

---

# Chatbot para la recuperación de información personal

---



Trabajo de Fin de Grado  
Curso 2021–2022

Autora  
Lucía Latorre Magaz

Directores  
Gonzalo Méndez Pozo  
Pablo Gervás Gómez-Navarro

Grado en Ingeniería Informática  
Facultad de Informática  
Universidad Complutense de Madrid



# Chatbot para la recuperación de información personal

Trabajo de Fin de Grado en Ingeniería Informática  
Departamento de Ingeniería del Software e Inteligencia  
Artificial

**Autora**  
Lucía Latorre Magaz

**Directores**  
Gonzalo Méndez Pozo  
Pablo Gervás Gómez-Navarro

**Convocatoria:** *Junio 2022*  
**Calificación:** *Nota*

Grado en Ingeniería Informática  
Facultad de Informática  
Universidad Complutense de Madrid

14 de junio de 2022



# Dedicatoria

Texto de la dedicatoria...



# Agradecimientos

Texto de los agradecimientos





# Resumen

Resumen en español del trabajo

## **Palabras clave**

Máximo 10 palabras clave separadas por comas chatbot, reminiscencia, demencia, recuerdo



# Abstract

Abstract in English.

## **Keywords**

10 keywords max., separated by commas.



# Índice

<b>1. Introduction</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
1.1. Motivación . . . . .	3
1.2. Objetivos . . . . .	3
1.3. Plan de trabajo . . . . .	4
<b>2. Estado de la Cuestión</b>	<b>5</b>
2.1. Demencia . . . . .	5
2.2. Terapia ocupacional . . . . .	5
2.2.1. Terapia ocupacional basada en reminiscencia . . . . .	5
2.3. Proyecto CANTOR . . . . .	6
2.4. TFGs 2021 . . . . .	6
2.5. Chatbots . . . . .	6
2.5.1. Plataformas de desarrollo . . . . .	6
2.6. PLN . . . . .	6
<b>3. Arquitectura</b>	<b>7</b>
3.1. Prototipo de análisis de texto . . . . .	7
3.2. Preguntas basadas en la respuesta . . . . .	8
<b>4. Conclusiones y Trabajo Futuro</b>	<b>9</b>
<b>4. Conclusions and Future Work</b>	<b>11</b>
<b>Bibliografía</b>	<b>13</b>



# Índice de figuras





# Índice de tablas



# Chapter 1

## Introduction

Introduction to the subject area.



# Capítulo 1

## Introducción

*“Frase célebre dicha por alguien inteligente”*

— Autor

### Introducción temporal

Las personas con Alzheimer u otros tipos de demencia pueden beneficiarse del uso de la llamada terapia basada en reminiscencia, que se basa en la construcción de un libro de vida del paciente que recopila recuerdos positivos de su vida que se pueden utilizar posteriormente par ejercitar su memoria y retrasar el deterioro, además de permitir aumentar el bienestar de los pacientes.

En el presente proyecto se propone el desarrollo de un chatbot que permita recopilar y estructurar esta información para ayudar a los terapeutas en la elaboración de los libros de vida.

### 1.1. Motivación

Introducción al tema del TFG.

### 1.2. Objetivos

Este proyecto tiene como objetivo desarrollar un chatbot mediante el cual recabar información personal sobre la vida del paciente con demencia, clasificarla y estructurarla siguiendo un esquema que pueda facilitar la tarea de los terapeutas a la hora de construir un libro de vida.

La clasificación de recuerdos se hará en base a unos criterios predefinidos por expertos en terapia ocupacional del proyecto CANTOR. Se clasificará en función de:

- **Emoción:** Se clasificarán los recuerdos en positivos y negativos, siendo estos últimos para identificar qué recuerdos no deben tratarse en las

terapias por afligir al paciente. Los positivos se puntuarán de 0 a 10 en función de la felicidad que le traen al paciente.

- **Etapas:** Los recuerdos pertenecerán a una de las siguientes etapas: infancia, adolescencia, edad adulta o tercera edad según el periodo temporal en que aconteció.
- **Categorías:** Cada recuerdo entrará dentro de una o varias categorías que recojan una característica del recuerdo. Ejemplos de categorías: guerra civil, bailes, ocio, familia, aficiones...

### 1.3. Plan de trabajo

El plan de trabajo ha consistido de tres etapas:

- Investigación y construcción de prototipo en la que se ha consolidado la idea del TFG, investigado sobre demencia y terapia ocupacional basada en reminiscencia, elegido tecnologías y creado un prototipo de análisis de texto usando spaCy para identificar si un recuerdo es positivo o negativo.
- Programación del Chatbot y desarrollo de la memoria
- Pruebas, revisión de la memoria y entrega

# Capítulo 2

## Estado de la Cuestión

### 2.1. Demencia

La demencia (dem) es una condición neurodegenerativa con tendencia a crecer, caracterizada por un deterioro cognitivo que interfiere con la vida cotidiana afectando a la memoria, al pensamiento, al lenguaje, al juicio y al comportamiento. La demencia no es una enfermedad específica aunque la mayor parte de los casos de demencia son provocados por la enfermedad de Alzheimer. Muchas veces se confunde la demencia con una consecuencia más del envejecimiento cuando la senilidad no implica sufrir misma.

Hay muchos síntomas asociados a la demencia pero, en este trabajo, nos vamos a centrar en la pérdida de memoria. Trabajar los recuerdos de una persona que sufre demencia ayuda a retrasar los efectos de la demencia. Hablaremos para ello de la terapia ocupacional basada en reminiscencia.

### 2.2. Terapia ocupacional

La terapia ocupacional (ter) se centra en que el paciente sea capaz de participar en las actividades de la vida cotidiana. Es decir, se basa en ayudar al individuo a llevar una vida lo más normal posible adaptando las tareas cotidianas a realizar o el entorno para que pueda llevarlas a cabo.

#### 2.2.1. Terapia ocupacional basada en reminiscencia

Sacar información del seminario cantor La terapia ocupacional basada en reminiscencia se centra en mejorar la calidad de vida de la persona con demencia. Se trata de una técnica basada en la recuperación de recuerdos dentro de un periodo de tiempo en la vida de la persona con el objetivo de construir la historia de vida del sujeto. La historia de vida surge de la

sucesión de acontecimientos que componen la totalidad de la vivencia del sujeto.

### **2.3. Proyecto CANTOR**

(can)

### **2.4. TFGs 2021**

Herramienta de ayuda guiada para la reminiscencia (rem) : Generación de historias a partir de una base de conocimiento: recomendación de temas a tratar en la terapia + aplicación web que enlaza situaciones y vivencias mediante grafos y luego permite añadir recursos fotográficos asociado a un tema. (Más como un chatbot que va sugiriendo temas a tratar)

Sistema de asistencia para cuidados de enfermos del Alzheimer (asi) : Página que guarda información sobre pacientes y terapeutas asociados. Información relevante + historia de vida formada por instancias de recuerdos

Extracción de preguntas a partir de imágenes para personas con problemas de memoria mediante técnicas de Deep Learning (pre) : chat desplegado con telegram. De las fotos que tiene archivadas va preguntando al usuario cosas relacionadas con la imagen

Generación de resúmenes de video-entrevistas utilizando redes neuronales (res) : transcripción de video-entrevistas a texto

Extracción de información personal a partir de redes sociales para la creación de un libro de vida (rrs)

### **2.5. Chatbots**

#### **2.5.1. Plataformas de desarrollo**

### **2.6. PLN**

El procesamiento de lenguaje natural



# Capítulo 3

## Arquitectura

### 3.1. Prototipo de análisis de texto

Se ha probado una forma de clasificar texto utilizando TextCategorizer (cat) de la librería spacy de python. Se trata de entrenar un modelo para saber identificar si un recuerdo es positivo o negativo. Se ha utilizado el código del artículo pro para probar cómo se podría entrenar el modelo para que supiese diferenciar los recuerdos negativos de los positivos y así tener la primera clasificación que pidieron los terapeutas especializados del proyecto CANTOR. Se ha probado usando una batería de recuerdos positivos encontrados entre los archivos del TFG de Laura Castillo (rem) y negativos inventados. Antes de probar con recuerdos se probó con un dataset mucho más grande de reseñas de moda que se ponía como ejemplo en pro.

El prototipo funciona siguiendo una serie de pasos. En primer lugar, se añade el componente TextCategorizer (textcat) a un modelo en blanco del idioma español. Un modelo en blanco es un modelo que no tiene ningún componente de spaCy definido (como serían NER,), es decir, el texto que se almacena en los documentos de spaCy no es analizado por ninguna cadena de procesos porque la tubería de componentes está vacía. Si hubiésemos cogido un modelo pre-entrenado como "es\_core\_news\_sm", el clasificador de textos se sumaría al trabajo previo de los procesos que analizan el texto como serían ['tok2vec', 'morphologizer', 'parser', 'attribute\_ruler', 'lemmatizer', 'ner']. Por ejemplo, el componente ner, más reconocido como Named Entity Recognizer, reconoce entidades dentro del texto como nombres de personas, fechas o lugares. Al usar el modelo en blanco empezamos con el español de cero, sin analizar. Tras añadir el componente textcat, al crear un documento de spaCy, automáticamente, pasará por el proceso de clasificación que le indiquemos. Para que textcat funcione como categorizador de recuerdos, se le añade a nuestro nuevo componente dos etiquetas, NEGATIVO y POSITIVO, que definirán cuánto de negativo es un recuerdo y

cuánto de positivo. Para seguir configurándolo, el prototipo coge el texto que usaremos para entrenar nuestro modelo y lo prepara adecuándolo al formato que entiende el clasificador, siendo éste una lista de tuplas (texto, etiqueta). Además, cogerá un porcentaje de esta batería de recuerdos (% que previamente hemos definido) y lo reservará para la evaluación de las predicciones, es decir, para analizar cómo de bien predice nuestro modelo tras entrenarlo.

Ya tenemos casi todo preparado para comenzar a entrenar el modelo, se dividen los casos de entrenamiento en lotes que se analizarán y evaluarán un número definido de iteraciones para asegurar que el entrenamiento es lo más preciso posible sin pasarnos de vueltas para que sea óptimo. Se trata de un proceso iterativo en el que las predicciones del modelo se comparan con las etiquetas de referencia para estimar el gradiente de la pérdida. El modelo se entrena utilizando una función que lo analiza y actualiza en cada iteración, la función `update()`. También se comprueban las predicciones del modelo en cada vuelta, se comparan con las etiquetas de referencia para estimar la desviación de la pérdida.

Una vez afinado el modelo mediante el entreno previo, ya podemos ponerlo a prueba. El prototipo sacará la probabilidad de que un texto procesado por `spacy` sea un recuerdo positivo y la probabilidad de que sea uno negativo.

### 3.2. Preguntas basadas en la respuesta

Esta sección consiste en conseguir que las preguntas que se hagan al interrogado tengan sentido con el resto de la conversación previa. Conseguir que sean lo más inteligentes y adecuadas posible.

Para ello vamos a utilizar el analizador de texto `Spacy` utilizando el código del TFG de Laura Castilla Castellano (Generación de historias a partir de una base de conocimiento), en concreto el código que aparece en `analysis.py` que contiene funciones para el procesamiento y síntesis de los textos.

Para encontrar la pregunta más adecuada primero se han eliminado las preguntas que ya habían sido contestadas para no repetirlas. Después con las funciones del TFG de Laura, de sintetizar y categorizar, se han desmembrado tanto las posibles preguntas como la respuesta anterior que el interrogado ha dado. Una vez desmembradas en lemmas, se han comparado los lemmas de ambas frases, tanto respuesta como posible pregunta. Dentro de la lista de posibles preguntas, la que más lemmas comunes tenga con la respuesta, gana como siguiente pregunta a formular.

# Capítulo 4

## Conclusiones y Trabajo Futuro

Conclusiones del trabajo y líneas de trabajo futuro.



# Chapter 4

## Conclusions and Future Work

Conclusions and future lines of work.



# Bibliografía

*Y así, del mucho leer y del poco dormir,  
se le secó el cerebro de manera que vino  
a perder el juicio.*

Miguel de Cervantes Saavedra

*Concesión del Proyecto CANTOR.* Versión electrónica, Disponible en <https://www.ucm.es/itc/noticias/40943>.

*Extracción de información personal a partir de redes sociales para la creación de un libro de vida.* Versión electrónica, Disponible en <https://eprints.ucm.es/id/eprint/68328/>.

*Extracción de preguntas a partir de imágenes para personas con problemas de memoria mediante técnicas de Deep Learning.* Versión electrónica, Disponible en <https://eprints.ucm.es/id/eprint/66857/>.

*Generación de resúmenes de video-entrevistas utilizando redes neuronales.* Versión electrónica, Disponible en <https://eprints.ucm.es/id/eprint/68333/>.

*Herramienta de ayuda guiada para la reminiscencia.* Versión electrónica, Disponible en <https://eprints.ucm.es/id/eprint/68332/>.

*How to Train Text Classification Model in spaCy?*. Versión electrónica, Disponible en <https://www.machinelearningplus.com/nlp/custom-text-classification-spacy/>.

*¿Qué es la demencia?*. Versión electrónica, Disponible en <https://www.alz.org/alzheimer-demencia/que-es-la-demencia?lang=es-MX>.

*¿Qué es la terapia ocupacional.* Versión electrónica, Disponible en <https://aptoca.org/terapia-ocupacional/que-es-la-terapia-ocupacional-2/>.

*Sistema de asistencia para cuidados de enfermos del Alzheimer.* Versión electrónica, Disponible en <https://eprints.ucm.es/id/eprint/67069/>.

*TextCategorizer.* Versión electrónica, Disponible en <https://spacy.io/api/textcategorizer>.