**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

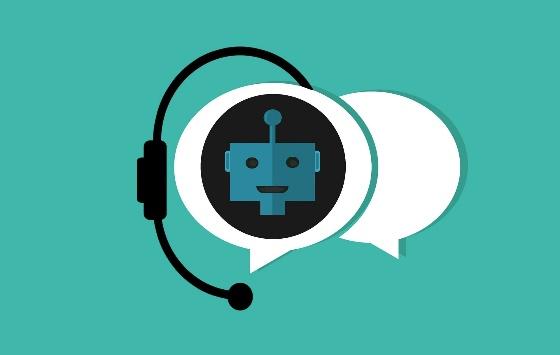
**INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS**

Logotipo, nombre de la empresa

Descripción generada automáticamenteDibujo en blanco y negro

Descripción generada automáticamente con confianza media

**CHATBOT SERVICIO SOCIAL**



**Arellano Carmona Axel**

**Rodriguez Reyes Rodrigo Eduardo**

**Objetivo:** Realizar un chatbot el cual proporcione información acerca de los procesos a seguir para iniciar el servicio social y/o terminarlo. También tendrá la función de recibir documentos para inicio del servicio social para agilizar y automatizar la misma, hicimos un estudio en donde se encontró las categorías de preguntas en las cuales los alumnos encuentran mayor número de dudas por dar ejemplos (cancelación de servicio, conclusion de servicio, descarga de solicitud del servicio, documentos necesarios, documentos para concluir)

**Contexto: Chatbot**

En el nivel más básico, un chatbot es un programa informático que simula y procesa conversaciones humanas (ya sea escritas o habladas), permitiendo a los humanos interactuar con dispositivos digitales como si se estuvieran comunicando con una persona real. Los chatbots pueden ser tan sencillos como programas rudimentarios que responden a consultas sencillas con una respuesta de una sola línea, o bien tan sofisticados como los asistentes digitales que pueden aprender y evolucionar para ofrecer niveles de personalización cada vez mayores a medida que reúnen y procesan información.

**PLN**

El procesamiento de lenguaje natural (PLN) aúna dos disciplinas tan aparentemente lejanas como la lingüística y la inteligencia artificial. En la actualidad, este campo de las ciencias de la computación, que consiste en transformar el lenguaje natural en un lenguaje formal como el de programación que los ordenadores puedan procesar, no deja de evolucionar y sus aplicaciones son cada vez mayores.

Según la definición de IBM, el procesamiento de lenguaje natural es la rama de la informática —más concretamente, de la inteligencia artificial— que se ocupa de dotar a los ordenadores de la capacidad de entender lenguaje hablado y escrito del mismo modo que los seres humanos. Esta tecnología ha alcanzado un alto nivel en la actualidad gracias a la aplicación de tecnologías como el machine learning (aprendizaje automático), el big *data,* el internet de las cosas o las redes neuronales.

Algunas de las aplicaciones más importantes se centran en la inteligencia de negocio (business intelligence*)*, que permite analizar automáticamente las reacciones de los clientes a través de lo que publican en Internet o de las preguntas que realizan para extraer información. Los *chatbots* son otra aplicación que, aunque aún cuenta con un gran margen de mejora, agiliza la interacción con los clientes a través de chats o contestadores telefónicos ofreciendo respuestas rápidas y automatizadas gracias al procesamiento de lenguaje natural.

***Desarrollo***

**Uso de API PyTorch (Transformers)**

PyTorch-Transformers (anteriormente conocido como pytorch-pretrained-bert) es una biblioteca de modelos preentrenados de última generación para el procesamiento del lenguaje natural (NLP).

***Documentacion PyTorch***[PyTorch-Transformers | PyTorch](https://pytorch.org/hub/huggingface_pytorch-transformers/)

**Roberta Model**

El modelo RoBERTa fue propuesto en RoBERTa: A Robustly Optimized BERT Pretraining Approach por Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. Se basa en el modelo BERT de Google lanzado en 2018.

Se basa en BERT y modifica hiperparámetros clave, eliminando el objetivo de preentrenamiento de la siguiente oración y entrenando con mini lotes y tasas de aprendizaje mucho más grandes.

***Documentación Modelo***  
[RoBERTa (huggingface.co)](https://huggingface.co/docs/transformers/model_doc/roberta)

**Chatbot SS**

Se desarrolló un chatbot haciendo uso de Python, uno de los lenguajes mas usados para desarrollo de procesos que ocupen Machine Learning, el cual en esencia requirió de dos archivos esenciales, uno que cuenta con poco mas de 100 ejemplos de preguntas (xlx) y un archivo json el cual está compuesto por las categorías y el contexto de cada una de ellas.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**Cargándolos en código**

Texto

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Nuestro algoritmo consistió en tokenizar cada una de las palabras tanto del archivo xlx como el archivo json, con el objetivo primeramente de tener una cat of words (frecuencia de palabras) y de misma forma detectar stop words (signos de ortografía, mayúsculas, acentos).

Gráfico, Gráfico de barras

Descripción generada automáticamente

Texto

Descripción generada automáticamente



Una vez teniendo una mezcla tanto de categorías y preguntas, asi como el contexto hacemos uso del modelo transformer, el cual con ayuda de un contexto es capaz de clasificar y dar una respuesta a un tipo de pregunta en especifico

Texto

Descripción generada automáticamente

Para ello desarrollamos una funcion de distancia coseno la cual nos ayudara a tener con mayor certeza un rango de similitud entre las palabras usadas en la pregunta y las que radican en el contexto.

Captura de pantalla de computadora

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Después hacemos uso del modelo de Roberta que nos ayudara a generar la respuesta que en esencia requiere de un contexto y de la pregunta para generar la respuesta.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

**Conclusiones Generales**

Durante el desarrollo del proyecto se reforzaron los conocimientos y aprendimos nuevos en el lenguaje Python, asi mismo se adquirieron nuevas habilidades para el Procesamiento del Lenguaje Natural, se tomaron diversas decisiones en base a los resultados que íbamos obteniendo, teniendo en cuenta diversos factores como el numero de preguntas con el que contábamos para el entrenamiento del modelo, hicimos uso de ambos modelos con tal de congeniar una mejor respuesta al usuario, agrademos al Dr Ivan Vladimir Meza el cual nos apoyó con documentación y consejos para conseguir el objetivo, repasamos temas con diversas librerías de Python para el manejo de los archivos, su lectura, manejo de diccionarios y diversas APIS que nos dan el uso de modelos ya entrenados por grandes desarrolladores alrededor del mundo en la plataforma de Hugging Face.

***Fuentes Consultadas***

[Qué es un chatbot | Oracle México](https://www.oracle.com/mx/chatbots/what-is-a-chatbot/)

[PyTorch-Transformers | PyTorch](https://pytorch.org/hub/huggingface_pytorch-transformers/)

[PLN: Procesamiento de lenguaje natural - Iberdrola](https://www.iberdrola.com/innovacion/procesamiento-lenguaje-natural-pln)

[RoBERTa (huggingface.co)](https://huggingface.co/docs/transformers/model_doc/roberta)

[pandas - Python Data Analysis Library (pydata.org)](https://pandas.pydata.org/)

[Python Dictionaries (w3schools.com)](https://www.w3schools.com/python/python_dictionaries.asp)

[(6) Ivan Vladimir Meza Ruiz - YouTube](https://www.youtube.com/c/IvanVladimirMezaRuiz/videos)

[Qué son las distancias, tipos y cómo calcularlas con Python 2/2. | by Rubiales Alberto | Medium](https://rubialesalberto.medium.com/qu%C3%A9-son-las-distancias-tipos-y-c%C3%B3mo-calcularlas-con-python-2-2-f1fbef9556c2)