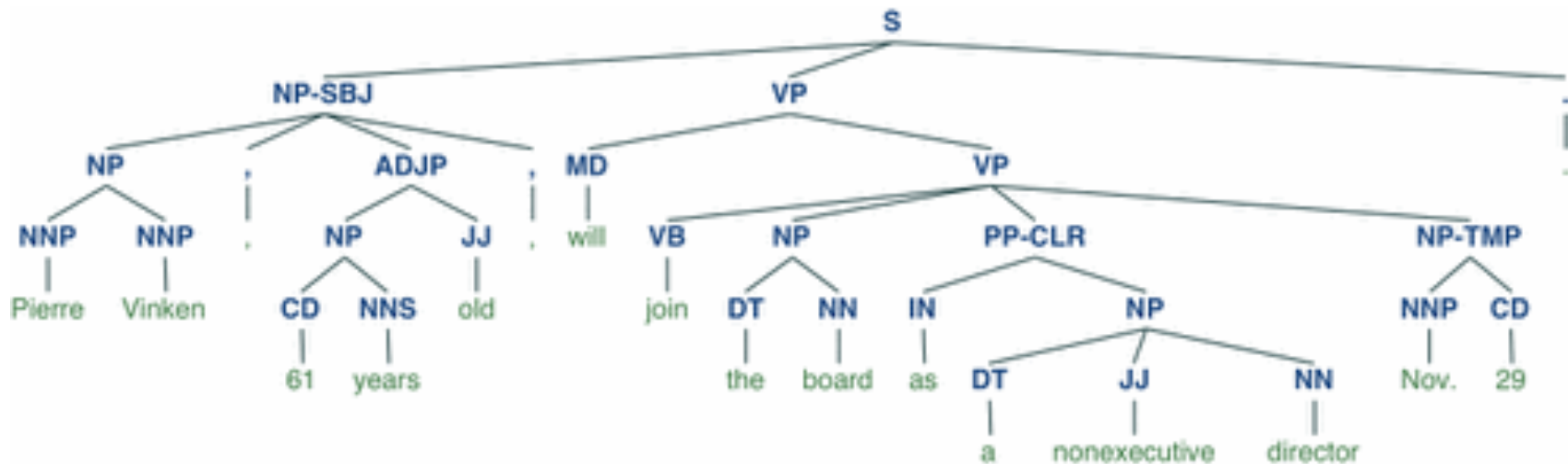
Usos del NLP

- Análisis de sentimiento. Identificar el sentimiento o opiniones subjetivas en grandes cantidades de texto incluyendo sentimiento promedio y minería de opiniones
- Predecir el género de un libro
- Responder preguntas automáticamente (chatbots)
- Traductor automático y reconocimiento de habla
- Resumen de documentos

Librerías principales de NLP:

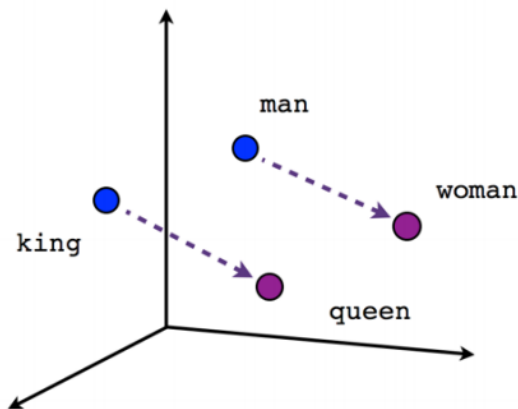
- Natural Language Toolkit (NLTK)
- SpaCy
- Stanford NLP
- OpenNLP

NLP

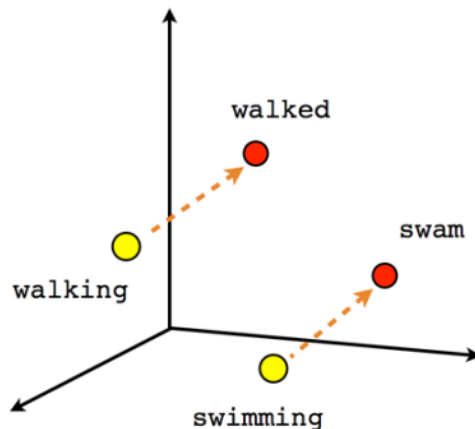


<http://www.nltk.org>

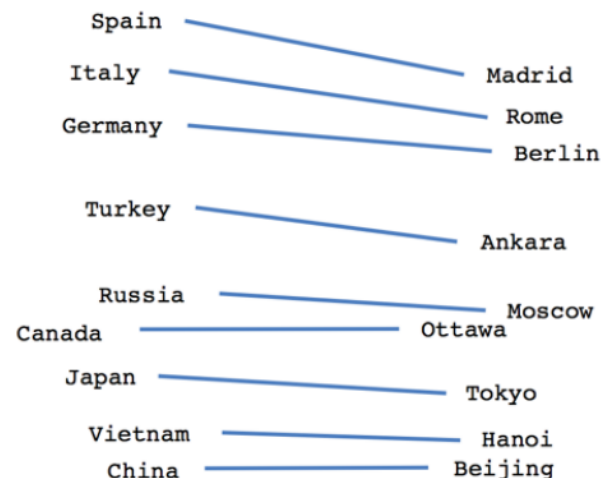
NLP



Male-Female



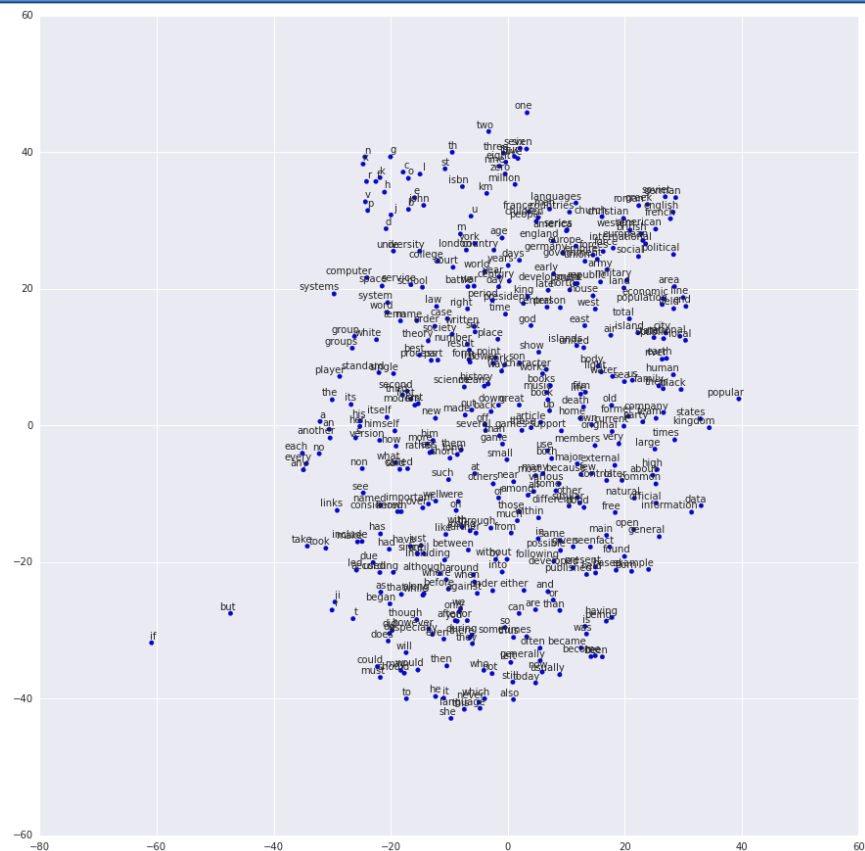
Verb tense



Country-Capital

<https://www.tensorflow.org/tutorials/representation/word2vec>

NLP



NLP - Bag of Words

Modelo de NLP muy popular. Se utiliza para pre procesar textos a ser clasificados antes del proceso de ajustar el modelo a las observaciones que contienen algún tipo de texto.

Involucra dos aspectos:

1. Un vocabulario de palabras conocidas
2. Una medida de la presencia de las palabras conocidas

En esta parte del curso nos centraremos en aprender a:

1. Limpiar textos y prepararlos para aplicarles algoritmos de ML
2. Crear un modelo de Bag of Words
3. Aplicar modelos de ML a nuestra Bag of Words