

# Evaluación e implementación de estrategias de asociación de voz o texto con contenido multimedia para la generación de repositorios digitales mediante técnicas de Inteligencia Artificial.

Jaime Emilio Zavala Barrios, Víctor Manuel Zamudio Rodríguez

*División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México/ IT de León*  
León Gto. México  
{M2020116, vic.zamudio}@leon.tecnm.mx

Carlos Lino Ramírez and David Asael Gutiérrez Hernández

*División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México/IT de León*  
León Gto. México  
{carloslino, david.gutierrez}@leon.tecnm.mx

**Resumen**— En México se estima que aproximadamente un millón 300 mil personas padecen la enfermedad de Alzheimer, cuya cifra representa ser entre el 60 y 70 por ciento de los mexicanos diagnosticados con demencia. El Alzheimer afecta con mayor frecuencia a las personas mayores de 65 años. [1] Actualmente no existe una cura para la enfermedad. En esta investigación se plantea la implementación del procesamiento del lenguaje natural para el análisis de recuerdos y la extracción de información con las técnicas de identificación de partes de discurso y reconocimiento de identidades, con la finalidad de preservar los recuerdos en un repositorio digital y enriquecer las historias con la sincronización de material multimedia.

**Palabras Clave**—*machine learning, data mining, text mining, storytelling, tokenization y named-entity recognition.*

## I. INTRODUCCIÓN

En el presente artículo se presenta un trabajo en progreso relativo al almacenamiento de información (basado en recuerdos) en repositorio digitales, apoyado con estrategias de procesamiento de lenguaje natural y sincronización de contenido multimedia. La motivación de la investigación en desarrollo tiene como finalidad la preservación de recuerdos de usuarios con algún tipo de demencia (incluyendo Alzheimer), a través de un repositorio digital. Este repositorio permite la aplicación de técnicas de procesamiento de lenguaje natural para la clasificación y extracción de la información contenida.

El Alzheimer es un tipo de demencia que afecta directamente a la memoria, el comportamiento y el pensamiento, al ser una enfermedad crónica los síntomas en la última fase avanzan hasta el punto de que logran interferir con las tareas cotidianas.

A inicios del siglo XX, se describió por primera vez por Alois Alzheimer las lesiones neuropatológicas asociadas con un tipo particular de demencia senil. Este hecho fue la pauta para establecer un nuevo concepto etiológico que en la actualidad se le conoce como demencia por enfermedad de Alzheimer [1]

Dicha enfermedad aumenta gradualmente, por lo cual desde su detección se observa como los síntomas de demencia empeoran progresivamente durante varios años. Una de las características en etapas tempranas es la pérdida de memoria sobre detalles pequeños y es considerada como una pérdida de memoria tenue. Sin embargo, las personas que llegan a padecerla en las etapas más avanzadas desarrollan la pérdida de capacidad para mantener una conversación y posteriormente, en la última etapa de la enfermedad, la posibilidad de responder a su entorno [2].

En promedio, una persona con Alzheimer vive de 4 a 8 años después del diagnóstico, pero con una probabilidad de hasta 20 años, dependiendo de múltiples factores como el cuidado, la detección temprana de la enfermedad, el uso de los medicamentos entre otros, pero como factor principal la evolución de la enfermedad [3].

Es entonces que este trabajo gira en torno a la preservación de la memoria de las personas diagnosticadas con la enfermedad, mediante la ejecución de estrategias digitales e implementación de algunos algoritmos de inteligencia artificial para el análisis del contenido, extracción de información, sincronización de contenido multimedia con la narrativa automatizada (videos, fotos, grabaciones, etc.), así como el mapeo de relaciones de datos.

La conservación de la memoria de una persona es un tema muy relevante en el contexto del Alzheimer, pues tal como se abordó con antelación, en las primeras etapas de la enfermedad, la persona comienza a olvidar detalles menores de su día y mientras que la enfermedad avanza, de manera progresiva, aquello que parecía claro y fácil de recordar ahora cuesta completar o recuperar fragmentos y datos de su memoria [1].

La memoria puede sufrir alteraciones por diferentes factores tales como los emocionales [4], entendiéndose estos como aquellos que protegen a la persona de una mala experiencia, por otro lado, el tiempo también permite agregar o remover detalles del suceso original [5].

La memoria puede estimularse y recordar fragmentos olvidados gracias a estímulos sensoriales, o bien con ayuda de material audiovisual, por ejemplo; al observar una fotografía, escuchar una canción o audios, así como ver un algún video pues estos materiales funcionan como detonantes que permiten revivir o recordar experiencias [6].

Con el avance de las herramientas digitales, se puede plantear la concepción y generación de nuevas estrategias para el almacenamiento digital, permitiendo a la persona almacenar y proteger el momento. Aunque el eje de la investigación se concentra en pacientes con Alzheimer, la herramienta estará disponible para cualquier persona, sin importar si es o no diagnosticada con esta enfermedad.

## II. JUSTIFICACIÓN

“La demencia es un nombre colectivo para los síndromes cerebrales degenerativos progresivos que afectan la memoria, el pensamiento, el comportamiento y las emociones” [7].

A partir de los datos oficiales se observa como México se está convirtiendo en un país con más personas mayores que infantes menores de 5 años, se prevé que para el año 2050, la proporción de las juventudes de 15 años en la población total disminuirá de 30.0% a 17.4%. De este modo, el incremento en la población de personas adultas mayores iría en aumento con la utilización de los servicios de salud destinados a la atención de enfermedades propias del envejecimiento como las demencias [8].

La demencia es uno de los factores que más contribuyen a la discapacidad ya la dependencia en las personas adultas mayores. La Enfermedad de Alzheimer (EA), es el tipo de demencia más frecuente y actualmente se conoce que existe una prevalencia del 7.3% y una incidencia de 27.3 (1000 personas/año) de la población adulta mayor mexicana para la EA [9]. En el sitio oficial de la secretaria de salud de México se indica que existe aproximadamente un millón 300 mil personas padecen la enfermedad de Alzheimer, cifra que representa entre 60 y 70 por ciento de los diagnósticos de demencia en el país y que afecta con mayor frecuencia a las personas mayores de 65 años [1]. La enfermedad de Alzheimer se genera debido a cambios en el cerebro por la presencia de la proteína llamada beta amiloide, que se acumula frecuentemente en el lóbulo temporal. Dicha toxina provoca inflamación y muerte progresiva de neuronas [3]. La Organización Mundial de la Salud (OMS) calcula una cifra aproximada de 60 millones de personas a nivel global que viven con Alzheimer, de las cuales 8.1 por ciento son mujeres y 5.4 por ciento, hombres mayores de 65 años [3].

Para una atención adecuada de la persona con enfermedad de Alzheimer, es necesario el diagnóstico temprano y tratamiento integral de alta especialidad que se basa en medicamentos que estimulan y ayudan a prolongar la vida de las neuronas de la memoria para mejorar su calidad de vida. [3].

En el presente trabajo, el procesamiento de recuerdos permitirá a pacientes con Alzheimer poder acceder a recuerdos aun con en el avance de la enfermedad, extraer información de las transcripciones que serán enriquecidas con material multimedia para estimular experiencias y hacerlas más vividas. Las memorias almacenadas estarán disponibles para ser compartidas con familiares y descendencia, de tal modo que generaciones distintas de una familia puedan conectarse, siempre que la persona dé su consentimiento.

Al preservar las memorias del paciente, a pesar de la pérdida de detalles sobre sus recuerdos, podrá revivirlos por medio de la aplicación, como la primera vez que fueron subidos al ser complementados con el contenido multimedia.

El procesamiento de lenguaje natural es una técnica indispensable que permite analizar las transcripciones de los recuerdos, extraer la información y complementarlos con material multimedia relacionado a los datos extraídos. Estos contenidos multimedia pueden ser enriquecidos con archivos subidos de otras experiencias previas, información disponible en la red o con la que contenida en el sistema.

Al almacenar las memorias de la persona dentro de un repositorio digital, se evitará que el paso del tiempo u otros factores como enfermedad, medicamentos, lesiones modifiquen la historia y el recuerdo.

El procesamiento de lenguaje natural se llevará a cabo valiéndose de las transcripciones en formato texto de los audios, controlándose desde la aplicación el formato de en mp4, y con la captura de recuerdos de manera escrita se aceptarán archivos en diferentes formatos: PDF, Text. y. Doc, para su posterior conversión en formato de texto plano.

La conservación del audio permitirá validar la precisión de las transcripciones escritas con respecto al audio original e identificar palabras que al sistema le cuesta diferenciar por la velocidad del habla de cada persona.

## III. MARCO TEORICO

### A. Parte del discurso / (Part of Speech Tagging - POS)

Parte del discurso se refiere a la categoría de palabras o los términos léxicos en un idioma.

El Español contiene 8 grupo de clases de palabras que hacen referencia a estos términos léxicos: sustantivo, verbo, adjetivo, adverbio, pronombre, preposición, conjunciones y nombres. En la siguiente tabla se muestra un ejemplo de clases de palabras, su abreviación basada en “*The Penn TreeBank Tagset*” [3] y en la tercera columna se logra apreciar un conjunto de palabras de cada categoría.

TABLA I. TABLA DE EJEMPLO DE ASIGNACIÓN DE ETIQUETAS.

Término léxico	Etiqueta estándares	Ejemplos
Sustantivo	NN	Algo, nada
Verbo	VB	Comer, correr
Preguntas	WRB	Cómo, Cuándo
Artículos	DT	El, la, los

El *Penn Treebank*, en sus ocho años de funcionamiento (1989–1996), produjo aproximadamente 7 millones de palabras de texto etiquetado de parte del discurso, 3 millones de palabras de texto analizado esqueléticamente, más de 2 millones de palabras de texto analizado para la estructura de argumento predicativo y 1,6 millones de palabras de texto hablado transcrito y anotado por falta de fluidez en el habla [4].

TABLA II. THE PENN TREEBANK TAGSET. ADAPTADO DE [10]

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner

EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRPS	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund, or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WPS	Possessive wh-pronoun
WRB	Wh-adverb

**B. Reconocimiento de entidad nombrada.**

Con la tarea de identificación de identidades, se pueden reconocer patrones numéricos para ser categorizados como entidades de tiempo, formulas químicas, entidades monetarias, expresiones de porcentajes y patrones de fechas.

Anterior a los modelos de aprendizaje profundo se utilizaba modelos probabilísticos tales como los Modelos Ocultos de Markov [11] y Campos Aleatorios Condicionales [12].

A partir de la implementación del Aprendizaje profundo o Deep Learning se han ido desarrollando diferentes enfoques para poder resolver la tarea del reconocimiento de entidades nombradas, algunos son implementando modelos con incrustaciones de vectores *Word2Vec*, *Glove*, *Senna* y *FastText* [13].

**IV. METODOLOGÍA EXPERIMENTAL PROPUESTA**

El desarrollo de la aplicación se llevó por medio del software Android Studio, utilizando el *SDK Software Development Kit Flutter*, el cual es de código fuente abierto creado por Google [14].

Se desarrolló una aplicación híbrida para sistemas operativos de Android y IOS. Esta aplicación busca conservar los recuerdos por medio del discurso oral y escrito del usuario contemplando en el desarrollo de la interfaz, las brechas [15] generacionales para simplificar sus funciones.

El sistema operativo de Apple iOS, y diseñado exclusivamente para el hardware producido por dicha compañía fue el primero en implementar el interfaz usuario multitouch y, en buena medida, se puede ver como el responsable de la explosión y universalización en el uso de dispositivos móviles. Al igual que el sistema operativo que emplean para sus equipos de escritorio, MacOS X, iOS está basado en el núcleo Darwin, derivado de FreeBSD, un sistema libre tipo Unix [16].

El sistema operativo Android Diseñado por la compañía Google, basa la mayor parte de su operación en software libre (un núcleo Linux, máquina virtual Java, y muchas de las bibliotecas de sistema comunes en sistemas Linux), agregando una capa de servicios propietarios. La estrategia de Google ha sido inversa a la de Apple: en vez de fabricar sus propios dispositivos, otorga licencias para el uso de este sistema operativo a prácticamente todos los fabricantes de hardware, con lo que la amplia mayoría de los modelos de teléfonos inteligentes y tabletas corren sobre Android [16].

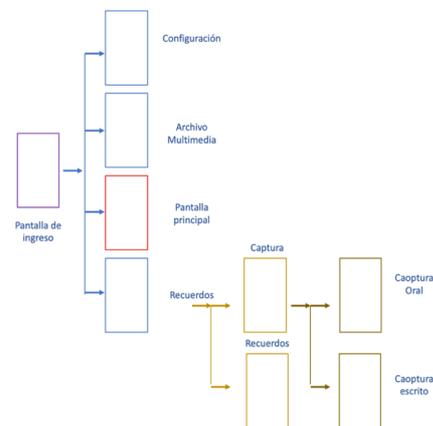


Figura. 1. Diagrama de estructura de aplicación.

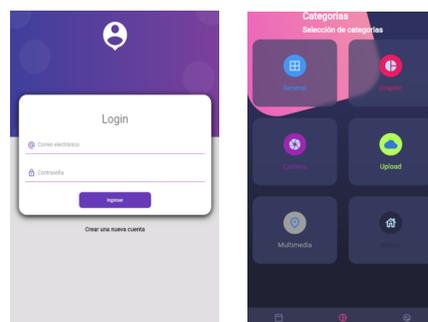


Figura. 2. Pantalla de ingreso a la aplicación y categorías.

Flutter funciona con Dart, un lenguaje optimizado para aplicaciones rápidas en cualquier plataforma que permite que desde una única base de código se la aplicación se pueda implementar en múltiples dispositivos: dispositivos móviles, web, de escritorio e integrados [14].

La aplicación móvil desarrollada en la investigación busca ser la herramienta que funcione como puente entre la parte de la captura de los recuerdos en su versión oral o escrita, con el repositorio digital y el sistema de narrativa inteligente.

Entre las características que se diseñaron para la aplicación se encuentran aquellas que toman en cuenta al paciente:

- La aplicación debe ser responsiva (se adapta a las diferentes dimensiones de equipos).
- Debe contar con características de usos intuitivo para el usuario.
- Tener las funciones integradas para poder actualizarse de manera remota.
- La herramienta debe estar diseñada para ser compatible con el mayor número de dispositivos Android y permitir el uso de código para el sistema IOS.

El primer paso del proyecto se enfoca en la captura de los recuerdos de las personas para la extracción y generación de transcripciones en texto plano “.txt” para su posterior procesamiento, como se muestra en la Figura 3.



Figura 3. Diagrama de captura de recuerdo.

En la Figura 3. Se observa la captura de los recuerdos por parte de grabaciones, las cuales se almacenarán ya que son un medio de validar si las transcripciones se están realizando correctamente.

En la parte de captura, el proyecto busca mantener la integridad de la historia en el caso de que la persona suba una igual posteriormente, por lo cual se considerará la primera historia subida al sistema, como la original. A partir de este relato se pueden realizar comparaciones con cargas posteriores, cuyo objetivo es detectar las variaciones entre ellas, como cambios, la disminución de detalles, la falta de participantes entre otros aspectos.

Teniendo en formato de texto plano los recuerdos o textos a procesar se puede implementar las técnicas de identificación de partes de discurso y reconocimiento de entidades.

Entender la identificación de discurso como la generación de un etiquetado, que generalmente se modela como una tarea estructurada: la etiqueta de la primera palabra puede depender de la etiqueta de la tercera, pero se puede aproximar bastante bien clasificada cada palabra de forma aislada en una etiqueta (“*part-of-speech*” en inglés, siglas *POS*), basada en una ventana de dos palabras a cada lado de la palabra. Si se etiquetan las palabras en un orden fijo, por ejemplo, de izquierda a derecha, también se puede condicionar cada predicción de etiquetado a las predicciones de etiquetas realizadas en etiquetas anteriores [17].

En la NER se proporciona un documento y se necesita encontrar entidades nombradas, así como categorizarlas en un conjunto predefinido de categorías como locación, organización, personas y otras categorías [17]. Algunas palabras pueden recibir una categoría diferente de acuerdo con el contexto en el que se emplean dentro del texto.

Se optó por un arreglo que almacene la información para cada palabra. Esta información va a estar separada por categorías relacionada a la implementación en las tareas de sincronización como se aprecia en la Figura 4.

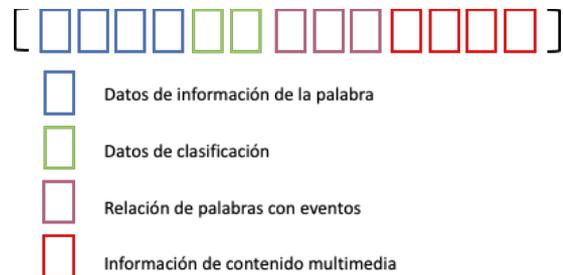


Figura 4. Diagrama de estructura de cada elemento del texto, excluyendo signos de puntuación.

En la Figura 5 Se desglosa más a detalle las funciones dentro de cada una de las categorías.

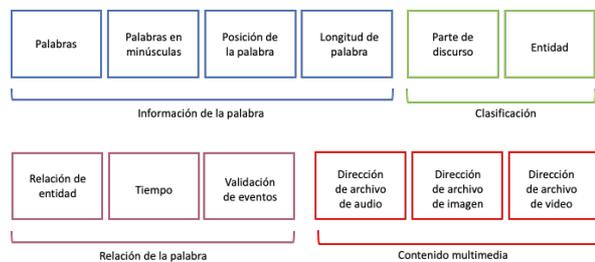


Figura 5. Desglose de procesamiento de cada palabra dentro del recuerdo.

Para algunas de las funciones de la aplicación y del sistema se consideró el uso de la librería SpaCy.

SpaCy es una biblioteca gratuita de código abierto para el procesamiento avanzado del lenguaje natural (NLP) en Python, diseñada para ayudar a crear aplicaciones NLP, cuyo uso no esté restringido por cuota de palabras o eventos, haciendo que no sea un servicio consumible y siendo una opción gratuita [18].

Por medio de la librería referida, se pueden construir sistemas de extracción de información o comprensión del lenguaje natural, procesar y entender grandes volúmenes de texto.

En lo que corresponde a la parte experimental, se utilizó el modelo en español de nivel medio, sin desestimar las palabras vacías (“*stop words*” en inglés) las cuales son palabras muy frecuentes en el idioma, pero cuyo uso no genera un significado importante al contenido.

Se asume que el conjunto de textos fue revisado y ajustado para que su composición tenga sentido y cumpla con las reglas gramaticales del idioma implementado.

El proceso anterior hace referencia a la corrección de errores en la escritura o transcripción de determinadas palabras para proceder con el procesamiento y extracción de información de los documentos.

Las palabras son posteriormente separadas en elementos individuales denominados tokens mismos que se obtienen al separar cada frase en sus elementos individuales sin considerar los signos de puntuación. Esto se puede llevar a cabo por medios de ciclos que recorran e identifiquen cada espacio vacío, o la aparición de un signo de puntuación en el caso del español.

En la siguiente Figura 7 se puede apreciar un diagrama del recorrido que permitirá identificar cada token y encapsularlo en frases o párrafos según la función. Como se ha mencionado anteriormente, se debe de ajustar el proceso para considerar las reglas gramaticales del idioma trabajado.

Dentro del español se implementa un recorrido de izquierda a derecha, cada espacio en blanco delimita a una palabra, al igual que la presencia de ciertos signos de puntuación: puntos, comas, entre otros y dentro del texto plano se reconoce a la combinación de “.\n” como saltos de párrafo y a la presencia de “\n” como un salto de línea, que puede o no removerse dentro de los procesos o implementarse para dar una estructura.

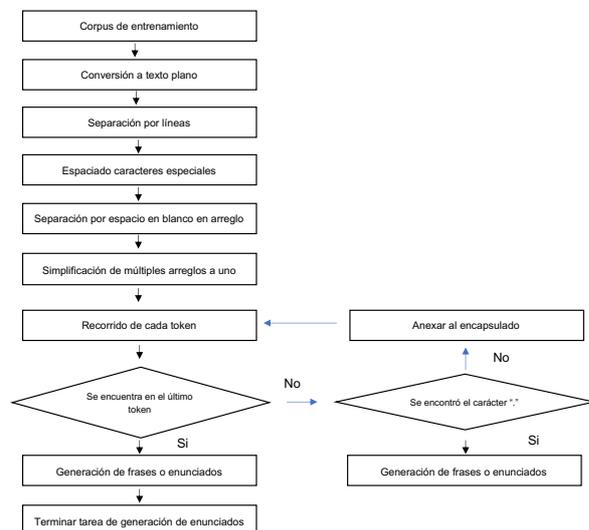


Figura 6. Diagrama de procesamiento de texto.

En las siguientes líneas de pseudocódigo se observa el desglose de los pasos para la selección y sincronización del material de las transcripciones.

Es importante mencionar que no todas las categorías de POS en que serán clasificadas las palabras requieran una sincronización con el material multimedia, como son los casos de artículos, verbos, conjunciones entre otras categorías, pero la presencia anterior o posterior puede aumentar el tiempo del contenido a mostrarse de la palabra que si requiere su sincronización. Ello se refleja en el código de la Figura 7.

```

    Selección de texto (recuerdo).
    Extracción de frases.
    Duplicado de frase
    Eliminación de signos ortográficos dentro de duplicado.
    Generación de bolsa de palabras.
    Incorporación de palabras nuevas a la bolsa de palabras del perfil del usuario.
    Recorrido de todas las frases dentro del texto.
  
```

```

    Extracción de Tokens por frase.
    Clasificación de partes de discurso
    Clasificación de entidades.
    Recorrer todos los tokens dentro de una frase.
    Rellenado de arreglo de datos por token.
    Si se requiere material por tipo de POS.
    Revisión de material multimedia.
    Si existe material dentro del perfil del usuario.
    Selección de archivos multimedia.
    Si no se consulta la base de datos de la aplicación.
    Si existe material.
    Selección de archivos multimedia.
    Si no existe material
    Descarga de la red los archivos multimedia.
    Guardar el archivo con descripción del Token.
    Selección de archivos multimedia.
    Si no se requiere material por tipo de POS.
    Continuar con siguiente Token.
    Subir nuevo archivo al servidor.
    Ligar recuerdo del usuario a recuerdo dentro de la aplicación.
  
```

Figura 7. Diagrama de procesamiento de recuerdos y sincronización.

Los recorridos por cada frase permiten identificar las expresiones regulares por medio de n-gramas, controlando la cantidad de palabras que se agrupan.

Un n-grama es un conjunto de n elementos consecutivos en un documento de texto, que puede incluir palabras, números, símbolos y puntuación. Los modelos de n-gramas son útiles en muchas aplicaciones de análisis de texto en que la secuencia de palabras es pertinente, tales como análisis de sentimiento, clasificación de texto y generación de texto. El modelado de n-gramas es una de las técnicas utilizadas para convertir texto de un formato no estructurado a un formato estructurado [19].

Una de las ventajas de la identificación de las expresiones regulares, es que se puede poseer una mayor precisión sobre el dominio de patrones con los que se está trabajando. Entre más recuerdos va ingresando la persona, más se podrá personalizar el sistema. Se puede determinar la frecuencia de algunas palabras, la extensión de su vocabulario, los patrones del habla, etc., lo cual permitirá facilitar la extracción de información.

El proceso permite identificar una entidad que va agrupada en una expresión regular, aumentando la certeza de que se ha hecho una clasificación con una precisión más elevada. Se puede llegar a aumentar la precisión de que todas las entidades que se capturan coincidan con determinado patrón, son del mismo tipo.

Aunado a esto, el hacer adaptable el sistema al uso de expresiones ahorra los reentrenamientos constantes del modelo de reconocimiento.

El uso de las expresiones regulares se considera para patrones individuales de una persona ya que es susceptible a patrones definidos y no llegan a ser reglas que puedan repetirse en cada caso.

Es importante destacar que los n gramas son susceptibles a palabras con: caracteres extras, repetición de letras, errores ortográficos entre otros, por eso se hace énfasis en la revisión y correcciones como pre-procesamiento de los textos para posteriormente hacer su análisis y extracción de información.

Mediante la creación de un proyecto de reconocimiento de entidades con nombre personalizado, se pueden etiquetar datos de forma iterativa, entrenar, evaluar y mejorar el rendimiento del modelo, ayudando al mapeo y extracción de datos.

La obtención de bolsa de palabras por el procesamiento de cada memoria permitirá realizar una actualización recurrente dentro del perfil del usuario. Y al segregarse a usuarios de la misma zona geográfica se podrá detectar regionalismos dentro de un mismo idioma e incorporar estas nuevas palabras al sistema para mejorar los niveles de precisión.

Para todas aquellas palabras desconocidas se optó por implementar un modelo de extracción de datos de la web que ingresa a la página oficial de la Real Academia de la Lengua Española (RAE), buscando la palabra deseada y extrayendo la implementación de la palabra en el español. El código mostrado en la Figura 8 ejemplifica la secuencia de pasos para realizar esta tarea.

La RAE, es una institución cultural dedicada a la regularización lingüística entre el mundo hispanohablante [20].

```

Selección de la palabra o lista de palabras
Abrir navegador predefinido.
Ingreso a sitio de la Real Academia Española (RAE)
Si es solo una palabra
    Búsqueda por palabra
    Extracción de categorías del discurso de las palabras
    Extracción de la entidad.
Si no es solo una palabra
    Recorrido de toda la lista de palabras.
    Búsqueda por palabra
    Extracción de categorías del discurso de las palabras
    Extracción de la entidad.

Generación de bolsa de palabras.
Incorporación de palabras nuevas a la bolsa de palabras del perfil del usuario.
Recorrido de todas las frases dentro del texto.
    Extracción de Tokens por frase.
    Clasificación de partes de discurso
    Clasificación de entidades.
Cerrar navegador.
    
```

Figura. 8. Seudocódigo de búsqueda de categorías y entidades por medio de Spacy.

En la Figura 9 se muestra el proceso de sincronización de contenido multimedia con la información obtenida de cada una de las palabras al ir recorriendo cada una de ellas de izquierda a derecha.

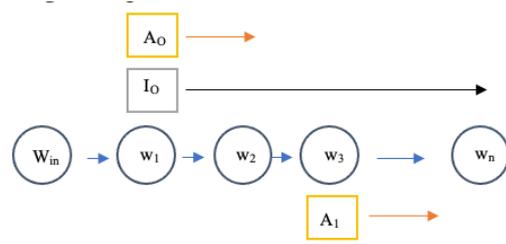


Figura. 9. Sincronización de elementos con narrativa.

Explicado más a detalle, los algoritmos deberán ser capaces de generar un preprocesamiento y llevar a cabo el análisis del contenido dentro del recuerdo antes de comenzar de que la persona comience la narrativa, se deben estimar los tiempos de duración de los elementos a aparecer dentro del recuerdo y de manera automática los llamados.

Se puede hacer una comparación al proceso de separar la oración, que se convertirá de discurso escrito a oral en palabras individuales y considerarlas como estados independientes de otros dentro de un arreglo secuencial. Al visitar cada uno de estos estados individualmente, y de manera ordenada se puede llamar a los eventos que contienen el estado de manera paralela a otros eventos generados por otros estados o pausar un evento que inicio en un estado q<sub>x</sub>. de la lista de palabras que conforman una frase como en la Figura 10.



Figura. 10. Diagrama secuencial de una frase

Nota: cada uno de los estados “q” hace referencia a una palabra de una frase, siendo el estado q<sub>n</sub> la palabra final o el signo especial “.” que indica el final de la oración, esto puede variar si el diseño contempla a los signos de puntuación como estados. Dicha secuencia se aprecia en la Figura 11.

El llamado de los elementos multimedia debe ser un proceso paralelo asíncrono ejemplificado en la Figura 10. De tal manera que no dependan el llamado de un elemento de la aparición de otro si no del contexto y el momento en el que se encuentra la historia así pudiendo complementar los contenidos de imágenes, audio o video sin requerir que estén todos u otro elemento, como se muestra en la Figura 11.

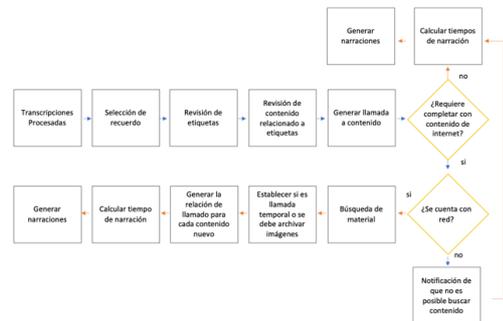


Figura. 11. Diagrama de flujo de procesamiento de recuerdo

A partir de este punto al generar la narrativa del recuerdo, una imagen puede ser mostrada en la pantalla del dispositivo

sin pausar la narrativa y a su vez, otro evento puede solicitar un efecto sonoro o arrancar de una pista audiovisual sin opacar la voz principal o parar los otros dos eventos, siendo esto calculado por el sistema.

Los eventos se pueden poner a dormir o concluir según indique el arreglo la configuración predefinida para cada palabra de acuerdo con el texto. Por lo cual, la manera que se comparta un elemento relacionado a una palabra en texto va a variar al contexto de la historia y a los otros elementos que se encuentran dentro de ella.

La opción de complementar las historias con elementos fuera de los archivos, el usuario o de los archivos generales cargados al momento queda a discreción de usuario. De igual modo queda a discreción pues no se requiere la función del llamado a elementos multimedia para la narración por defecto.

Al activarse la función de llamada a elementos multimedia, el algoritmo será capaz de revisar con el primer preprocesamiento los elementos con los que cuenta en su base de datos interna del dispositivo que tengan una relación al recuerdo, complementarlo con elementos del repositorio general o permitir una descarga de elementos de la red tal como imágenes, videos, elementos de audio, etc., para complementar los vacíos de elementos.

Si el dispositivo cuenta con una conexión a la red y estos vacíos son de carácter universal como ciudades, países, lugares públicos, por medio de algoritmos de raspado de web (“*scraping*” en inglés) se podrá descargar los elementos que complementen la historia y añadirlos a una base de datos temporal de donde serán llamados. Si, por el contrario, son de carácter privado no se llamará a ningún archivo o elemento haciendo referencia a fotografía personales con las que el sistema no cuente.

El raspado web es la práctica de recopilar datos a través de cualquier otro medio que no sea un programa que interactúa con una API también considerado a través de un humano que usa un navegador web. Comúnmente se logra este proceso al escribir un programa automatizado que consulta un servidor web determinado, solicita los datos ayudándose de identificadores de lenguaje HTML y de otros archivos que compongan la página o páginas web seleccionadas y luego analiza los datos para extraer la información necesaria [21].

El web scraping abarca una amplia variedad de técnicas y tecnologías de programación, como el análisis de datos y la seguridad de la información [21].

## V. RESULTADOS PRELIMINARES.

Como puede apreciarse, existen numerosas tareas y múltiples aproximaciones para el procesamiento del lenguaje natural necesario para poder llevar a cabo la implementación de contenido multimedia en las historias.

Los primeros desarrollos de la aplicación se enfocaron en realizar el proceso de transcripciones en tiempo real de los recuerdos, pero la configuración del uso de librerías generó limitantes. La primera es que la aplicación, al detectar una pausa el relato, detenía por completo la transcripción y se debía iniciar de nuevo, por ende, se optó por la captura de un audio en formato MP4 con la aplicación y por medio de código escrito en Python generar la conversión a WAV para poder

aplicar la librería de Google de conversión de discurso hablado a escrito.

La preservación del formato de audio permitirá medir la precisión en el proceso de conversión del discurso oral al escrito en futuros trabajos (“*speech to text*”, en inglés). Es entonces que se limitó la duración de las grabaciones a un máximo de 30 min para los archivos de audio subidos por la aplicación.

Los archivos de audio de la aplicación fueron cargados y almacenados en una base de datos de Firebase, de donde se extraían para la conversión a formato WAV y la generación de las transcripciones a formato de texto plano por medio del lenguaje Python, en la conversión de mp4 a WAV no se modificaba los tiempos de duración de los archivos.

En el caso del uso de carga de transcripciones por medio de documentos, o de escritura de recuerdos directamente en la aplicación los textos ingresados pasaron por un proceso de conversión a texto plano por medio de la librería de Python Tika, la cual detecta el formato en el que se encuentra y lo convierte a texto plano. Para esta tarea es necesaria el uso de conexión a internet ya que la librería procesa el documento de manera remota. Para la implementación de pruebas se destaca la implementación de dicha librería debido que preserva caracteres especiales. Pero el uso de servidores de terceros podría ser sensible en cuestiones de la seguridad en el manejo de datos personales.

En la parte de sincronización al utilizar un estándar de 150 palabras por minutos y una relación de 7.5 ms por sílaba, con un espacio entre las palabras de 8 ms de espera, se logró mostrar el material en forma lineal su narrativa obteniendo los recursos del repositorio de imágenes.

La mayoría de las pruebas de procesamiento se han realizado con cuentos cortos y fábulas que no contienen material sensible y que sus entidades nombradas no son similares entre sí. Logrando observar en el proceso de clasificación de cada texto, la generación de los arreglos de datos de palabras para la asignación de llamado a recursos multimedia, y el cálculo de los tiempos por intervención en la narrativa de cada palabra con la relación de tiempos antes mencionadas.

La relación mostrada entre la sincronización de los elementos y el audio por separado de cada narrativa demuestra que se han consolidado bien los tiempos y elementos entre el momento que se les llama y el momento que dejan de ser requeridos.

La librería SpaCy, permite reentrenar con nuevos elementos las identidades nombradas para mejorar los resultados de clasificación y su relación con otros elementos que preceden o anteceden las palabras como son los casos de adjetivos, abreviaciones o acrónimos. Este proceso por el momento debe ser realizado de manera manual debido a que, a diferencia de trabajar con documentos especializados en un solo tema con patrones o características comunes, el manejo de un corpus que cuyos documentos pueden tener un gran número de elementos distintos requiere reentrenamientos periódicos para ajustar en base a nuevas necesidades.

## VI. CONCLUSIONES Y FUTUROS TRABAJOS.

En el presente trabajo (actualmente en investigación) se plantea el desarrollo de una estrategia que permite la recolección de historias, el análisis de sus contenidos, la extracción de información y la sincronización con contenido multimedia almacenado en bases de datos.

La identificación y extracción con técnicas de procesamiento natural es susceptible a la interpretación en el proceso de generación de transcripciones del discurso hablado al escrito por la velocidad y modo de pronunciación.

El cálculo en los tiempos de aparición por palabra con una relación de 7.6 ms por sílaba permitió sincronizar de manera funcional los tokens con los elementos dentro del repositorio de imágenes mientras se desarrollaba la narrativa del texto a discurso oral a través del lenguaje Python.

Gracias a la realización de este trabajo se pudieron implementar y obtener resultados en el análisis de corpus con un conjunto de documentos que variaba los temas dentro de su contenido. Esto textos permitieron la sincronización automática de los documentos en el set de evaluación.

En trabajos futuros se plantea dos fases, siendo la primera la generación de repositorios con el material multimedia de la red cuya descarga sea de manera automática al no detectar elementos dentro del repositorio personal del usuario y del repositorio general de la aplicación. Dicho proceso se realizaría por medio de técnicas de extracción de datos de la red.

La segunda fase consistirá en implementar un analizador de imágenes por *Machine Learning* para asegurar la confiabilidad en el proceso de la generación de repositorios de manera automática con contenido de la red. Se plantea entrenar un modelo que se encargue de revisar que la imagen descargada se encuentre relacionada al token del contenido y de esta manera, evitar la descarga de imágenes incorrectamente etiquetadas en la red y preservar la relación del material multimedia con el nombre del token dentro de los repositorios.

## AGRADECIMIENTOS

J.E.Z.B. agradece al CONACYT el apoyo brindado para la realización de estudios de Maestría. Los autores agradecen al Tecnológico Nacional de México/ IT León las facilidades brindadas para la realización de este proyecto.

## REFERENCIAS

- [1] M. d. C. García Peña, L. M. Gutiérrez Robledo, P. A. Roa Rojas y A. Martínez Ruiz, «La Enfermedad de Alzheimer y otras demencias como problema nacional de salud.» Academia Nacional de Medicina de México (ANMM), México, 2017.
- [2] F. A. ESPAÑA, «<http://www.alzfae.org/>,» 13 04 2015. [En línea]. Available: <http://www.alzfae.org/actualidad/mundo-alzheimer/1158/30-congreso-annual-de-alzheimer-europe-la-demencia-en-un-mundo-cambiante-congreso-virtual-celebrado-entre-20-22-de-octubre-de-2020>. [Último acceso: 02 10 2022].
- [3] S. d. S. d. México, «<https://www.gob.mx/>,» Secretaría de Salud, 05 Octubre 21. [En línea]. Available: <https://www.gob.mx/salud/es/articulos/enfermedad-de-alzheimer-demencia-mas-comun-que-afecta-a-personas-adultas-mayores?idiom=es>. [Último acceso: 01 Octubre 22].
- [4] N. Justel, M. Psyrdellis y E. Ruetti, «Modulación de la memoria emocional: una revisión de los principales factores que afectan los recuerdos,» *Suma Psicológica*, vol. 20, n° 2, 2013.
- [5] E. Echeburúa y P. J. Amor, «Memoria traumática: estrategias de afrontamiento adaptativas e inadaptativas,» *Terapia psicológica*, vol. 37, n° 1, p. 2, 2019.
- [6] D. A. Muñoz González, «La estimulación cognitiva como estrategia para la atención psicogerontológica a los adultos mayores con demencia,» *Revista Cubana de Salud Pública*, vol. 44, n° 3, 2018.
- [7] C. d. I. p. D. e. S. Pública, «Instituto Nacional de Salud Pública,» Instituto Nacional de Salud Pública, 21 Septiembre 2020. [En línea]. Available: <https://www.insp.mx/avisos/hablemos-demencia#sup3>. [Último acceso: 04 Octubre 2022].
- [8] T. Álvarez Cisneros, S. Torres Castro, B. Mena Montes y . N. M. Torres Carrillo, «Factores de protección para la demencia en adultos mayores en México: ¿dónde estamos? En el libro: Gerontología para la vida activa.,» *Género y salud en cifras*, vol. 15, n° 3, p. 13, 2016.
- [9] M. d. L. Reyna Carrizales, M. V. González Rubio, F. J. López Esqueda y G. R. González González, «Demencias, Una visión panorámica,» de *V Simposio de Medicina Geriátrica*, San Luis Potosí, 2014.
- [10] A. Taylor, B. Santorini y M. Marcus, «The Penn Treebank: An overview,» Springer, Dordrecht, 2003.
- [11] D. Jurafsky y . J. H. Martin, *Hidden Markov Models*, Palo Alto: Stanford, 2021.
- [12] R. Cabeza Ruiz, «Aplicando los Campos Aleatorios Condicionales en la Segmentación de Textos por Idiomas,» de *Conditional Random Fields in Text Segmentation by Language*, Holguin, 2019.
- [13] S. K. Siencnik, «Adapting word2vec to named entity recognition,» *Proceedings of the 20th nordic conference of computational linguistics*, vol. 20, p. 239–243., 2015.
- [14] Google, «flutter.dev,» Google, 01 Octubre 2022. [En línea]. Available: <https://flutter.dev/>. [Último acceso: 01 octubre 2022].
- [15] A. M. M. S. B. Taylor, «The Penn Treebank: An Overview,» *Springer, Dordrecht*, Vols. %1 de %2In: Abeillé, A. (eds) Treebanks. Text, Speech and Language Technology, vol 20. , n° 20, p. 18, 2003.
- [16] G. Wolf, E. Ruiz, F. Bergero y E. Meza, FUNDAMENTOS DE SISTEMAS OPERATIVOS, Ciudad de México: UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, 2015.
- [17] Y. Goldberg, *Neural Network Methods for Natural Language Processing*, Toronto: Graeme Hirst, University of Toronto, 2017.
- [18] Spacy, «Spacy.io,» 01 Octubre 2022. [En línea]. Available: <https://spacy.io/usage/spacy-101>. [Último acceso: 02 Octubre 2022].
- [19] mathworks, «mathworks,» 01 Octubre 2022. [En línea]. Available: <https://la.mathworks.com/discovery/ngram.html>. [Último acceso: 01 Octubre 2022].
- [20] R. A. Española, «Real Academia Española,» 01 Octubre 2022. [En línea]. Available: <https://www.rae.es>. [Último acceso: 01 Octubre 2022].
- [21] R. Mitchell, *Web Scraping with Python Collecting. Data from the Modern Web*, Sebastopol: 2015, Reilly Media.