

Recolección algoritmos - Prototipo web que genera exámenes usando Procesamiento del Lenguaje Natural para el curso de Inglés IV de Ingeniería de Sistemas

#	Tema	Nombre	Autores	Descripción	Enlace	Técnicas PLN	Información adicional (Estrategias, red neuronal, Datasets, otros)	Relevancia (1-5)	Ventajas	Desventajas
1	Text Generation	GPT-2, Language Models are Unsupervised Multitask Learners	Radford, Alec and Wu, Jeff and Child, Rewon and Luan, David and Amodei, Dario and Sutskever, Ilya	Es un modelo de aprendizaje automático transformador, GPT-2 utiliza el aprendizaje profundo para traducir texto, responder preguntas, resumir pasajes y generar resultados de texto	https://github.com/openai/gpt-2	Tokenization, Word Embedding (word2vec), Byte Pair Encoding (BPE)	Transformers, self-Attention	4	- Puede ser reentrenado con otro conjunto de datos	- Puede volverse repetitivo o sin sentido al generar pasajes largos (Más de dos párrafos)
2	Question Generation	Question Generation using transformers	Suraj Patil, Neeraj Varshney, Priyank Soni, Manuel Romero, Tom Wilde	Este proyecto está dirigido como un estudio de código abierto sobre la generación de preguntas con transformadores previamente entrenados (específicamente modelos seq-2-seq) utilizando métodos sencillos de extremo a extremo sin tuberías muy complicadas	https://github.com/patil-suraj/question_generation	Tokenizer, Word Embedding, Masked LM (MLM), Next Sentence Prediction (NSP)	Tranformers, T5, Seq2Seq, sQUAD	4	- Código abierto y permite hacer Fine tuning - Documentación detallada - Disponible en cuaderno de Google Colab para probar - Artículo en Medium	-
3	Question generation	Question Generator	Adam Montgomerie	Question Generator es un sistema de PNL para generar preguntas de estilo de comprensión lectora a partir de textos como artículos de noticias o extractos de páginas de libros. El sistema está construido utilizando modelos previamente entrenados de HuggingFace Transformers. Hay dos modelos: el generador de preguntas en sí mismo y el evaluador de control de calidad, que clasifica y filtra los pares de preguntas y respuestas en función de su aceptabilidad.	https://github.com/AMontgomerie/question_generator	Tokenizers	Tranformers, T5, Seq2Seq, sQUAD	4	- Permite cambiar algunos parámetros del modelo, como el número de preguntas, el estilo de preguntas - Fácil ejecución - Genera preguntas de opción múltiple a partir de texto	- No tiene bien especificado la estructura del modelo, sin embargo se podría obtener inspeccionando el código fuente
4	Question generation	Question Generation	Kristiyan Vachev	La idea es generar respuestas de opción múltiple a partir del texto, dividiendo este complejo problema en pasos más simples: Identifique las palabras clave del texto y utilícelas como respuestas a las preguntas. Reemplaza la respuesta de la oración con un espacio en blanco y úsala como base para la pregunta. Transforme la oración con un espacio en blanco para responder a una oración más similar a una pregunta. Genere distractores, palabras similares a la respuesta, como respuestas incorrectas.	https://github.com/KristiyanVachev/Question-Generation	Part of speech, Named entity, Word count	sQUAD, TF-IDF score and cosine similarity, scikit-learn's Gaussian Naive Bayes	4	- Buena documentación - El algoritmo está dividido en pasos más simples y entendibles (modularidad y simplicidad)	- Las preguntas generadas no son realmente aptas para utilizar directamente en exámenes

Recolección algoritmos - Prototipo web que genera exámenes usando Procesamiento del Lenguaje Natural para el curso de Inglés IV de Ingeniería de Sistemas

#	Tema	Nombre	Autores	Descripción	Enlace	Técnicas PLN	Información adicional (Estrategias, red neuronal, Datasets, otros)	Relevancia (1-5)	Ventajas	Desventajas
6	Text Generation	GPT2-Pytorch with Text-Generator	Tae Hwan Jung (Jeff Jung)	Es un modelo de aprendizaje automático transformador, GPT-2 utiliza el aprendizaje profundo para traducir texto, responder preguntas, resumir pasajes y generar resultados de texto	https://github.com/graykode/gpt-2-Pytorch	Tokenization, Word Embedding (word2vec), Byte Pair Encoding (BPE)	Transformers, self-Attention	4	- Fácil ejecución y ajuste de parámetros	- Puede volverse repetitivo o sin sentido al generar pasajes largos (Más de dos párrafos) - El rendimiento es menor al modelo original
7	Text Generation	Textgenrnn	Max Woolf	Entrene fácilmente su propia red neuronal generadora de texto de cualquier tamaño y complejidad en cualquier conjunto de datos de texto con unas pocas líneas de código, o entrene rápidamente en un texto usando un modelo previamente entrenado. textgenrnn es un módulo de Python 3 sobre Keras / TensorFlow para crear char-rnn s, con muchas características interesantes	https://github.com/minimaxir/textgenrnn	-	Recurrent Neural Network (RNN), Long Short Term Memory (LSTM)	3	- Se describe gráficamente la estructura del modelo - Entrainable fácilmente	-
10	Text Generation	GPT-2	Radford, Alec and Wu, Jeff and Child, Rewon and Luan, David and Amodei, Dario and Sutskever, Ilya	GPT-2 es un modelo de transformadores previamente entrenado en un corpus muy grande de datos en inglés de una manera auto-supervisada. Esto significa que se entrenó previamente solo en los textos en bruto, sin que los humanos los etiquetaran de ninguna manera (por lo que puede usar muchos datos disponibles públicamente) con un proceso automático para generar entradas y etiquetas a partir de esos textos.	https://huggingface.co/gpt2	Lowercasing, Tokenization, Stemming, out-of-vocabulary tokens	Transformers	5	- Se puede importar fácilmente desde HuggingFace Transformers - Su rendimiento es bueno en comparación de otros algoritmos modificados basados en GPT2	-
9	Text Generation	Text generation with LSTMs.	Samuel Rochette, Wassel Alazhar	Es un generador de texto basado en Long Short Term Memory (LSTM), el proyecto está basado en un artículo del 2015 llamado "The Unreasonable Effectiveness of Recurrent Neural Networks" de Andrej Karpathy	https://github.com/Saxamos/text-generator	-	Long Short Term Memory (LSTM)	3	- Permite entrenar el modelo de manera "fácil" - Permite cambiar algunos parámetros del modelo	Poca documentación respecto al modelo y funcionamiento del algoritmo
5	Text Generation	GPT-2 Pre-training and text generation, implemented in Tensorflow 2.0	Abhay Singh	Es un modelo implementado originalmente en tensorflow 1.14 por OpenAI: - "openai / gpt-2". El repositorio tiene implementación de generación de secuencia y preentrenamiento OpenAI GPT-2 en tensorflow 2.0	https://github.com/akanyaani/gpt-2-tensorflow2.0	Part of speech, Named entity, Word count	Transformers, self-Attention	2	- Se puede entrenar previamente el modelo usando datos de muestra, dándole el enfoque necesario para la generación de texto - Gráfico del cálculo de GPT-2	- Se debe terminar de entrenar el modelo, lo cual demanda recursos computacionales

Recolección algoritmos - Prototipo web que genera exámenes usando Procesamiento del Lenguaje Natural para el curso de Inglés IV de Ingeniería de Sistemas										
#	Tema	Nombre	Autores	Descripción	Enlace	Técnicas PLN	Información adicional (Estrategias, red neuronal, Datasets, otros)	Relevancia (1-5)	Ventajas	Desventajas
8	Text Generation	GPT2	Connor Leahy, Lu Zhihe, Anh Hong and Cong	Es un modelo de aprendizaje automático transformador, GPT-2 utiliza el aprendizaje profundo para traducir texto, responder preguntas, resumir pasajes y generar resultados de texto	https://github.com/ConnorJL/GPT2	Tokenization, Word Embedding (word2vec), Byte Pair Encoding (BPE)	Transformers, self-Attention	2	<ul style="list-style-type: none"> - Buena documentación - Permite terminar de entrenar el modelo con corpus de preferencia propia - Ofrece la posibilidad de descargar entre dos modelos preentrenados 	<ul style="list-style-type: none"> - No sigue el mismo rendimiento del modelo GTP 2 original por causa desconocida por el desarrollador - Es una modificación del modelo original