

**UNIVERSIDAD AUTONOMA DEL ESTADO DE MÉXICO**



.

**INGENIERÍA EN SISTEMAS INTELIGENTES**

**MATERIA:**

**PROCESAMIENTO DE LENGUAJE NATURAL**

**NO. ACTIVIDAD:**

**ACTIVIDAD 1ª**

**NOMBRE:**

**(PROYECTO FINAL)**

PROFESOR(A):

**CALDERÓN NEPAMUCENO DORA MARÍA**

**ALUMNA(O):**

BONILLA LEYVA WENDY

RAMIREZ LEAL CARMEN

TAPIA GUZMAN ERICK OMAR

**TURNO:** **VESPERTINO****GRUPO:** **V9**

# **TITULO:**

ASISTENTE VIRTUAL A TRAVÉS DE RECONOCIMIENTO DE VOZ EN UN PROCESO DE LENGUAJE NATURAL

# **OBJETIVO:**

Diseñar una aplicación que sea nuestro asistente virtual, es decir, que, al momento de hablar, el programa pueda obtener lo que has dicho y por medio de una serie de condiciones, pueda ofrecerle un resultado. Como por ejemplo hacer que reproduzca videos en YouTube, que diga la hora y que pueda buscar información.

# **ALCANCE:**

El reconocimiento de voz tiene varias aplicaciones útiles en el dominio de la interacción humana con la computadora y la transcripción automática de voz. Este proyecto explicara brevemente el proceso de transcripción de voz en Python a través de **speech\_recognition** biblioteca creando un programa en Python y explicando cómo traducir voz a texto cuando la fuente de audio es un archivo de audio o un micrófono.

# **INTRODUCCIÓN:**

Los **Lenguajes Naturales** son los utilizados en la comunicación humana, ya sean escritos, hablados o signados.

El castellano, el inglés o el francés son ejemplos de lenguaje natural (lenguas). Poseen una sintaxis y una gramática, y cumplen con los principios de economía y optimización, a pesar de que contienen muchas ambigüedades. Han evolucionado junto a la humanidad: las lenguas han sido creadas entre todos.

Por el contrario, los Lenguajes Formales se utilizan para transferir información, sin dejar lugar a ambigüedades. Algunos ejemplos de lenguajes formales son las matemáticas, XML, SQL y PHP.

Los ordenadores pueden tratar los lenguajes formales sin problemas, pero uno de sus principales retos es entender el lenguaje natural. Con este propósito, hay un área informática dedicada a la interacción entre ordenadores y las lenguas habladas por los humanos denominada **Procesamiento del Lenguaje Natural (PLN)**.

El **Procesamiento del Lenguaje Natural** es una disciplina de la Inteligencia Artificial que se ocupa de la formulación e investigación de mecanismos computacionales para la comunicación entre personas y máquinas mediante el uso de **Lenguajes Naturales**.

Algunas de las aplicaciones del **Procesamiento del Lenguaje Natural**:

* Comprensión del lenguaje.
* Recuperación de la información.
* Extracción de la información.
* Búsqueda de respuestas.
* Generación de discurso.
* Traducción automática.
* Reconstrucción de discurso.
* Reconocimiento del habla.
* Síntesis de voz.
* …

En las últimas décadas, gracias al acceso a Internet, el desarrollo de la Inteligencia Artificial, la interconectividad, y con el auge del Internet de las Cosas, (Internet of Things IoT, por sus siglas en inglés) han surgido una gran variedad de dispositivos inteligentes. Entre estos dispositivos interconectados se destacan, aquellos que son capaces de interpretar y procesar comandos de voz, los cuales están siendo utilizados en ambientes cada vez más íntimos, como oficinas, hogares, lugares de descanso, entre otros. No cabe duda que el avance tecnológico que ha experimentado la sociedad con el IoT ha sido positivo, ayudando a establecer una comunicación más efectiva entre el hombre y este tipo de dispositivos inteligentes. Su función de alguna manera está enfocada en cumplir con una serie de tareas útiles, las cuales tratan de facilitar la vida de las personas que lo utilizan. Sin embargo, es evidente que el énfasis que se ha dado a estos dispositivos se basa en añadir y mejorar características y funcionalidades, antes que reforzar la seguridad en los mismos; dejando así brechas o puertas abiertas que pueden ser aprovechadas por individuos malintencionados o por errores involuntarios, comprometiendo la privacidad de quienes lo utilizan.

# **METODOLOGÍA.**

Para el desarrollo del proyecto, a nivel de herramientas de software, se utilizó:

* **Python versión 3.5** (Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código.​ Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional).

Las librerías de Python que se utilizaran para el desarrollo de la interfaz en el reconocimiento de voz son:

* **SpeechRecognition**: Librería para realizar el reconocimiento de voz, con soporte para varios motores y API, en línea y fuera de línea.
* **Pyaudio**: Enlaces para PortAudio v19, librería de flujo de entrada / salida de audio multiplataforma.

Y para la Síntesis de voz:

* **Pyttsx3**: Librería de texto a voz (TTS) para Python 2 y 3 la cual funciona sin conexión a Internet.

**Phyton y PLN. ¿Por qué utilizar Phyton para el PLN?**

El ecosistema de Python es diverso y admite una amplia variedad de bibliotecas, frameworks y módulos en diversos dominios. Existen frameworks y bibliotecas dedicadas para el procesamiento del lenguaje natural y el análisis de texto, que puede instalar y comenzar a usar, como cualquier otro módulo integrado en el estándar Python. Estos frameworks se han construido durante un largo período de tiempo y, por lo general, todavía están en desarrollo activo. A menudo, la forma de evaluar un framework es ver qué tan activa es su comunidad de desarrolladores.

Cada framework contiene varios métodos, capacidades y características para operar en texto, obtener información y preparar los datos para un análisis posterior, como aplicar algoritmos de aprendizaje automático en datos textuales preprocesados. Aprovechar estos frameworks ahorra mucho esfuerzo y tiempo que se habría invertido en escribir código estándar para manejar, procesar y manipular datos de texto. Por lo tanto, esto permite a los desarrolladores e investigadores centrarse más en resolver el problema real y la lógica y los algoritmos necesarios.

**Ejemplos reales de casos de éxito relacionados al uso de Phyton para el PLN.**

**GPT – 3.**

GPT-3, es un modelo de lenguaje autorregresivo que utiliza el aprendizaje profundo para producir texto similar al de un humano. Es el modelo de predicción de lenguaje de tercera generación de la serie GPT-n creada por OpenAI, un laboratorio de investigación de inteligencia artificial con sede en San Francisco.

La versión completa de GPT-3 tiene una capacidad de 175 mil millones de parámetros de aprendizaje automático, que es más de dos órdenes de magnitud mayor que la de su predecesor, GPT-2.

GPT-3, que se introdujo en Mayo de 2020 y se encuentra en prueba beta a partir de julio de 2020, es parte de una tendencia en los sistemas de procesamiento del lenguaje natural (PLN) de representaciones de lenguaje previamente entrenadas. Antes del lanzamiento de GPT-3, el modelo de lenguaje más grande era Turing NLG de Microsoft, presentado en febrero de 2020, con una capacidad diez veces menor que la de GPT-3.

**Google Bert.**

**BERT es el acrónimo para Bidirectional Encoder Representations from Transformers**(Representaciones de Codificador Bidireccional de Transformadores). Se trata de un sistema basado en inteligencia artificial (IA) para ayudar a los algoritmos de Google Search a entender mejor el lenguaje que utilizamos los usuarios en el momento de realizar una búsqueda mediante oraciones.

En este sentido, la nueva actualización en los algoritmos de Google es tan innovadora como la realizada en 2015 con RankBrain, el primer mecanismo de IA utilizado para analizar las consultas de los usuarios y clasificar mejor los resultados en las SERP. Ambos sistemas, BERT en mayor medida, aplican un método de análisis que permite contextualizar de forma más natural cada consulta.

**BERT posee una característica que se llama “bidireccionalidad”**, que consiste en analizar una oración en dos direcciones. Es decir, analiza las palabras que se encuentran tanto a la izquierda como a la derecha de una palabra clave, y esto le permite entender en profundidad el contexto y la temática de toda la frase que introduce un usuario para la búsqueda en Google. De esta manera, después de comprender muy bien de qué trata determinada consulta, los algoritmos de Google, gracias a sus más de 200 factores de clasificación, seleccionan aquellos contenidos que mejor responden a esa búsqueda, con lo cual los resultados en las SERP se vuelven cada vez más relevantes para los usuarios.

**Text to Text.**

**VODAFONE.**

Vodafone es el claro ejemplo de una compañía pionera en el uso de la Inteligencia Artificial y el Lenguaje Natural en su contact center. La empresa telefónica, que ya **había experimentado resultados muy positivos tras la instalación de un Voicebot en su contact center y un bot en su aplicación**, ha querido sumar un canal de contacto más mediante la instalación de un bot de texto en WhatsApp para mejorar la experiencia de sus clientes, acercarse a ellos y ampliar su capacidad de autogestión.

**Speech to Speech.**

**CASER SEGUROS.**

Caser es el claro ejemplo de una aseguradora que ha aprovechado la ola de transformación digital para **mejorar la experiencia de sus clientes y optimizar el trabajo** de sus gestores con la modernización de su anterior IVR incorporando el reconocimiento de lenguaje natural (NLP), pregunta abierta y la implementación de autogestiones telefónicas en el nuevo Voicebot, así como la creación de un chatbot interno para mediadores.

**VODAFONE.**

**MyVoice** es el proyecto que ha permitido redefinir el **VoiceBot** de **Vodafone** modernizando su interfaz para acercarla aún más al cliente, hacerlo inteligible, confiable y empático.

**MÉTODO Y TIPO DE ESTUDIO.**

La metodología propuesta para el desarrollo del proyecto está basada en el

método Inductivo, teniendo un enfoque cualitativo y con alcances experimentales.

Se partirá de un estudio bibliográfico de sistemas similares o afines donde se

empleen señales de voz para el comando de interfaces humano-máquina.

Posteriormente, se definirán los requerimientos del sistema de procesamiento de

voz para captar comandos y ofrecer resultados. Como por ejemplo hacer que

reproduzca videos en YouTube, que diga la hora y que pueda buscar información.

# **DESARROLLO.**

Como comienzo a nivel de software utilizaremos:

• **Sistema operativo Windows 10** (Windows 10 es un sistema operativo desarrollado por Microsoft como parte de la familia de sistemas operativos Windows NT).

• **Visual Studio Code 1.4** (Visual Studio Code es un editor de código fuente desarrollado por Microsoft para Windows, Linux y macOS. Incluye soporte para la depuración, control integrado de Git, resaltado de sintaxis, finalización inteligente de código, fragmentos y refactorización de código).

A continuación, el hardware a utilizar es:

* **Computador**: Procesador de 1,8 GHz o superior. Doble núcleo o superior recomendado. 4 GB de RAM recomendado, Espacio en disco duro: hasta 130 GB de espacio disponible, en función de las características instaladas; las instalaciones típicas requieren entre 20 y 50 GB de espacio libre.
* **Audífonos y micrófonos**: En el proceso se utilizó una diadema inalámbrica que unifica estos dos dispositivos, pero no hay problema al usarse por separado.
* **Conexión a Internet.**