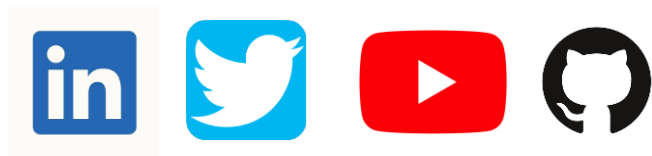




Introducción a NLP

Juan Manuel Gutiérrez Henao

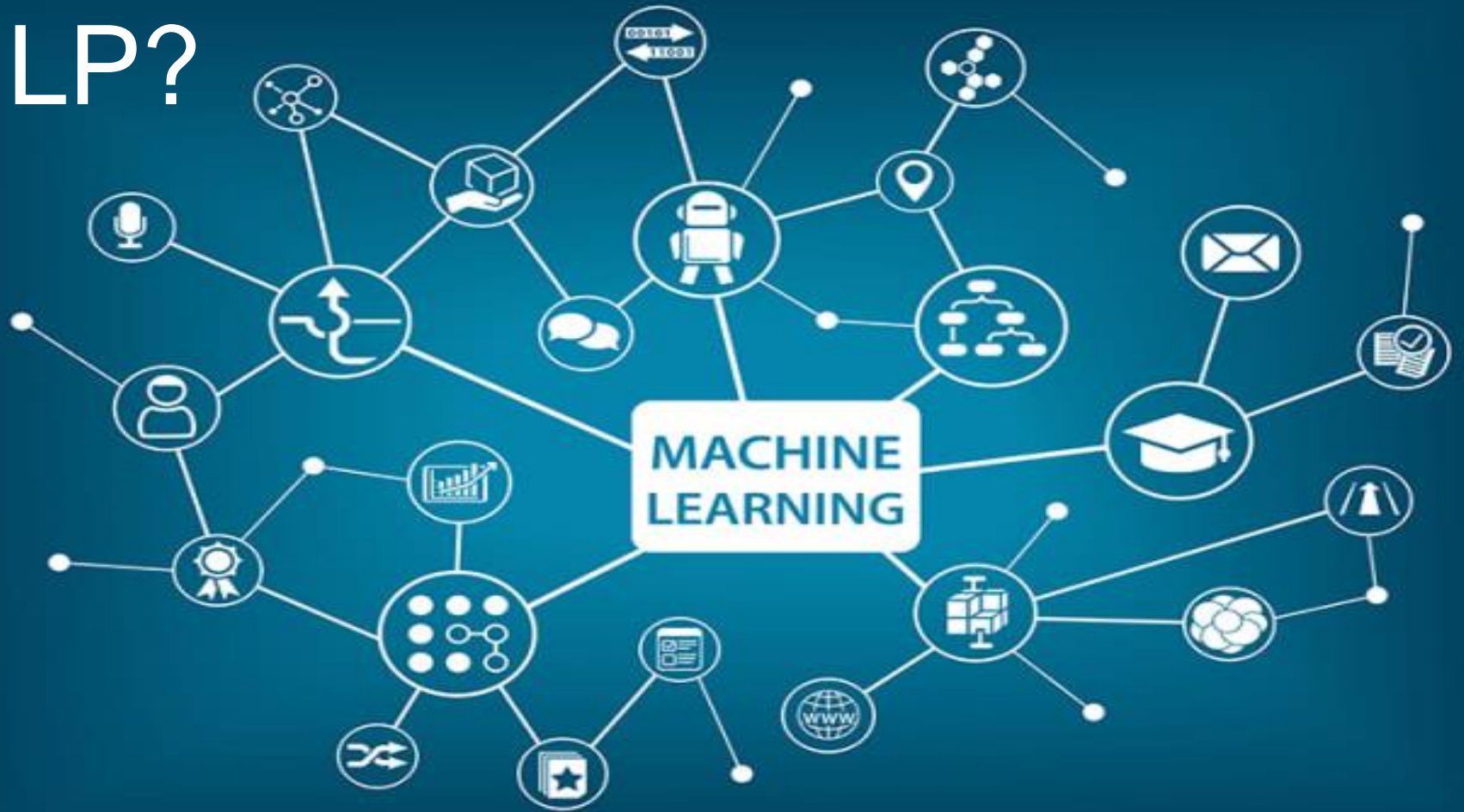




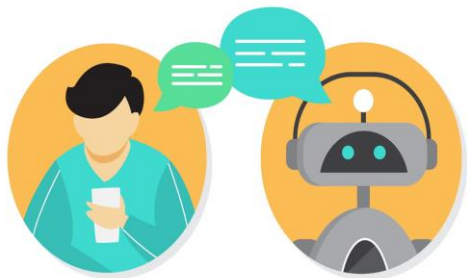
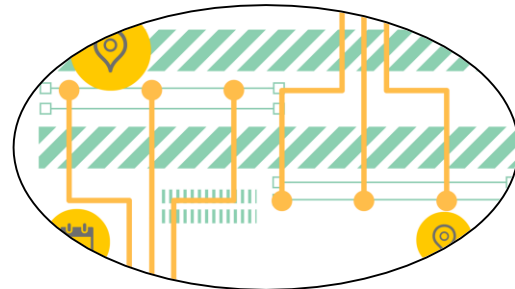
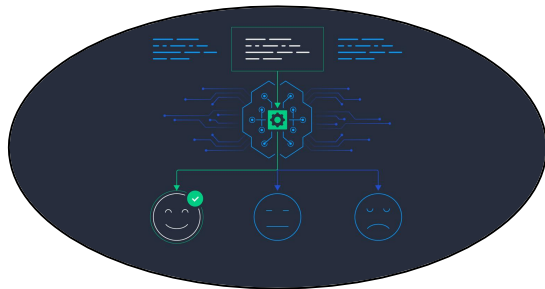
Transformar el mundo a través
de la Inteligencia Artificial...



NLP?



¿Qué se puede hacer con NLP?





¿Cómo representar el texto
en forma numérica?

Representaciones basadas en conteos

- Conteos
- Binarios
- TF-IDF



- Fáciles



- Semántica
- Dimensionalidad

	data	science	...	cambridge	spark
document 1	5	0	...	1	0
document 2	0	10	...	0	3
...
document 49	20	13	...	5	5
document 50	0	0	...	0	0

Representaciones semánticas

- LSA
- Word2Vec
- GloVe



- Semántica
- Dimensionalidad
- Transfer learning



- Contexto

		royalty ↓	femininity ↓	intelligence ↓
王	king	0.9	-0.9	0.5
女王	queen	0.9	0.9	0.5
男人	man	0.1	-0.9	0.5
女人	woman	0.1	0.9	0.5
聪明的	smart	0.5	0	0.87
智能的	intelligent	0.5	0	0.9

*Words are simply discrete states,
and we are simply looking for the
transitional probabilities between
those states: the likelihood that
they will co-occur*

Representaciones contextuales

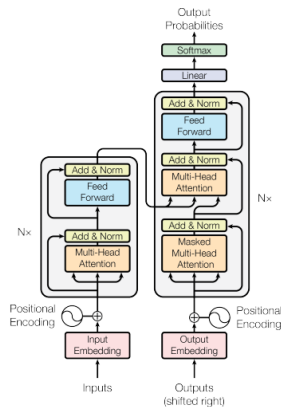
- BERT (2018/10)
- GPT-2 (2019/02)
- XLNET (2019/06)



- Contexto
- Zero-shot learning
- SotA



- Complejos
- Fine-tuning

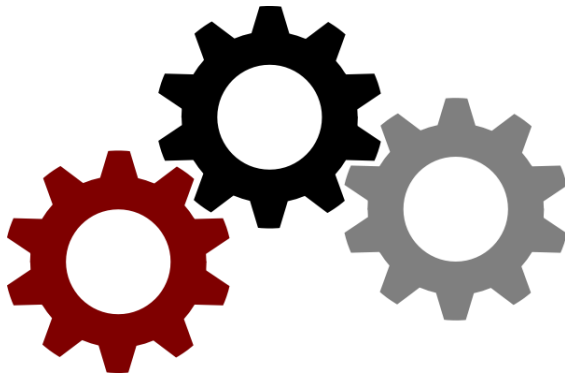


*Language models are
unsupervised multitask learners*

Preprocesamiento de Texto

Normalización

- Casing
- Acentos
- Números
- Stopwords
- Stemming
- Lemmatizing
- Colocaciones



Aumentación

- Parts-of-Speech
- Dependency parsing
- Chunking

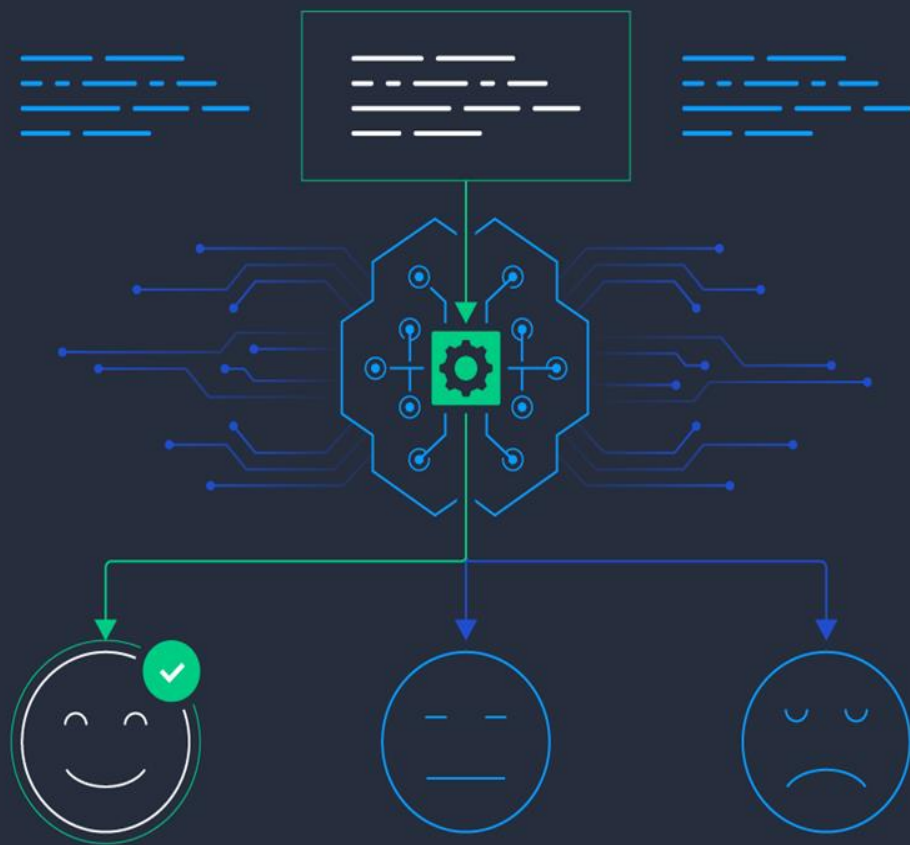


Existen muchas formas de representar la información de texto.

Nosotros somos los responsables de decidir cómo representarla.

Aplicaciones

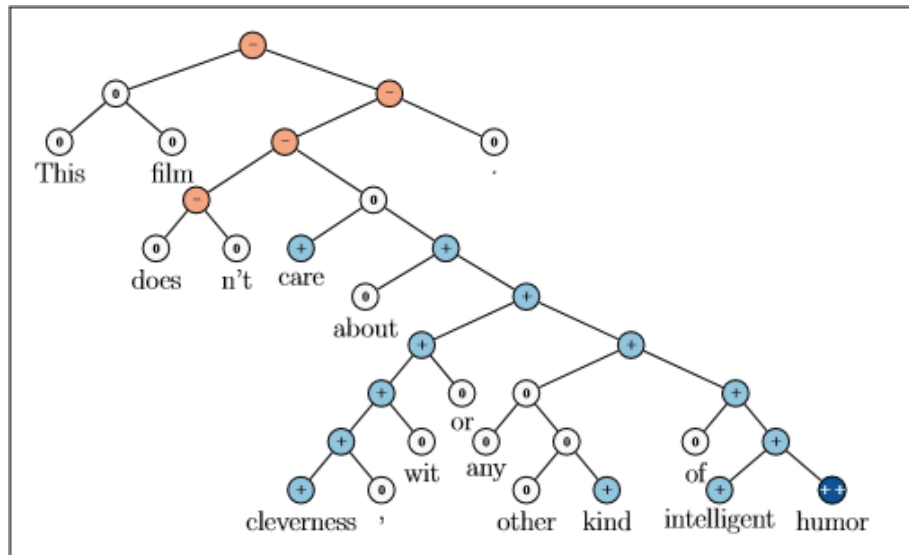
- Clasificación de Tópicos
- Análisis de Sentimiento
- Reconocimiento de Entidades Nombradas

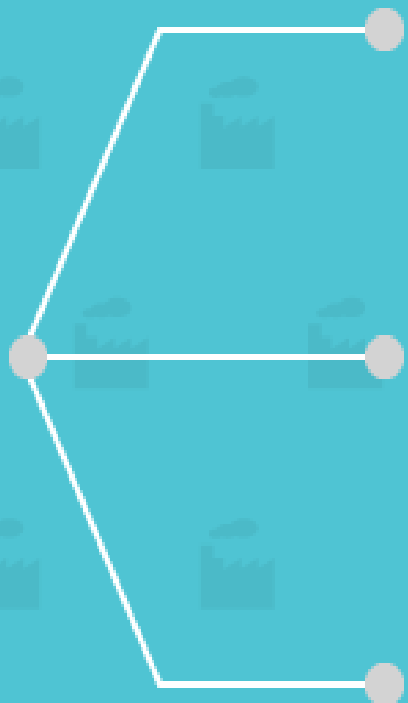


Análisis de Sentimiento

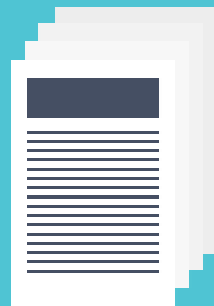
- Objetivo: clasificación binaria o en pocas categorías
- Keypoints: sintaxis
- SotA: XLNet

- Identificación de comentarios negativos
- Gestión personalizada de requerimientos basado en el sentimiento

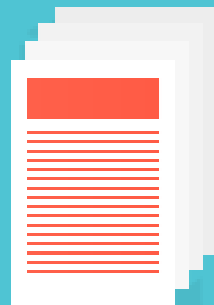




Technology



Sports

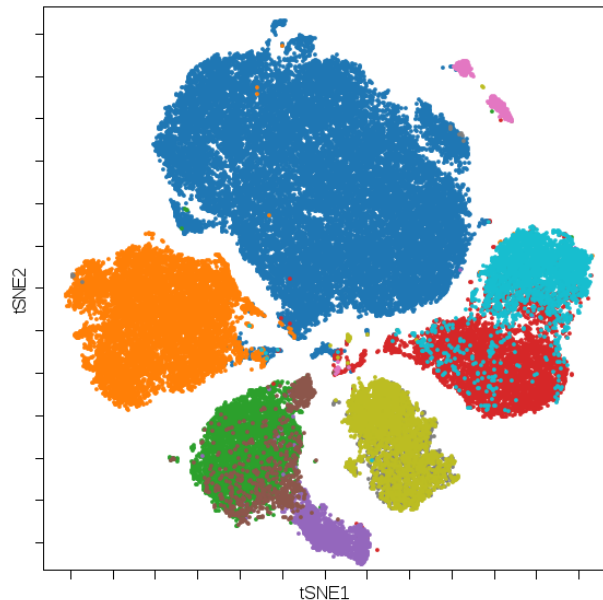


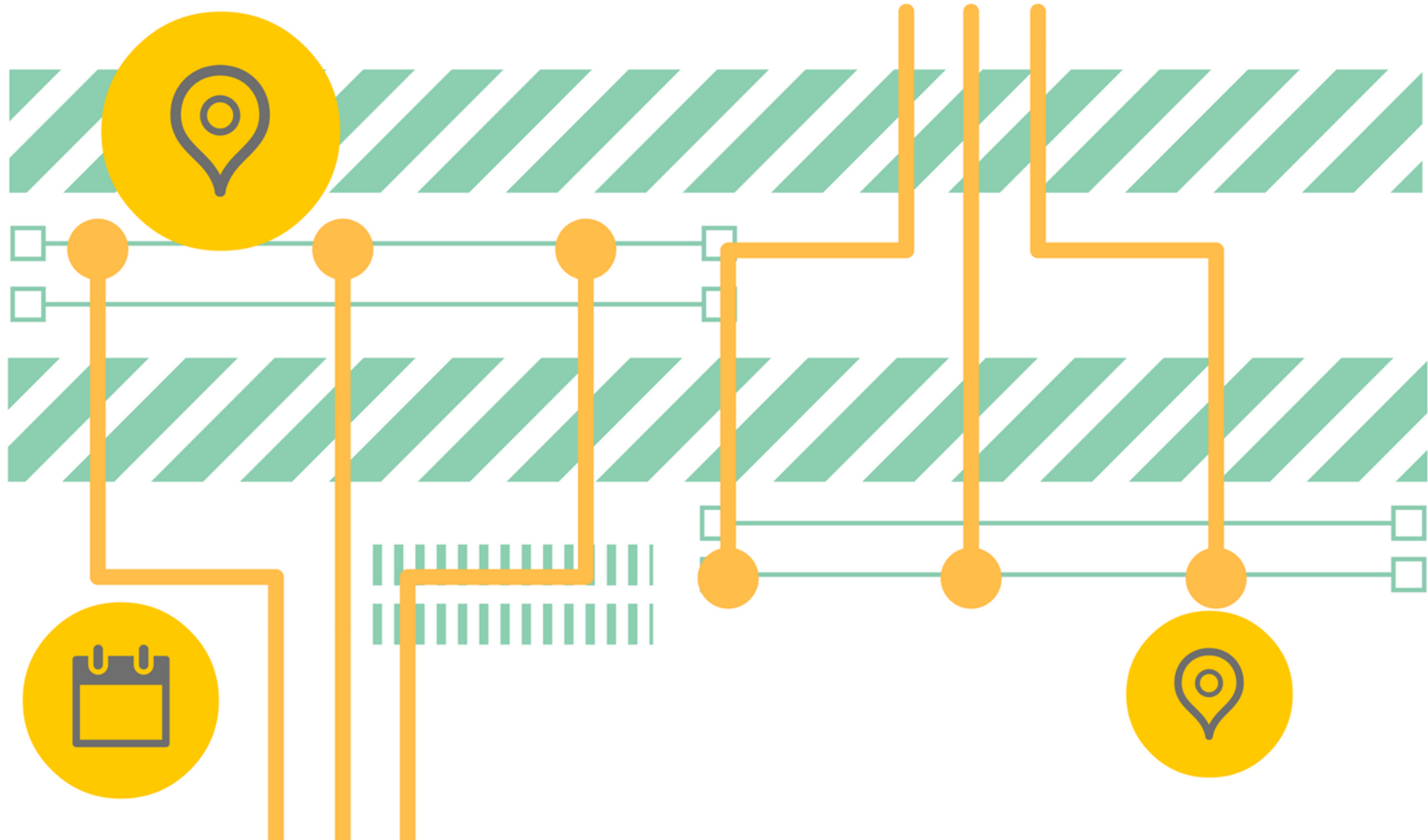
Fashion

Clasificación de Texto

- Objetivo: clasificación de muchas categorías
- Keypoints: semántica > sintaxis
- SotA: XLNet

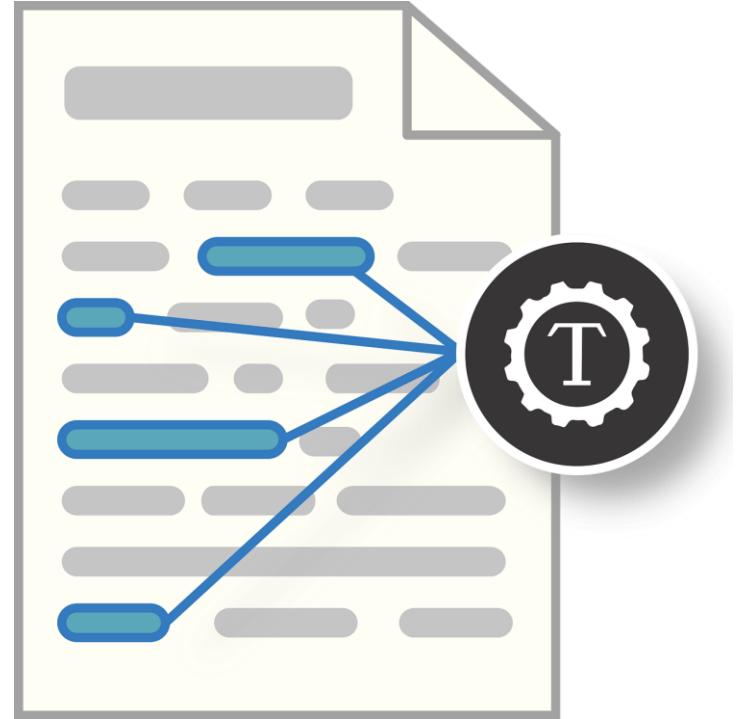
- ¿De qué temas están hablando mis clientes?
Social Network Listening.

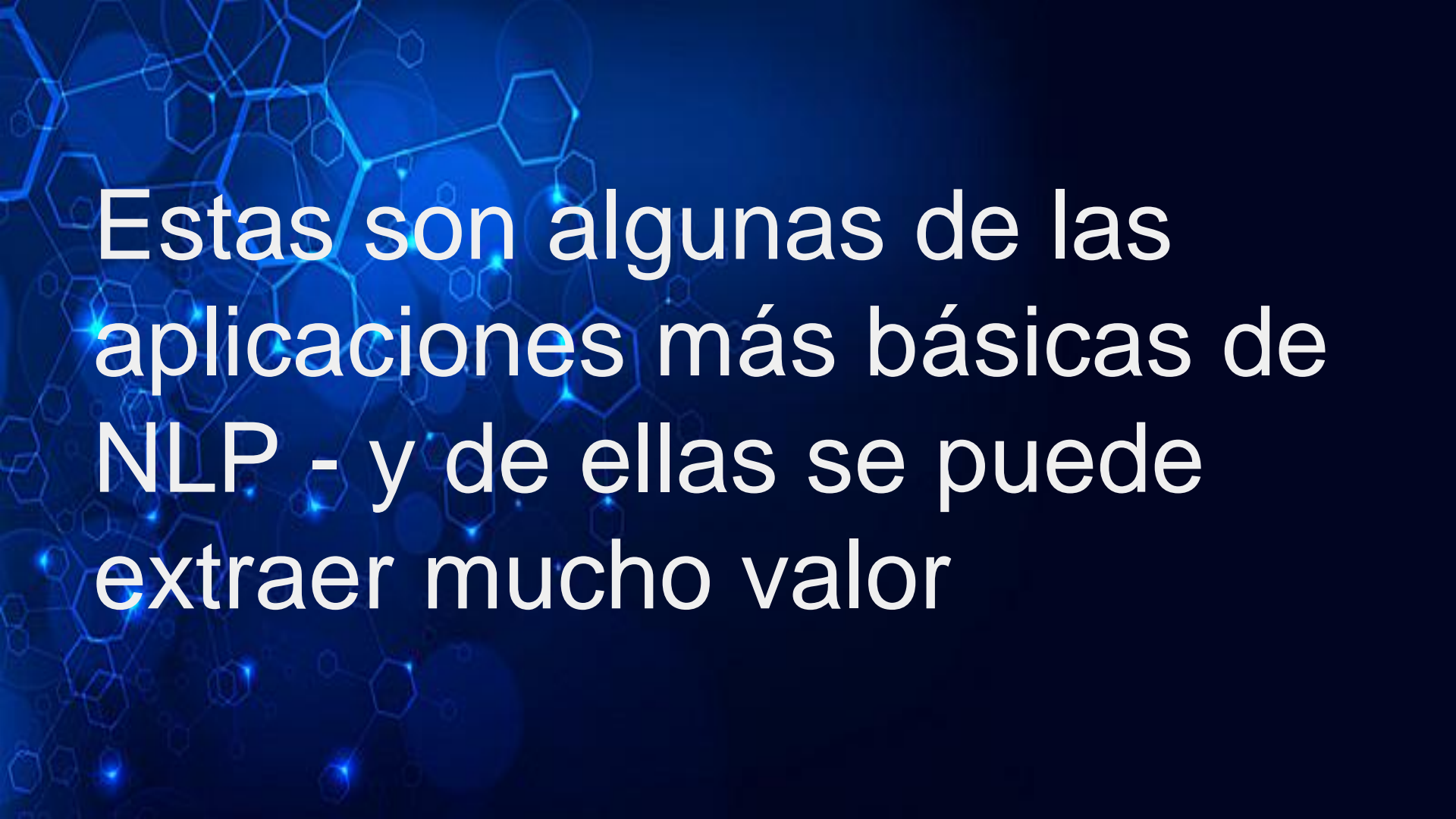




Reconocimiento de Entidades Nombradas

- Objetivo: clasificación
 - Keypoints: sintaxis, aumentación
 - SotA: ~BERT
-
- Inteligencia competitiva
 - Servicio al cliente: identificación de las entidades involucradas en un caso específico

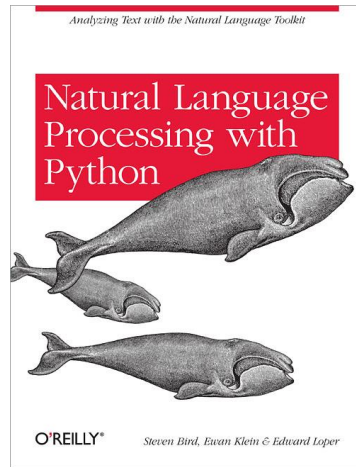




Estas son algunas de las aplicaciones más básicas de NLP - y de ellas se puede extraer mucho valor

¿Cómo sigo aprendiendo de NLP?

- Aplicándolo
 - Realizar tutoriales hasta sentirse cómodo con los conceptos, ideas y arquitecturas.
 - Intentar nuevos proyectos con datasets propios o públicos.
 - No tener miedo de aplicar nuevas cosas!
 - Mentalidad científica: ¿cómo pruebo si esto realmente funciona?



spaCy

Tutoriales

- [Tutorial Sentiment Analysis](#)
- [Tutorial Text Classification](#)
- [Tutorial NER \(1\)](#)
- [Tutorial NER \(2\)](#)



Referencias

- [Word2Vec](#)
- [Intro to Word2Vec](#)
- [GloVe](#)
- [BERT](#)
- [GPT-2](#)
- [XLNET](#)
- [Sebastian Ruder - NLP Progress](#)

