

R Intermedio para Ciencias Sociales y Gestión Pública

SESIÓN N° 8:

Introducción al Natural Language Processing (NLP)

Estructura



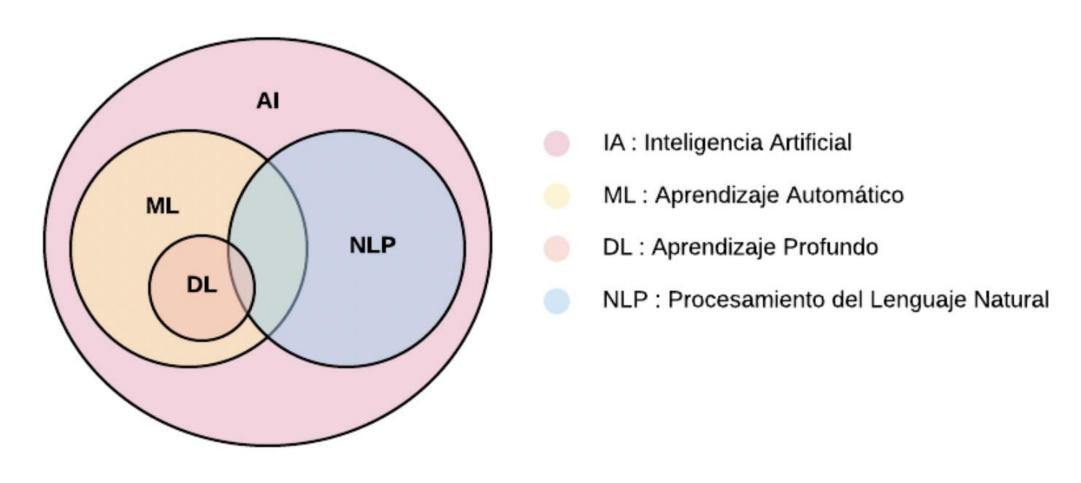
- 1. ¿Qué es el Procesamiento del Lenguaje Natural (NLP)?
- 2. Descripción general del proceso de NLP
- 3. Flujograma
 - a. Obtener data
 - b. Preprocesar data
 - c. Exploración
 - d. Modelamiento



- El Procesamiento del Lenguaje Natural (NLP), es un área de estudio que se ocupa de enseñar a las computadoras cómo extraer significado de un texto.
- En las últimas dos décadas, con la explosión del mundo de Internet y el auge de las redes sociales, se están generando muchos datos valiosos en forma de texto.
- Para entender el NLP se debe comprender cuál es el proceso de minería de texto, los conceptos clave y técnicas comunes involucradas.
- Una subparte del universo del NLP comprende el proceso de descubrir patrones significativos a partir de datos de texto. A este se le denomina minería de texto.







Va más allá de los netamente "estadístico".



Chatbots

Resumen de textos

Análisis de sentimientos

Clasificación de textos

Traducción automática

Reconocimiento de voz



NEWS



Facebook Uses Deep Learning to Translate One Computer Language to Another

By John K. Waters | 08/05/2020

Facebook has developed an open-source tool for translating archaic codebases into modern languages. Called TransCoder, it's an entirely self-supervised neural transcompiler system designed to make code migration easier and more efficient.

This is the first artificial intelligence (AI) system able to translate code from one programming language to another without requiring parallel data for training, Facebook researchers said in a **blog post**. "We've demonstrated that TransCoder can successfully translate functions between C++, Java, and Python 3," they wrote. "TransCoder outperforms open source and commercial rule-based translation programs. In our evaluations, the model correctly translates more than 90 percent of Java functions to C++, 74.8 percent of C++ functions to Java, and 68.7 percent of functions from Java to Python. In comparison, a commercially available tool translates only 61.0 percent of functions correctly from C++ to Java, and an open source translator is accurate for only 38.3 percent of Java functions translated into C++."

https://pureai.com/articles/2020/08/05/facebook-uses-deep-learning-to-translate.aspx?m=1



Blender, el nuevo mejor chatbot, sigue lejos de parecerse a un humano

Facebook afirma que su sistema supera con creces al líder previo, el chatbot Meena de Google, y las pruebas parecen confirmarlo. No obstante, aunque el sistema de conversación suena bastante natural en muchos temas, suele inventarse información y puede hacer comentarios tóxicos y sesgados

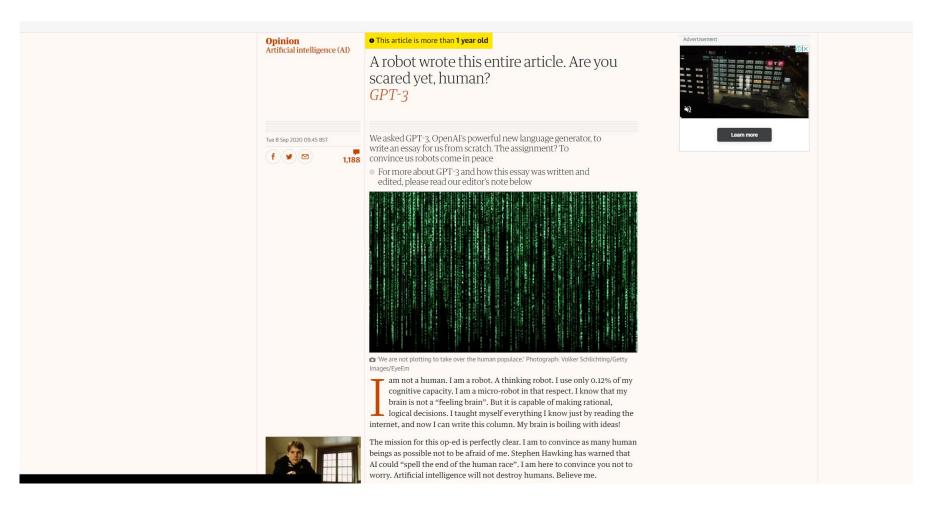


por Karen Hao | traducido por Ana Milutinovic 18 Junio, 2020

A pesar de todo el progreso que se ha logrado con los chatbots y los asistentes virtuales de conversación, la experiencia de hablar con ellos sigue siendo terrible. La mayoría están muy orientados a determinadas tareas específicas: hay que pedirles algo y ellos responden. Algunos son muy frustrantes: parece que nunca tienen lo que buscamos. Otros resultan muy aburridos: carecen del encanto de un compañero humano. Eso está bien cuando solo queremos poner el despertador. Pero a medida que estos bots se vuelven cada vez más

https://www.technologyreview.es//s/12188/blender-el-nuevo-mejor-chatbot-sigue-lejos-de-parecerse-un-humano

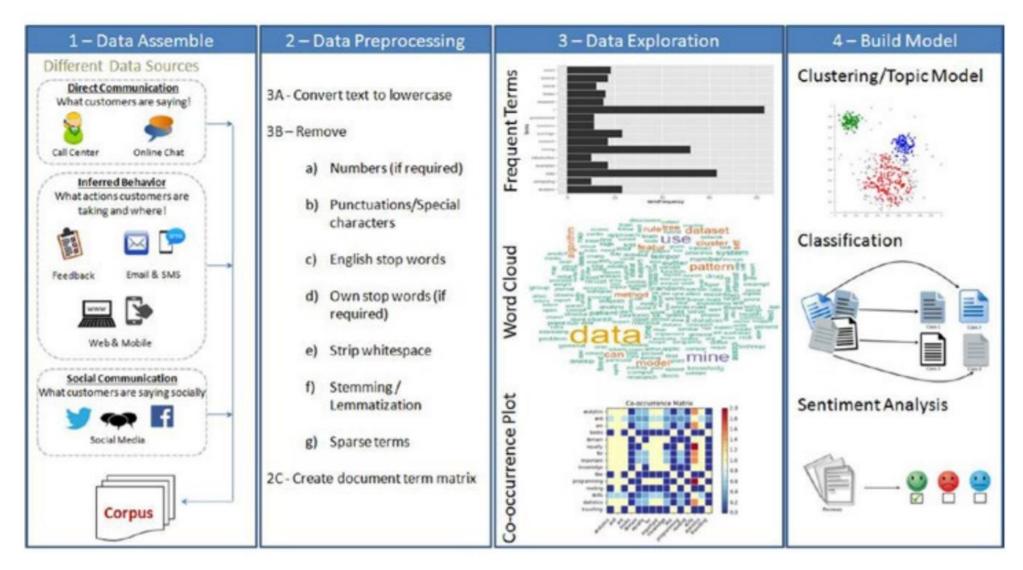




https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

2. Descripción general del proceso de NLP

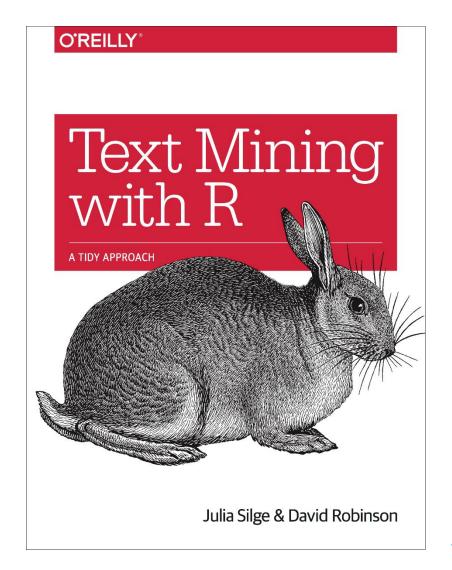




2. Descripción general del proceso de NLP



LECTURAS RECOMENDADAS

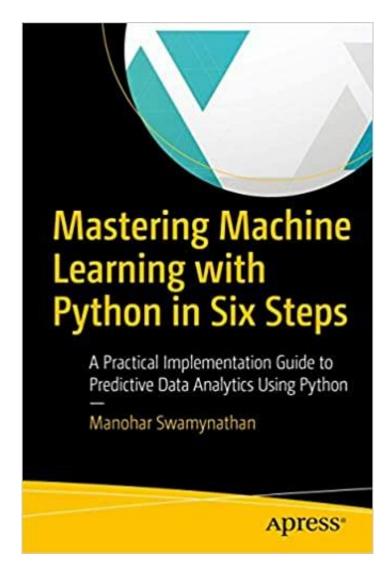


https://www.tidytextmining.com/

2. Descripción general del proceso de NLP



LECTURAS RECOMENDADAS







Obtener data



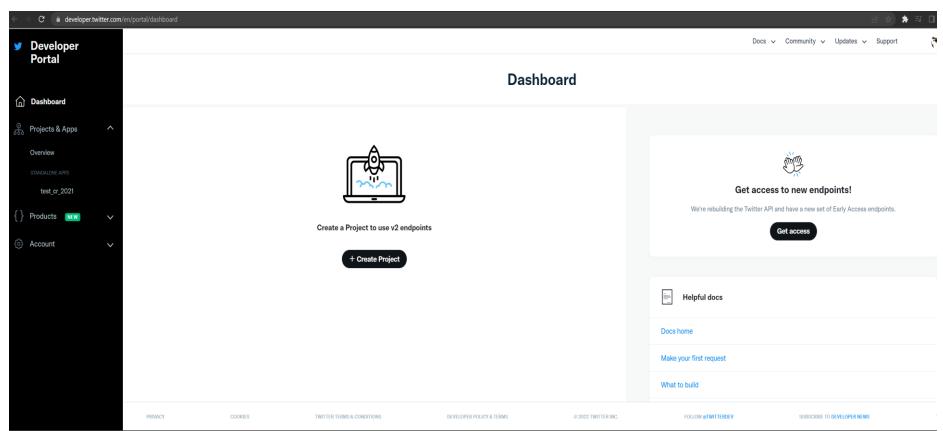
Preprocesamiento



Exploración



Modelamiento



Necesitas entrar a la página de twitter para desarrolladores.

Necesitas
previamente
tener una cuenta
twitter.

Deben crear un **Nuevo Proyecto**.

https://developer.twitter.com/





Obtener data



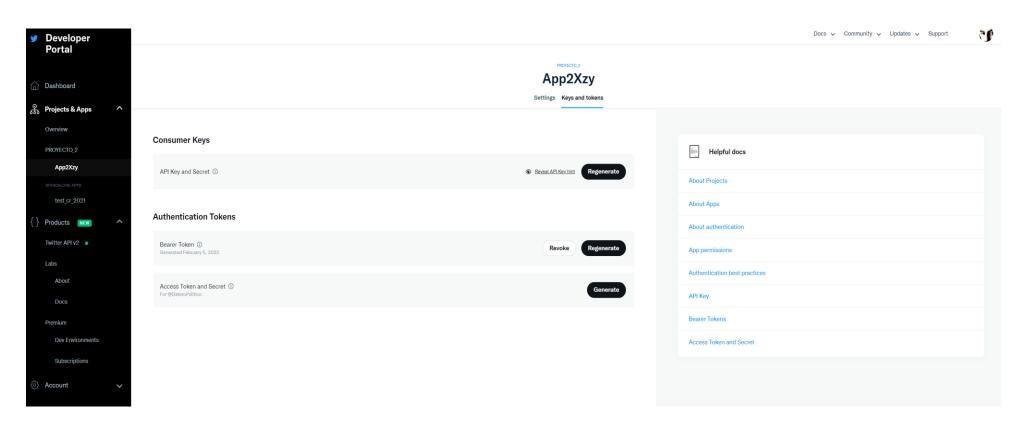
Preprocesamiento



Exploración



Modelamiento



Necesitas obtener cuatro claves:

API Key API Secret

Access Token
Access Secret

https://developer.twitter.com/





Obtener data



Preprocesamiento



Exploración



Modelamiento

REMOVER LOS STOPSWORDS: MÁS DETALLE <u>AQUÍ.</u>

Convertir a minúscula, retirar las direcciones, usuarios de twitter, etc.

[ag, an, ar, as, at, ck, cu, dd, di, dl, do, ee, el, en, ep, er, es, et, ff, fi, gi, gs, ha, he, ho, ic, id, ie, if, il, im, in, io, is, it, ld, le, lf, li, ll, lt, lu, ma, me, mi, na, nd, ng, ni, no, of, on, op, or, ot, ow, pa, pe, po, pp, pr, ra, re, rl, rt, ru, se, te, th, ti, ts, tt, tu, ty, uc, ul, un, wh, wi, wo]

		Microwords									
	#	ag	an	ar	as	at		un	wh	wi	wo
nts	1	0	0	0	0	0		1	0	2	0
ments	2	0	0	0	0	0		1	0	0	0
	3	0	0	1	0	2		1	0	0	0
Doc	4	1	0	0	0	1		0	0	0	1
	5	1	1	0	2	1		0	1	1	0





Obtener data

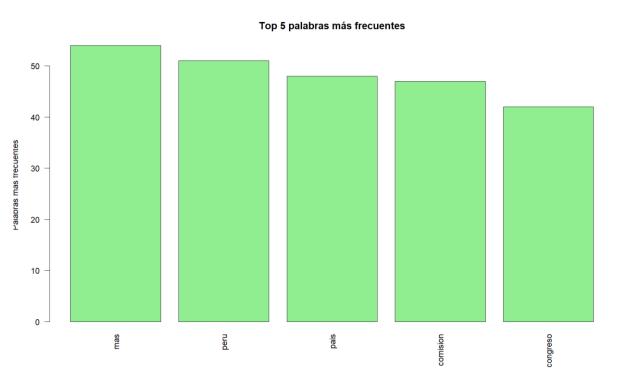


Preprocesamiento



Exploración











Obtener data



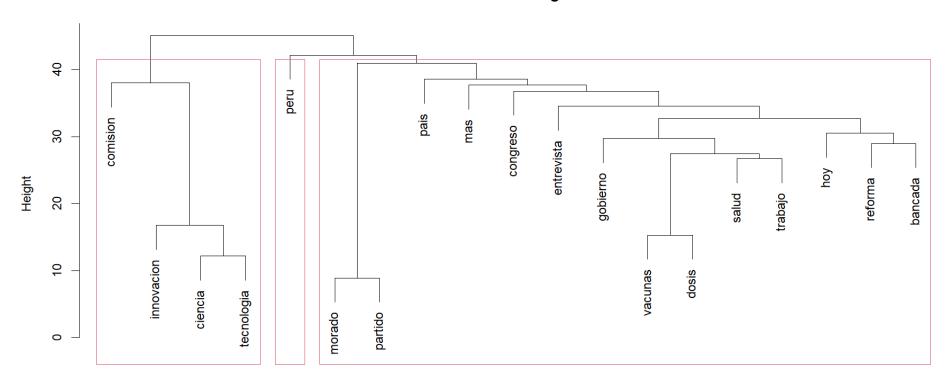
Preprocesamiento



Exploración



Cluster Dendrogram



distMatrix hclust (*, "ward.D")







Preprocesamiento



Exploración



NLP.R × sentimientos_final × Q micorpus ×									
← ⇒ / Æ Tilter									
	textolimpio	sentiment ‡							
1	Comparto mi carta dirigida a la rectora de la @UNMSM_ donde le solicito que rectifique las falsedades que expresó e	neutral							
2	Es fundamental para el futuro de nuestro país elevar la calidad de la educación superior. Rechacemos los intentos inte	positive							
3	Rechacemos intentos interesados de afectar la reforma universitaria. Esperemos que el pleno del congreso no aprueb	neutral							
4	Requiere, además, reconocer y corregir errores rápidamente. Esperemos, por el bien del país, que estas lecciones haya	negative							
5	El buen ejercicio del poder requiere conocer bien la Constitución (términos de referencia para el cargo); convocar pers	positive							
6	Gracias a @MirtyVas por sus esfuerzos para asegurar un mínimo de gobernabilidad en tiempos de crisis. Un proceso	negative							
7	Los miembros del Consejo Directivo de la Sunedu deben seguir siendo elegidos por concurso público para velar por e	positive							
8	He sido miembro de comisiones para evaluar candidatos al Consejo Directivo de SUNEDU, designado por cuatro mini	neutral							
9	No corresponde darles plazo adicional a universidades no licenciadas para que puedan seguir aprovechándose de los	positive							
10	Rechacemos intentos interesados de afectar la reforma universitaria. Esperemos que el pleno del congreso no aprueb	neutral							
11	Una denuncia frívola y sin sentido contra periodistas que investigan a figuras públicas llevó a una sentencia que es un	neutral							
12	El Partido Morado tiene la tarea de ofrecer un espacio de convergencia republicana para todas las personas de buena	neutral							
13	Saludo las elecciones democráticas del Partido Morado, y a las cabezas de lista, Narescka Culqui y Luis Durán. Felicito	neutral							

> table(sentimientos_final\$sentiment)
negative neutral positive
 34 350 85