

Traductor de texto audio para ayuda en la enseñanza de débiles visuales

Armando Martinez Rios, Adriana Sosa Rojas,
Ruben Naftanael Palomino Flores

Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME) Zacatenco,
México

{armartinezr, asosaro, rubennaf03}@gmail.com

Resumen. Este trabajo muestra el desarrollo de una aplicación destinada a promover la educación inclusiva para personas con discapacidad visual en la Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), unidad Zacatenco, del Instituto Politécnico Nacional (IPN). Diseñada como un traductor, convierte material didáctico como presentaciones, documentos en formato docx y PDF, en archivos de audio MP3. La aplicación desarrollada presenta una interfaz gráfica de usuario que ayuda a cargar sus documentos en formatos pptx, docx y pdf, tanto desde su computadora como desde Google Drive con el apoyo de un asistente virtual. Los documentos son procesados para convertir los caracteres en archivos de audio. Python es el lenguaje de programación principal debido a la gama de herramientas y bibliotecas disponibles y a su código abierto.

Palabras clave: Traductor, asistente, procesamiento de lenguaje natural.

Text-to-Audio Translator for Assistance in the Education of the Visually Impaired

Abstract. This work shows the development of an application aimed at promoting inclusive education for visually impaired people at the Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), Zacatenco unit, of the Instituto Politécnico Nacional (IPN). Designed as a translator, it converts didactic material such as presentations, documents in docx and PDF format, into MP3 audio files. The developed application features a graphical user interface that helps upload documents in pptx, docx, and pdf formats, both from the user's computer and from Google Drive with the support of a virtual assistant. The documents are processed to convert the characters into audio files. Python is the main programming language due to the range of available tools and libraries and its open source nature.

Keywords: Translator, assistant, natural language processing.

1. Introducción

1.1. Debilidad visual

La debilidad visual es una deficiencia que afecta la capacidad de las personas en su agudeza y campo visual, motilidad ocular, distinción de colores o profundidad en los objetos a diferencia de una persona con ceguera que tiene nula visión de su entorno. Sin embargo aunque se tiene un poco de visión, la posibilidad de asistir a la escuela (desde su nivel básico hasta el nivel profesional), se ve afectada por muchos factores. La accesibilidad a los textos que suele ocuparse en las instituciones ya sea en formato digital o en formato físico, se convierte en un obstaculo mas. Estevez, [4] define debilidad visual como cualquier restricción o carencia (resultado de una deficiencia) de la capacidad de realizar una actividad en la misma forma o grado que se considera normal para un ser humano. Se refiere a actividades complejas e integradas que se esperan de las personas o del cuerpo en conjunto, como pueden ser las representadas por tareas, aptitudes y conducta.

En México se promulgó el 29 de junio de 2021 la ley General de Educación donde se adicionaron diversas disposiciones para asegurar la educación inclusiva, tomando para ello los artículos 61 a 68. Agregado a esto, en el Diario oficial de la Federación (DOF) con fecha del 6 de enero del 2023, se menciona que para asegurar una educación inclusiva deben aplicarse métodos, materiales de apoyo y técnicas específicas propiciando la integración de personas con alguna discapacidad [14].

1.2. Políticas públicas en instituciones de educación superior

En el ámbito de las políticas públicas y el apoyo financiero para fomentar la inclusión en México, las Instituciones de Educación Superior (IES) deben alinearse con las directrices y normativas de las entidades decisorias, que son las responsables de asignar presupuestos y recursos adicionales. La modernización de los mecanismos gubernamentales ha motivado a las IES a impulsar la inclusión educativa mediante programas específicos como becas dirigidas a estudiantes de grupos vulnerables y fondos adicionales. Aunque algunas IES desarrollan programas de inclusión independientes de estos incentivos, muchas participan en concursos para obtener recursos adicionales, lo que refleja una dependencia parcial de estas ayudas. Según un informe de Vadillo y Casillas en 2017, solo 17 IES en México implementan programas de inclusión, distribuidas en varios estados del país. A nivel federal, una de las medidas gubernamentales para fomentar la inclusión educativa en la educación superior es el Programa para la Inclusión y la Equidad Educativa (PIEE), establecido en 2015 y aún en vigor. Desde su inicio, este programa ha asignado un promedio anual de 70,075,958 pesos, lo que representa el 0.06 % del presupuesto total destinado a la educación superior en 2018. Estos fondos se distribuyen anualmente entre las diversas IES públicas que compiten con proyectos de inclusión [6].

En la ESIME Zacatenco, no se cuenta con evidencia de algún programa de apoyo destinado a la elaboración de material didáctico o la adaptación de trámites administrativos dirigidos específicamente a personas con debilidad visual. Esta situación es notable debido a la naturaleza educativa de la institución, donde la falta de

estos recursos puede representar una barrera significativa para el acceso al conocimiento por parte de la población estudiantil que presenta esta discapacidad. Además, la ausencia de herramientas adecuadas en los trámites administrativos puede tener un impacto negativo en el proceso académico. Aunque la población con discapacidad visual en el instituto pueda considerarse como una minoría, tanto a nivel nacional como internacional existen normativas que promueven los derechos de estas personas a la educación inclusiva.

En este contexto, surge la necesidad de desarrollar un programa informático que apoye a los docentes en la traducción de sus textos a formato de audio, facilitando así la inclusión de estudiantes con debilidad visual en el ámbito educativo de la ESIME Zacatenco. La carencia de herramientas específicas y adaptadas a las necesidades de esta población en la institución evidencia la urgencia de abordar este problema y contribuir a la mejora del acceso a la educación para todos los estudiantes, independientemente de sus habilidades visuales. Por lo que surge la siguiente pregunta **¿Cómo desarrollar un asistente para que por medio de reconocimiento óptico de caracteres que permita la traducción de texto a audio en estudiantes con debilidad visual?**

1.3. Objetivo general

Desarrollar un prototipo para traducción de información escrita a audio basada en un asistente virtual que ayude a estudiantes de nivel superior con deficiencia visual

2. Estado del arte

El trabajo *"Soluciones digitales de inclusión educativa para personas con discapacidad visual"*, describe la discapacidad visual, abarcando sus tipos, características, causas y prevalencia global. El estudio también examina los obstáculos que enfrentan las personas con discapacidad visual para acceder a la educación superior y explora cómo el desarrollo tecnológico, mediante el uso de herramientas digitales, puede ofrecer soluciones [7]. Mientras que en *"Implementación de un asistente virtual para los estudiantes de pregrado de una universidad peruana"* se muestra la implementación de un asistente virtual destinado a los estudiantes mencionados. Se plantearon tres objetivos específicos: analizar el tiempo de respuesta del asistente, evaluar su eficiencia, y comprobar la percepción de los estudiantes sobre su uso. La evaluación del asistente se realizó en un enfoque cuantitativo con un diseño no experimental y de nivel descriptivo. La población objetivo consistió en 3,228 estudiantes, de los cuales se seleccionó una muestra de 78 estudiantes de ingeniería industrial [11].

El dispositivo vestible lector de textos impresos para niños con discapacidad visual se concibe como un recurso de apoyo en el contexto educativo. Aprovecha la visión artificial, la tecnología de reconocimiento óptico de caracteres (OCR) y la síntesis de voz (TTS) para ofrecer acceso a la información escrita, transformándola en información hablada [8]. Mientras que en España, un equipo desarrolló una plataforma para mejorar la calidad educativa, en la cual implementaron la conversión de texto a

audio (TTS). Su objetivo era crear un entorno digital que ofreciera diversos servicios, permitiendo capturar la información transmitida por el profesor en forma de voz para luego transformarla en datos digitales que pudieran ser procesados, modificados o compartidos con los distintos dispositivos utilizados por los estudiantes. De esta manera, los estudiantes podrían acceder a una representación tanto escrita como auditiva de las clases impartidas por el profesor [15].

El artículo *"Asistente Virtual Académico Utilizando Tecnologías Cognitivas De Procesamiento De Lenguaje Natural"* proporciona un análisis del estado actual del uso de tecnologías de Procesamiento de Lenguaje Natural (PLN) en el desarrollo de asistentes virtuales académicos. Los autores detallan el uso de Dialogflow, una plataforma de desarrollo de agentes conversacionales de Google, como base para la creación del asistente virtual. Destacan cómo esta herramienta permite la interacción fluida entre el asistente y los estudiantes, capacitándolo para responder a una amplia gama de consultas [12].

En el Estado de Puebla, México, opera un grupo académico conocido como ITTEH-CA-09, que ha estado trabajando durante los últimos tres años en la investigación y desarrollo de software destinado a fomentar la inclusión de personas con discapacidad visual y auditiva. Han desarrollado una aplicación que utiliza tecnologías accesibles, como dispositivos móviles, para la captura y procesamiento de imágenes, lo que facilita el acceso de personas con discapacidad visual a entornos educativos y les ayuda a comprender textos con nomenclatura matemática [9].

El artículo *"Asistente Virtual Académico Utilizando Tecnologías Cognitivas De Procesamiento De Lenguaje Natural"* proporciona un análisis exhaustivo del estado actual del uso de tecnologías de Procesamiento de Lenguaje Natural (PLN) en el desarrollo de asistentes virtuales académicos. Los autores detallan el uso de Dialogflow, una plataforma de desarrollo de agentes conversacionales de Google, como base para la creación del asistente virtual. Asimismo, el artículo resalta la integración del asistente virtual con Fulfillment, una herramienta que posibilita la generación dinámica de respuestas y la ejecución de acciones contextuales [12].

En síntesis la comunidad científica ha estado interesada en desarrollar herramientas tecnológicas que apoyen en la educación y en la vida misma de comunidades vulnerables, sin embargo no han permeado lo suficiente y en la ESIME no se tiene ningún tipo de este apoyo.

3. Marco teórico

3.1. Débil visual

La debilidad visual surge cuando una condición ocular afecta al sistema visual y sus funciones relacionadas con la visión. Esta afección puede ser el resultado de diversas enfermedades oculares, como el debilitamiento de los vasos sanguíneos que alimentan la retina, el daño al nervio óptico debido al aumento de la presión ocular, lesiones oculares, o incluso puede ser una condición congénita. A nivel global, las principales causas de debilidad visual y ceguera son los errores de refracción y las cataratas. Los errores de refracción implican problemas con la forma del ojo que

causan una refracción incorrecta de la luz, resultando en visión borrosa. Por otro lado, las cataratas son opacidades en el cristalino del ojo que obstruyen el paso de la luz a la retina [13]. Se estima que al menos 2,200 millones de personas en el mundo presentan algún tipo de deterioro visual, de las cuales al menos 1,000 millones podrían haber prevenido o tratado su discapacidad visual. Sin embargo, solo el 36% de las personas con deterioro visual por errores de refracción y el 17% de aquellas con discapacidad visual a causa de cataratas han tenido acceso a una intervención adecuada. La discapacidad visual representa una carga económica considerable a nivel global, con un costo anual en términos de productividad que se estima en 411,000 millones de dólares estadounidenses [10].

3.2. Educación inclusiva

La educación inclusiva es un enfoque pedagógico que busca garantizar que todos los estudiantes, sin importar sus características individuales, tengan acceso y participación en un entorno de aprendizaje que promueva la igualdad, la diversidad y la tolerancia. Este enfoque implica una revisión profunda de las políticas, prácticas y cultura de las instituciones educativas, y necesita la participación activa de toda la comunidad escolar. En la educación inclusiva, se busca eliminar las barreras que puedan restringir el acceso y la participación de los estudiantes, ya sea por discapacidad, género, origen étnico, orientación sexual u otras diferencias. Se fomenta un entorno en el que todos los alumnos se sientan valorados, respetados y apoyados en su proceso de aprendizaje, así como el desarrollo de habilidades sociales y emocionales para convivir en la diversidad. La educación inclusiva no solo beneficia a los estudiantes con necesidades particulares, sino que también enriquece la experiencia educativa de todos al promover la comprensión mutua, el respeto por las diferencias y la colaboración entre compañeros. Es un proceso continuo que requiere un compromiso constante hacia la equidad y la justicia en la educación [13].

3.3. Aplicaciones de software y herramientas tecnológicas en el ámbito educativo

El empleo de herramientas tecnológicas en el ámbito educativo abarca desde la creación de plataformas web para la formación de estudiantes, hasta la utilización de la inteligencia artificial para adaptar el proceso de aprendizaje, enriquecer la experiencia educativa, automatizar tareas administrativas y analizar datos con el fin de mejorar la experiencia estudiantil en la escuela [2]. Soluciones digitales de inclusión, que aprovechan el avance de la tecnología para mejorar la experiencia educativa de las personas con discapacidad visual. Estas herramientas facilitan una mayor interacción entre estudiantes y profesores, permitiendo diseñar estrategias de trabajo adaptadas a diferentes estilos de aprendizaje y modelos educativos inclusivos [7].

En el caso de hardware, se han desarrollado dispositivos como los de tecnología vestible para asistir el desplazamiento de personas con discapacidad visual, mejorando así su orientación y seguridad en entornos interiores y en algunos casos en exteriores [5].

3.4. Lenguaje Natural y su procesamiento

El Procesamiento del Lenguaje Natural (PLN) es una área de la inteligencia artificial que se enfoca en facilitar la comunicación entre las computadoras y el lenguaje humano. Su objetivo principal es dotar a las máquinas de la capacidad de entender, interpretar y generar lenguaje humano de forma natural. El PLN tiene diversas aplicaciones prácticas, como la traducción automática, la creación de resúmenes, la clasificación de textos y la respuesta automática a preguntas. Para alcanzar estos fines, el PLN se basa en técnicas de disciplinas como la lingüística computacional, el aprendizaje automático y la minería de datos, lo que permite procesar grandes volúmenes de datos en lenguaje natural y extraer información relevante [3]. La Generación Natural del Lenguaje (NLG) es el proceso mediante el cual los sistemas informáticos producen texto de manera automática, con una estructura y contenido que se asemejan al lenguaje humano. Este proceso implica varias tareas, como determinar el contenido del texto, estructurarlo, seleccionar palabras y frases, agregar información relevante, generar expresiones de referencia, asegurar la gramaticalidad y presentar el texto de manera coherente [1].

3.5. Población

La Escuela La Superior de Ingeniería Mecánica y Eléctrica (ESIME) en su unidad Zacatenco, es una de las escuelas fundadoras del Instituto Politécnico Nacional (IPN), con una matrícula de 12,500 estudiantes inscritos para el periodo 2024/2 en dos turnos; este número se divide en cinco carreras: Ingeniería en Comunicaciones y Electrónica (ICE), con casi la mitad de la matrícula; le siguen ingeniería eléctrica (IE) e ingeniería en control y automatización (ICA) con otro 40 % y el otro 10 % dividido en ingeniería en sistemas automotrices (ISISA) e Ingeniería Fotónica (IF) de nueva creación. Esta cantidad de estudiantes la ubica entre las tres escuelas más grandes, en cuanto a población de todo el Instituto Politécnico Nacional (IPN). Esta matrícula se fue incrementando de forma paulatina de 2004 a 2015. En lo que respecta a docentes, entre las cuatro carreras se tiene casi un millar de ellos, de los cuales alrededor de 600 dan servicio a la carrera de ICE y el resto se divide en las otras cuatro carreras. Hasta esta fecha no se tiene un dato exacto oficial de la población con problemas visuales pero si se tiene evidencia no oficial a través de los compañeros docentes.

4. Propuesta de diseño

El sistema propuesto se compone de cinco etapas como son el asistente virtual, la interfaz gráfica de usuario, la conversión, la integración con Google y el empaquetado. Las etapas se muestran en la figura 1. Estas etapas no necesariamente ocurren una después de otra sino que se dividieron así para una mejor explicación.

4.1. Asistente virtual

Esta etapa fué diseñada para interactuar con los usuarios y guiarlos a través del proceso de la selección de archivos para la conversión en audio. Cuando el



Fig. 1. Diagrama a bloques de la propuesta.

usuario inicia el asistente, se le da la bienvenida con un mensaje introductorio que explica las funcionalidades disponibles y las opciones para comenzar. Este asistente está programado para responder a los comandos de voz del usuario utilizando el reconocimiento de voz para interpretar las solicitudes y preguntas del usuario.

Una de las primeras decisiones que el asistente guía al usuario a tomar es la selección de la ubicación del archivo que desea convertir. Esta ubicación puede ser de forma local en el dispositivo del usuario o en alguna cuenta de Google Drive. El asistente recoge la respuesta y procede según la opción seleccionada. Si se elige la opción de la computadora local, el asistente muestra las rutas disponibles y permite al usuario seleccionarla; Luego, muestra los archivos disponibles en esa ruta y permite al usuario elegir el archivo que desea convertir en audio.

Por otro lado, si el usuario elige la opción de Google Drive, el asistente solicita la autenticación y el acceso a la cuenta de Google Drive del usuario. Una vez autenticado, el usuario puede proporcionar un enlace al archivo que desea convertir, y el asistente descarga automáticamente el archivo para su conversión.

4.2. Interfaz gráfica de usuario

La interfaz de la aplicación de conversión de texto a audio se ha diseñado con varias pantallas y elementos interactivos para proporcionar al usuario una experiencia fluida y eficiente. Al iniciar la aplicación, la pantalla de portada se presenta como el punto de entrada, ofreciendo dos opciones principales: "Conversión Manual" o "Conversión por Asistente Virtual". Cada opción se presenta mediante botones que el usuario puede seleccionar con un clic. Si el usuario elige la opción "Conversión Manual", se despliega la pantalla correspondiente (InitialScreen), donde puede seleccionar un archivo de texto de su sistema local. Esta pantalla facilita la selección del archivo mediante un botón de búsqueda, habilita una ventana de diálogo para que el usuario elija el archivo deseado.

En caso de elegir la opción de "Conversión por Asistente Virtual", la pantalla de Asistente Virtual se inicia. Aquí, el usuario es guiado a través del proceso de conversión por el asistente virtual que utiliza reconocimiento de voz. El asistente le permite seleccionar la ubicación del archivo que desea convertir, ya sea en su computadora o en Google Drive, y luego inicia la conversión a audio.

Durante el proceso de conversión, la pantalla de Conversión (PowerPointToAudioConverter) proporciona una vista previa de su contenido y un botón para iniciar la conversión a formato de audio. Una vez que la conversión se completa, se ofrece al usuario la opción de reproducir el archivo de audio resultante para su revisión. Además, la aplicación incluye una Ventana de Ruta (VentanaRuta) que permite al usuario ingresar rutas de archivos locales y enlaces de Google Drive para facilitar la búsqueda y selección de archivos en futuras sesiones.

4.3. Conversión

Esta etapa consiste en la extracción del contenido de archivos de texto o imagen mediante un script, donde se se abren los archivos en formatos como pptx, PDF o docx, y se extrae información de sus diapositivas o páginas según corresponda. Luego, se lleva a cabo el reconocimiento óptico de caracteres (OCR) en las imágenes y, finalmente, se procedía a la lectura en voz alta del contenido de cada diapositiva o página.

El programa correspondiente escrito para el lenguaje python V3.x; consta de dos clases, la clase principal que efectúa el reconocimiento del texto y lo convierte a audio y la clase que muestra el estado de la conversión mediante una interfaz. La clase `AudioConversionThread`, hereda de `QThread` y gestiona las tareas en segundo plano. Su método `init` inicializa la clase con la información necesaria, mientras que el método `run` es donde se lleva a cabo el trabajo principal. Este trabajo implica acciones como extraer texto de archivos PDF, convertir documentos DOCX a PDF, o extraer información de presentaciones PowerPoint, dependiendo de la extensión del archivo.

Por otro lado, la clase `PowerPointToAudioConverter` hereda de `QWidget` y se encarga de interactuar con la interfaz gráfica de usuario. Su método `init` inicializa la clase y establece las variables necesarias, mientras que el método `init_ui` configura la interfaz gráfica.

Estas clases están conectadas de manera que cuando se inicia la conversión de audio desde la interfaz gráfica, se crea una instancia de `AudioConversionThread` en un hilo separado. La señal `finished` de `AudioConversionThread` está conectada a la función `conversion_completed` de `PowerPointToAudioConverter`, lo que permite actualizar la interfaz gráfica de usuario según el estado de la conversión.

4.4. Integración con Google

Tras el desarrollo inicial del asistente virtual, se integró la funcionalidad para gestionar archivos almacenados directamente en la nube, específicamente en Google Drive con el fin de brindar un mayor apoyo al usuario en la obtención de los archivos que desea convertir.

La integración de Google Drive como una opción de almacenamiento proporciona a los usuarios un acceso a sus archivos desde cualquier dispositivo con conexión a internet. Ahora, el asistente virtual permite al usuario seleccionar archivos tanto de su sistema local como de su cuenta de Google Drive. Esta mejora busca que la experiencia del usuario al hacer que el proceso de conversión de texto a audio sea más rápida.

4.5. Empaquetado

`Auto-Py-to-Exe` es una herramienta que permite convertir scripts de Python en ejecutables independientes, incluyendo todas las bibliotecas y archivos necesarios para su correcto funcionamiento. Este proceso de empaquetado es necesario para garantizar que la aplicación sea fácilmente distribuable y ejecutable en sistemas que no tengan instalado Python ni las bibliotecas utilizadas. La creación del ejecutable con `Auto-Py-to-Exe` generalmente implica seleccionar el script principal de la aplicación, así como especificar cualquier archivo adicional que sea necesario incluir, ya sea



Fig. 2. Pantalla de bienvenida al iniciar el asistente.

imágenes, archivos de configuración o bases de datos. Después, AutoPy-to-Exe se encarga de analizar las dependencias del script y empaquetar todo en un único archivo ejecutable que puede ser distribuido y ejecutado en sistemas Windows sin necesidad de tener instalado Python. Ya generado el archivo ejecutable, la aplicación está lista para ser distribuida a los usuarios finales, lo que facilita su uso en diferentes entornos sin requerir configuraciones adicionales.

5. Resultados

La versión final de la aplicación desarrollada con PyQt5 cuenta con las dos opciones de conversión ya comentadas, manual y asistida; llamada asistente virtual, como se muestra en la figura 2.

La conversión manual permite a los usuarios buscar o arrastrar el archivo deseado directamente en la aplicación, ofreciendo una interacción intuitiva y eficiente. La opción asistida, por otro lado, permite cargar el archivo mediante comandos de voz, ya sea desde la computadora o desde un servicio de almacenamiento en la nube, (Google Drive). La ventana de salida se muestra en la figura 3.

Para simplificar la búsqueda de archivos, se pueden añadir rutas específicas de la computadora y estas se van a ir guardando por si se requieren nuevamente o modificar el enlace de Google Drive para seleccionar otro archivo. Además, se integran algoritmos para el Reconocimiento Óptico de Caracteres y la conversión de texto a voz; lo que permite extraer y procesar datos de archivos en formato PPTX, PDF y DOCX, incluyendo tanto texto como imágenes.

Utilizando el motor de Tesseract y el lenguaje de programación Python, el OCR transforma las imágenes de las presentaciones o de los archivos en un formato legible por máquina, que posteriormente se convierte en audio MP3. Este formato de salida es ampliamente compatible y facilita la reproducción en diversos dispositivos compatibles con el mismo.

Cuando se realiza una conversión automática dentro de la aplicación, el usuario tiene que seleccionar el archivo, ya sea PDF, PPTX o DOCX y no realizar ningún



Fig. 3. Pantalla de salida de conversión en proceso.



Fig. 4. Pantalla de fin del proceso.

proceso técnico subyacente. Básicamente, el usuario solo necesita saber la ubicación de su archivo en el dispositivo o en su Google Drive.

Los archivos PDF, PPTX y DOCX son formatos ampliamente utilizados en el ámbito educativo, lo que garantiza que la aplicación sea compatible con los tipos de documentos más comunes empleados por los docentes y estudiantes. El archivo de audio resultante se guarda automáticamente en la misma ubicación que el archivo ejecutable de la aplicación, lo que facilita la tarea de búsqueda para el usuario. Además, la interfaz de la aplicación permite reproducir directamente el archivo una vez que ha sido creado. Cuando termina el proceso se informa al usuario con la ventana mostrada en la figura 4.

La aplicación es completamente portátil y puede utilizarse en cualquier dispositivo con sistema operativo Windows 10 o superior. Sin embargo, es importante destacar que su funcionamiento requiere una conexión a internet, lo que permite acceder a los

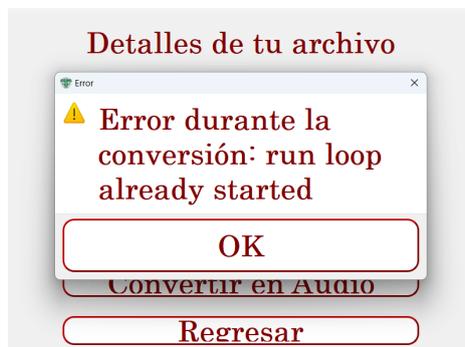


Fig. 5. Pantalla de salida de conversión en proceso.

servicios en la nube y a las actualizaciones necesarias para la conversión de archivos y el procesamiento de datos. Esta combinación de portabilidad y conectividad asegura que los usuarios puedan aprovechar al máximo las capacidades de la aplicación en cualquier momento y lugar, siempre que dispongan de acceso a la red.

Cabe señalar que el asistente está programado para manejar errores durante el reconocimiento de voz, como el que se muestra en la figura 5. Si el usuario menciona un número de ruta o documento fuera de rango, como "ruta siete" por ejemplo, cuando solo hay cinco rutas disponibles, el asistente informará que la selección está fuera de rango y pedirá una nueva selección. También, si no se reconoce la selección de la ruta o el documento, el asistente pedirá nuevamente al usuario que diga la ruta o el documento seguido de su número. El asistente también maneja errores en la detección de audio, ya sea por que el usuario esté hablando muy bajo o porque hubo algún ruido externo que impidió que se escuchara correctamente lo que el usuario dijo, si el asistente no pudo detectar correctamente el audio, dará el siguiente mensaje: "No se detectaron voces claras. Por favor, hable más fuerte y vuelva a intentarlo".

El asistente tiene un límite de intentos para seleccionar rutas y documentos. Si se excede este límite, el asistente despliega un mensaje que indica "Se ha excedido el número de intentos. Por favor, inicia de nuevo el asistente", y el usuario deberá presionar nuevamente el botón de Asistente Virtual para reiniciar el proceso, esto en caso de no poder seleccionar un documento o ruta, o no se haga bien el reconocimiento de voz el programa no siga preguntando de manera continua y detener el proceso. El asistente también incluye una opción para cancelar el proceso en cualquier momento. Si el usuario dice la palabra "cancelar", el asistente dará el mensaje: "Asistente cancelado. Por favor, inicia de nuevo si deseas utilizarlo", y el usuario deberá presionar nuevamente el botón de Asistente Virtual para reiniciar el asistente.

6. Conclusiones

Este proyecto ha alcanzado con éxito el desarrollo de una herramienta de software en forma de un asistente virtual, para convertir material didáctico en formato de audio MP3 en el sistema operativo Windows 10 o superior; cumpliendo la meta al integrar algoritmos como el Reconocimiento Óptico de Caracteres y la conversión de texto

a voz, facilitando así la interpretación eficiente de archivos multimedia. Además, el desarrollo de un asistente virtual es capaz de reconocer voz y ejecutar comandos específicos relacionados con la conversión de material educativo ha enriquecido significativamente la funcionalidad de la herramienta. La implementación exitosa de tecnologías como Tesseract OCR y pyttsx3 ha permitido crear archivos de audio MP3 a partir de contenido textual y gráfico.

Esta aplicación ofrece una solución para la conversión de contenido educativo visual en narraciones auditivas, apoyando a docentes y estudiantes en el proceso de enseñanza y aprendizaje inclusivo. La implementación de múltiples rutas de búsqueda y la compatibilidad con archivos almacenados en la nube mejoran la flexibilidad y accesibilidad del software, al tener disponibles los archivos de conversión en todo momento.

Se cumplió el objetivo planteado al lograr desarrollar una plataforma para conversión de texto a audio en el lenguaje de programación Python y compatible con el sistema operativo Windows 10 o superior; alcanzando el registro de derechos de autor no. 03-2024-122013344100-01 por el Instituto Nacional de Derechos de autor.

Referencias

1. Bernardos-Galindo, M. d. S.: ¿Qué es la generación de lenguaje natural? Una visión general sobre el proceso de generación. *Inteligencia Artificial, revista Iberoamericana de inteligencia artificial*, vol. 11, pp. 105–128 (2007) doi: 10.4114/IA.V11I34.911.
2. Berrones, L., Salgado, S.: La aplicación de la inteligencia artificial para mejorar la enseñanza y el aprendizaje en el ámbito educativo. *Esprint Investigación*, vol. 2, pp. 52–60 (2023) doi: 10.61347/ei.v2i11.52.
3. Durán-Chinchilla, C. M., García-Quintero, C. L., Rosado-Gómez, A. A.: Aplicación del procesamiento del lenguaje natural como técnica de análisis en la producción textual, caso estudiantes de Ingeniería de Sistemas UFPSO. *Revista Boletín Redipe*, vol. 11, no. 1, pp. 489–497 (2022) doi: 10.36260/rbr.v11i1.1656.
4. Estévez-Alarcón, J. A., Ortiz-Vázquez, P. E., Ticante-Calderón, C. E.: Discapacidad visual como causa de depresión en adolescentes. *Ciencias de la Salud en la Ibero Puebla*, (2017)
5. Fariña-Molina, N. R., Ibarrola-Chamorro, C. R., Ramírez-Aquino, A. T.: Desarrollo tecnológico de un brazalete sensorizado (CiegNest) para apoyo a personas con discapacidad visual implementando software y hardware libre. In: *Anais do XIX Congresso Latino-Americano de Software Livre e Tecnologias Abertas*, pp. 149–152 (2022) doi: 10.5753/latinoware.2022.228046.
6. González-García, J. A., Zúñiga-Llamas, A., Arce-Casas, P. O.: Un panorama sobre la cobertura educativa a nivel superior en México para personas con discapacidad. *IE Revista de Investigación Educativa de la REDIECH*, vol. 12, pp. 1–17 (2021) doi: 10.33010/ie_rie_rediech.v12i0.1171.
7. Hernández-Bermejo, M. A., Leon-Pacheco, D., Minchala-González, J., Lema-Lema, E.: Soluciones digitales de inclusión educativa para personas con discapacidad visual. *Killkana Técnica*, vol. 7, pp. 19–32 (2024) doi: 10.26871/killkanatecnica.v7i2.1067.
8. Mejía-Campos, R. G.: Dispositivo Wearable lector de textos impresos para niños con discapacidad visual como herramienta de apoyo en el proceso de enseñanza y aprendizaje en el ámbito escolar usando técnicas de visión artificial, OCR y TTS. Tesis de Ingeniería, Benemérita Universidad Autónoma de Puebla, (2022)

9. Olguín-Gil, L. E., Vázquez-Zayas, E., Vázquez-Guzmán, F., García-Ortega, I., Rodríguez-Ramírez, F.: Aplicación móvil que traduce ecuaciones matemáticas a voz mediante OCR. *Pistas Educativas*, vol. 42, pp. 74–90 (2020)
10. OMS: Ceguera y discapacidad visual, (Feb 2024)
11. Ortiz-Elías, A. N., Dávila-Morán, R. C.: Implementación de un asistente virtual para los estudiantes de pregrado de una universidad peruana. *Revista Conrado*, vol. 19, pp. 121–128 (2023)
12. Torres-Echeverri, M. M., Manjarrés-Betancur, R.: Asistente virtual académico utilizando tecnologías cognitivas de procesamiento de lenguaje natural. *Revista Politécnica*, vol. 16, no. 31, pp. 85–96 (2020) doi: 10.33571/rpolitec.v16n31a7.
13. Vallejos-Garcías, V., Castro-Durán, L.: Educación inclusiva. *Revista Internacional de Humanidades*, vol. 12, pp. 1–11 (2023) doi: 10.37467/revhuman.v12.4726.
14. Villa-Guardiola, V. J., Lorenzo-Loredo, A. G., Sayago Gonzalez, I. N.: Educación superior incluyente para personas con discapacidad visual en la Universidad Autónoma de Guerrero, México. *Saber, Ciencia y Libertad*, vol. 17, pp. 474–493 (2022) doi: 10.18041/2382-3240/saber.2022v17n2.9332.
15. Villarrubia-González, G., de-Paz-Santana, J. F., Moreno-García, M. N., et al.: E-SPEECH: plataforma de apoyo a la comprensión oral mediante tecnologías disruptivas, (2022)