

# Introducción

## Curso de procesamiento de lenguaje natural

Julio Weissman

Maestría en Tecnologías de la Información  
Universidad Nacional de Misiones

9 de marzo de 2018



# Algo acerca de mi

- Profesor-Investigador Universidad de Sonora
- Doctorado en *Sistemas Automáticos* INPT
- Grupo de trabajo en Ciencia de Datos
- Algunos proyectos de investigación
- Algunos proyectos de vinculación

# ¿Que es PLN?

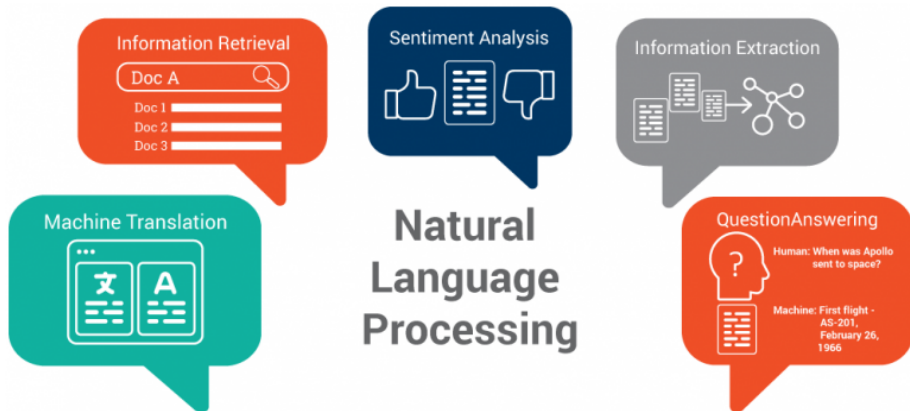
- *Segun Wikipedia*

*Campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre las computadoras y el lenguaje humano.*

- *Segun Copestake*

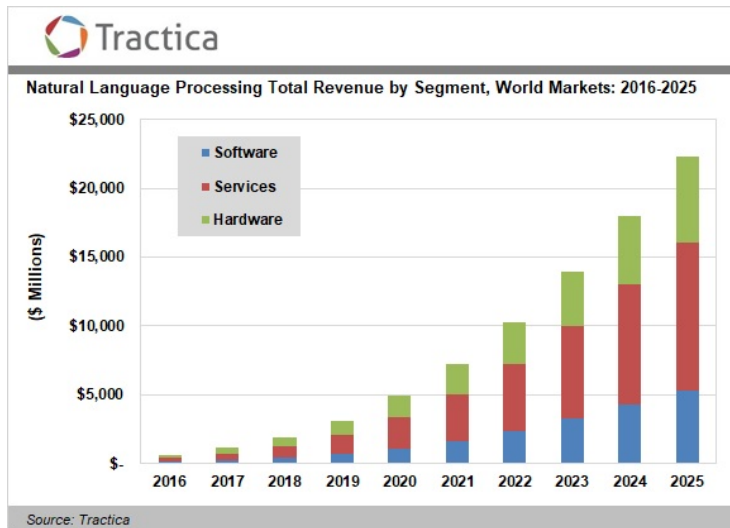
*Procesamiento automático (o semi-automático) del lenguaje humano*

- La mayoría de los autores de libros de texto (y cursos) más que definirlo lo contextualizan.
- Tambien se conoce como *Lingüística computacional*.



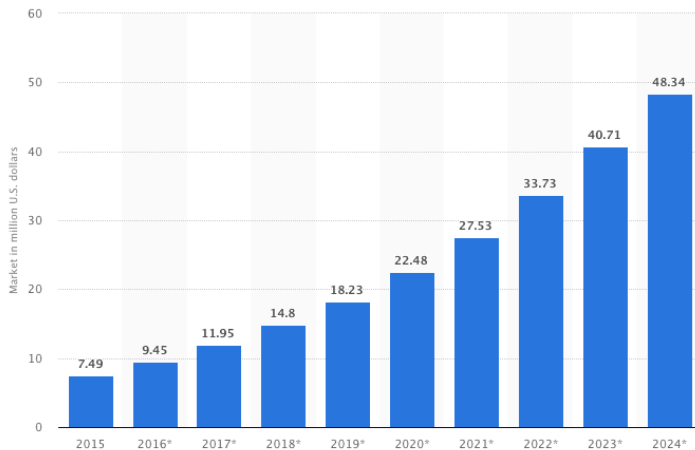
Fuente <https://ontotext.com/top-5-semantic-technology-trends-2017/>

# ¿Porqué estudiar PLN?



Fuente <https://www.tractica.com/newsroom/press-releases/natural-language-processing-market-to-reach-22-3->

# ¿Porqué estudiar PLN?



Fuente: <https://www.statista.com/statistics/607931/latin-america-natural-language-processing-market-revenues/>

- Basado en reglas
  - Expresiones regulares
  - Gramáticas libres de contexto
- Probabilista
  - Modelos de  $n$ -gramas
  - Clasificadores lineales
  - Métodos basados en maximización de la verosimilitud
- Aprendizaje profundo
  - Redes convolucionales
  - Redes recurrentes

# ¿Que esperar de este curso?

- 1 Una noción general del PLN
- 2 Una idea general de las técnicas y herramientas más populares para las tareas de PLN seleccionadas en el curso.
- 3 Conocimiento superficial de *como* funcionan algunos de los modelos más importantes (o más populares) de PLN
- 4 Conocer y saber aplicar las herramientas existentes en *python* para casos de estudio



- 1 Normalización de la información
- 2 Clasificación de documentos
- 3 Modelos probabilistas secuenciales
- 4 Modelos de semántica distribuida
- 5 Modelado de tópicos

- Explicación general de los temas
- Ejemplos de aplicación y/o modelado al final de cada explicación
- Trabajos prácticos para reforzar cada uno de los temas principales
- Un proyecto final que permita enfrentarse a la búsqueda y/o generación de datos

